Magnolia phenology model report

Temporal Ecology and UBC Botanical Garden

By Justin Ngo and E. M. Wolkovich

This summary of the Stan model we constructed for the Magnolia phenology data will go through our observations of:

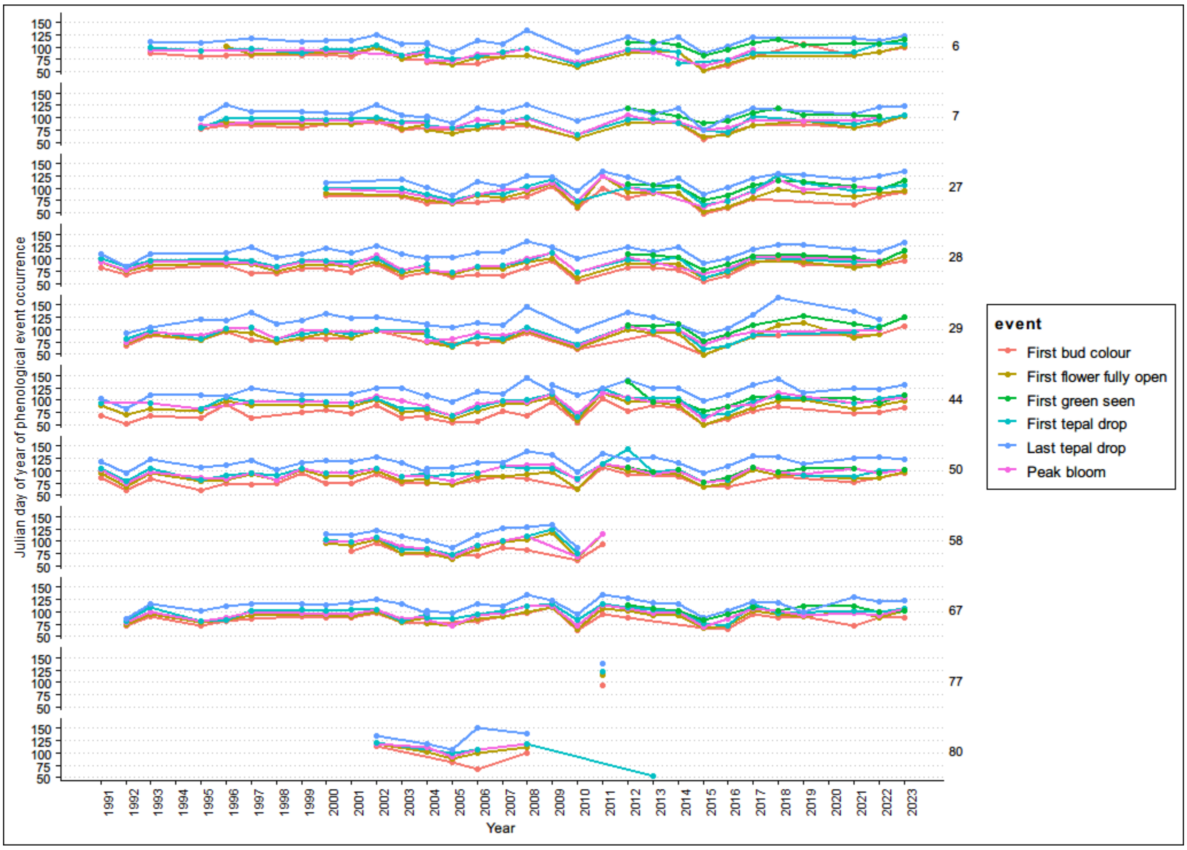
1. Foreword
2. Timing of phenological events in *Magnolia* blooming across the current duration of the project
3. Temperature trends in Vancouver
4. Raw data and other metrics
5. Nested model outputs
6. Future steps
7. Supplementary data
8. References

1] Foreword

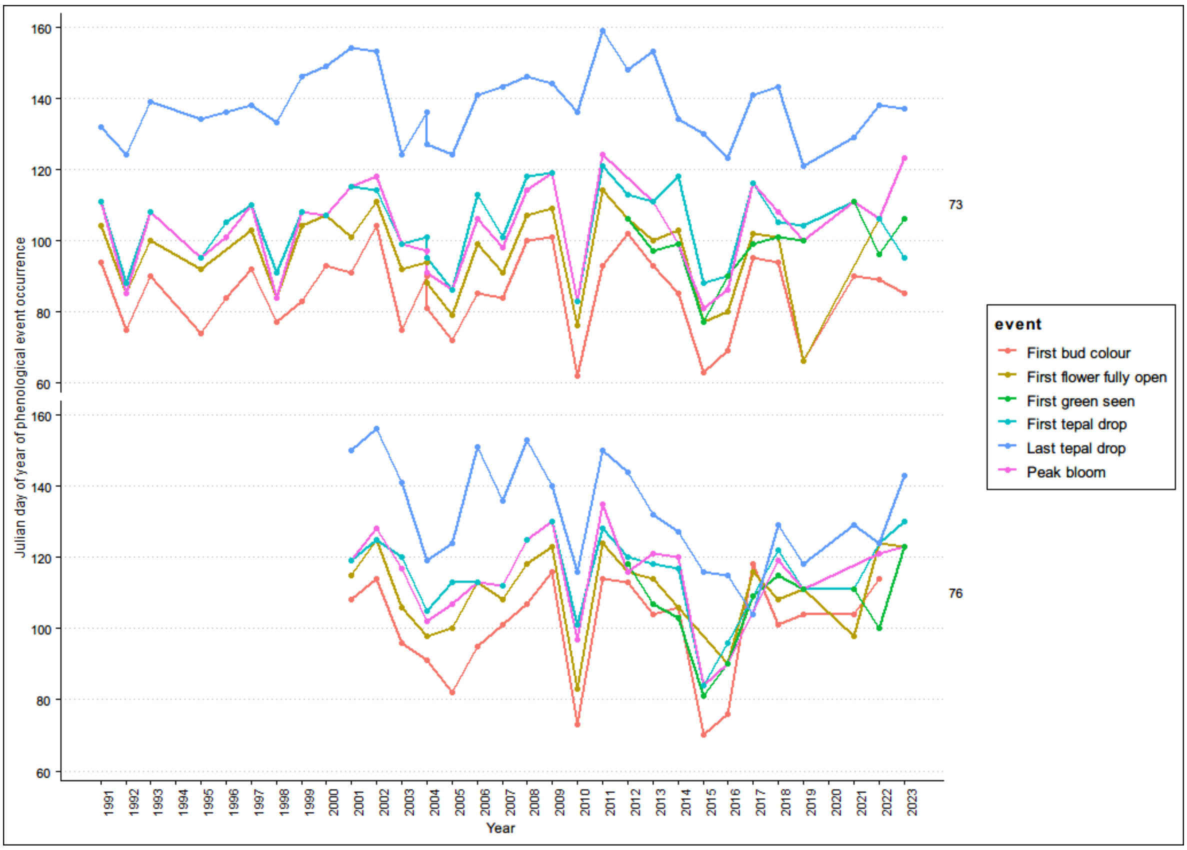
Firstly, we would like to emphasize the quality of your dataset with which we have generated this model. That the model was able to run fairly smoothly already indicates that the data is in good shape. With so many cultivars involved in this project spanning multiple different species all within a single genus, there is probably some genetic influence that can be teased out if there has been any comparative molecular work on the species represented in this dataset. Furthermore the extent and detail of the phenological sequences recorded is also of great use, as it is incredibly rare to find studies observing temporal data at such a fine-scale. Within B.C. there is hardly any data like this, especially compared to the relatively more popular research done on provenance trials (like common garden experiments) or on communities within the Arctic, where the severity and frequency of warming events has been shown to play a major role in modulating the phenologies of high-latitude species. Even if this kind of drastic change to climate is not currently seen in Vancouver (as shown in section 3), it is clear through climate projections of the Pacific Northwest that extreme warming will indeed be observed in the future; as such, having this extensive data and model will serve as a very useful preliminary testing ground and practice tool that can be applied to many other species that are predicted to respond sensitively to climate change.

2] Timing of phenological events in *Magnolia*

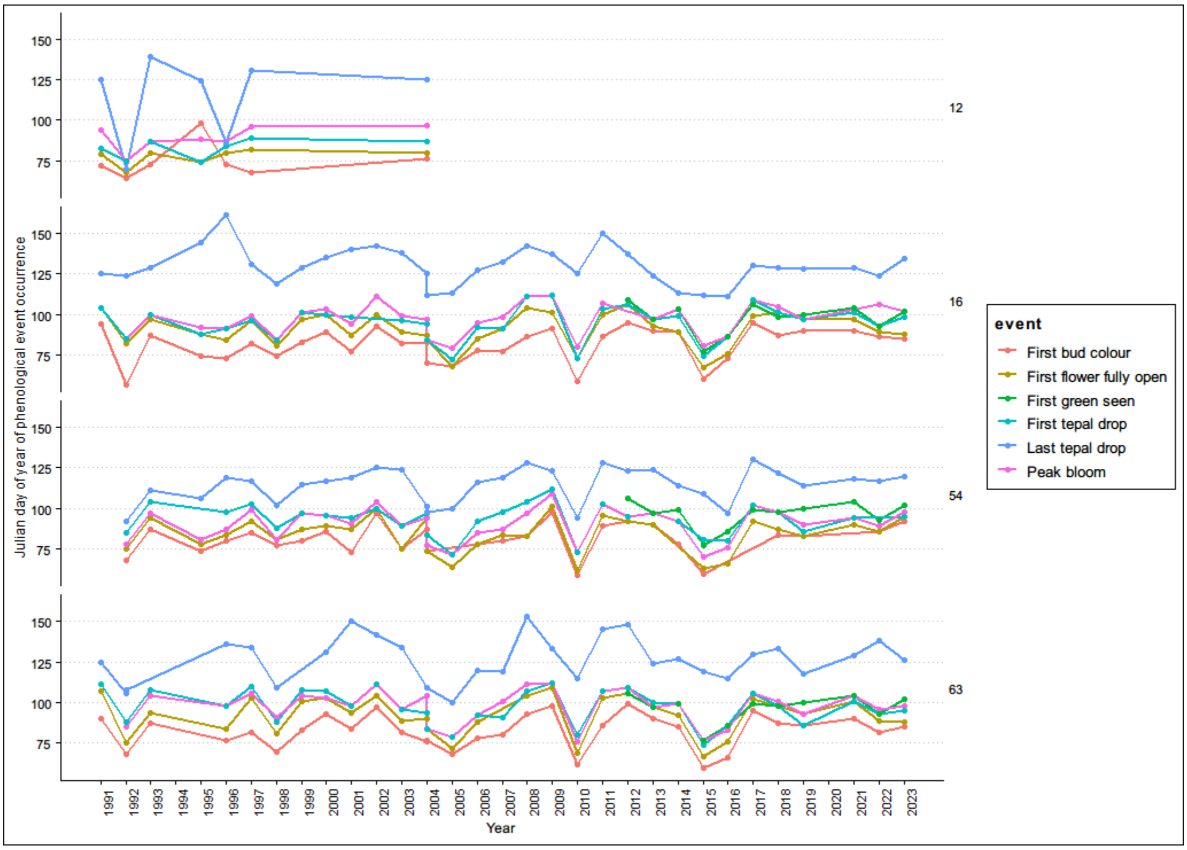
These plots below are visualizations of the day of year in which a given phenological event occurs for each of the species/cultivars. These are first bud colour, first flower fully open, peak bloom, first tepal dropped, and last tepal dropped, which will from now on be called “bud”, “anthesis”, “peak”, “ftepal”, and “ltepal”, respectively. “First green seen” has been dropped from the model due to its more recent introduction to the dataset, though it has been shown in these plots below purely for visualization of the full dataset. Not pictured here are *Magnolia amoena*, *M. conifera*, *M. sapaensis*, *M. biondii*, *M. maudiae*, *M. cavaleriei*, *M. chevalieri*, and *M. laevifolia*.



