

Daniel Buonaiuto

OEB 201

November 29, 2017

Is Bud Size a Predictor of Phenology in Temperate Woody Plants?

Project Aim

Phenology, the study of organisms' annual life cycle events, has received increased attention in the science of the past decades as phenological changes have appeared as one of the most apparent hallmarks of a changing climate (Menzel *et al.*, 2006). Temperature and photoperiod have been identified as the major drivers of variation in phenology for temperate zone plants (Forrest & Miller-Rushing, 2010), but as our understanding of the implications of changing phenology grows, it is critical to explore other drivers that serve to fine-tune phenological expression in organisms. It is possible that both intra-specific variation in phenology could be linked with other physiological differences between individuals. One promising physiological trait for fine-tuning predictions of the timing of budburst in temperate woody plants is the size of their winter buds. It would be expected that larger buds would burst earlier than their small counterparts, assuming that larger buds are a sign of greater initial investment in leaf tissue. The paper below details an experiment designed to test this hypothesis. In the following section I will discuss the methods of study, with an emphasis on the statistical approach I used. I will then discuss preliminary results of the analysis and make recommendations for a more complete approach.

Methods

The data

In January of 2015, 40 cm cuttings from 27 temperate woody plant species were collected from two sites in Eastern North America (Harvard Forest, Massachusetts, USA and Saint Hippolyte, Quebec, Canada). The collection included 6 replicates per species, with buds per replicate varying, resulting in a data set of 6,467 rows with data at the individual bud level. For each cutting, bud length and volume were measured

with calipers. For many individuals, buds were massed, fixed in a formalin-ethanol solution and dissected with the intention of evaluating other bud characteristics that might affect timing of bud burst such as bud determinism and structure, but these procedures are beyond the scope of this analysis and will not be discussed in detail. Bud width and length were used to calculate bud volume using the equation:

$$V = \pi(r^2)(h/3)$$

where r is one half of the bud width and h is bud length. It should be noted that bud measurements were taken over a 3 month period, and most individuals from Harvard Forest measured before those from Saint Hippolyte. Because it is biologically possible that buds would grow between the time when they were cut from the tree and when they were measured, this measurement pattern produces a potential temporal bias in these data (see figure 1). Much of the statistical analysis presented below will detail the development of the model used to correct for this bias.

Statistical Analysis

To correct for the temporal bias in the data, I developed a model predicting how much bud volume could change per unit time. This model was built from a subset of the full data set which included some repeat measurements from Harvard Forest, was restricted to include only measurements Harvard Forest measurement recorded in a 30 day window, and the few Saint Hippolyte measurements recorded at the same time. Since the data were unbalanced at the species level and it is likely that species would differ considerably both in their absolute bud volume and rate of change, I used a Bayesian multilevel, varying slope, varying intercept at the species level model to predict bud volume and a function of day of year of measurement. All computing was done using R statistical software, and models were specified using the package `rstanarm` (Team, 2016). The model is specified below:

$$y_{[i]} = \alpha_{[j][i]} \beta_{[j]} X_{[i]} + e$$

where α is a species specific intercept and β is a species specific coefficient for the predictor day of measurement, represented here by X . First, to test this model's ability to return true values, I simulated a data set of bud volumes for 20 species with 100 observations each, and simulated measurements over a time span of 30 days. In this data set. I set the true global intercept of bud volume to be 8 and effect of date

of measurement to be 0.1, and let these values vary by species. The model fit was evaluated with posterior predictive checks (see supplement).

Once the model returned the proper values for the simulated data (see next section for more details), I applied the model to the real data set. To fit the modeling frame work, I transformed bud volume distribution to normal using a logarithmic transformation.

Just as with the simulated data, the model's accuracy was evaluated using posterior predictive checks (see supplement).

The model output a species specific coefficient for the effect of day of measurement on bud volume, and these coefficients were used to produce a true bud volume estimate for the full data set, which can be best described as the expected bud volume if all individuals had been measured on the day 40 of the year. The formula used for the bias correct is specified below:

$$\text{True}_{budvolume} = measured_{budvolume} - \beta_{[i]} * (measured_{day} - 40)$$

where β is the species specific model coefficient for day of measurement, representing the expected change in volume per day.

