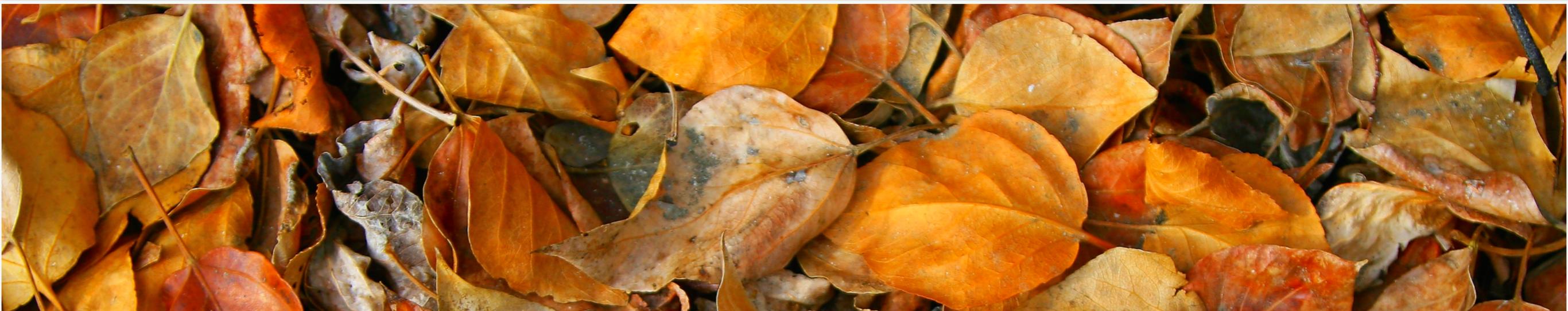
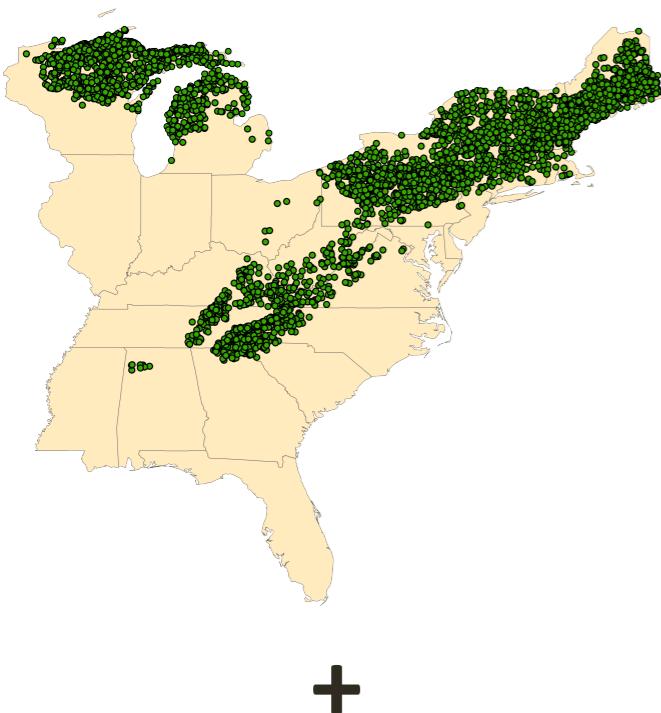


Spatial modeling of biodiversity composition



Species distribution modeling

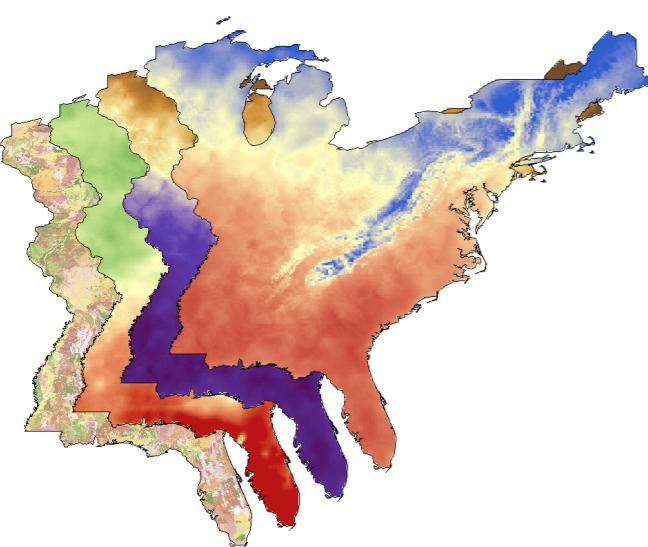
Tsuga canadensis



Biological
survey data

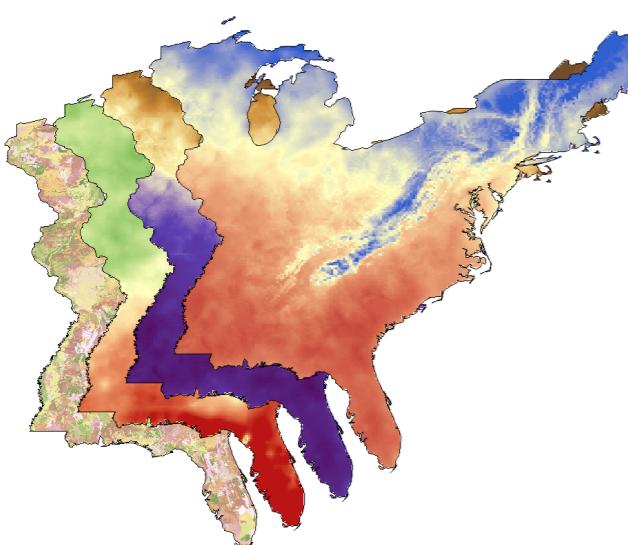
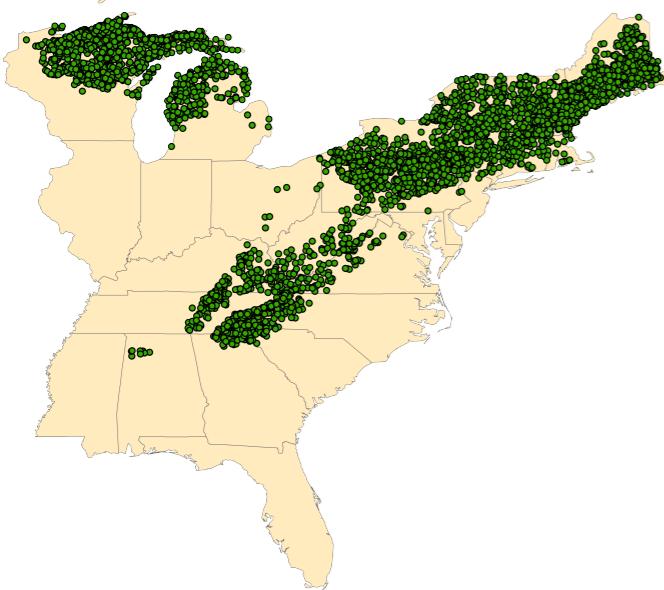
$$\begin{bmatrix} Spp_1 \\ Site_1 & 0 \\ Site_2 & 1 \\ Site_3 & 1 \\ \dots & \dots \\ Site_j & 1 \end{bmatrix} = f \left\{ \begin{bmatrix} Env_1 & Env_2 & Env_3 & \dots & Env_k \\ Site_1 & 23.4 & 545.5 & 0.64 & \dots & 4.1 \\ Site_2 & 22.1 & 89.0 & 0.22 & \dots & 8.0 \\ Site_3 & 24.9 & 439.5 & 0.61 & \dots & 3.4 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Site_j & 25.3 & 321.7 & 0.88 & \dots & 3.9 \end{bmatrix} \right\}$$

as a function of environmental covariates



Species distribution modeling

Tsuga canadensis



+

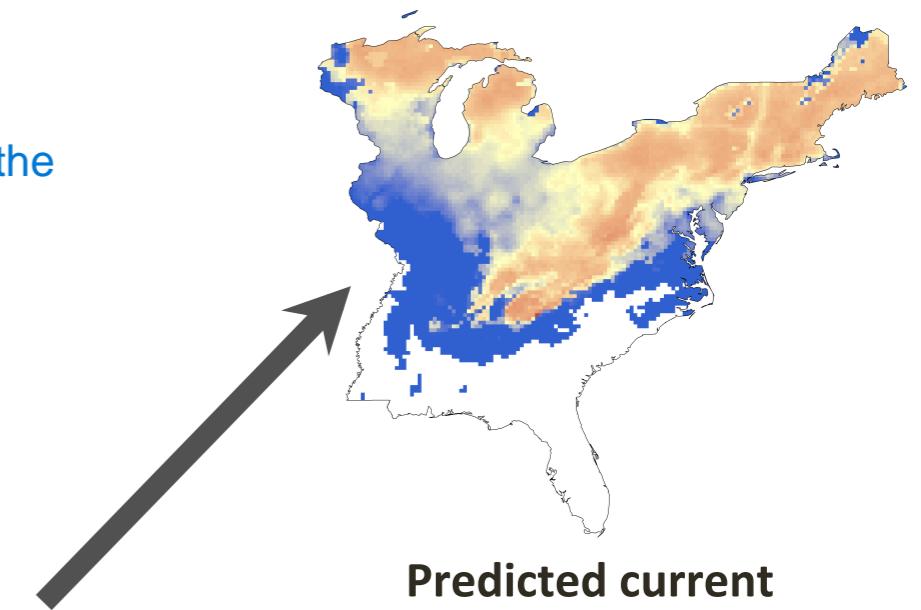


can also do this for more than one species by making a stack of the species response functions

Modeled species response functions

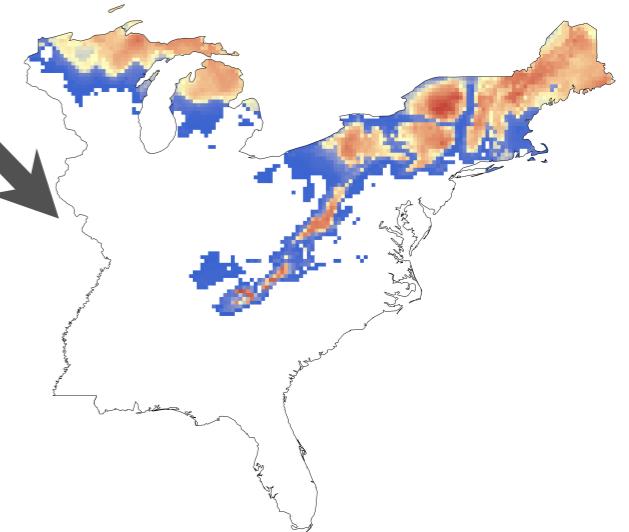


Species
Distribution
Model

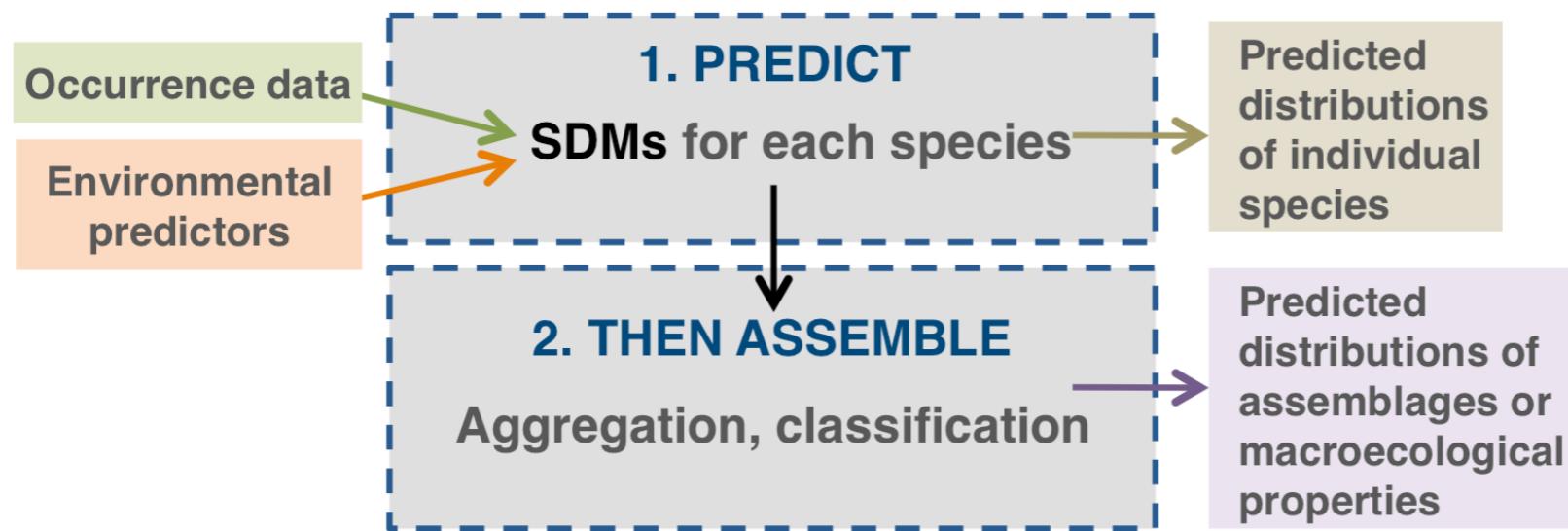


Predicted current range

Potential range under new conditions



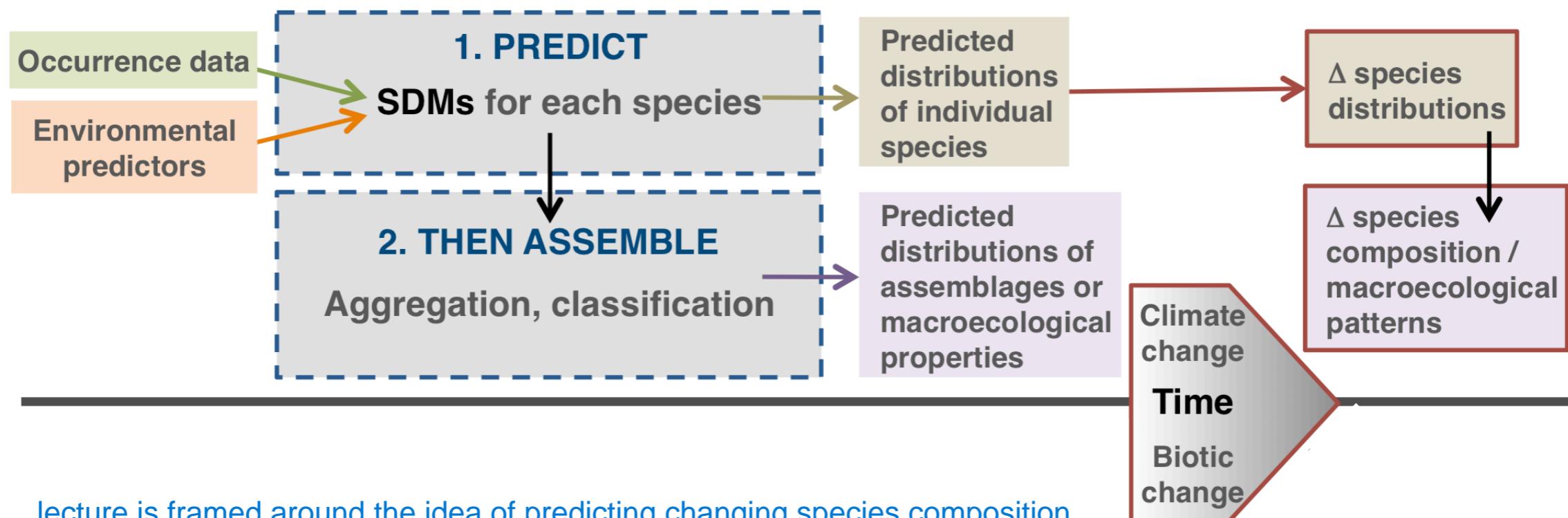
Strategies for modeling biodiversity



build the individual models in complete isolation and then assemble the models of the different organisms afterwards