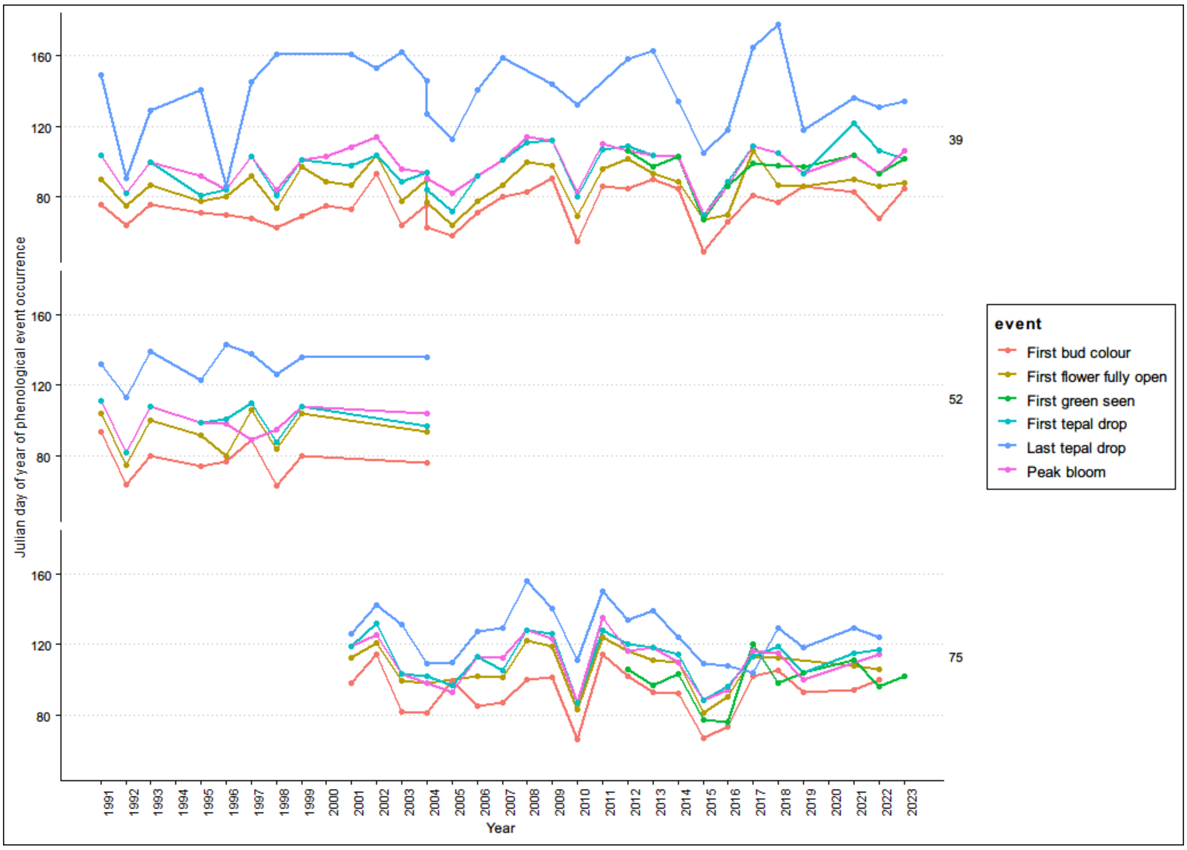
M. campbellii



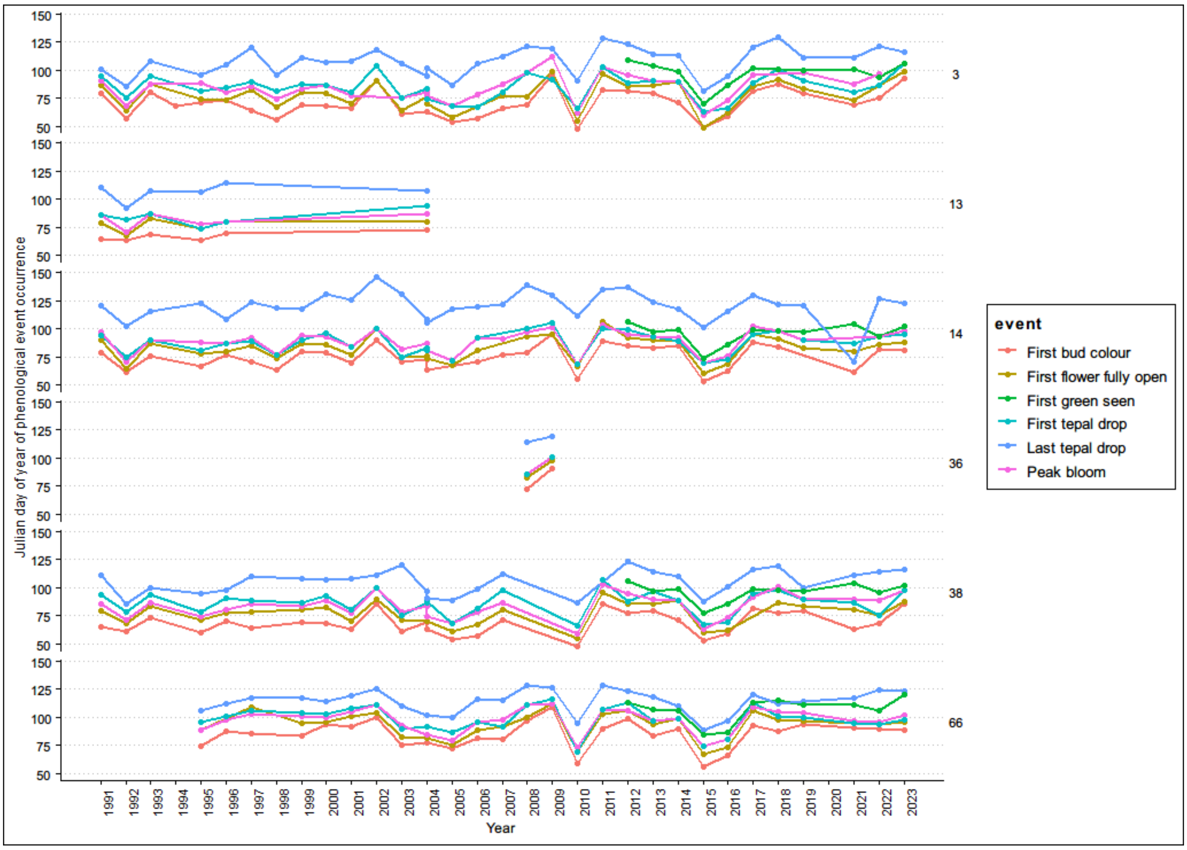
M. cylindrica



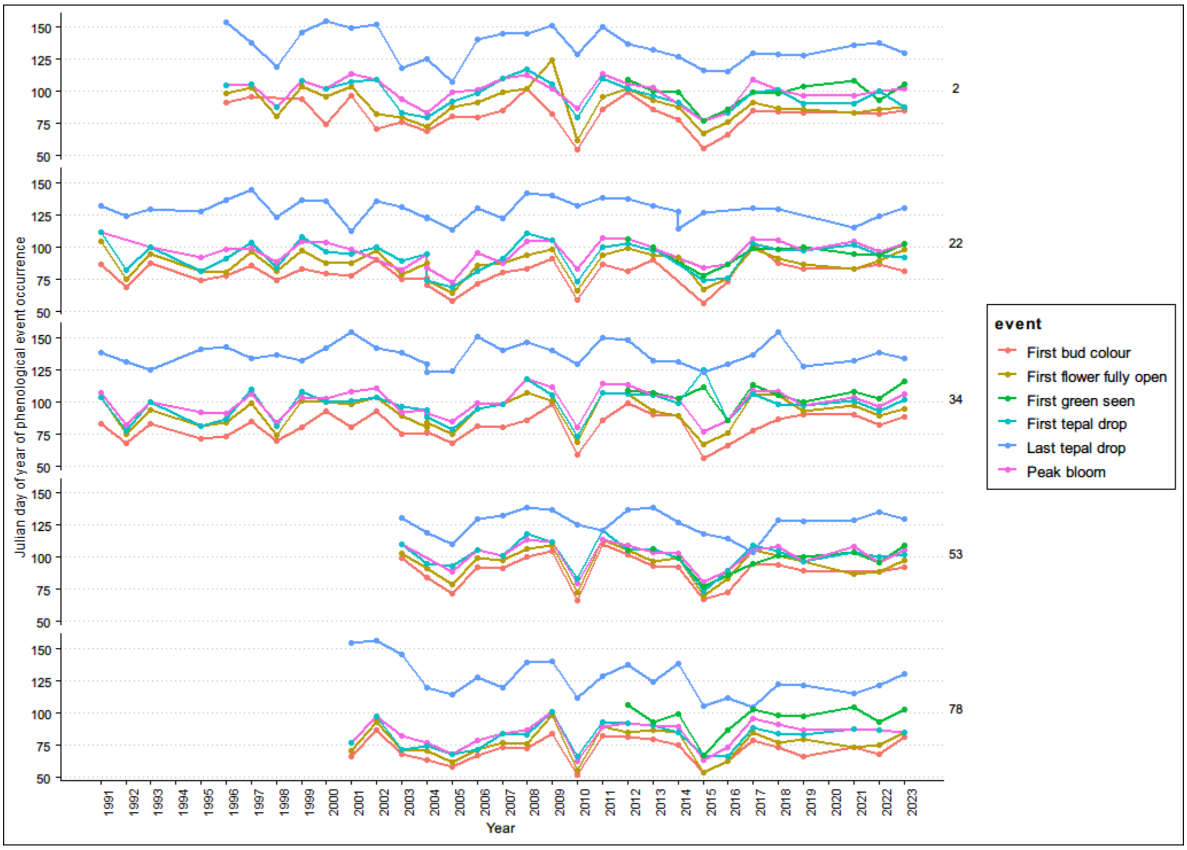
M. dawsoniana



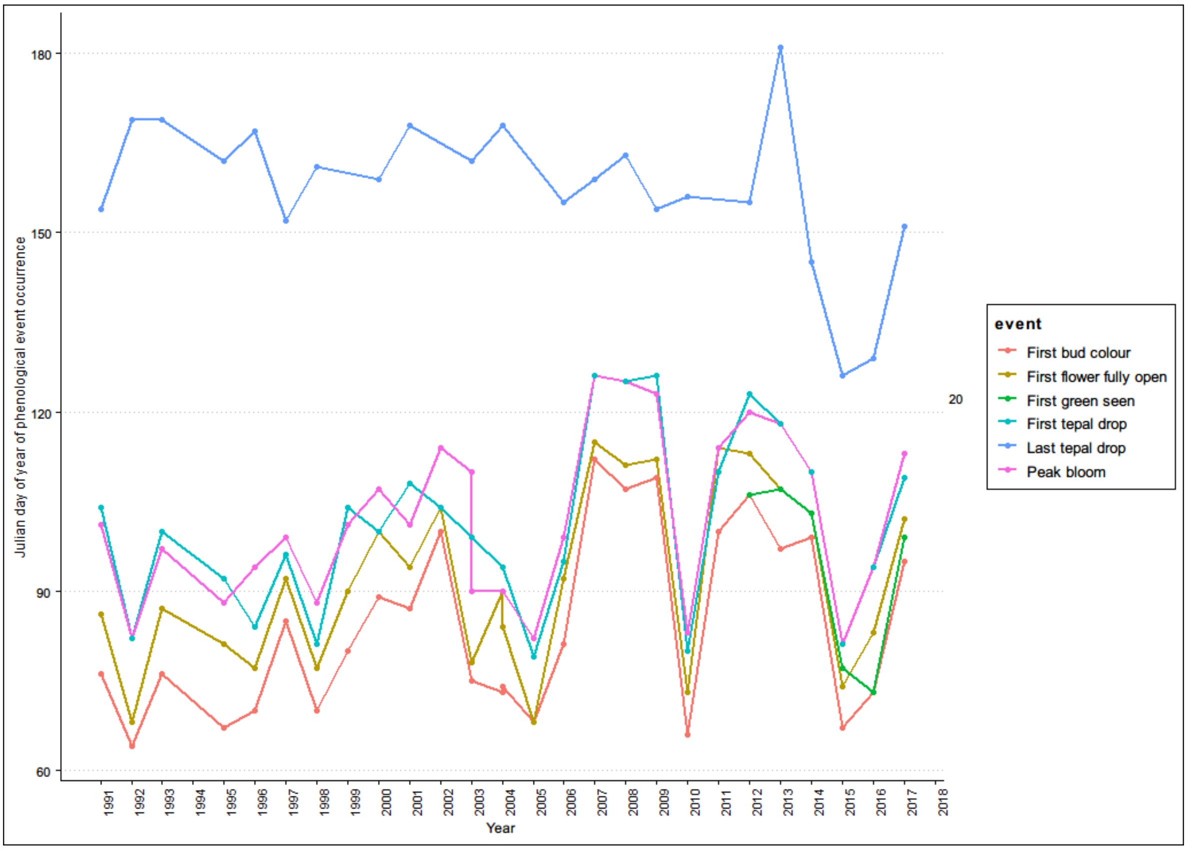
M. denudata



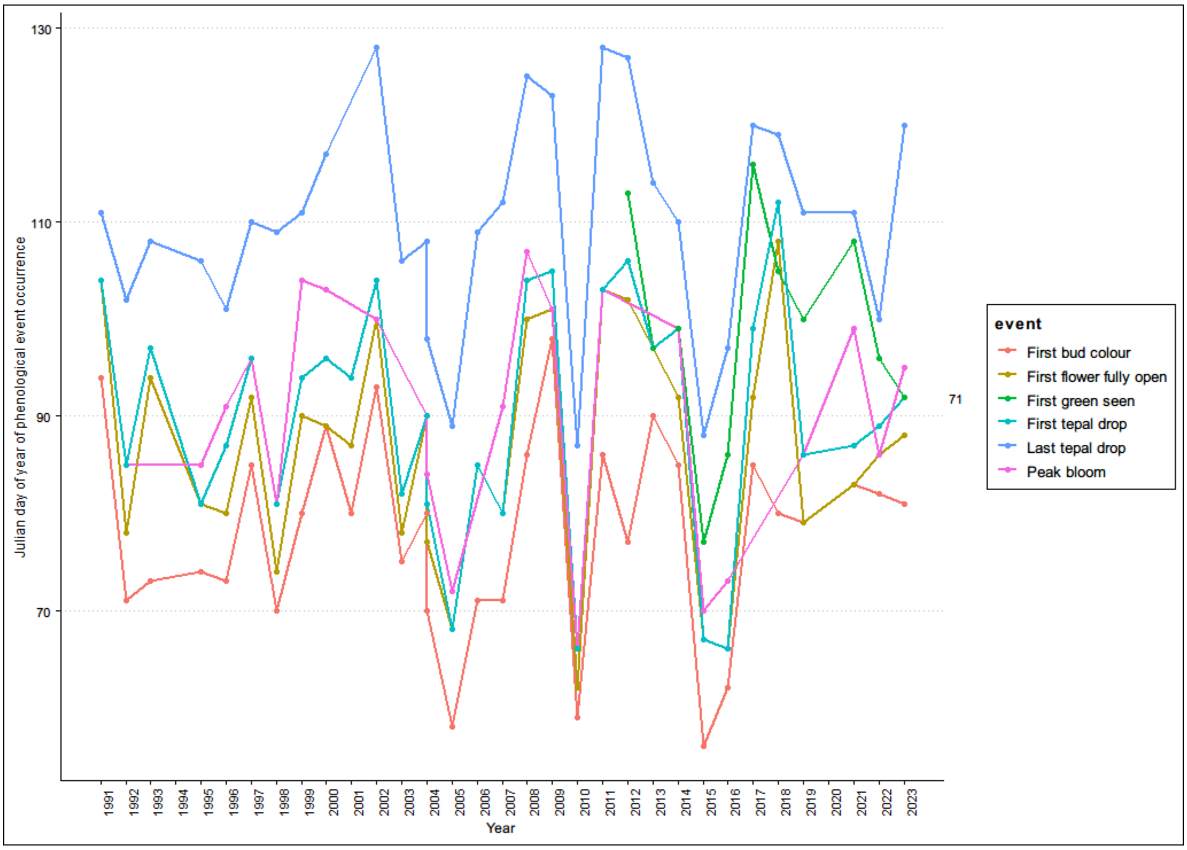
M. sargentiana



M. sprengeri



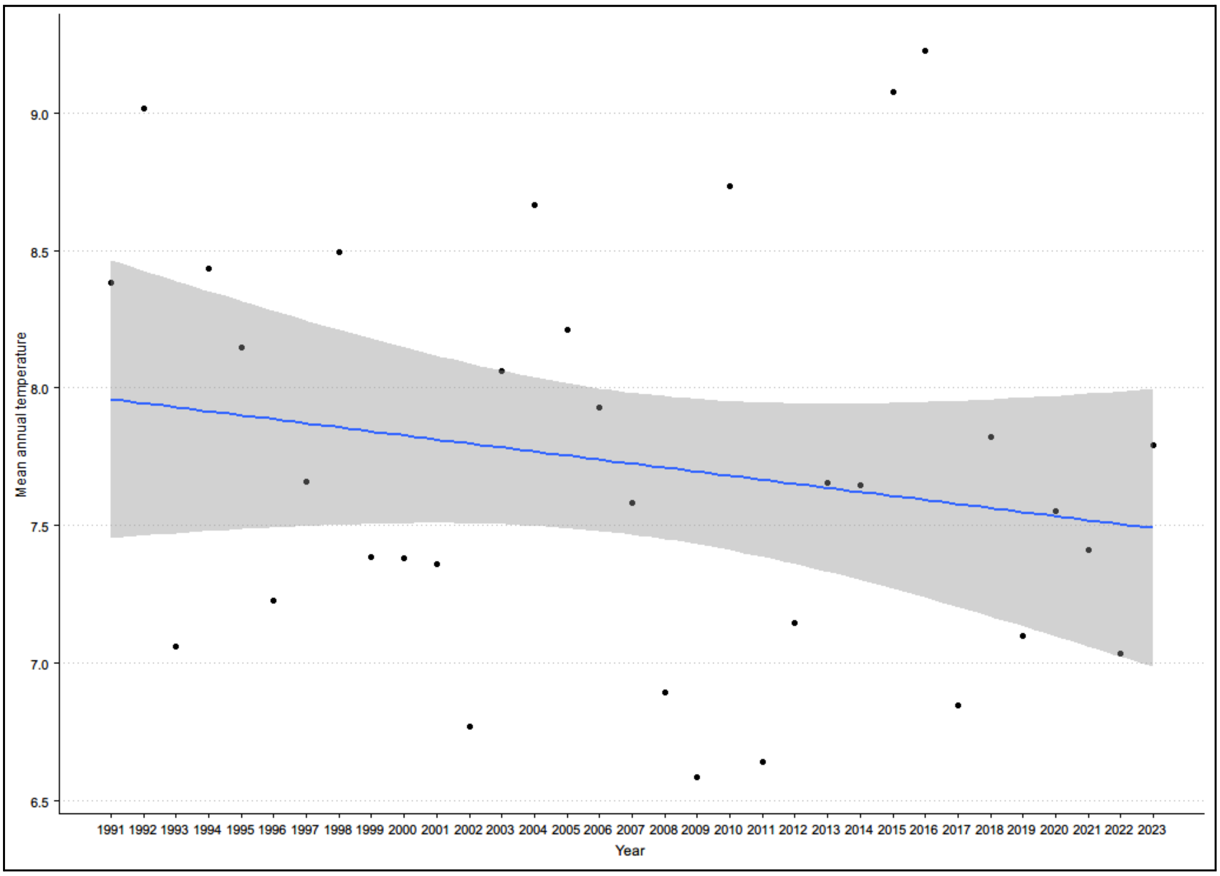
M. stellata

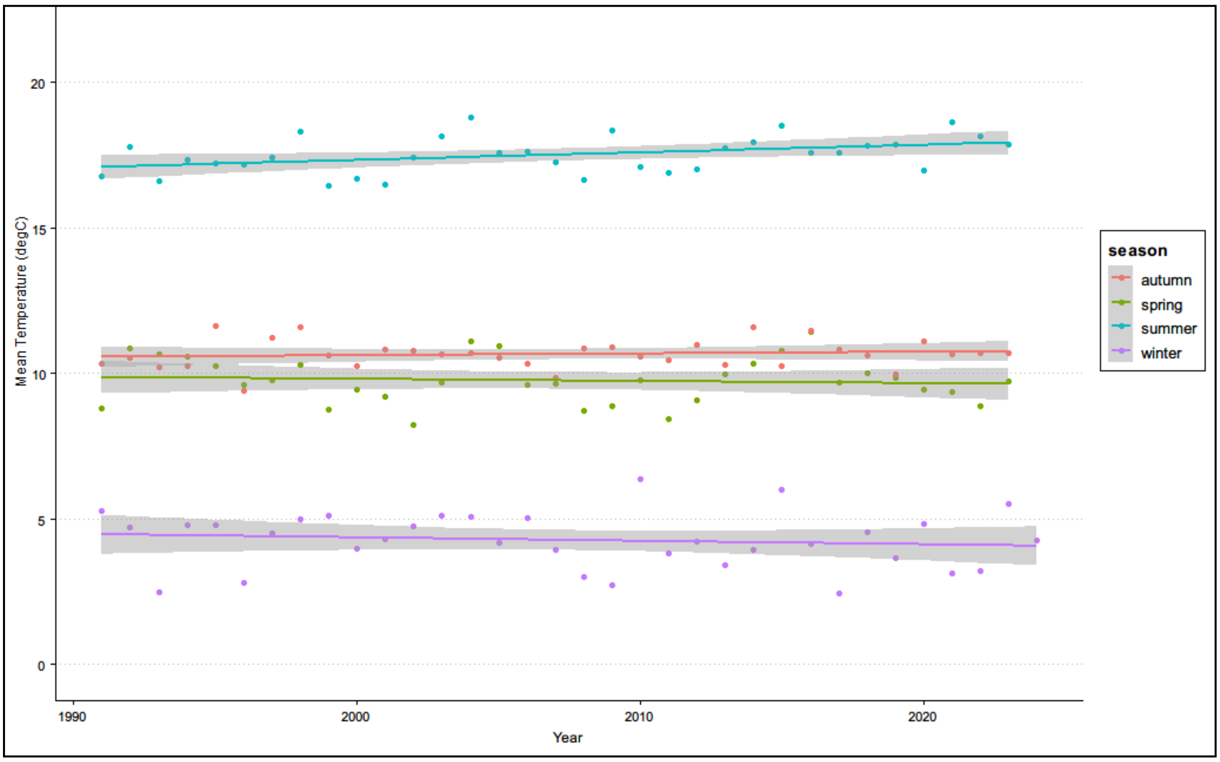


M. zenii

3] Temperature trends in Vancouver

First we wanted to assess if temperature was indeed changing significantly over time in Vancouver. We used daily data from Environment Canada for the year range 1991 to 2023 and performed a simple linear regression both with annual mean averages and also after grouping all dates into 4 seasons, where spring was the consolidated data for March, April, and May, and so on for the other 3 seasons. We set our significance threshold at a value of $\alpha = 0.05$ with the null hypothesis that there is no change in temperature across the years and an alternative hypothesis that temperature change instead indeed occurred.

