Pieris virginiensis progress report 04

Jeff Oliver

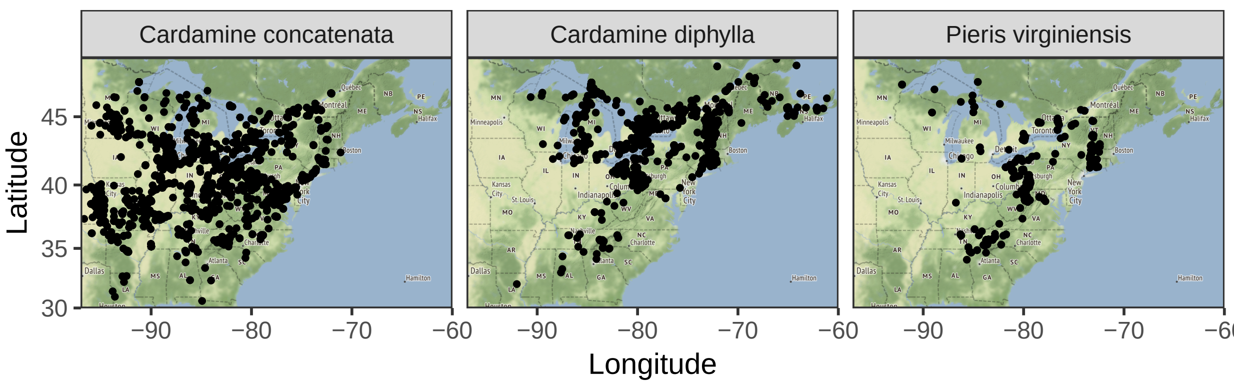
06 August, 2019

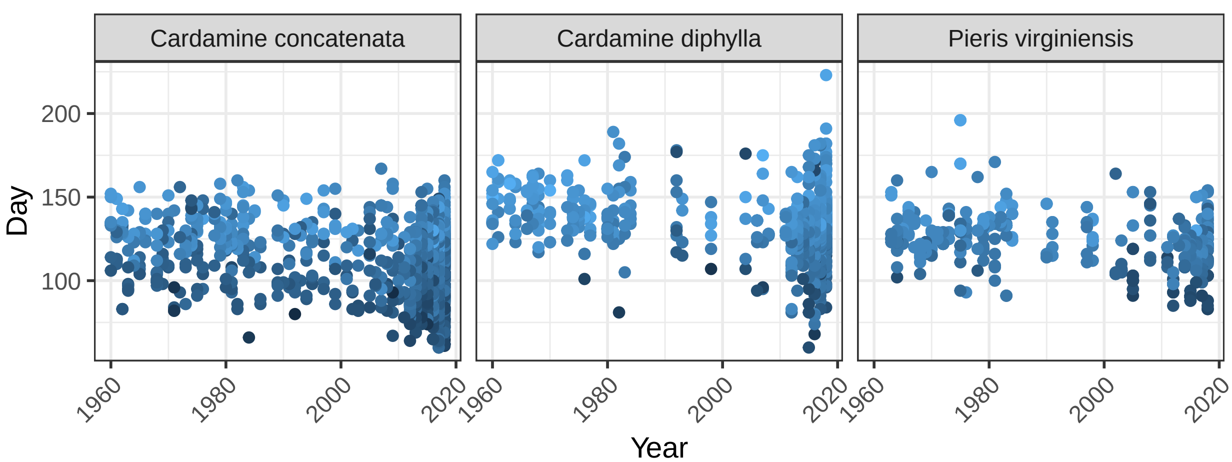
Retrieval from iNaturalist and GBIF returned 3369 total observations.

Following date and sample size filtering, observation counts for species:

- *Pieris virginiensis*: 544 observations
- *Cardamine concatenata*: 1963 observations
- *Cardamine diphylla*: 862 observations

Observations for 1960 - 2018 (total of 3369 observations following date and sample size filtering):





Note in lower plot, **darker** points are from lower latitudes, while **lighter** points are from higher latitudes.

