CoringTreespotters Model notes

Christophe

September 8, 2025

15 April 2025

Quick notes after meeting with Victor, where we discussed the preliminary steps for building a model to address how juvenile trees respond to climate change. I'd like to frame this in terms of the recruitment capacity of tree saplings—looking at how their growth varies across years and how it's influenced by temperature and the length of the growing season (or perhaps not the length specifically... more on that below).

1 Come up with a model

What are my predictors?

We discussed which of the following would be the most relevant variable that relates to growth:

1. **Growing Degree Days (GDD)** which would be calculated between the budburst and budset date of each individual.

2. **Growing Season Length (GSL)** which could be calculated by:

   (a) Substracting budset DOY by budburst DOY or

   (b) By taking the number of days of each year when the mean (or max/min) was above 5°C (or maybe the nb of consecutive days when the temperature was above 5 or something like that).

<div align="center">

Likelihood:

$$y_i \sim \mathcal{N}(\mu_i, \sigma_y^2),$$
$$\mu_i = \alpha + \alpha_{\text{spp}} + \alpha_{\text{spp[ids]}} + \beta \cdot gdd$$

Simulated coefficients:

$$\alpha = 1.5$$
$$\alpha_{\text{spp}} \sim \mathcal{N}(0, \sigma_{\text{spp}})$$
$$\alpha_{\text{spp[ids]}} \sim \mathcal{N}(\alpha_{\text{spp}}, \sigma_{\text{ids}})$$
$$\beta = 0.4,$$
$$\sigma_y = 0.2$$
$$\sigma_{\text{spp}} = 0.3$$
$$\sigma_{\text{ids}} = 0.15$$
$$gdd \sim \mathcal{N}(1800, 200)$$

</div>

Where:

α: overall intercept

$\alpha_{\text{spp},j}$: each species intercept deviation from the overall intercept

$\alpha_{\text{ids},i(j)}$: each ids interecept deviation from the overall intercept, nested within each species

β: overall slope gdd: growing degree days. Measured from budburst to budset (theoretical start and end of tree primary growth)

Lizzie's notation after bayesian working group:

Nested on the intercept without provenance

$y_i \sim \text{normal}(\mu, \sigma_y)$

$\mu = \alpha + \alpha_{sp[id]} + \beta * gdd$

$\alpha_{sp} \sim \text{normal}(0, \sigma_{\alpha sp})$

$\alpha_{sp[id]} \sim \text{normal}(\alpha_{sp}, \sigma_{\alpha id})$

Nested on the intercept with provenance—Option 1

$y_i \sim \text{normal}(\mu, \sigma_y)$

$\mu = \alpha + \alpha_{sp[id]} + \alpha_{prov} + \beta * gdd$

$\alpha_{sp} \sim \text{normal}(0, \sigma_{\alpha sp})$

$\alpha_{sp[id]} \sim \text{normal}(\alpha_{sp}, \sigma_{\alpha id})$

$\alpha_{prov} \sim \text{normal}(0, \sigma_{\alpha prov})$

Nested on the intercept without provenance—Option 2

$y_i \sim \text{normal}(\mu, \sigma_y)$

$\mu = \alpha + \alpha_{sp} + \alpha_{prov} + \alpha_{id} + \beta * gdd$

$\alpha_{sp} \sim \text{normal}(0, \sigma_{\alpha sp})$

$\alpha_{prov} \sim \text{normal}(0, \sigma_{\alpha prov})$

$\alpha_{id} \sim \text{normal}(0, \sigma_{\alpha id})$

Notes on option 1 vs option 2: Charles pointed out during the working group that opt 1 and 2 were actually the same thing. After chatting with Lizzie on 8-09-2025, we concluded that he was right and to simplify the code, we will use option 2.

Model update after simulating the data

Model Specification

Data:

- $i = 1, \ldots, N$ (observations)

- $j[i] = 1, \ldots, J$ (site indices for each observation)

- y_i: Tree ring width measurement

- x_i: Growing Degree Days (GDD) covariate

Parameters:

- Site-level intercepts: α_j $(j = 1, \ldots, J)$

- Site-level slopes: β_j $(j = 1, \ldots, J)$

- Population means: μ_α, μ_β

- Population variances: σ_α, σ_β

- Measurement error: σ_y

Model Structure

Measurement model:
$$y_i \sim \mathcal{N}(\hat{y}_i, \sigma_y^2)$$
$$\hat{y}_i = \alpha_{j[i]} + \beta_{j[i]} \cdot x_i$$

Site-level parameters:
$$\alpha_j \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$$
$$\beta_j \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2)$$

Hyperpriors:
$$\sigma_y \sim \text{Uniform}(0, 100)$$
$$\sigma_\alpha \sim \text{Uniform}(0, 100)$$
$$\sigma_\beta \sim \text{Uniform}(0, 100)$$
$$\mu_\alpha \sim \mathcal{N}(0, 100) \quad \text{(weakly informative)}$$
$$\mu_\beta \sim \mathcal{N}(0, 100) \quad \text{(weakly informative)}$$

Key Features

- **Partially pooled slopes:** Each site j has its own GDD response β_j, drawn from a shared normal distribution

- **Adaptive shrinkage:** The estimated σ_β controls pooling strength (smaller values $\Rightarrow$ stronger pooling toward μ_β)

- **Interpretation:**
  - μ_β: Overall average effect of GDD on tree growth
  - σ_β: Between-site variability in GDD sensitivity
  - β_j: Site-specific growth response to GDD

2 Simulate data

I arbitrarily decided to sum up the GDD between DOY 100 and 300 for data simulation. yhat: combine the effects that I generated from the grand mean, the year effect (other than variation of GDD), individual effect, species response to GDD and the overall error, which returns my yhat i.e. ring width in mm.

3 Set your priors

4 Run model on empirical data

5 Perform retrodictive checks using the model fit to your empirical data

6 Mis

Below I will be stating the o bvious because I want to make sure I understand the notations and stuff. The steps will be notes and thoughts that I have when I code my model in R.