Modeling phenology as a function of bud volume

As stated in the introduction, the ultimate goal of this project was to evaluate the role that resting bud size plays in fin-tuning phenological event dates. The following section details the beginning states of developing such a model. To this aim, I obtained phenological event information from another experiment that used cuttings from the same parent plants that were used to create the bud measurement data. In this experiment, cuttings were grown in growth chambers under varying winter chilling temperature, photoperiod and spring warming temperature regimes, and phenological responses (days to budburst) were observed, resulting in 2136 rows of data recorded at the twig level.

As a first pass, for each individual parent, I calculated the mean true bud volume (data set I) and mean phenological response across all experimental treatments (data set II). I built a preliminary, multi level, varying slope and intercept at the species level model to evaluate the possible influence of physiological predictors such as bud volume and twig diameter on phenological rank. The model is specified below:

$$y[i] = \alpha_{[j][i]} + \beta_1[j]X_{[i]} + \beta_2[j]Z_{[i]} + e$$

where predictors X and Z are mean bud volume and stem diameter respectively, α is the species specific intercept and the β 's are the species specific coefficients for bud volume and stem diameter respectively. Finally, it should be noted that I did not heavily interpret this model and do not feel the results are robust. Rather, these model results should be view as a step in the process of creating a more robust model to address this question in the future.

Results and Discussion

Bias Correction

The bias correction model, when run on simulated data returned the true values of the parameters (see figure 2, 3). The model, when applied to the real data, predicts a non-detectable global effect of date of measurement on bud volume. There is high predictive uncertainty associated with this model (sigma= 4.1 see figure 4) and small effect size estimates which likely contribute to this result. Because I expect there to be substantial differences between species, the global intercept and coefficient estimates are not of high values in this analysis. When viewing the species specific coefficients for day of measurement (see figure 5), the effect sizes are all very small, however, it is important to note that these estimates represent change in bud volume per day, so for example, the .017 effect size of *Acer rubrum* would mean that a bud measured 90 days after abother, may have grown 1.53 cubic millimeters in that time, which would introduce significant bias if measurement day was not account for.

Most species' 95 percent credible intervals were overlapping zero. For such species, there is high uncertainty that and change in bud volume takes place overtime, and the model indicates that the temporal bias is not relevant. Some species, such as *Quercus alba*, *Acer saccharum*, *Populus grandidentata* and *Spirea alba* have negative estimated effect sizes and credible intervals that do not overlap zero. This result fails to support my original hypothesis. While it is possible that bud volume may decrease with time if, for example buds become more desiccated over time, or lose protective structures over the course of the season, the effect of shrinking buds with time runs counter to my biological hypothesis and the observation that buds swell as they approach budburst. Several other species, *Acer pensylvanicum*, *Acer rubrum*, *Fagus grandifolia* and *Lonicera canadensis* show the biologically expected trend, with positive coefficient estimates and credible

intervals that do not overlap zero. Because the sources of bias was based on the expectation that buds grow over time, it may be tempting to only perform bias correction on the 4 species that fit the expected pattern, but with high predictive uncertainty and no mechanisms for conclusively saying that the positive coefficients are the result of a biological mechanism and the negative coefficients are a statistical artifact, I determined to use all coefficient values, and project adjusted bud volumes for each individual (see figure 6).

Phenological predictions

As mentioned above in the methods section, no serious conclusion should be drawn from this model, but it should rather be considered a building block for future analysis. Global coefficients estimates for the effect sizes of both stem diameter and bud volume are negative, which is in line with my hypothesis that larger buds (and thicker stems) are more primed for budburst and can burst earlier. Larger buds may indicated more advanced development, and thicker stems larger reservoirs of stored nutrients that can be utilized by developing leaves. However, for both parameters, the 95 credible intervals overlap zero, so it can be determined there is considerable uncertainty in these predictions. As seen in the posterior checks (see supplement), the model tends to miss a dip and spike in the data on the higher side of the distribution and overestimate phenological response in this region.