Test for change in peak over time

Ultimately, we want to know if the changes in insect phenology are matching the changes in host phenology. That is, is the effect of year on day of year for observations the same for insect and host(s). In this case, there are two host plants, and one question is: do we analyze them separately, or do we combine them? The latter is easier to do, and, in some ways, easier to interpret:

$$Julian\ day = \beta_0 + \beta_1 Year + \beta_2 Latitude + \beta_3 Category + \beta_4 Year \times Category$$

where there are two levels to $Category$: insect and host

Model visualization

We can plot these in a couple of ways. First, we'll plot the predicted lines for several latitudes, with separate lines for insect and host. To calculate the slopes and intercepts, we need to look at the model formula and do some rearrangements. For the insect lines, the "insect" category is the reference, so we can drop all terms from the model that include $Category$ (because $Category = 0$ for all insect observations), and the model becomes:

$$Julian\ day = \beta_0 + \beta_1 Year + \beta_2 Latitude$$

And rearranging:

$$Julian\ day = \beta_0 + \beta_2 Latitude + \beta_1 Year$$

For a single value of $Latitude$, the first two terms constitute the intercept ($\beta_0 + \beta_2 Latitude$) and the last coefficient is the slope (β_1).

For the host lines, $Category = 1$, so the model becomes:

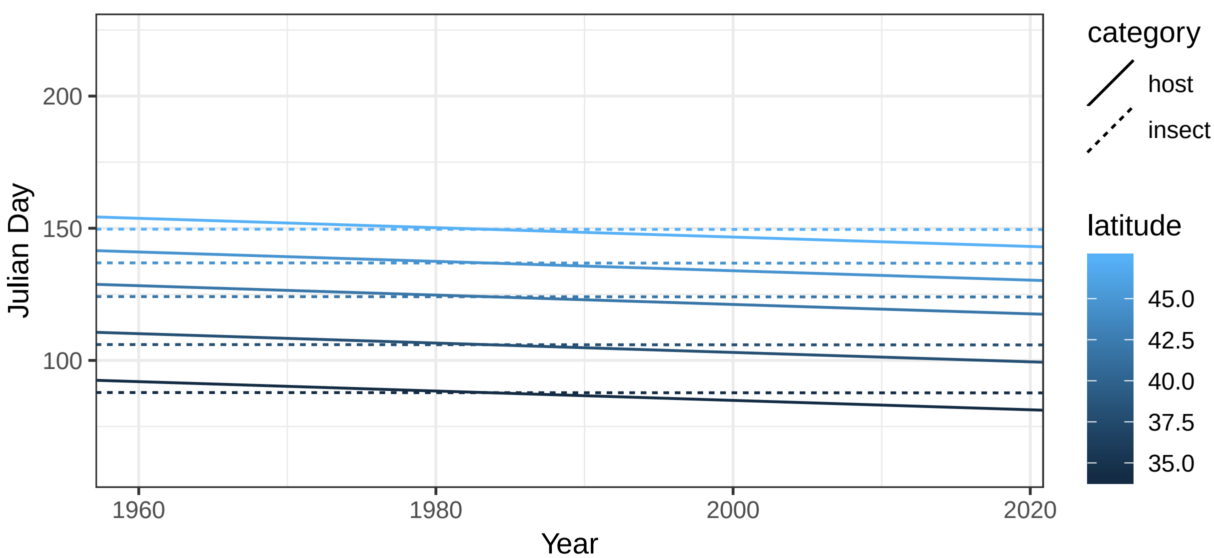
$$Julian\ day = \beta_0 + \beta_1 Year + \beta_2 Latitude + \beta_3 + \beta_4 Year$$

Rearranging:

$$Julian\ day = \beta_0 + \beta_3 + \beta_2 Latitude + (\beta_1 + \beta_4) Year$$

For a single value of $Latitude$, the intercept is defined by the sum of the first three terms ($\beta_0 + \beta_3 + \beta_2 Latitude$) and the slope is the sum of the coefficients β_1 and β_4.

Using those formulae, we can draw regression lines for a series of latitudes for insect and host responses through time.



Separating hosts

There is question (in my mind at least) about the lumping together of both hosts. It may be more appropriate to replace the $Category$ predictor with $Species$, keeping *P. virginensis* as the reference for ease of interpretation.

$$Julian\ day = \beta_0 + \beta_1 Year + \beta_2 Latitude + \beta_3 Species + \beta_4 Year \times Species$$

However, because there are **three** levels to $Species$, a more accurate representation of this model would be

$$Julian\ day = \beta_0 + \beta_1 Year + \beta_2 Latitude + \beta_3 concatenata + \beta_4 diphylla + \beta_5 Year \times concatenata + \beta_6 Year \times diphylla$$

(so long as *P. virginensis* is reference).

