Comparing Flora Flowering Data between iNaturalist and Experts Databases

Evelin Reyes

12/8/2021

[1]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA

email: ereye11@lsu.edu (mailto:ereye11@lsu.edu); 232 Life Science Building, Baton Rouge, LA 70803

Introduction

Climate is driving changes in how plants and animals experience their life history. Environmental changes and global climate change are a big dilemma due to their influence on many organisms. (David W. Inouye, 2008). Research projects like this are essential to understand better how species adapt to climate changes over long periods and help us predict how species will fare in the future. iNaturalist is a unique tool available to scientists that allow for accumulating large datasets without the need for challenging fieldwork. This project will compare the two datasets, one from iNaturalis and one from experts from 1967 to 2020. Data for the iNaturalist was obtained from worldwide observations. Thus, flowering time is measured annually; it can vary from year to year. (Inouye, 2008; Diekmann, 1996). Even though the flowering time is different from species to species, climate changes affect them. Also, the changes in temperature have been documented many times by various studies around the world (Menzel et al., 2006, Parmesan, 2007)

Materials and Methods

Data Collection

Data were collected by Dr. Daijiang Li through an open-source website called iNaturalist. https://www.inaturalist.org/ (https://www.inaturalist.org/). The data were collected from the global distribution and included 133 flowering plant species. The expert data were obtained from 66 flowering plant species.

Data Management

All the data were analyzed using R (R Core Team, 2021) For the t-test, I compared the overlap in the flowering time predicted by the expert data and the flowering time observed in the iNaturalist data. For this, I only considered species that we found in both datasets (i.e., 66 species).

For the linear mixed effect models I used all species in the iNaturalist data (i.e., 133 species) to compare whether there are changes in species flowering times over the years. This model used mean flowering day as the response variable and year as the fixed effect. I used species as random effects in our linear mixed effect models.

To analyze the temporal trends for each species individually, I ran linear regressions separately for each species (133) in the iNaturalist data, where the response variable was mean flowering day, and the predictor variable was the year.

Results

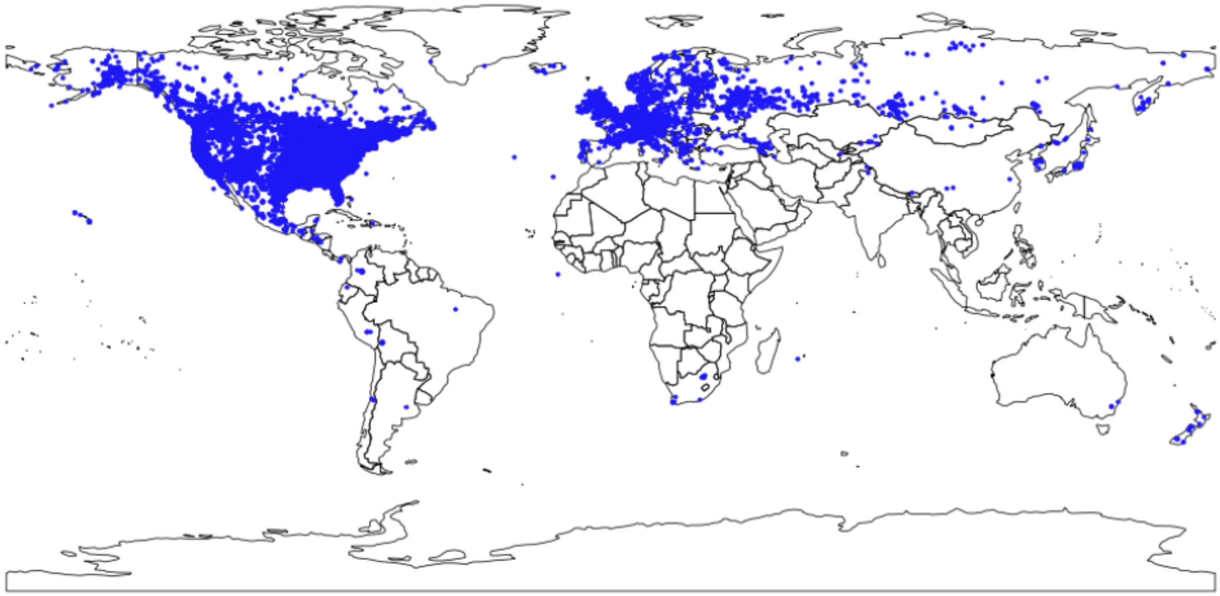

Figure 1.R was used to plot all the iNaturalist observations, showing a worldwide distribution (figure inaturalistdata.pdf). Overall, most of the flowering observations made in the inaturalist dataset were recorded during the summer months (i.e., from March to August), while fewer observations were made during the winter months (figure numberofobservations.pdf)

