Generating Priors of metabolic index parameters using a taxonomic hierarchical model

2023-07-10

Background and Motivation

Several recent papers have used compilations of laboratory experiments to estimate parameters that govern the metabolic index: the ratio of oxygen supply to demand. Because taxonomic coverage of these experiments is limited, existing species distributions are also commonly used to make inferences about the critical values of this index below which organisms avoid habitatis, and to predict changes in physiologically-favorable habitat.

Generally, information sharing between these two approaches have been ad hoc i.e., using the laboratory data to set reasonable bounds on unknown parameters, or using mean values of parameters. A formal statistical approach (for example, an explicit Bayesian analysis of species distributions and how they are govern by temperature and oxygen) would permit this information sharing to be more explicit and adaptive as more information is collected.

For that reason, an analysis of laboratory data that can provide prior probabilities for in situ observations could enhance our ability to assess how changing temperature and oxygen will change species distributions and ranges.

Metabolic index

The metabolic index depicts the ratio of oxygen supply to oxygen demand, assuming that both are goverrned by an Arrhenius equation. A common form of the metabolic index is:

$$\phi = pO_2 A_o \left(\frac{B}{B_{ref}} \right)^n \exp \left(\frac{E_o}{k_b} (\frac{1}{T} - \frac{1}{T_{ref}}) \right)$$

where ϕ is the metabolic index, pO_2 is the partial pressure of oxygen (kP), B is body size, B_{ref} is a reference body size, k_b is Boltzmann's constant, T is temperature (K) and T_ref is a reference temperature, and A_o, n, and E_o are the parameters to be estimated.

The parameters of the metabolic index can be estimated from laboratory studies that expose organisms to combinations of temperature and pO_2 and estimate the oxygen level at which organisms cannot survive or otherwise exhibit stress that are indicative of insufficient oxygen supply. By setting these empirically-derived thresholds as the values of pO_2 and T that make ϕ equal 1, the expression can be solved through ordinary least squared regression on log-transformed rearrangement of the metabolic index.

Because all observations and experiments are subject to random error, fitting each experimental fitting independently will propagate these errors into the metabolic index parameter estimates. One way to lessen these effects is through hierarchical modeling. Hierarchical models presume that each parameter for each species follows some probability distribution with a mean and variance set at a higher taxonomic level.

Recently, Thorson developed a taxonomic hierarchical model to predict life history traits in the absence of independent measurements. Briefly, a vector of j traits among organisms at taxonomic level g, $\mathbf{x}_g$ follows a multivariate normal distribution. The model assumes that these traits evolve, and that taxonomic levels are

a proxy for evolutionary potential. Thus, given a "parent" taxonomic group (the taxonomic group above the group under consideration), the vector x_g is:

$$\mathbf{x}_g \sim \text{MVN}(\mathbf{x}_{[p]}, \Sigma_{l(g)})$$

where $\mathbf{x}_{[p]}$ are the mean values of the parent taxonomic grouping, and $\Sigma_{l(g)}$ is the variance - covariance matrix for taxonomic level g. The latter is calculated as

$$\Sigma_{l(g)} = \lambda_{l(g)}\Sigma$$

where Σ is the variance-covariance matrices among the parameters, and $\lambda_{l(g)}$ expresses how the variability grows or expands at each taxonomic level. Thorson describes the latter as the proportion of evolutionary covariance occuring at taxonomic level $l(g)$.

Taxonomic Summary of the data

The following tables depict the data organized by Class, Order, and Family.

Table 1: Summary of data by Class

Class	NoOrder	NoFamily	NoSpecies
Actinopteri	14	21	28
Ascidiacea	1	1	1
Bivalvia	3	3	3
Cephalopoda	3	3	3
Copepoda	2	4	4
Elasmobranchii	1	1	1
Gastropoda	1	1	1
Malacostraca	3	15	28
Ophiuroidea	1	2	2
Polychaeta	1	1	1
Thaliacea	1	1	1

Table 2: Summary of data by Order

Order	Class	NoFamily	NoSpecies
Acipenseriformes	Actinopteri	1	1
Amphipoda	Malacostraca	1	1
Anguilliformes	Actinopteri	1	1
Atheriniformes	Actinopteri	1	1
Blenniiformes	Actinopteri	1	2
Calanoida	Copepoda	3	3
Carcharhiniformes	Elasmobranchii	1	1
Cypriniformes	Actinopteri	1	2
Decapoda	Malacostraca	13	26
Eupercaria incertae sedis	Actinopteri	4	5
Euphausiacea	Malacostraca	1	1
Gadiformes	Actinopteri	1	2
Gobiiformes	Actinopteri	1	2
Harpacticoida	Copepoda	1	1
Littorinimorpha	Gastropoda	1	1

