

Supplementary Material

Phylogenetic estimates of species-level phenology improve
ecological forecasting

December 29, 2022

Authors:

The Wolkovich Lab in 2019 & collaborators ^{1,2,3,4}

Author affiliations:

¹Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4;

²Arnold Arboretum of Harvard University, 1300 Centre Street, Boston, Massachusetts, USA;

³Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts, USA;

⁴Edificio Ciencias, Campus Universitario 28805 Alcalá de Henares, Madrid, Spain

*Corresponding author: ignacio.moralesc@uah.es

Extended Methods

Interpretation of λ_j and σ_j^2 on slopes and intercepts

Most current phylogenetic regression approaches aimed at controlling for phylogenetic non-independence of analysis units (i.e. species, see Revell, 2010) assume the λ scaling parameter is constant across the full set of predictors in the model. Thus, λ is estimated as a single parameter based on one single residual term VCV matrix. While useful for correcting for phylogenetic non-independence this approach does not allow the phylogeny to differentially affect different predictors (i.e. environmental cues in our example). In models with multiple cues, species responses to all cues are estimated as similarly phylogenetically structured, but this may not be the case. For example, in a PGLS model with three cues, it would be possible to have a high (i.e. close to 1) value of λ , due to either a strong phylogenetic signal in the response, but no phylogenetic structuring in the cues, or one or more predictors being strongly phylogenetically structured. In the latter case, phylogenetic structuring of responses to cues could be correlated (i.e., responses to cues evolving in a correlated fashion) or uncorrelated (i.e., independent evolution of responses to cues). Discerning these different situations is not trivial as they would inform whether responses to predictors configure in a structured fashion along the evolutionary process. However, most current approaches act as a black box regarding this information; they simply inform whether or not model residuals are phylogenetically structured (i.e. in PGLS) or the amount of model variance attributable to the phylogeny and independent from other sources of variation (i.e., in PMM, see Housworth et al. (2004)).

Because we are specifically interested in estimating the phylogenetic structure of each cue, our approach explicitly partitions variance into specific components relative to the model intercept and predictor (cue) slopes (see equation ??). The multivariate normal distributions of the intercept and slope terms include each a variance term (see equation ??), modelled with a λ scaling parameter. The interpretation of λ s in our models are analogous to Pagel’s λ (Pagel, 1999) parameter (Housworth et al., 2004), constrained to range from 0 to 1, with values of 0 indicating absence of phylogenetic relatedness, and values of 1 indicating *Brownian Motion* evolution (BM). Estimated λ s are not fully equivalent to computing phylogenetic signal of the slopes of each cue separately (i.e., fitting a multilevel regression model with species as a grouping factor on intercepts, and subsequently estimating phylogenetic signal for model slopes). Instead, they are a relative metric of phylogenetic relatedness allowing us to compare among responses known to interact with each other and estimated simultaneously. This approach has the further benefit of adjusting our partial pooling (‘random effect’ of species) based on evolutionary distance, more strongly pooling closely related species, and only weakly pooling distantly related species (see Gaussian process models in Gelman et al., 2014).

A traditional interpretation of σ^2 s under Brownian Motion evolution, is an ‘evolutionary rate’ or phenotypic accumulation over time (Revell et al., 2008). In PGLS, σ_ϵ^2 is estimated for the model error term, which is distributed as a multivariate normal with VCV matrix given by $\sigma_\epsilon^2 \mathbf{\Sigma}_i$. Here, similar to our approach to λ , we estimate four σ^2 values, corresponding to each model parameter. In our particular case (i.e., modelling a phenological response to three environmental cues), σ_α^2 for the intercept could be interpreted as the phenological variation across species accumulated along evolution independently from the cues. The $\sigma_{\beta_1}^2$, $\sigma_{\beta_2}^2$, and $\sigma_{\beta_3}^2$, corresponding to model slopes, would represent the phylogenetic variance linked to species responses to each of the modelled cues (i.e., forcing, chilling, and photoperiod, respectively). This is, the variability in how species shift their phenology responding to temperature and light, accumulated along the evolutionary process and considered in concert.

References

- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. Bayesian Data Analysis. 3rd ed. CRC Press, New York.
- Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. *The American Naturalist* 163:84–96.
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.
- Revell, L. J. 2010. Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* 1:319–329.
- Revell, L. J., L. J. Harmon, and D. C. Collar. 2008. Phylogenetic signal, evolutionary process, and rate. *Systematic biology* 57:591–601.

