Estimating Tree Growth Models from Complex Forest Monitoring Data: Appendix C

Melissa Eitzel, John Battles, Robert York, Jonas Knape, Perry de Valpine

Appendix C: Additional details on statistical model and algebra regarding standardization of explanatory variables

This appendix includes a more detailed version of the model specification, including explicitly stating normal distributions for random effects, as well as the algebra for how the variables were centered and scaled, and then later unscaled (but left centered) in the results.

Full model description

Recall that subscript i is for compartment, j is for plot, and k is for tree; superscript m indicates one of several explanatory variables. The process model relating size at $(t+1)$ to the other variables is:

$$x_{ijk}(t+1) \sim \mathcal{N}(\alpha_{ijk}(t) + \beta_{ijk}(t)x_{ijk}(t) + \sum_m \gamma^m z_{ij}^m(t) + \sum_m \kappa^m z_{ij}^m(t)x_{ijk}(t), \sigma_\varepsilon^2) \tag{C.1}$$

x_{ijk} is size, while $z_{ij}^m(t)$ are the other explanatory variables. These are measured at plot and/or year level: insolation, topographic slope, elevation, and soil category are all measured at plot level, i.e. z_{ij}^{insol}, z_{ij}^{tslope}, z_{ij}^{elev}, and a group of five indicator variables representing a tree's soil type (z_{ij}^C, z_{ij}^H, z_{ij}^{HB}, z_{ij}^{HM}, and z_{ij}^J) ; basal area is measured at plot and year level, $z_{ij}^{ba}(t)$; and annual water deficit is measured at year level, $z^{def}(t)$. The variables z^m, x, and y are centered and scaled as detailed in the next section.

We assume that size in the next year is a linear function of the $z_{ij}^m(t)$ and have parameters for slope γ^m and interaction with size κ^m. We also assume that size in the next year $x_{ijk}(t+1)$ is a linear function of size in the previous year, $x_{ijk}(t)$, with soil type-dependent slope and intercept (e.g., for Cohassett: slope $a_C = \beta_{ijk}(t) + \kappa^C$ and intercept $b_C = \alpha_{ijk}(t) + \gamma^C$). We report the average growth increment $\bar{b}$ and average effect of size on growth increment $\bar{a} - 1$, which are weighted averages over soil types and for average values of explanatory variables. Residual error, with variance σ_ε^2, accounts for additional unexplained variation in growth.

At the next hierarchical level, we model the collective random effects on intercept $\alpha_{ijk}(t)$ and the random effects on size slope $\beta_{ijk}(t)$ as a combination of random effects for tree (q_{ijk}^α and q_{ijk}^β), plot (p_{ij}^α and p_{ij}^β), compartment (c_i^α and c_i^β), and year ($w^\alpha(t)$ and $w^\beta(t)$). The intercept effects reflect differences in overall growth increment while the slope effects reflect differences in growth as a function of size. The random effect intercept and slope for a specific tree is determined by the random tree, plot, compartment and year effects as follows:

$$\alpha_{ijk}(t) = q_{ijk}^\alpha + c_i^\alpha + p_{ij}^\alpha + w^\alpha(t) \quad \beta_{ijk}(t) = q_{ijk}^\beta + c_i^\beta + p_{ij}^\beta + w^\beta(t) \tag{C.2}$$

The random effects for years, compartments, plots, and individuals follow normal distributions,

$$\begin{aligned} w^{\alpha}(t) \sim \mathcal{N}(0, \sigma^2_{\alpha,w}), \quad c_i^{\alpha} \sim \mathcal{N}(0, \sigma^2_{\alpha,c}), \quad p_{ij}^{\alpha} \sim \mathcal{N}(0, \sigma^2_{\alpha,p}), \quad q_{ijk}^{\alpha} \sim \mathcal{N}(0, \sigma^2_{\alpha,q}), \\ w^{\beta}(t) \sim \mathcal{N}(0, \sigma^2_{\beta,w}), \quad c_i^{\beta} \sim \mathcal{N}(0, \sigma^2_{\beta,c}), \quad p_{ij}^{\beta} \sim \mathcal{N}(0, \sigma^2_{\beta,p}), \quad q_{ijk}^{\beta} \sim \mathcal{N}(0, \sigma^2_{\beta,q}) \end{aligned} \tag{C.3}$$

respectively. At each level of nesting, random effects are assumed to be independent.

$x_{ijk}(t)$ is not measured perfectly (a latent state), and so $y_{ijk}(t)$ is the size actually measured with some observation error:

$$y_{ijk}(t) \sim \mathcal{N}(x_{ijk}(t), \sigma^2_{\mathrm{DBH}}) \tag{C.4}$$

Standardizing and unscaling of explanatory variables

One can combine equations 1 and 4 in a more algebraic form:

$$y_{ijk}(t+1) = \alpha_{ijk} + \beta_{ijk} x_{ijk}(t) + \sum_m \gamma^m z_{ij}^m(t) + \sum_m \kappa^m z_{ij}^m(t) x_{ijk}(t) + \epsilon_{res} + \epsilon_{obs} \tag{C.5}$$

(recall that ϵ_{res} and ϵ_{obs} are normally distributed with mean zero and variance σ^2_{ε} and σ^2_{DBH}, respectively)

We can expand this using equation C.2:

$$\begin{aligned} y_{ijk}(t+1) = q_{ijk}^{\alpha} + c_i^{\alpha} + p_{ij}^{\alpha} + w^{\alpha}(t) + \left(q_{ijk}^{\beta} + c_i^{\beta} + p_{ij}^{\beta} + w^{\beta}(t)\right) x_{ijk}(t) \\ + \sum_m \gamma^m z_{ij}^m(t) + \sum_m \kappa^m z_{ij}^m(t) x_{ijk}(t) + \epsilon_{res} + \epsilon_{obs} \end{aligned} \tag{C.6}$$

In reality, we are estimating those slopes, intercepts, random effects, and random effect standard deviations for standardized covariates. For true latent size x and measured size y, we have standardized by subtracting the mean and dividing by the standard deviation for the measured sizes y (μ_y and σ_y). We have done the same for the covariates z^m (with σ_{z^m} and μ_{z^m}). When we unscale (but leave the variables centered) for interpretation, each of the continuous covariates will need to be unscaled by its respective standard deviation.

This is how we have scaled the original variables:

$$x'_{ijk}(t) = \frac{x_{ijk}(t) - \mu_y}{\sigma_y} \quad y'_{ijk}(t) = \frac{y_{ijk}(t) - \mu_y}{\sigma_y} \quad {z^m}'_{ij}(t) = \frac{z_{ij}^m(t) - \mu_{z^m}}{\sigma_{z^m}} \tag{C.7}$$

The equation we have estimated the parameters for is in the primed variables:

$$\begin{aligned} y'_{ijk}(t+1) = {q_{ijk}^{\alpha}}' + c_i^{\alpha\prime} + {p_{ij}^{\alpha}}' + w^{\alpha\prime}(t) + \left({q_{ijk}^{\beta}}' + c_i^{\beta\prime} + {p_{ij}^{\beta}}' + w^{\beta\prime}(t)\right) x'_{ijk}(t) \\ + \sum_m \gamma^{m\prime} z_{ij}^{m\prime}(t) + \sum_m \kappa^{m\prime} z_{ij}^{m\prime}(t) x_{ijk}(t)' + \epsilon'_{res} + \epsilon'_{obs} \end{aligned} \tag{C.8}$$

Therefore, we wish to know the relationship between these primed parameters and equations in unscaled variables. When we unscale the covariates, we only wish to multiply by the standard deviation to restore the units of each covariate. We leave the variables centered, such that interpretation of the parameter estimates refers to the covariates at their mean values. Thus, we want an equation in the following double-primed variables:

$$x''_{ijk}(t) = x_{ijk}(t) - \mu_y \quad y''_{ijk}(t) = y_{ijk}(t) - \mu_y \quad z_{ij}^m(t)'' = z_{ij}^m(t) - \mu_{z^m} \tag{C.9}$$

First let us write the equation whose parameters we estimated (equation C.8) in terms of the unprimed original variables:

$$\begin{aligned}\tfrac{y_{ijk}(t+1)-\mu_y}{\sigma_y} &= {q^{\alpha}_{ijk}}' + c^{\alpha\prime}_i + {p^{\alpha}_{ij}}' + w^{\alpha\prime}(t) + \left({q^{\beta}_{ijk}}' + c^{\beta\prime}_i + {p^{\beta}_{ij}}' + w^{\beta\prime}(t)\right)\tfrac{x_{ijk}(t)-\mu_y}{\sigma_y} \\ &+ \sum_m \gamma^{m\prime}\frac{z^m_{ij}(t)-\mu_{z^m}}{\sigma_{z^m}}(t) + \sum_m \kappa^{m\prime}\frac{x_{ijk}(t)-\mu_y}{\sigma_y}\frac{z^m_{ij}(t)-\mu_{z^m}}{\sigma_{z^m}} + \epsilon'_{res} + \epsilon'_{obs}\end{aligned} \tag{C.10}$$

