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

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

Appendix D: Details on model estimation and the evaluation of model results

This appendix gives details on model estimation and the evaluation of model results, including specification of priors and selection of initial values for chains, results of Gelman-Rubin convergence assessments; comparisons of posteriors and priors; and tradeoffs between observation and residual standard deviations.

Prior specification

All but one of the priors used in estimation of the model were uninformative. Uninformative priors for means were a normal distribution with a high standard deviation (low precision: 1.0e-6) and for standard deviations, uniform priors over the range [0,100]. The first latent size of each tree was given a non-informative (normal, low precision) prior.

For the observation error, we observed very poor mixing for priors which allowed small values (including zero), as the system struggled to move away from the scenario in which the latent states exactly matched the observations. The resulting observation error posteriors were sensitive to the degree to which the priors excluded small values (Fig. D1).

Because some observation error is certainly present in our system, both due to the precision of the diameter tape and due to bark loss and human error, we chose two different reasonably flat, otherwise uninformative priors with a nonzero lower limit. We chose this lower limit based on the rounding error inherent in the diameter tape, which we considered to be a minimum. Reasoning that the 0.1 inch increment labeled on the tape leads to true observations that are 0.05 inches in either direction being uniformly attributed to the mark on the tape in the center of that interval, and assuming x is the truth, while y is the measurement, we have:

$$\begin{aligned} y &= x + u \\ u &\sim \text{uniform}(-a/2, a/2) \end{aligned} \tag{D.1}$$

(Where $a = 0.1$) We are essentially taking the case where the only error is due to rounding on the tape to mean that $\epsilon_{obs} \sim \text{uniform}(-a/2, a/2)$. To establish the correct standard deviation associated with this model for observation error (to then use as a lower bound in a prior for this node in the larger model), we turn to the definition of the uniform distribution from $-a/2$ to $a/2$:

$$f(u) = \begin{cases} 1/a, & \text{for } -a/2 \leq u \leq a/2 \\ 0, & \text{otherwise} \end{cases} \tag{D.2}$$

And the definition of variance as the second central moment:

$$\begin{aligned} V[u] &= \int_{-\frac{a}{2}}^{\frac{a}{2}} u^2 f(u) du \\ V[u] &= \int_{-\frac{a}{2}}^{\frac{a}{2}} u^2 \tfrac{1}{a} du \\ V[u] &= \tfrac{u^3}{3} \tfrac{1}{a} \Big|_{-\frac{a}{2}}^{\frac{a}{2}} \\ V[u] &= \tfrac{a^2}{24} - \tfrac{-a^2}{24} \\ V[u] &= \tfrac{a^2}{12} \end{aligned} \tag{D.3}$$

So, the standard deviation is $a/\sqrt{12}$. Plugging in $a = 0.1$ and converting from inches to our standardized scale, we have the lower limit for the standard deviation set at 0.0036. For our two candidate priors, we therefore use a uniform distribution from [0.0036, 100] on the observation error standard deviation and a gamma prior with shape and rate (.01,0.0001) on the observation error precision. These parameters for the gamma on precision gave rise to a threshold in the standard deviation at approximately the right location. The gamma prior on the precision results in an inverse gamma prior on the standard deviation.

We also tried a uniform prior on observation error standard deviation from [0,100] on a simplified model with several covariates left out (elevation, insolation, and annual deficit), which has the effect of allowing values as low as zero (and which mixes slightly better due to the simpler covariate structure). Finally, we tried a gamma with shape and rate (0.0001,0.0001), which has the effect of excluding values below 0.005 on the standardized scale (Fig. D1).