Supporting Figures and Tables

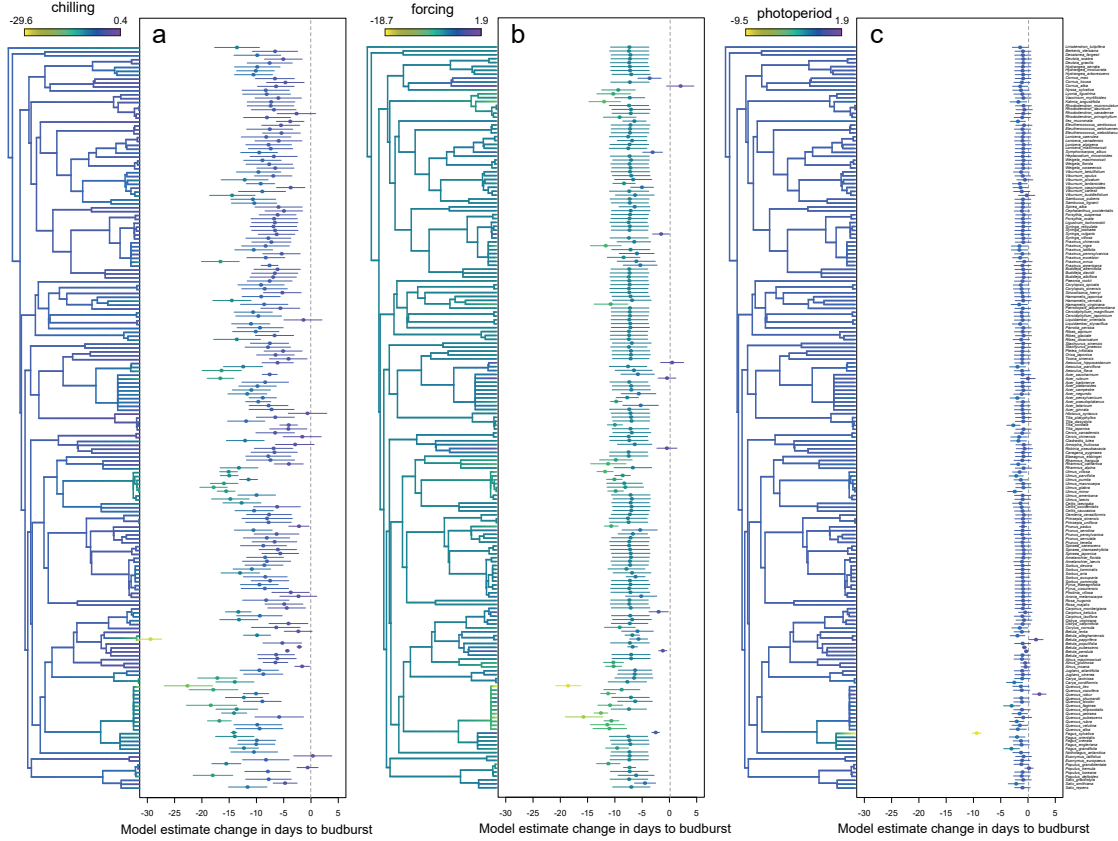


Figure 1: Non-phylogenetic phenological sensitivity to three environmental cues, chilling (a), forcing (b) and photoperiod (c) measured in change in days to budburst per standardized unit (z-transformation) of the cues across 191 tree species. Sensitivity estimates are computed by commonly used hierarchical model where phylogenetic distances are not accounted for ($\lambda = 0$). The same phylogenetic tree is shown in each panel, colored according to an estimation of ancestral character states, being the states at the tips the species' sensitivities to a cue. Species sensitivities are shown along with 50% Credible Intervals in the diagrams. Note that the color scale varies in each panel. Total tree depth is 81. My.

