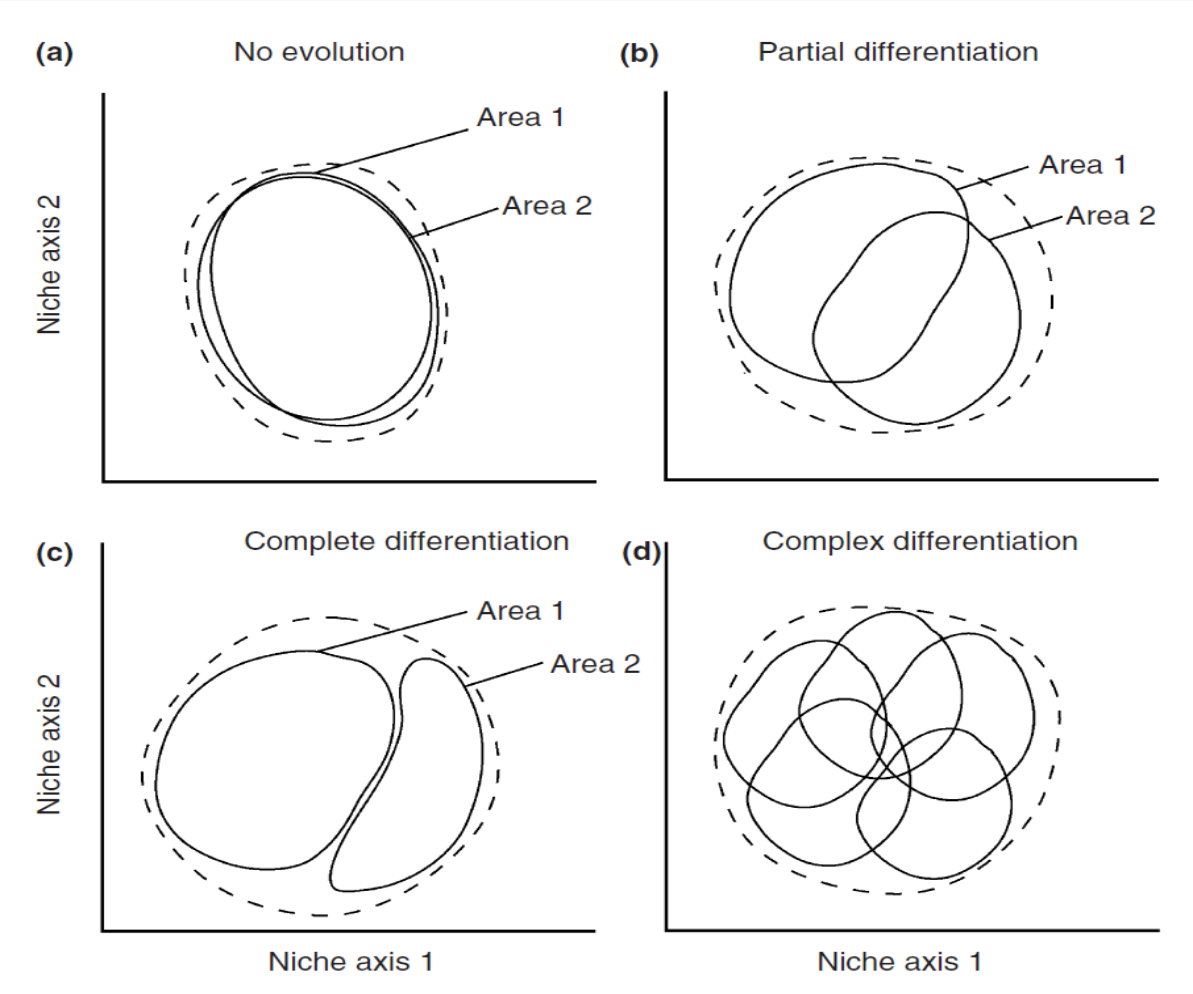
Incorporating evolution in ecological niche modelling

Francisco Rodríguez-Sánchez

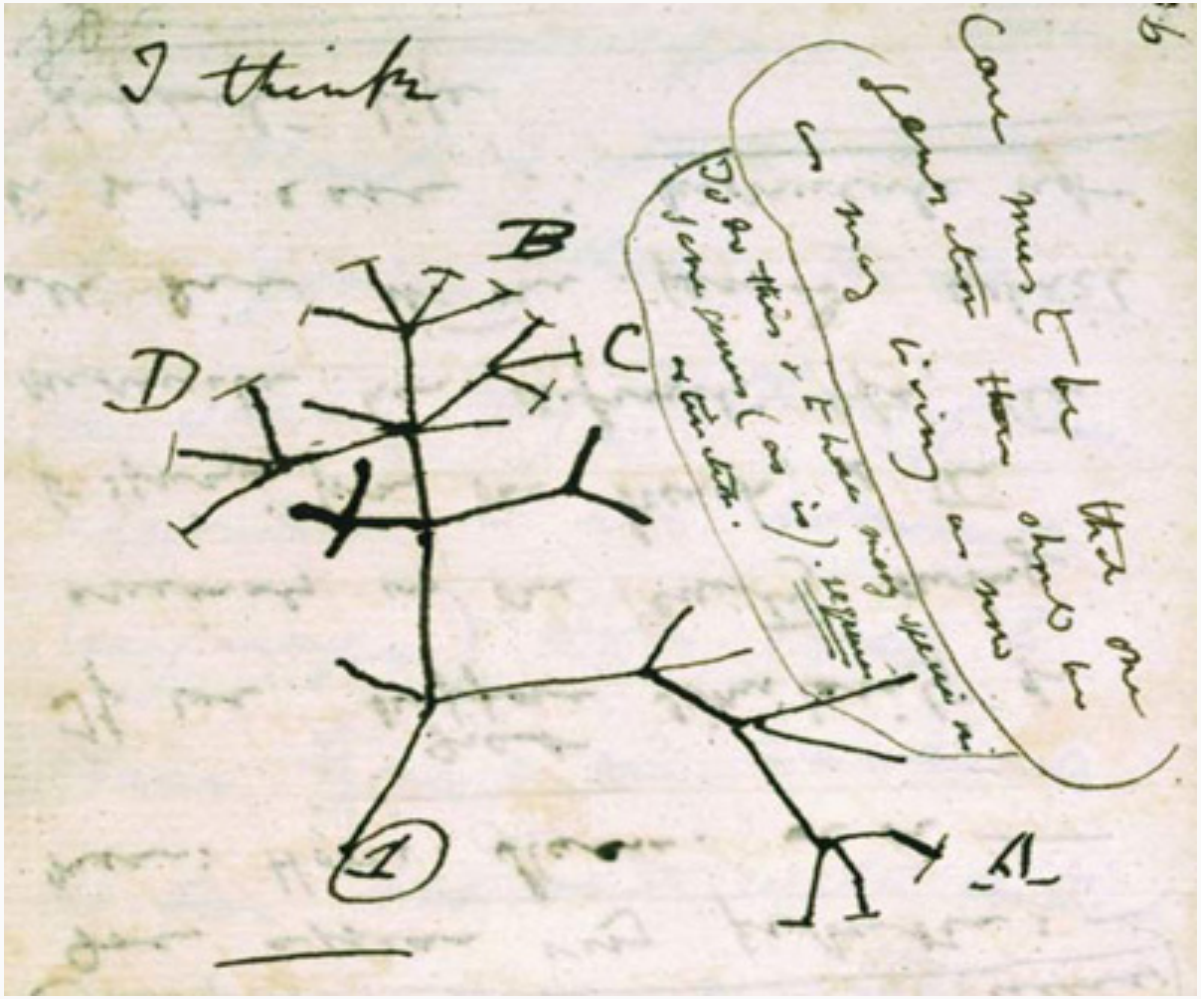
@frod_san

Species: aggregates of (different) populations



Peterson & Holt 2003

No species emerged from vacuum

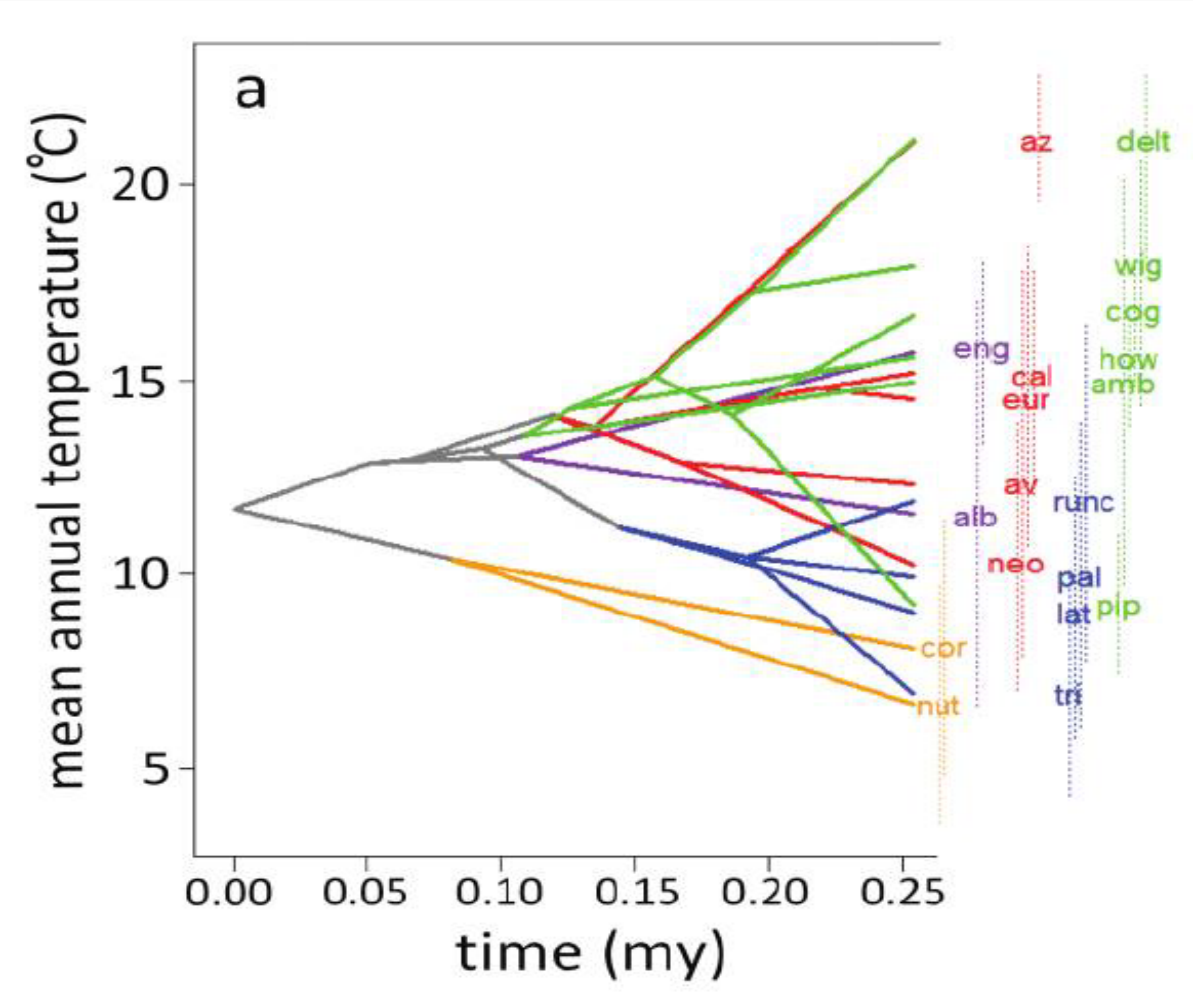


SPECIES

distribution

modelling

Incorporating evolution into niche modelling



Evans et al. 2009

How can we best model niches. . .

- accounting for **local adaptation**/intraspecific variation

How can we best model niches. . .

- accounting for **local adaptation**/intraspecific variation
- exploiting knowledge of evolutionary relations (**niche conservatism**)

How can we best model niches...

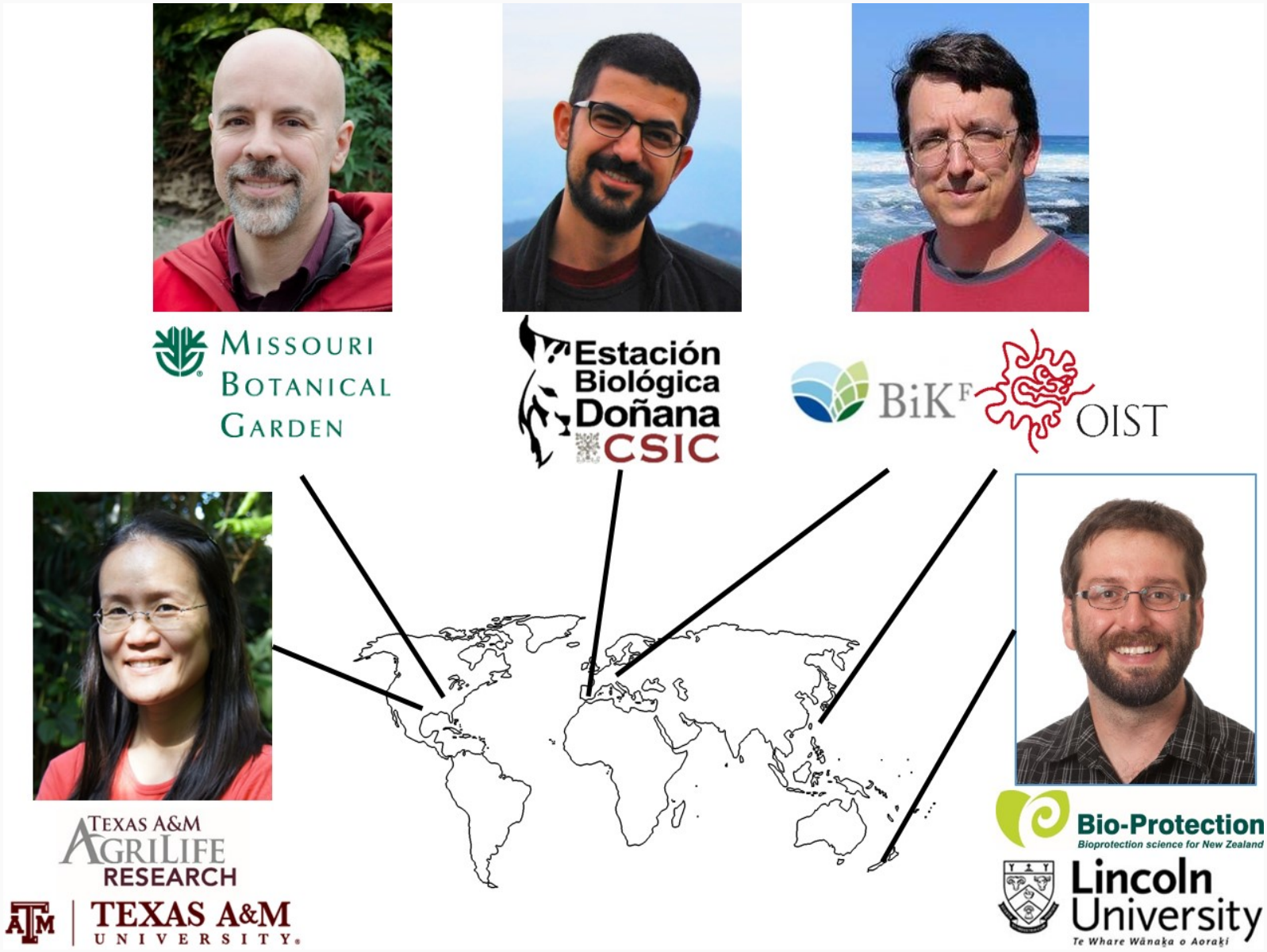
- accounting for **local adaptation**/intraspecific variation
- exploiting knowledge of evolutionary relations (**niche conservatism**)
- minimising **bias** from noisy distribution data

