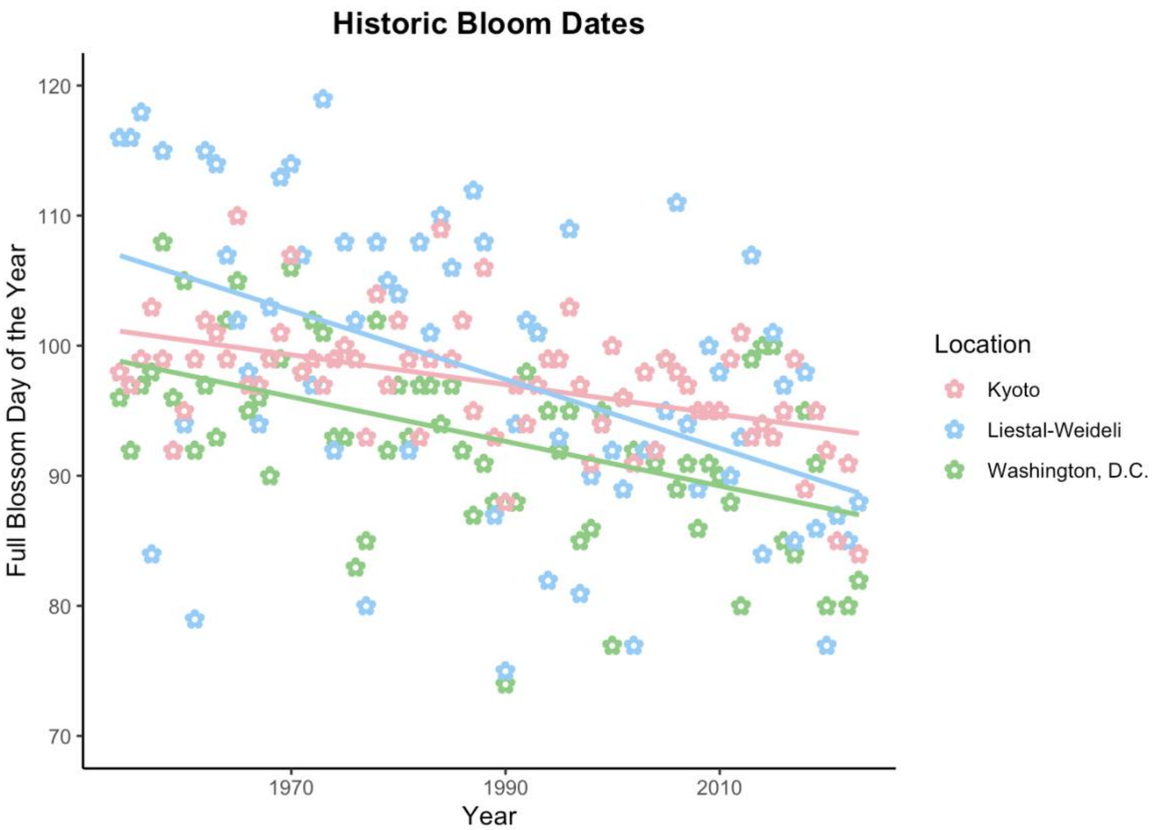
Cherry Blossom Prediction Contest: Narrative

Many people still view climate change as something distant and abstract which they won't come across in their daily lives (McDonald et al., 2015). However, there are many changes which can be noticed without complicated climate models—for example the fact that cherry trees tend to bloom earlier every year. Figure 1 shows the dates on which cherry trees bloomed during the past 100 years in three different cities around the world. As one can see, full blossom tends to occur earlier each year. Regarding these three cities, this is most pronounced for Liestal-Weideli in Switzerland, where the full bloom date prematures on average by 0.19 days every year.

Figure 1

Historic Bloom Dates in Kyoto, Liestal-Weideli, and Washington, D.C.



