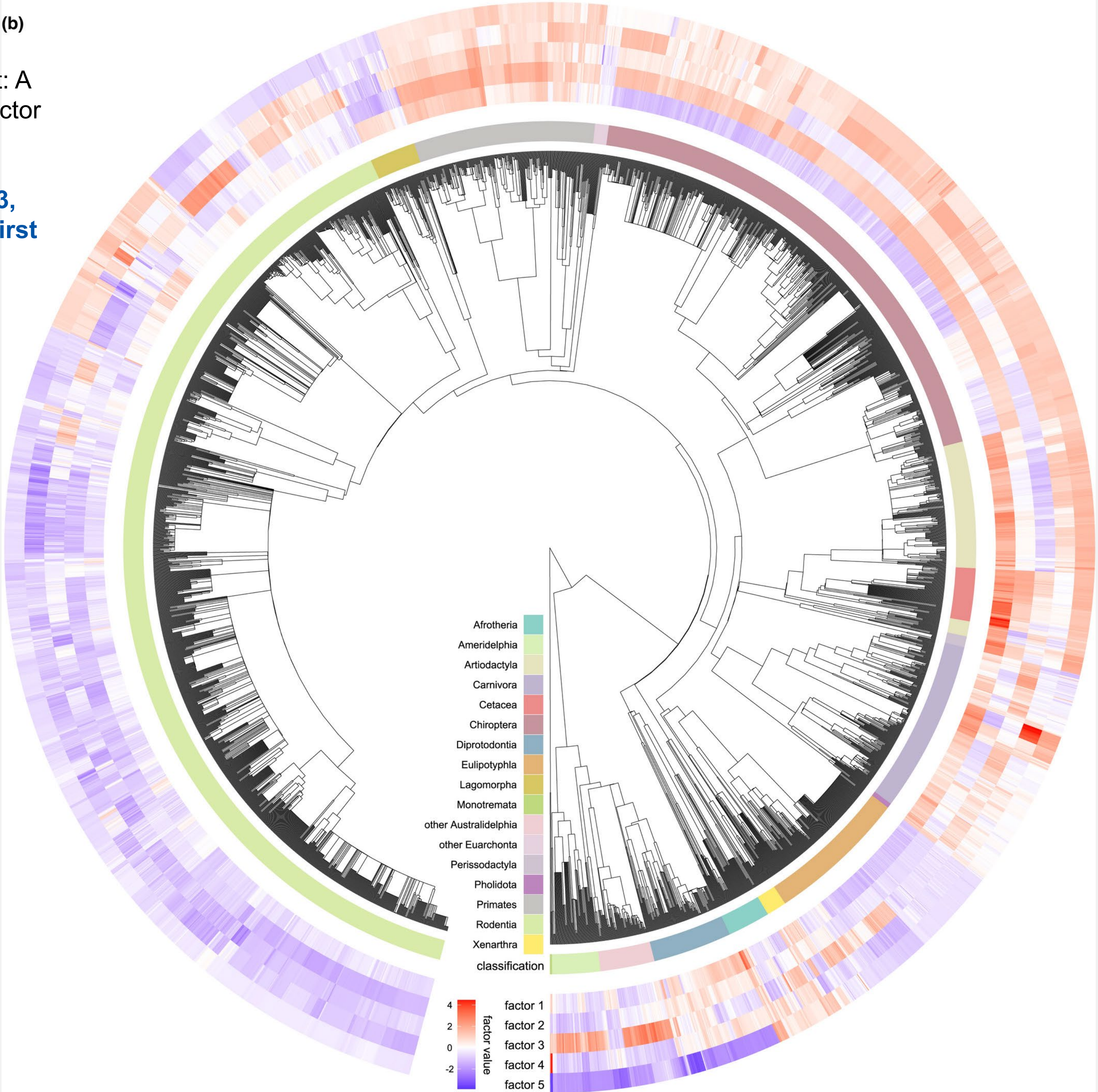
Estimating trade-offs in life-history, morphometric, behavioral, reproductive, and trophic traits for all fishes

Thorson, Maureaud, Frelat, Mérigot, Friedman,
Palomares, Pinsky, Price, Wainwright

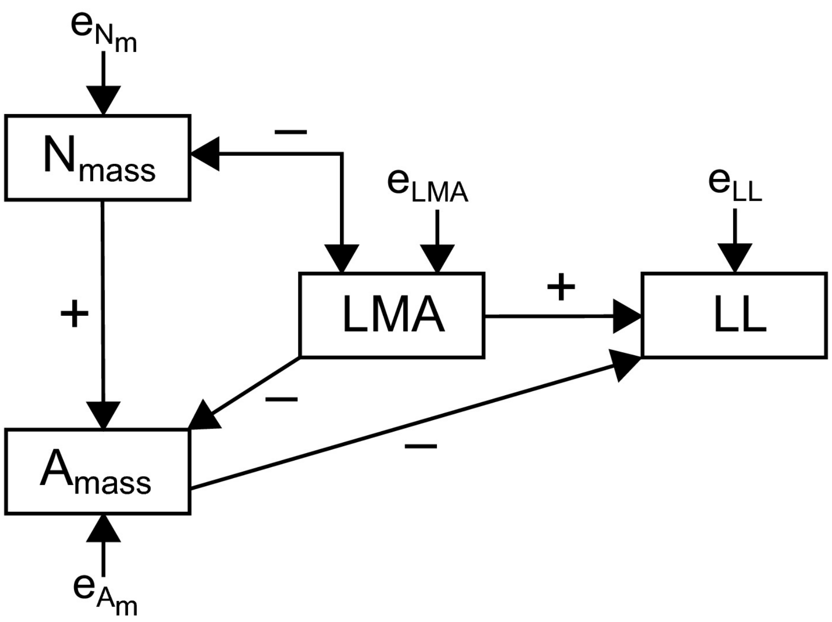
© Jonny Armstrong

Principled, practical, flexible, fast: A new approach to phylogenetic factor analysis

Methods Ecol Evol, Volume: 13, Issue: 10, Pages: 2181-2197, First published: 19 June 2022, DOI: (10.1111/2041-210X.13920)

