Homework 4 PLS 298

Elena Suglia

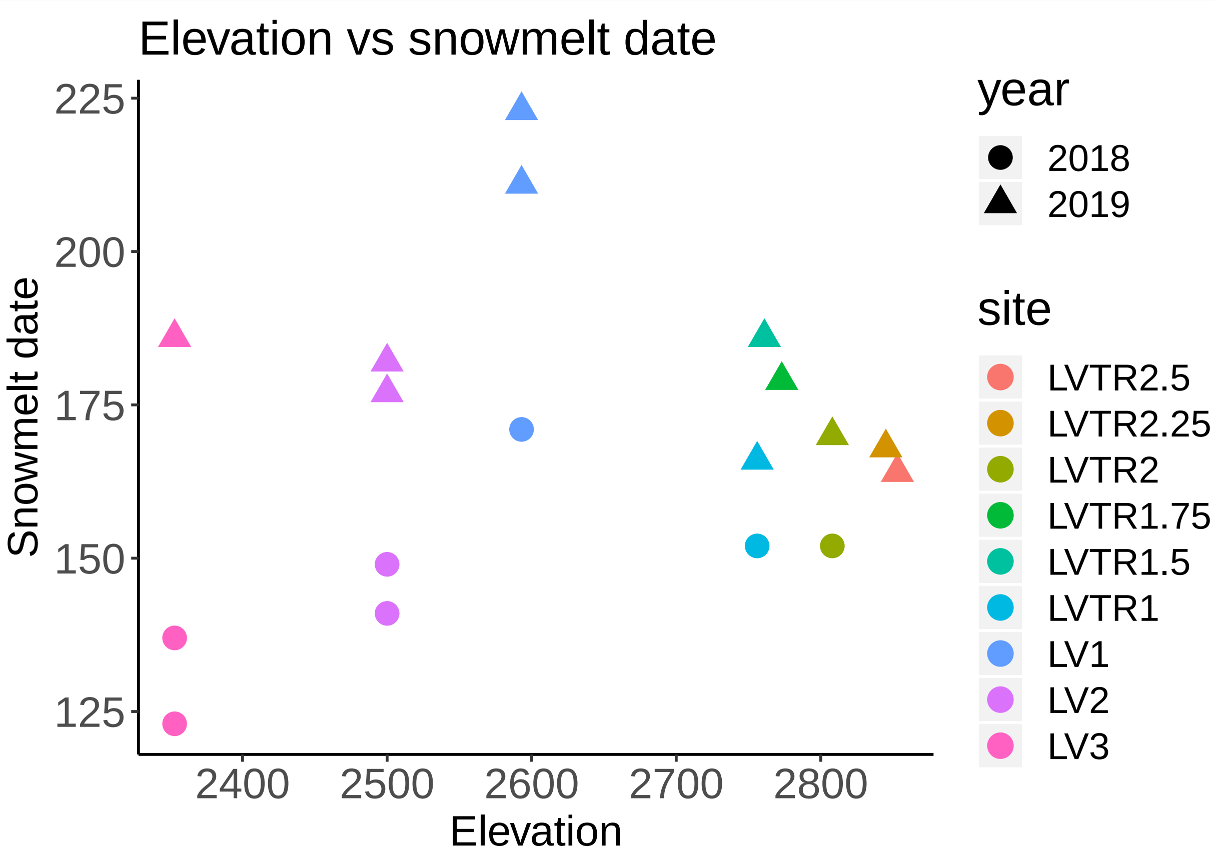
12/11/2019

The relationship between environment and phenology in the native California wildflower *Streptanthus tortuosus* at Lassen Volcanic National Park

This work represents preliminary analyses for a chapter of my dissertation proposal. In the following analyses, I'm investigating variation in phenology of *Streptanthus tortuosus* plants in relation to environmental variables at Lassen Volcanic National Park along an elevation gradient. I collected data every 7-10 days throughout the growing season from 2017-2019 at four populations with varying numbers of plots within those populations and varying numbers of marked individuals within those plots (ranging from 20-100, but fewer in some years). Plants at these populations are biennials or iteroparous perennials that germinate in the spring, grow into vegetative rosettes their first year and overwinter until spring the next year, when they melt out and flower. This data includes number of buds, flowers, and fruits on each plant every week, as well as some size and morphological metrics taken less frequently like height, stem diameter, number of branches, canopy width, and leaf size. I also measured temperature every 3 hours in 2018 and 2019 using small temperature sensors (iButtons) installed in the field (more were installed in 2019 than in 2018). The temperature data also provides me with metrics related to snowpack, an important abiotic factor in alpine environments. These data include snowmelt timing and length of snow-free period.