I will now wrap up this section by proposing strategies that could be used to improve this model moving forward. This model does not acknowledge many factors that may be important biologically in influencing budburst timing. Adding predictors such as site, tree dbh (as a proxy for age), location of bud on the twig, or other physiological predictors might serve to absorb some of the uncertainty. A sound analysis of these data also calls for a more sophisticated hierarchical model, which would allow for the predictors (for example bud volume) to be evaluated on the individual bud level and the response (days to budburst) at the twig level rather than averaging as I did in this preliminary analysis. This would allow for the treatment effect to be maintained in the model, which would be expected to be a major source of variation.

In conclusion, in this project I have developed a fairly reliable estimation of true bud volume from a temporally biased data set. An improvement to this method would be to take more repeat measures over a similar time window, so changes in size could be more accurately observed rather than inferred. I also began to develop a model to determine if bud volume and stem diameter is a predictor for phenological rank within

species, and laid out possible next steps for building a more robust model.

References

- Forrest, J. & Miller-Rushing, A.J. (2010) Toward a synthetic understanding of the role of phenology in ecology and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 3101–3112.
- Menzel, A., Sparks, T.H., Estrella, N., Koch, E., AASA, A., AHAS, R., ALM-KÜBLER, K., BISSOLLI, P., BRASLAVSKÁ, O., BRIEDE, A. & et al. (2006) European phenological response to climate change matches the warming pattern. *Global Change Biology* **12**, 1969–1976.
- Team, S.D. (2016) rstanarm: Bayesian applied regression modeling via Stan.

Figures

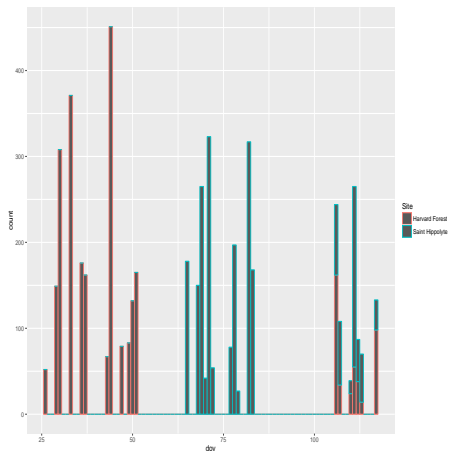


Figure 1: Measurement days of buds colored by site. Twigs from the different site were measured non-randomly, producing a temporal bias

	Actual	Median	MAD SD	CI ₅	CI ₉₅
Intercept	8	8.3	1.4	5.99	10.67
doy	0.1	0.1	0.0	0.0337	0.159
sigma	4	4.1	0.1		

Figure 2: real and model predicted effects for simulated data with standard deviation and 95 percent credible intervals

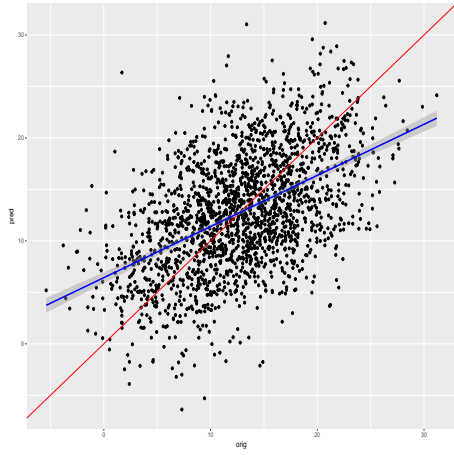


Figure 3: Predicted vs. original values for fake data. The red line represents the 1:1 line which would represent perfect prediction. The blue line is the actual slope, demonstration some deviation

	Median	MAD _{SD}	CI ₅	CI ₉₅
Intercept	0.3	0.3	-1.8e-01	0.82
doy	0.0	0.0	-1.03e-03	0.008
sigma	4.1	0.0	1.4042e+00	1.45046