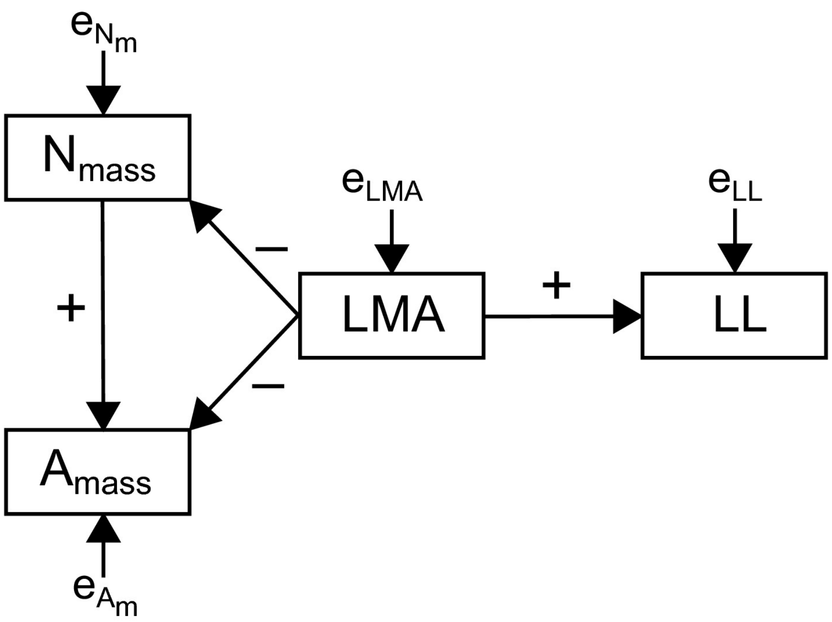


(a) Intuitive model 1

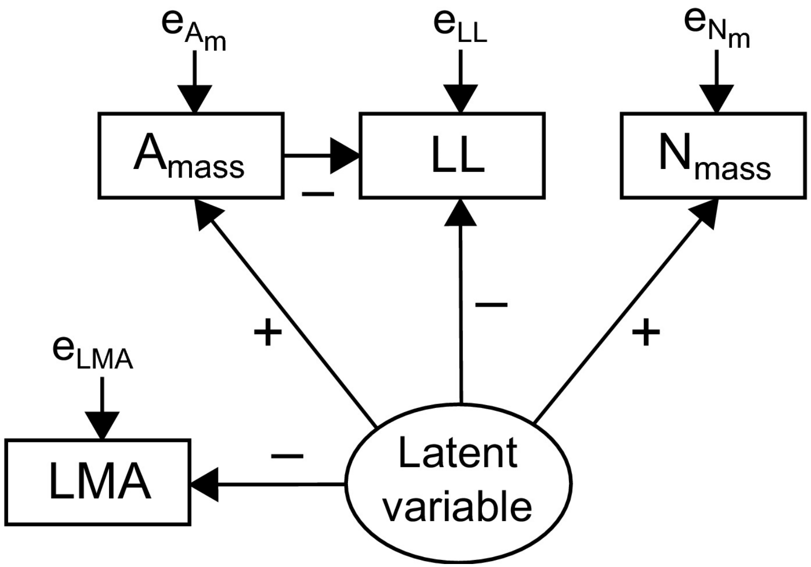


N_mass: Nitrogen concentration
A_mass: Photosynthetic rate
LMA: Leaf mass per area
LL: Leaf lifespan

(b) Intuitive model 2



(c) Shipley latent model

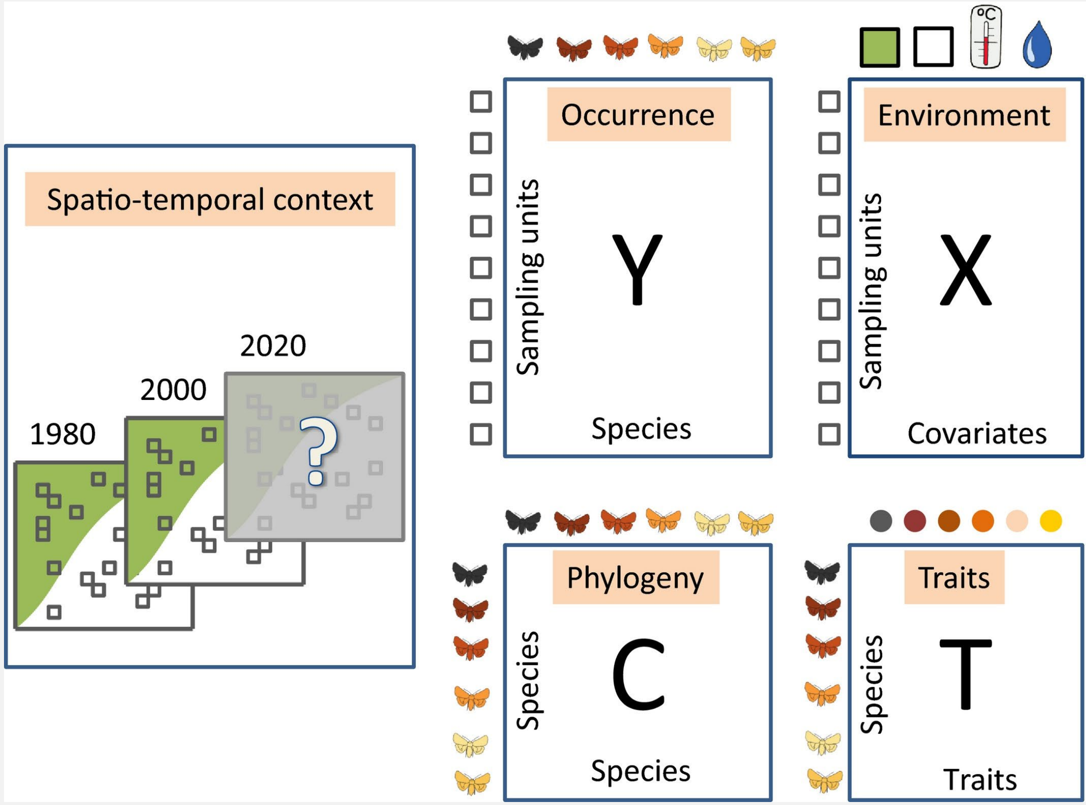


Phylogenetic structural equation modelling reveals no need for an ‘origin’ of the leaf economics spectrum

Ecology Letters, Volume: 19, Issue: 1, Pages: 54-61, First published: 13 November 2015, DOI: (10.1111/ele.12542)

How to make more out of community data? A conceptual framework and its implementation as models and software

Ecology Letters, Volume: 20, Issue: 5, Pages: 561-576, First published: 20 March 2017, DOI: (10.1111/ele.12757)

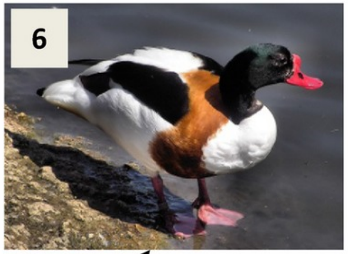


(a) **Study design**

Presence-absence of 7 waterbirds surveyed for 7 years on 215 irrigation ponds in southeast Spain.