Order	Class	NoFamily	NoSpecies
Myctophiformes	Actinopteri	1	1
Myopsida	Cephalopoda	1	1
Mytilida	Bivalvia	1	1
Octopoda	Cephalopoda	1	1
Oegopsida	Cephalopoda	1	1
Ophiacanthida	Ophiuroidea	2	2
Ostreida	Bivalvia	1	1
Ovalentaria incertae sedis	Actinopteri	1	1
Pectinida	Bivalvia	1	1
Perciformes	Actinopteri	4	5
Phyllodocida	Polychaeta	1	1
Pleuronectiformes	Actinopteri	2	2
Salmoniformes	Actinopteri	1	2
Salpida	Thaliacea	1	1
Stolidobranchia	Ascidiacea	1	1
Tetraodontiformes	Actinopteri	1	1

Table 3: Summary of data by Family

Family	Order	NoGenus	NoSpecies
Acanthephyridae	Decapoda	2	6
Acipenseridae	Acipenseriformes	1	1
Anguillidae	Anguilliformes	1	1
Anoplopomatidae	Perciformes	1	1
Atherinopsidae	Atheriniformes	1	1
Benthesicymidae	Decapoda	1	2
Bythograeidae	Decapoda	1	1
Cancridae	Decapoda	1	1
Carcinidae	Decapoda	1	1
Cyclopteridae	Perciformes	1	1
Cyprinidae	Cypriniformes	2	2
Euphausiidae	Euphausiacea	1	1
Gadidae	Gadiformes	1	2
Gammaridae	Amphipoda	1	1
Gobiidae	Gobiiformes	1	2
Harpacticidae	Harpacticoida	1	1
Labridae	Eupercaria incertae sedis	1	1
Loliginidae	Myopsida	1	1
Lucicutiidae	Calanoida	1	1
Majidae	Decapoda	1	1
Megacalanidae	Calanoida	1	1
Metridinidae	Calanoida	1	1
Moronidae	Eupercaria incertae sedis	1	1
Munididae	Decapoda	1	1
Myctophidae	Myctophiformes	1	1
Mytilidae	Mytilida	1	1
Nereididae	Phyllodocida	1	1
Octopodidae	Octopoda	1	1
Ommastrephidae	Oegopsida	1	1
Ophiocomidae	Ophiacanthida	1	1

Family	Order	NoGenus	NoSpecies
Ophiodermatidae	Ophiacanthida	1	1
Oplophoridae	Decapoda	2	3
Ostreidae	Ostreida	1	1
Palinuridae	Decapoda	1	1
Pandalidae	Decapoda	1	2
Paralichthyidae	Pleuronectiformes	1	1
Pectinidae	Pectinida	1	1
Penaeidae	Decapoda	2	2
Pleuronectidae	Pleuronectiformes	1	1
Pomacentridae	Ovalentaria incertae sedis	1	1
Portunidae	Decapoda	1	1
Salmonidae	Salmoniformes	2	2
Salpidae	Salpida	1	1
Sciaenidae	Eupercaria incertae sedis	1	1
Scyliorhinidae	Carcharhiniformes	1	1
Sergestidae	Decapoda	3	4
Serranidae	Perciformes	1	1
Sparidae	Eupercaria incertae sedis	2	2
Strombidae	Littorinimorpha	1	1
Styelidae	Stolidobranchia	1	1
Tetraodontidae	Tetraodontiformes	1	1
Tripterygiidae	Blenniiformes	1	2
Zoarcidae	Perciformes	2	2

By far, the most well represented groups are the ray-finned fishes (Class Actinopteri) and the Malacostraca crustaceas (Classs Malacostraca). Of the latter, roughly half were in the Order Decapoda. For the vast majority of Orders, the dataset included only a single Family. Within families, there were at most two genera and two species represented.

Because of the limited number of organisms within lower taxonomic groupings in the database, we ran the hierchircal model using four taxonomic groupings Class, Order, Family and Species (i.e, we omitted Genus)

Hierarchical Model Findings Summary

Point 1: n does not vary, and is approximately -0.1.