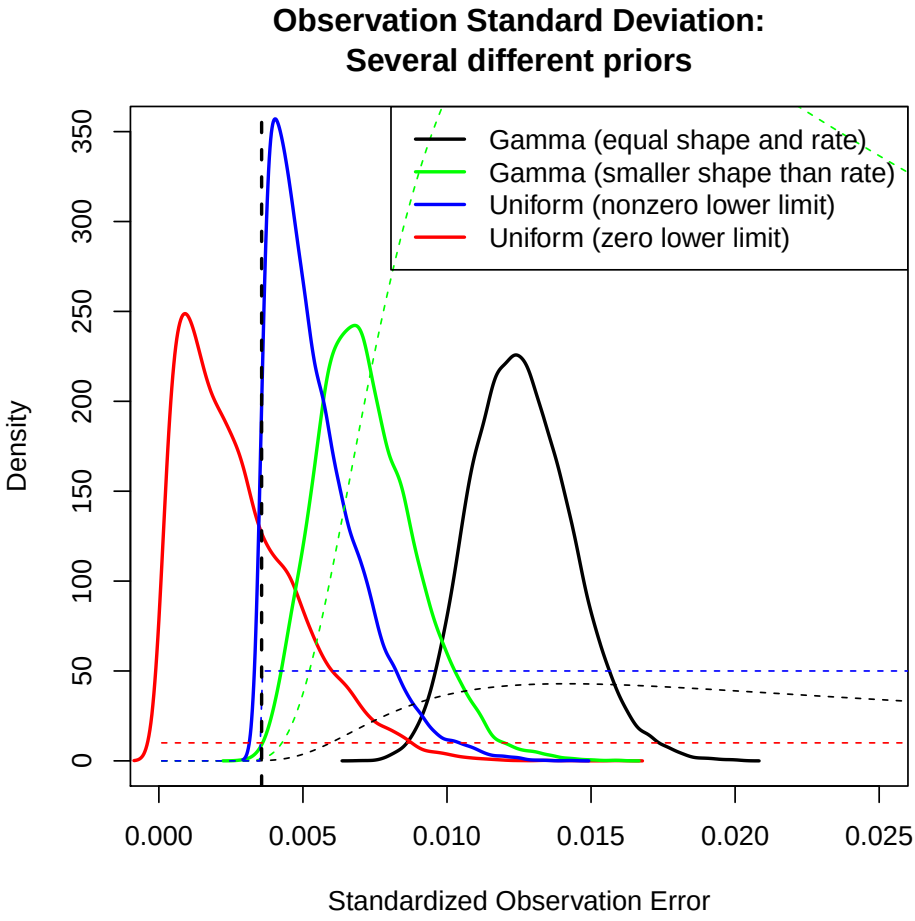

Figure D1: Four different models with different observation error σ_{DBH} priors. The blue, black, and green lines are priors (dashed) and posteriors (solid) for the full model described in Appendix C; the red lines are for a reduced model which does not include annual water deficit, elevation, or insolation, but is otherwise similar to the other models. The green and black posteriors correspond to gamma priors, and the red and blue to uniform priors. The priors (dashed lines) have been multiplied by powers of ten in order to be visible on the same scale as the posteriors; they are not all multiplied by the same factor in order to be visible alongside each other as well. Note that the green gamma prior results in a much more quickly increasing allowance of small values, hence the position of the green posterior closer to zero than the black posterior. Similarly, note that the red posterior (corresponding to the prior which does not exclude zero) is very close to zero, while the blue posterior is very close to the diameter tape-based lower limit established above, which is shown in a dashed black vertical line. Our model and estimation procedure biases the observation error as close to zero as we allow it.

Though we do not show chains for all these models in these appendices, the general behavior was that mixing was better the higher the threshold of exclusion (the more small values excluded). The lower the threshold, the worse the mixing. So mixing was best with the inverse gamma with equal shape and rate, and worst for the uniform prior allowing zero values.

We take this to mean that the observation error cannot truly be estimated with our model, data, and MCMC sampler (OpenBUGS); but for the results in the main paper we chose the uniform prior with the cutoff at 0.0036 because we believe that at least some observation error is reasonable and will improve other estimates.

Estimation

OpenBUGS was run through R using R2WinBUGS (Sturtz et al. 2005) long enough to achieve sufficient mixing. The results in the main paper are from a five day run on a Dell Studio XPS with an Intel Core i7 CPU, 2.93 GHz (4 cores) and 12 GB of RAM, running 64-bit Windows Vista Ultimate. In order to use R2WinBUGS, we had to use the 32-bit version of R. See online supplement for model code, data, and initial values and Fig. D2 for trace plots of MCMC chains. With the included supplements, it is possible to reproduce our parameter estimates, though the MCMC may need to be run for several weeks depending on computer speeds. 95% credible intervals were calculated using the coda package (Plummer et al. 2010).

Initial values

Due to slow mixing, randomly generated initial values from BUGS tended to produce chains that would not converge. Thus, initial values were primarily generated from interpolating the sizes of trees using smoothing splines between inventories (Wood 2006) and fitting linear mixed effects models for random effects using R package lme4 (Bates and Maechler 2010). For the two chains used to test convergence, initial values were generated by adding a small amount of randomly generated variation to the original initial values (normal in the case of the slopes and intercepts, uniform in the case of the variances).

Assessment of convergence: Gelman-Rubin statistics and trace plots

Assessing the mixing and convergence of chains is central to MCMC. If a chain is not mixing and has not converged, then the results are not interpretable. Gelman-Rubin diagnostics require running multiple chains and comparing them to assess convergence through a plot or single diagnostic value. These functions are built into OpenBUGS; we do not show the plots here but do display the convergence statistics, which should be close to 1 (Table D1). We also give examples in Fig. D2 of some of our trace plots which show the actual values in the chains. In this case, the analyst is looking for a lack of pattern in the trace: no trends and a trace that looks like noise. Based on the Gelman-Rubin diagnostics and our trace plots, we conclude that mixing is adequate.

Table D1: Gelman-Rubin statistics for tracked nodes. Names of variables are internal to the model. "s" refers to a main effect γ^m for a covariate z^m, or to a slope random effect standard deviation σ_β; "int.size" refers to interactions κ^m between size x and covariates z^m, and "i" refers to an intercept random effect σ_α or to the overall mean for Cohassett soils, "i.overall" (γ^C). See supplement for model code, where these names are used.

Name	PointEstimate	UpperCredibleInterval
res.sd	1.00129	1.00611
obs.sd	1.00028	1.00156
i.overall	1.0049	1.02105
s.size	1.0005	1.00072
i.year.sd	1.00633	1.03022
i.tree.sd	1.00004	1.00006
i.plot.sd	1.00015	1.0008
i.comp.sd	1.00396	1.00886
s.year.sd	1.00259	1.01073
s.tree.sd	1.0281	1.07996
s.plot.sd	1	1.00005
s.comp.sd	1.00029	1.00149
s.slope	0.999966	0.999998
s.ba	1.00008	1.00059
soil.mean	1.00418	1.00423
s.anndef	1.0018	1.00714
int.size.ba	1.00001	1.00012
int.insol.anndef	1.0036	1.00934
s.elev	0.99999	1.00006
s.insol	1.00046	1.00238
int.size.insol	1.00073	1.00326
int.size.slope	0.999984	1.00002
int.size.elev	0.999958	0.999959
int.size.anndef	1.02702	1.11724
int.size.soil	1.00132	1.0006
deviance	1.00029	1.00164