NIMBioS

The National Institute for Mathematical and Biological Synthesis
1122 Volunteer Blvd, Suite 106, University of Tennessee, Knoxville, TN 37996
www.nimbios.org

NIMBioS Investigative Workshop

Species' Range Shifts in a Warming World


Missouri Botanical Garden
Estación Biológica Doñana CSIC
BiKF
OIST
Texas A&M AgriLife Research
Texas A&M University
Bio-Protection
Bioprotection science for New Zealand
Lincoln University
Te Whare Wānaka o Aoraki

Review

Niche Estimation Above and Below the Species Level

Adam B. Smith [1,*] William Godsoe,[2] Francisco Rodríguez-Sánchez,[3] Hsiao-Hsuan Wang,[4] and Dan Warren[5,6]

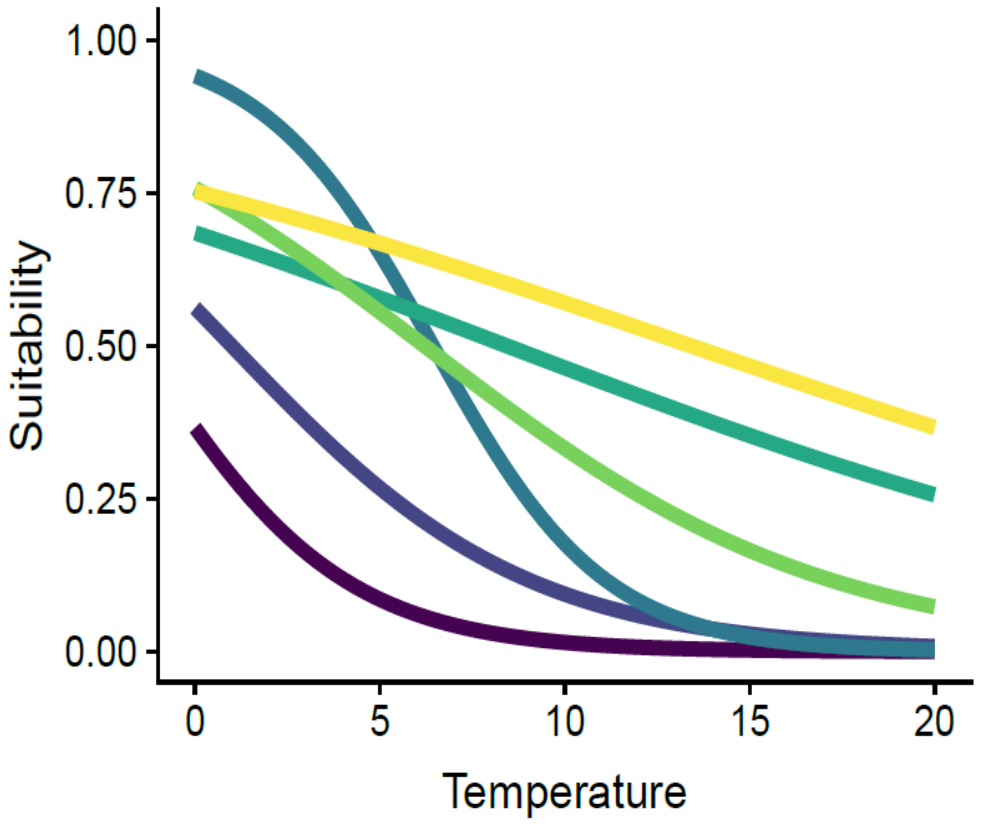
https://doi.org/10.1016/j.tree.2018.10.012

Splitting

Splitting

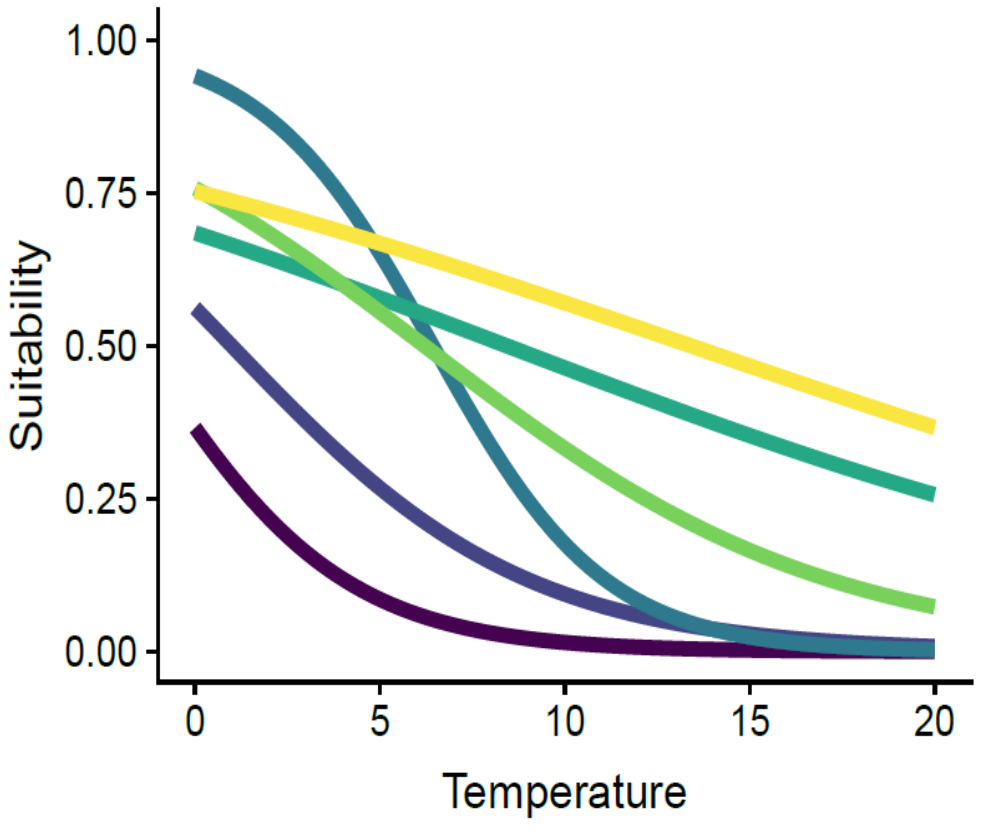
Shirley Grant

Splitting: fitting independent models



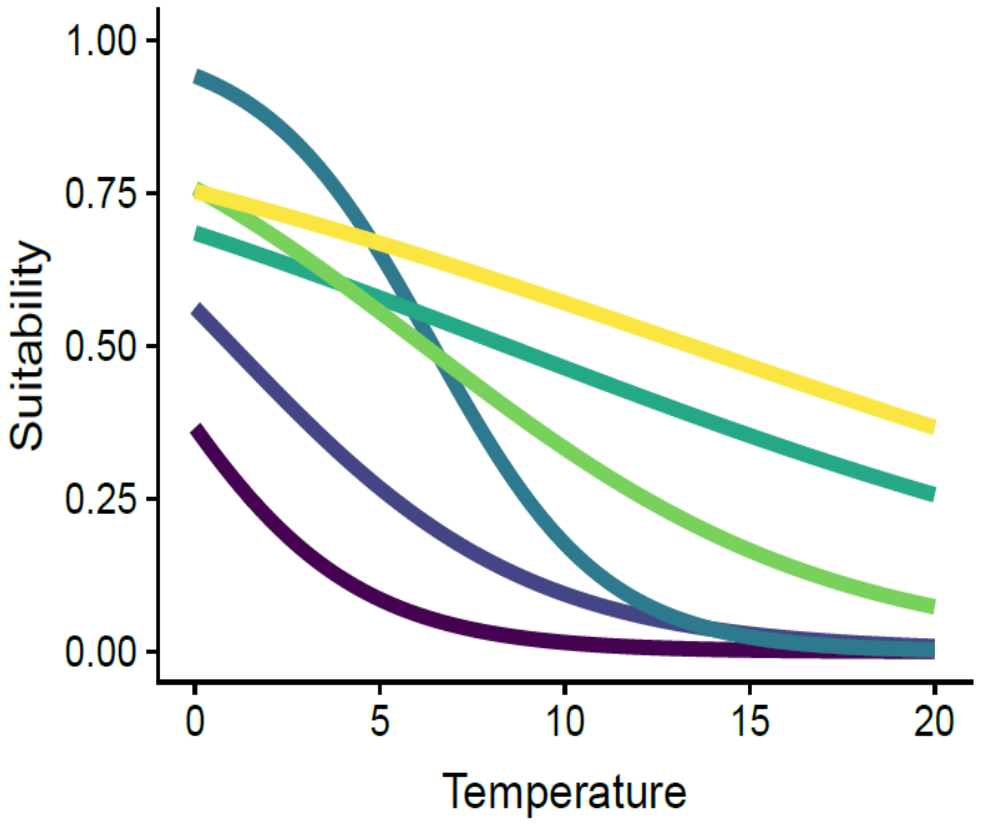
- Accounts for differentiation/LA

Splitting: fitting independent models



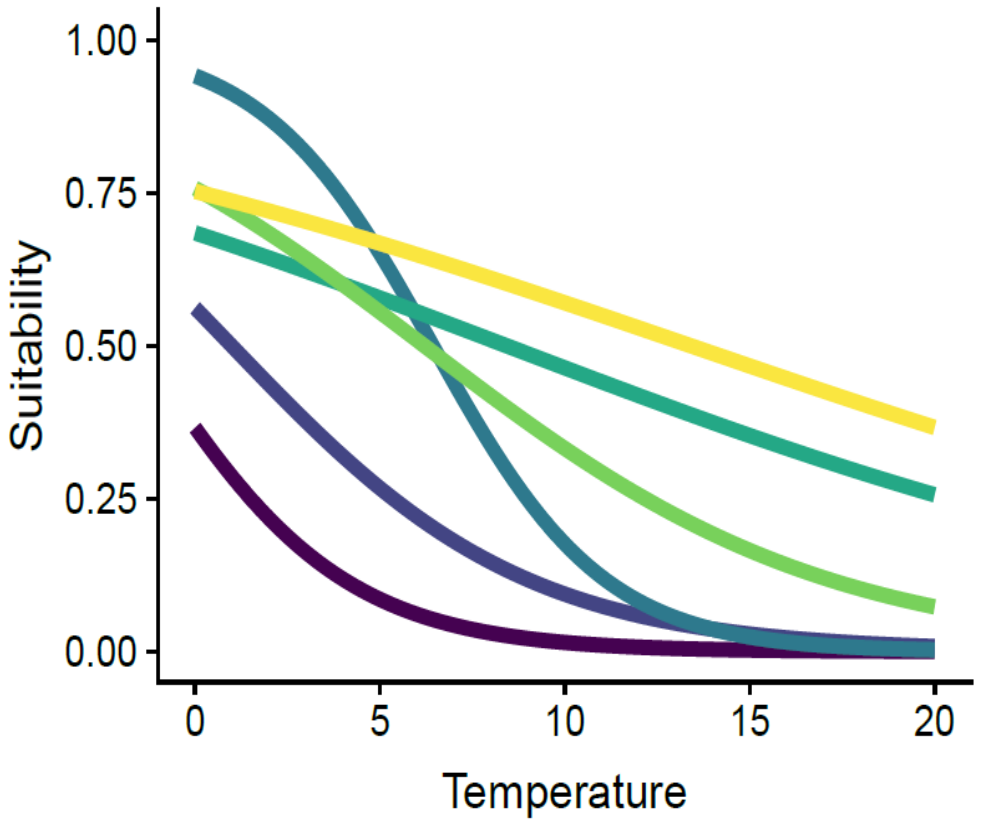
- Accounts for differentiation/LA
- Data-demanding

Splitting: fitting independent models



- Accounts for differentiation/LA
- Data-demanding
- Bias/Overfitting

Splitting: fitting independent models

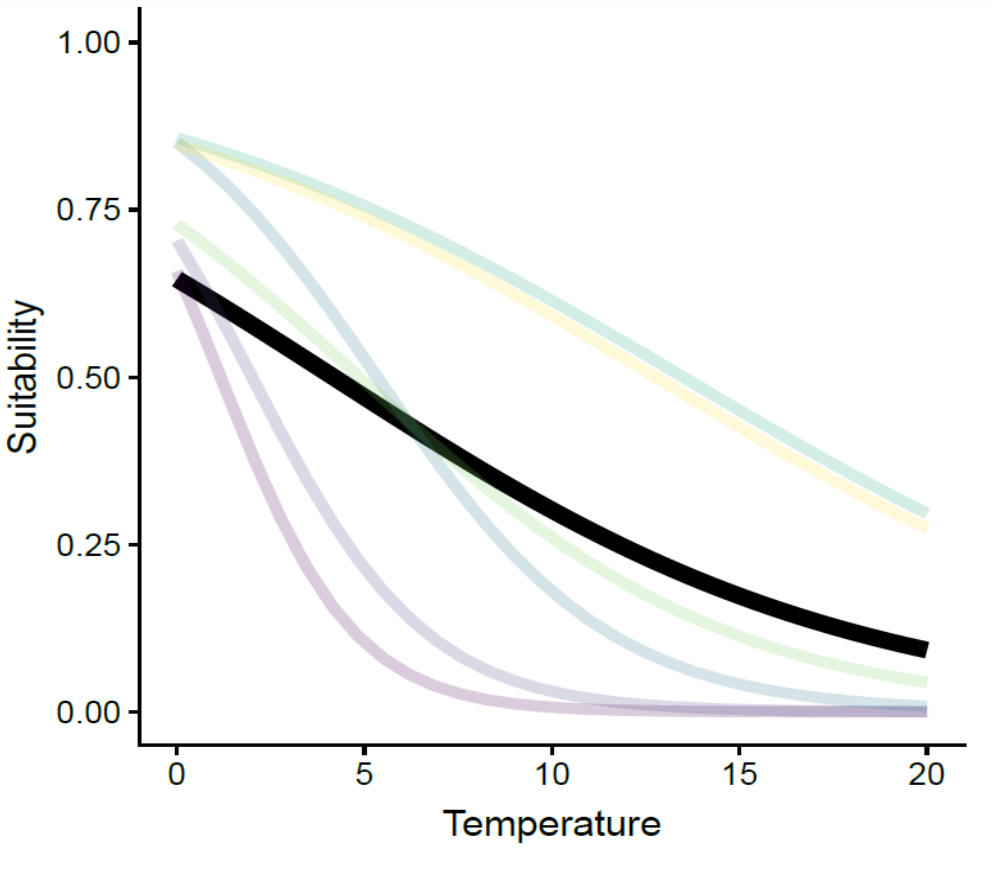


- Accounts for differentiation/LA
- Data-demanding
- Bias/Overfitting
- Where to split?