At nearly every taxonomic level (Class, Order, Family), the mean value of n did not vary substantially from the grand mean of -0.08. A small number of Family-level estimates showed some deviation, but these were all cases in which a single species and experimental study was represented.

Point 2: Few Differences in Ao and Eo among classes and orders

Below I show results of simulating 10,000 random species within each taxomic grouping, where the group-level means were drawn from multivariate normal disributions, and then then the estimated lower-level values down to species were simulated based on the estimated mean values,estimated variance covariance matrix and the $\lambda_{l(j)}$. These would be similar to a prior probabilty if one knew the taxonomic identity of an organism at a high level (Class, Order, or Family) and wished assign a prior probability at the species level.

```
## Warning: Removed 2595 rows containing non-finite values
## (`stat_density2d_filled()`).
```
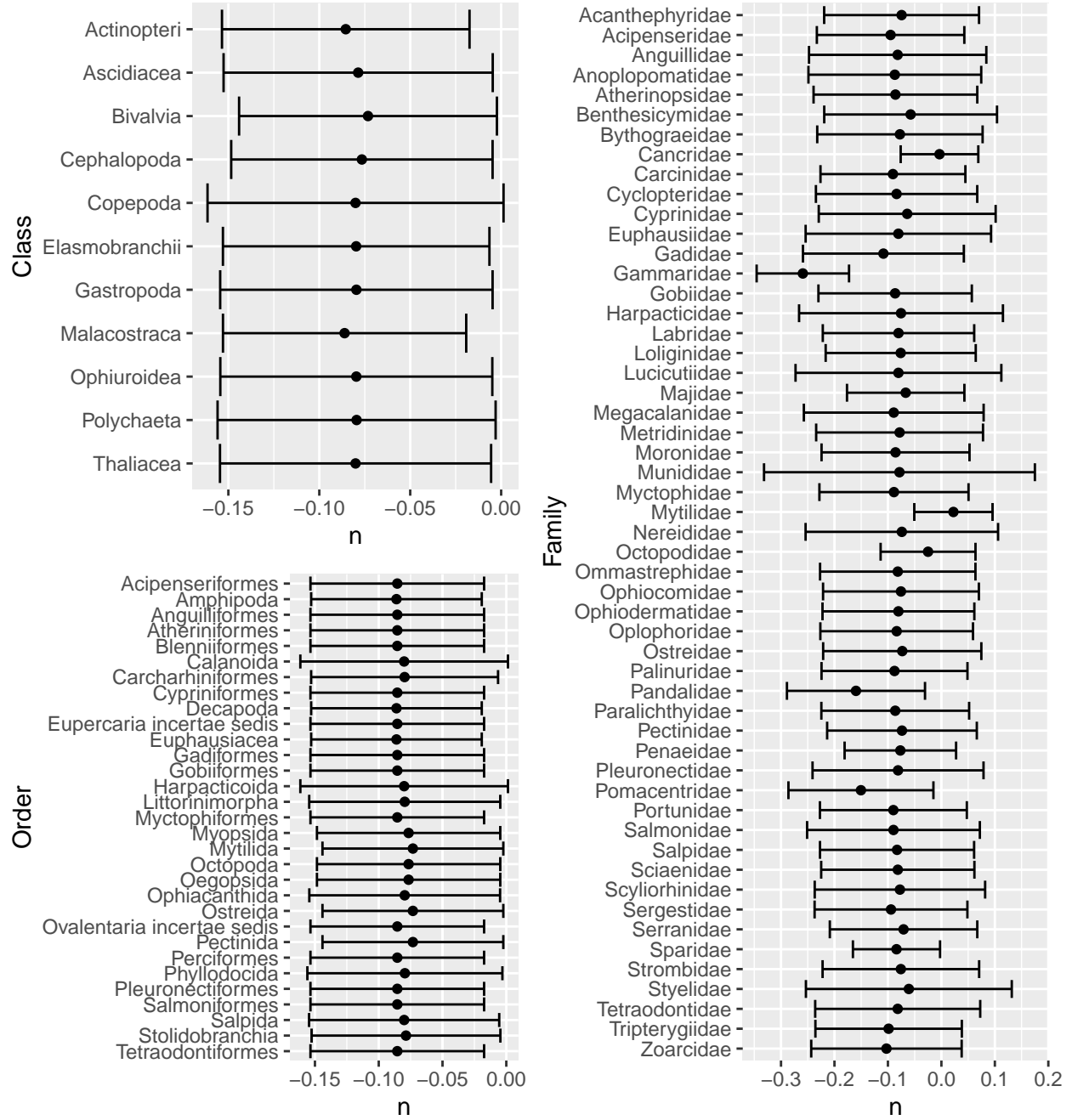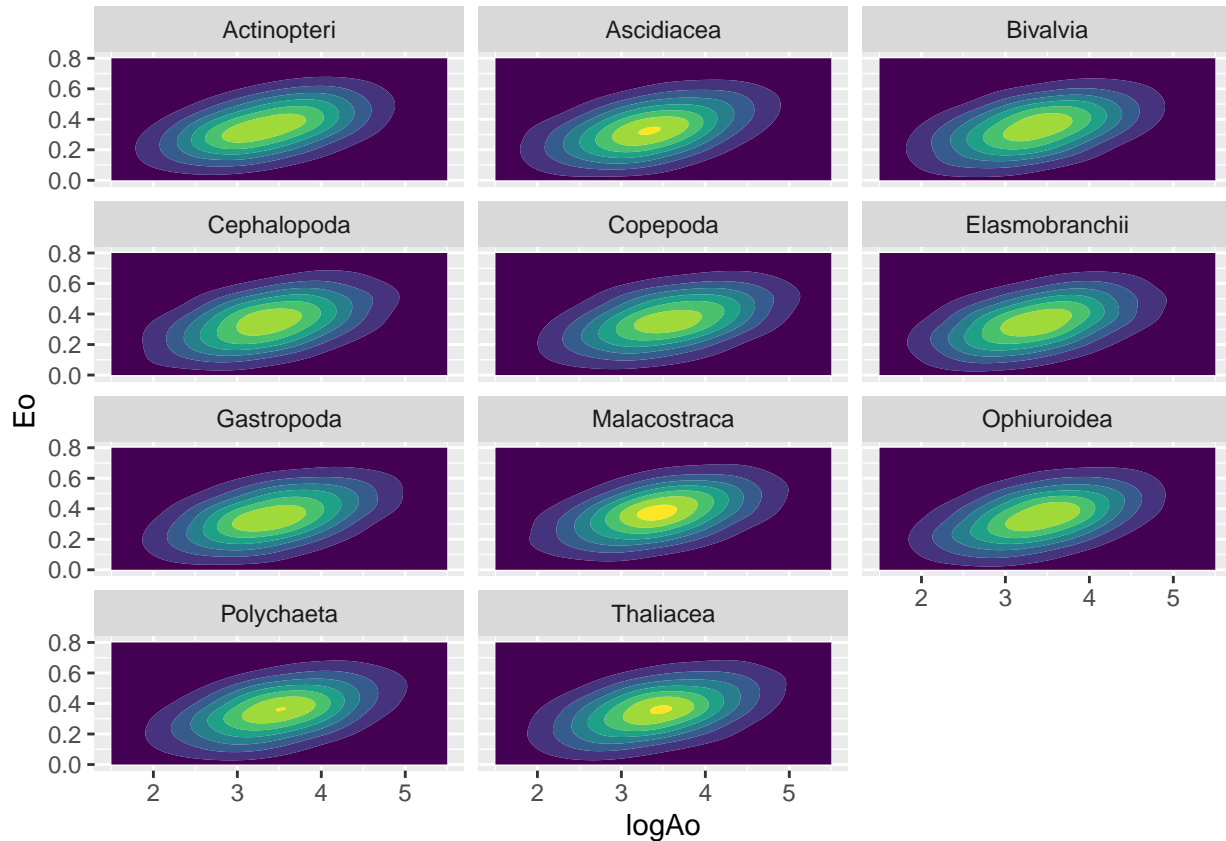
Figure 1: Estimated mean (SE) of n for Class, Order, and Family