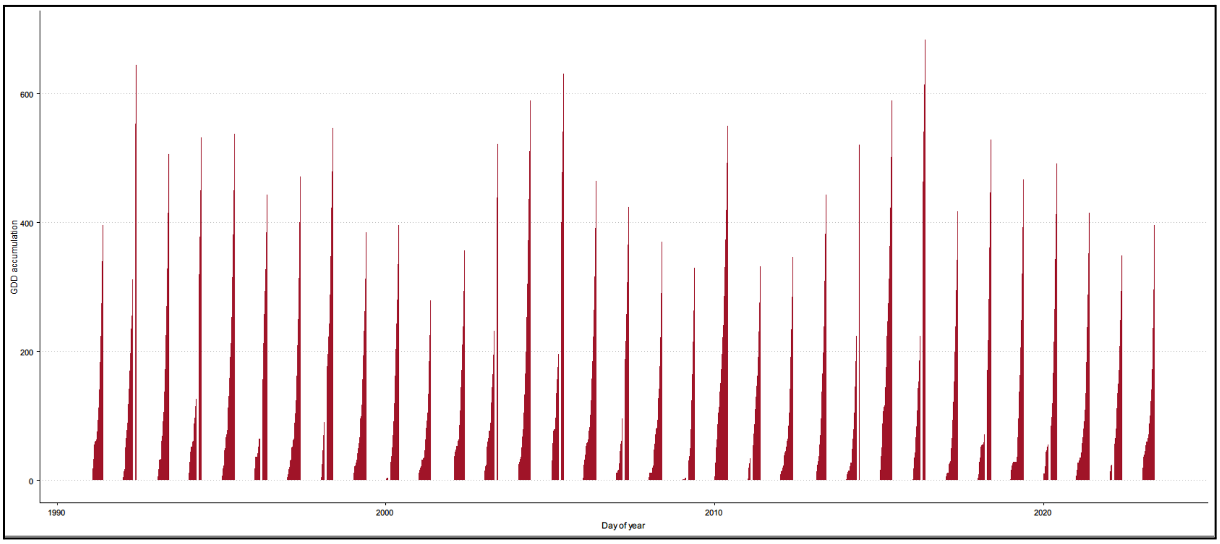




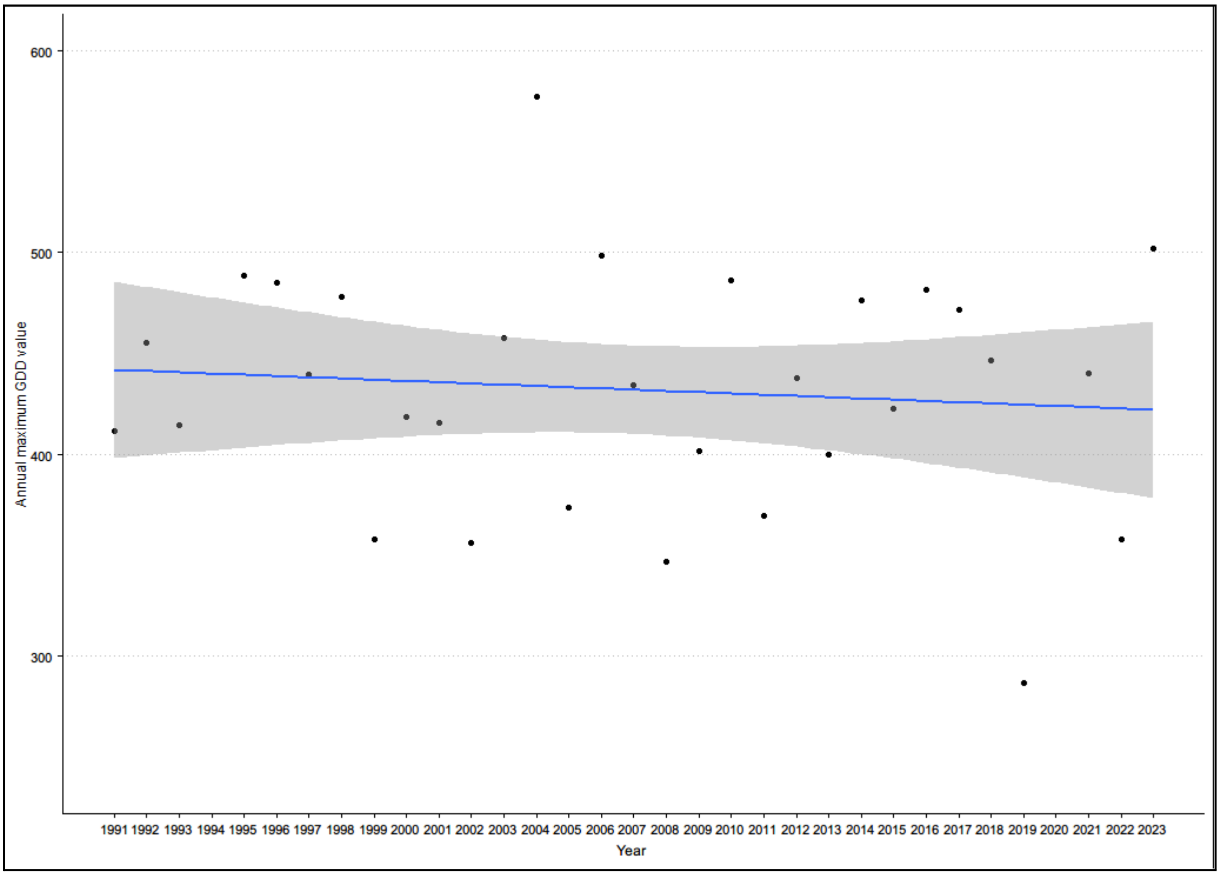
No significant change in temperature over time was detected in the mean annual temperature linear regression model. Furthermore, of the four seasons, only summer displayed a significant change, with an increase of approximately 0.261 degrees Celsius per decade (F-statistic = 1.218, df = 31, p = 2.782). Likewise, the only season in which a significant change in temperature over the years occurred was summer, where the maximum temperature appears to be increasing by 0.3600 degrees Celsius per decade (F-statistic = 5.713, df = 31, p = 0.0231). For other linear regression outputs, please see section 7: Supplementary data.

These findings are very similar to other temperature trends we have observed in another project done in the lab (here, done on wine grapes), revealing that 3 of 4 seasons do not display a significant trend, nor is there an annual trend in general. However, there is indeed a slight trend of warming in summer. There is certainly much to be discussed about this and climate change is clearly a real phenomenon, but for the purposes of this project, we will assume that any trends in *Magnolia* blooming are probably not due to annual or seasonal temperature patterns for the time being. Additionally, the IPCC highlights that the presence and magnitude of these temperature trends is dependent on the focal region in the world, suggesting that there is more heterogeneity in warming effects that one may expect (IPCC [Core Writing Team, Lee, H. & Romero, J (eds.)], 2023).

Because of this, we also wanted to look into total accumulated growing degree days (GDDs) each year. Plotted below is the cumulative GDD value per day per year;



The maximum GDD value can be visualized as such (the exclusion of a data point for 1994 is due to the presence of only one GDD value available, as 1994 only had one phenological event recorded for one cultivar);



We decided to go forward with GDD as it is based on our current understanding of plant development in response to threshold temperatures above which growth can occur, and has been a rather good predictor of phenology thus far. These linear regressions might not show something dramatic now, but they certainly give us insight as to what we might expect to see for the overall temperature trends as they pertain to phenology, especially in the future where we might face a warmer climate than now. There is an ongoing discussion regarding our knowledge of chilling and warming and their effects on bud development, and it is this apparent requirement for certain temperatures to be met upon which we decided that GDDs made the most sense. We set our baseline temperature at 5 degrees Celsius which seems to be a good starting point, though any further clarification on how magnolias respond to certain temperatures would be valuable to the model we are about to outline.

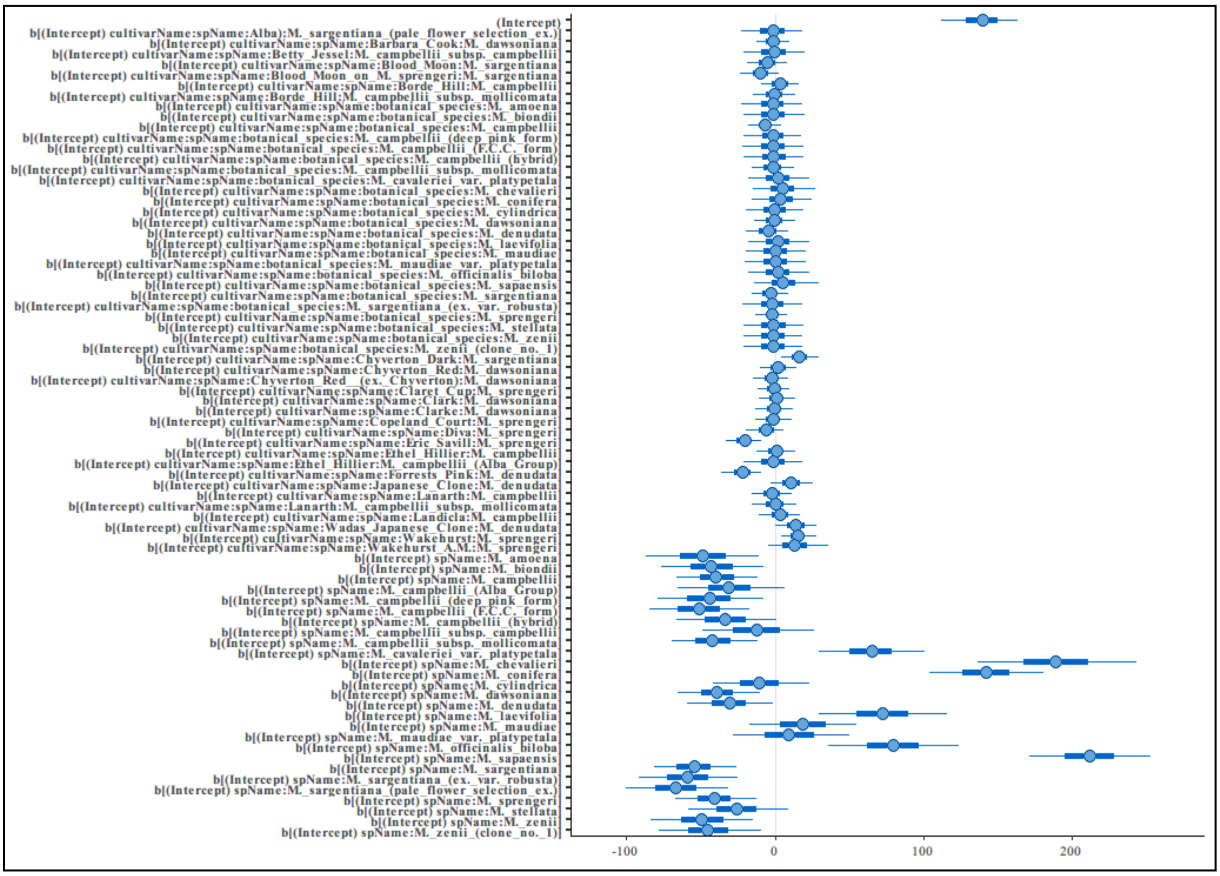
4] Raw data and other metrics

[This section links to the tabulated the parameter posterior estimate values](#) generated from the model, divided into 5 tables (according to the floral phenological events). There are 5 sheets in total, corresponding to each of the floral phenological events.