Lumping

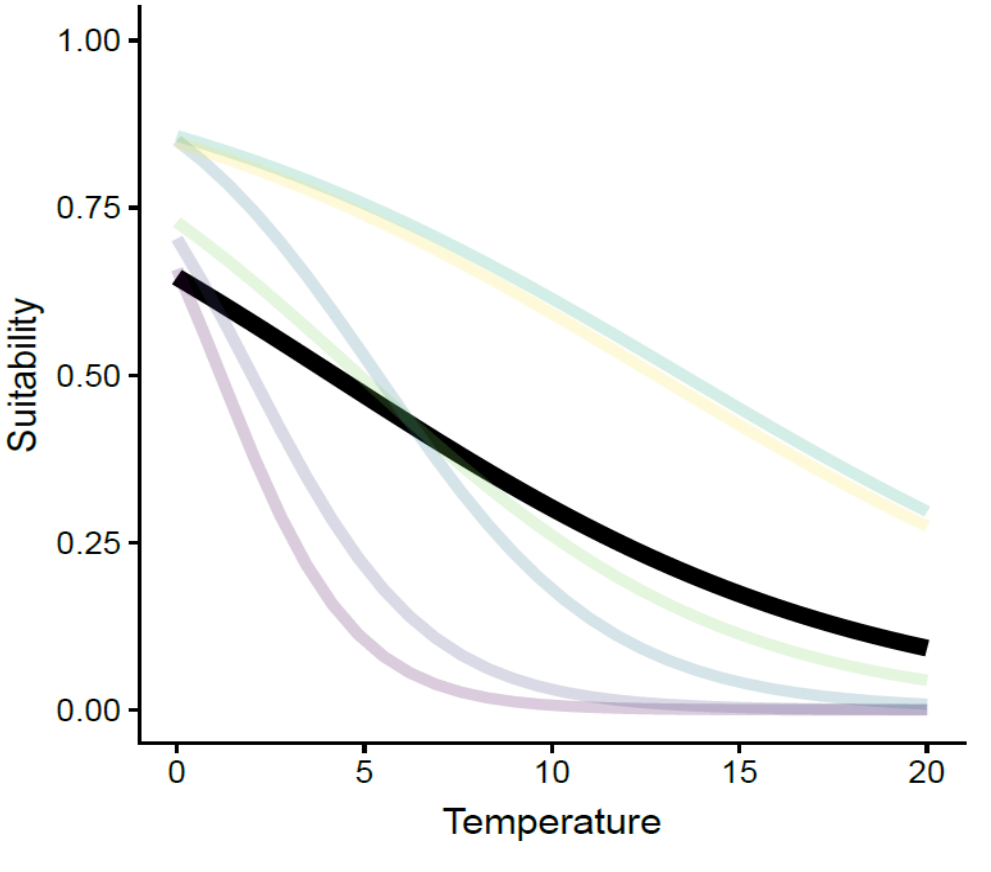
Lumping

Lumping: model taxa as a group



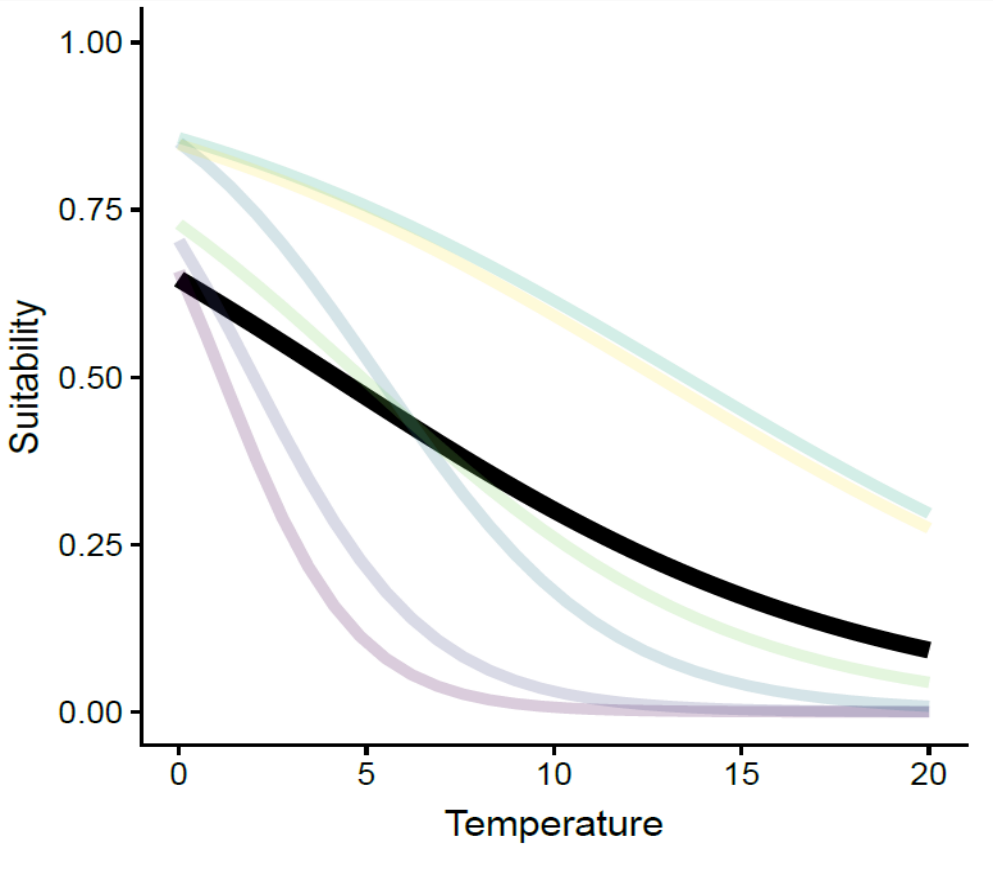
- Niche conservatism

Lumping: model taxa as a group



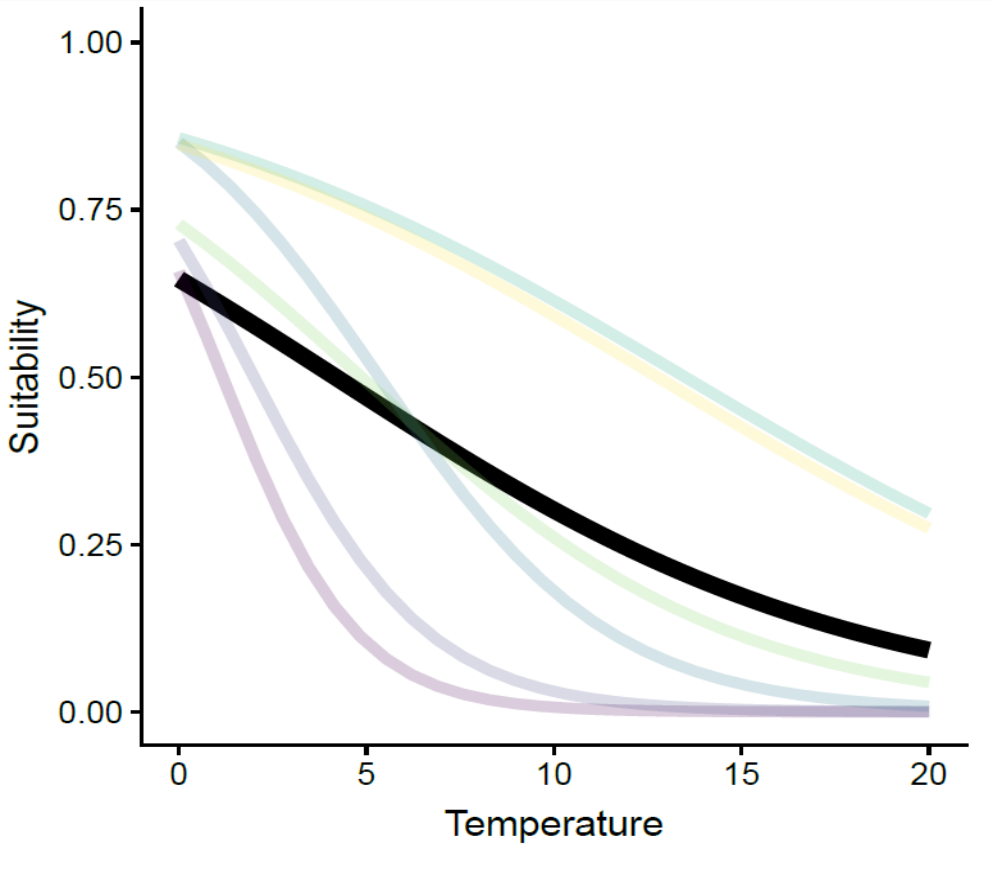
- Niche conservatism
- Cryptic taxa (fossils)

Lumping: model taxa as a group



- Niche conservatism
- Cryptic taxa (fossils)
- Robust

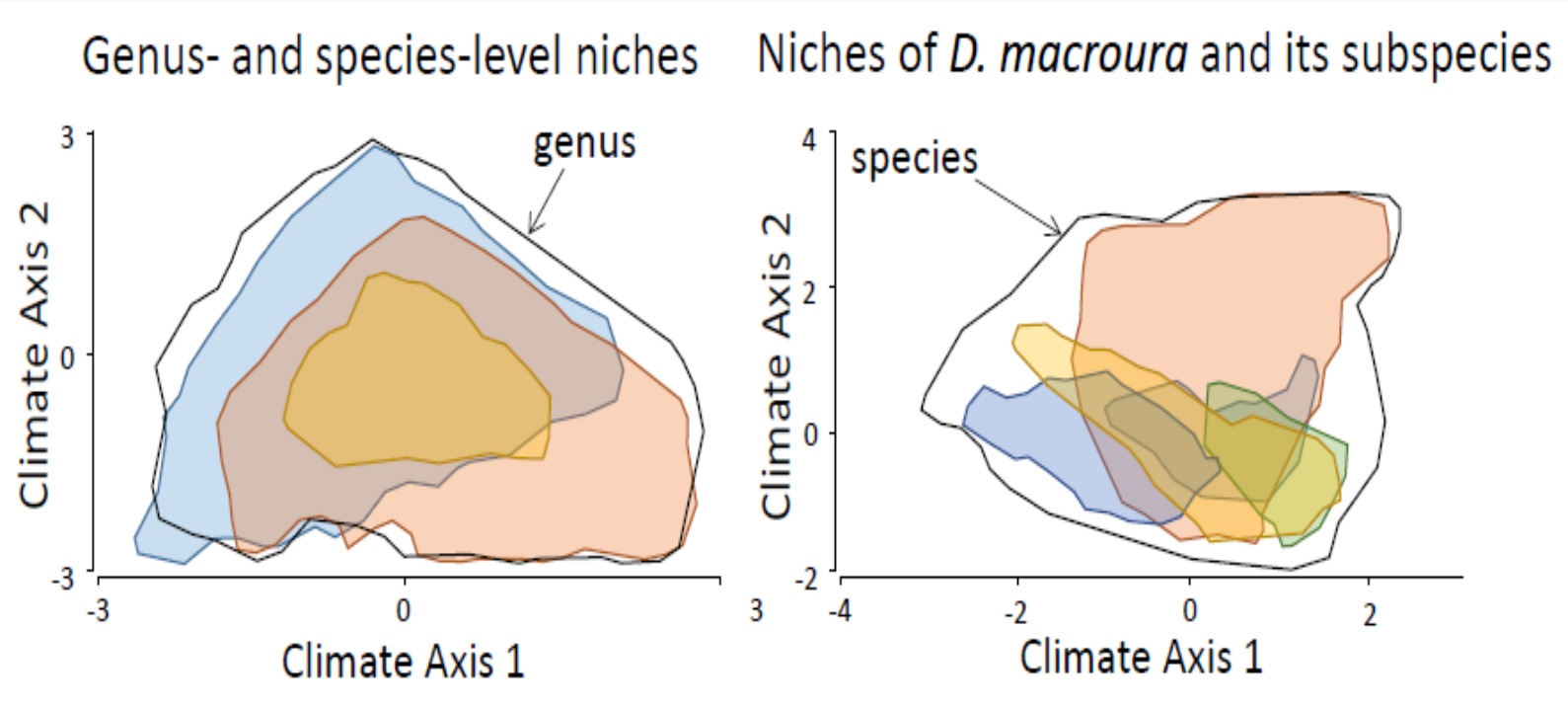
Lumping: model taxa as a group



- Niche conservatism
- Cryptic taxa (fossils)
- Robust
- Underfitting: ignores differentiation

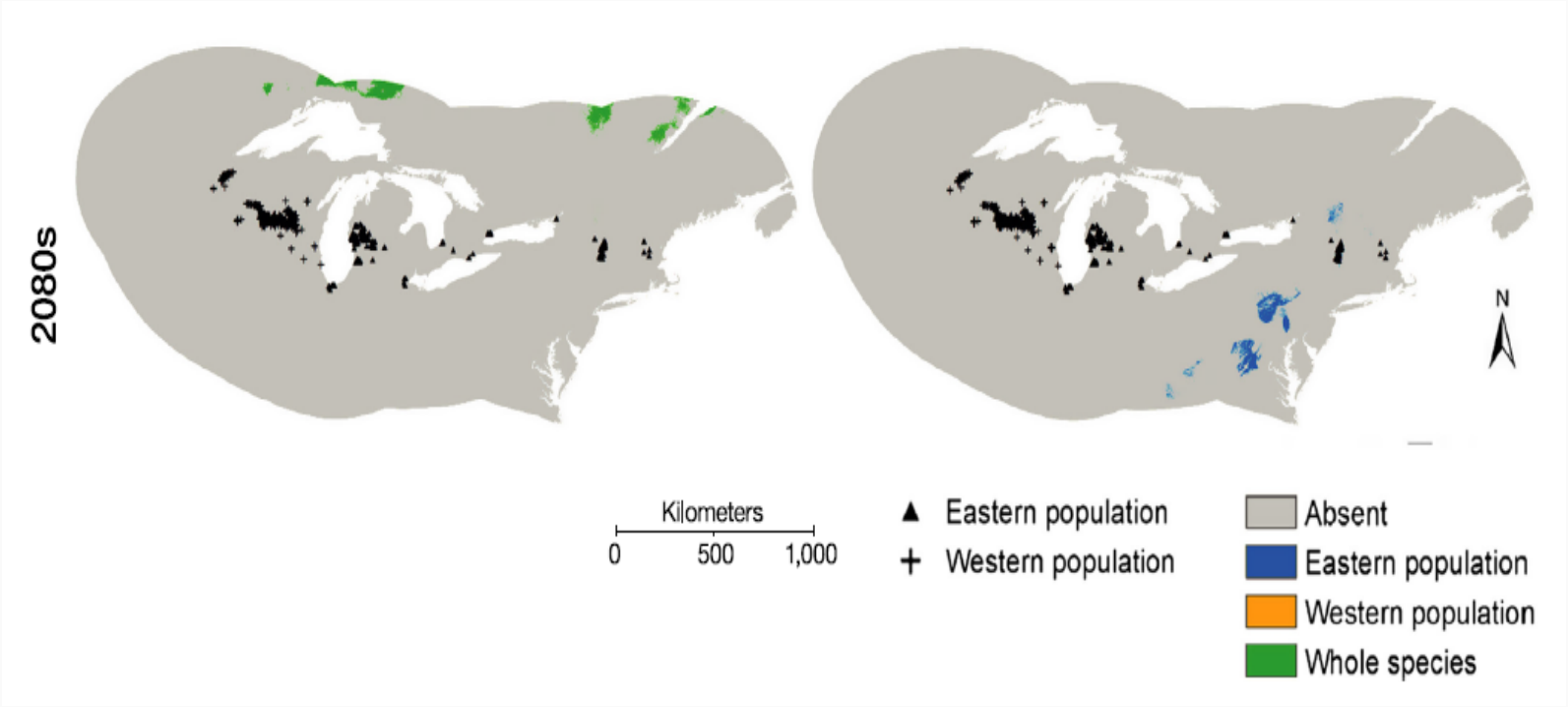
Does it matter?