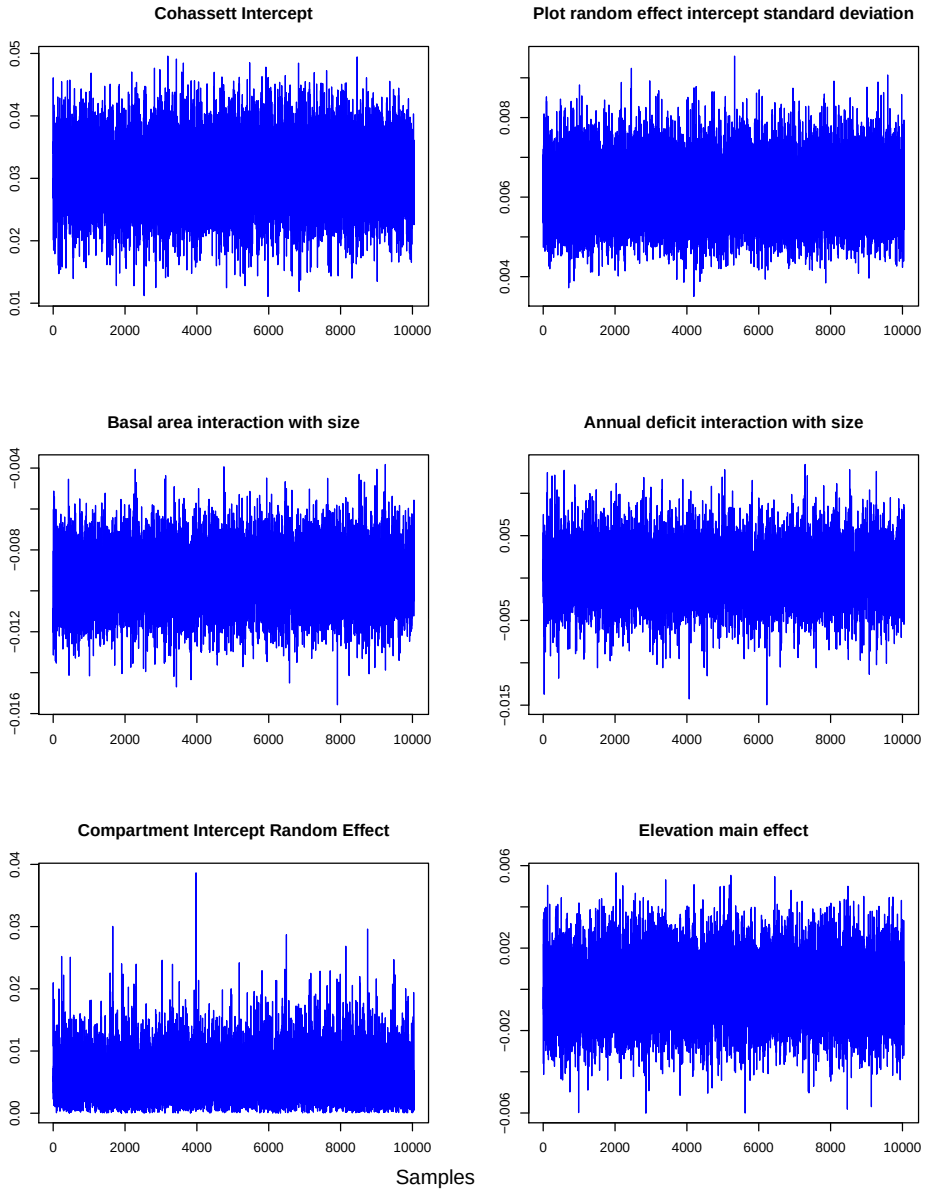

Figure D2: Trace plots showing the "white noise" behavior indicative of good mixing for six parameters.

Comparison of posteriors and priors: Bayesian learning figures

One way to assess the information in the data and the relative influence of the priors is to compare posterior distributions for parameters with their prior distributions and initial values. All but one of our priors (observation error) are uninformative, and the posteriors and priors for that estimate are shown in Fig. D1. All the other posteriors are clearly distinguished from their priors. Initial values, especially for the standard deviations, are typically clearly outside the main support of the posterior (the range of values where the posterior is concentrated), indicating that sensitivity to initial values was minimal within this range. The posteriors for both chains tend to overlap. When the initial values are very different from the posteriors, the two chain posteriors are indistinguishable due to their overlap.

Fig. D3 shows residual error and observation error standard deviations, Figs. D4 and D5 show random effect standard deviations, and Figs. D6 and D7 show covariate main effects and interactions with size.