Generally, the middle of the simulated species within Classes are close to the overall mean values (Ao = 3.41, Eo = 0.350), and 95% of species in these classes will fall between 1.95 and 4.88 (Ao) and 0.036 and 0.66 (Eo).

There were slight differences among Orders, where only Orders with at least 3 separate species are shown.

```
## Warning: Removed 765 rows containing non-finite values
## (`stat_density2d_filled()`).
```

Slightly larger differences were projected among some families, although many of these groups included only two species (three families had three species represented). Zoocaridae (eelpouts) and Gadidai (cods) had relatively low Ao (or lower tolerance for low $pO2$), while species in the families Penaidae and Triptyrygiidae had a higher Ao. There was more variability in E_o among families, with estimated temperature dependence being low among Salmonidae and exceptionally high among Penaeidae.

```
#### Plot families with at least 2 species ####
sim_beta_df <- NULL
redFamilyEst <- dplyr::filter(FamilyEst, NoSpecies >=2)
ngroups <- nrow(redFamilyEst)
for (i in 1:ngroups) {
  Group.2.use <- redFamilyEst$Family[i]
  sim_beta_df_group <- sim_spc_in_group(obj  = obj,
                                        rep = rep,
                                        n.sims = n.sims,
                                        ParentChild_gz = ParentChild_gz,
                                        Group = Group.2.use ,