t-test

```
        Welch Two Sample t-test

data:  df.out[, 2] and df.out[, 5]
t = -10.047, df = 105.36, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.154433 -2.784961
sample estimates:
mean of x mean of y
 3.242424  6.712121
```

Figure 2. The t-test shows that the flowering season in the iNaturalist dataset (mean 6.7 months) is almost twice as large as the one described by the expert dataset (mean 3.2 months) ($t=-10.047$, $df=105.36$, $p<0.001$; figure comparison.pdf).

The linear mixed effect model

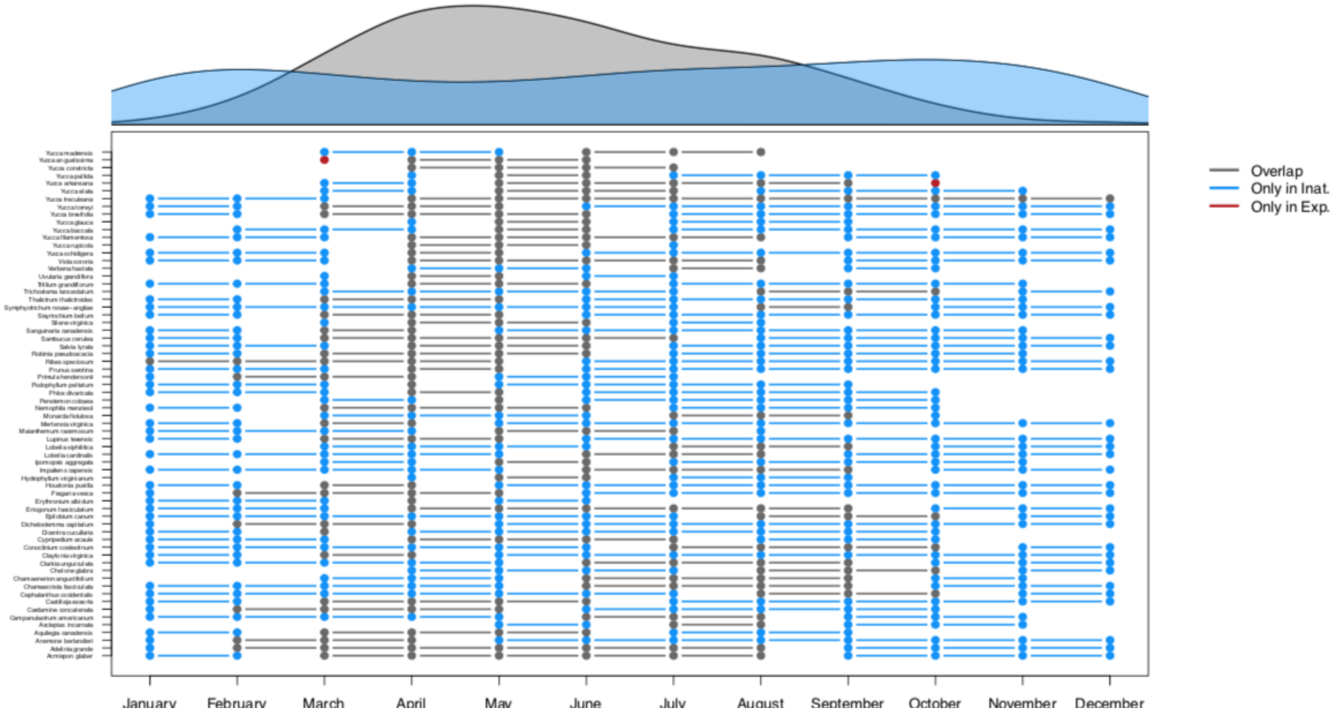

Figure 3. includes species from the iNaturalist and Expert datasets. The general temporal flowering trend is for the species to have their flowering time half a day earlier with each subsequent year (beta = -0.4616, t-value = -4.464, p<0.001). Thus, as the first collecting date was 1967, it means that by 2020 the species were flowering on average 25 days earlier than in 1967.