Prediction Strategy

To predict this year's bloom dates, I have chosen a multiple linear regression approach with four predictors: Year, mean January temperature, the estimated dormancy release date (DRD), and the modeled temperature at dormancy release. While the regression approach seems fairly simple, I compared its fit indices with those of a random forest approach, using leave-one-out cross validation as well as (for the random forest model) out-of-bag error. The multiple regression model performed better at all instances. Furthermore, I have put a lot of effort into modeling the yearly dormancy release dates and temperature. After searching for some literature on cherry trees, I believe this to be an important predictor of bloom dates.

Dormancy Release Date (DRD)

Dormancy release denotes the time in late Winter or early Spring when the temperature trend reverses from going down to getting warmer again. Since cherry blossoms need both periods of cold and warmer temperatures, the dormancy release date can be an important predictor of the cherry bloom date (Chung et al., 2011). To estimate past DRDs, I used the GHCNd temperature data and fitted quadratic regressions to each year's temperature data between October 1st and March 31st. I then took the coefficients b_i of this regression and put them into a quadratic function of temperature values[1]:

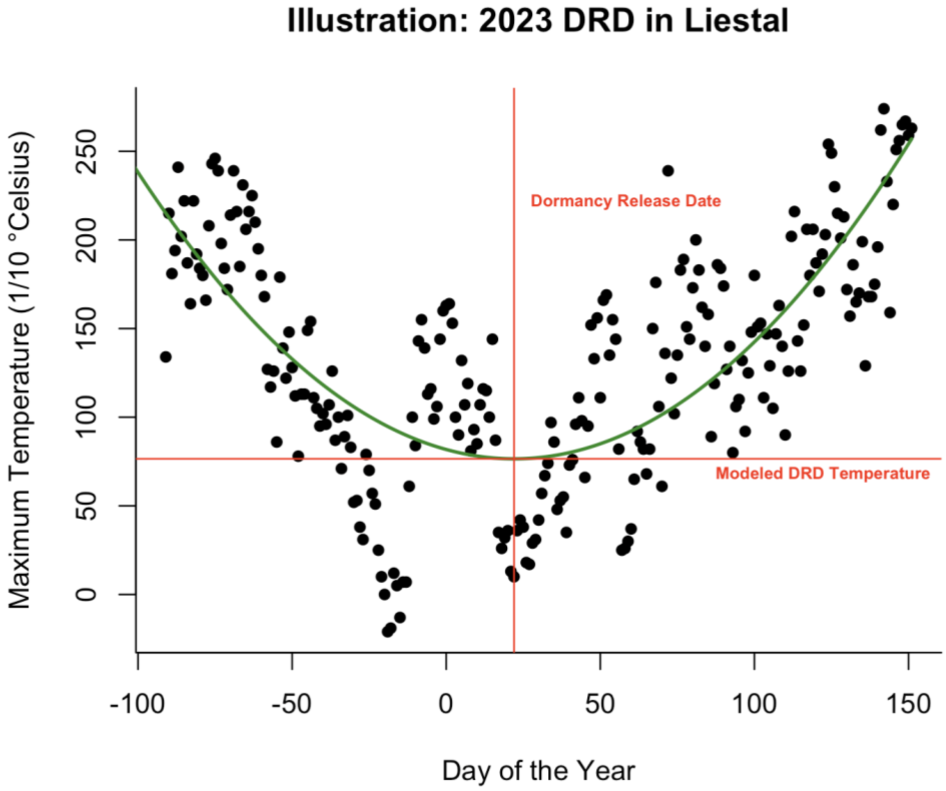
$f_{year}(temp) = b_0 + b_1 * temp + b_2 * temp^2$
$\widehat{DRD}_{\text{year}} = \min(f_{year}(temp))$

I then repeated this for every year for which I have temperature values. The minimum of this parabola indicates the modeled dormancy release date and temperature on this day. Figure 2 illustrates the estimation of the 2023 dormancy release date for Liestal.

[1] Please excuse if the notation is not formally correct since I have no background in mathematics.