One thing to note here is that snowmelt timing does not vary linearly with elevation, which offers an opportunity to decouple snowmelt, elevation, and phenology in ways that are usually not feasible:

```
elsnowmelt
```



- Multiple points for some populations in some years indicate several plots.

The specific questions I would like to answer in this chapter are:

1. What is the relationship between environment and phenology across years and populations?
2. Do fitness and selection vary in relation to phenology and environment?
3. How does the environment relate to synchronization of flowering across populations?

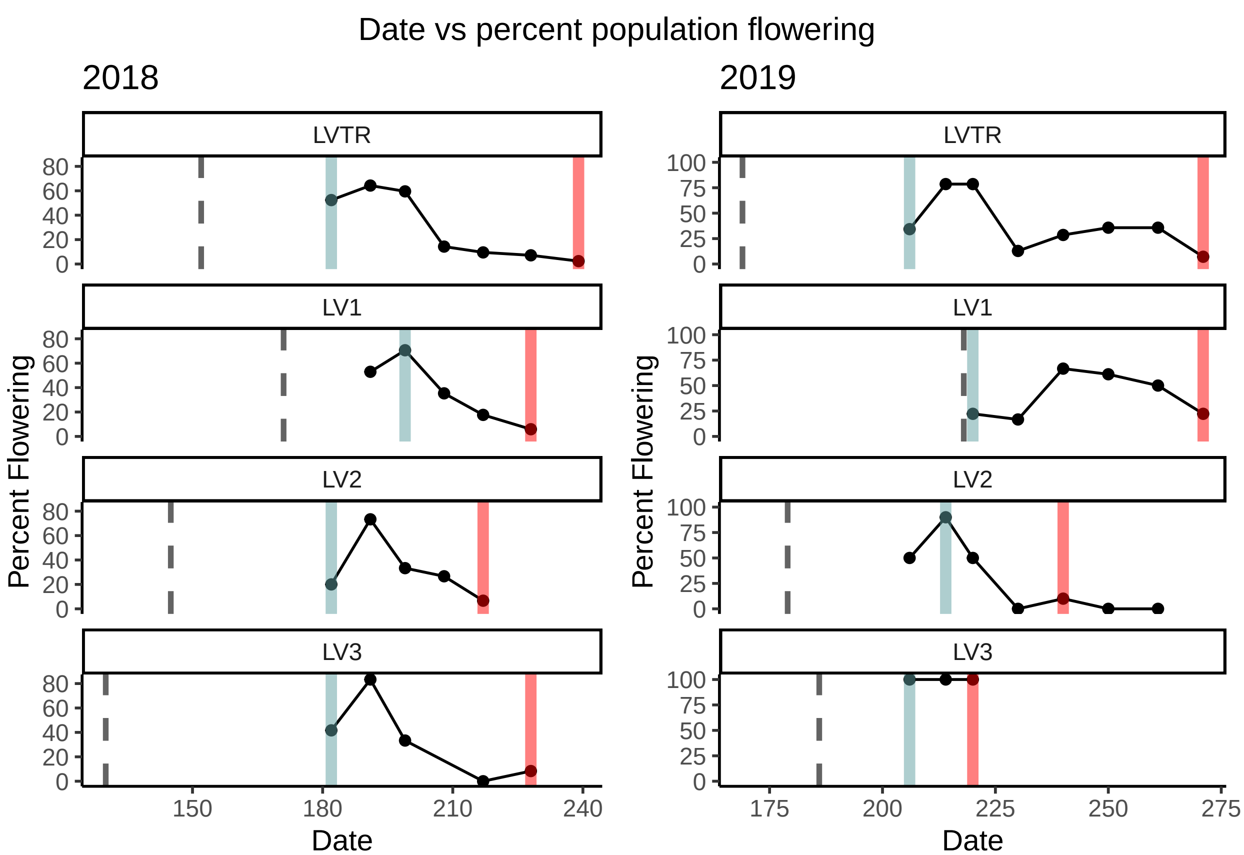
I realize the term "environment" is rather vague. This is the order of priority of environmental variables I'm interested in investigating, and I will complete as many of them as I can before my dissertation proposal is due:

- snowmelt timing
- temperature
- growing degree days
- photoperiod (daylength) accumulation
- photothermal accumulation

1. Relationship between environment and phenology

Let's visualize how populations are timing their flowering in relation to snowmelt timing:

```
require(grid)
grid.arrange(d5plot18, d5plot19, nrow = 1, top=textGrob("Date vs percent population flowering"))
```



- Populations are ordered from highest elevation at the top to lowest at the bottom
- The x-axis is day of the year (Julian day)

- The y-axis is percent of the reproductive individuals in the population that are flowering
- The dashed gray vertical lined is snowmelt timing
- The green vertical line is "peak flowering," which is the day on which that population has the greatest total number of flowers
- The red vertical line is the last day that the population had any flowers (included to provide a visualization of flowering duration between the green and red lines)
- I do not include first flowering in this figure because I missed it for several populations (it's too challening to reach certain populations before the snow melts out)
- Ignore LV3 in 2019 (bottom right population); it experienced a landslide halfway through the 2019 growing season that wiped out all my marked individuals

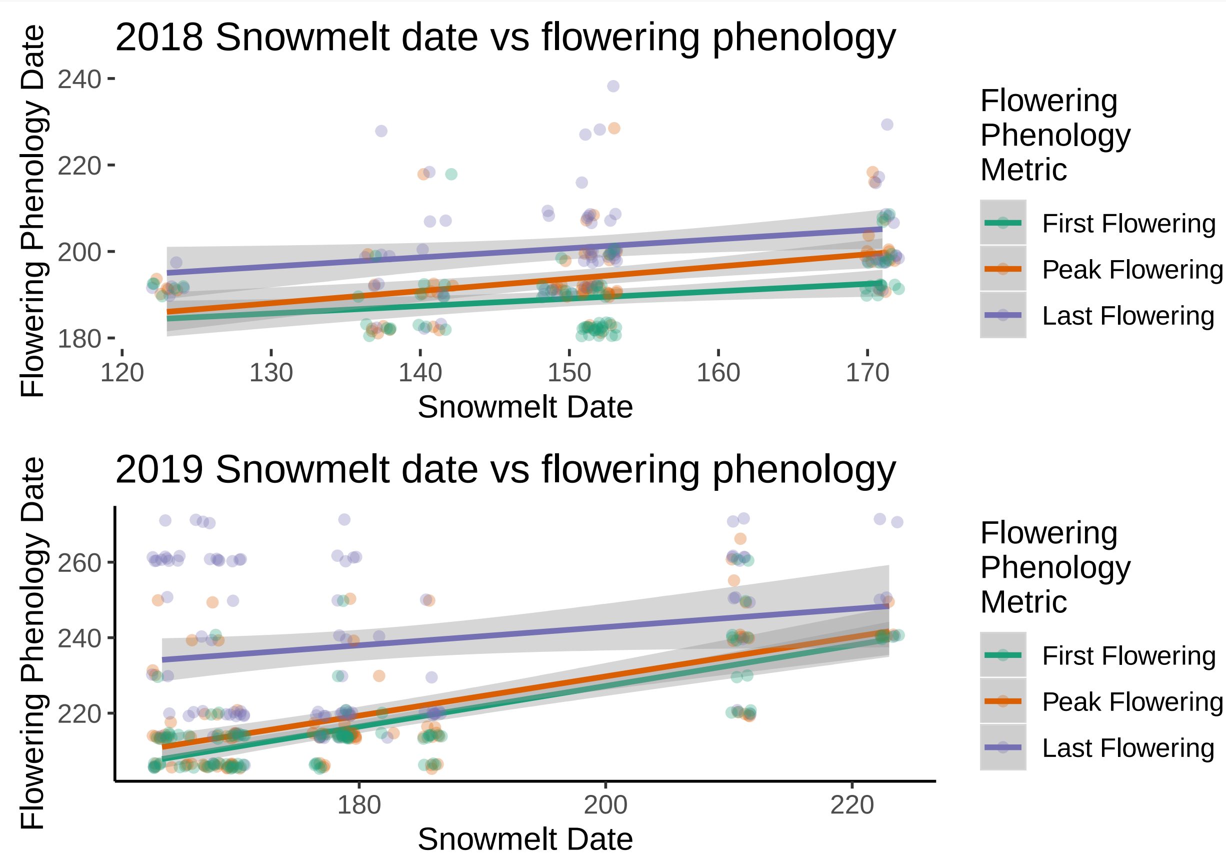
Here, we see that there is variation in snowmelt timing and flowering time across populations and years (note the different x axes in 2018 and 2019; the snowmelt timing and therefore the growing season were much delayed in 2019 compared to 2018). We also see what appears to be a relationship between snowmelt timing and days from snowmelt to peak flowering: the later the snow melts, the faster the population flowers.