Splitting/Lumping: niche estimates differ

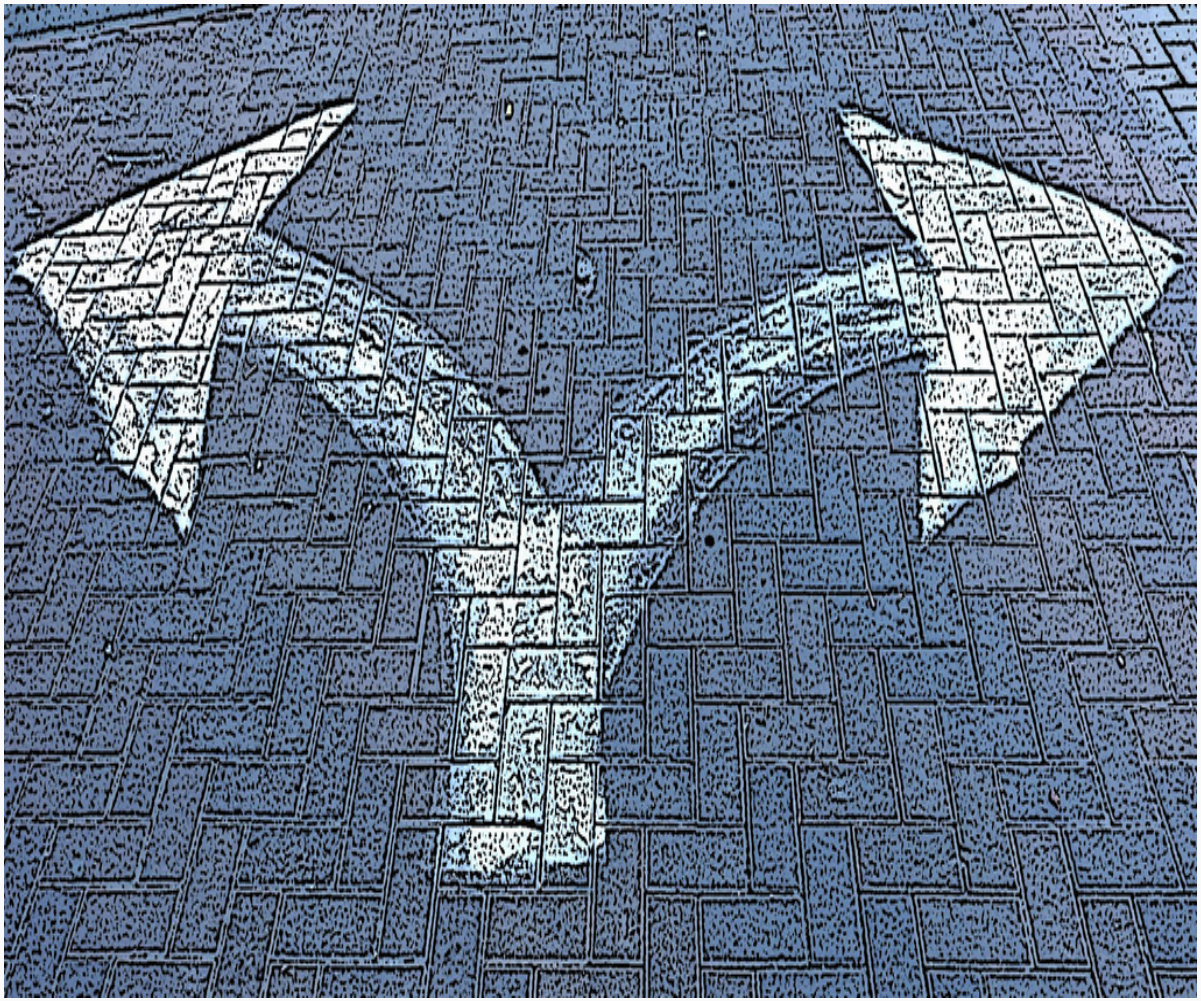


Mota-Vargas & Rojas-Soto 2016

Splitting/Lumping: projected distributions differ



Hallfors et al. 2016



Partial pooling

Partial pooling

Partial pooling: many methods

Making more out of sparse data: hierarchical modeling of species communities

Otso Ovaskainen[1,3] and Janne Soininen[2]

Generalized linear mixed models for phylogenetic analyses of community structure

Anthony R. Ives[1,3] and Matthew R. Helmus[2]

Finite Mixture of Regression Modeling for High-Dimensional Count and Biomass Data in Ecology

Piers K. Dunstan, Scott D. Foster, Francis K.C. Hui, and David I. Warton

Spatio-phylogenetic multispecies distribution models

Arne Kaldhusdal[1], Roland Brandl[2], Jörg Müller[3,4], Lisa Möst[1] and Torsten Hothorn[5,*]

Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring

James T. Thorson[1,*], James N. Ianelli[2], Elise A. Larsen[3], Leslie Ries[4], Mark D. Scheuerell[5], Cody Szuwalski[6,7] and Elise F. Zipkin[8,9]

Combining phylogeny and co-occurrence to improve single species distribution models

Ignacio Morales-Castilla[1,2,3] | T. Jonathan Davies[2,3,4] | William D. Pearse[3,5] | Pedro Peres-Neto[3,6]

Partial pooling

Estimating all niches through single joint model

- Phylogenetic

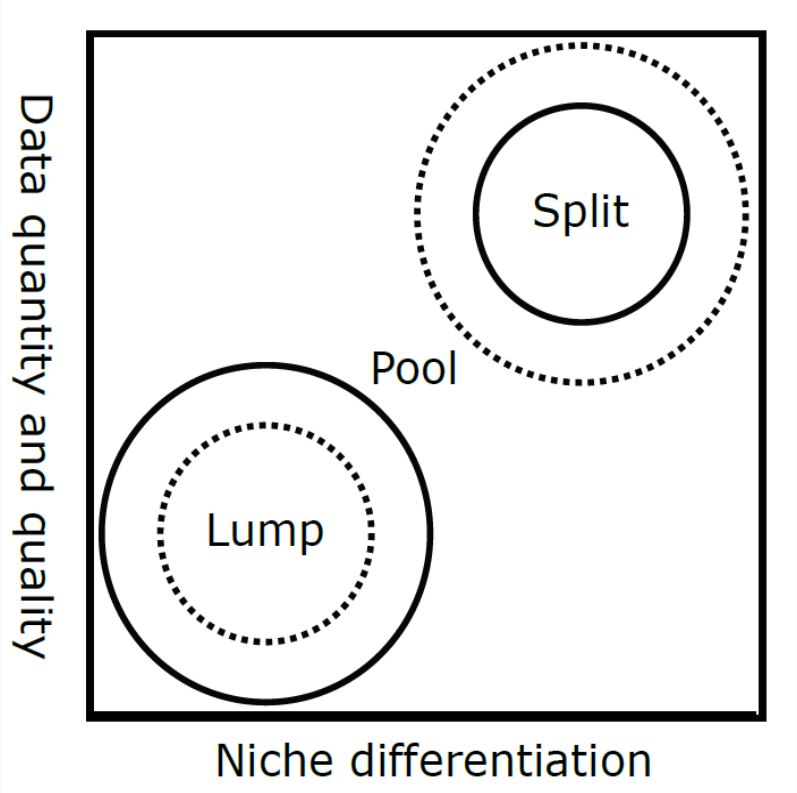
Partial pooling

Estimating all niches through single joint model

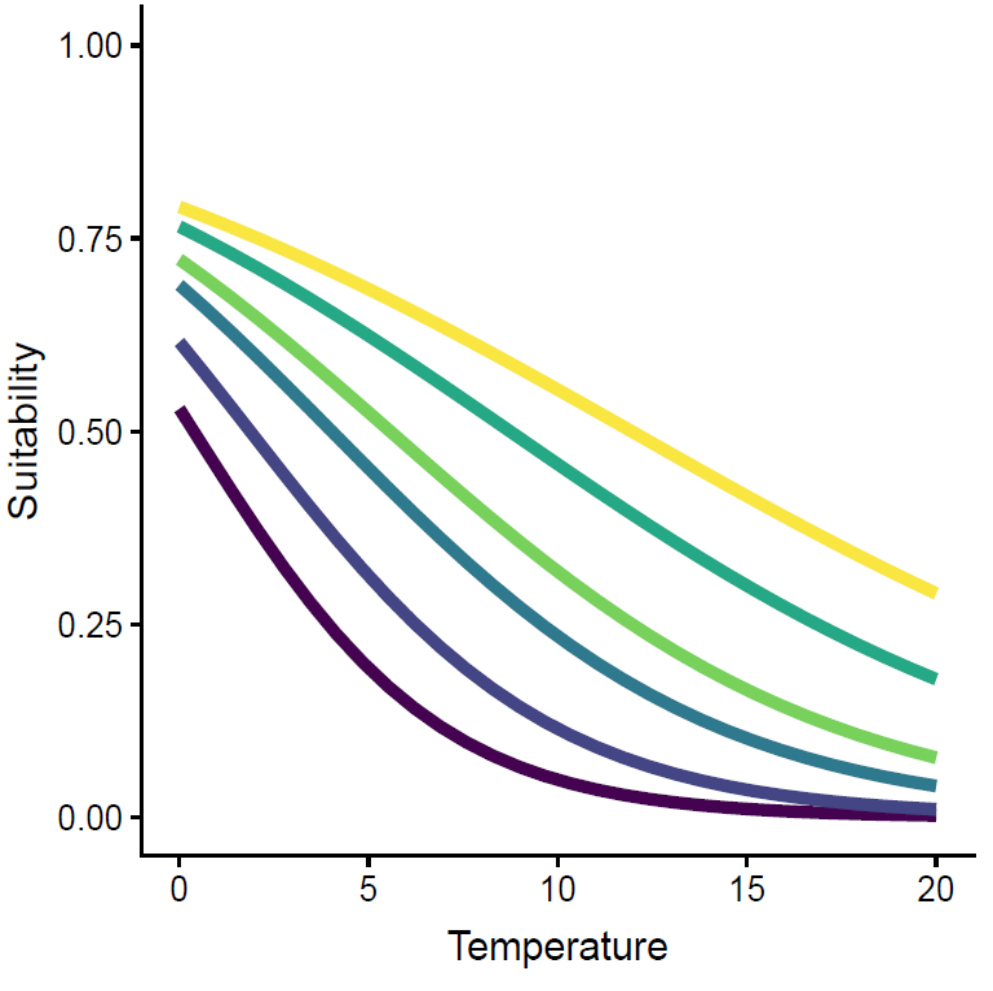
- Phylogenetic
- Non-Phylogenetic

Partial pooling: middle ground

Degree of niche differentiation determined by data (and priors if Bayesian)