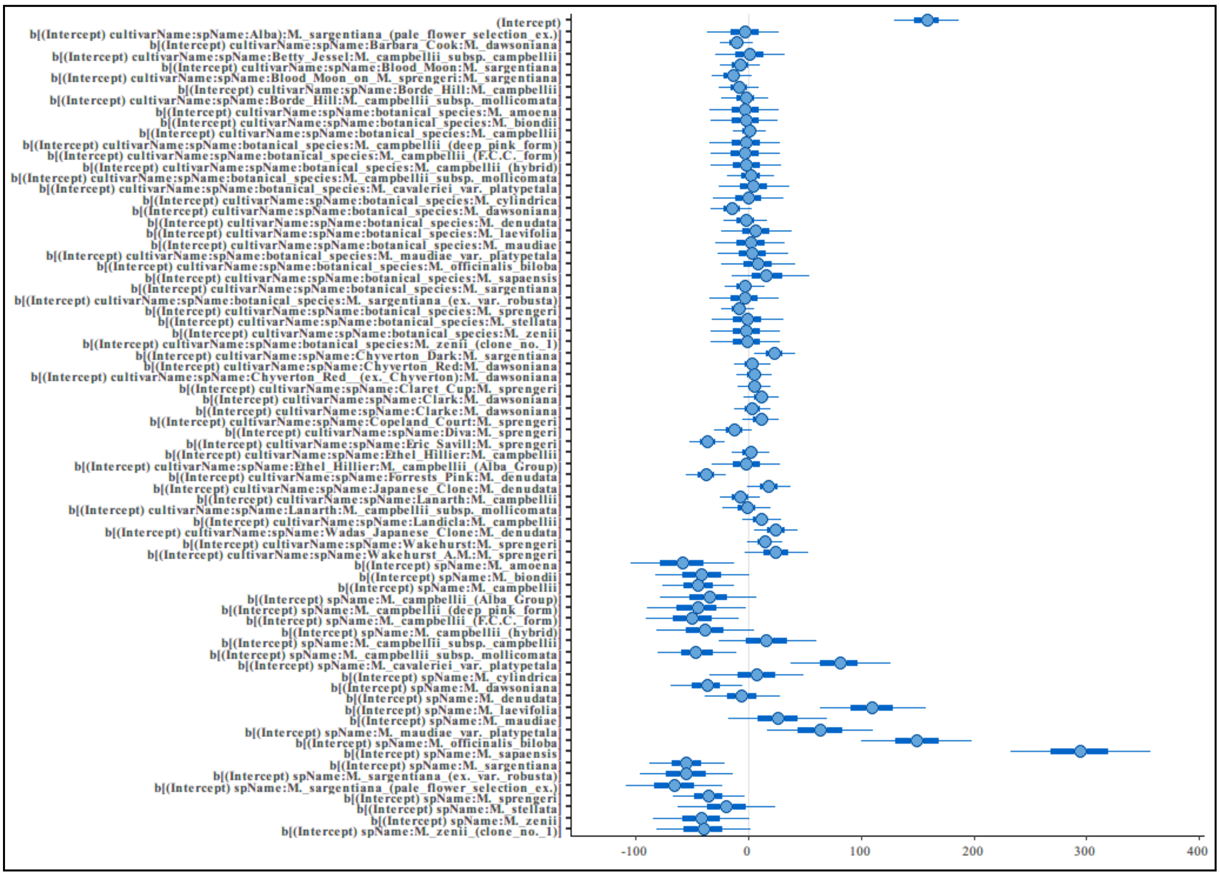
Any data used to generate temperature trend linear regressions is included within the github repository for this project.

5] Nested model outputs

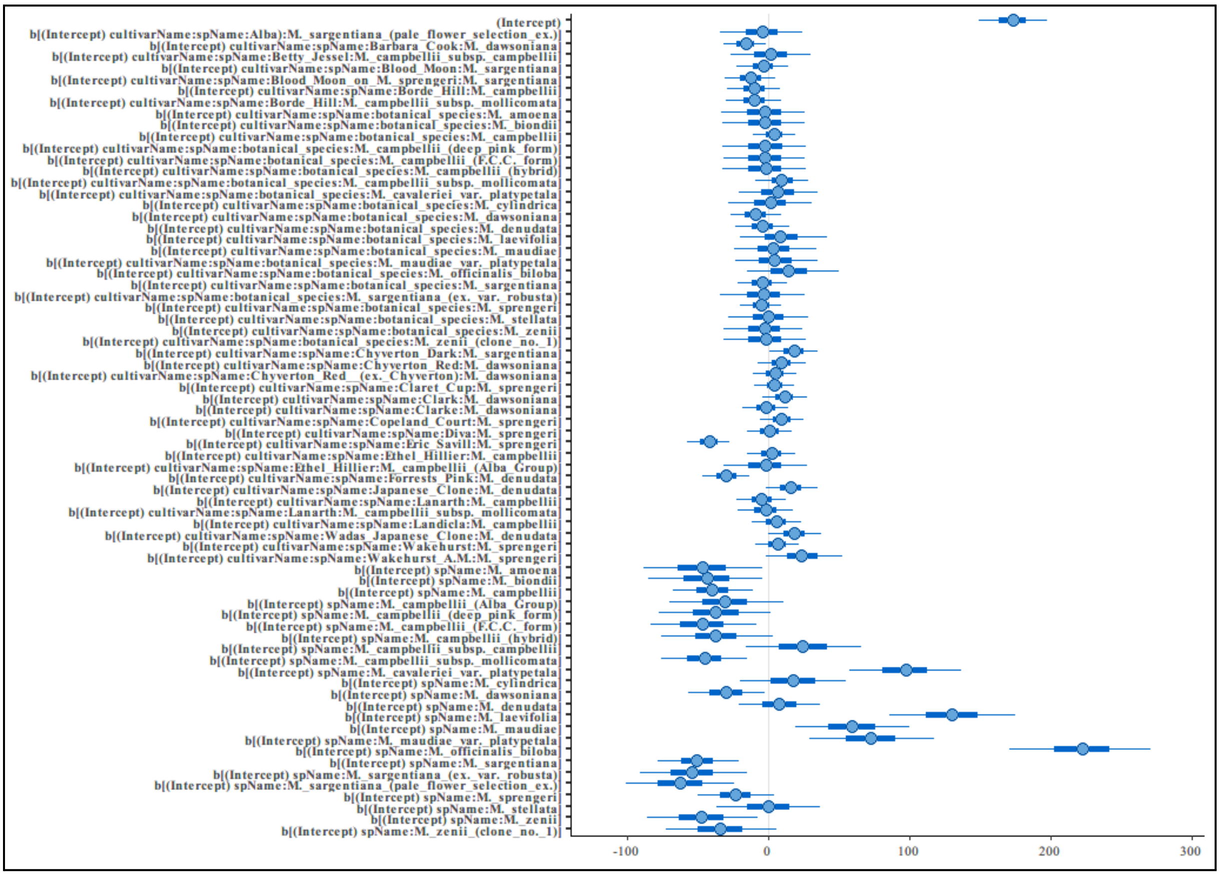
We decided to use a nested structure in which magnolias of a particular species have their respective cultivars clustered within them, treating cultivars as subsets of the botanical species. Models were run separately on each of the 5 phenological events, with the uncertainty intervals appearing like this for bud, anthesis, peak, ftepal, and ltepal events, respectively;



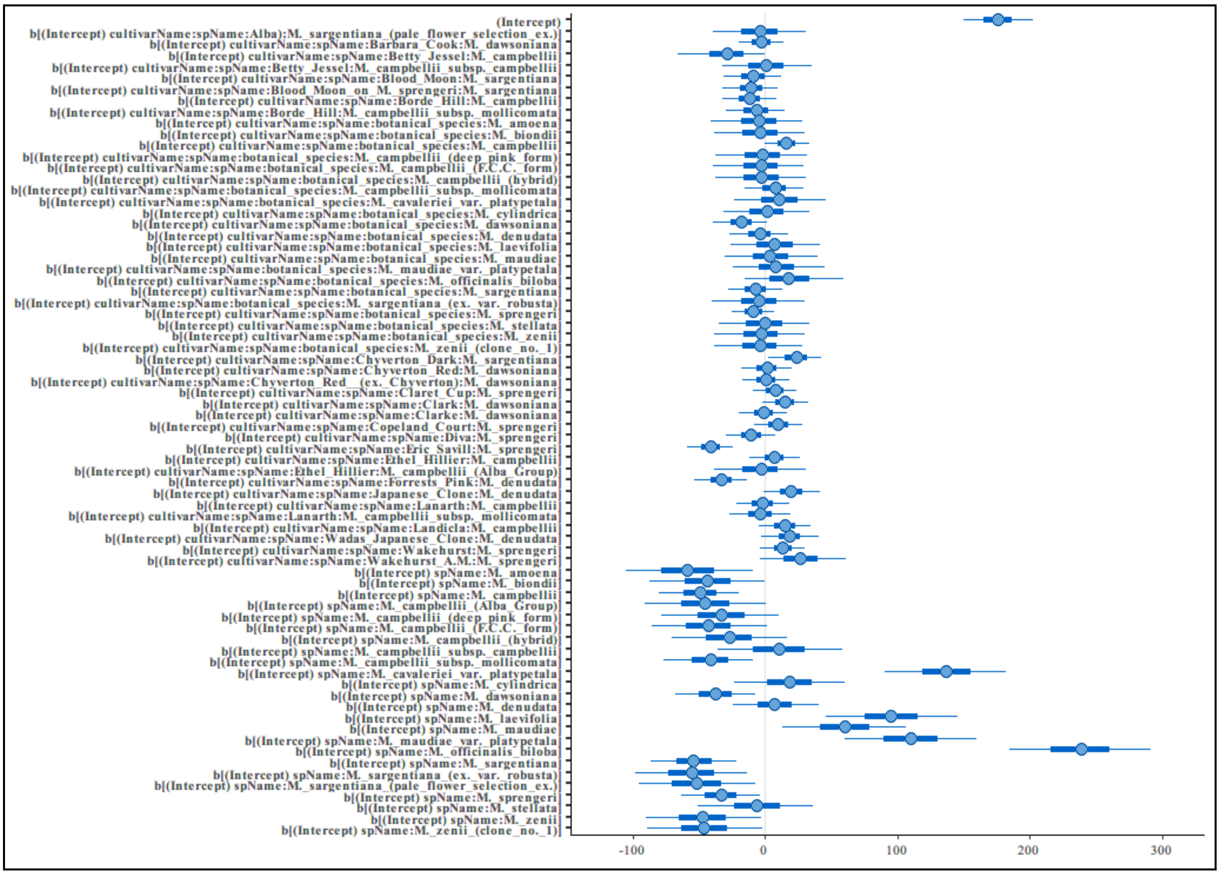
*Uncertainty intervals for **first bud colour (bud)** for cultivars nested within species as well as the species themselves.*



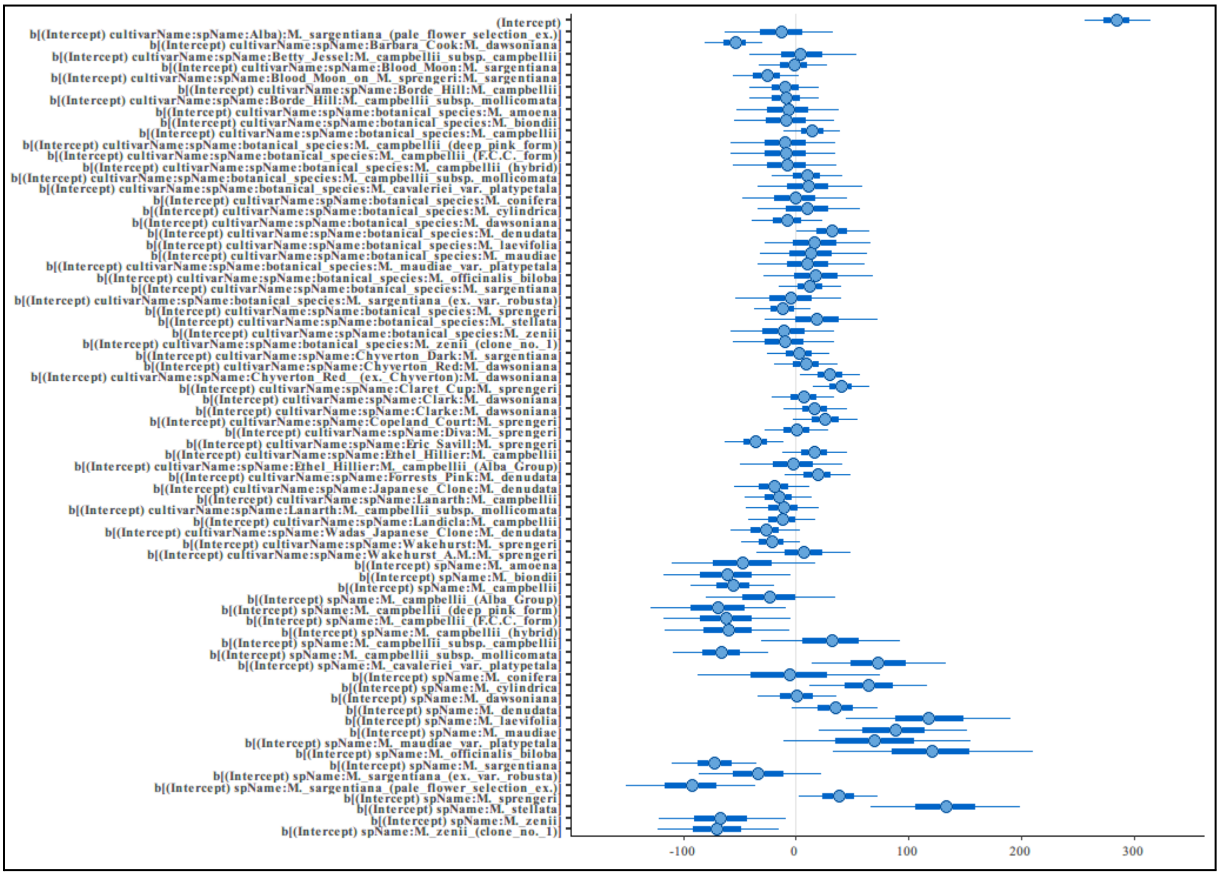
Uncertainty intervals for ***first flower fully opened (anthesis)*** *for cultivars nested within species as well as the species themselves.*



Uncertainty intervals for ***peak bloom (peak)*** *for cultivars nested within species as well as the species themselves.*



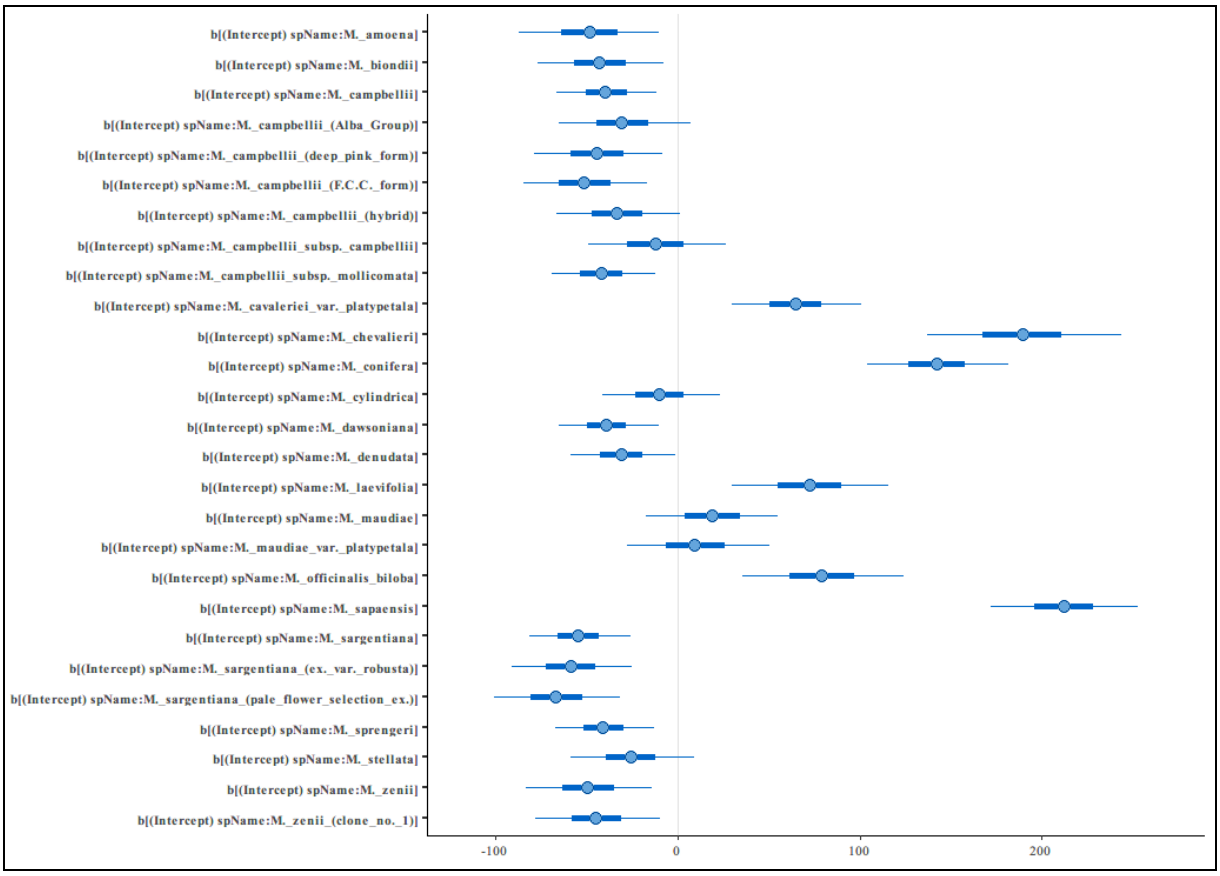
*Uncertainty intervals for **first tepal dropped (ftepal)** for cultivars nested within species as well as the species themselves.*



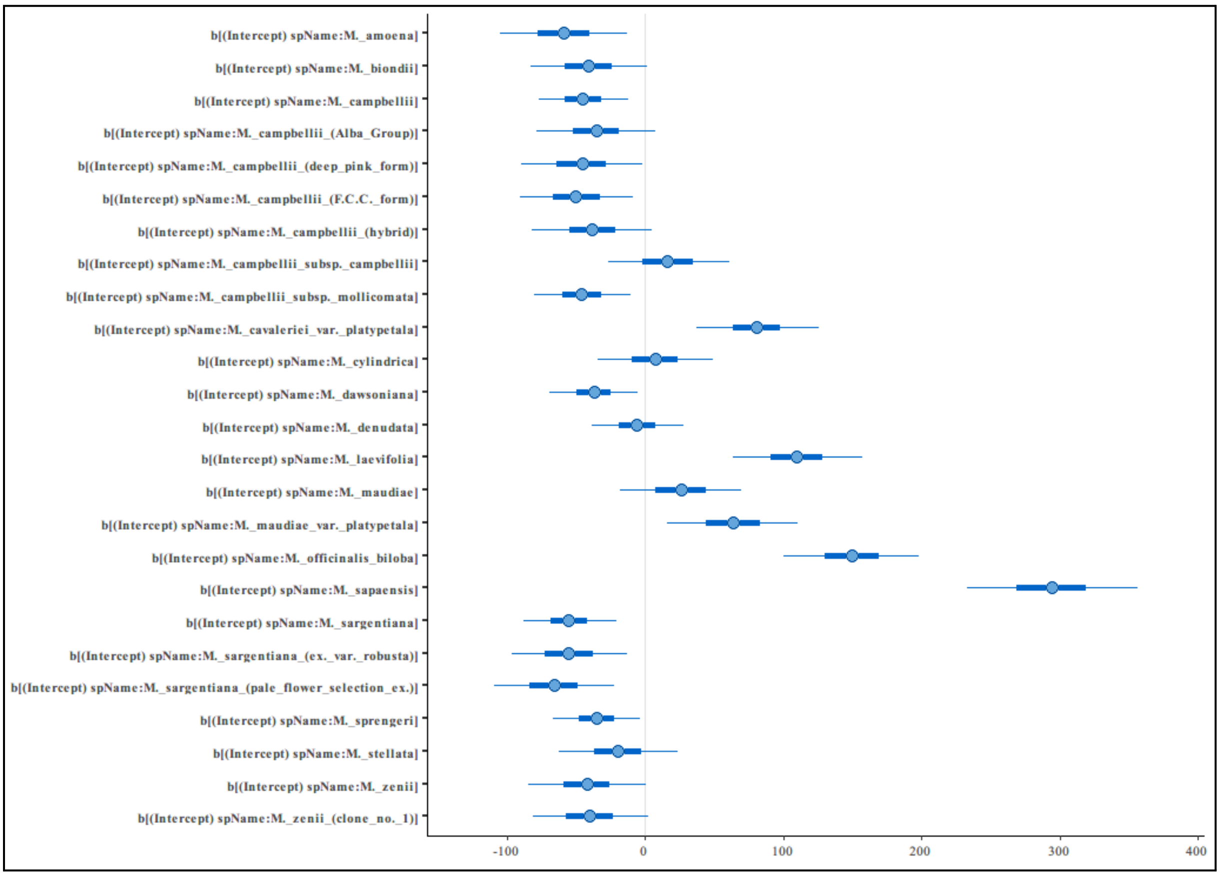
*Uncertainty intervals for **last tepal dropped (ltepal)** for cultivars nested within species as well as the species themselves.*

The blue bars and their corresponding margins of error represent the posterior interval estimates for each of the parameters, which in turn represent the offset value against an intercept GDD value of each of the species, or species by cultivar, by year. Blue dots at the centers represent the mean estimate of GDD accumulation needed for a given floral phenological event, relative to an intercept value (**139.3**). Thicker proximal blue bars and thinner distal blue bars adjacent represent the 50% and 90% uncertainty intervals, respectively. These intervals hint toward the observation that phenological events that occur later in the year tend to require a greater GDD accumulation; the same can thus be said for species and cultivars whose overall floral phenological processes occur later in general.