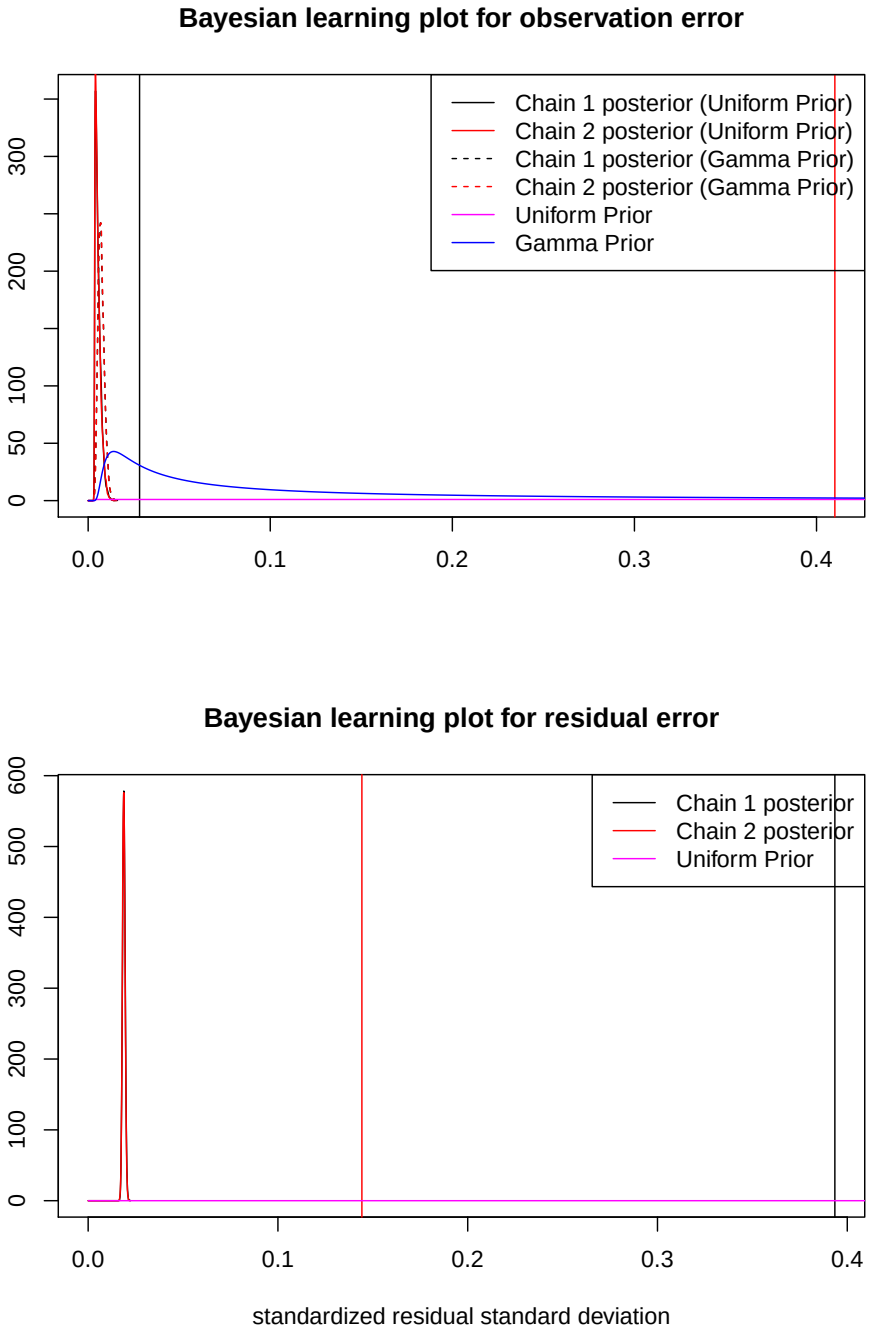

Figure D3: Bayesian learning figures for observation error σ_{DBH} and residual error σ_{ε}. Priors in observation error figure have been multiplied by 100 for visibility. The details of the priors for observation error are difficult to discern (See Fig. D1 for a closer look), but the initial values can clearly be quite large and the model will still converge to values close to zero. The red and black are two different chains; posteriors are shown with corresponding initial values as vertical lines in matching colors. Priors are shown in magenta (uniform), or blue (gamma). All posteriors are given on the standardized unitless scale. The two chain posteriors are indistinguishable due to their overlap on both figures, even with the different priors on observation error.