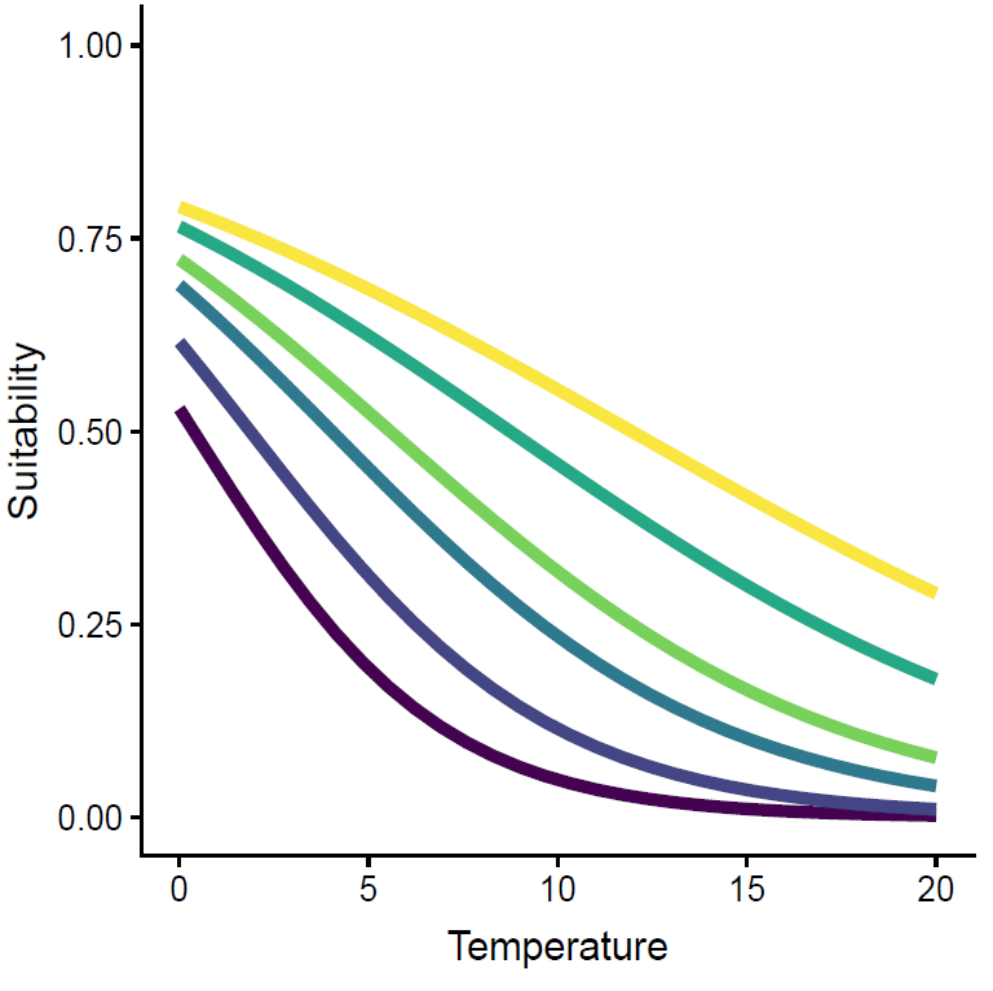


Partial pooling: pros and cons



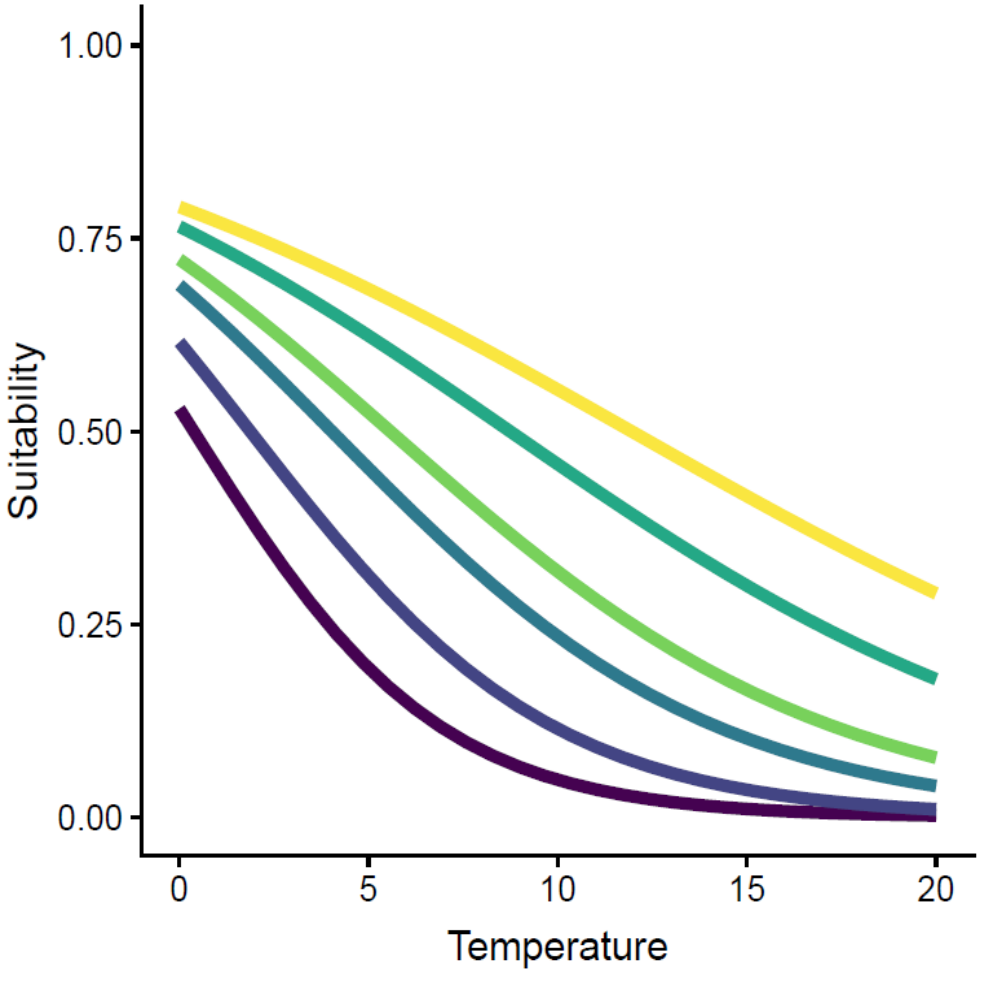
- 'Borrowing strength'

Partial pooling: pros and cons



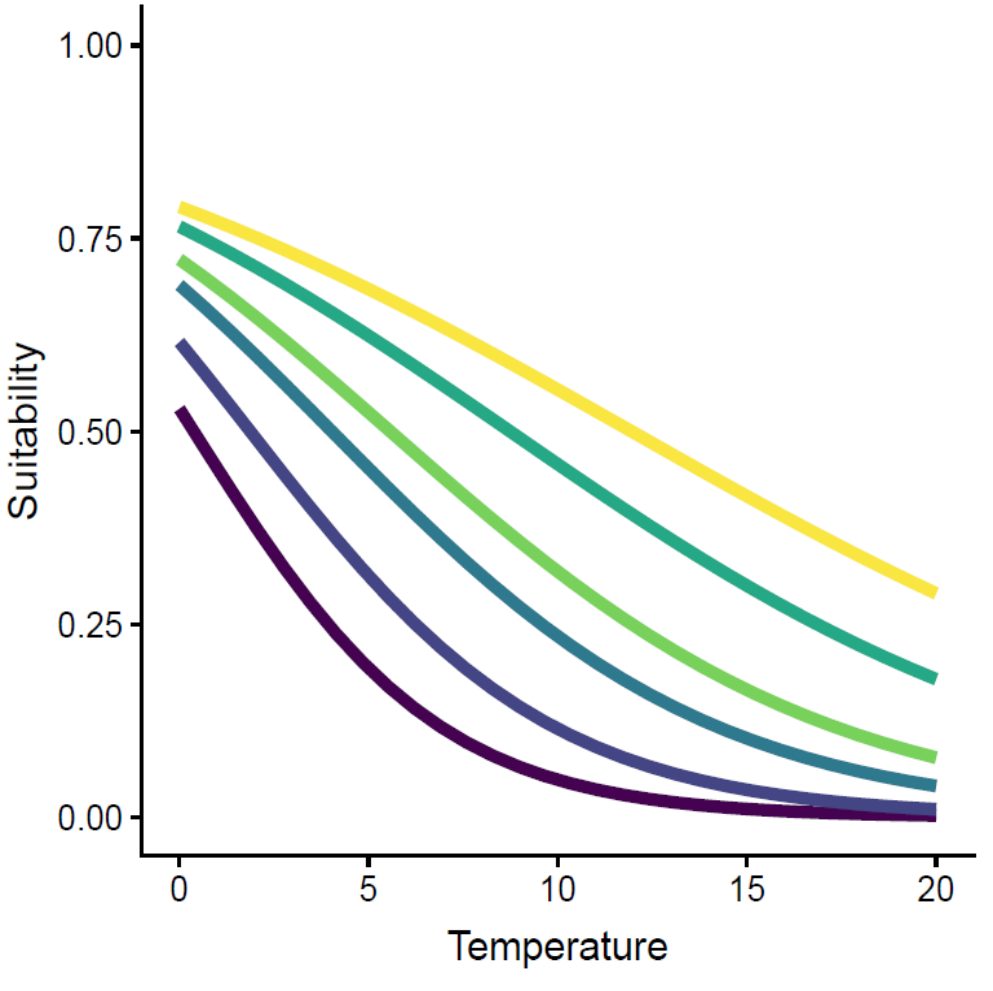
- ‘Borrowing strength’
- Best for rare taxa

Partial pooling: pros and cons



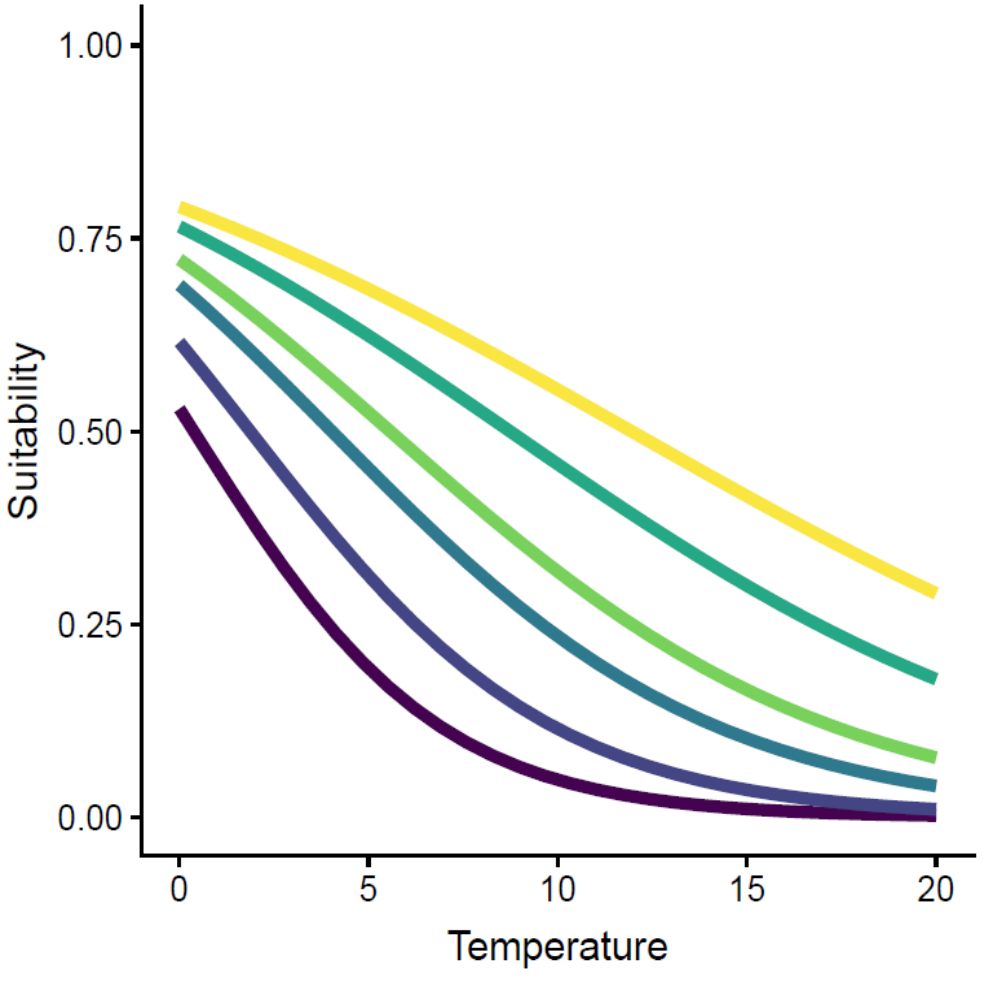
- 'Borrowing strength'
- Best for rare taxa
- Reduce overfitting

Partial pooling: pros and cons



- 'Borrowing strength'
- Best for rare taxa
- Reduce overfitting
- Complex models

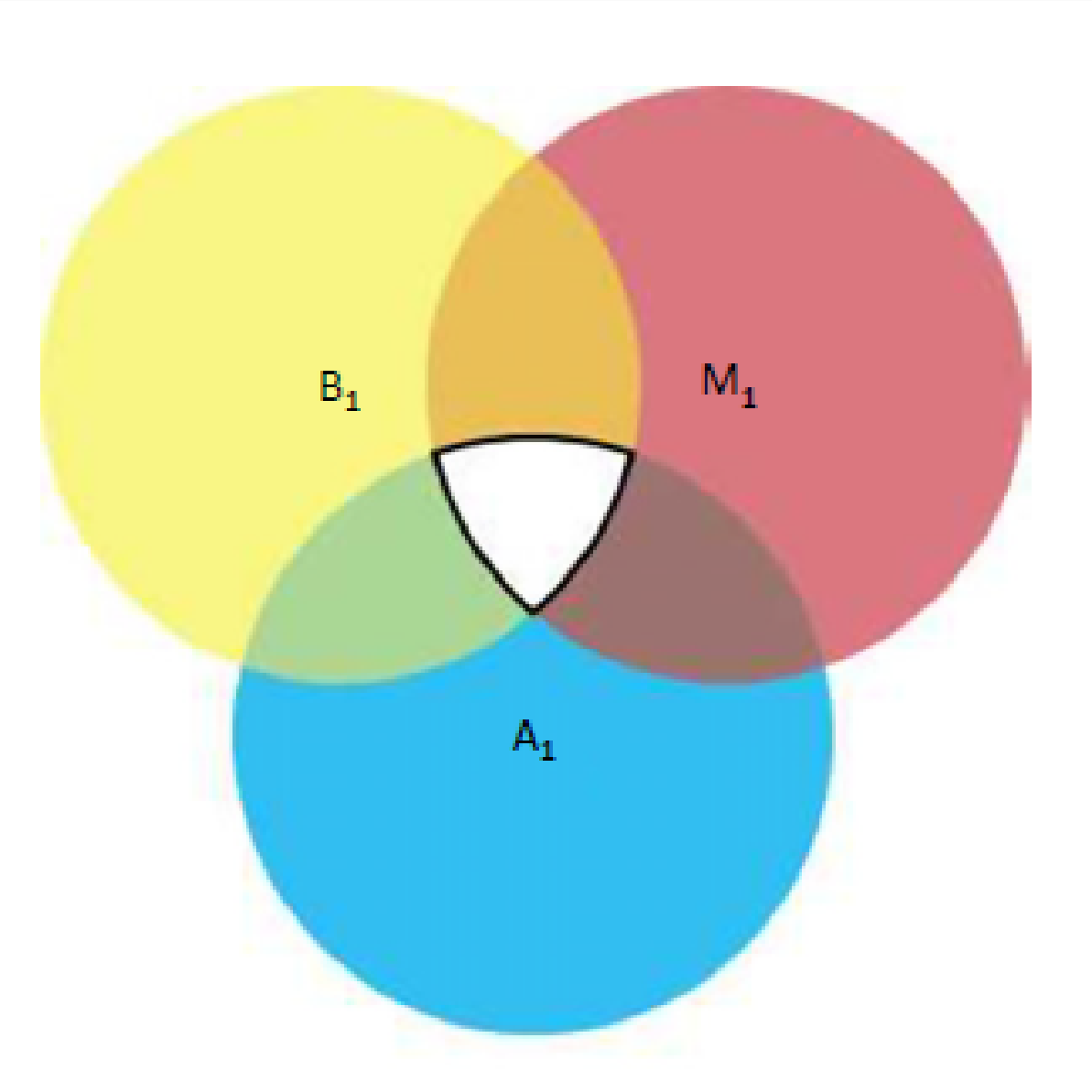
Partial pooling: pros and cons



- 'Borrowing strength'
- Best for rare taxa
- Reduce overfitting
- Complex models
- Oversmoothing divergent taxa

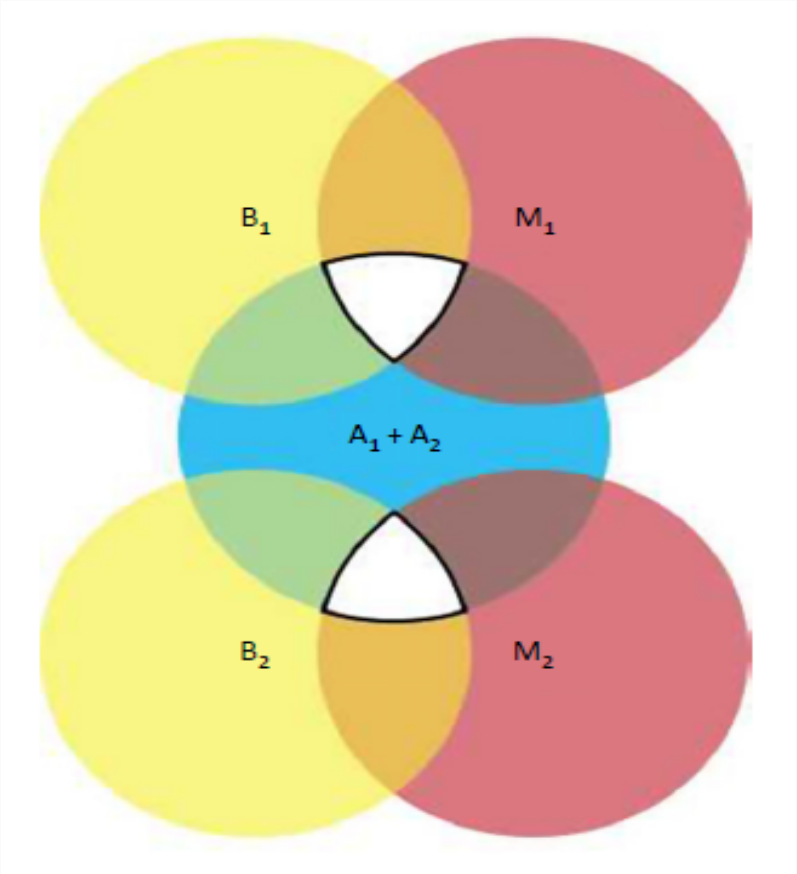
When to use what?