The model for *P. virginensis* remains the same:

$$Julian\ day = \beta_0 + \beta_1 Year + \beta_2 Latitude$$

And rearranging:

$$Julian\ day = \beta_0 + \beta_2 Latitude + \beta_1 Year$$

For a single value of $Latitude$, the first two terms constitute the intercept ($\beta_0 + \beta_2 Latitude$) and the last coefficient is the slope (β_1).

For the host lines, there are two models, one for each species of host. For *C. concatenata*:

$$Julian\ day = \beta_0 + \beta_1 Year + \beta_2 Latitude + \beta_3 + \beta_5 Year$$

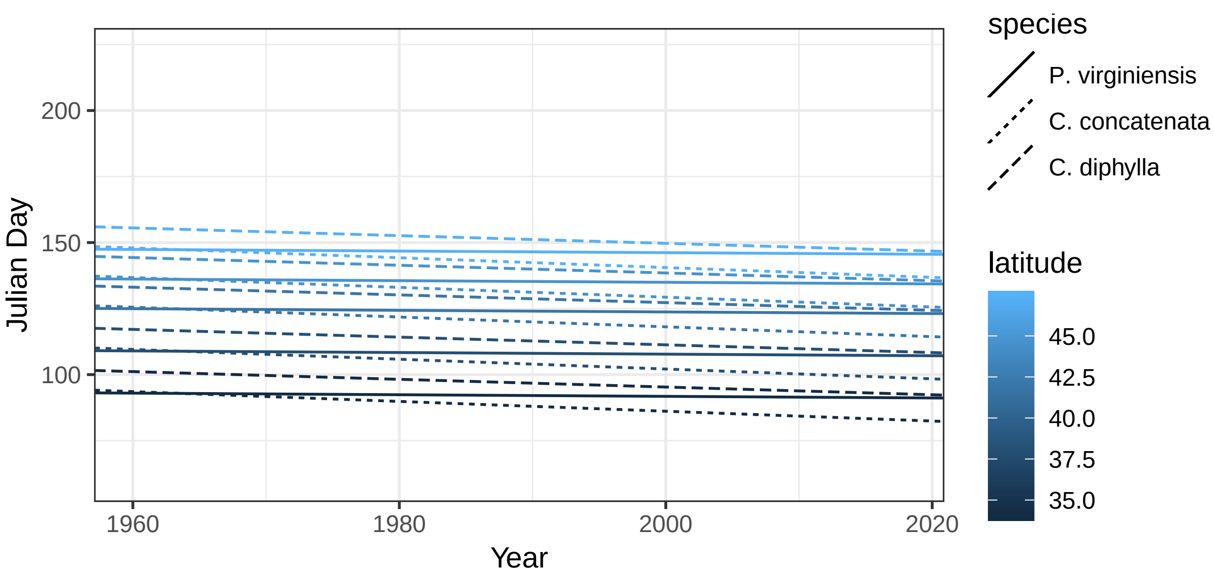
Rearranging:

$$Julian\ day = \beta_0 + \beta_3 + \beta_2 Latitude + (\beta_1 + \beta_5) Year$$

So for a single value of $Latitude$, the model for *C. concatenata* has an intercept of $\beta_0 + \beta_3 + \beta_2 Latitude$ and a slope of β_1 and β_5

Similarly, the model for *C. diphylla* can be simplified to an intercept of $\beta_0 + \beta_4 + \beta_2 Latitude$ and a slope of $\beta_1 + \beta_6$

Visualizing the model:



Considering the slopes, we see that controlling for latitude, observations are getting earlier, but the change is happening faster in the host plants than in *P. virginiensis*:

Species	Change
P. virginiensis	0.031 days/year
C. concatenata	0.186 days/year
C. diphylla	0.146 days/year