Species:

1. Little grebe
2. Black-winged stilt
3. Little-ringed plover
4. Mallard
5. Moorhen
6. Common shelduck
7. Common coot



(b) **Variance partitioning**

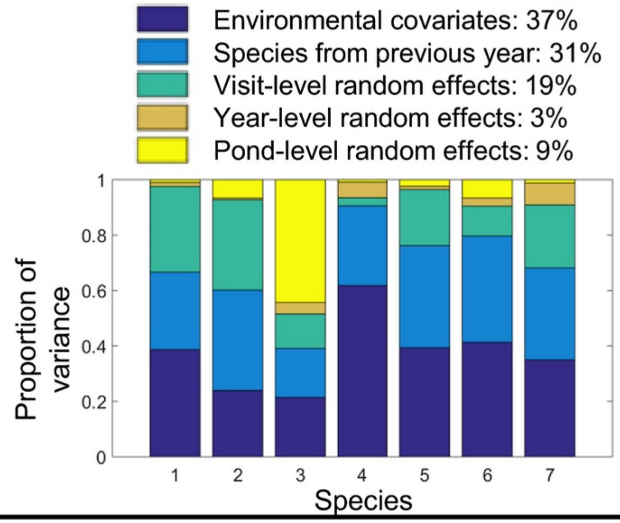
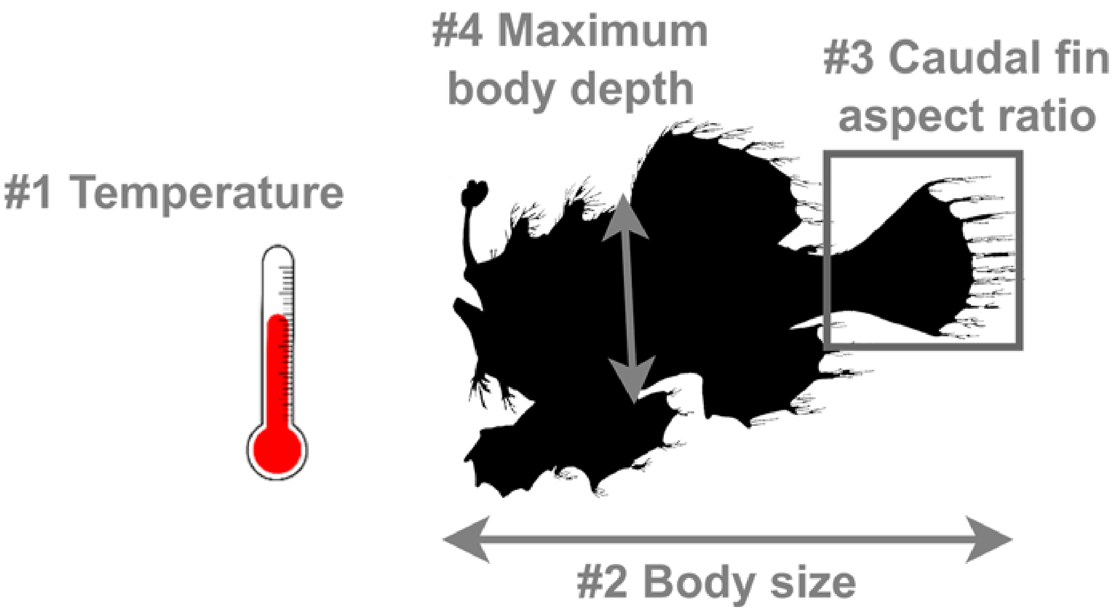
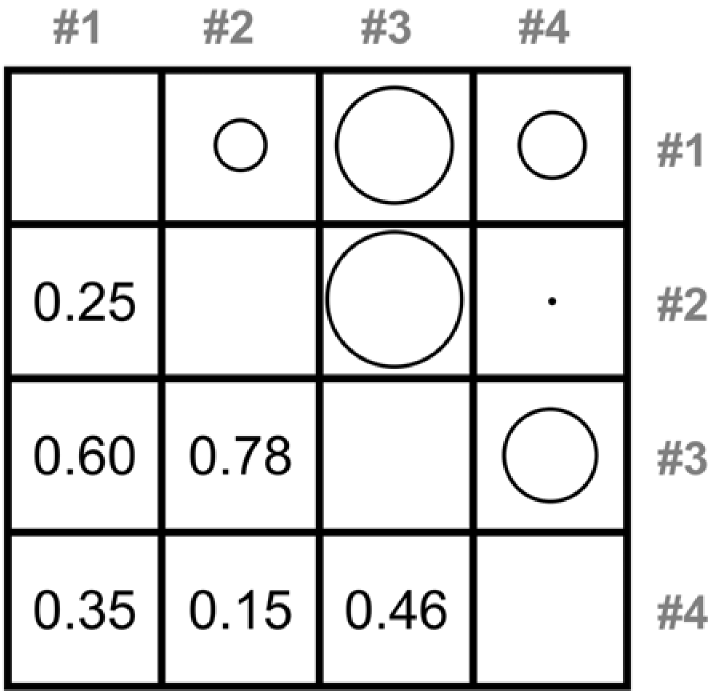


