

Project Aim

The aim of this projects is to build a regression model that allows for the exploration of the relationship between the size of winter buds and the time of key phenological phases such as budburst and leafout in temperate woody plants. As will be discussed below, such analysis may be possible given the data limitations. If this is the case, I will utilize these data to understand other important questions about winter bud character and dynamics, such as range of variability of bud size among sites, species, individuals and positions on twigs.

Data Structure

My primary dataset contains 6467 rows of bud width and length measurement, recorded at the individual bud level for 27 tree species at two field sites (Saint Hippolyte and Harvard Forest) in the Eastern North America. There are 6 replicates of individual per species and the number of buds per replicate are in the range of 3-25. This data set also includes important secondary descriptions such as the date of collect from the field and date that measurements occurred. From the bud width and length, I have calculated bud volume using the formula for the volume of a cone ($\pi \cdot r^2 \cdot (h/3)$).

The second dataset comes from a different experiment in which twig cuttings from many of the same individuals were collected and grown in growth chambers, under 12 different varying temperature and day length treatments. The dataset records the day on which the phenophases of budburst and leaf out were observed. The dataset contains 2136 rows of data recorded at the twig level.

There are two major challenges to working with these data:

- 1) Is it appropriate to combine these two data sets? The fact that the bud size was not measured on the same individual twigs on which the phenological observations were made calls into question the validity of my proposed model. Furthermore, the effects of the experimental conditions on the phenological response would have to be accounted for in the model, and it is likely the large effect sizes of the experiment treatments would swamp my ability to observe any effects in my predictor of interest. Finally, the fact that these data are recorded at different levels of observation would make inference challenging and approximate.
- 2) The second, is that winter bud size is not static, and buds will grow overtime, so the date of measurement is a confounding variable. Unfortunately in these data, buds were not measured randomly, but each site measured sequentially (see figure 1), yielding a temporal bias. Repeat measurements made on some of the Harvard forest individuals allow for a general inference about the degree of change over time. These effects do not seem to be consistent between species, but could be used as a baseline to perform bias correction for the data. Another approach would be to subset the data and only consider records for comparable time window.

Model and other considerations

If these challenges can be overcome my basic proposed model is:

$\text{budburstday} \sim \beta \text{budvolume} + \epsilon$

A possible framework for account for the treatment effects, and species and site differences would be to include the model in a hierarchical framework:

$\text{budburstday} \sim \text{budvolume} + (1|\text{Treatment}) + (1|\text{species}) + (1|\text{site})$

As mentioned above, if such models are invalid due to the lack of a relationship between the predictor and response variables other models could be built. For example, I could use the repeat measures to evaluate the change in budsize as a function of time. Such a model could be written:

$\text{budvole} \sim \text{time} + (1|\text{species})$

par