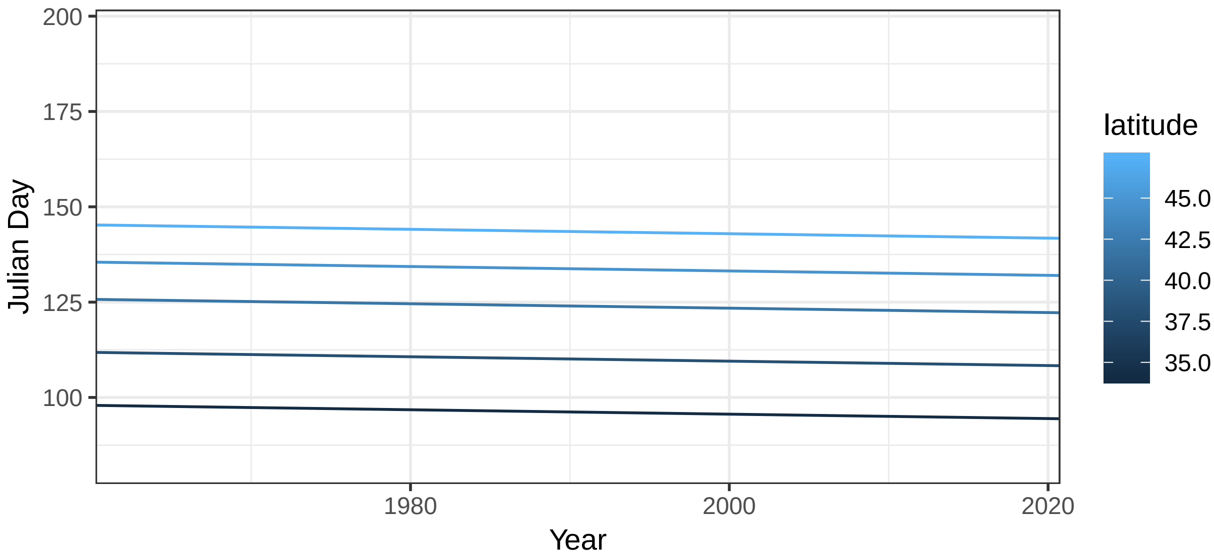
Separate analyses

There is some question about the sample size issues among the three taxa. Is the relationship between year and yday for *P. virginiensis* being heavily influenced by the considerably larger sample sizes of the host plants?

P. virginiensis data alone.

Consider the model

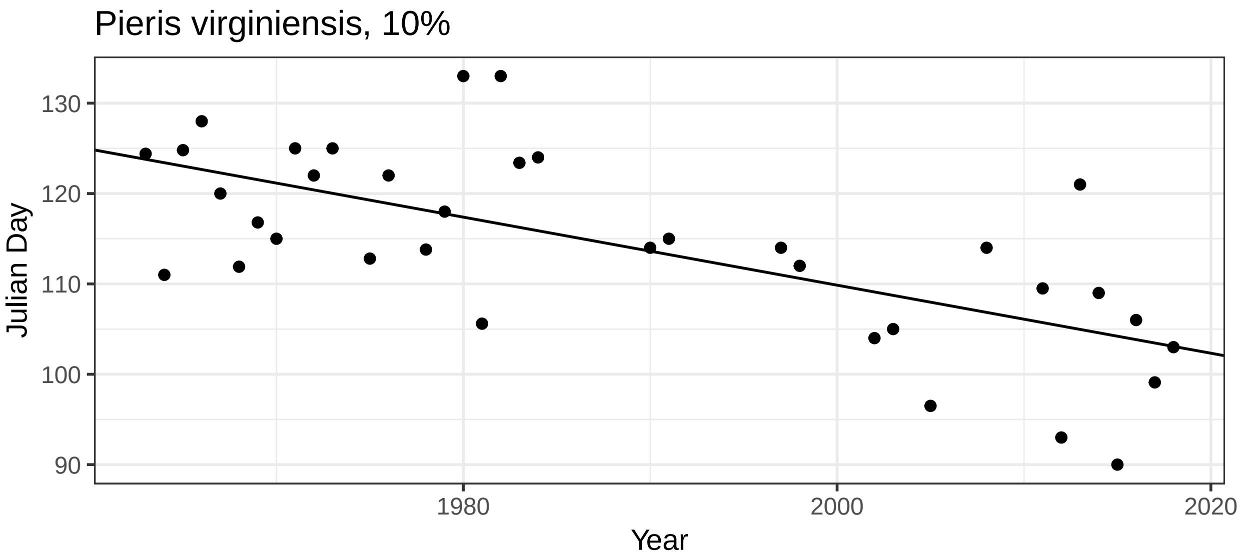
$$Julian\ day = \beta_0 + \beta_1 Year + \beta_2 Latitude$$