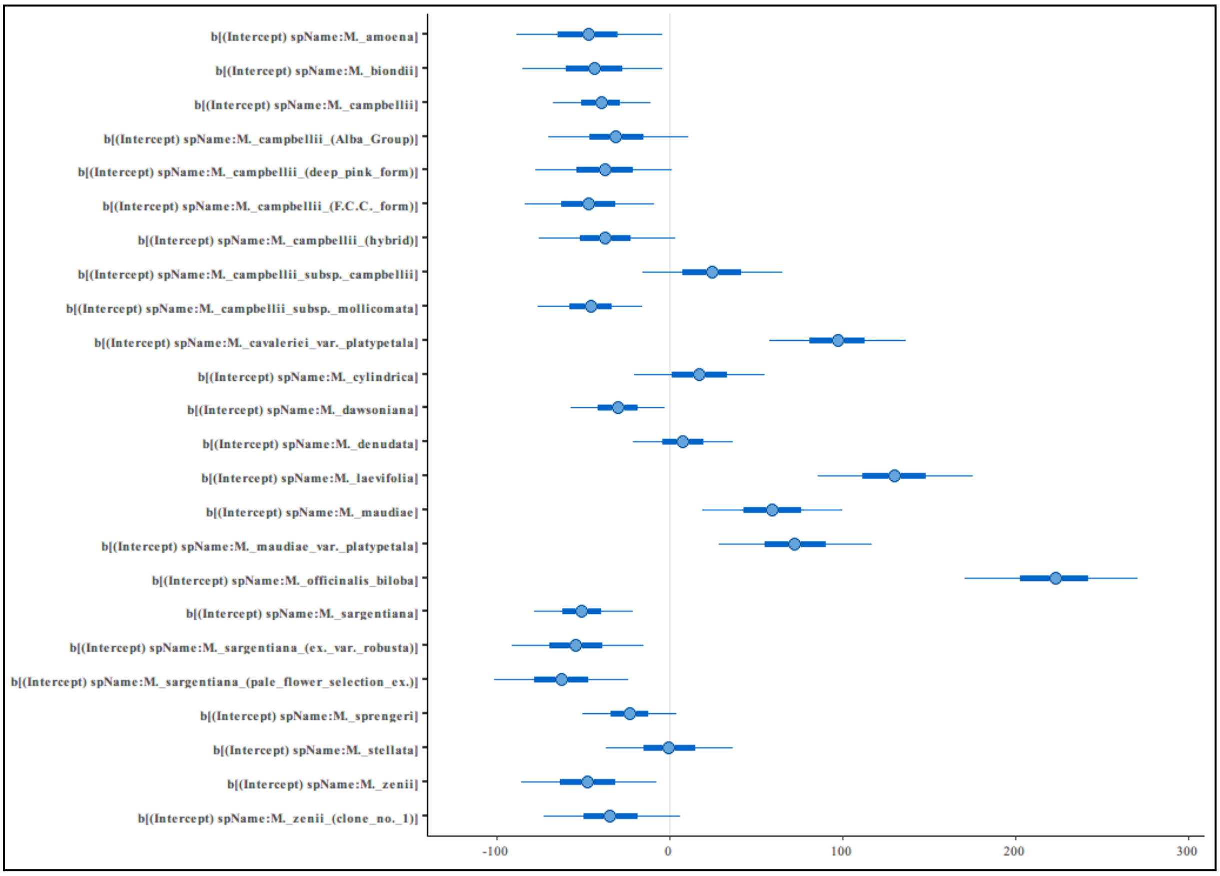
The parameters in the bottom third of the list represent the entire species altogether; that is, they include the variation in GDD offset of all cultivars of a particular species, nested under that species. These can be seen in the magnified uncertainty intervals below for bud, anthesis, peak, ftepal, and ltepal events, respectively.



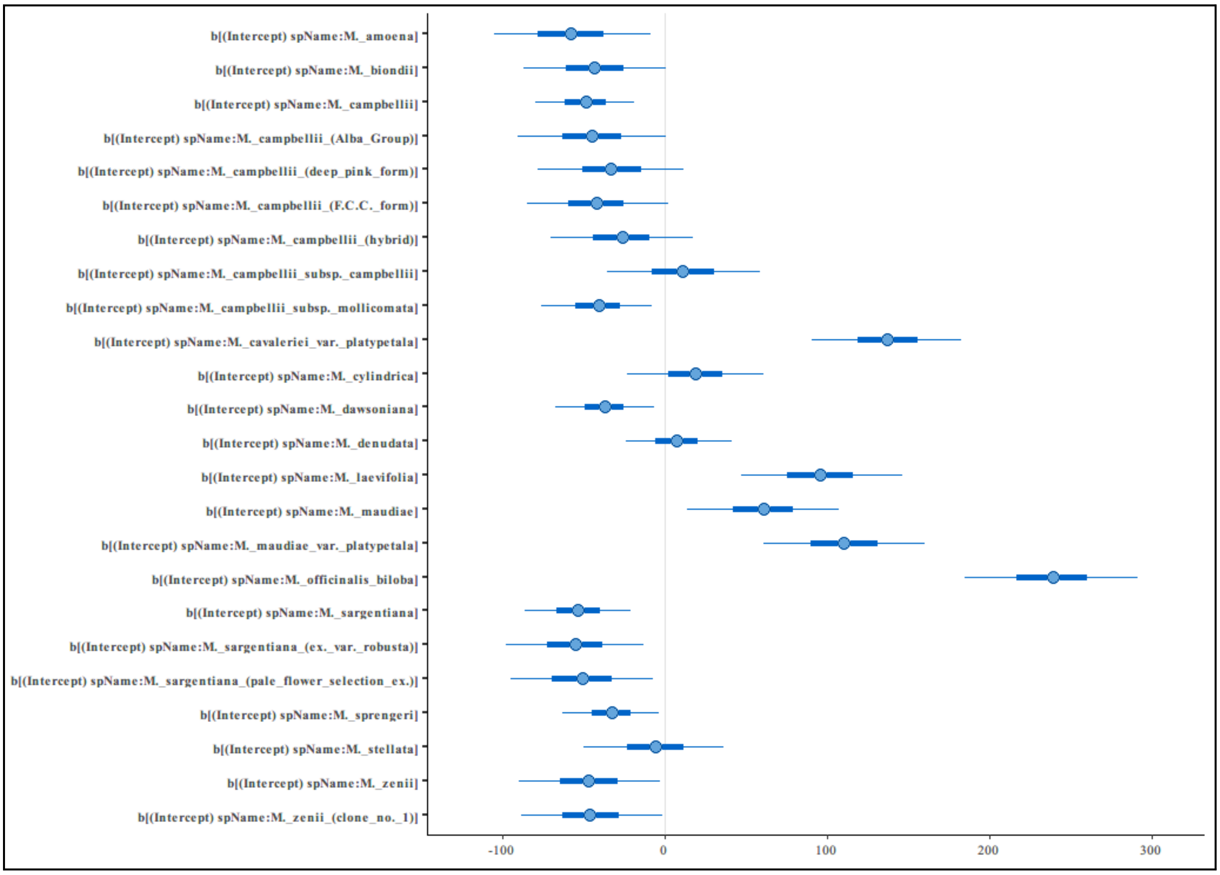
*Uncertainty intervals for **first bud colour (bud)** for just the species magnolias.*



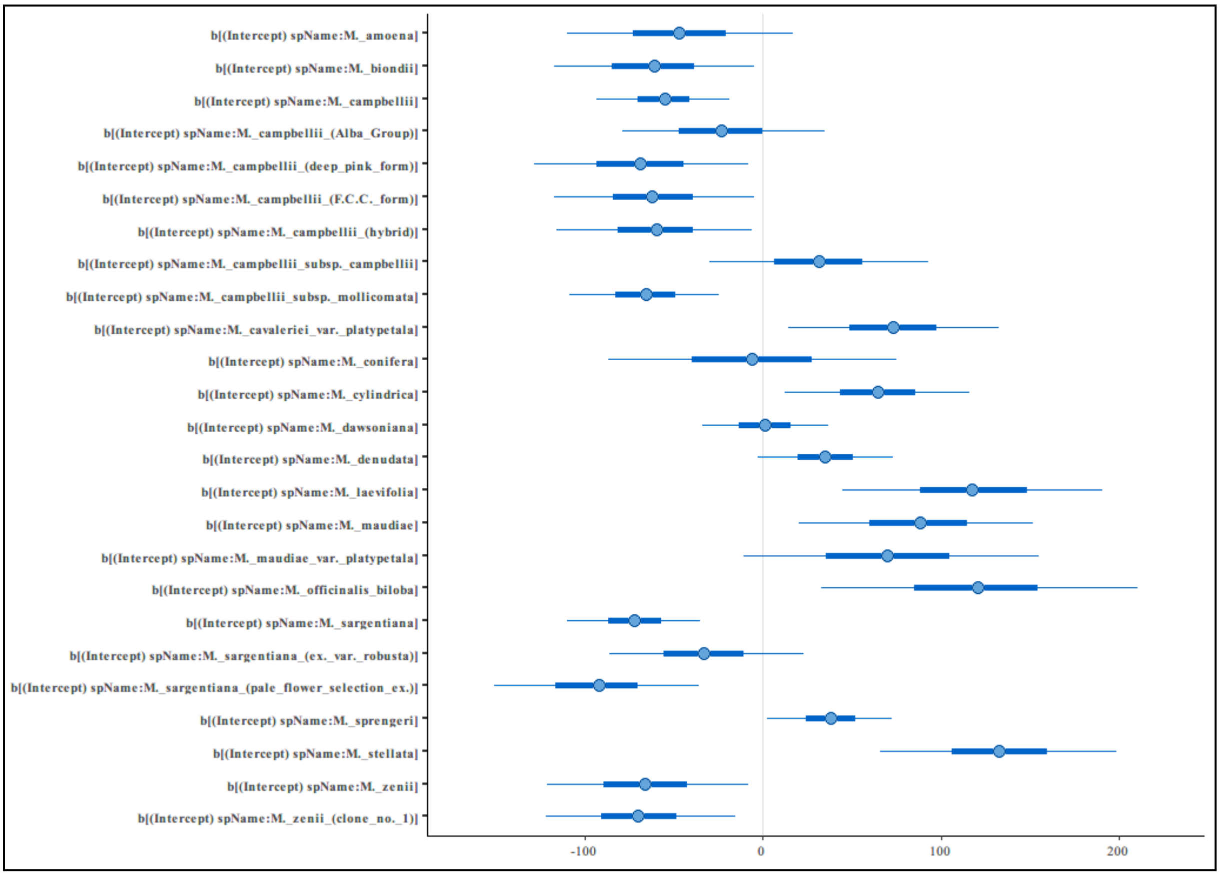
*Uncertainty intervals for **first flower fully opened (anthesis)** for just the species magnolias.*



*Uncertainty intervals for **peak bloom (peak)** for just the species magnolias.*



*Uncertainty intervals for **first tepal dropped (ftepal)** for just the species magnolias.*



*Uncertainty intervals for **last tepal dropped (ltepal)** for just the species magnolias.*

Because these central blue dots (mean GDD estimate) and bars (50% and 90% uncertainty intervals) represent the offset value relative to the intercept and not the whole GDD itself, the value of this intercept can be added onto the values of each of these parameters to show the posterior interval estimate for the actual GDD at which an individual reflected by a given parameter undergoes a given phenological event. As a result, the estimated GDD for a floral phenological event of a given species or its cultivar is the sum of its relative value - shown above - and the overall intercept's value. For example, if we are to calculate the overall GDD required for the last tepal dropped for species *Magnolia amoena*, we would take the value of the intercept (**139.3**) and add it to the value of the offset (**-47.1**), giving us a total GDD estimate of **92.2**. For the full tabulated data from which operations like this can be done, please see the data linked in section 4: Raw data and other metrics.