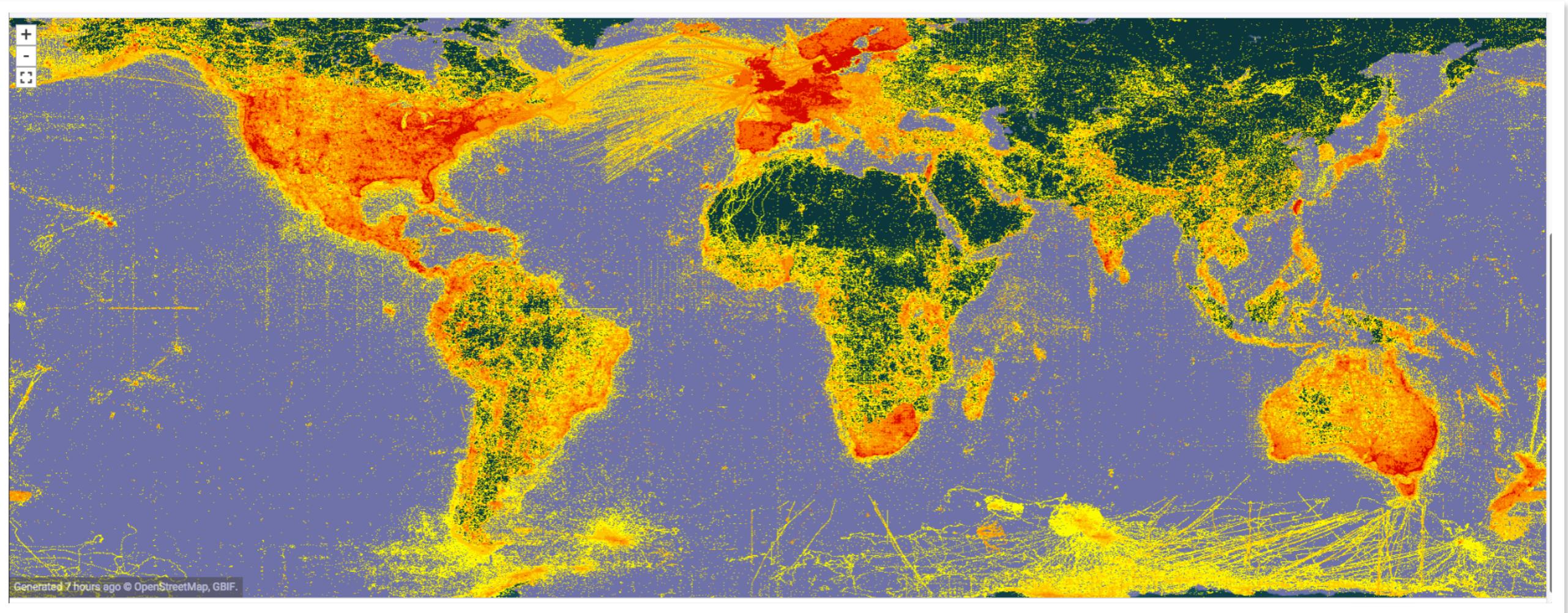
Predict, then assemble

Strategies for modeling biodiversity



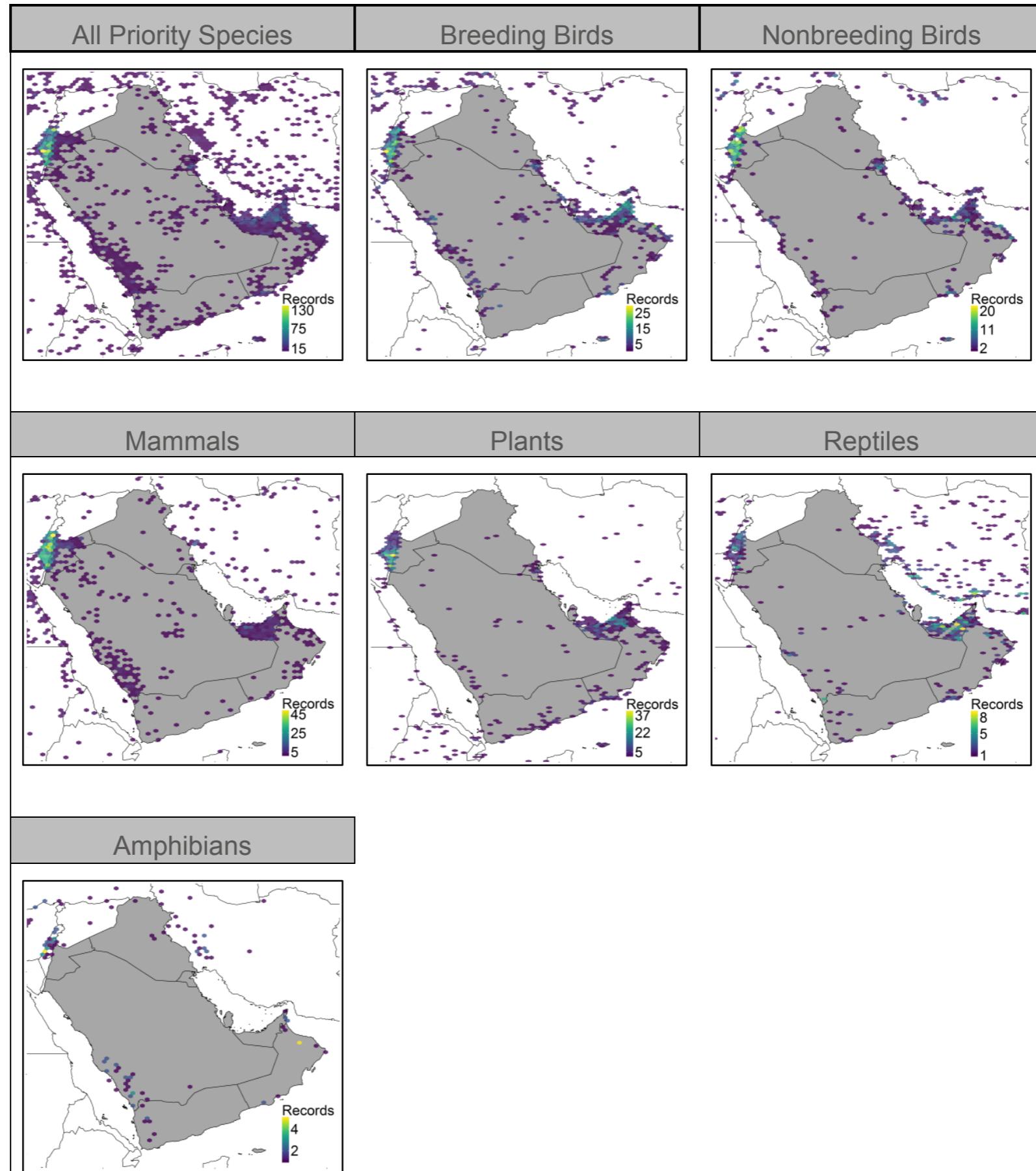
Predict, then assemble

this is the dominant approach (bc we have species-level data not community-level), but community-level approaches are becoming more common



Middle East:

it can be hard to model single species due to:
-sparsity (some areas with lots of records some with none)

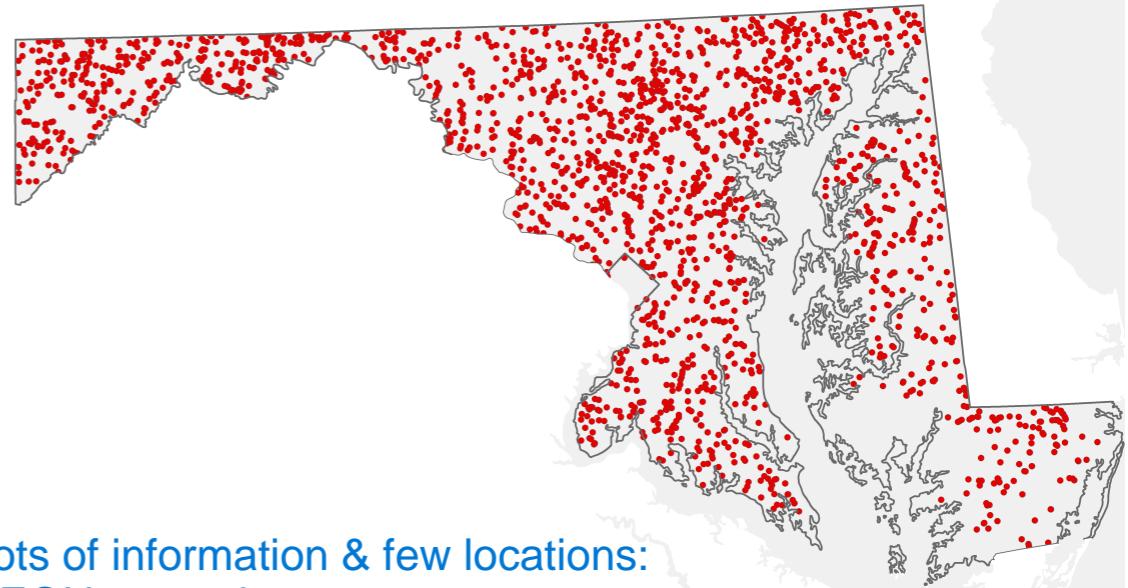


Large number of species, but few if any samples

tropics are hard bc so many species and so few samples >> misestimate biodiversity (under estimate)



(relatively) few locations, but data rich

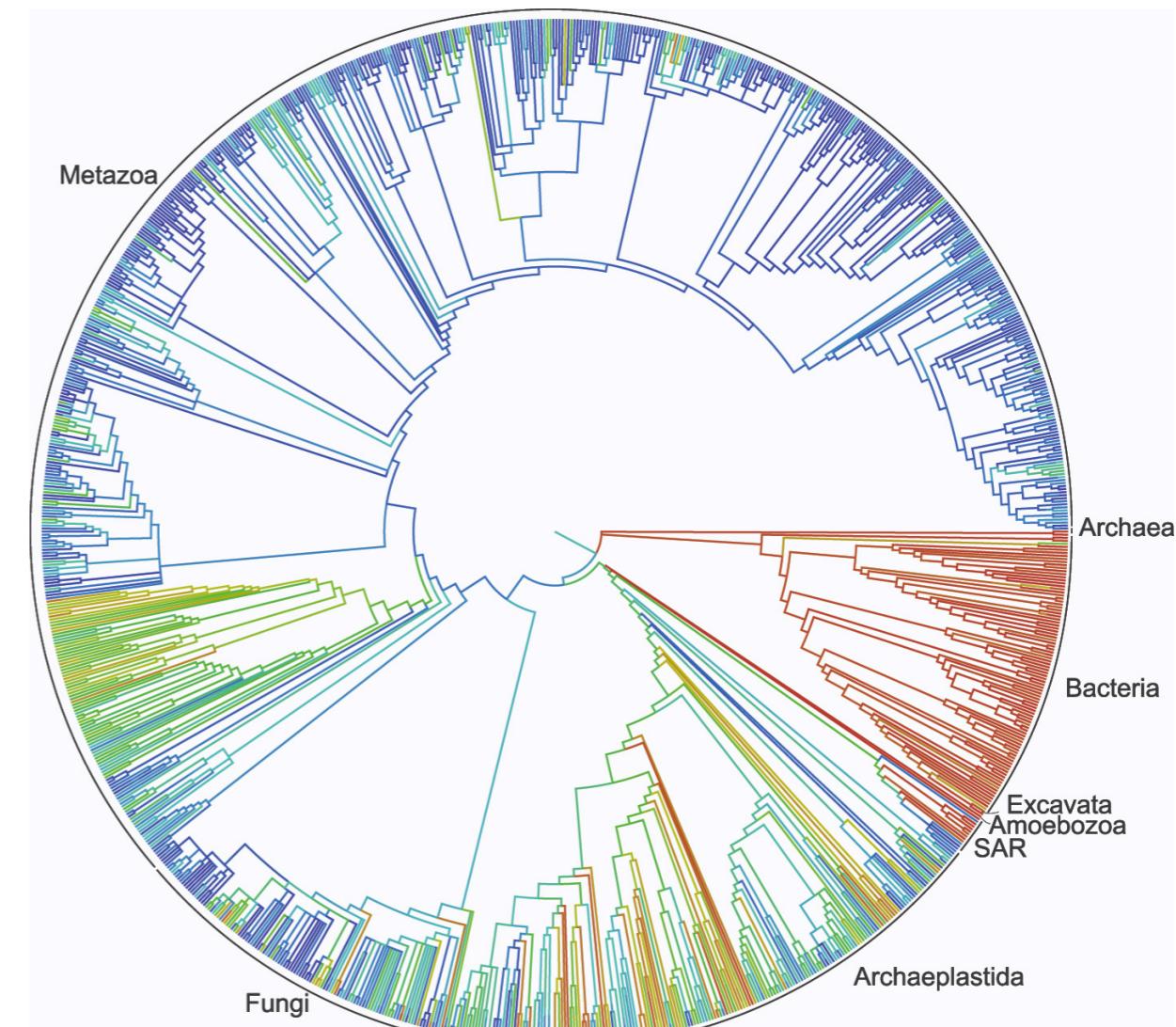
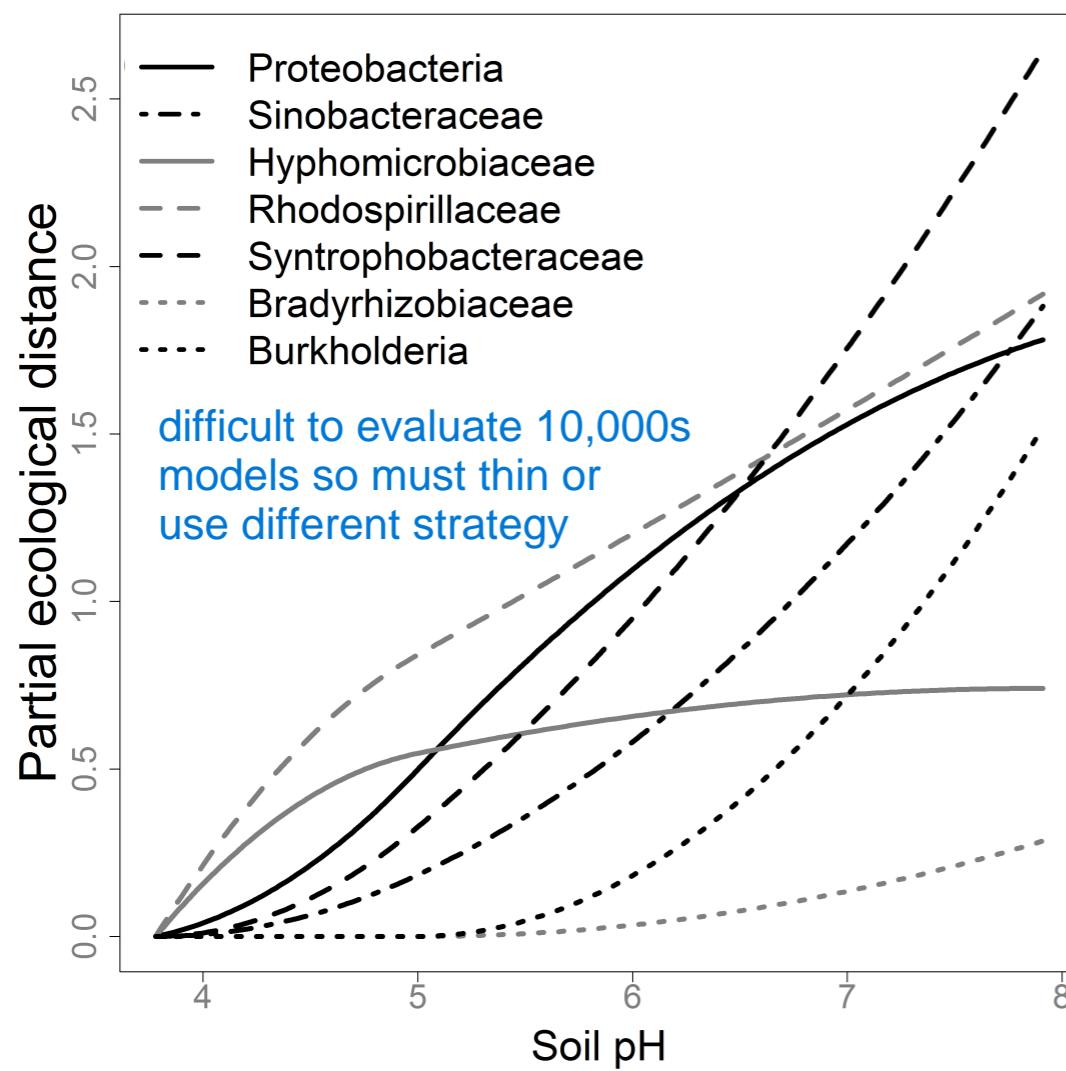


Lots of information & few locations:
NEON network
is very heavily sampled



Many locations & data rich

genomics, acoustics, eDNA all give tons of samples



Strategies for modeling biodiversity

beyond the species level

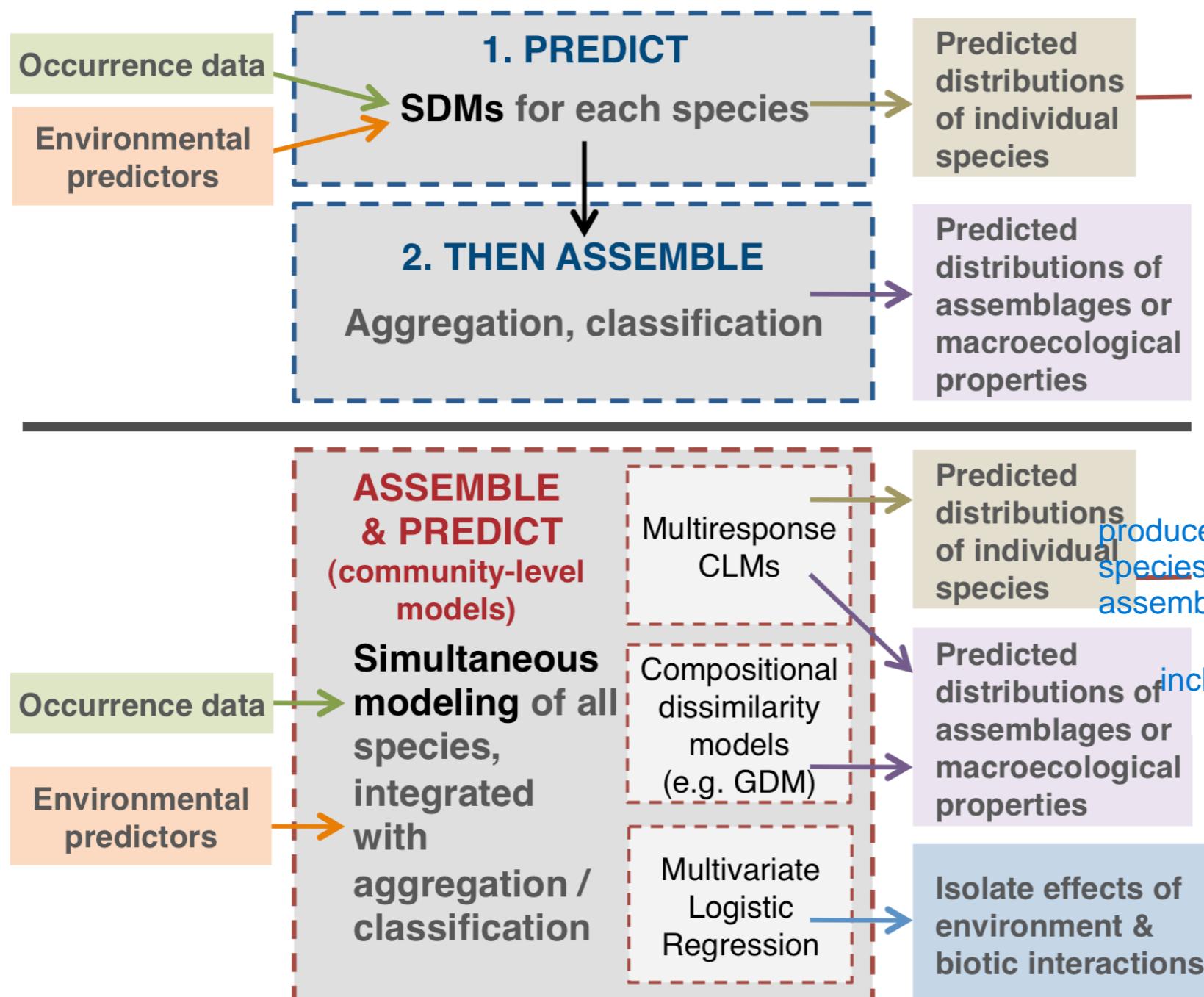
- Gleason vs. Clements: Competing theories regarding communities
 - **Gleasonian** - Communities are the result of interactions between individual species & the environment (abiotic and biotic factors) + chance historic events
 - Species have their own unique tolerances & therefore respond individualistically to environmental changes
 - “Communities” are arbitrary human constructs as a result of humans trying to make sense of things
 - **Clementsian** - Communities are highly organized “superorganisms” made up of mutually interdependent species
- neither one is correct -- nature is somewhere in the middle of the two ideas

for modeling species richness, community types, assume unchanged by environmental variables -- unlikely to get published