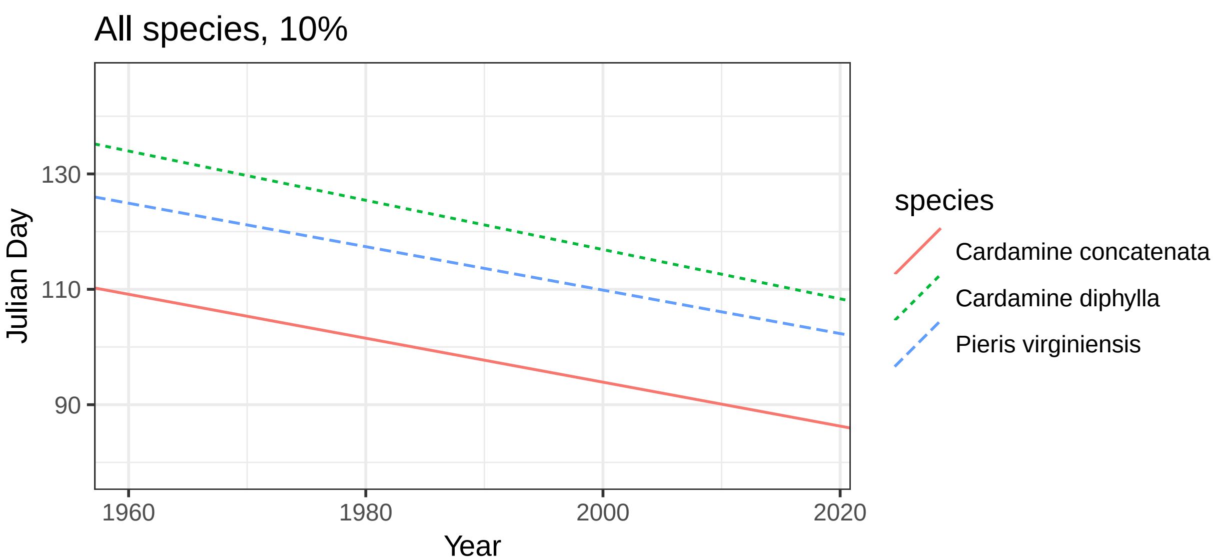
Controlling for latitude, the change in *P. virginiensis* is 0.058 days earlier per year. This is a larger change than estimated in the model that includes the host plants (0.031 days/year).

Interpolation

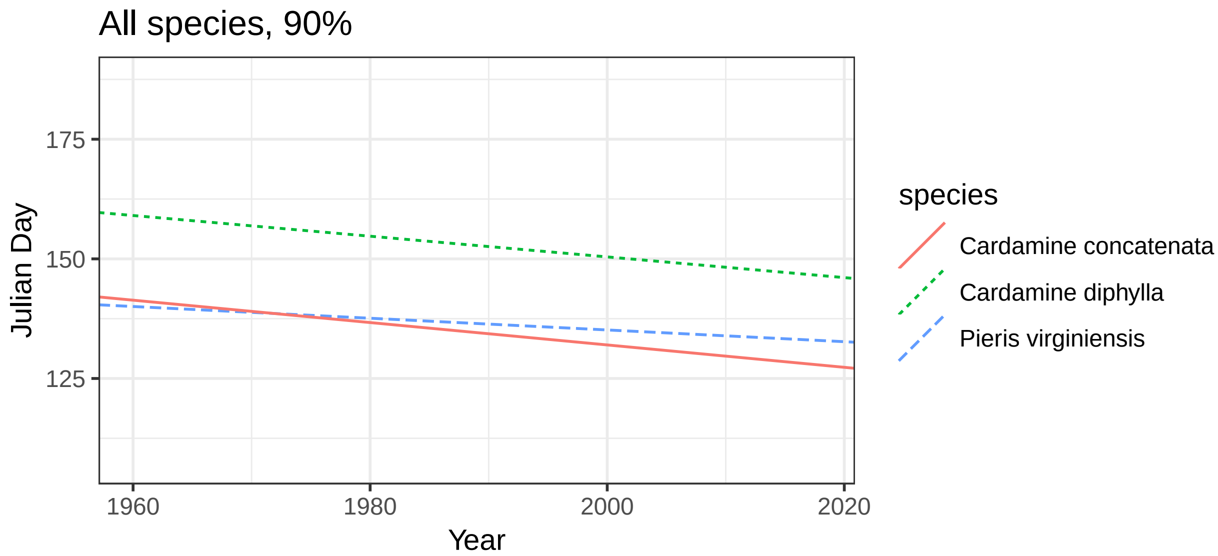
One approach used in other phenological studies uses interpolation to estimate the day on which 10% of the observations for that year occurred (and 50% and 90%). This approach should be less influenced by sample size. For example, we can look at the bottom 10% estimate for each year for *Pieris virginiensis*:



We can also investigate the same statistic (the day of they year on which 10% of the observations for that year were observed) for the two host plants.



Now consider the *end* of the season, with the 90% quantile.



Considering both 10% and 90% on the same plot: