

Joint modeling notes and thoughts

For modeling trait values estimated at the species-level from one model to help predict phenology data from OSPREE, which model makes the most sense?

Update thoughts, 26 March 2020

Joint trait and phenology mode formulated like this:

$$\begin{aligned}\hat{y}_{trait,i} &= \alpha_{trait,grand} + \alpha_{trait,sp[i]} + \alpha_{study[i]} \\ \alpha_{trait,sp} &\sim N(0, \sigma_{\alpha,trait}) \\ \alpha_{study} &\sim N(0, \sigma_{\alpha,study}) \\ y_{trait} &\sim N(\hat{y}_{trait}, \sigma_{trait,y}^2)\end{aligned}$$

$$\begin{aligned}\hat{y}_{pheno,i} &= \alpha_{pheno,sp[i]} + \beta_{forcing_{sp[i]}} * F_i \\ \beta_{forcing_{sp}} &= \alpha_{forcing_{sp}} + \beta_{trait:pheno} * \alpha_{trait,sp} \\ \alpha_{pheno,sp} &\sim N(\mu_{\alpha,pheno}, \sigma_{\alpha,pheno}) \\ \alpha_{forcing_{sp}} &\sim N(\mu_{\alpha,forcing}, \sigma_{\alpha,forcing}) \\ y_{pheno} &\sim N(\hat{y}_{pheno}, \sigma_{y,pheno}^2)\end{aligned}$$

Side note ... this model:

$$\begin{aligned}\hat{y}_{trait,i} &= \alpha_{trait,grand} + \alpha_{trait,sp[i]} + \alpha_{study[i]} \\ \alpha_{trait,sp} &\sim N(0, \sigma_{\alpha,trait}) \\ \alpha_{study} &\sim N(0, \sigma_{\alpha,study}) \\ y_{trait} &\sim N(\hat{y}_{trait}, \sigma_{trait,y}^2)\end{aligned}$$

And this model:

$$\begin{aligned}y_{trait,i} &= \alpha_{trait,sp[i]} + \alpha_{study[i]} + \epsilon_{trait,i} \\ \alpha_{trait,sp} &\sim N(\mu_{\alpha,trait}, \sigma_{\alpha,trait}) \\ \alpha_{study} &\sim N(\mu_{\alpha,study}, \sigma_{\alpha,study}) \\ \epsilon_{trait} &\sim N(0, \sigma_{trait,y}^2)\end{aligned}$$

I believe are the same, but I was clearer about \hat{y}_{trait} (predicted y) versus the data (y_{trait}) in the first version, and in the first version I showed that you need to break out $\alpha_{trait,grand}$ when you have multiple intercepts.

By Lizzie, from before March

Example here is whether traits may help predict phenology from growth chamber studies. Even though we generally estimate chilling, forcing and photoperiod from growth chamber studies I just focus on forcing (F) here to keep things simple (but we could start with even simpler models ... e.g., intercept-only, if needed).

One option: Keep each model separate and then try to put together posteriors in new model (this is similar to what Heather Kharouba and I did in our 2018 paper, though only from a single model):

Trait model:

$$\begin{aligned}y_{trait,i} &= \alpha_{trait,sp[i]} + \alpha_{study[i]} + \epsilon_{trait,i} \\ \alpha_{trait,sp} &\sim N(\mu_{\alpha,trait}, \sigma_{\alpha,trait}) \\ \alpha_{study} &\sim N(\mu_{\alpha,study}, \sigma_{\alpha,study}) \\ \epsilon_{trait} &\sim N(0, \sigma_{trait,y}^2)\end{aligned}$$

Phenology model:

$$\begin{aligned}y_{pheno,i} &= \alpha_{pheno,sp[i]} + \beta_{forcing_{sp[i]}} * F_i + \epsilon_{pheno,i} \\ \alpha_{pheno,sp} &\sim N(\mu_{\alpha,pheno}, \sigma_{\alpha,pheno}) \\ \beta_{forcing_{sp}} &\sim N(\mu_{forcing}, \sigma_{forcing}) \\ \epsilon_{pheno} &\sim N(0, \sigma_{y,pheno}^2)\end{aligned}$$

Put them together somehow like:

$$\begin{aligned}\beta_{forcing_{sp}} &= \alpha + \beta_{phenoxtrait} * \alpha_{trait,sp} + \epsilon_{phenoxtrait} \\ &\dots\end{aligned}$$

Alternatively, we could joint model them. In this case, these three models would be run together in Stan but—because all the parameters are estimated at once—you have to be cautious in formulating a model so that you don’t force correlations (in BUGS I am told the ‘cut’ function was designed for just this issue, from openbugs.net: “[use cut] when we want evidence from one part of a model to form a prior distribution for a second part of the model, but we do not want ‘feedback’ from this second part”). The best approach of course is to try to build your question more clearly into the model.

In discussion with Michael Betancourt (18 December 2019), he suggested a good option: add a latent parameter in the phenology model, which can go to zero, if there is no correlation due to traits. Here’s my first stab at this (trait model is unchanged but carries through to phenology model):

$$\begin{aligned}
y_{trait,i} &= \alpha_{trait,sp[i]} + \alpha_{study[i]} + \epsilon_{trait,i} \\
\alpha_{trait,sp} &\sim N(\mu_{\alpha,trait}, \sigma_{\alpha,trait}) \\
\alpha_{study} &\sim N(\mu_{\alpha,study}, \sigma_{\alpha,study}) \\
\epsilon_{trait} &\sim N(0, \sigma_{trait,y}^2)
\end{aligned}$$

$$\begin{aligned}
y_{pheno,i} &= \alpha_{pheno,sp[i]} + \beta_{forcing_{sp[i]}} * F_i + \epsilon_{pheno,i} \\
\beta_{forcing_{sp}} &= \alpha_{forcing_{sp}} + \beta_{traitxpheno} * \alpha_{trait,sp} \\
\alpha_{pheno,sp} &\sim N(\mu_{\alpha,pheno}, \sigma_{\alpha,pheno}) \\
\alpha_{forcing_{sp}} &\sim N(\mu_{\alpha,forcing}, \sigma_{\alpha,forcing}) \\
\epsilon_{pheno} &\sim N(0, \sigma_{y,pheno}^2)
\end{aligned}$$

Ideally (if we write this correctly, which I am not sure I have), the model would test our question about how much traits may inform phenology via the $\beta_{traitxpheno}$ term; if the traits don't correlate with phenology then $\beta_{traitxpheno}$ can go to zero (and variation can load on $\alpha_{forcing_{sp}}$).