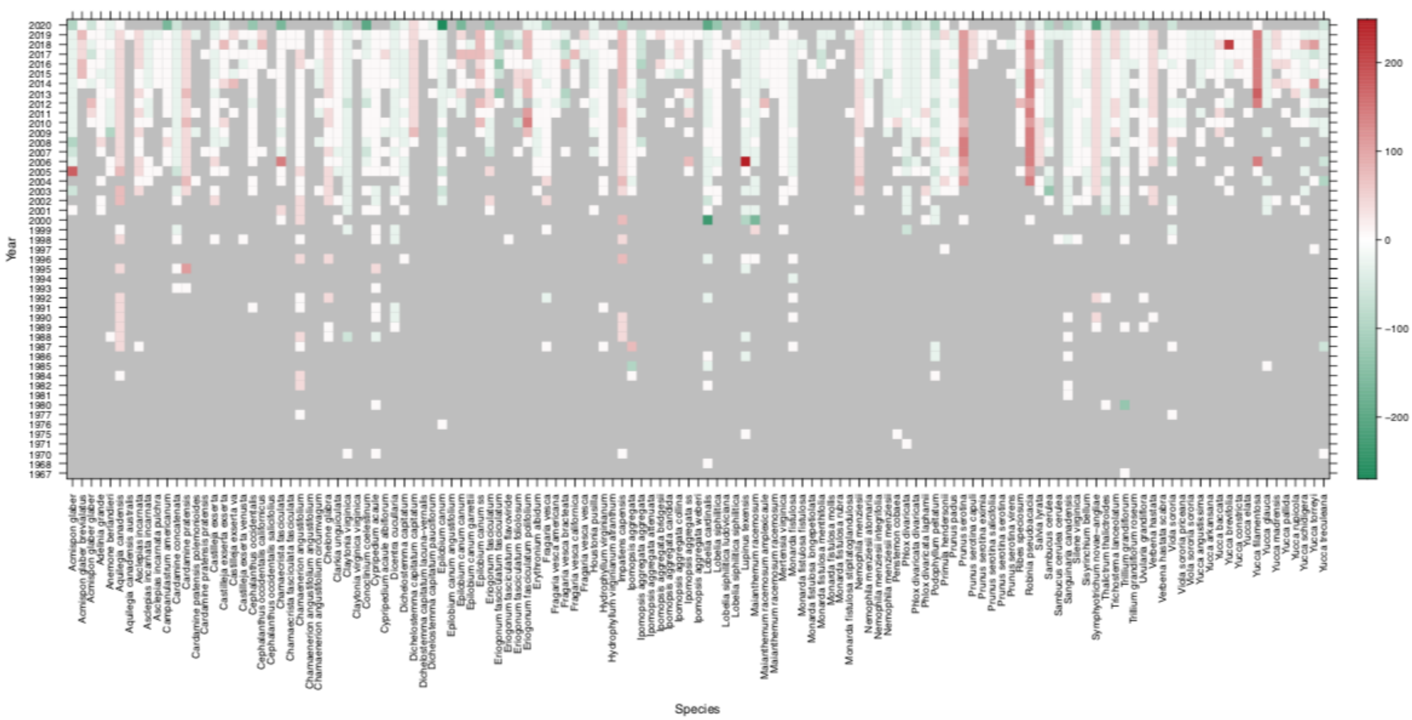

Figure 4. I analyzed the flowering time for 113 plant species over the length of the study period, from 1967 to 2020. Out of the 113 species studied, we found statistically significant results for 94 species, indicating that the flowering period has changed over time. For 54 species, it was observed that they are having their flowering time earlier, whereas 40 species are having their flowering time delayed.