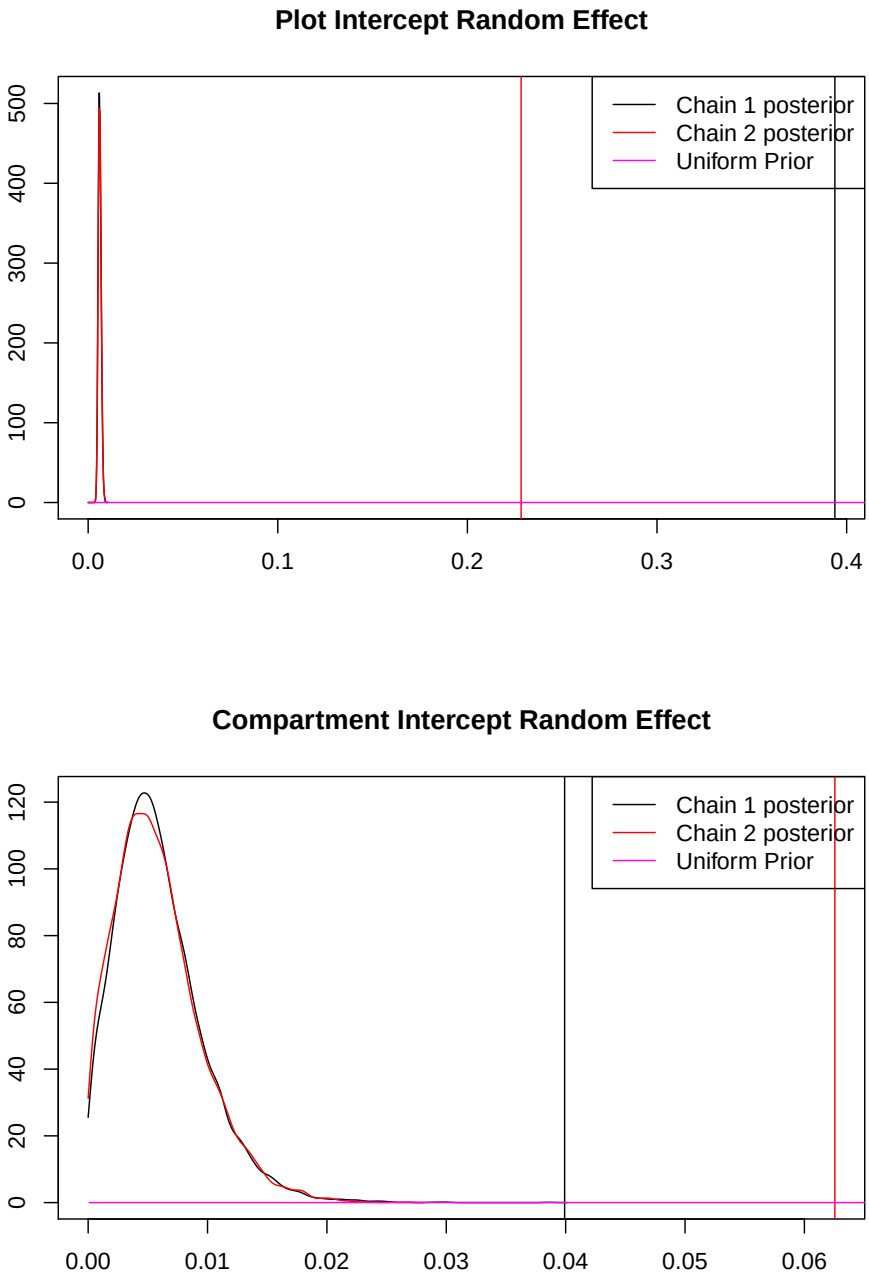

Figure D4: Bayesian learning figures for plot and compartment intercept random effect standard deviations ($\sigma_{\alpha,p}$ and $\sigma_{\alpha,c}$). The red and black are two different chains; posteriors are shown with corresponding initial values as vertical lines in matching colors. Uniform prior is shown in magenta. All posteriors are given on the standardized unitless scale. The two chain posteriors are indistinguishable due to their overlap for the plot intercept standard deviation.

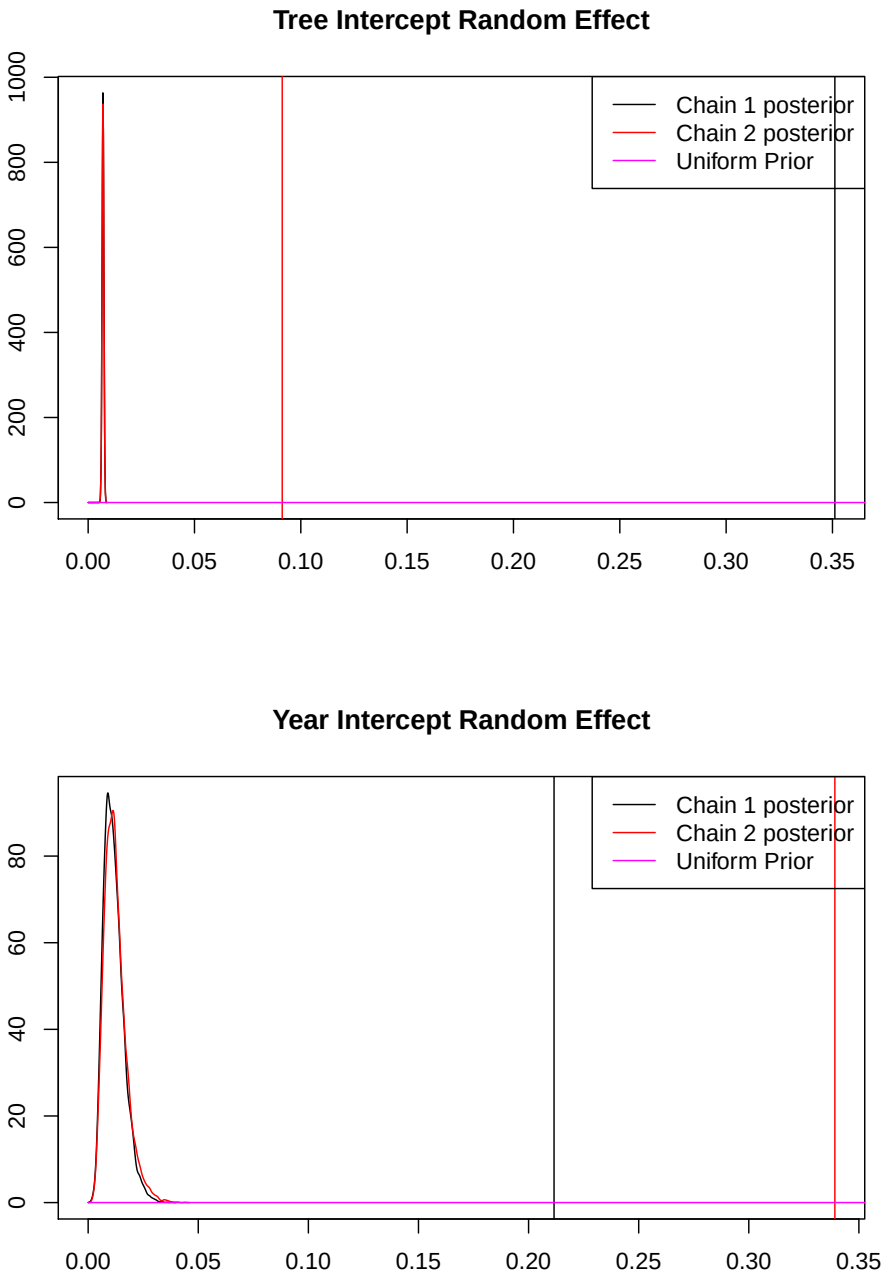

Figure D5: Bayesian learning figures for tree and year intercept random effect standard deviations ($\sigma_{\alpha,q}$ and $\sigma_{\alpha,w}$). The red and black are two different chains; posteriors are shown with corresponding initial values as vertical lines in matching colors. Uniform prior is shown in magenta. All posteriors are given on the standardized unitless scale. The two chain posteriors are indistinguishable due to their overlap for the tree intercept standard deviation.