Assemble first, then predict

Strategies for modeling biodiversity

CLM = Community Level Models



much more likely to be published

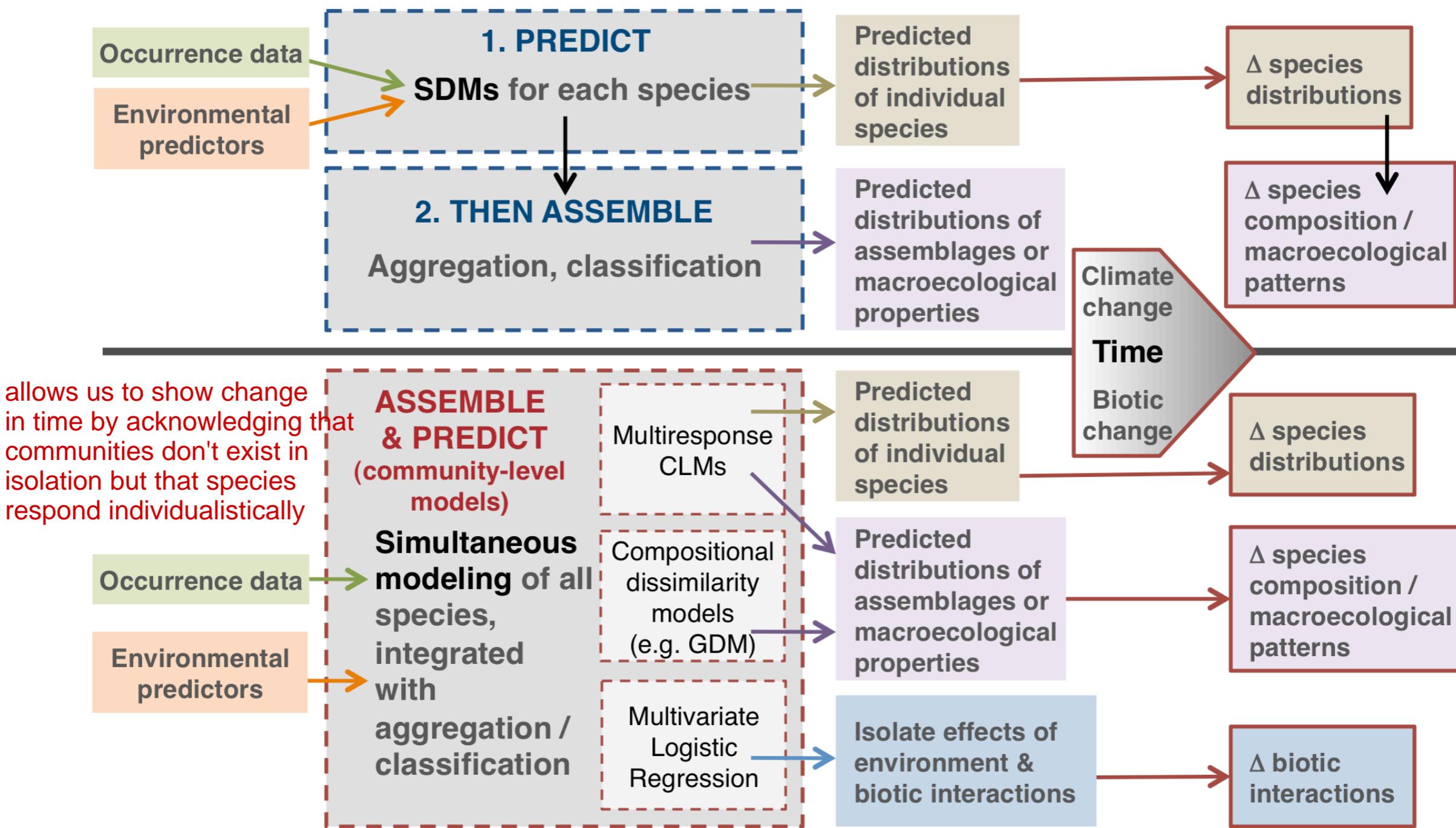
**Assemble &
predict together**

produce predicted distributions of individual species that are conditional upon other assemblages of species

including beta diversity

understanding what residual covariates between the species mean

Strategies for modeling *biodiversity*





$$\begin{bmatrix} Spp_1 \\ Site_1 & 0 \\ Site_2 & 1 \\ Site_3 & 1 \\ \dots & \dots \\ Site_j & 1 \end{bmatrix} = f \left\{ \begin{bmatrix} Env_1 & Env_2 & Env_3 & \dots & Env_k \\ Site_1 & 23.4 & 545.5 & 0.64 & \dots & 4.1 \\ Site_2 & 22.1 & 89.0 & 0.22 & \dots & 8.0 \\ Site_3 & 24.9 & 439.5 & 0.61 & \dots & 3.4 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Site_j & 25.3 & 321.7 & 0.88 & \dots & 3.9 \end{bmatrix} \right\}$$



$$\begin{bmatrix} Spp_1 & Spp_2 & Spp_3 & \dots & Spp_n \\ Site_1 & 0 & 1 & 0 & \dots & 0 \\ Site_2 & 1 & 1 & 1 & \dots & 0 \\ Site_3 & 1 & 0 & 0 & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Site_j & 1 & 0 & 0 & \dots & 1 \end{bmatrix} = f \left\{ \begin{bmatrix} Env_1 & Env_2 & Env_3 & \dots & Env_k \\ Site_1 & 23.4 & 545.5 & 0.64 & \dots & 4.1 \\ Site_2 & 22.1 & 89.0 & 0.22 & \dots & 8.0 \\ Site_3 & 24.9 & 439.5 & 0.61 & \dots & 3.4 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Site_j & 25.3 & 321.7 & 0.88 & \dots & 3.9 \end{bmatrix} \right\}$$

Community-level modeling



$$\left[\begin{array}{ccccc} & Spp_1 & Spp_2 & Spp_3 & \dots & Spp_n \\ Site_1 & 0 & 1 & 0 & \dots & 0 \\ Site_2 & 1 & 1 & 1 & \dots & 0 \\ Site_3 & 1 & 0 & 0 & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Site_j & 1 & 0 & 0 & \dots & 1 \end{array} \right] = f \left\{ \left[\begin{array}{ccccc} & Env_1 & Env_2 & Env_3 & \dots & Env_k \\ Site_1 & 23.4 & 545.5 & 0.64 & \dots & 4.1 \\ Site_2 & 22.1 & 89.0 & 0.22 & \dots & 8.0 \\ Site_3 & 24.9 & 439.5 & 0.61 & \dots & 3.4 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Site_j & 25.3 & 321.7 & 0.88 & \dots & 3.9 \end{array} \right] \right\}$$

↗

Species within an assemblage

every site must have a presence/absence or abundance measure (presence only has many assumptions)

Community-level modeling

Assemble & predict together

J-SDMs must consider the covariates between species

“Joint” Species distribution models (J-SDMs)

1. Multivariate regression response is multivariate
 - i. GLMs / Latent variable models
 - ii. GLMMs
 - iii. Bayesian hierarchical models
 - iv. Constrained gradient ordination
 - v. Multi-species occupancy models

Assemble & predict together

The image shows the cover of the journal *Ecology Letters*. The title "ECOLOGY LETTERS" is at the top in a large, serif font. Below it, the journal's name "Ecology Letters" is followed by the volume and page number "(2017) 20: 561–576". To the right is the digital object identifier "doi: 10.1111/ele.12757". On the left, a black box contains the text "IDEA AND PERSPECTIVE". The main title of the article is "How to make more out of community data? A conceptual framework and its implementation as models and software". The abstract section begins with "Community ecology aims to understand what factors determine the assembly and dynamics of species assemblages...". In the bottom right corner of the journal cover, there is a small box containing the text "Trends in Ecology & Evolution".

introductions to J-SDMs to read

Review

So Many Variables: Joint Modeling in Community Ecology

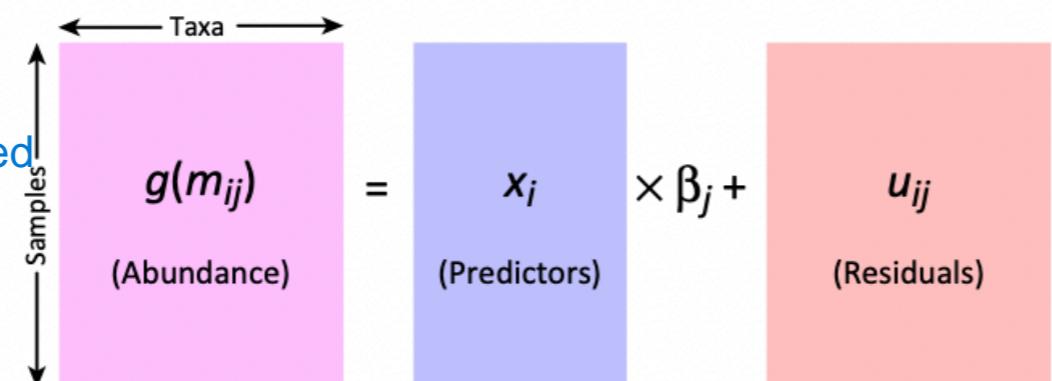
David I. Warton,^{1,*} F. Guillaume Blanchet,² Robert B. O'Hara,³ Otso Ovaskainen,^{4,5} Sara Taskinen,⁶ Steven C. Walker,² and Francis K.C. Hui⁷

Assemble & predict together

Key Figure

Schematic Diagram of Joint Models for Response of Abundance to Predictors, and Residual Correlation Across Taxa

(A) Multivariate generalised linear mixed model (GLMM)



pink = response
purple = coefficients fitted to the response terms
red=residuals

(B) Latent variable model (LVM)

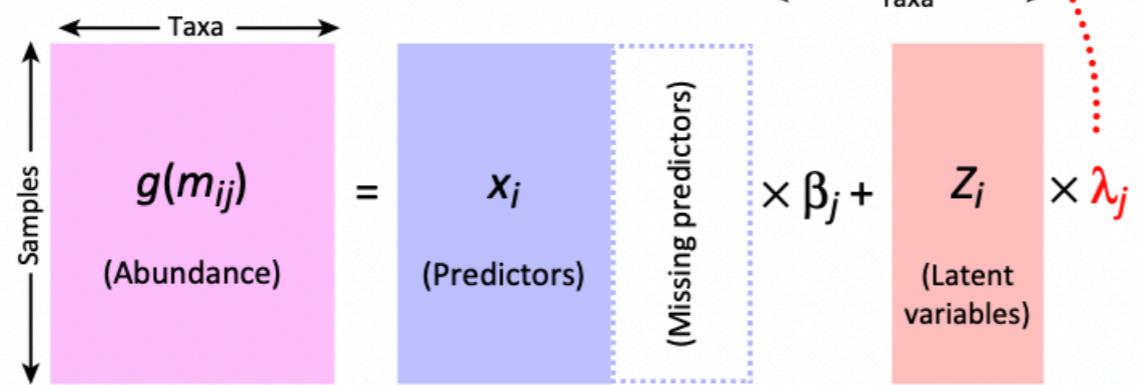
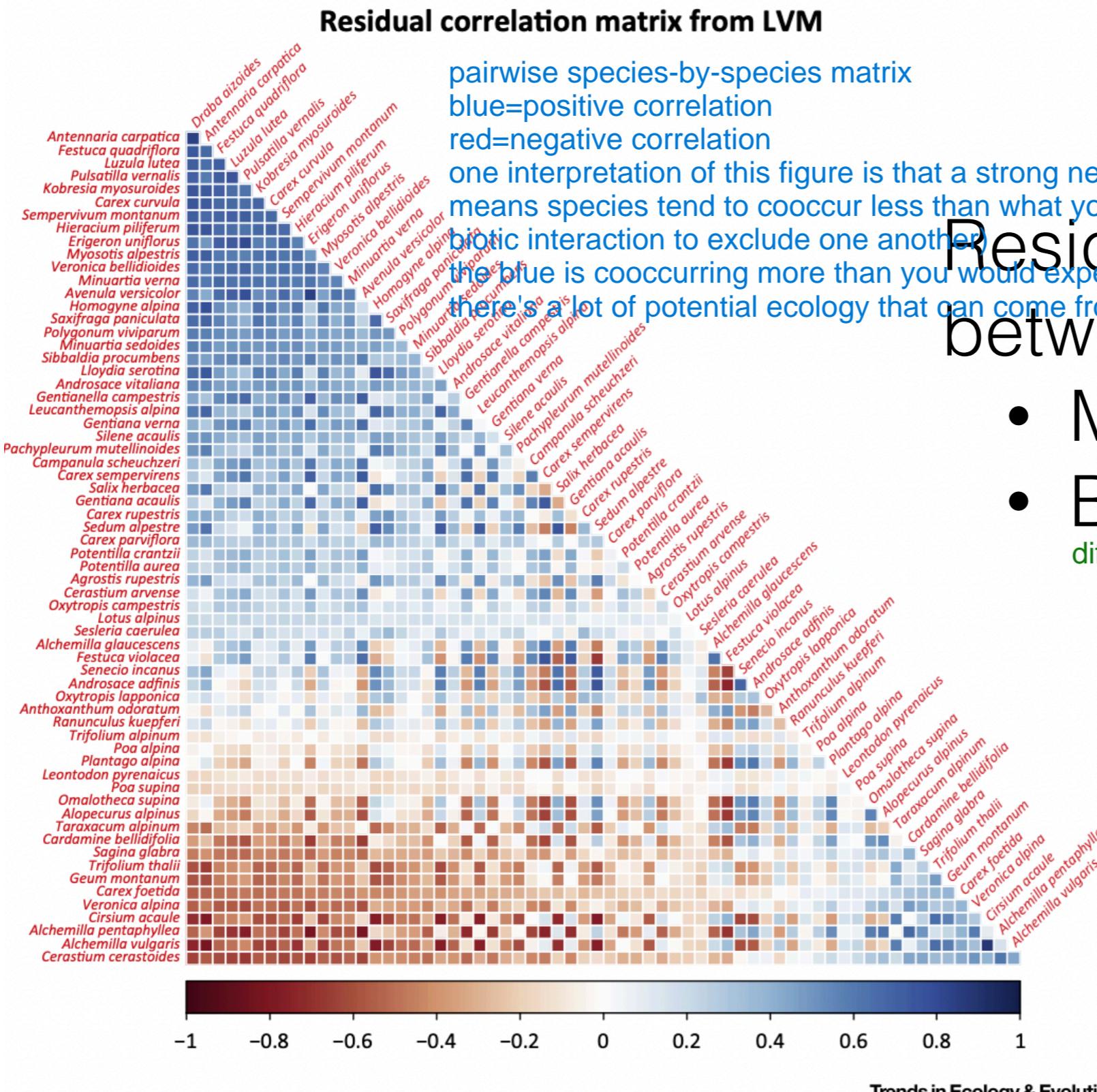


Figure 1. Correlation can be handled in different ways: (A) A multivariate GLMM uses correlated multivariate random effects, u_{ij} , to estimate correlation. (B) A LVM includes a smaller number of latent variables, z_i , which play the role of missing predictors. Their factor loadings, λ_j , approximate the correlation across taxa, but use fewer parameters than the GLMM (because the matrix of z_i has fewer columns than the matrix of u_{ij}).

residual correlation after fitting these models: any residual covariates between the species indicates either a missing covariate OR these remaining correlations are the signals of biotic interactions

Assemble & predict together



Residual correlations between species

- Missing covariates?
- Biotic interactions?

difficult to isolate these two conditions

Example paper:

"Do large-scale associations in birds imply biotic interactions or environmental filtering"

Check out, published 2022:

<https://doi.org/10.1111/jbi.14520>

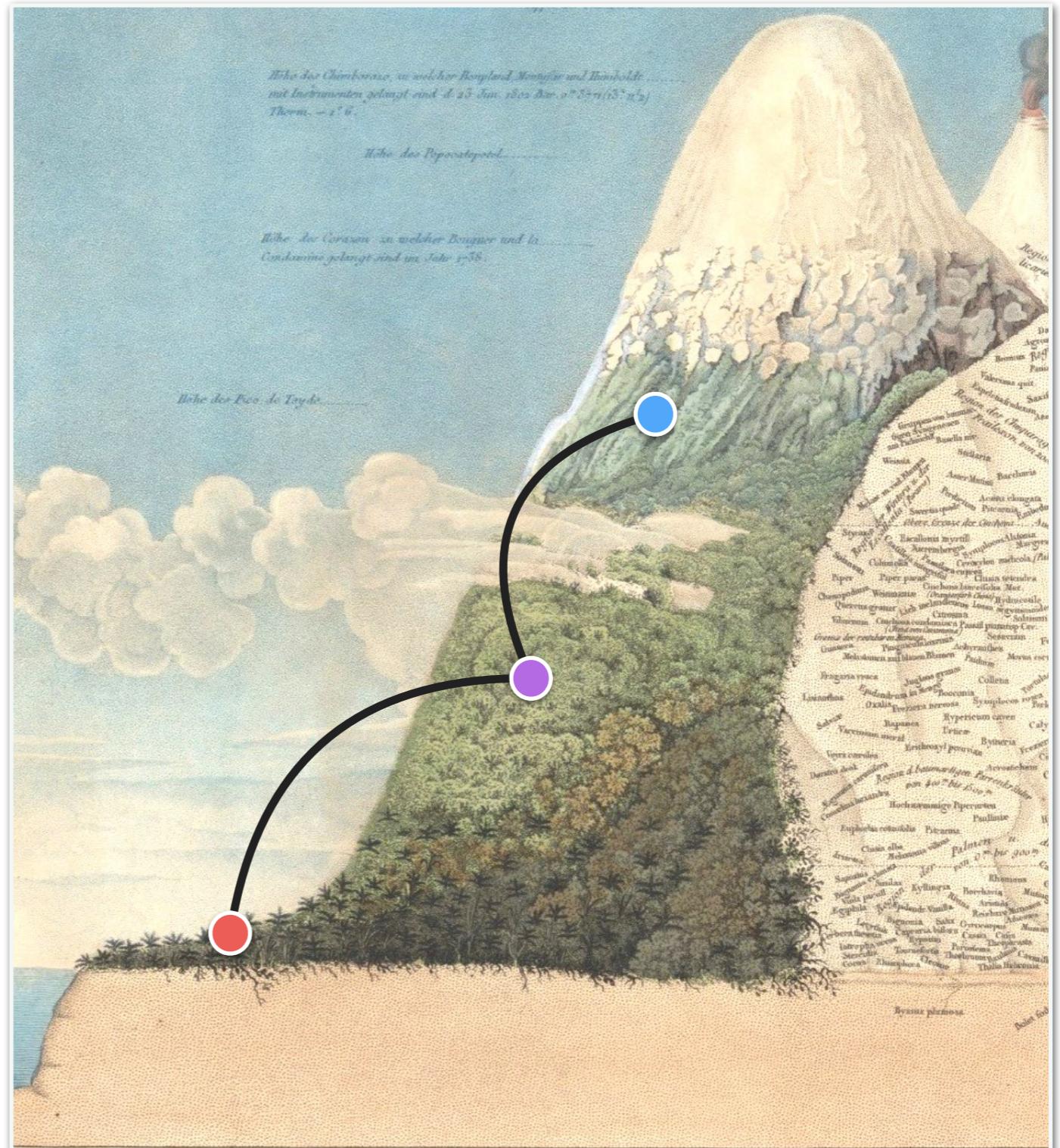
for R code, see:

https://ecology.ghislainv.fr/jSDM/reference/plot_residual_cor.html

Assemble & predict together

as you go from one habitat to another (ie can be thought of as space and time)

β -diversity: change
(turnover) in community
structure along spatial,
temporal or environmental