Substituting x'', y'' and $z^{m\prime\prime}$ into the primed equation, we have:

$$\begin{aligned}\tfrac{y''_{ijk}(t+1)}{\sigma_y} &= {q^{\alpha}_{ijk}}' + c^{\alpha\prime}_i + {p^{\alpha}_{ij}}' + w^{\alpha\prime}(t) + \left({q^{\beta}_{ijk}}' + c^{\beta\prime}_i + {p^{\beta}_{ij}}' + w^{\beta\prime}(t)\right)\tfrac{x''_{ijk}(t)}{\sigma_y} \\ &+ \sum_m \gamma^{m\prime}\frac{{z^m_{ijk}}''(t)}{\sigma_{z^m}} + \sum_m \kappa^{m\prime}\frac{x''_{ijk}(t)}{\sigma_y}\frac{{z^m_{ijk}}''(t)}{\sigma_{z^m}} + \epsilon'_{res} + \epsilon'_{obs}\end{aligned} \tag{C.11}$$

Now, multiply through by σ_y:

$$\begin{aligned}y''_{ijk}(t+1) &= \sigma_y\left({q^{\alpha}_{ijk}}' + c^{\alpha\prime}_i + {p^{\alpha}_{ij}}' + w^{\alpha\prime}(t)\right) + \sigma_y\left({q^{\beta}_{ijk}}' + c^{\beta\prime}_i + {p^{\beta}_{ij}}' + w^{\beta\prime}(t)\right)\tfrac{x''_{ijk}(t)}{\sigma_y} \\ &+\sigma_y\sum_m \gamma^{m\prime}\frac{z^{m\prime\prime}_{ij}(t)}{\sigma_{z^m}} + \sigma_y\sum_m \kappa^{m\prime}\frac{x''_{ijk}(t)}{\sigma_y}\frac{z^{m\prime\prime}_{ij}(t)}{\sigma_{z^m}} + \sigma_y\epsilon'_{res} + \sigma_y\epsilon'_{obs}\end{aligned} \tag{C.12}$$

After canceling σ_y, explanatory variable slope and intercept parameters scaled as follows:

$$\alpha_{ijk}''(t) = \sigma_y\alpha_{ijk}'(t) \quad \beta_{ijk}''(t) = \beta_{ijk}'(t) \quad \gamma^{m\prime\prime} = \tfrac{\sigma_y}{\sigma_{z^m}}\gamma^{m\prime} \quad \kappa^{m\prime\prime} = \tfrac{1}{\sigma_{z^m}}\kappa^{m\prime} \tag{C.13}$$

Observation error and residual error epsilons scale like intercepts:

$$\epsilon''_{res} = \sigma_y\epsilon'_{res} \quad \epsilon''_{obs} = \sigma_y\epsilon'_{obs} \tag{C.14}$$

Random effects scale as follows (slope effects remain unscaled, intercept effects are multiplied by size standard deviation σ_y):

$$\begin{aligned}w^{\alpha}(t)'' = \sigma_y w^{\alpha}(t)', \quad c^{\alpha\prime\prime}_i = \sigma_y c^{\alpha\prime}_i, \quad {p^{\alpha}_{ij}}'' = \sigma_y {p^{\alpha}_{ij}}', \quad {q^{\alpha}_{ijk}}'' = \sigma_y {q^{\alpha}_{ijk}}', \\ w^{\beta}(t)'' = w^{\beta}(t)', \quad c^{\beta\prime\prime}_i = c^{\beta\prime}_i, \quad {p^{\beta}_{ij}}'' = {p^{\beta}_{ij}}', \quad {q^{\beta}_{ijk}}'' = {q^{\beta}_{ijk}}'\end{aligned} \tag{C.15}$$

We assume that standard deviations for random effects (i.e. $\sigma^2_{\alpha,w}$, $\sigma^2_{\beta,w}$, $\sigma^2_{\alpha,q}$, $\sigma^2_{\beta,q}$, $\sigma^2_{\alpha,c}$, $\sigma^2_{\beta,c}$, $\sigma^2_{\alpha,p}$, $\sigma^2_{\beta,p}$, equation C.3) scale in the same manner as the random effects themselves.