The BAM diagram



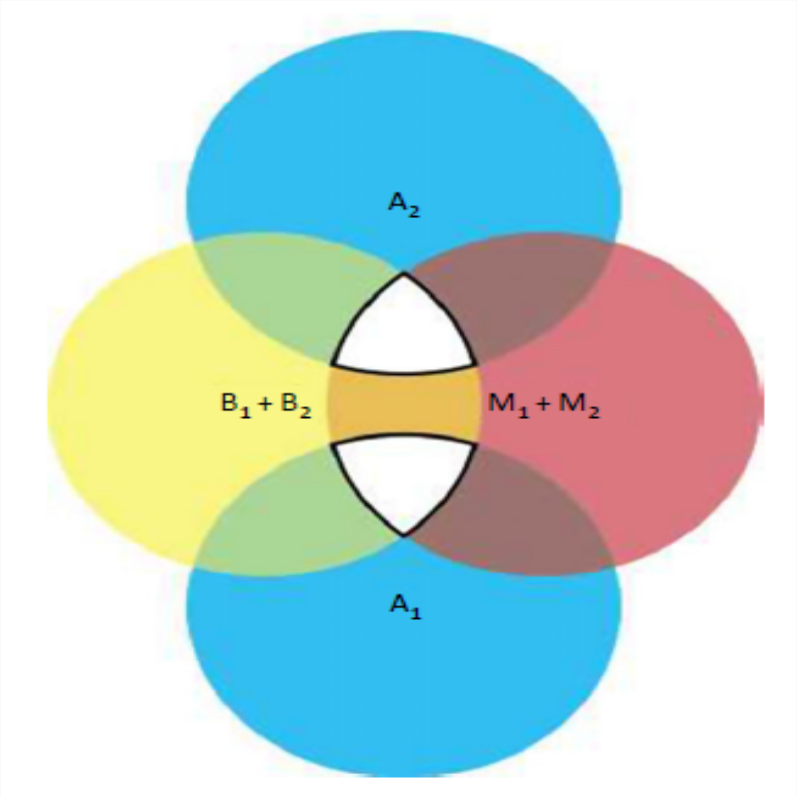
Partial pooling/lumping can reduce overfitting. . .

e.g. if separate ranges actually driven by dispersal (M) or biotic interactions (B), not niche differences (A)



Splitting can reduce bias. . .

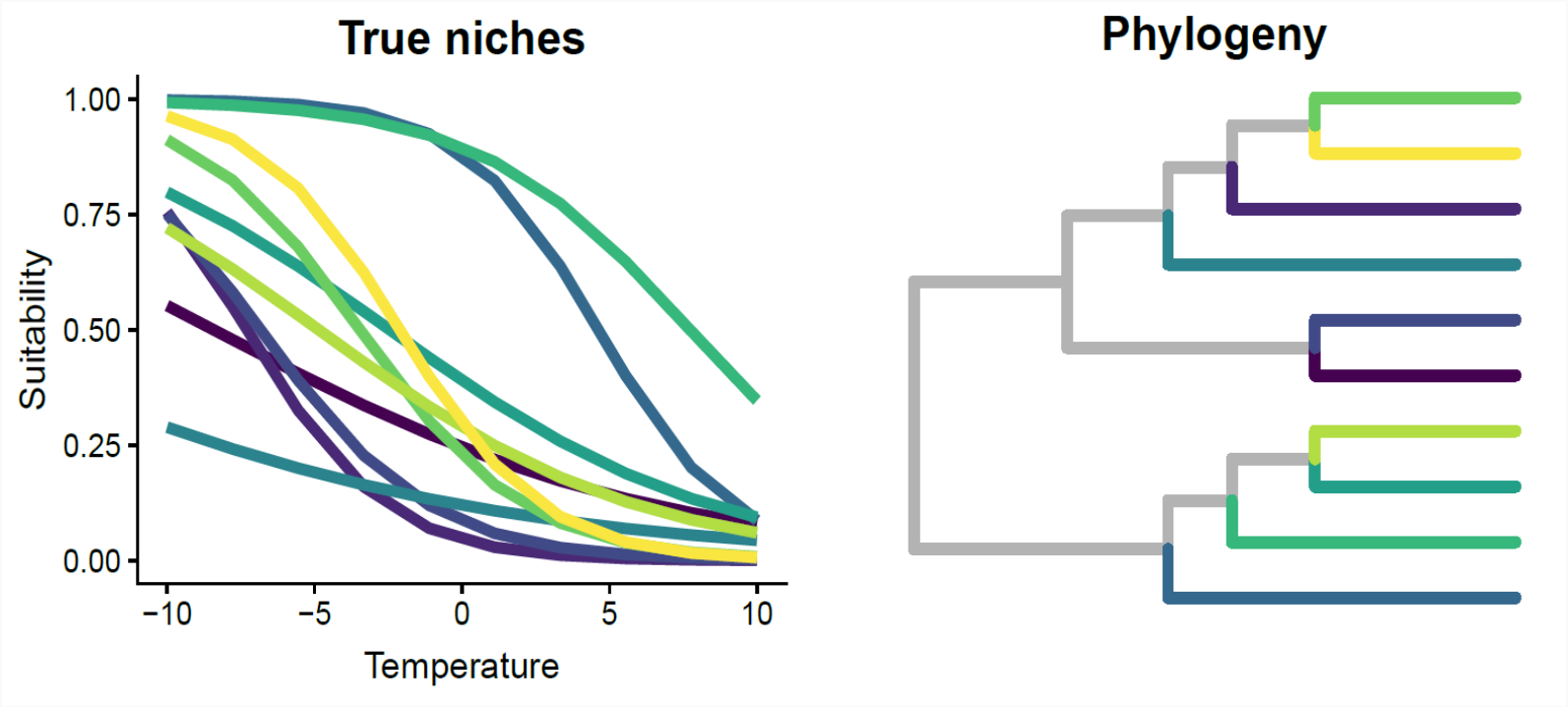
e.g. if co-occurrence driven by biotic interactions (B) or dispersal (M), not niche differences (A)



So... when to use what?

Lets's ask our computer: simulations

Simulated niches and phylogeny



Model comparison

- **Splitting**: GLM

Model comparison

- **Splitting**: GLM
  - `glm(presabs ~ env)` (for each taxon)

Model comparison

- **Splitting**: GLM
  - `glm(presabs ~ env)` (for each taxon)
- **Lumping**: GLM

Model comparison

- **Splitting**: GLM
  - `glm(presabs ~ env)` (for each taxon)
- **Lumping**: GLM
  - `glm(presabs ~ env)` (all taxa together)

Model comparison

- **Splitting**: GLM
  - `glm(presabs ~ env)` (for each taxon)
- **Lumping**: GLM
  - `glm(presabs ~ env)` (all taxa together)
- **GLMM** (partial pooling)

Model comparison

- **Splitting**: GLM
  - `glm(presabs ~ env)` (for each taxon)
- **Lumping**: GLM
  - `glm(presabs ~ env)` (all taxa together)
- **GLMM** (partial pooling)
  - `lme4::glmer(presabs ~ env + (1+env | taxon))`

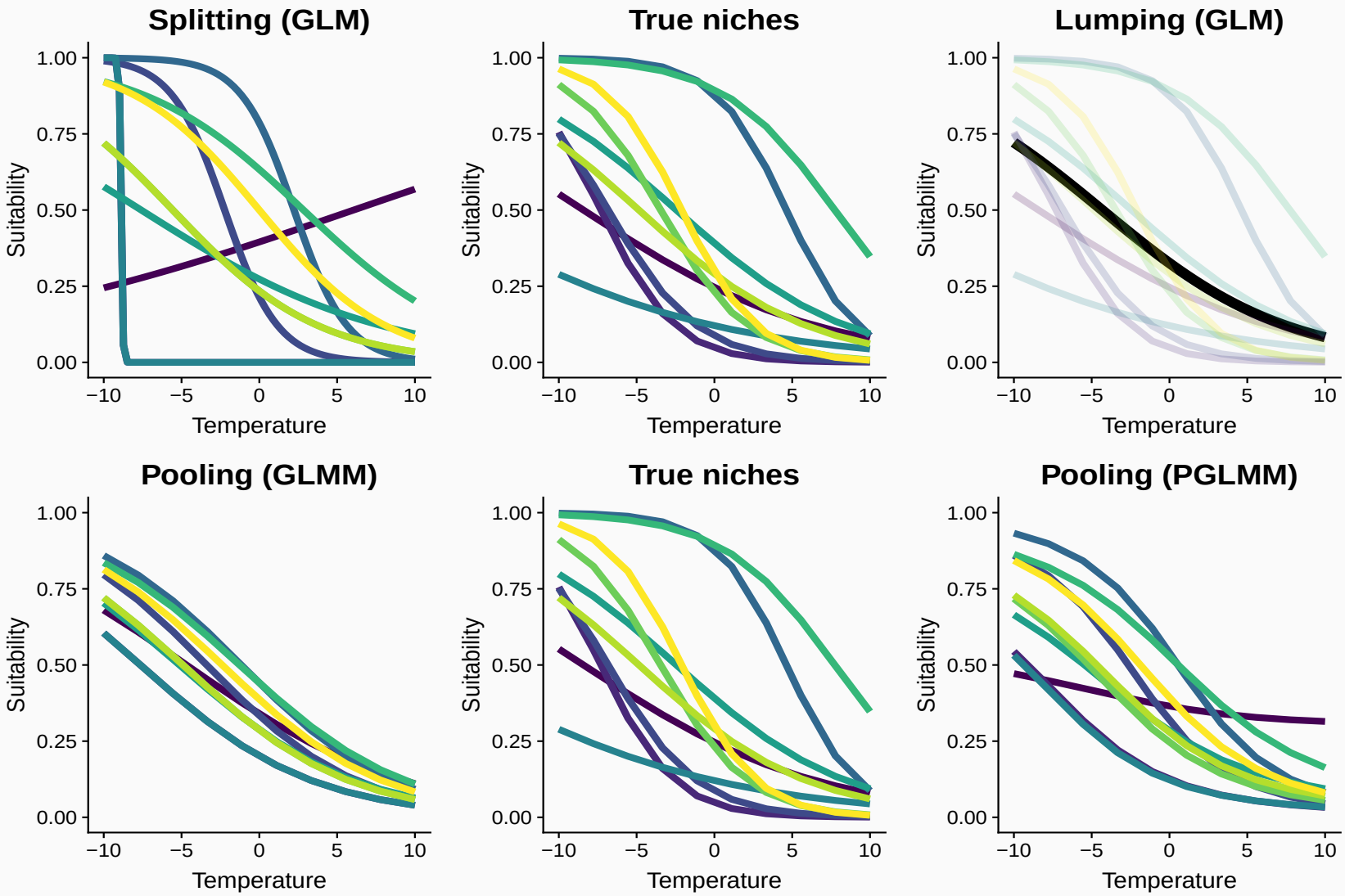
Model comparison

- **Splitting**: GLM
  - `glm(presabs ~ env)` (for each taxon)
- **Lumping**: GLM
  - `glm(presabs ~ env)` (all taxa together)
- **GLMM** (partial pooling)
  - `lme4::glmer(presabs ~ env + (1+env | taxon))`
- **PGLMM**: Phylogenetic GLMM (partial pooling)