CLMs based on non-linear compositional turnover functions

1. Gradient Forests based on random forest
2. Generalized Dissimilarity Modeling based on generalized linear model

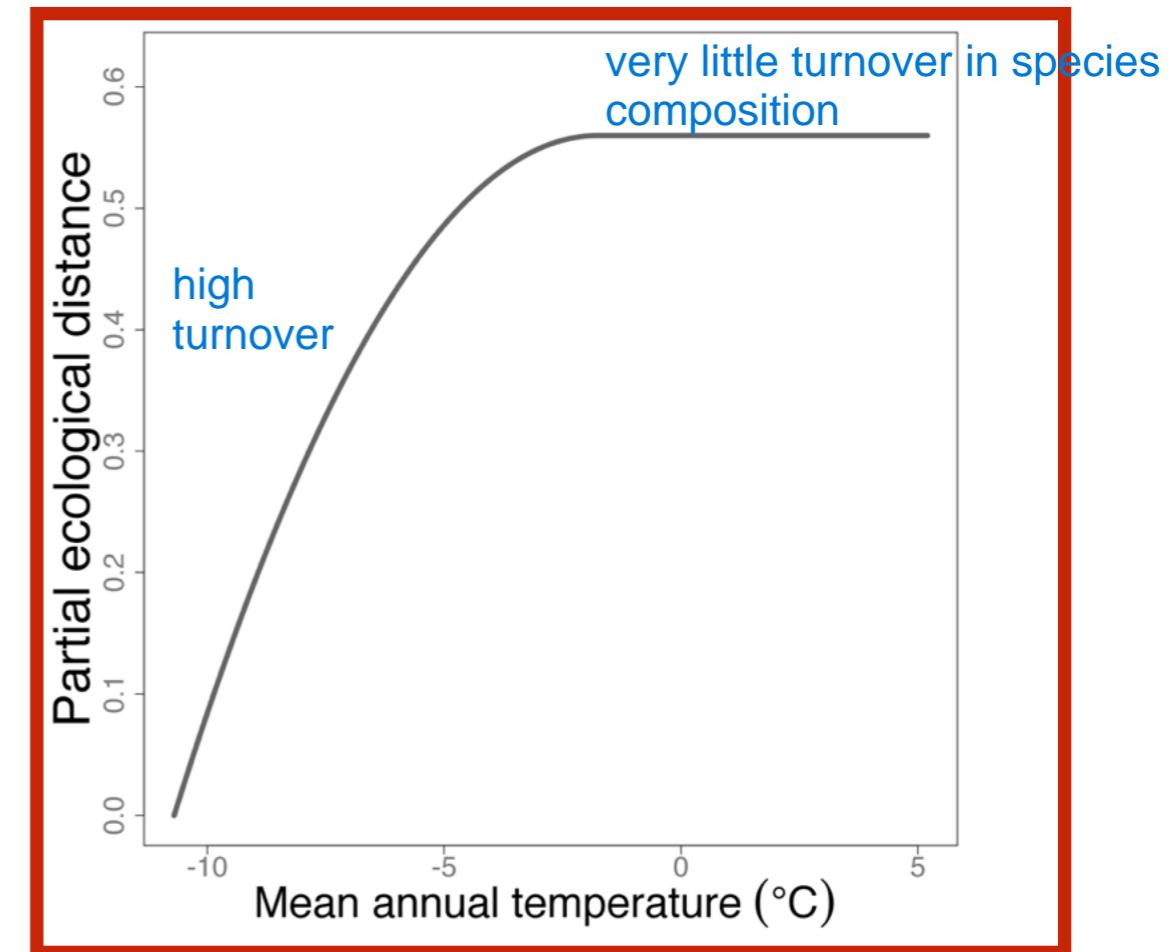
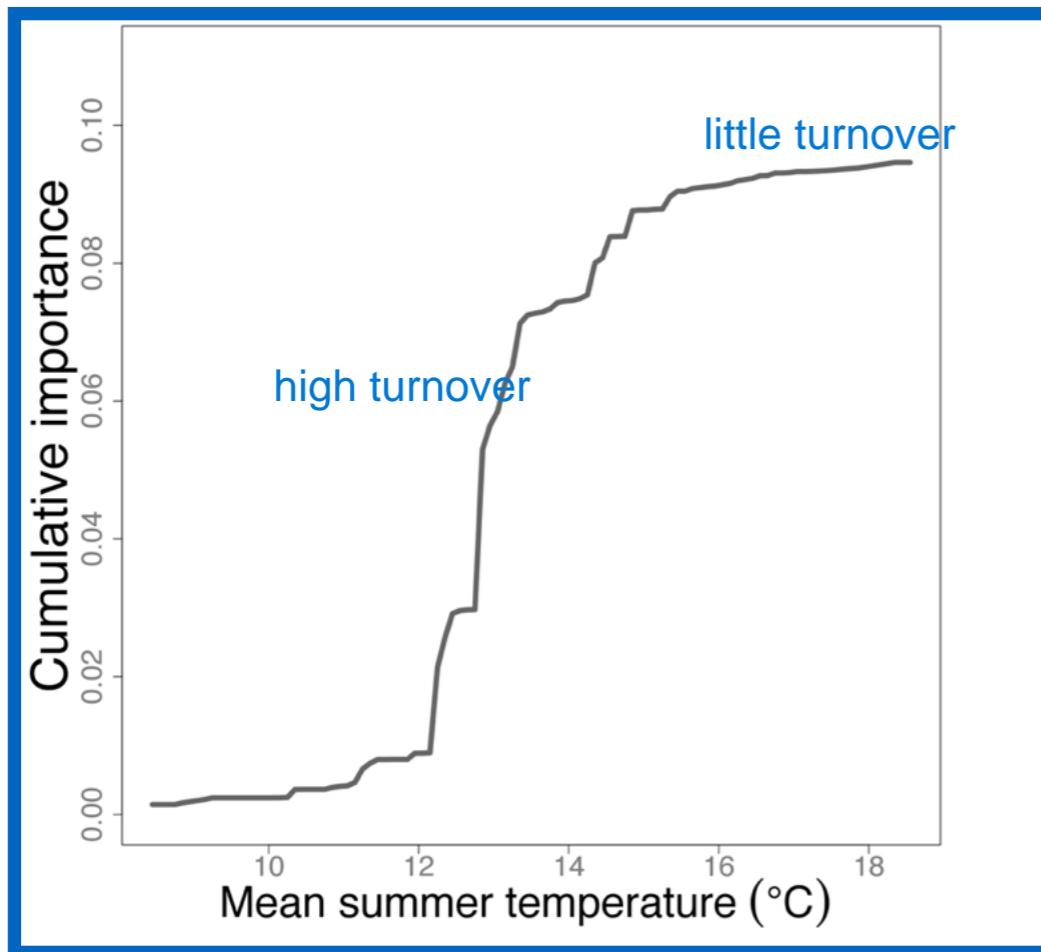
CLMs based on non-linear compositional turnover functions

1. Gradient Forests

shape = rate of biological change along this gradient

height = relative importance of that variable by comparison to other variables in the model

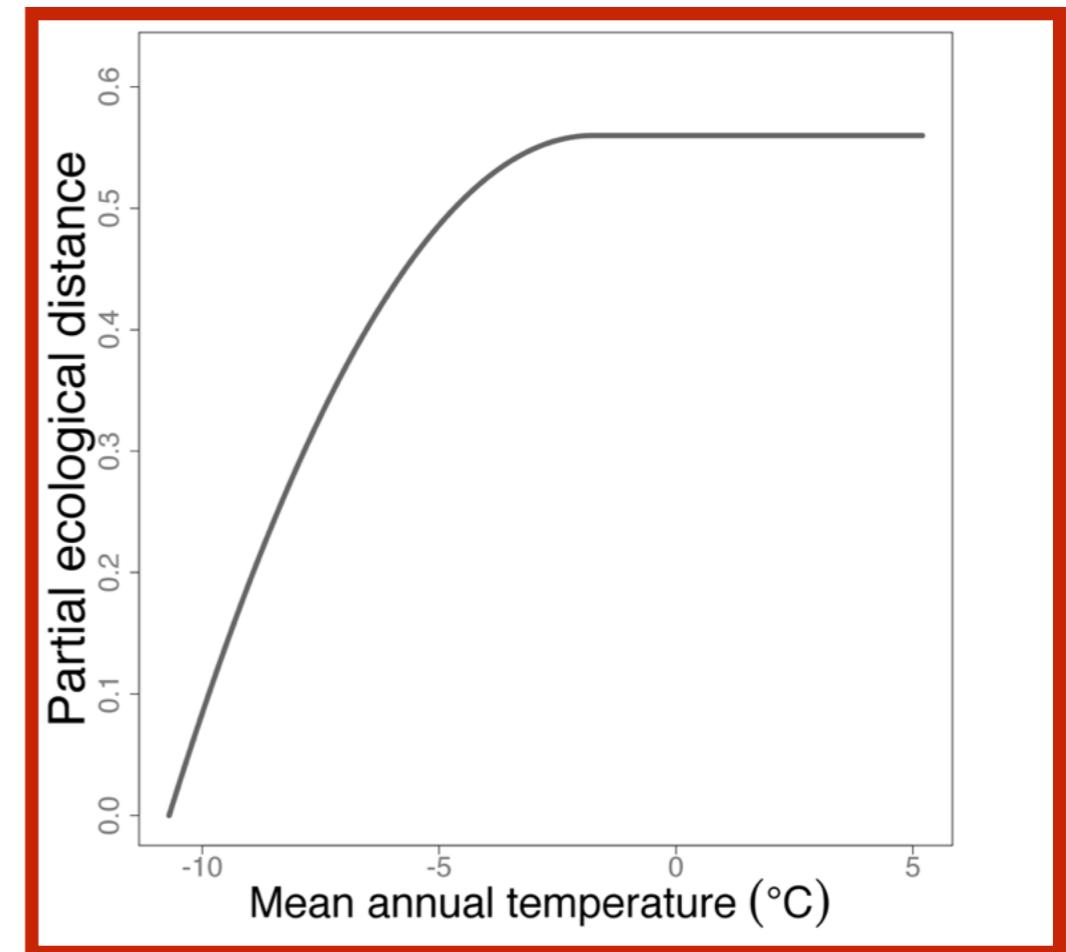
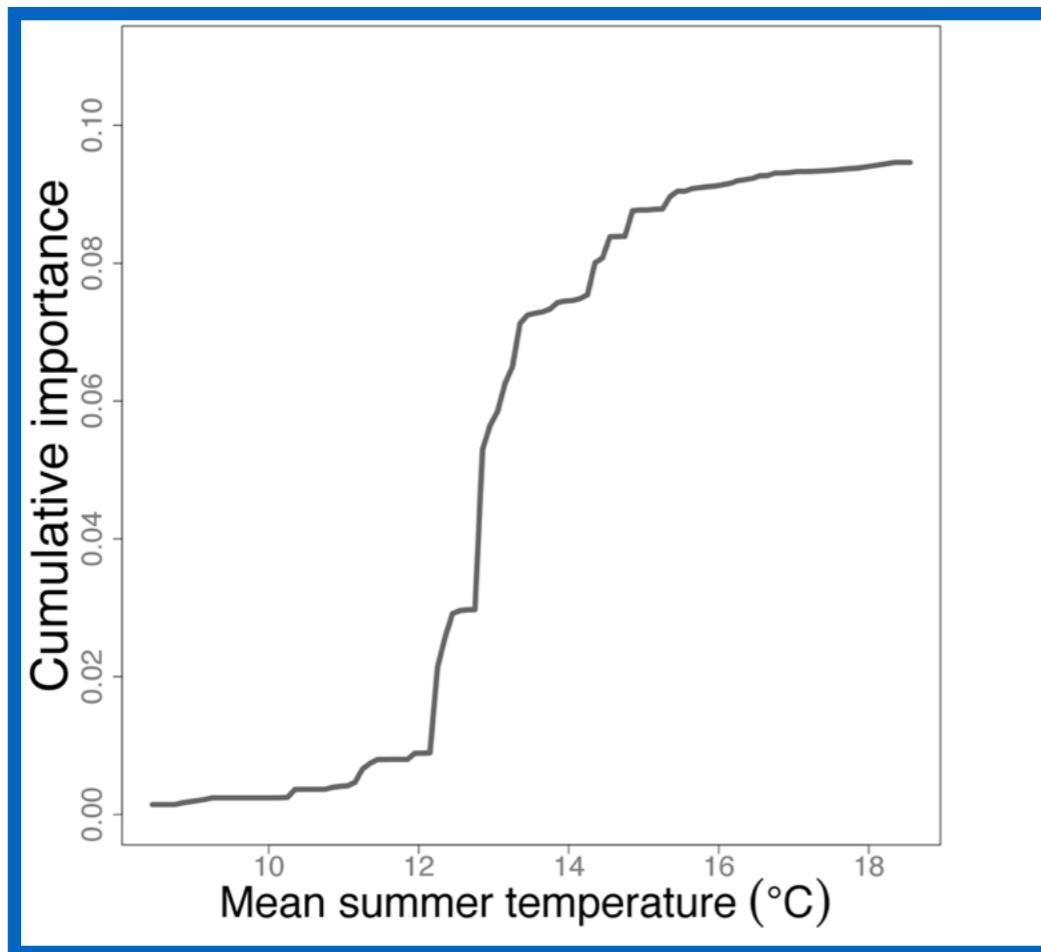
2. Generalized Dissimilarity Modeling



CLMs based on non-linear compositional turnover functions

Shape = rate of biological change

Height = relative importance

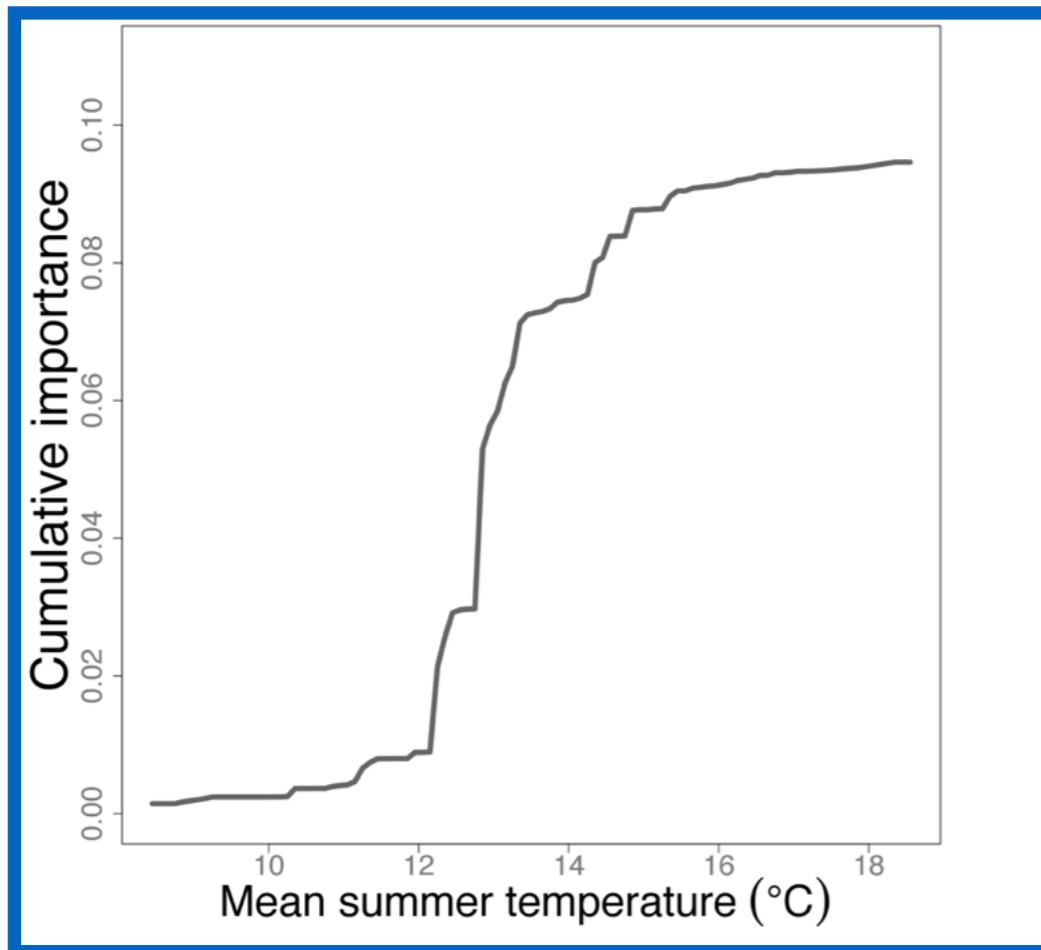


CLMs based on non-linear compositional turnover functions

1. Gradient Forests (GF; Ellis *et al.* 2012)

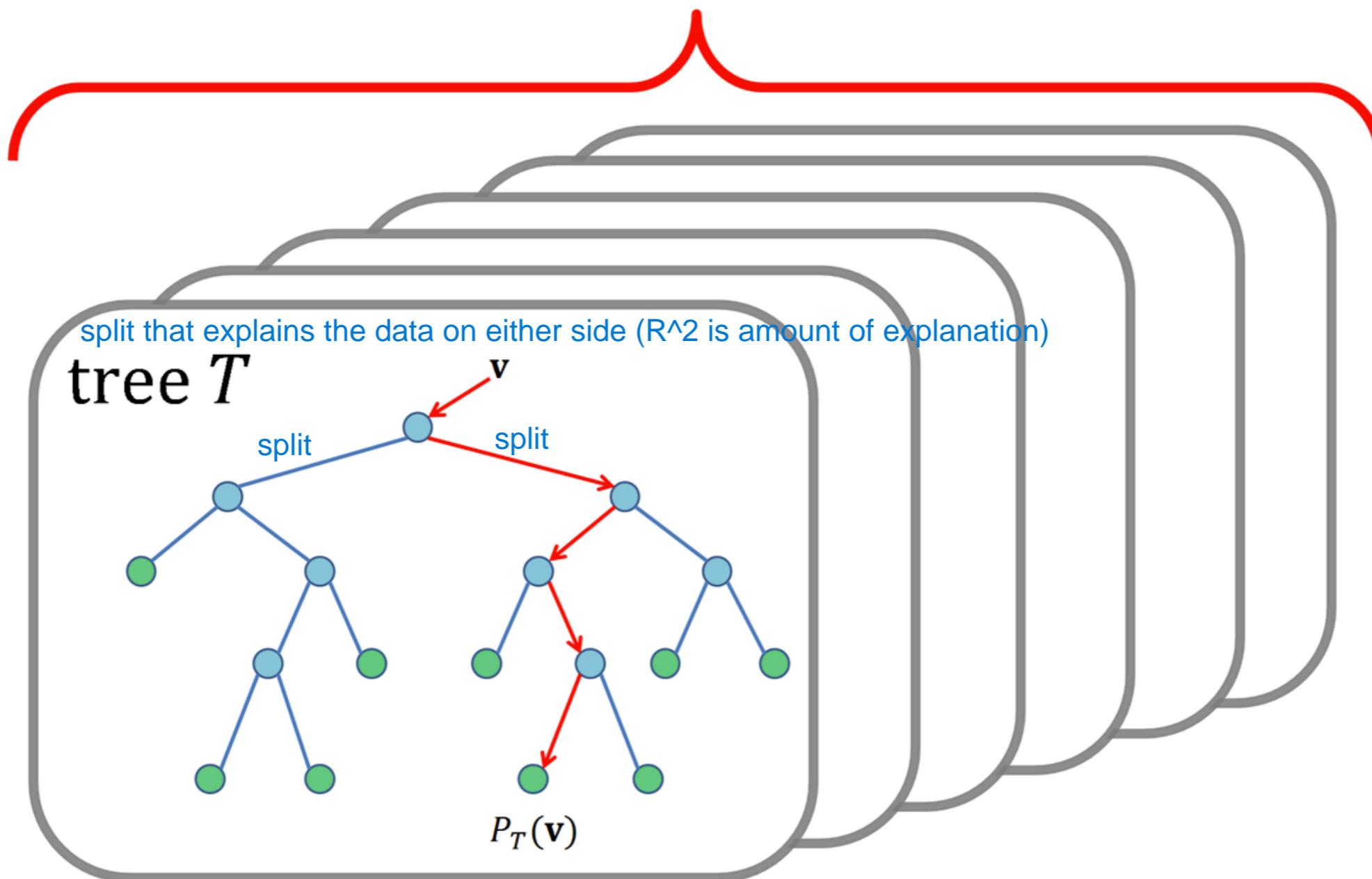
- Nonparametric, machine learning, regression tree method (randomForests)
- Turnover functions for individual taxa and for all taxa combined

ex: curve for each plant and then an overall vegetation model

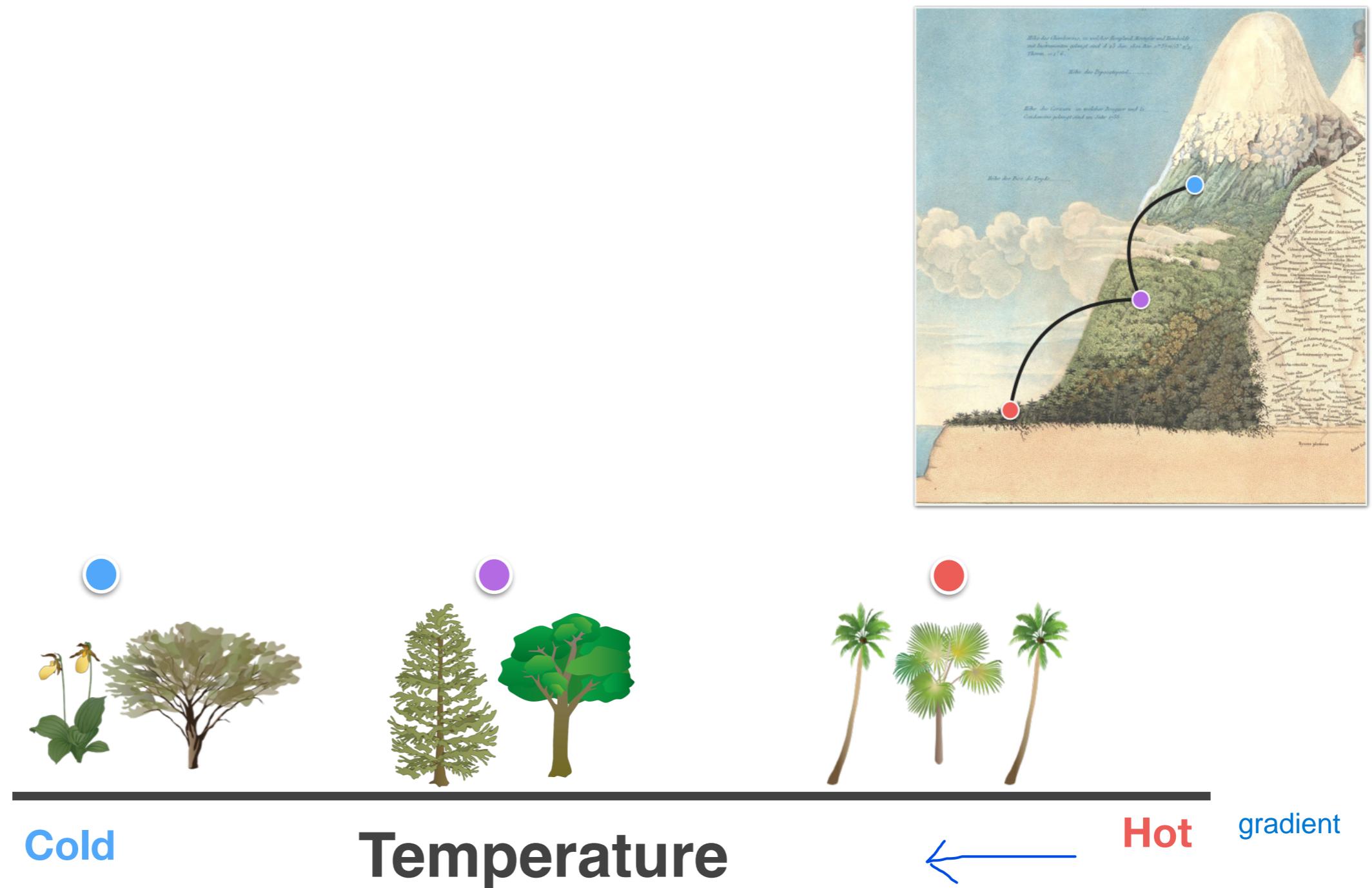


Random Forest

Decision Forest



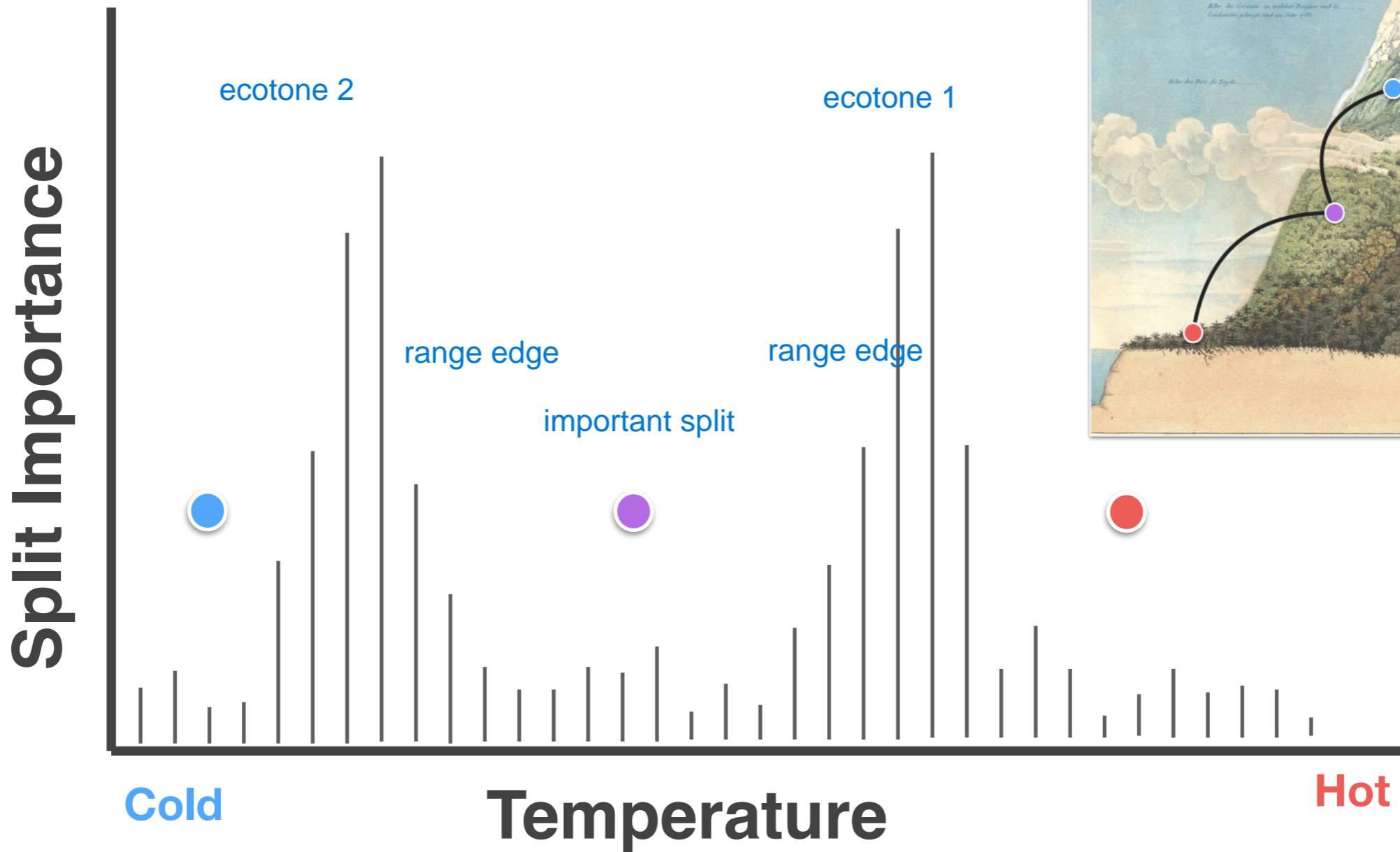
GF: Turnover as a function of “split importance”



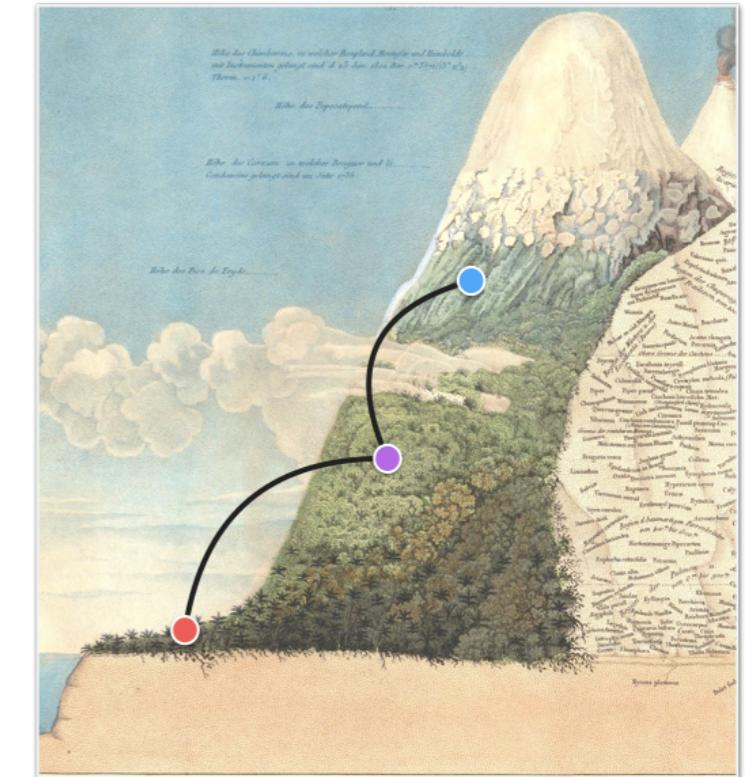
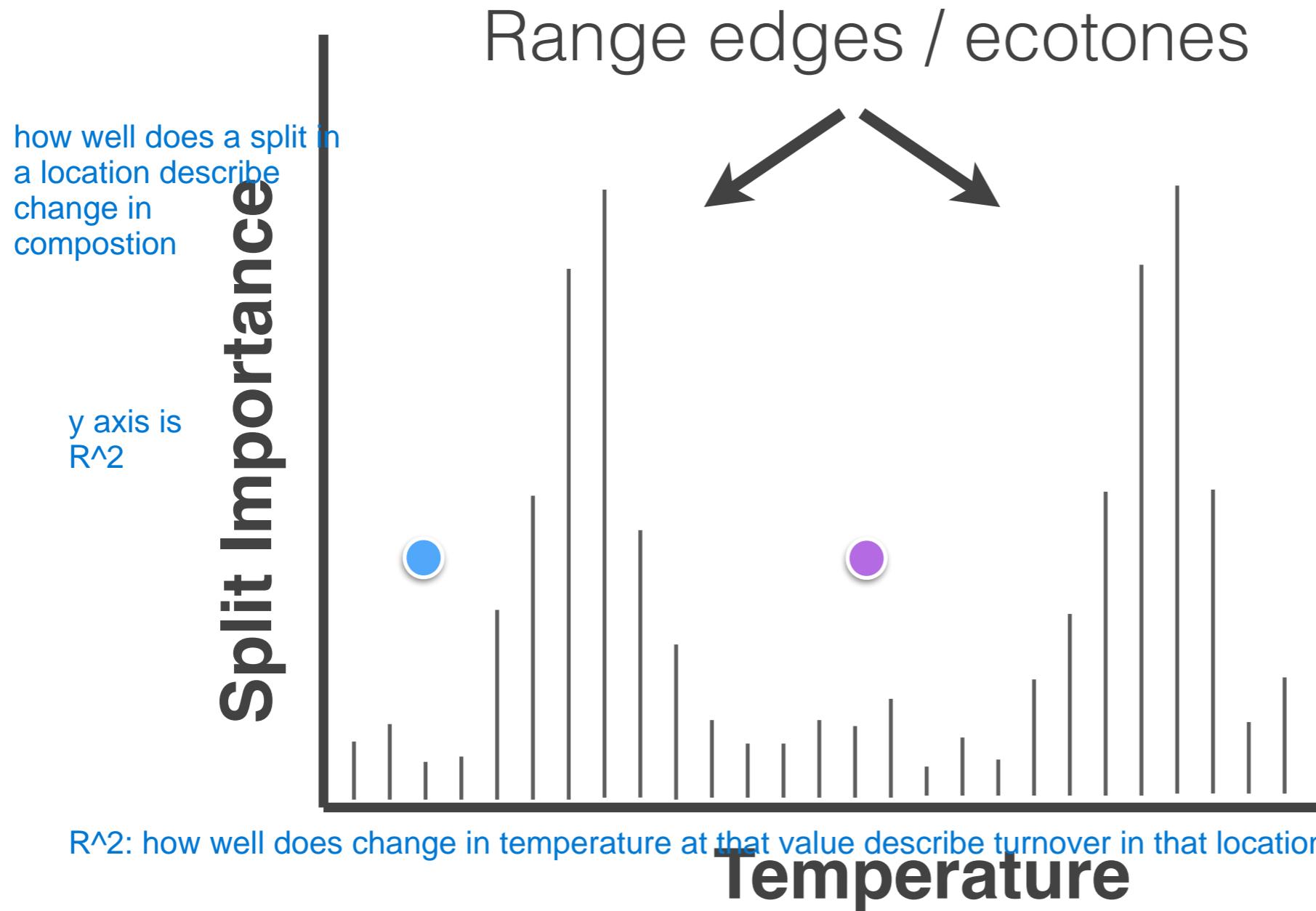
Symbols courtesy of the Integration and Application Network,
University of Maryland Center for Environmental Science
(ian.umces.edu/symbols/)

GF: Turnover as a function of “split importance”

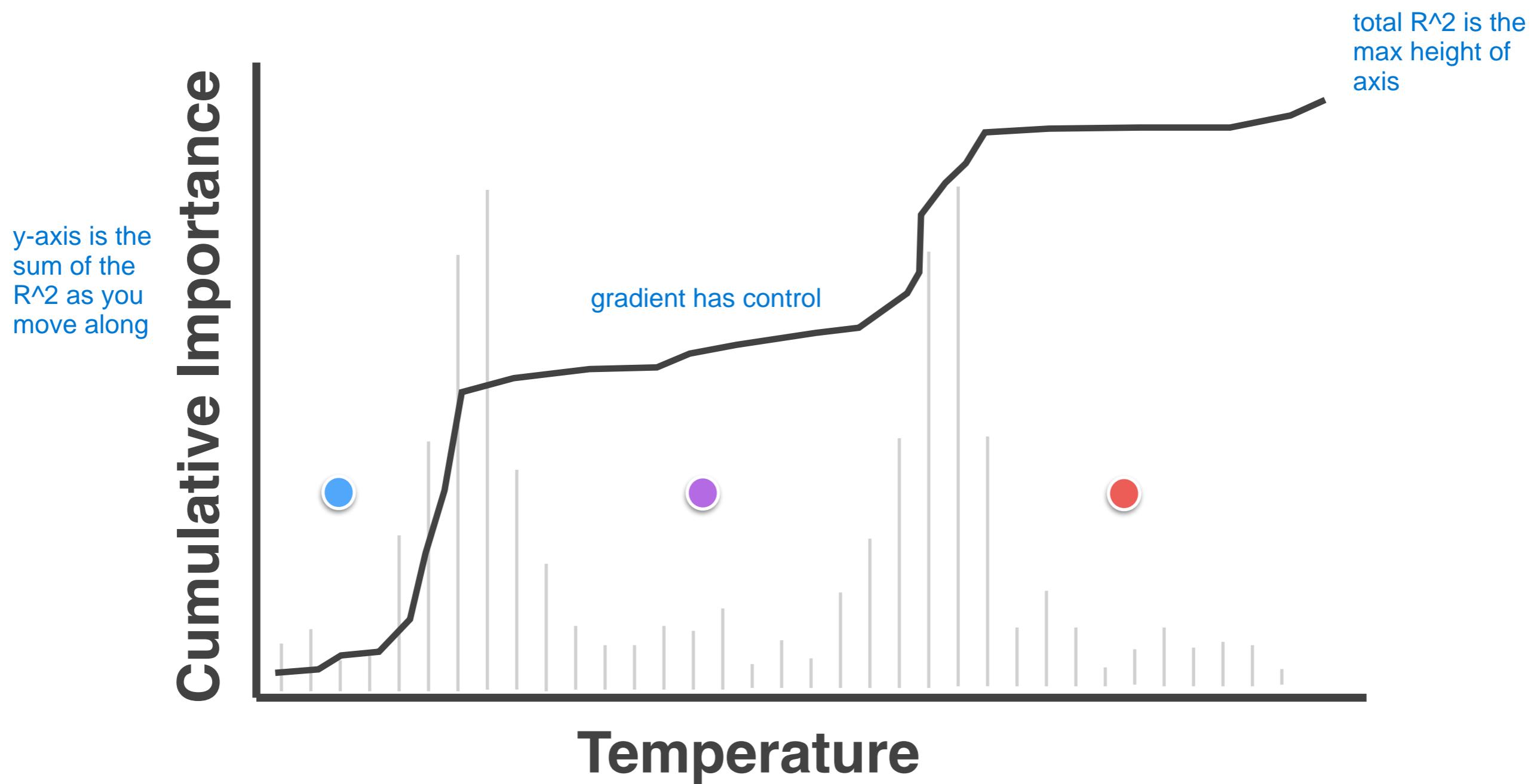
decision trees to describe presence and absence or abundance
how important is a split in a decision tree to species composition



GF: Turnover as a function of “split importance”



GF: Turnover as a function of “split importance”



GF Inputs

- Biological data
 - site x (species) table
 - Abundance, allele frequencies - basically any continuous response (regression)
 - presence-absence (binomial; classification)
 - or other units like allele frequencies (can be any continuous response)
- Environmental data
 - site x environment table
 - Continuous, categorical
 - can use both continuous and categorical variables in the model

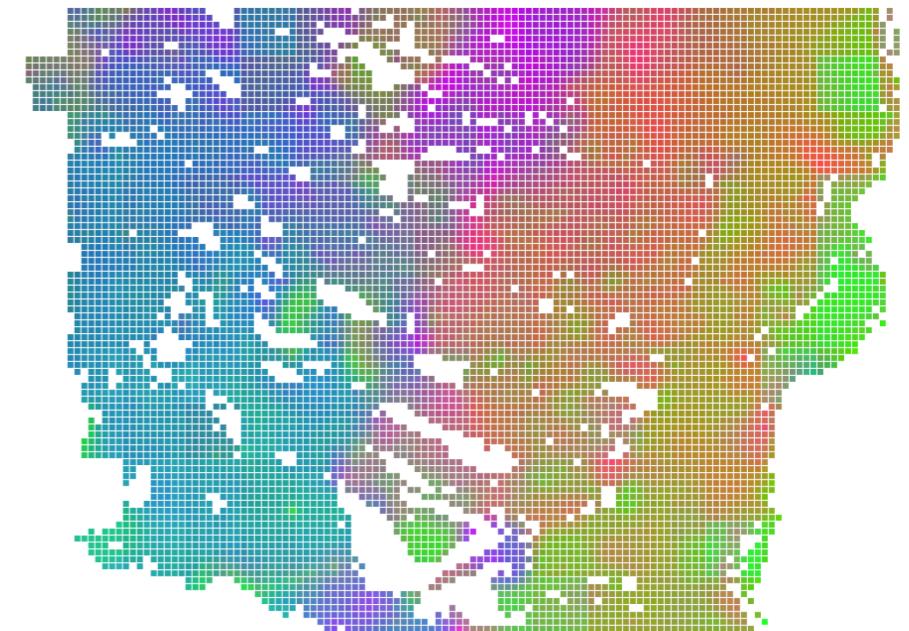
GF Outputs

- Turnover function for each species, for each predictor
- Combined, community-level turnover function for each predictor
- Variable importance (species & community)

GF ‘predictions’

rescaling variables is in essence rescaling environmental data to reflect the biological data

- “Rescaled” environmental data
- Spatial mapping of composition turnover (variation in species composition across space)
- Magnitude of change expected under climate change (Euclidean distance between locations in current and future environmental space)
 - can make conclusions like this community is going to be more effected by environmental change than another community



the more similar the color, the more similar in species composition

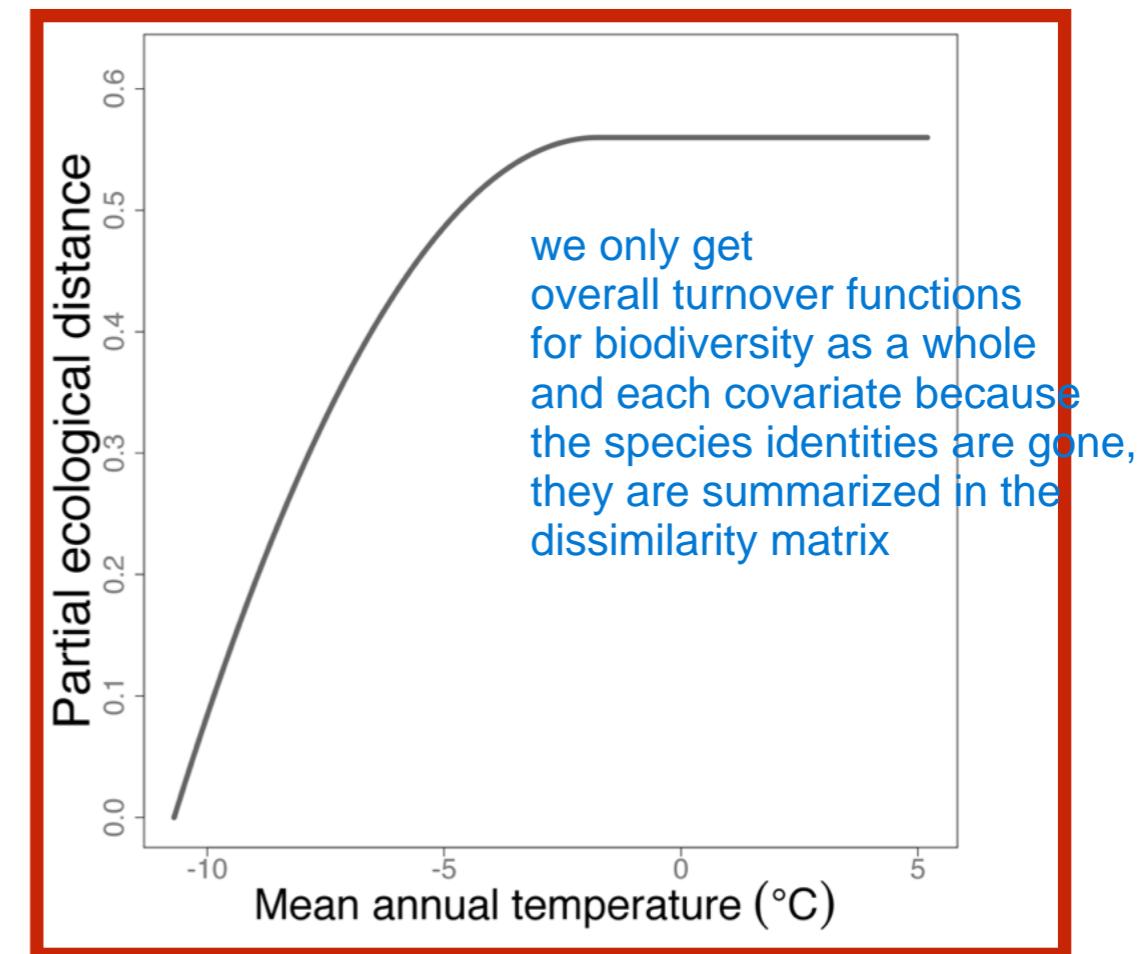
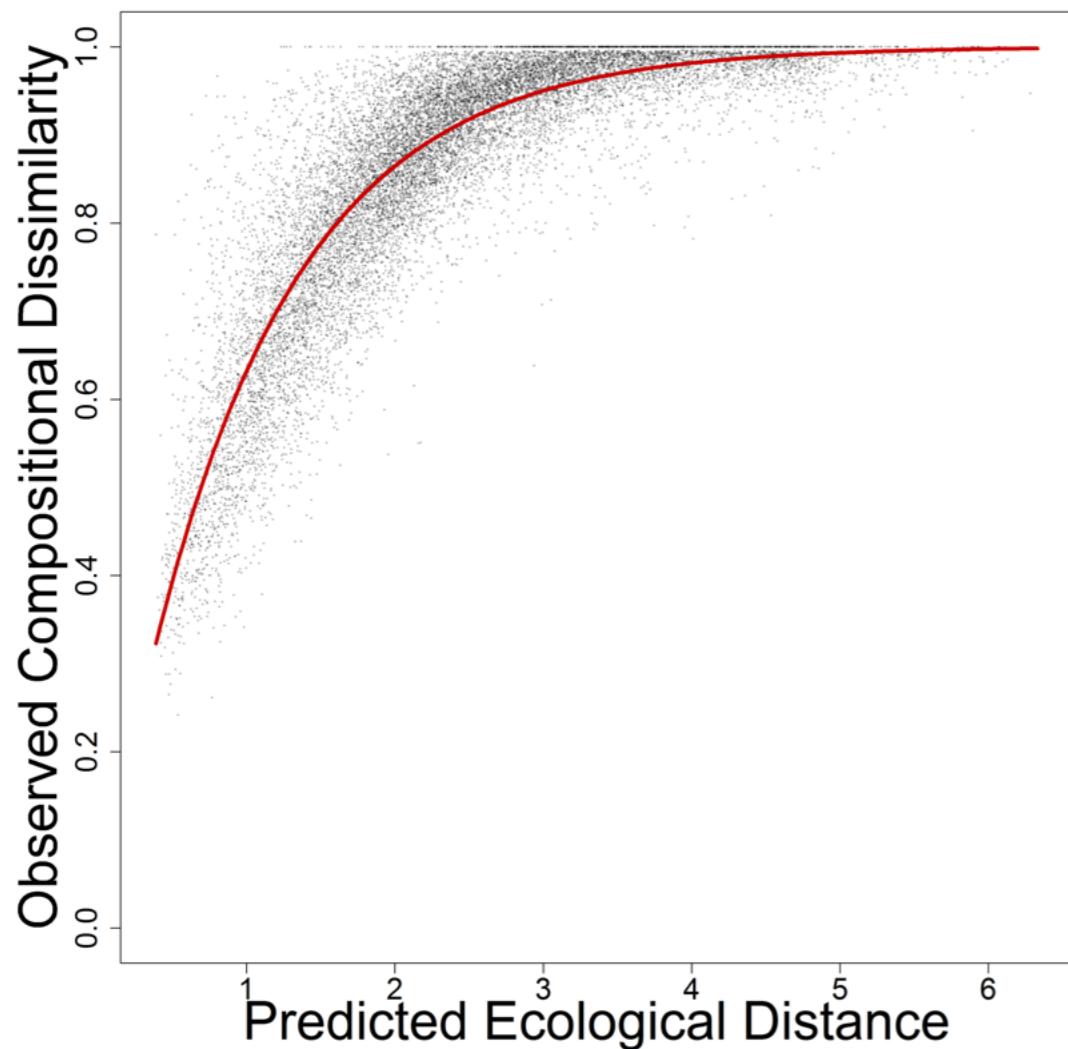
if we had new climate predictions, we could fit the model to those environmental predictions and then calculate the euclidian distance between the two to get the approximate change between the time 1 scenario and the time 2 scenario

GDM: Turnover as a function of geographic and environmental distance

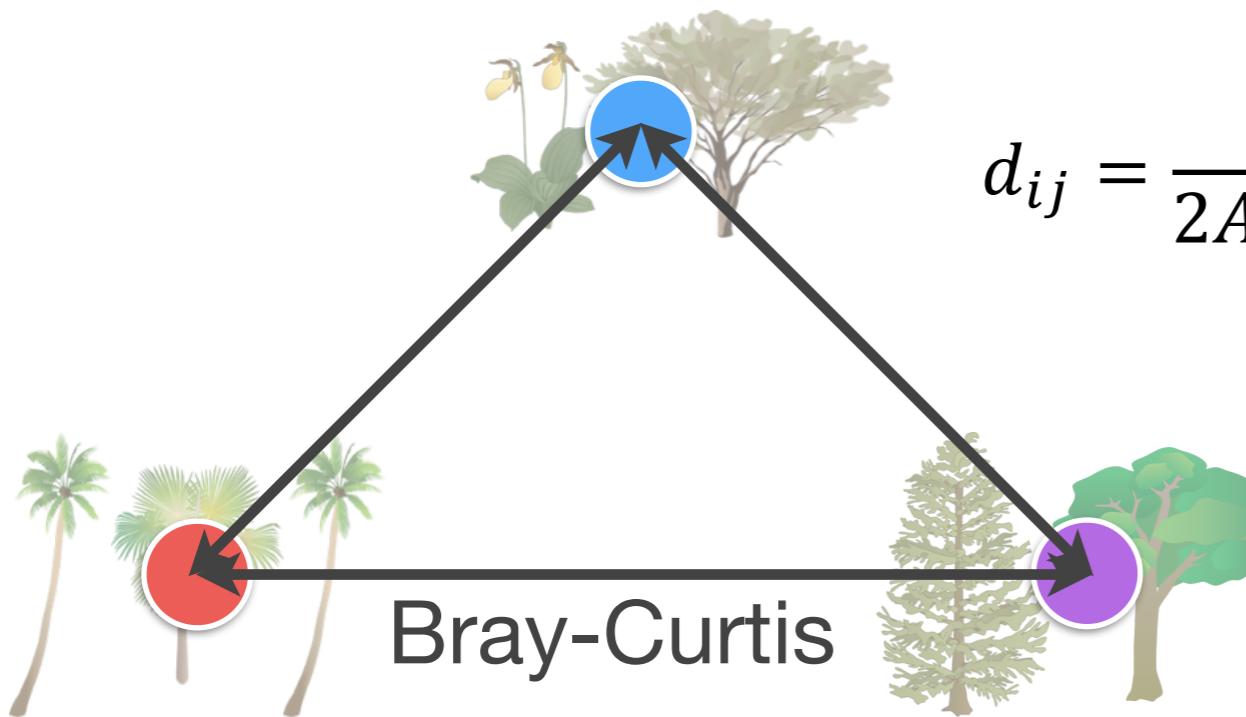
2. Generalized Dissimilarity Modeling (GDM; Ferrier et al. 2007)

- GLM-like framework that models pairwise biological dissimilarity as a function of environmental & geographic distance

convert the site-by-species matrix into a dissimilarity matrix

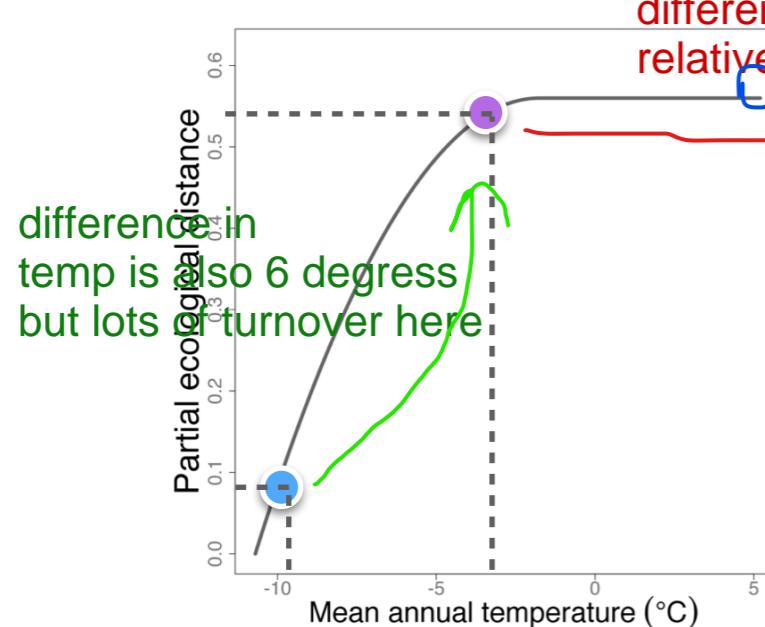


GDM: Turnover as a function of geographic and environmental distance

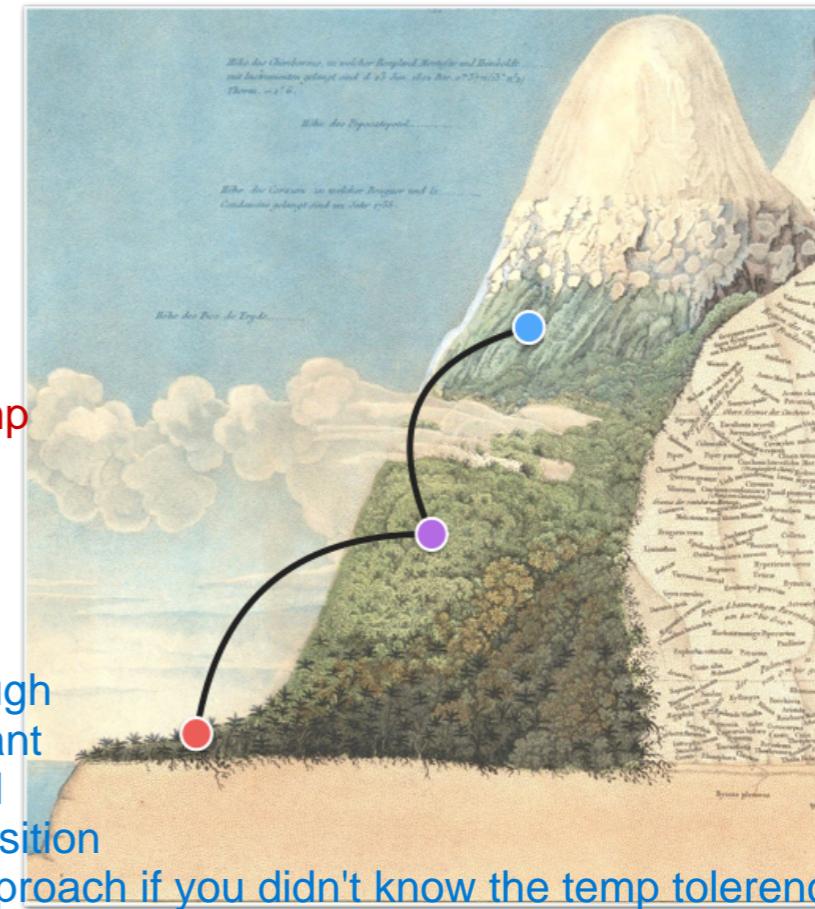


$$d_{ij} = \frac{2A}{2A + C + B}$$

A = number of species common to both sites *i* & *j*
B = number of species present only at site *i*
C = number of species present only at site *j*



assumes the space through
 time distribution is constant
 and that this can be used
 to predict species composition
 turnover (ie. this is an approach if you didn't know the temp tolerance)



GDM

$$-\ln\left(1 - d_{ij}\right) = a_0 + \sum_{p=1}^n \left| f_p(x_{pi}) - f_p(x_{pj}) \right|$$

GDM

Biological distance
metric

$$-\ln(1 - d_{ij}) = a_0 + \sum_{p=1}^n |f_p(x_{pi}) - f_p(x_{pj})|$$

GDM

Biological distance
metric

$$-\ln(1 - d_{ij}) = a_0 + \sum_{p=1}^n |f_p(x_{pi}) - f_p(x_{pj})|$$

Bray-Curtis
Dissimilarity

$$d_{ij} = \frac{2A}{2A + C + B}$$

GDM

Biological distance
metric

$$-\ln(1 - d_{ij}) = a_0 + \sum_{p=1}^n |f_p(x_{pi}) - f_p(x_{pj})|$$

Model intercept

GDM

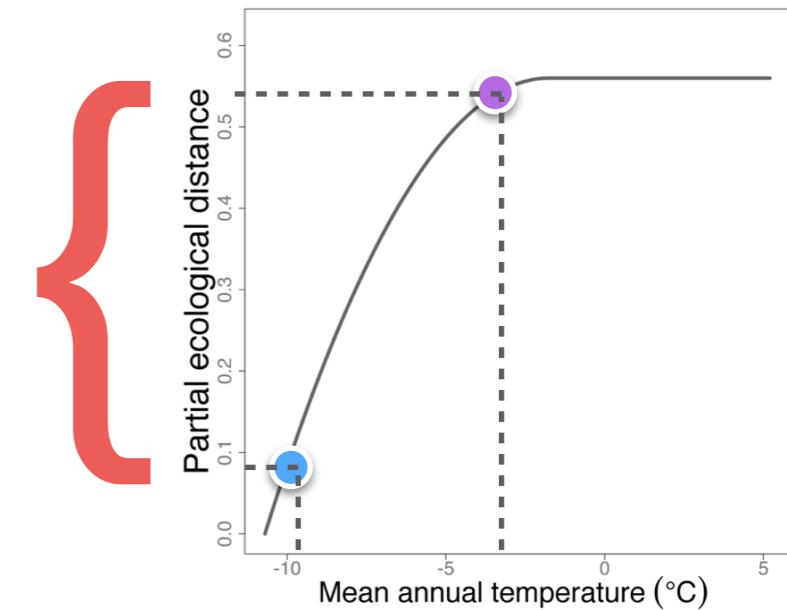
Biological distance metric

$$-\ln(1 - d_{ij}) = a_0 + \sum_{p=1}^n |f_p(x_{pi}) - f_p(x_{pj})|$$

Model intercept

Scaled Environmental Distance

can use Euclidian distance (how far away in geographic space), travel-cost distance or travel distance (ie for elevation) both possible for environmental distance



Predicting biological distance between any pair of locations

Biological distance metric

Model intercept

Scaled Environmental Distance

$$d_{ij} = 1 - e^{-(a + E_{ij})}$$

$$E_{ij} = \sum_{p=1}^n |f_p(x_{pi}) - f_p(x_{pj})|$$

GDM Inputs

- Environmental data
 - Continuous mainly
 - Categorical (risky) and only if ordinal risky because you must calculate the distance between variables
 - Geographic predictors common
 - Least cost paths
 - Resistance distance
 - Biological predictors common
 - Dissimilarity matrix for another taxonomic group ie habitat selection differences