Relationship between flowering time and snowmelt date

Let's look at the relationship between flowering date as response variables and snowmelt date as the explanatory variable. Here, I include population and year as random variables, and plot as a random variable nested within population.

First, we can visualize the relationships:

```
require(gridExtra)
grid.arrange(plotflrsnow18, plotflrsnow19, nrow=2)
```



There is a lot of individual variation, but later snowmelt appears to delay all metrics of flowering phenology.

Now let's do some statistical analyses. In the following, the day of year of snowmelt (Julian day) is denoted as "jmeltdate."

```r
# peakflr ~ jmeltdate model selection

# peakflr ~ jmeltdate fixed effects model & anova

peakflrlm = lm(peakflr ~ jmeltdate, data = i)
summary(peakflrlm)
```

```
##
## Call:
## lm(formula = peakflr ~ jmeltdate, data = i)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.383  -7.951  -1.901   4.377  40.377
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 103.53047    6.13232   16.88   <2e-16 ***
## jmeltdate     0.62777    0.03635   17.27   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.31 on 199 degrees of freedom
##   (1699 observations deleted due to missingness)
## Multiple R-squared:  0.5998, Adjusted R-squared:  0.5978
## F-statistic: 298.3 on 1 and 199 DF,  p-value: < 2.2e-16
```

```r
anova(peakflrlm)
```

```
## Analysis of Variance Table
##
## Response: peakflr
##            Df Sum Sq Mean Sq F value    Pr(>F)
## jmeltdate   1  38167   38167  298.29 < 2.2e-16 ***
## Residuals 199  25462     128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# later snowmelt significantly delays peak flowering

# mixed models for peakflr ~ jmeltdate
m6 = lmer(peakflr ~ jmeltdate + (1|site), data = i, REML = FALSE)
m7 = lmer(peakflr ~ jmeltdate + (1 + jmeltdate|site), data = i, REML = FALSE)
m8 = lmer(peakflr ~ jmeltdate + (1|year), data = i, REML = FALSE)
m9 = lmer(peakflr ~ jmeltdate + (1|site) + (1|year), data = i, REML = FALSE)
m10 = lmer(peakflr ~ jmeltdate*year + (1|site/plot), data = i, REML = FALSE) # best fit
m11 = lmer(peakflr ~ jmeltdate*site + (1|year), data = i, REML = FALSE)
m12 = lmer(peakflr ~ jmeltdate*site + (1|year) + (1|site), data = i, REML = FALSE)
m13 = lmer(peakflr ~ jmeltdate*site + (1 + site|year), data = i, REML = FALSE)

AIC(peakflrlm, m6, m7, m8, m9, m10, m11, m12, m13)
```

```
##           df      AIC
```

```
## peakflrlm  3 1549.586
## m6         4 1547.705
## m7         6 1534.054
## m8         4 1523.306
## m9         5 1525.306
## m10        7 1518.786
## m11       10 1505.357
## m12       11 1507.357
## m13       19 1523.357
```

```r
AICc(peakflrlm, m6, m7, m8, m9, m10, m11, m12, m13)
```

```
##           df     AICc
## peakflrlm  3 1549.707
## m6         4 1547.909
## m7         6 1534.487
## m8         4 1523.510
## m9         5 1525.614
## m10        7 1519.366
## m11       10 1506.515
## m12       11 1508.754
## m13       19 1527.556
```

```r
BIC(peakflrlm, m6, m7, m8, m9, m10, m11, m12, m13)
```

```
##           df      BIC
## peakflrlm  3 1559.495
## m6         4 1560.918
## m7         6 1553.874
## m8         4 1536.519
## m9         5 1541.822
## m10        7 1541.909
## m11       10 1538.390
## m12       11 1543.693
## m13       19 1586.120
```

```r
# best fit model is model 10 and is consistent across model comparison methods


#r.squaredGLMM(m6)
#r.squaredGLMM(m7)
#r.squaredGLMM(m8)
#r.squaredGLMM(m9)
r.squaredGLMM(m10)
```

```
##            R2m      R2c
## [1,] 0.4963072 0.695956
```

```r
#r.squaredGLMM(m11)
#r.squaredGLMM(m12)
#r.squaredGLMM(m13)

# including random effects improves all models; displayed results of best fit model only

# firstflr ~ jmeltdate model selection

# firstflr ~ jmeltdate fixed model & anova
```

```r
firstflrlm = lm(firstflr ~ jmeltdate, data = i)
summary(firstflrlm)
```

```
##
## Call:
## lm(formula = firstflr ~ jmeltdate, data = i)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.273  -8.053   0.275   8.275  44.275
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 93.46710    6.09162   15.34   <2e-16 ***
## jmeltdate    0.66425    0.03631   18.30   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.46 on 185 degrees of freedom
##   (1713 observations deleted due to missingness)
## Multiple R-squared:  0.6441, Adjusted R-squared:  0.6421
## F-statistic: 334.7 on 1 and 185 DF,  p-value: < 2.2e-16
```

```r
anova(firstflrlm)
```

```
## Analysis of Variance Table
##
## Response: firstflr
##            Df Sum Sq Mean Sq F value    Pr(>F)
## jmeltdate   1  36611   36611  334.74 < 2.2e-16 ***
## Residuals 185  20234     109
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# later snowmelt significantly delays first flowering

m14 = lmer(firstflr ~ jmeltdate + (1|site), data = i, REML = FALSE)
m15 = lmer(firstflr ~ jmeltdate + (1 + jmeltdate|site), data = i, REML = FALSE)
m16 = lmer(firstflr ~ jmeltdate + (1|year), data = i, REML = FALSE)
m17 = lmer(firstflr ~ jmeltdate + (1|site) + (1|year), data = i, REML = FALSE)
m18 = lmer(firstflr ~ jmeltdate*year + (1|site), data = i, REML = FALSE)
m19 = lmer(firstflr ~ jmeltdate*site + (1|year), data = i, REML = FALSE) # best fit AIC
m20 = lmer(firstflr ~ jmeltdate*site + (1|year) + (1|site), data = i, REML = FALSE)
m21 = lmer(firstflr ~ jmeltdate*site + (1 + site|year), data = i, REML = FALSE)
# some warning messages

AIC(firstflrlm, m14, m15, m16, m17, m18, m19, m20, m21)
```

```
##            df      AIC
## firstflrlm  3 1412.590
## m14         4 1399.690
## m15         6 1336.655
## m16         4 1345.345
## m17         5 1343.156
## m18         6 1314.124
```

```
## m19        10 1298.278
## m20        11 1300.278
## m21        19 1316.278
```

```r
# AICc(firstflrlm, m14, m15, m16, m17, m18, m19, m20, m21)
# BIC(firstflrlm, m14, m15, m16, m17, m18, m19, m20, m21) # BIC gives m18 as best fit model

#r.squaredGLMM(m14)
#r.squaredGLMM(m15)
#r.squaredGLMM(m16)
#r.squaredGLMM(m17)
#r.squaredGLMM(m18)
r.squaredGLMM(m19)
```

```
##            R2m       R2c
## [1,] 0.8215661 0.8215661
```

```r
r.squaredGLMM(m20)
```

```
##            R2m       R2c
## [1,] 0.8215661 0.8215661
```

```r