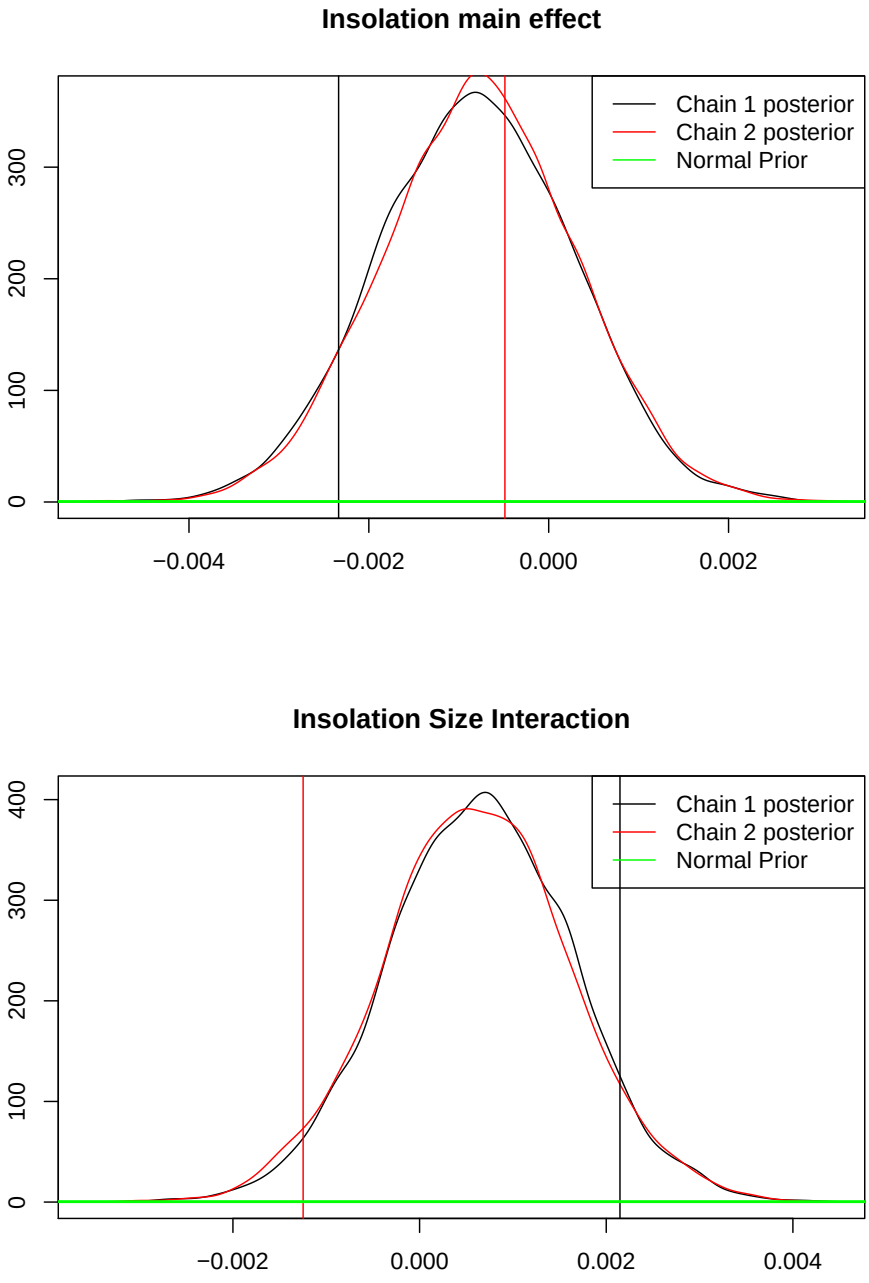

Figure D6: Bayesian learning figures for insolation main effect γ^{insol} and size interaction κ^{insol}. The red and black are two different chains; posteriors are shown with corresponding initial values as vertical lines in matching colors. Normal prior is shown in green. All posteriors are given on the standardized unitless scale.