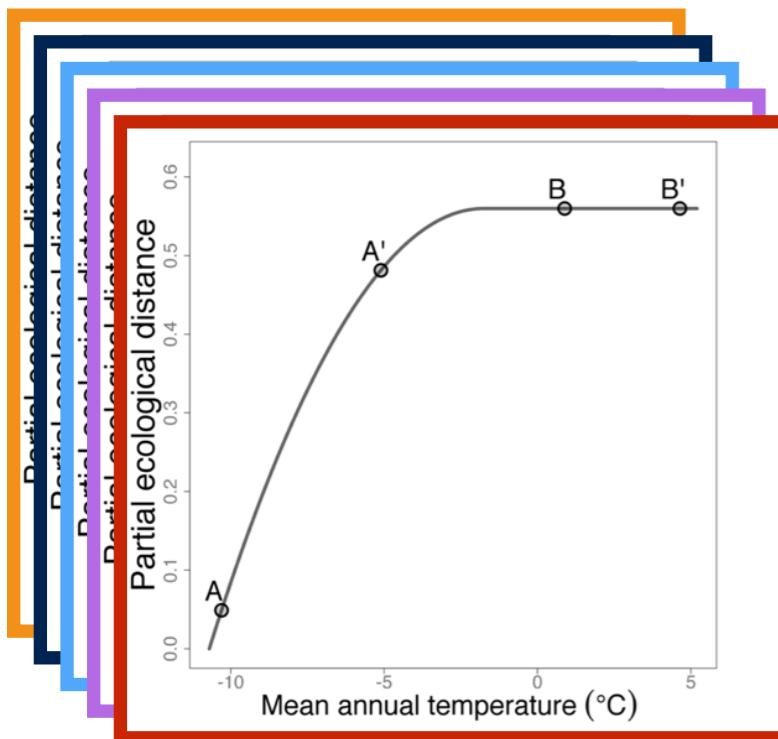
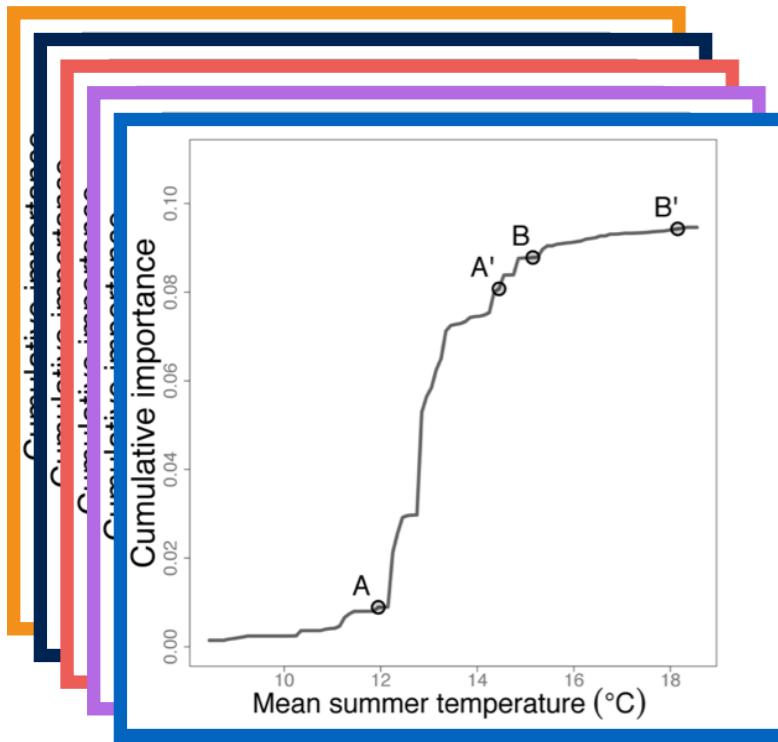
GDM Outputs

- Deviance explained how well the model explains the data
- Turnover function for each predictor
- Fitted model (predict pairwise distances)

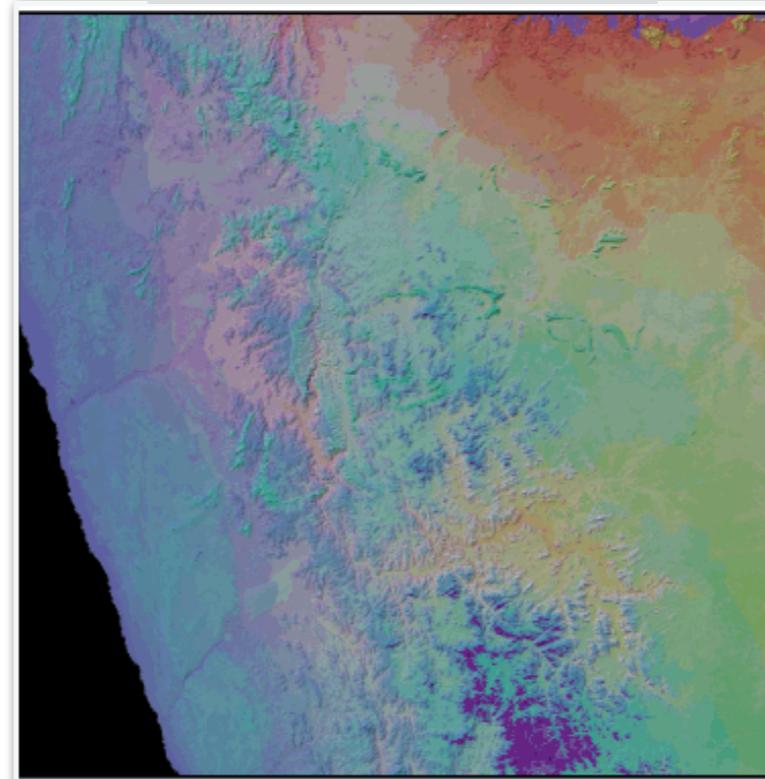
GDM predictions

- Transformed environmental data same as rescaled
- Predicted pairwise biological distance
 - In space
 - Proportion of species shared between locations
 - Or time
 - Expected magnitude of change in species composition

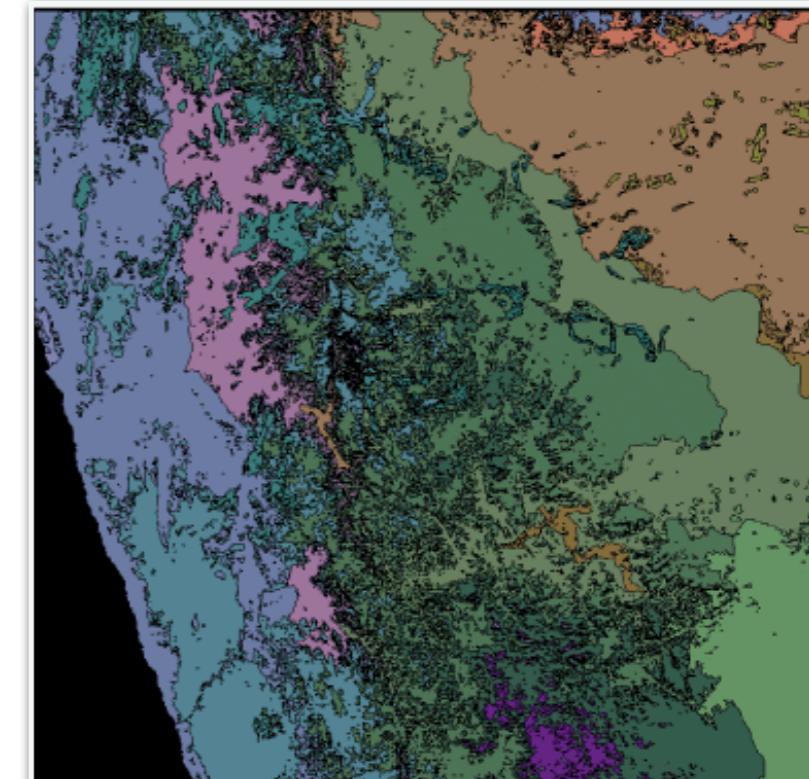
Mapping: Turnover functions convert environmental predictors into biological importance



Continuous



Discrete

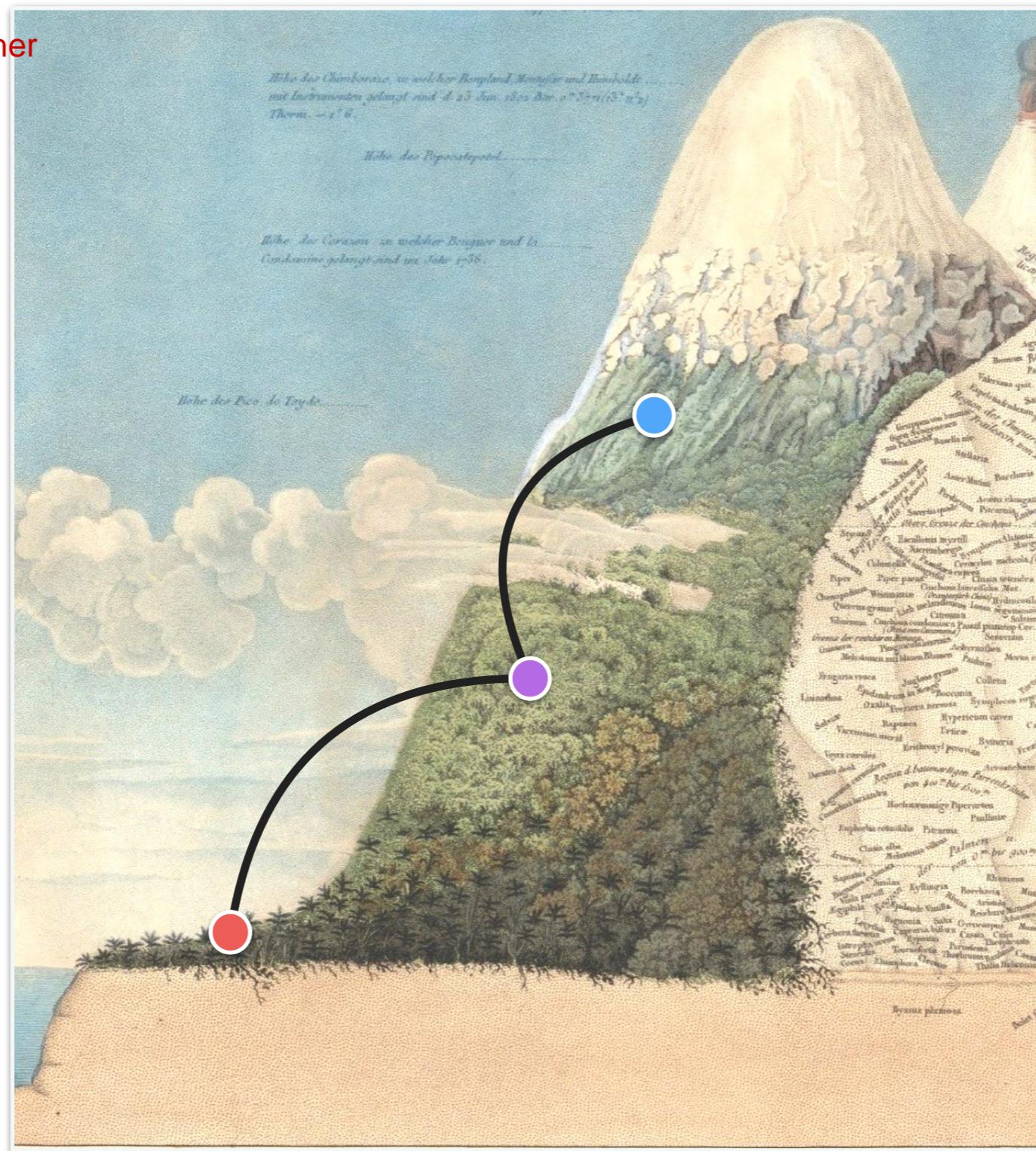


more similar the color of the pixel, the more similar the species composition

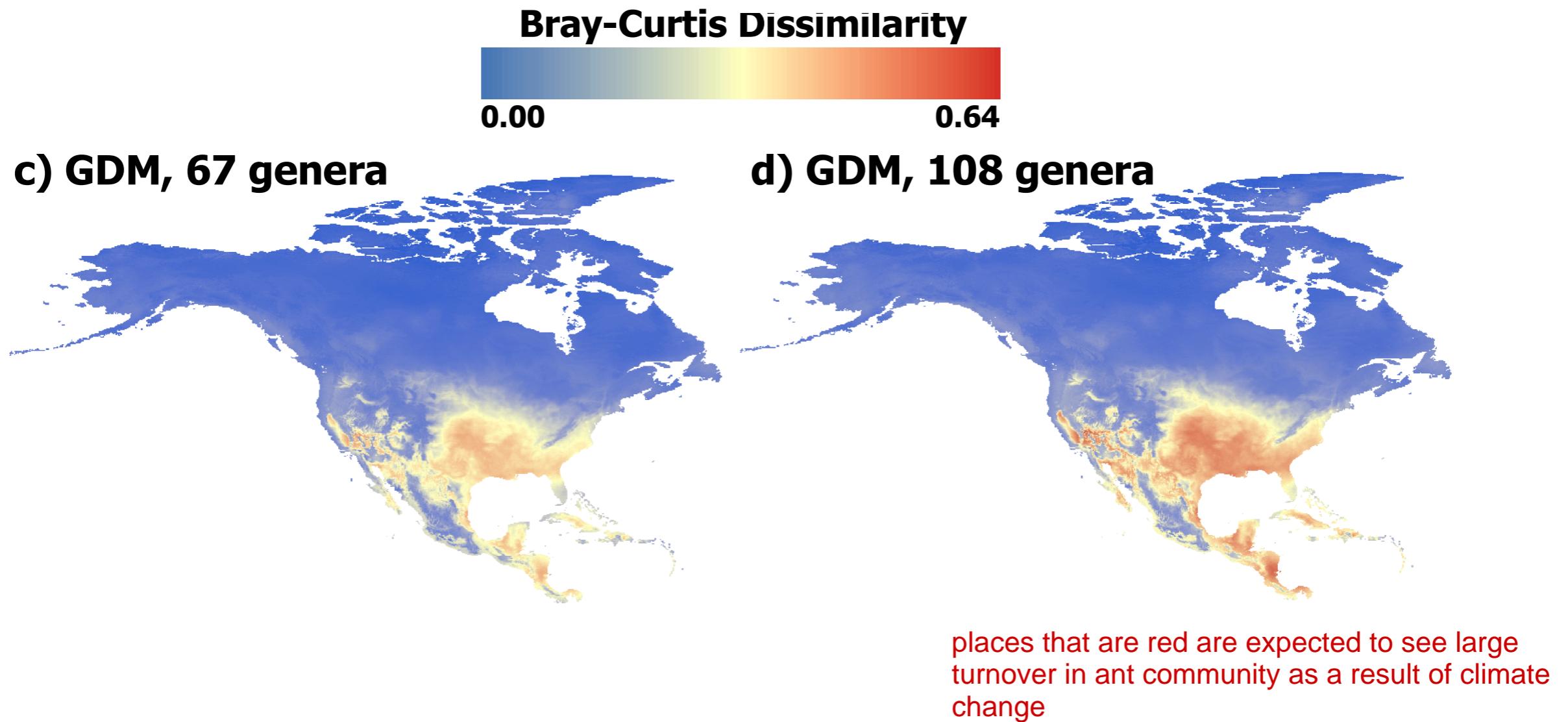
provides the community type discretely
ie. understory forest type

β -diversity: change (turnover) in community structure along spatial, **temporal** or environmental gradients

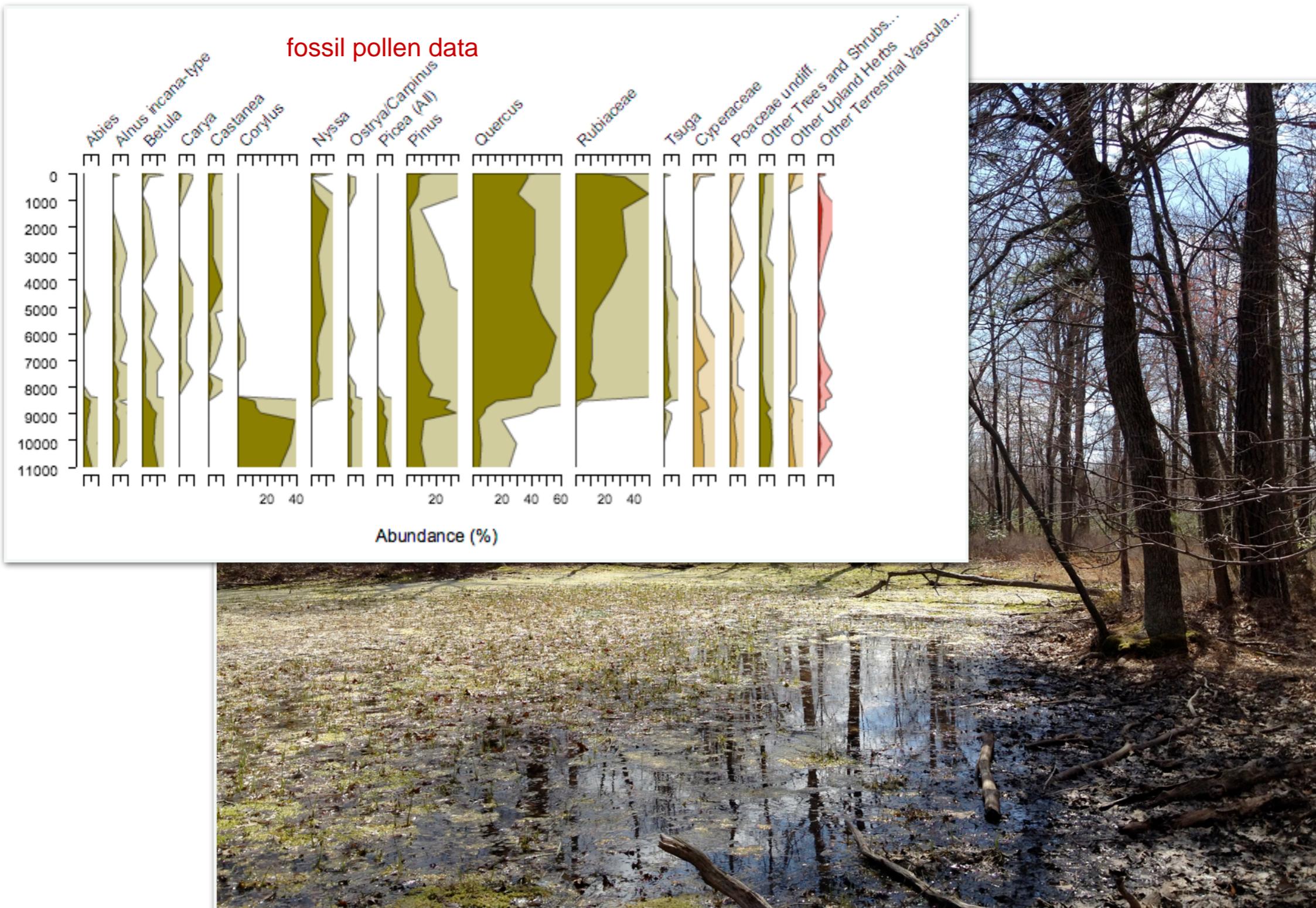
change in species composition either
moving along a spatial gradient
OR as the environment changes