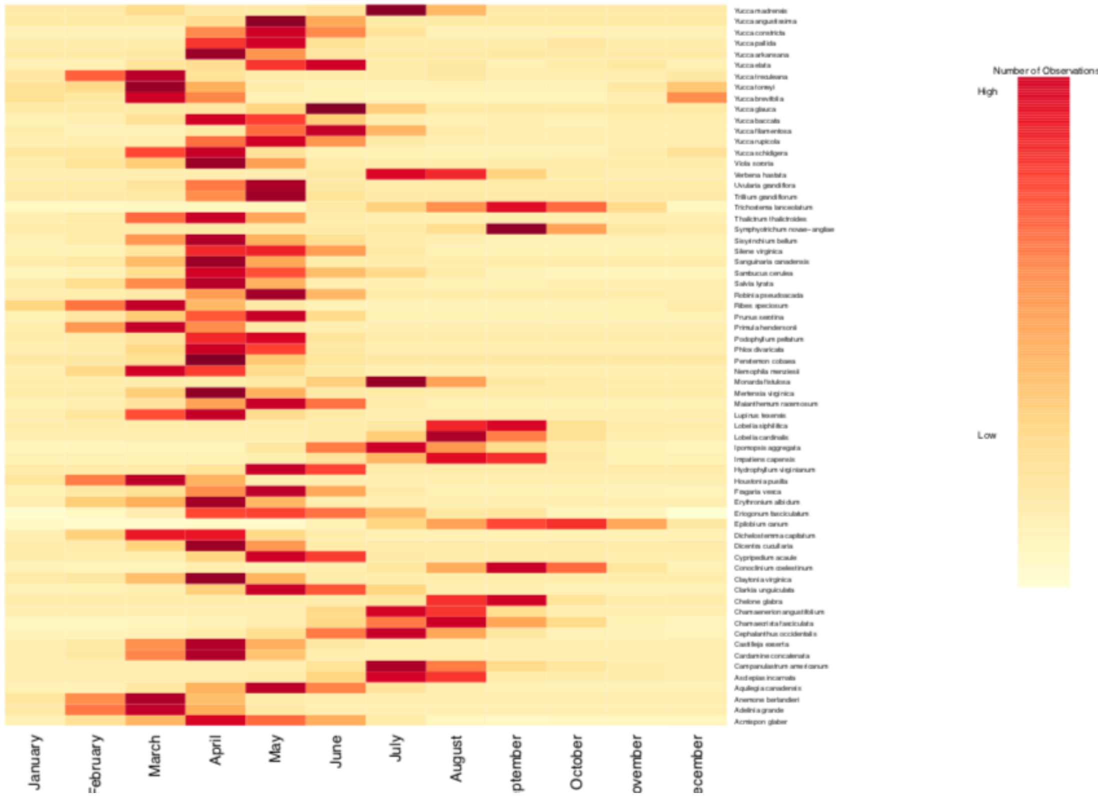

Figure 5. Number of observations, as predicted the flowering time happened during the months of april to september.

Discussion

Using a dataset from experts and iNaturlist, significant differences were found in the 53 species found in both datasets.

Of 53 species of flower observed during the period from 1967 to 2020 in 2 independent datasets, those of expert researchers versus publically-sourced iNaturalist contributions, a significant delay in the flowering time in 40 of these species was observed.

As a future implementation, I would like to run a new analysis and include different variables such as nutrients, soil, sun exposure, and greenhouse plants.

Conclusion

In summary, the present project showed the effect of climate change in flowering time in a worldwide range. For example, in 2020, the flowering times for most of the species happened 20 days earlier. Future work will include the total species in both datasets, iNaturalist, and Experts.

References

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (https://www.R-project.org/).

Menzel A, Sparks TH, Estrella N et al. (2006) European phenological response to climate change matches the warming pattern. Global Change Biology, 12, 1969–1976.

Parmesan C (2007) Influences of species, latitudes, and methodologies on estimates of phenological response to global warming. Global Change Biology, 13, 1860–1872.

Inouye, D.W., 2008 Effects of climate change on phenology, frost damage and floral abundance of montane wildflowers. Ecology 89, 353-362.