Illustration of trait correlations

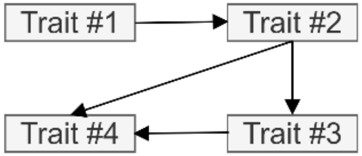




	#1	#2	#3	#4
#1		○	○	○
#2	0.25		○	·
#3	0.60	0.78		○
#4	0.35	0.15	0.46	

II - Specifying trait correlations

Structural Equation Model (SEM) specification

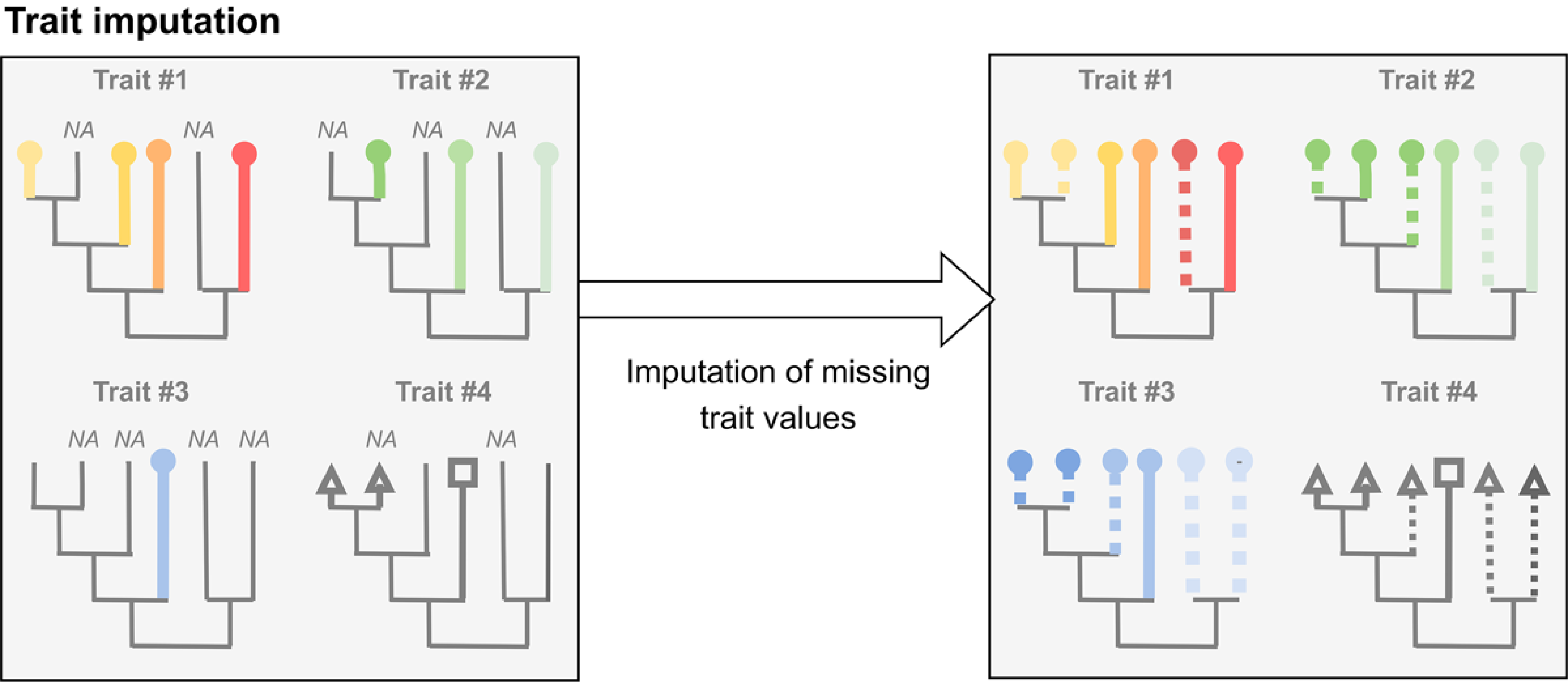


Script model entry

```
SEM = "
Trait #1 -> Trait #2
Trait #2 -> Trait #3
Trait #2 -> Trait #4
Trait #3 -> Trait #4
"
```

Resulting trait correlation matrix

$$\Gamma = 0.5 * \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$


Trait imputation
Trait #1
NA
NA
Trait #2
NA
NA
NA
Trait #3
NA
NA
NA
NA
Trait #4
NA
NA
Imputation of missing trait values
Trait #1
Trait #2
Trait #3
Trait #4

Simulation testing

Simulate a model:

$$y_i = 2 + x_i + \varepsilon_i$$

Where

$$\mathbf{x} \sim \mathrm{Brownian.Motion}(\mathrm{mean} = 1, \mathrm{sd} = 0.3)$$
$$\boldsymbol{\varepsilon} \sim \mathrm{Brownian.Motion}(\mathrm{mean} = 2, \mathrm{sd} = 0.3)$$

Three models:

1. Linear model (lm)
2. Phylogenetic linear model (*phylolm*)
3. Phylogenetic structural equation model (*FishLife*)

Two scenarios

- 100 taxa with complete data for x and y
- 100 taxa with 60% missing independently for x and y

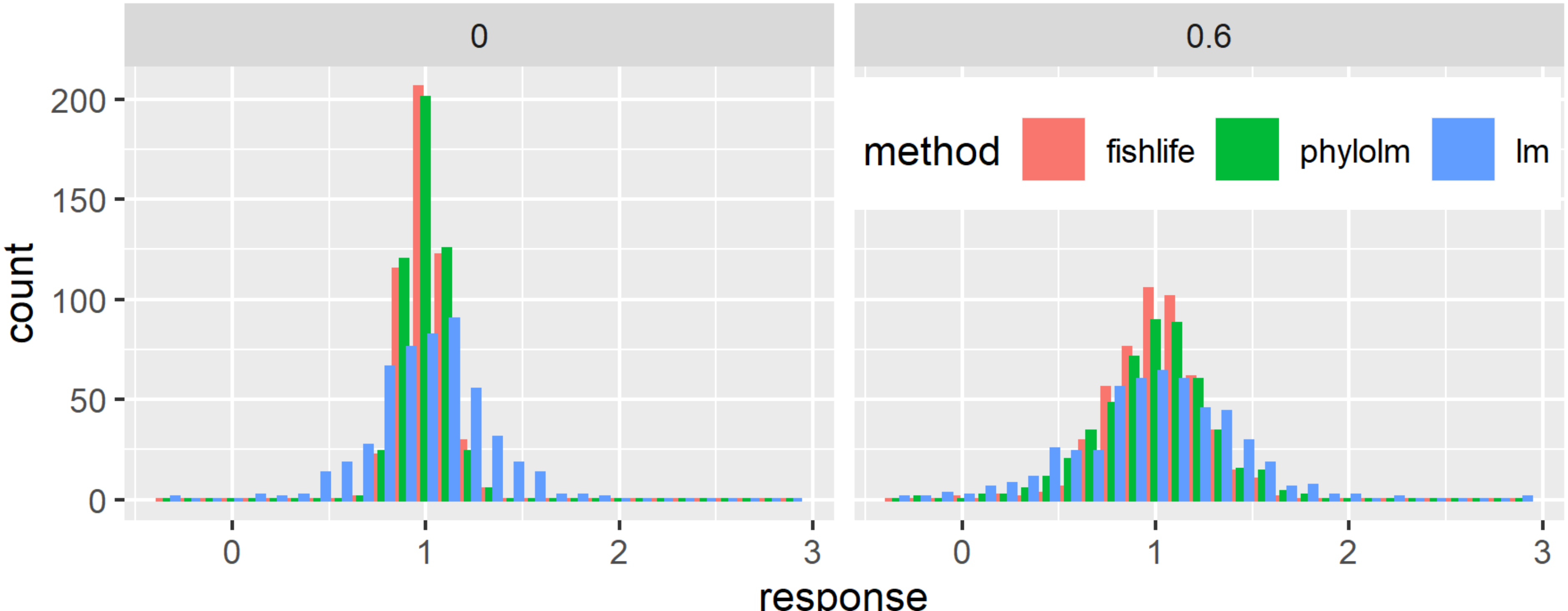
Simulation testing

Performance metric:

- Estimate of slope (true=1)

Results

- Linear model does worst
- FishLife better than phylolm with missing data



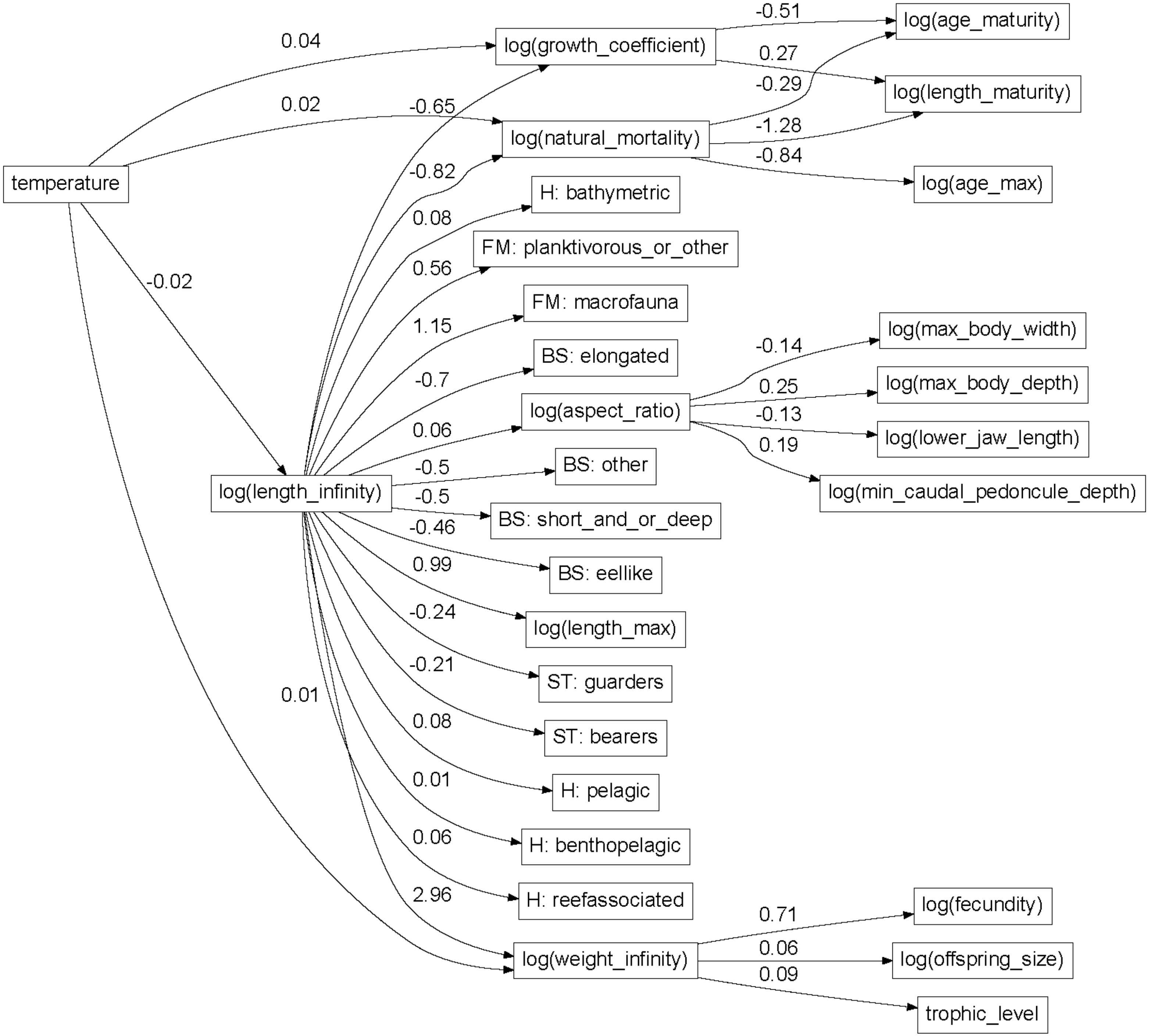
FishLife

Databases

- 53 variables
  - Expands each factor to N-1 variables, where N is the number of levels
- 34,342 taxa

Reduced version

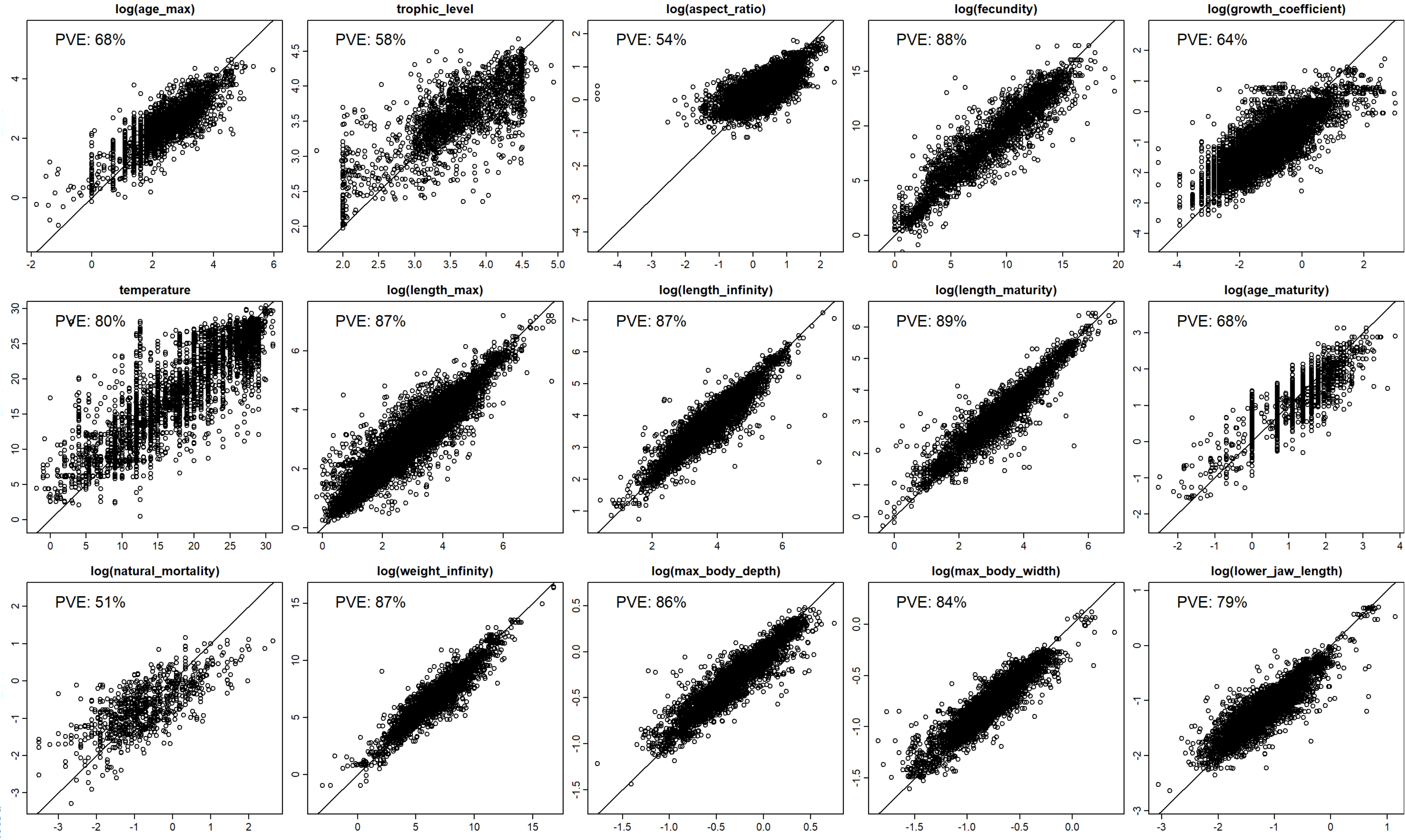
- 11 continuous variables
  - age_max, trophic_level, aspect_ratio, fecundity, growth_coefficient, temperature, length_max, max_body_depth, max_body_width, lower_jaw_length, min_caudal_pedoncule_depth
- 4 factor variables
  - spawning_type (3 levels), habitat (6 levels), feeding_mode (5 levels), body_shape (5 levels)
- 26 variables


temperature
0.04
0.02
-0.02
0.01
log(growth_coefficient)
log(natural_mortality)
log(length_infinity)
-0.65
-0.82
0.08
0.56
1.15
-0.7
0.06
-0.5
-0.5
-0.46
0.99
-0.24
-0.21
0.08
0.01
0.06
2.96
H: bathymetric
FM: planktivorous_or_other
FM: macrofauna
BS: elongated
log(aspect_ratio)
BS: other
BS: short_and_or_deep
BS: eellike
log(length_max)
ST: guarders
ST: bearers
H: pelagic
H: benthopelagic
H: reefassociated
log(weight_infinity)
-0.51
0.27
-0.29
-1.28
-0.84
log(age_maturity)
log(length_maturity)
log(age_max)
-0.14
0.25
-0.13
0.19
log(max_body_width)
log(max_body_depth)
log(lower_jaw_length)
log(min_caudal_pedoncule_depth)
0.71
0.06
0.09
log(fecundity)
log(offspring_size)
trophic_level

Cross-validation (subsampling data)

Continuous: Held-out vs. predicted and Percent-Variance-Explained (100% is good)

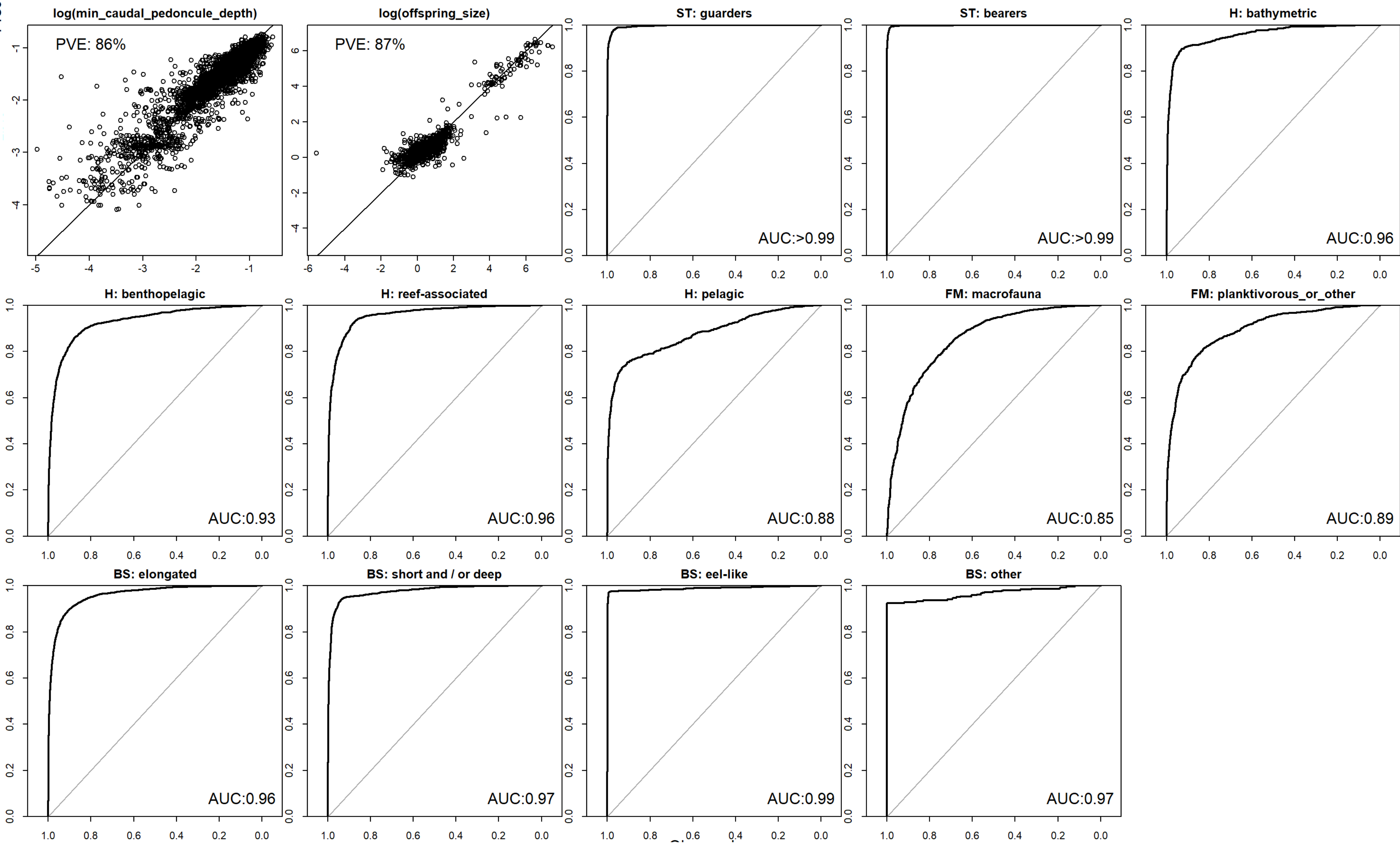
Categorical: Sensitivity vs. specificity and Area under the ROC curve (1 is good)

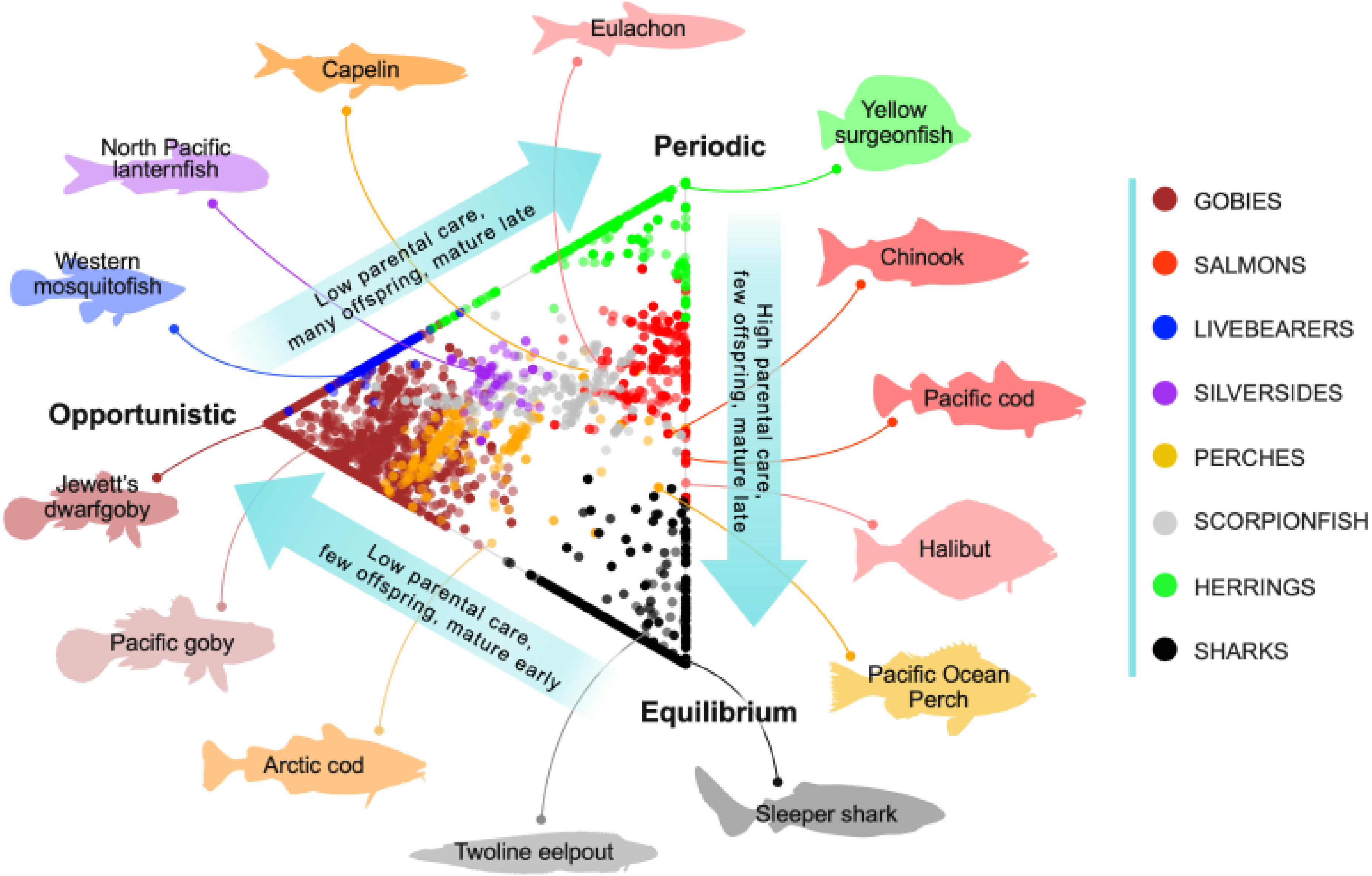


Cross-validation (subsampling data)

Continuous: Held-out vs. predicted and Percent-Variance-Explained (100% is good)

Categorical: Sensitivity vs. specificity and Area under the ROC curve (1 is good)

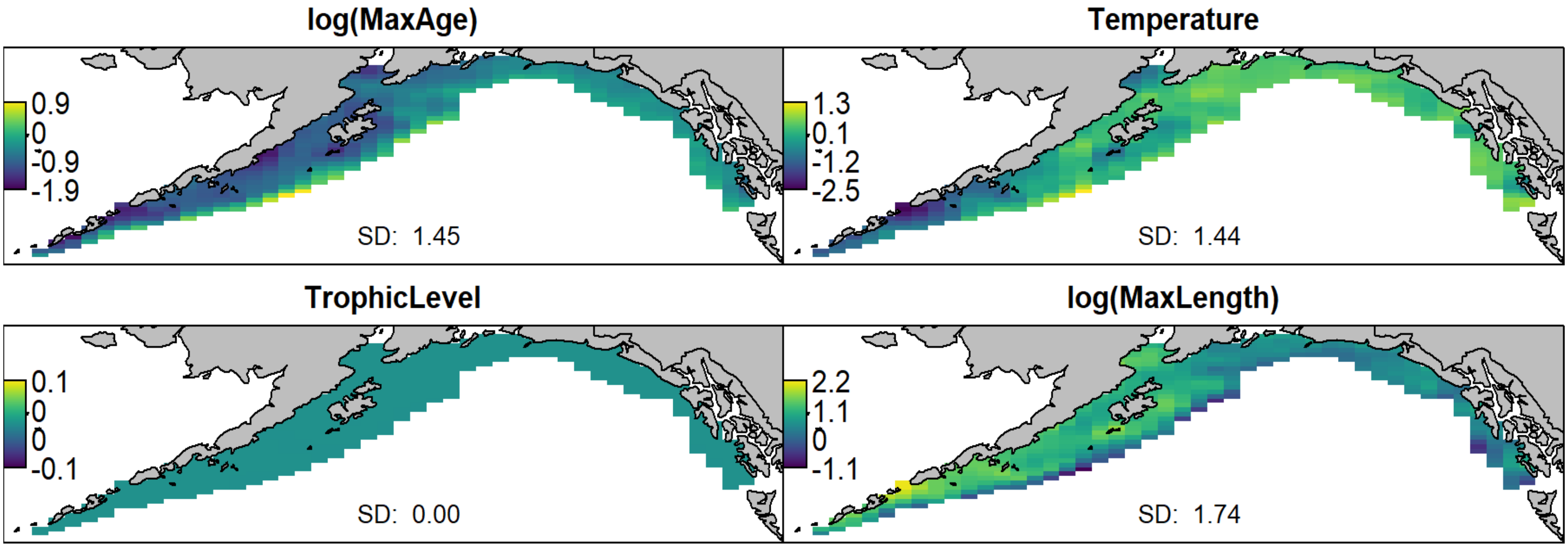



Eulachon
Capelin
Yellow surgeonfish
North Pacific lanternfish
Periodic
Western mosquitofish
Low parental care, many offspring, mature late
Chinook
High parental care, few offspring, mature late
Pacific cod
Opportunistic
Jewett's dwarfgoby
Halibut
Low parental care, few offspring, mature early
Pacific goby
Pacific Ocean Perch
Equilibrium
Arctic cod
Sleeper shark
Twoline eelpout
GOBIES
SALMONS
LIVEBEARERS
SILVERSIDES
PERCHES
SCORPIONFISH
HERRINGS
SHARKS

Why care and what's next?

© Jonny Armstrong

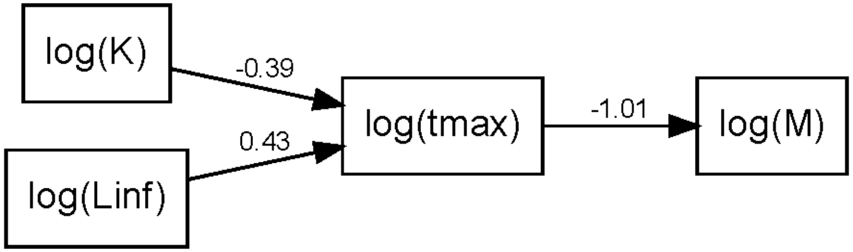
Trait-based joint species distribution models



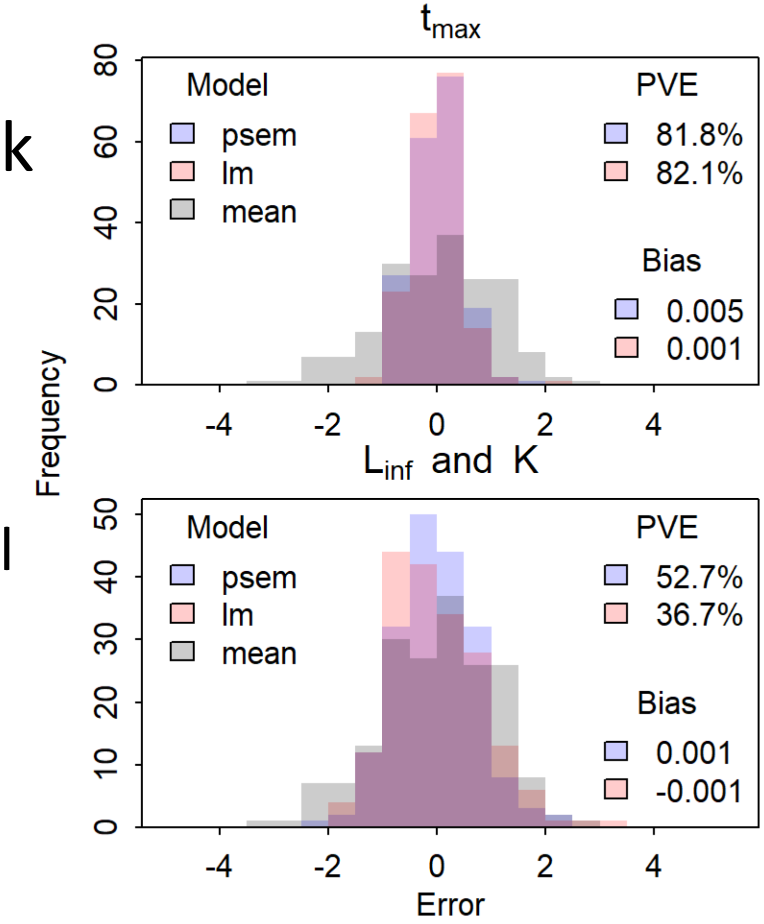
- Long-lived species: higher density on shelf-break
- Cold-water species: higher density near Aleutian Islands
- Trophic level: no estimated effect
- Large-bodied species: higher density in gullies and western GOA

Thorson, J., Barnes, C., Morano, J., Friedman, S., Siple, M., 2023. Spatially varying coefficients can improve parsimony and descriptive power for species distribution models. Ecography.

Improved life-history priors for stock assessment



Crossvalidation experiment



Improved life-history priors for stock assessment

1. Longevity → natural mortality
   - Precise and no phylogenetic signal
2. Linf / K → natural mortality
   - Less precise, but phylogeny improves things

Acknowledgements

Collaborators on related work

- Steve Munch
- Olaf Jensen
- Jason Cope
- Wesley Patrick
- Elise Zipkin
- Shijie Zhou
- Jin Gao
- Daniel Pauly
- Rainer Froese

FishGlob consortium

Discussions

- Wouter van der Bijl