Figure 4: Global intercept and coefficient estimates for bias correction model

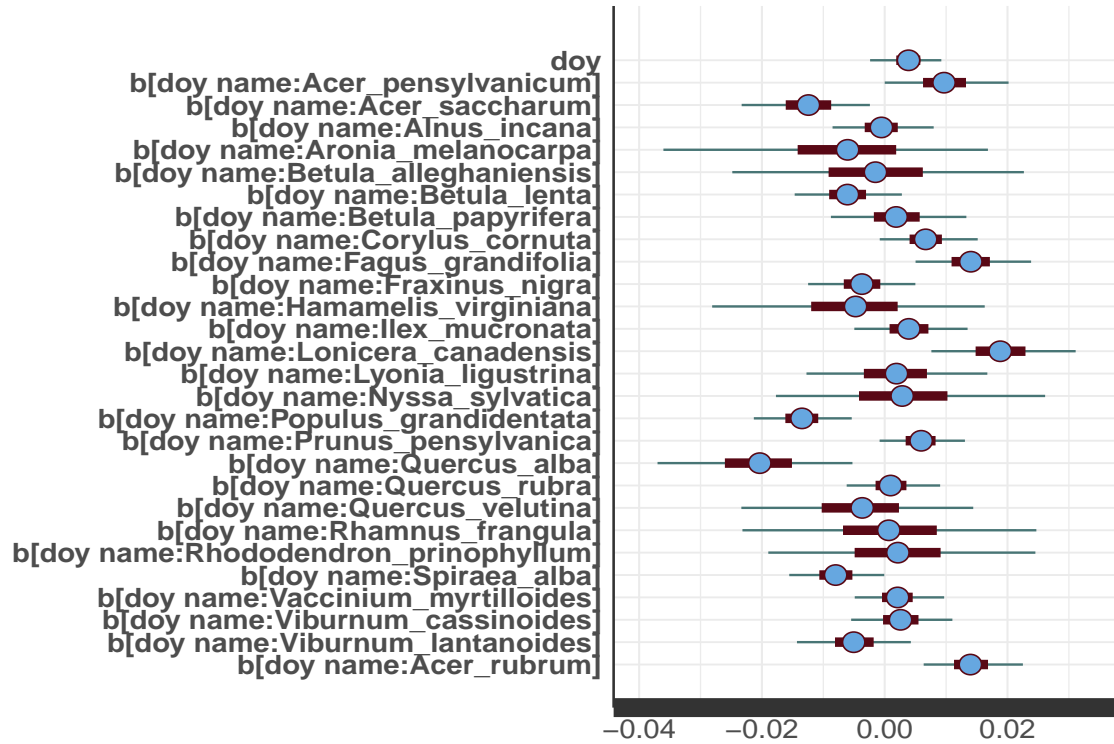


Figure 5: The mean effect sizes estimates and 50 and 95 credible intervals of day of measurement on each of the 27 study species from the model $bvol \sim doym + (doym | sp)$