6] Future steps

Building upon this extensive and high-quality dataset, the garden should continue in their efforts to collect data of this type while incorporating additional species and cultivars not yet represented alongside introducing individuals of the same species or cultivar to increase sample size. This fulfills 4 main reasons;

1) Larger sample sizes will allow the model to more clearly elucidate whatever relationship currently naturally exists between GDDs and magnolia floral phenology
2) Broader scope may reveal diversity and variability in response to GDD across different species and cultivars, which may then also enable reasonably large-enough clusters of species which might provide insight into shared environment-response traits across species complexes within the genus
3) Aforementioned importance of maintaining this longitudinal research as it stands unique among other phenology-oriented datasets, which are primarily performed on Arctic communities or provenance trials for forestry and agriculture
4) Long-term gratification and fulfillment for volunteers (FOGs) who have worked on this dataset for many years, and may provide validation, vindication, and encouragement that their efforts are useful to not just the public, but specifically the scientific community

This report still only represents a relatively shallow dive into phenological modelling compared to what has been published out there in this field, especially community ecology. There is also a great opportunity for expansion, especially considering the overall importance of pollinator interactions, floral and seed predation, germination, and so on. Other kinds of analytical deep-dives into how climate affects floral - and subsequently reproductive - phenology might be possible as future tasks for this rich dataset, especially for student workers or research assistants who are interested in coding, statistics, ecology, and/or climatology.

7] Supplementary data

Linear regression model on annual temperature across the years

```
> summary(tempmodel)

Call:
lm(formula = temp.annual ~ year, data = tempmodeltable)
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.11371 -0.47273  0.01742  0.42235  1.63625

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 37.19848   26.70290   1.393    0.174
year        -0.01469    0.01330  -1.104    0.278

Residual standard error: 0.7278 on 31 degrees of freedom
Multiple R-squared:  0.03781, Adjusted R-squared:  0.006776
F-statistic: 1.218 on 1 and 31 DF,  p-value: 0.2782
```

Linear regression done on spring temperature across the years, where spring = mean average temperature across the months of March, April, and May.

```
> summary(springmodel)

Call:
lm(formula = avg_temperature ~ year, data = szn.spring)

Residuals:
     Min       1Q   Median       3Q      Max
-1.55143 -0.58303 -0.05638  0.47804  1.74985

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 24.30453   29.70398   0.818    0.419
year        -0.00725    0.01480  -0.490    0.628

Residual standard error: 0.8095 on 31 degrees of freedom
Multiple R-squared:  0.007682,       Adjusted R-squared:  -0.02433
F-statistic:  0.24 on 1 and 31 DF,  p-value: 0.6277

> paste('y =', coef(springmodel)[[2]], '* x', '+', coef(springmodel)[[1]])
[1] "y = -0.00725038508486477 * x + 24.3045301117004"
```

Linear regression done on summer temperature across the years, where spring = mean average temperature across the months of June, July, and August.

```
> summary(summermodel)

Call:
lm(formula = avg_temperature ~ year, data = szn.summer)

Residuals:
     Min       1Q   Median       3Q      Max
-0.87576 -0.48210  0.03165  0.22950  1.35813

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -34.89129   21.92107  -1.592   0.1216
```

```
year        0.02611    0.01092   2.390   0.0231 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5974 on 31 degrees of freedom
Multiple R-squared:  0.1556,   Adjusted R-squared:  0.1284
F-statistic: 5.713 on 1 and 31 DF,  p-value: 0.0231

> paste('y =', coef(summermodel)[[2]], '* x', '+', coef(summermodel)[[1]])
[1] "y = 0.0261071916414787 * x + -34.8912871158839"
```

Linear regression done on autumn temperature across the years, where spring = mean average temperature across the months of September, October, and November.

```
> summary(autumnmodel)

Call:
lm(formula = avg_temperature ~ year, data = szn.autumn)

Residuals:
     Min       1Q   Median       3Q      Max
-1.19116 -0.33397 -0.05786  0.19778  1.06073

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.520003  18.461705  -0.082    0.935
year         0.006068   0.009199   0.660    0.514

Residual standard error: 0.5032 on 31 degrees of freedom
Multiple R-squared:  0.01384, Adjusted R-squared:  -0.01797
F-statistic: 0.4352 on 1 and 31 DF,  p-value: 0.5143

> paste('y =', coef(autumnmodel)[[2]], '* x', '+', coef(autumnmodel)[[1]])
[1] "y = 0.00606793941352754 * x + -1.52000275459813"
```

Linear regression done on winter temperature across the years, where spring = mean average temperature across the months of December, January, and February.

```
> summary(wintermodel)

Call:
lm(formula = avg_temperature ~ year, data = szn.winter)

Residuals:
     Min       1Q   Median       3Q      Max
-1.95367 -0.45007  0.07056  0.67437  2.13281

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 28.30223   34.39269   0.823    0.417
year        -0.01197    0.01713  -0.699    0.490

Residual standard error: 0.98 on 32 degrees of freedom
Multiple R-squared:  0.01503, Adjusted R-squared:  -0.01576
F-statistic: 0.4881 on 1 and 32 DF,  p-value: 0.4898
```

```
> paste('y =', coef(wintermodel)[[2]], '* x', '+', coef(wintermodel)[[1]])
[1] "y = -0.0119695588741282 * x + 28.3022270476177"
```

Linear regression model outputs for minimum and maximum temperatures across the years according to each season; shown here, spring.

```
> summary(lm(min_temperature~year,
szn.spring))

Call:
lm(formula = min_temperature ~ year, data
= szn.spring)

Residuals:
    Min      1Q  Median      3Q     Max
-1.36269 -0.48305 -0.01212  0.55348
1.63382

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 22.226849  27.480258   0.809
0.425
year       -0.008058   0.013692  -0.588
0.560

Residual standard error: 0.7489 on 31
degrees of freedom
Multiple R-squared:  0.01105, Adjusted
R-squared:  -0.02085
F-statistic: 0.3463 on 1 and 31 DF,  p-value:
0.5605
```

```
> summary(lm(max_temperature~year,
szn.spring))

Call:
lm(formula = max_temperature ~ year, data
= szn.spring)

Residuals:
    Min      1Q  Median      3Q     Max
-1.74018 -0.66461  0.01279  0.64879
1.86532

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 26.396168  34.182074   0.772
0.446
year       -0.006449   0.017031  -0.379
0.708

Residual standard error: 0.9316 on 31
degrees of freedom
Multiple R-squared:  0.004604,
Adjusted R-squared:  -0.02751
F-statistic: 0.1434 on 1 and 31 DF,  p-value:
0.7075
```

Linear regression model outputs for minimum and maximum temperatures across the years according to each season; shown here, summer.

```
> summary(lm(min_temperature~year,
szn.summer))

Call:
lm(formula = min_temperature ~ year, data
= szn.summer)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7776 -0.3275 -0.1072  0.3409  1.0277

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.316338  18.398789  -1.050
0.3019
year          0.016288   0.009167   1.777
0.0854 .

Residual standard error: 0.5014 on 31
```

```
> summary(lm(max_temperature~year,
szn.summer))

Call:
lm(formula = max_temperature ~ year, data
= szn.summer)

Residuals:
    Min      1Q  Median      3Q     Max
-1.10518 -0.56197  0.00009  0.40499
1.68852

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -50.51392   27.34951  -1.847
0.0743 .
year          0.03595    0.01363   2.638
0.0129 *
```

<table>
<tr><td>degrees of freedom
Multiple R-squared: 0.09243, Adjusted R-squared: 0.06315
F-statistic: 3.157 on 1 and 31 DF, p-value: 0.08541</td><td>Residual standard error: 0.7454 on 31 degrees of freedom
Multiple R-squared: 0.1834,   Adjusted R-squared: 0.157
F-statistic: 6.961 on 1 and 31 DF, p-value: 0.01292</td></tr>
</table>

Linear regression model outputs for minimum and maximum temperatures across the years according to each season; shown here, autumn.

<table>
<tr><td>> summary(lm(min_temperature~year, szn.autumn))

Call:
lm(formula = min_temperature ~ year, data = szn.autumn)

Residuals:
    Min   1Q   Median   3Q   Max
-0.95986 -0.48659 -0.01251 0.39160 1.11886

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.927662 21.517218 -0.229   0.820
year   0.006096   0.010721   0.569   0.574

Residual standard error: 0.5864 on 31 degrees of freedom
Multiple R-squared: 0.01032, Adjusted R-squared: -0.0216
F-statistic: 0.3233 on 1 and 31 DF, p-value: 0.5737</td><td>> summary(lm(max_temperature~year, szn.autumn))

Call:
lm(formula = max_temperature ~ year, data = szn.autumn)

Residuals:
    Min   1Q Median   3Q   Max
-1.5697 -0.3417 0.1006 0.2462 1.2242

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.884375 20.995753   0.090   0.929
year   0.006043   0.010461   0.578   0.568

Residual standard error: 0.5722 on 31 degrees of freedom
Multiple R-squared: 0.01065, Adjusted R-squared: -0.02126
F-statistic: 0.3337 on 1 and 31 DF, p-value: 0.5677</td></tr>
</table>

Linear regression model outputs for minimum and maximum temperatures across the years according to each season; shown here, winter.

<table>
<tr><td>> summary(lm(min_temperature~year, szn.winter))

Call:
lm(formula = min_temperature ~ year, data = szn.winter)

Residuals:
    Min   1Q Median   3Q   Max
-2.6643 -0.5558 0.2035 0.7282 2.0898

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.7953936 38.6204114   0.046   0.963
year   -0.0001778 0.0192378 -0.009</td><td>> summary(lm(max_temperature~year, szn.winter))

Call:
lm(formula = max_temperature ~ year, data = szn.winter)

Residuals:
    Min   1Q Median   3Q   Max
-1.5435 -0.5106 0.1900 0.4724 2.1768

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 54.88575   31.70064   1.731   0.093 .
year   -0.02380   0.01579 -1.507   0.142</td></tr>
</table>

0.993 Residual standard error: 1.101 on 32 degrees of freedom Multiple R-squared: 2.67e-06, Adjusted R-squared: -0.03125 F-statistic: 8.545e-05 on 1 and 32 DF, p-value: 0.9927	Residual standard error: 0.9033 on 32 degrees of freedom Multiple R-squared: 0.06627, Adjusted R-squared: 0.0371 F-statistic: 2.271 on 1 and 32 DF, p-value: 0.1416

Linear regression model output for annual GDDs across the years.

```
> summary(gddmodel)

Call:
lm(formula = gdd.annual ~ year, data = gddmodeltable)

Residuals:
     4      9     15
-50.51  92.59 -42.09

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -49886.23  29149.52  -1.711    0.337
year            25.09     14.58   1.721    0.335

Residual standard error: 113.6 on 1 degrees of freedom
  (29 observations deleted due to missingness)
Multiple R-squared:  0.7475,   Adjusted R-squared:  0.4951
F-statistic: 2.961 on 1 and 1 DF,  p-value: 0.3351
```

8] References

IPCC [Core Writing Team, Lee, H. & Romero, J. (eds.)]. (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. doi: 10.59327/IPCC/AR6-9789291691647.