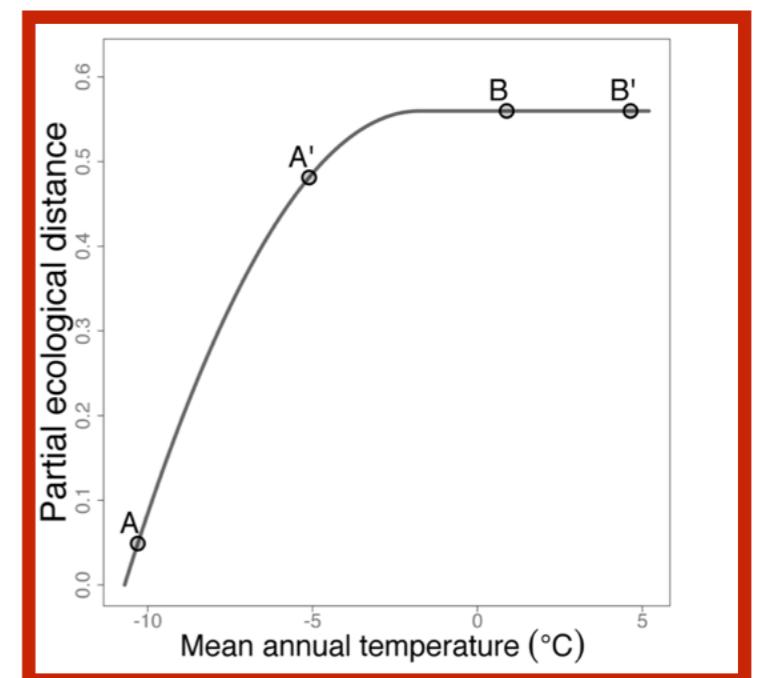
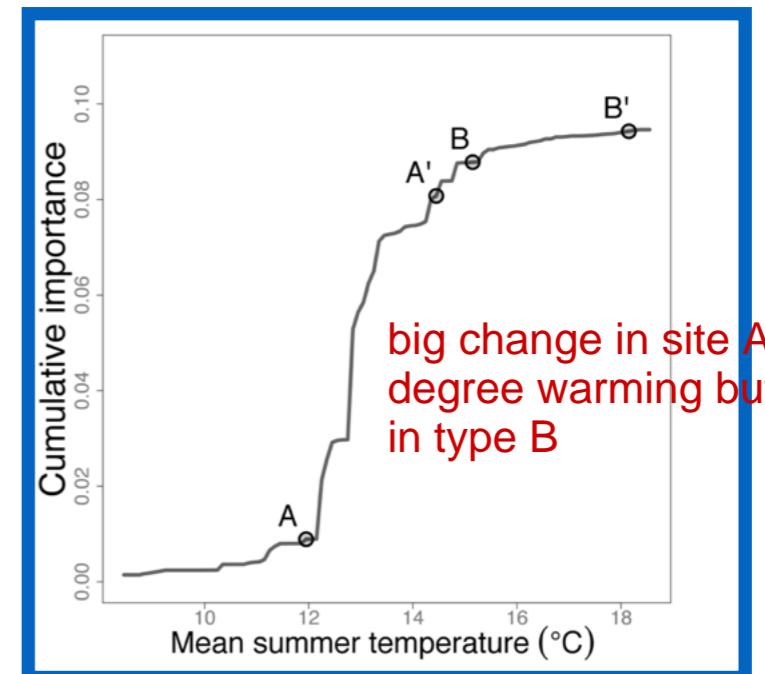
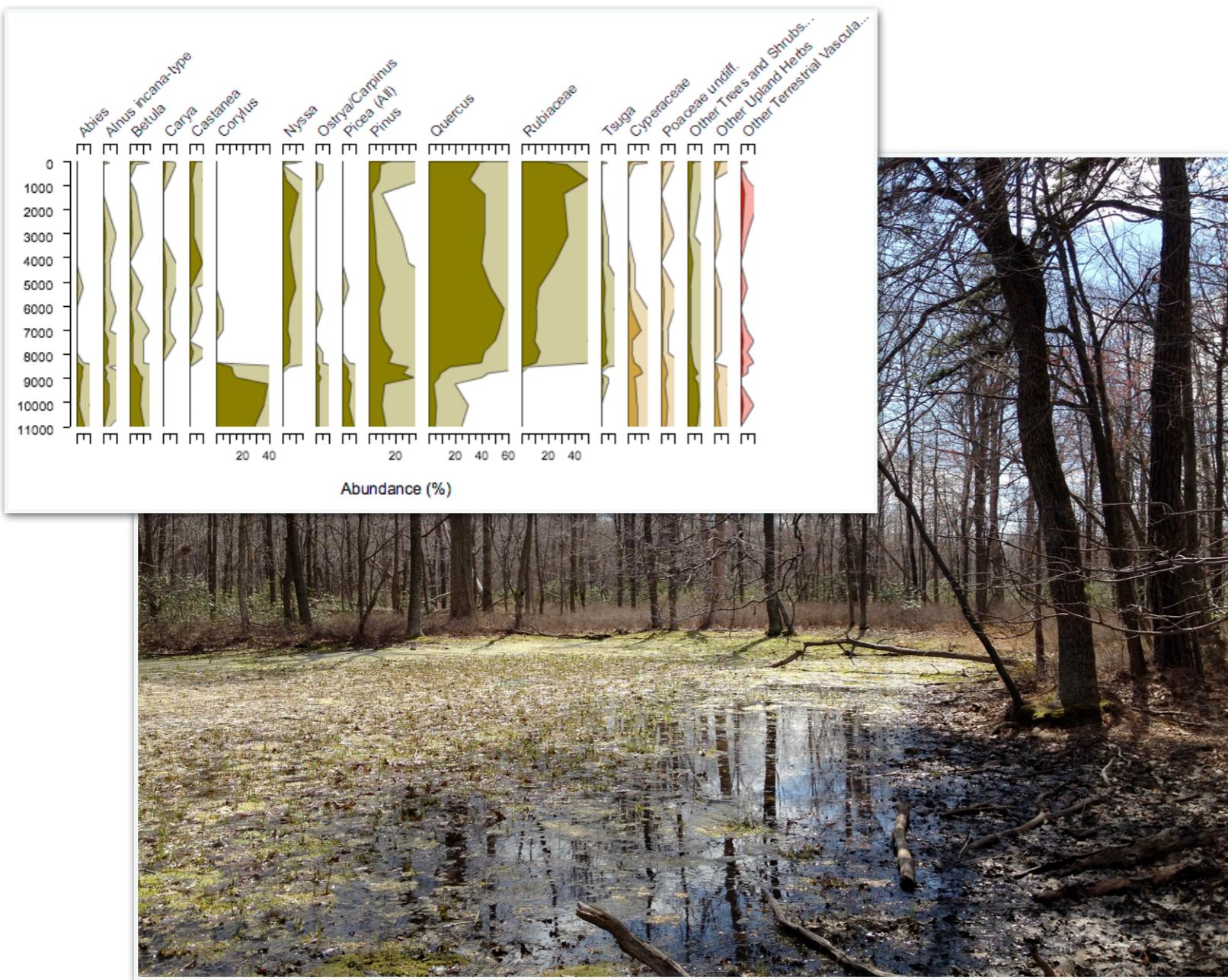
Turnover functions can be used to predict biological change through time in response to environmental change



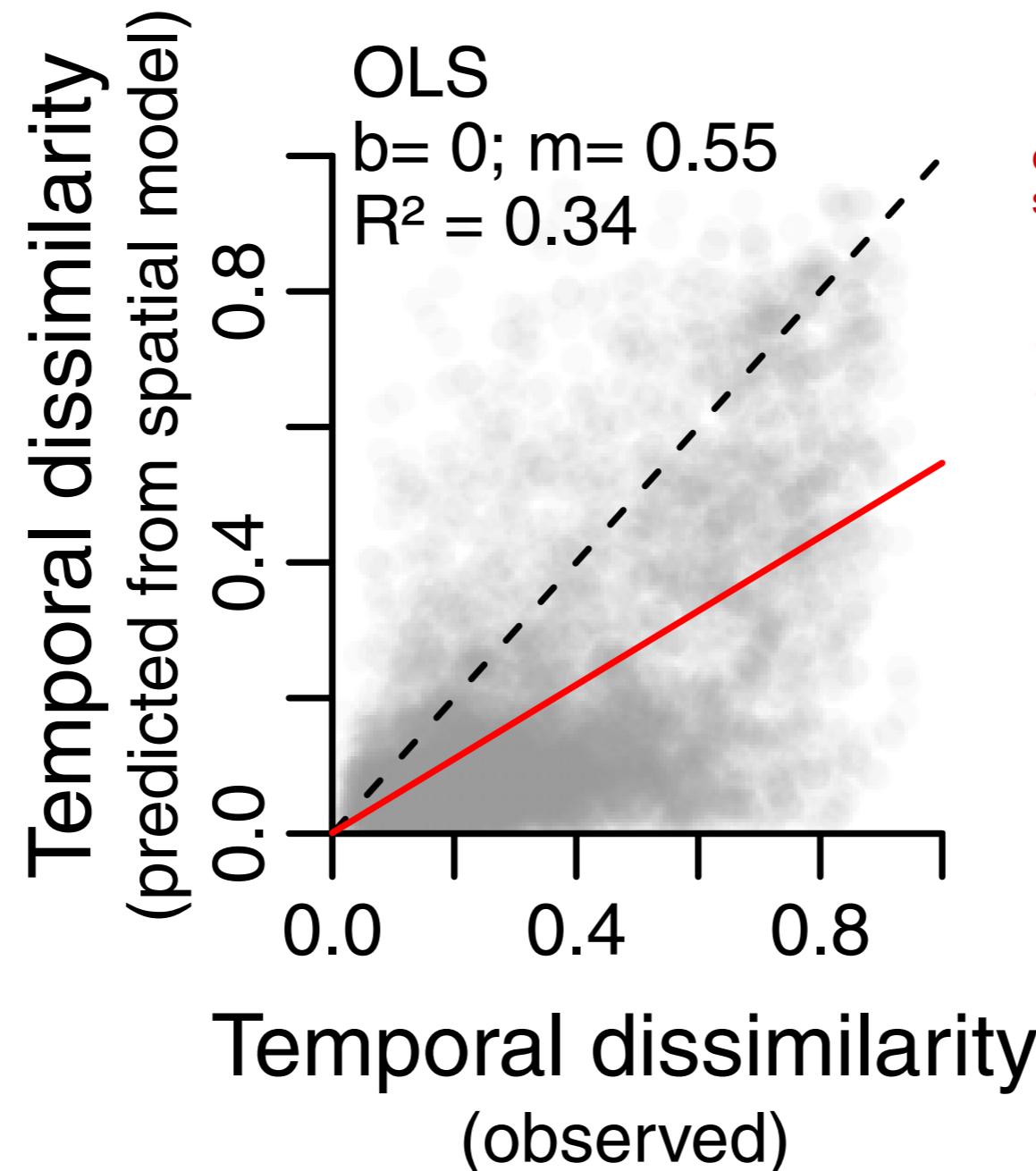
β -diversity: change (turnover) in community structure along spatial, **temporal** or environmental gradients



Does turnover in space predict turnover in time?



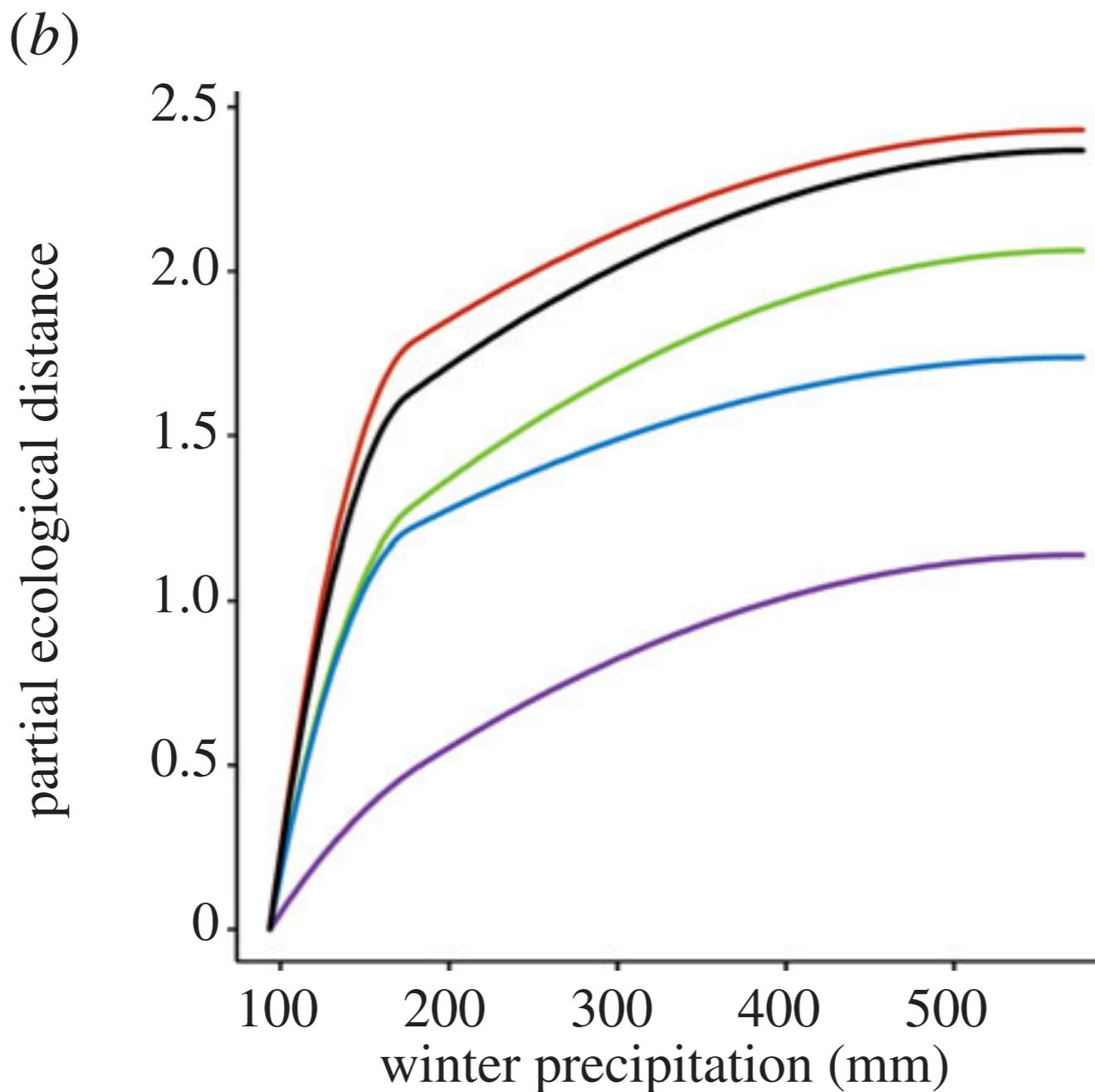
Does turnover in space predict turnover in time? YES



can predict using the space for time substitution

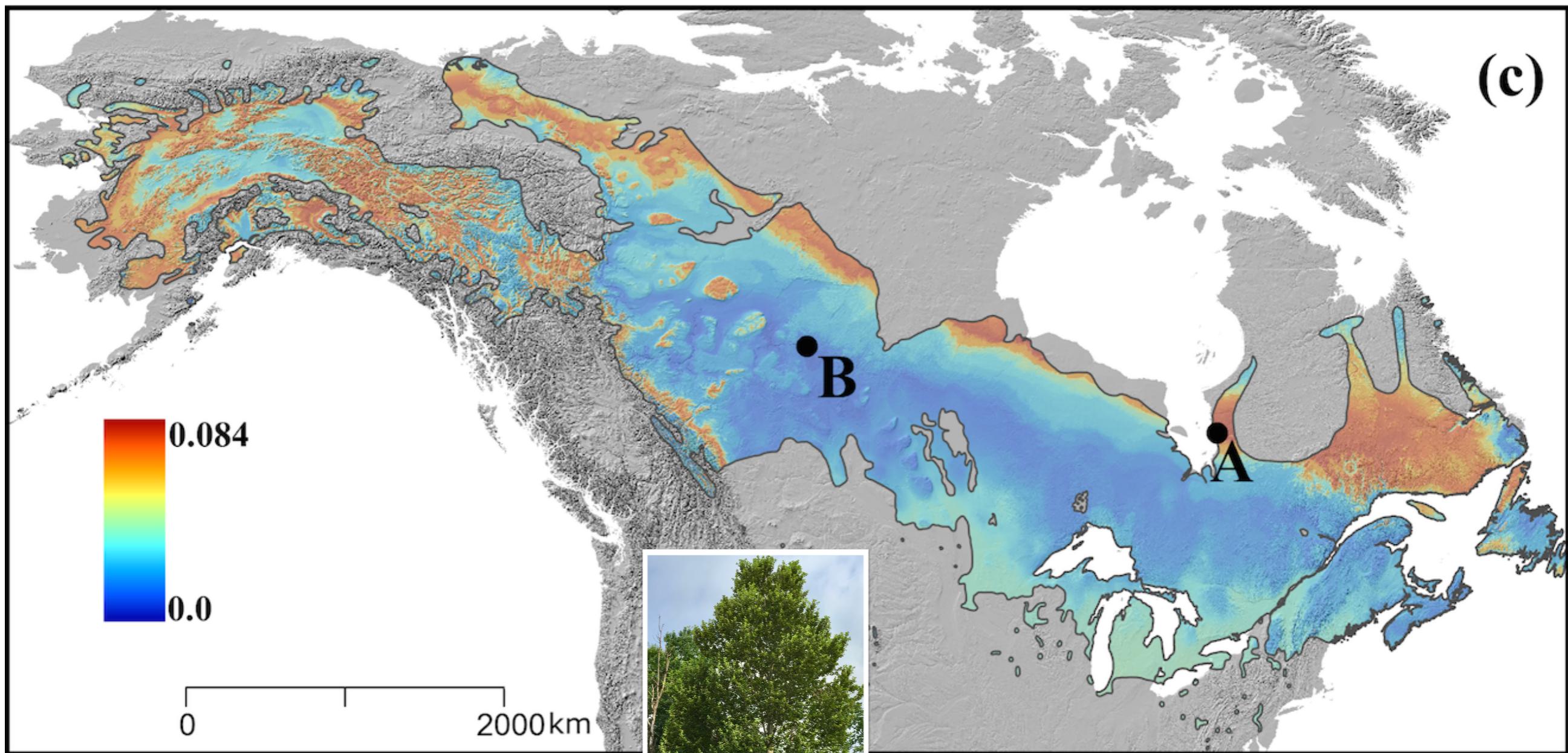
magnitude of change is easier to predict than changes in individual species -- ie. better at assessing the overall impact

Rapid change in abundance / composition
in space suggests sensitivity to
environmental change



“Genetic offset” under future climate

also for spatial and temporal changes in gene composition-- ie. population-level impact for species



trees in Alaska and New England respond similarly to climate change-- red areas more vulnerable to gene composition changes despite this happening at the population level