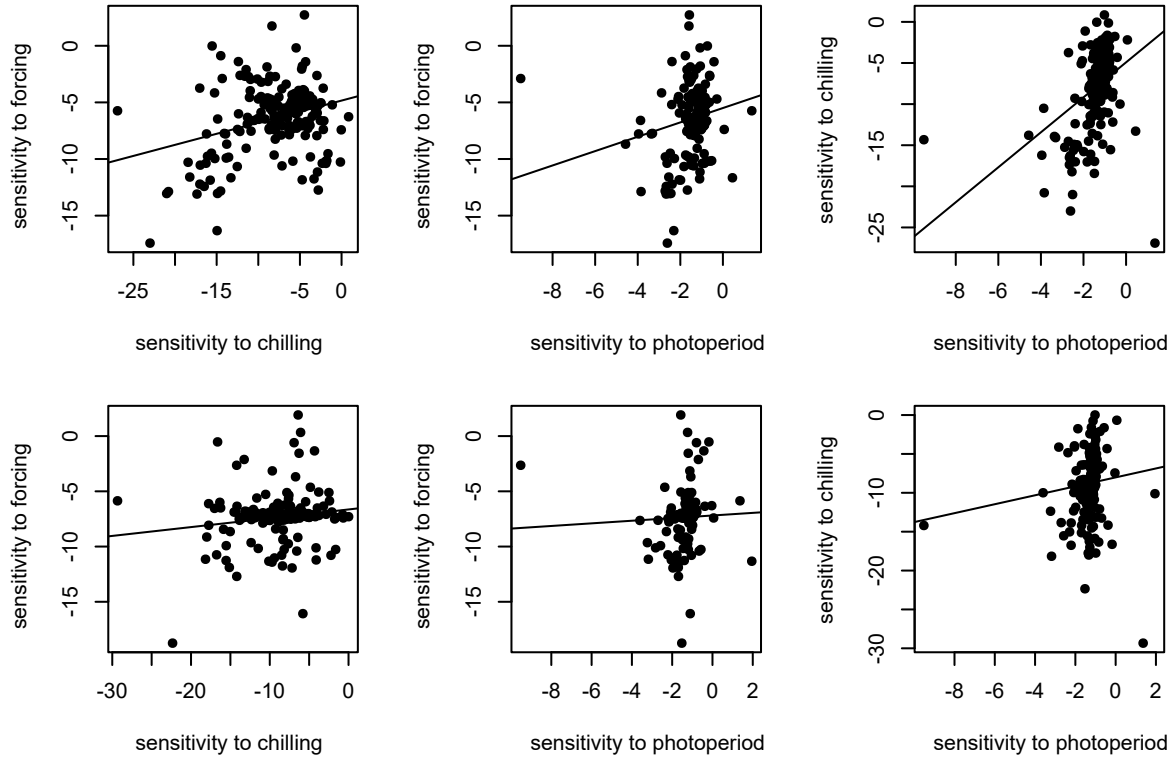


Figure 2: Correlations among estimated sensitivities to the environmental cues comparing forcing vs. chilling (a,d), forcing vs. photoperiod (b,e) and chilling vs. photoperiod (c,f). Upper panels show correlations among estimated sensitivities by the phylogenetic model and lower panels show results for the non-phylogenetic model.

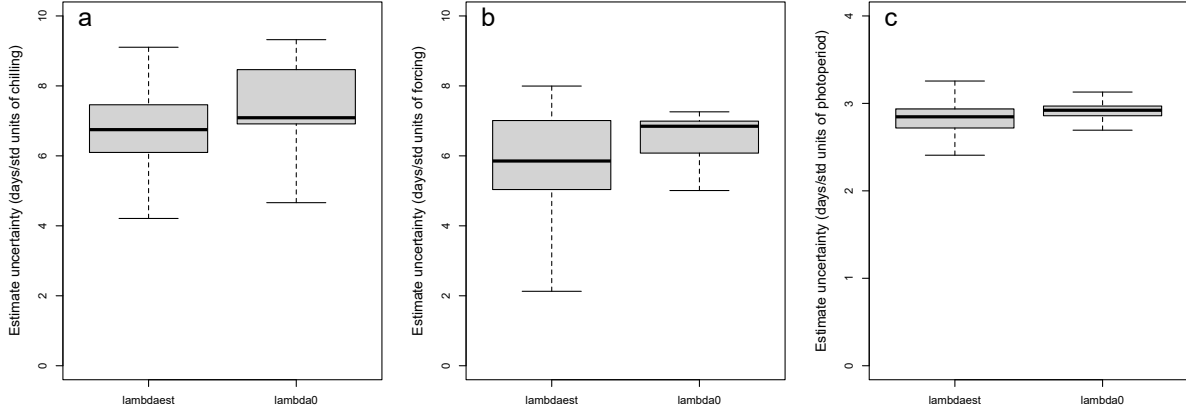


Figure 3: Comparison of uncertainty around estimated sensitivities to chilling (a), forcing (b) and photoperiod (c) of individual species between the phylogenetic model with estimated λ (lambdaest), and the non-phylogenetic model with $\lambda = 0$ (lambda0). The non-phylogenetic model increases uncertainty.

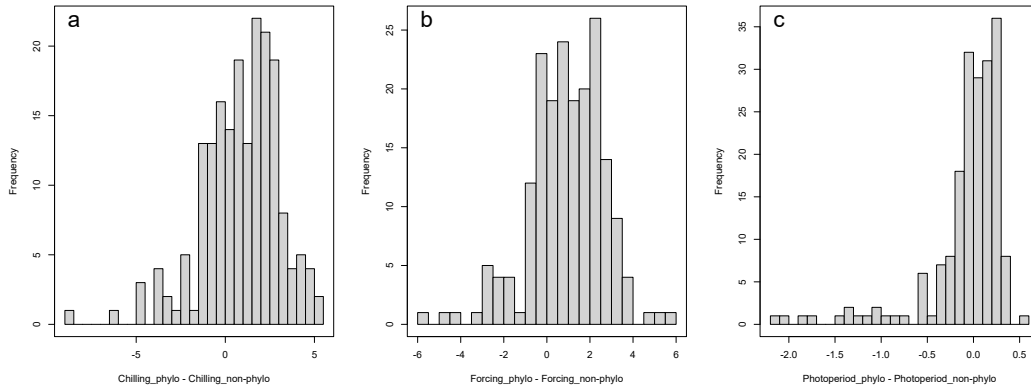


Figure 4: Bias in estimation of sensitivity to chilling (a), forcing (b) and photoperiod (c). Histograms show the difference between the phylogenetic model with estimated λ against the non-phylogenetic model with $\lambda = 0$. Positive values indicate that sensitivities estimated by the non-phylogenetic model are smaller than those estimated by the phylogenetic model.