

CoringTreespotters Model notes

Christophe

May 7, 2025

15 April 2025

Quick notes after meeting with Victor where we talked about preliminary steps of building a model to answer the question of how juvenile trees respond to climate change. I would like to approach this as a recruitment capacity of tree saplings and how their growth varies across years and how it's impacted by temperature and growing season length (or maybe not length... more on that below)

Below I will be stating the obvious 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.

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) Subtracting 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).

Preliminary model

$$\log(w_{i,t}) \sim \text{normal}(X, \sigma) \quad (1)$$

$$\alpha + \alpha_{sp[i]} + \beta_{sp[i]} X_{i,t} + \dots \quad (2)$$

What kind of pooling? Partial pooling since they are all within the same family?

2 Simulate data

I arbitrarily decided to sum up th GDD between 100 and 250 days for now in order to simplify data simulation.

Common Bayesian Notation

- θ — Parameters of the model (can be a vector of parameters)
- y — Observed data
- x — Predictor variables or covariates
- $p(\theta)$ — Prior distribution over parameters
- $p(y \mid \theta)$ — Likelihood: the probability of the data given parameters
- $p(\theta \mid y)$ — Posterior distribution: updated beliefs about parameters after observing data
- $p(y)$ — Marginal likelihood or evidence: probability of the data under the model

- $p(y \mid x, \theta)$ — Likelihood conditional on predictors
- $p(\theta, y)$ — Joint distribution of parameters and data
- μ — Mean (typically of a normal distribution or prior)
- σ — Standard deviation (scale of a distribution)
- τ — Precision (inverse of variance: $\tau = 1/\sigma^2$)
- $\mathcal{N}(\mu, \sigma^2)$ — Normal (Gaussian) distribution with mean μ and variance σ^2
- $\text{Beta}(\alpha, \beta)$ — Beta distribution (common prior for probabilities)
- $\text{Gamma}(\alpha, \beta)$ — Gamma distribution (common prior for positive variables)
- $\text{InvGamma}(\alpha, \beta)$ — Inverse Gamma distribution (common prior for variances)
- $\text{Bernoulli}(p)$ — Bernoulli distribution for binary data
- $\text{Binomial}(n, p)$ — Binomial distribution
- $\text{Poisson}(\lambda)$ — Poisson distribution for count data
- \sim — "Is distributed as" (e.g., $y \sim \mathcal{N}(\mu, \sigma^2)$)
- \propto — Proportional to (used in unnormalized posteriors: $p(\theta \mid y) \propto p(y \mid \theta)p(\theta)$)

3 Set your priors

4 Run model on empirical data

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