r.squaredGLMM(m21)
```

```
##           R2m       R2c
## [1,] 0.821566 0.8215661
```

```r
# including random effects does not necessarily appear to help
# why are the last three identical? What do the messages mean?

# lastflr ~ jmeltdate model selection

# lastflr ~ jmeltdate fixed effects model
lastflrlm = lm(lastflr ~ jmeltdate, data = i)
summary(lastflrlm)
```

```
##
## Call:
## lm(formula = lastflr ~ jmeltdate, data = i)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.049 -11.195  -2.911   7.013  48.221
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 97.68364   11.60434   8.418 1.03e-14 ***
## jmeltdate    0.74021    0.06916  10.703  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.92 on 185 degrees of freedom
##   (1713 observations deleted due to missingness)
## Multiple R-squared:  0.3824, Adjusted R-squared:  0.3791
## F-statistic: 114.5 on 1 and 185 DF,  p-value: < 2.2e-16
```

```r
anova(lastflrlm)
```

```
## Analysis of Variance Table
##
## Response: lastflr
##            Df Sum Sq Mean Sq F value    Pr(>F)
## jmeltdate   1  45463   45463  114.55 < 2.2e-16 ***
## Residuals 185  73426     397
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# mixed models for meltdate ~ lastflr
m22 = lmer(lastflr ~ jmeltdate + (1|site), data = i, REML = FALSE)
m23 = lmer(lastflr ~ jmeltdate + (1 + jmeltdate|site), data = i, REML = FALSE)
m24 = lmer(lastflr ~ jmeltdate + (1|year), data = i, REML = FALSE)
m25 = lmer(lastflr ~ jmeltdate + (1|site) + (1|year), data = i, REML = FALSE)
m26 = lmer(lastflr ~ jmeltdate*year + (1|site), data = i, REML = FALSE) # best fit BIC
m27 = lmer(lastflr ~ jmeltdate*site + (1|year), data = i, REML = FALSE) # best fit AIC
m28 = lmer(lastflr ~ jmeltdate*site + (1|year) + (1|site), data = i, REML = FALSE)
m29 = lmer(lastflr ~ jmeltdate*site + (1 + site|year), data = i, REML = FALSE)

AIC(lastflrlm, m22, m23, m24, m25, m26, m27, m28, m29)
```

```
##           df      AIC
## lastflrlm  3 1653.619
## m22        4 1620.777
## m23        6 1605.988
## m24        4 1594.386
## m25        5 1592.343
## m26        6 1581.786
## m27       10 1572.177
## m28       11 1574.177
## m29       19 1590.177
```

```r
AICc(lastflrlm, m22, m23, m24, m25, m26, m27, m28, m29)
```

```
##           df     AICc
## lastflrlm  3 1653.750
## m22        4 1620.996
## m23        6 1606.455
## m24        4 1594.606
## m25        5 1592.674
## m26        6 1582.252
## m27       10 1573.427
## m28       11 1575.686
## m29       19 1594.728
```

```r
BIC(lastflrlm, m22, m23, m24, m25, m26, m27, m28, m29)
```

```
##           df      BIC
## lastflrlm  3 1663.313
## m22        4 1633.701
## m23        6 1625.375
## m24        4 1607.311
## m25        5 1608.499
## m26        6 1601.172
```

```
## m27       10 1604.488
## m28       11 1609.720
## m29       19 1651.569
```

```r
# BIC says model 26 is best

#r.squaredGLMM(m22)
#r.squaredGLMM(m23)
#r.squaredGLMM(m24)
#r.squaredGLMM(m25)
#r.squaredGLMM(m26)
r.squaredGLMM(m27)
```

```
##            R2m       R2c
## [1,] 0.6305272 0.6305272
```

```r
#r.squaredGLMM(m28)
#r.squaredGLMM(m29)
```