                                        GroupEst = redFamilyEst,
                                        level = 3,
                                        log_lambda = log_lambda,
                                        sigma
```
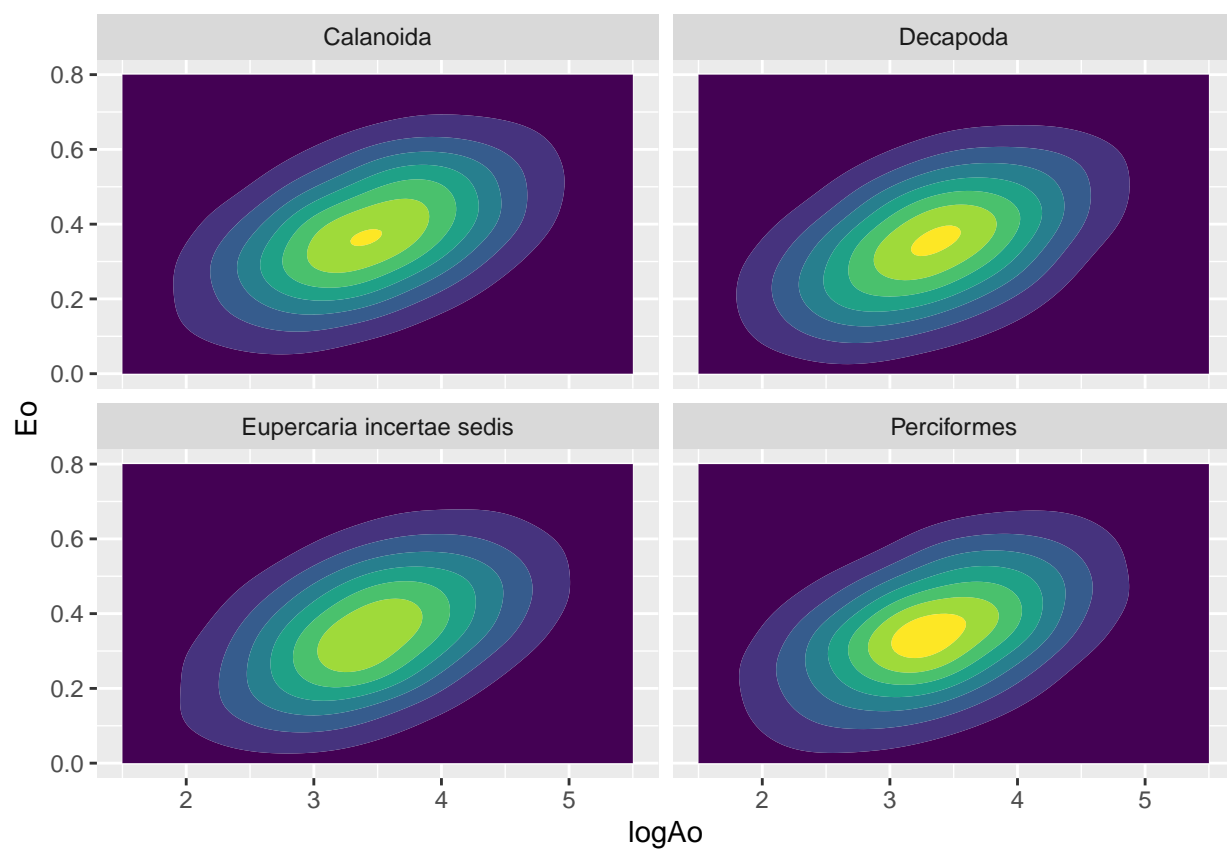
Figure 2: Same plot for Orders that had at least three species represented in the data

```
  )
  sim_beta_df <- rbind(sim_beta_df, sim_beta_df_group)
}


d <- ggplot(data = sim_beta_df, aes( x = logAo, y = Eo)) +
  geom_density_2d_filled( stat = "density_2d_filled", h = c(1, 0.2),
                          show.legend = F) +
  xlim(1.5, 5.5) +
  ylim(-0.1, 1.0) +
  facet_wrap(vars(Group), nrow = 4, ncol = 4)
print(d)
```
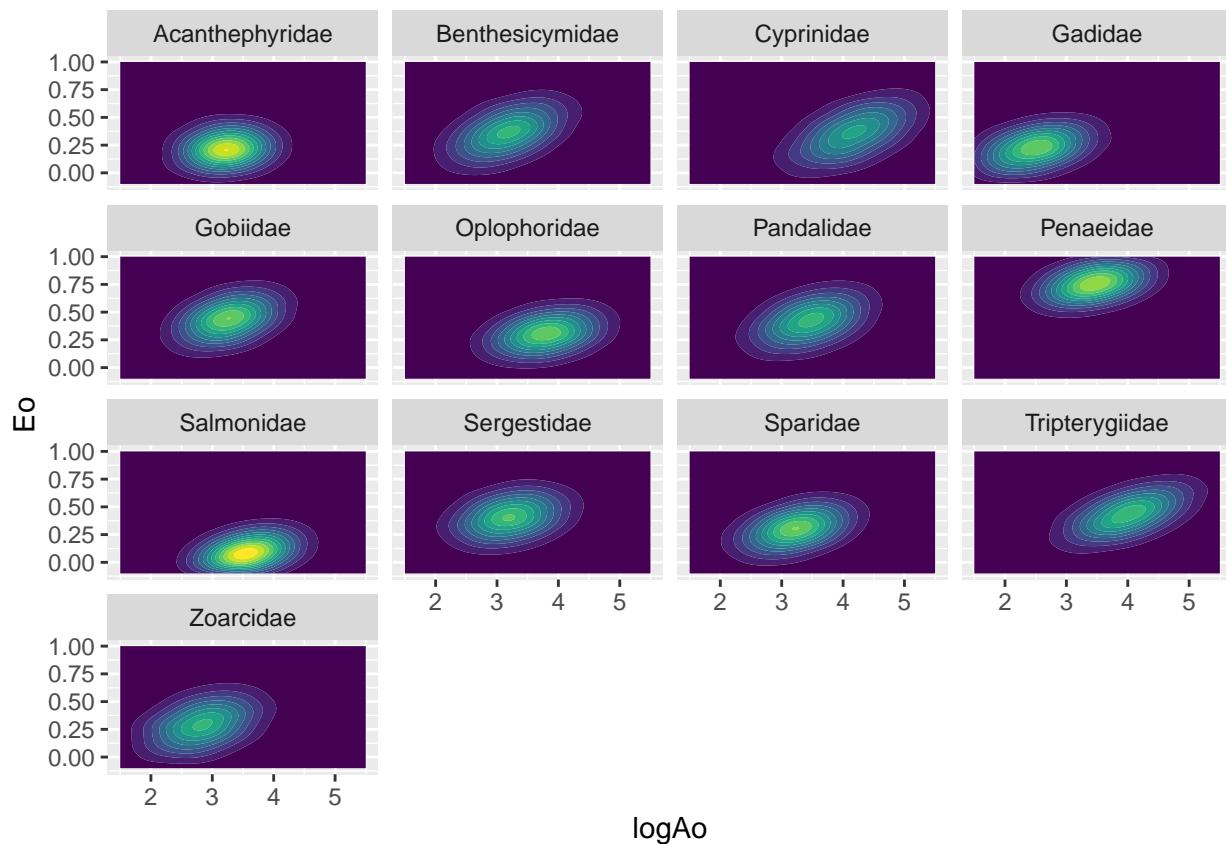
```
## Warning: Removed 2262 rows containing non-finite values
## (`stat_density2d_filled()`).
```


Point 3: Hierarchical model improves estimation through shrinkage

Fitting individual linear model to species-level data typically leads to wider variability of traits than hierarchical models. In other words, the hierarchical model dismisses exceptionally unusual parameter estimates to a species data and "shrinks" those estimates to the Parent group level means. In essence, this stabilizes parameter estimates.

To do: fit each species independently when there are adequate T, W data

Note: Cannot recreate parameter estimates in spreadsheet "MI_traits_Est_by_Curtis"