FishBase Nutrient Model Structure

AARON MACNEIL

June 15, 2021

The model that underlies our nutrient predictions is a modification of that presented in Hicks et al. 2019, where we removed a couple of covariates, depth and K (growth rate), that had the potential to induce spurious correlation in our posterior effect sizes, given their potential for collider bias in our asserted directed acyclic graph.

As an alternative to the GP phylogenetic covariance model (used in Vaitla et al. 2018 for example) we capitalized on the hierarchical nesting of phylogeny (sensu Thorson 2020), whereby species belong to a given genus, genera to specific families, and families to specific orders. This implies that species-level intercepts in the observed data come from a population related by genus group membership, genera represent samples from families, and families are samples from their parent orders, which can be represented in a hierarchical phylogenetic model that includes a global (overall) mean (0) at the top of a series of a non-centred, hierarchical relationships:

$$
\begin{aligned}
\gamma_0 &\sim N(0,1) \\
\sigma_{ord} &\sim Exp(1) \\
\beta_{oz} &\sim N(0,1) \\
\beta_{ord} =& \gamma_0 + \sigma_{ord}\beta_{oz} \\
\sigma_{fam} &\sim Exp(1) \\
\beta_{fz} &\sim N(0,1) \\
\beta_{fam} &= \beta_{ord} + \sigma_{fam}\beta_{fz} \\
\sigma_{gen} &\sim Exp(1) \\
\beta_{0,gz} &\sim N(0,1) \\
\beta_{0,gen} &= \beta_{0,fam} + \sigma_{gen}\beta_{0,gz} \\
\mu_i &= \beta_{0,gen} + \beta_x X \\
\beta_i &\sim N(\mu_i, \sigma)
\end{aligned}
$$

In both phylogenetic models the set of species level trait covariates was the same

$$\beta_x X = \beta_1 GZ + \beta_2 TL + \beta_4 FP + \beta_5 L_m ax + \beta_6 BS + \beta_8 A_m at + \beta_9 WC$$

Leading to an observation-scale model

$$\mu_{obs} = \mu_i + \gamma_1 FO + \gamma_2 PR$$

While Hicks et al. 2019 used a mix of Normal, Gamma, and Noncentral-t distributions for the data likelihood, we chose to model nutrients (except protein) on the log scale, and used either a Normal (selenium, omega-3)

$$\gamma_{obs} \sim N(\mu_{obs}, \sigma_{obs})$$

or Noncentral-t distribution (protein, zinc, calcium, iron, vitamin A)

$$\gamma_{obs} \ Nt(\mu_{obs}, \sigma_{obs,\tau})$$

Given regularizing priors

$$\begin{aligned} \beta_x, \gamma_x &\sim N(0, 1) \\ \sigma_{obs} &\sim Exp(1) \\ \tau &\sim U(0, 20) \end{aligned}$$

We ran the three models on each of the seven nutrients, using the Python package PyMC3. Models were run with four separately-initiated chains for 5,000 iterations using a No-U-Turn sampler (NUTS).