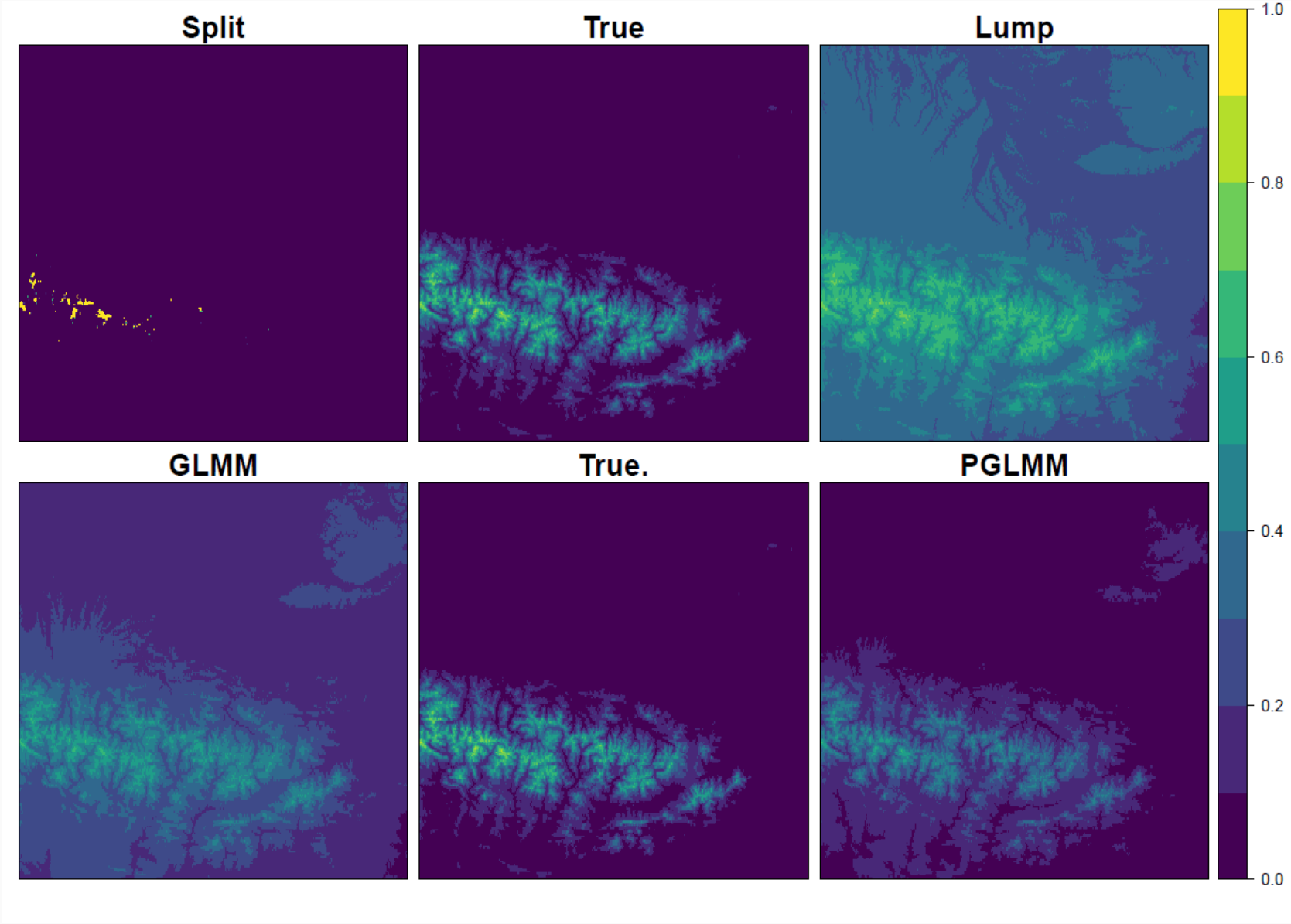
Model comparison

- **Splitting**: GLM
  - `glm(presabs ~ env)` (for each taxon)
- **Lumping**: GLM
  - `glm(presabs ~ env)` (all taxa together)
- **GLMM** (partial pooling)
  - `lme4::glmer(presabs ~ env + (1+env | taxon))`
- **PGLMM**: Phylogenetic GLMM (partial pooling)
  - `brms::brm(presabs ~ env + (1 + env | taxon) + (1 + env | phylo))`