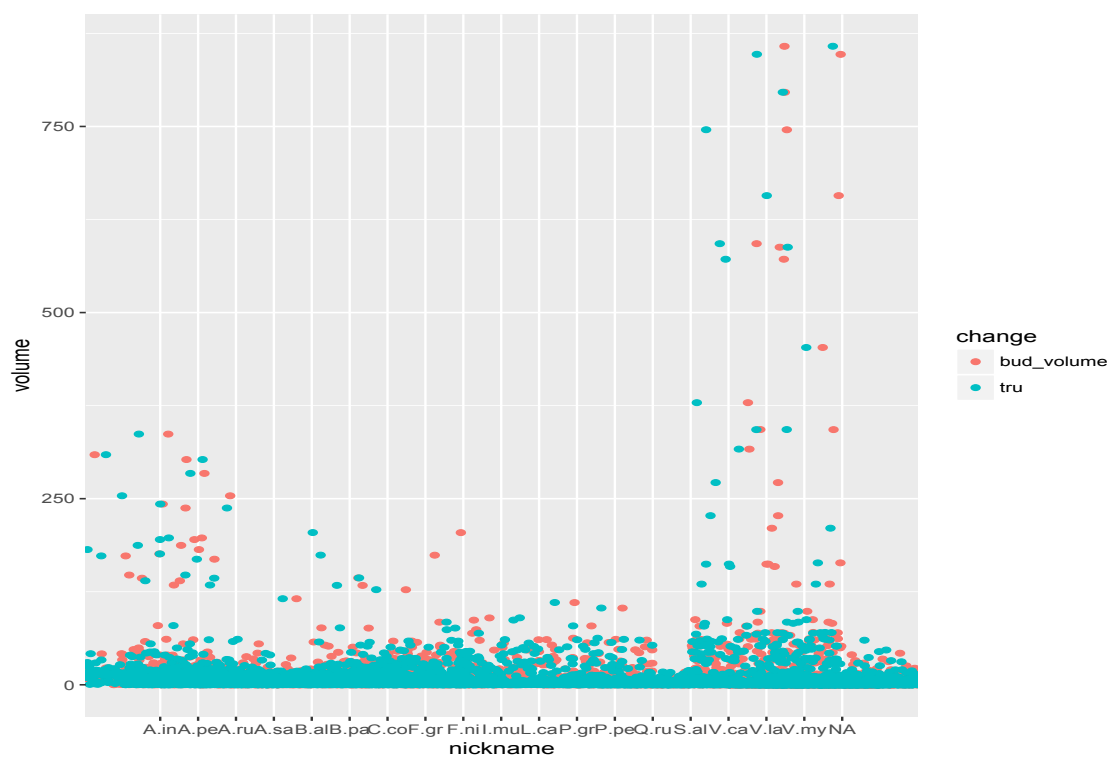


Figure 6: Temporal bias correction. Blue dots demonstrate the corrected values against the original (red). The plot has been horizontally jittered for better viewing

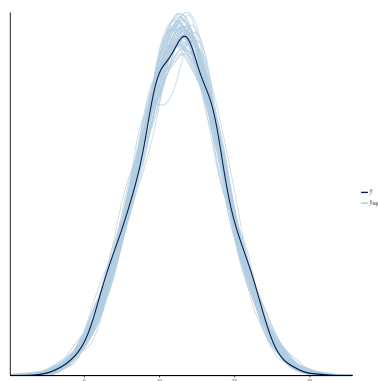


Figure 7: SUPPLIMENT: Posterior predictive checks for model $bvol \sim \text{doy} + (\text{doy} \mid \text{sp})$ on fake data

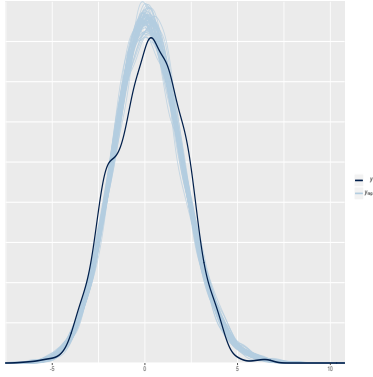


Figure 8: SUPPLIMENT: Posterior predictive checks for model $\text{bvol} \sim \text{doy} + (\text{doy} \mid \text{sp})$ on real data

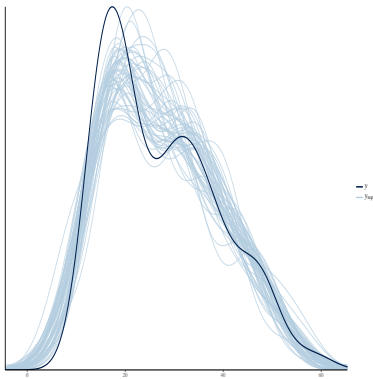


Figure 9: SUPPLIMENT: Posterior predictive check for exploratory model $\text{meanbday} \sim \text{meanstemdiam} + \text{meanbvol} + (\text{meanstemdiam} + \text{meanbvol} \mid \text{SP})$