What about the relationship between flowering date and elevation? Using peak flowering as the representative flowering phenology metric from now on.

```r
# peakflr ~ elevation
elevpeakflrlm = lm(peakflr ~ elevation, data = i)
summary(elevpeakflrlm)
```

```
##
## Call:
## lm(formula = peakflr ~ elevation, data = i)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.793 -16.652  -1.389  18.611  58.549
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 290.582135  18.211419  15.956  < 2e-16 ***
## elevation    -0.026685   0.006936  -3.847 0.000143 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.82 on 334 degrees of freedom
##   (1564 observations deleted due to missingness)
## Multiple R-squared:  0.04244,    Adjusted R-squared:  0.03957
## F-statistic:  14.8 on 1 and 334 DF,  p-value: 0.0001429
```

```r
anova(elevpeakflrlm)
```

```
## Analysis of Variance Table
##
## Response: peakflr
##            Df Sum Sq Mean Sq F value    Pr(>F)
## elevation   1   6416  6416.2  14.803 0.0001429 ***
## Residuals 334 144767   433.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# significant, but less so than snowmelt date
```

flowering duration ~ jmeltdate —-

meltdatedurationlm = lm(flrduration ~ jmeltdate, data = i) summary(meltdatedurationlm) anova(meltdatedurationlm) # this is a pretty interesting result. Flowering duration is more influenced when snowmelt is later (or is it something else about that year?) # days from snowmelt to first flr ~ jmeltdate —- meltdatesnowtoflrlm = lm(snowtoflr ~ jmeltdate, data = i) summary(meltdatesnowtoflrlm) anova(meltdatesnowtoflrlm) AIC(meltdatesnowtoflrlm) # flower much faster when snow melts later. This could either be developmental (growth is faster in warmer temps) or due to anticipatory/photoperiod cuing to synchronize flowering with other plants.

2.2 Relationships between environment, phenology, and fitness —-

fitness ~ snowmelt date —-

fixed effects model & anova

fitnesssnowlm = lm(fitness ~ jmeltdate, data = i) summary(fitnesssnowlm) anova(fitnesssnowlm) # site is important. not sure how to interpret this result

mixed model

m1 = lmer(fitness ~ jmeltdate + (1 + jmeltdate|year), data = i, REML = FALSE) m2 = lmer(fitness ~ jmeltdate + (1|site), data = i, REML = FALSE) m2.5 = lmer(fitness ~ jmeltdate + (1|jmeltdate), data = i, REML = FALSE) m3 = lmer(fitness ~ jmeltdate + (1|year), data = i, REML = FALSE) m4 = lmer(fitness ~ jmeltdate + (1 + jmeltdate|year) + (1|site), data = i, REML = FALSE) m5 = lmer(fitness ~ jmeltdate*year + (1|site), data = i, REML=FALSE)

display(m1) display(m2) AIC(fitnesssnowlm, m1, m2, m2.5, m3, m4, m5) AICc(fitnesssnowlm, m1, m2, m2.5, m3, m4, m5) BIC(fitnesssnowlm, m1, m2, m2.5, m3, m4, m5)

Do we include the random effects or not?

The marginal R2 value represents the variance explained by the fixed effects

The conditional R2 value is interpreted as the variance explained by the entire model, including both fixed and random effects

r.squaredGLMM(m1) # R2m R2c # 0.0560748 0.1495234 r.squaredGLMM(m2) r.squaredGLMM(m2.5) r.squaredGLMM(m3) r.squaredGLMM(m4) r.squaredGLMM(m5) # all models perform better when including random effects

fitness ~ peakflr —-

fitnesspeakflrlm = lm(fitness ~ peakflr, data = i) summary(flrlm) anova(flrlm)

fitness ~ firstflr —-

fitnessfirstflrlm = lm(fitness ~ firstflr, data = i) summary(flrlm) anova(flrlm)

fitness ~ days from snowmelt to first flr —-

snowfirstflrlm = lm(fitness ~ snowtoflr*year, data = i) summary(snowfirstflrlm) anova(snowfirstflrlm) # relationship between fitness and days from snowmelt to first flowering are totally dependent on year!

fitness ~ branchiness —-

branchfitnesslm = lm(fitness ~ branchiness, data = i) summary(branchfitnesslm) anova(branchfitnesslm) # Those with more (max) total branches had significantly higher fecundity.