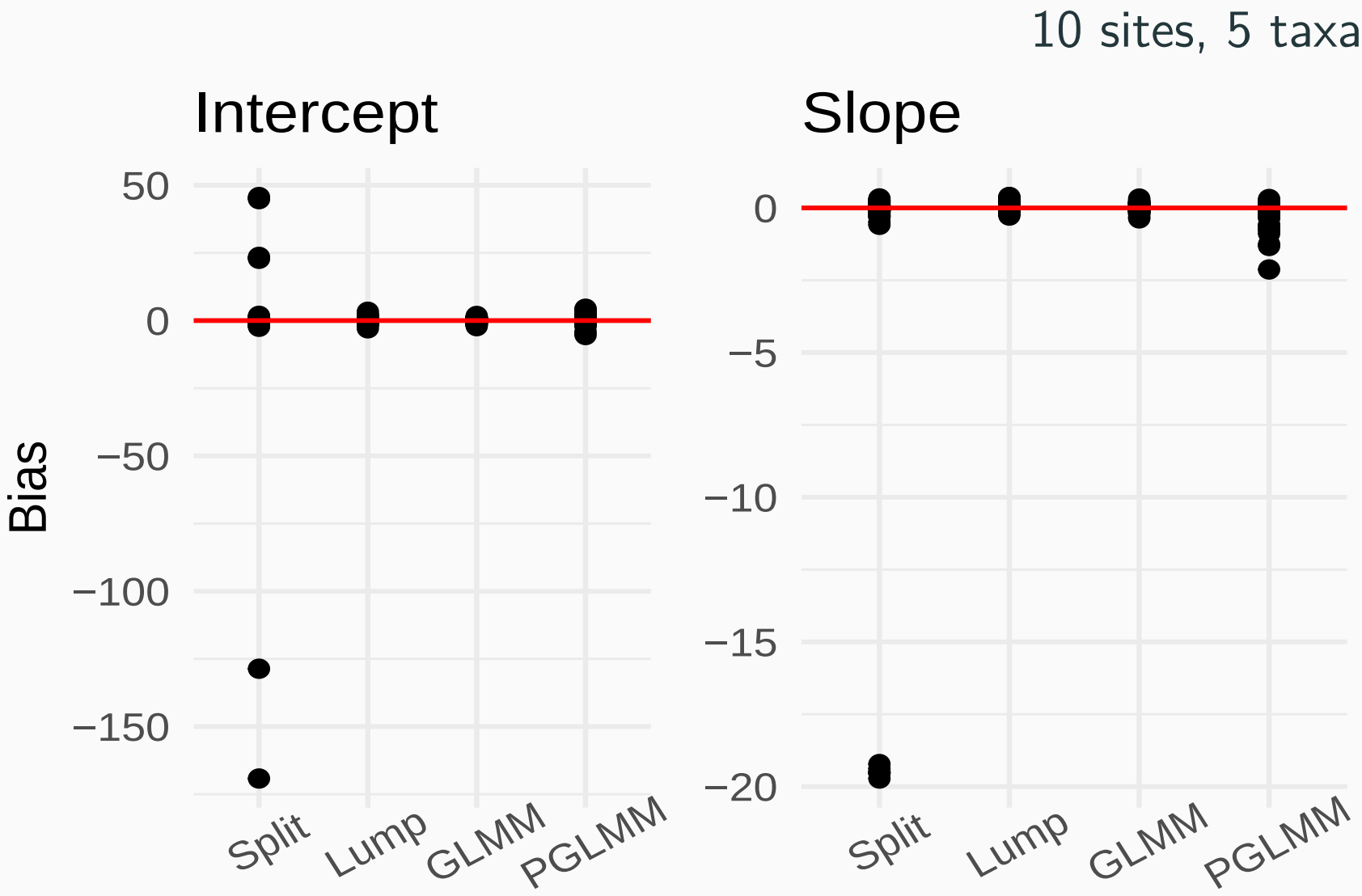
Widely different niche estimates



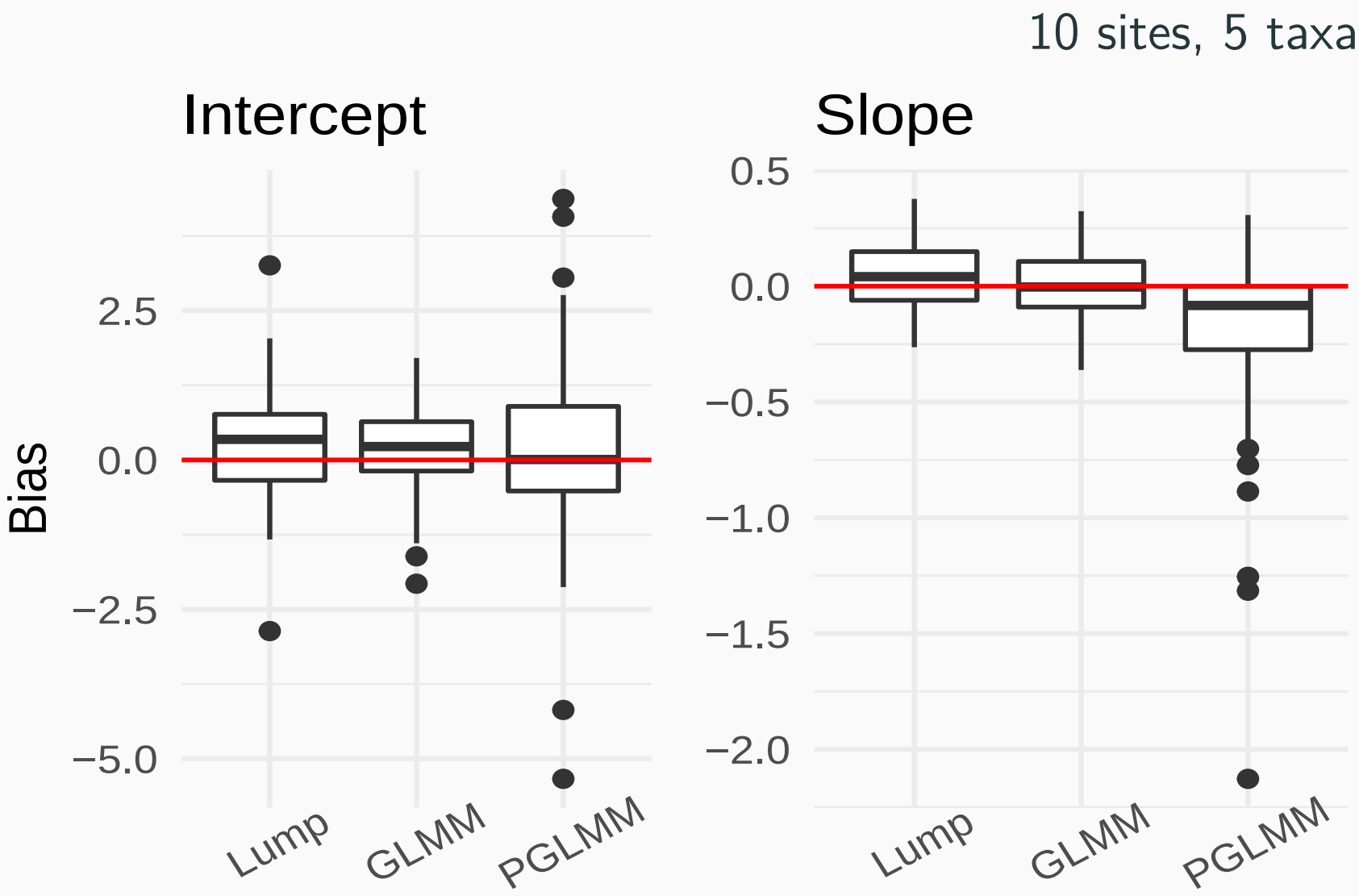
Different geographic projections



FEW sites, FEW taxa

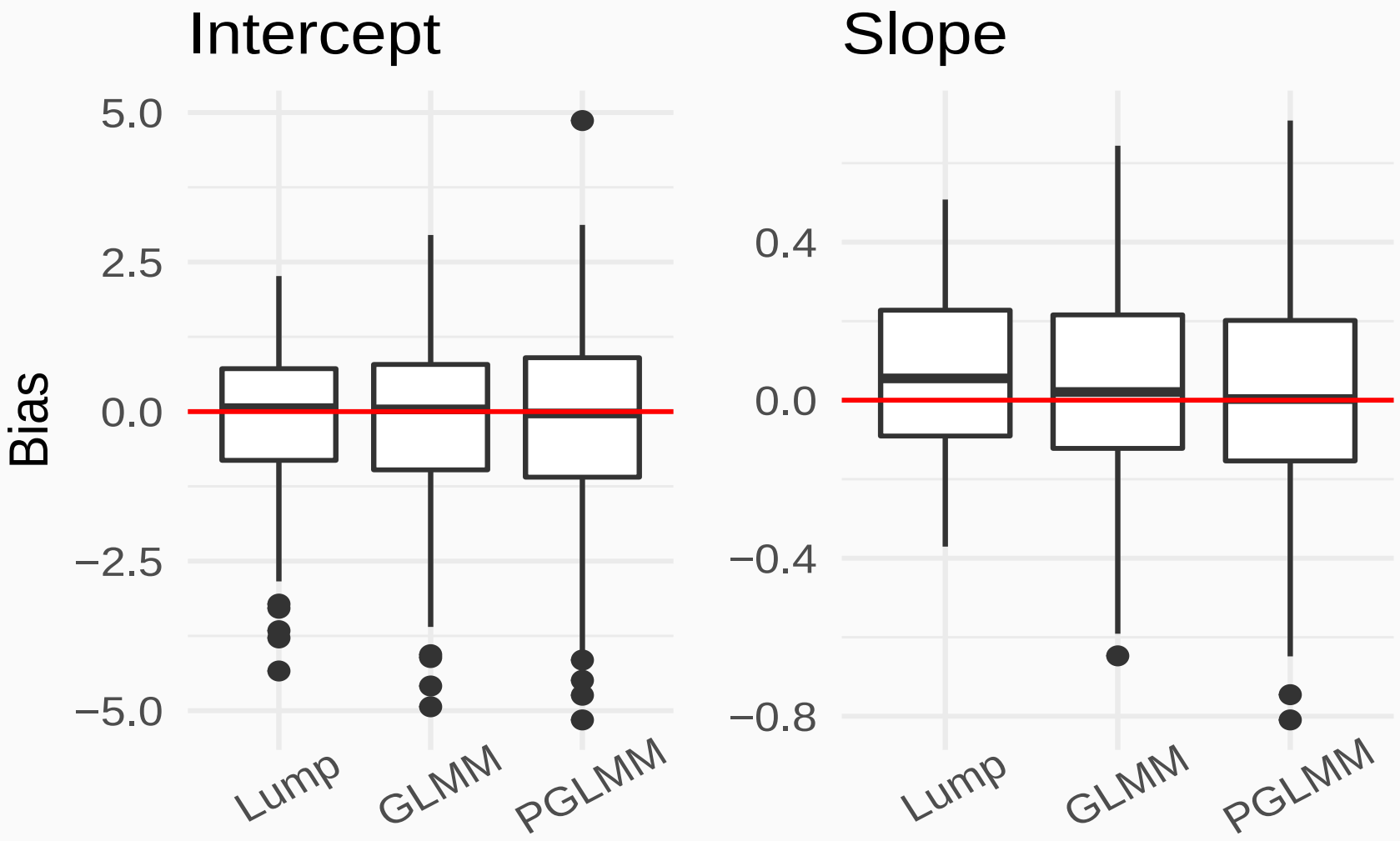


FEW sites, FEW taxa

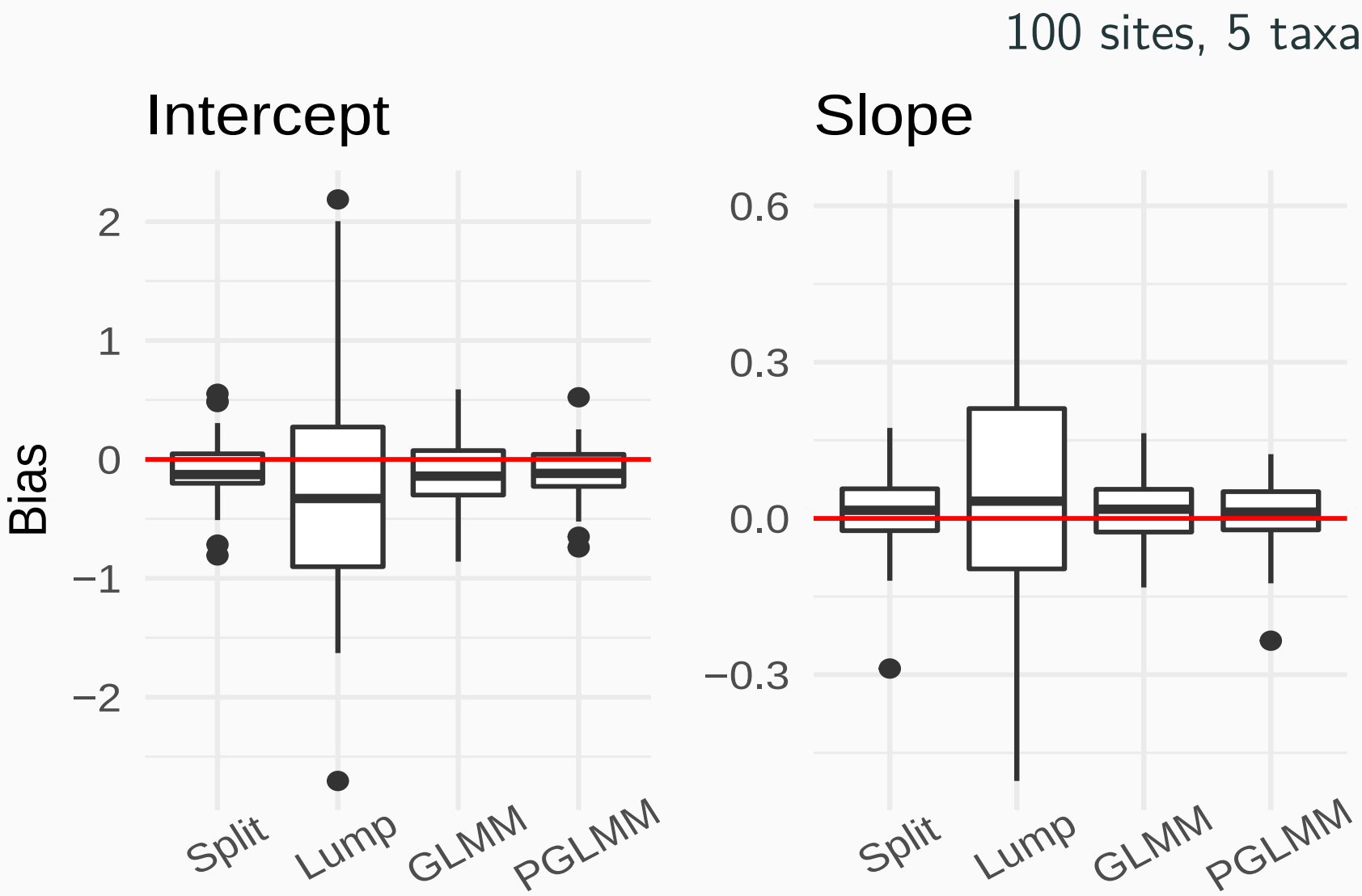


FEW sites, MANY taxa

10 sites, 30 taxa

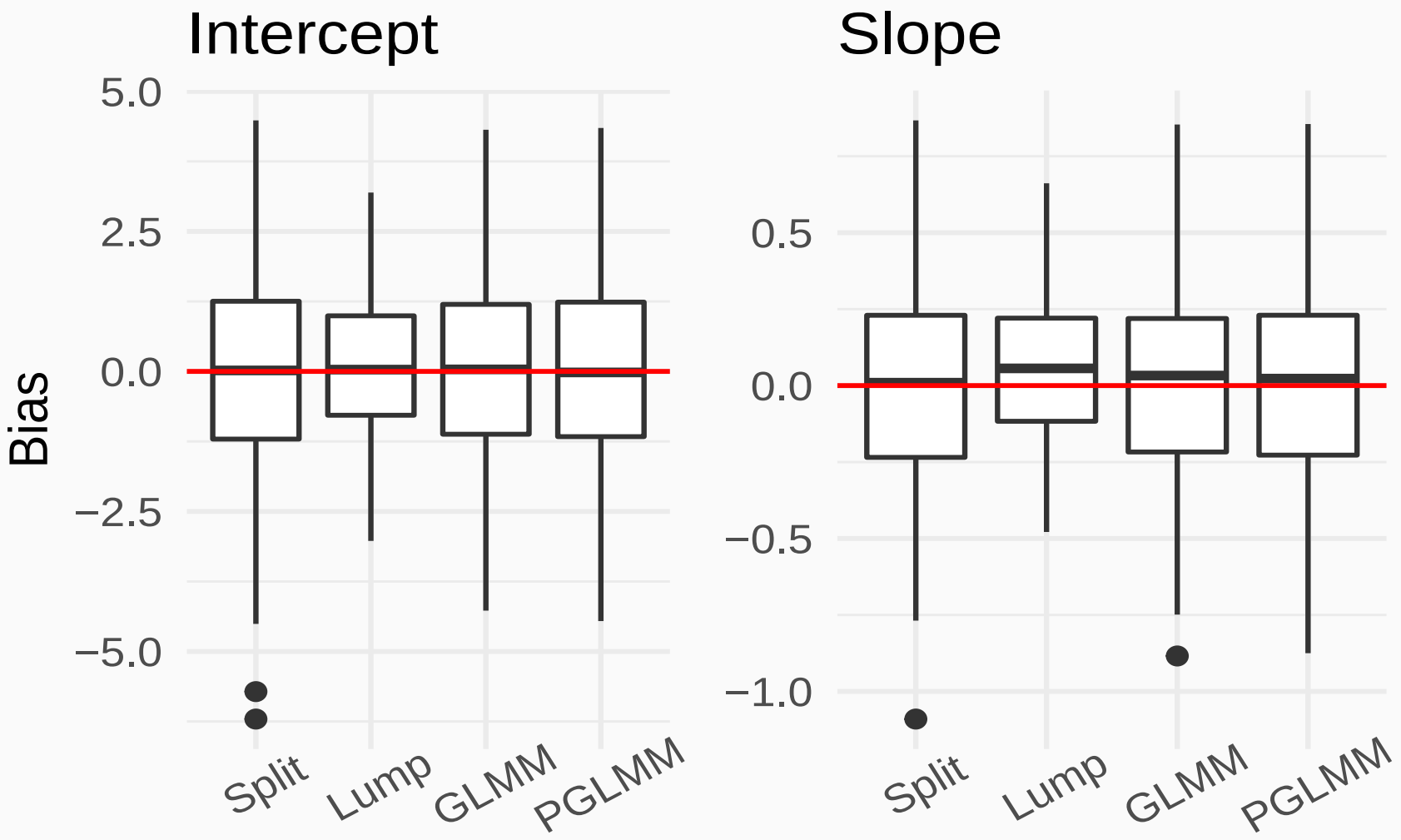


MANY sites, FEW taxa



MANY sites, MANY taxa

100 sites, 30 taxa

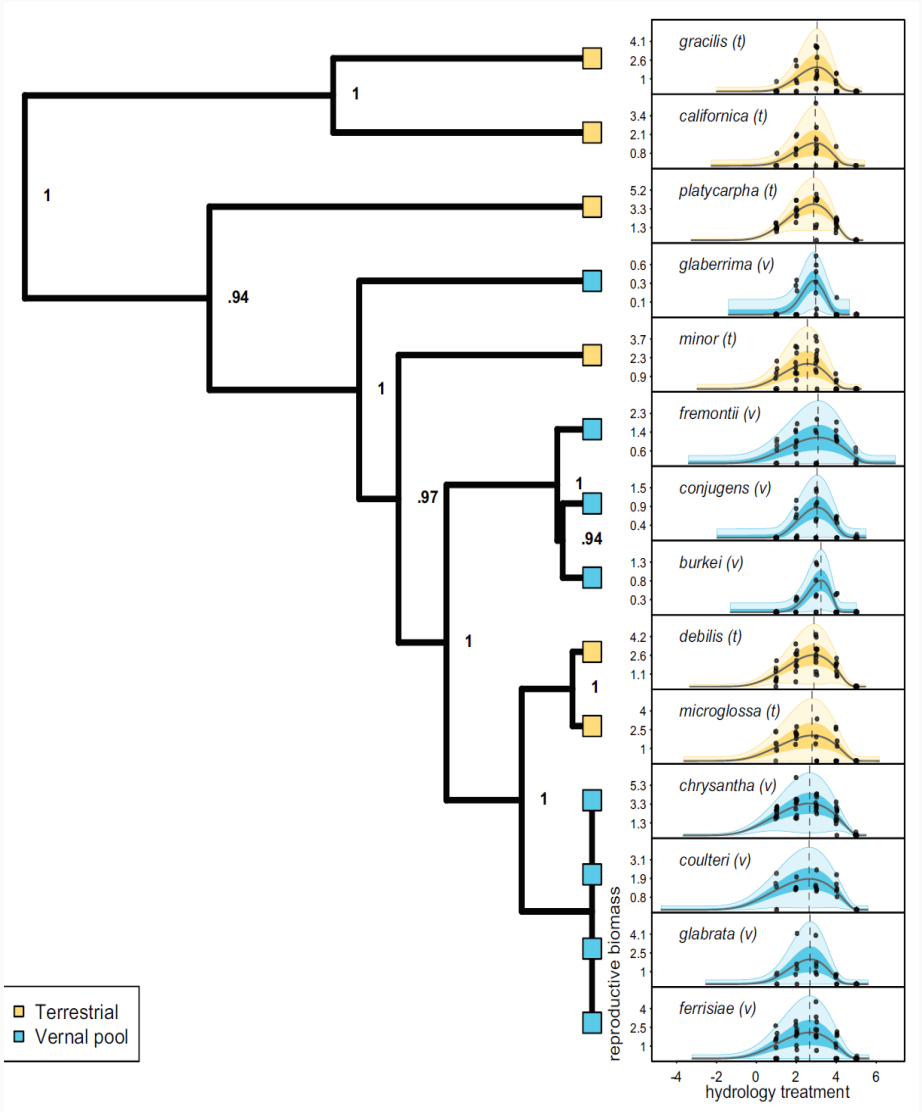


No free lunch

Data quantity and quality

	Low	High
Splitting	Fails	Good
Lumping	Good	Depends on data quality
Partial pooling	Good, but complex models require more data	Good

A bright future for partial pooling?



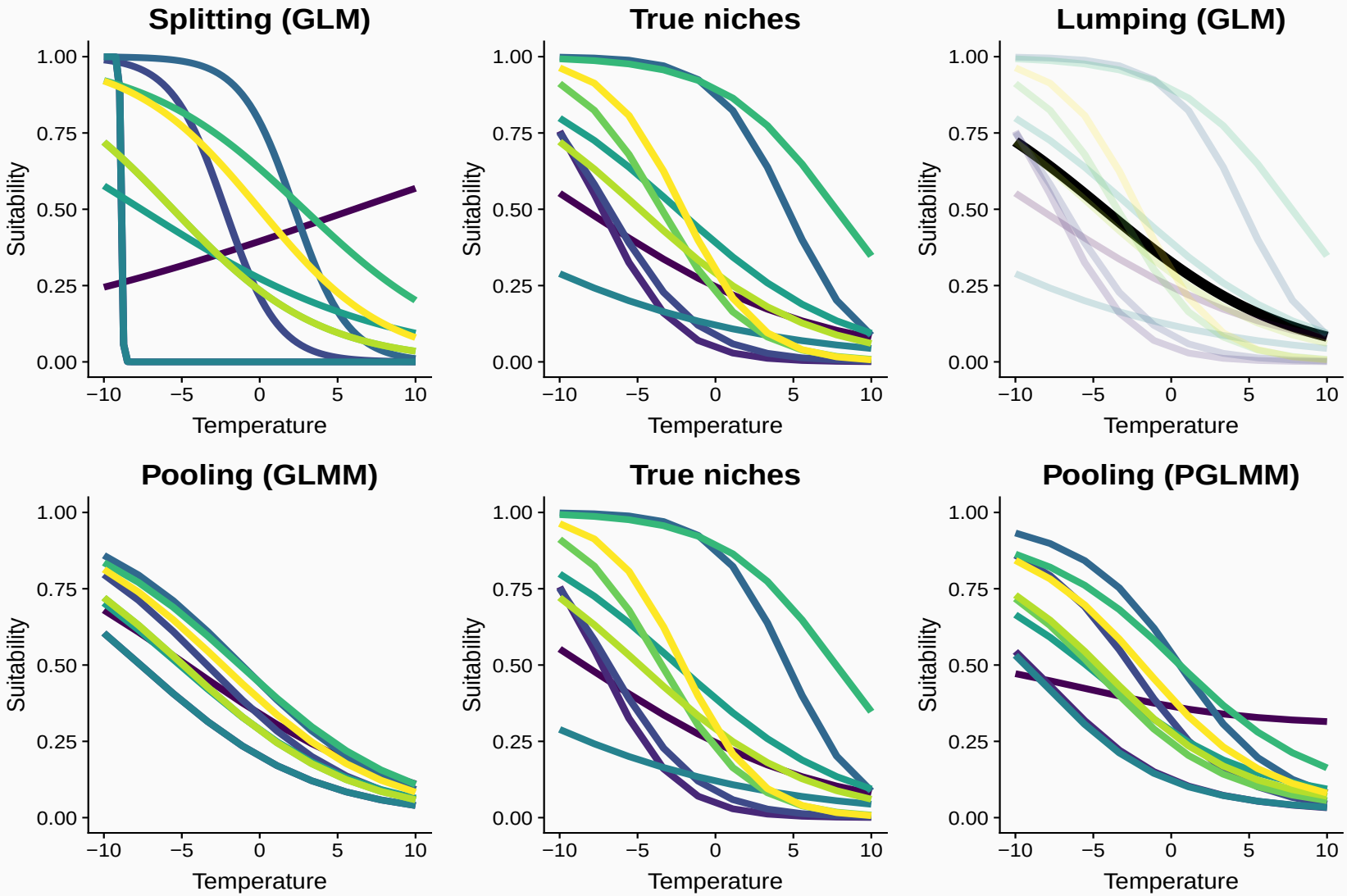
Tittes et al. 2019

Future developments/challenges

- Assess model trade-offs (sample size, bias, phylogenetic signal...)

Future developments/challenges

- Assess model trade-offs (sample size, bias, phylogenetic signal...)
- Model evaluation: Niche-Distribution duality


Splitting (GLM)
True niches
Lumping (GLM)
Pooling (GLMM)
True niches
Pooling (PGLMM)
Suitability
Temperature
1.00
0.75
0.50
0.25
0.00
−10
−5
0
5
10