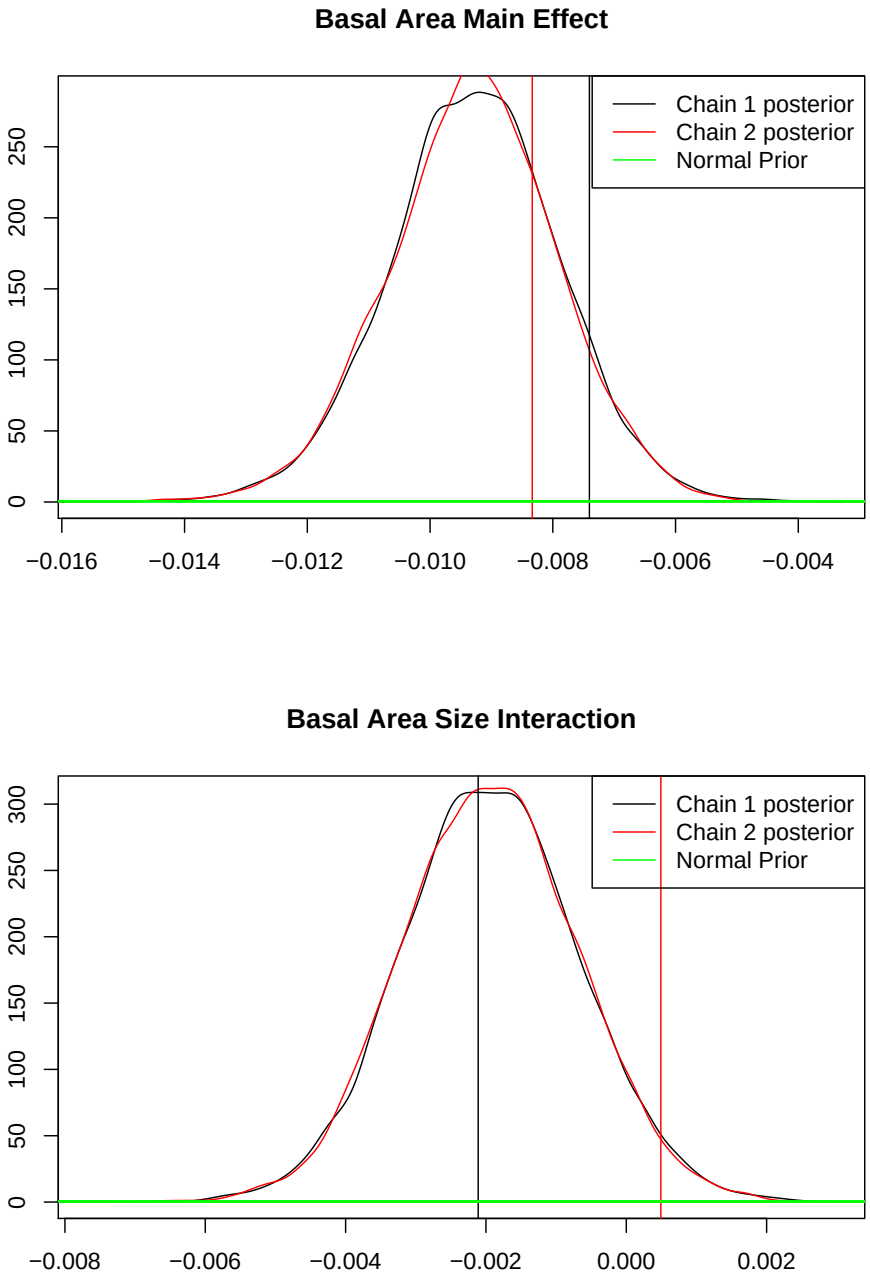

Figure D7: Bayesian learning figures for basal area main effect γ^{ba} and interaction with size κ^{ba}. The red and black are two different chains; posteriors are shown with corresponding initial values as vertical lines in matching colors. Normal prior is shown in green. All posteriors are given on the standardized unitless scale.

Tradeoff between observation and residual standard deviation

Observation standard deviation and residual standard deviation, while mathematically identifiable, practically speaking can be weakly identifiable, depending on the prior used for the observation error. For more informative priors (with a higher threshold, excluding a wider range of small values), these two nodes in the model do trade off in the MCMC. Fig. D8 shows MCMC samples from the two variables exhibiting this tradeoff, that when one node becomes larger, the other becomes smaller (they are slightly inversely related), but this effect is most pronounced in the model with the gamma prior with equal shape and rate. This correlation between observation and residual error is not extreme, however: the tradeoff is never so large that one node goes to zero when the other node is large, not even in the model where this tradeoff is most pronounced.

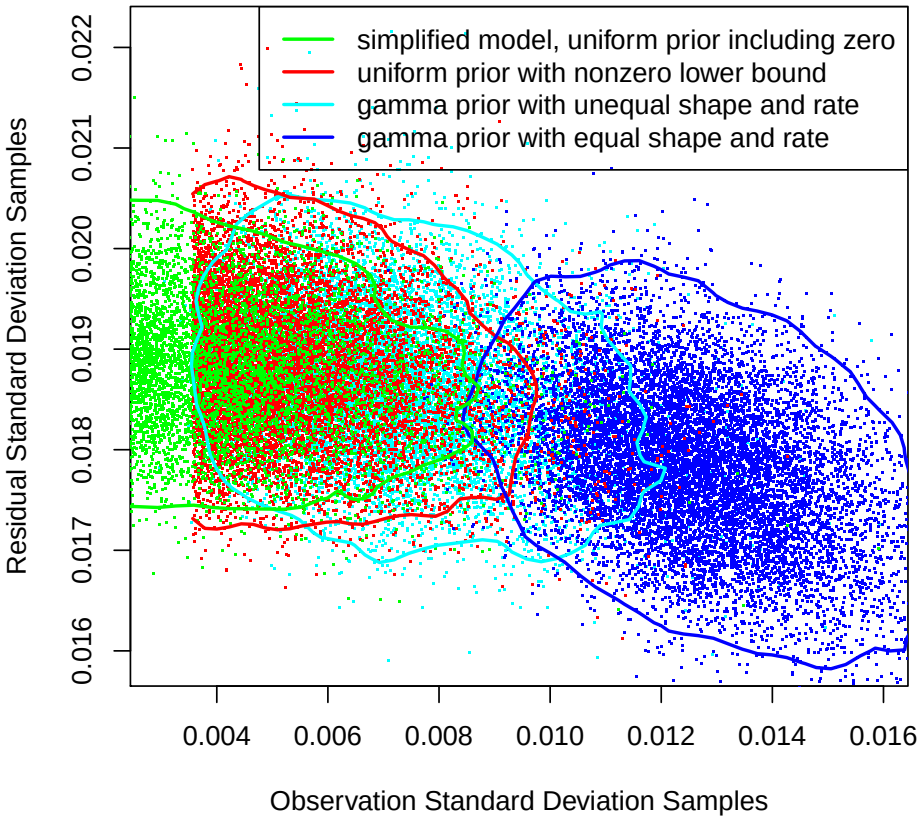

Figure D8: Scatterplots of MCMC samples for residual standard deviation σ_ε and observation standard deviation σ_{DBH} for four different models with different observation error priors. Contour lines are 95% credible contours (Bolker 2012). Cyan is the model with a gamma prior with unequal shape and rate; red is the model with a uniform prior with a nonzero lower bound (clearly visible!); blue is the model with a gamma prior with equal shape and rate; and green is the simplified model (missing annual water deficit, insolation, and elevation) with a uniform prior including zero. Clearly the tradeoff between residual and observation error is affected by the prior chosen for observation error, and how close we allow it to get to zero.

Analysts should watch for these kinds of tradeoffs in their models as well. In our case, this relationship means that when observation error standard deviation is estimated with different priors, residual standard deviation is affected, though not extremely so: the biggest differences in mean estimates for residual standard deviation in the four models is 0.001 on the standardized scale, or about 0.02 cm (Fig. D9), and the significance of the parameter does not change (it is still well separated from zero). In fact, it is a mark of the robustness of the residual standard deviation estimate. Despite this tradeoff relationship with observation error, the residual error estimate does not change substantially when the observation error standard deviation does. None of the other parameters in the model are affected by observation error's prior.

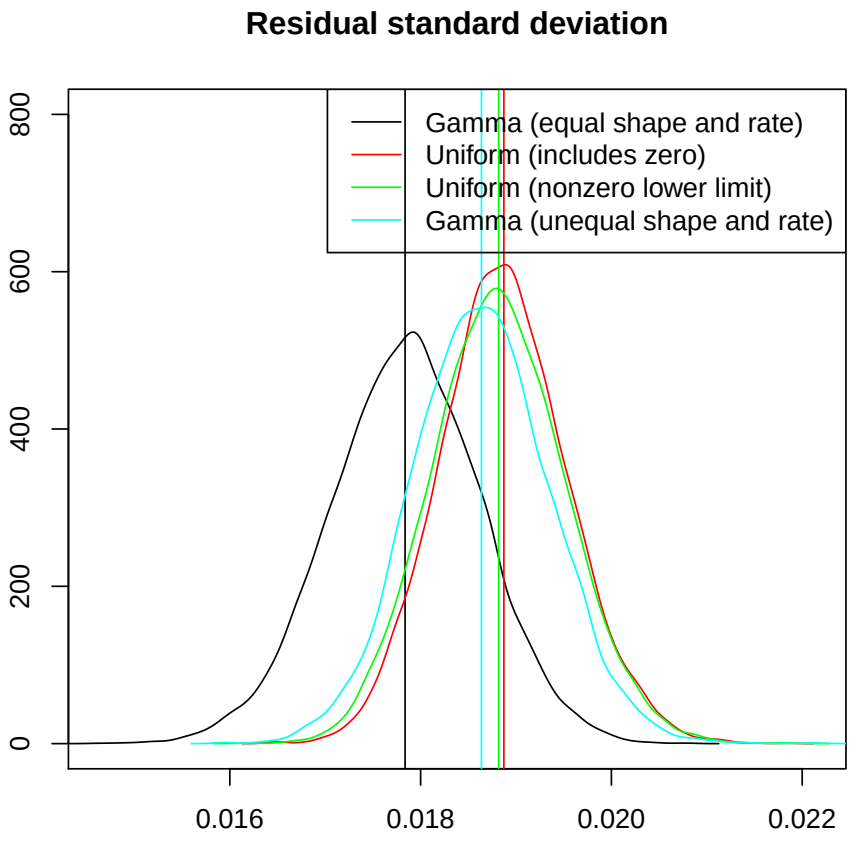

Figure D9: Posterior estimates for residual standard deviation σ_ε for four different observation error posteriors. Cyan is the model with a gamma prior with unequal shape and rate; green is the model with a uniform prior with a nonzero lower bound; black is the model with a gamma prior with equal shape and rate; and red is the simplified model (missing annual water deficit, insolation, and elevation) with a uniform prior including zero. Vertical lines are means for each posterior.

References

Bates, D., and M. Maechler, 2010. lme4: Linear mixed-effects models using S4 classes. URL `http://cran.r-project.org/package=lme4`.

Bolker, B., 2012. emdbook: Ecological Models and Data in R, R package version 1.3.2.

Plummer, M., N. Best, K. Cowles, and K. Vines, 2010. coda: Output analysis and diagnostics for MCMC. URL `http://cran.r-project.org/package=coda`.

Sturtz, S., U. Ligges, and A. Gelman. 2005. R2WinBUGS: A Package for Running WinBUGS from R. Journal of Statistical Software **12**:1–16.

Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC, Boca Raton, FL.