Figure 2

2023 Dormancy Release Date in Liestal, Switzerland

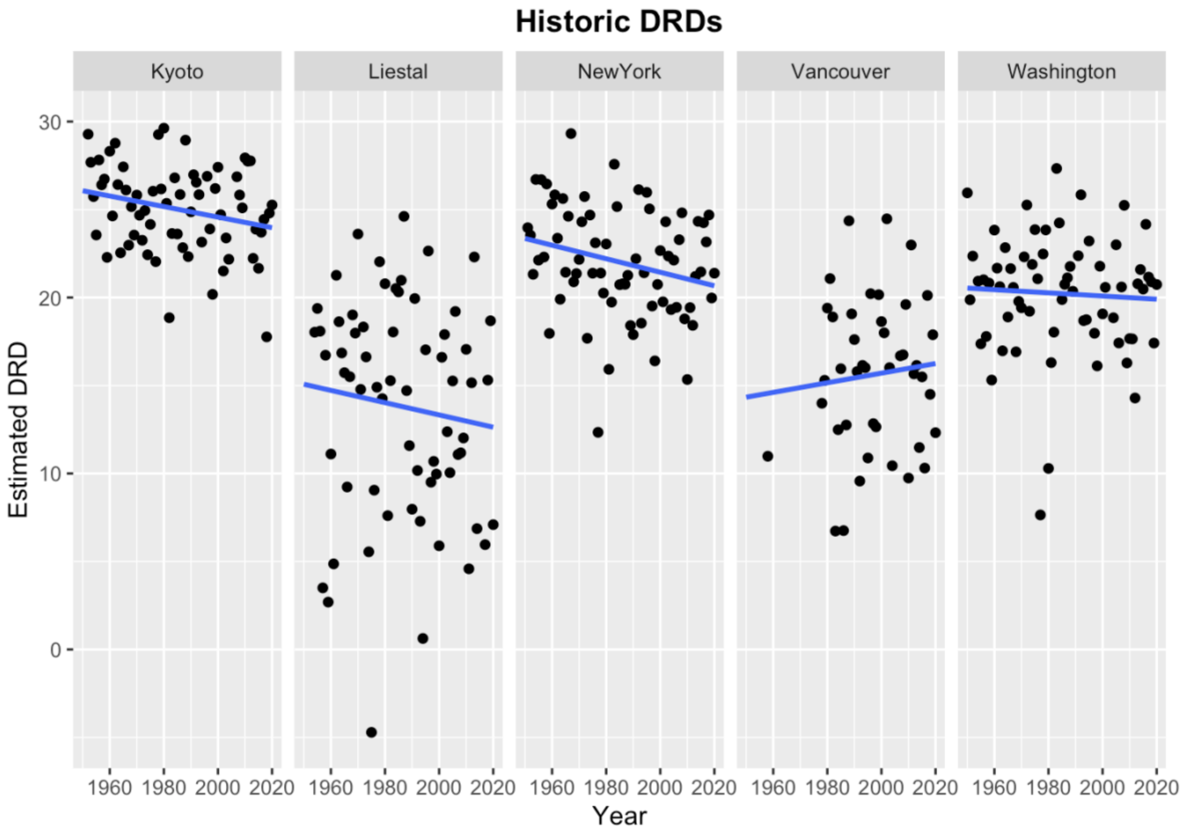


Analyses showed that DRD is, in fact, a very important predictor of full bloom dates. For Kyoto, DRD was the most informative predictor in my model (meaning it was more informative than mean January temperature and year), explaining 40% of variance in bloom dates, and for Liestal the modeled DRD temperature was most informative, explaining even 57% of variance. This supports the notion that (a) DRD is important to consider when predicting cherry bloom dates and (b) that my way to model DRDs was successful.

It is also interesting to see that the Dormancy Release Dates which I modeled from past temperature data tend to become smaller every year (see figure 3). This means that, on average, the beginning of a positive temperature trend after winter gets earlier every year. This fits to the trend of earlier-blooming cherry trees.

Figure 3

Historic dormancy release dates modeled using past temperature data



In Vancouver, this trend seems to be reversed. However, this most likely is due to missing DRD data for early years as well as to some influential data points (see figure 3, Vancouver).

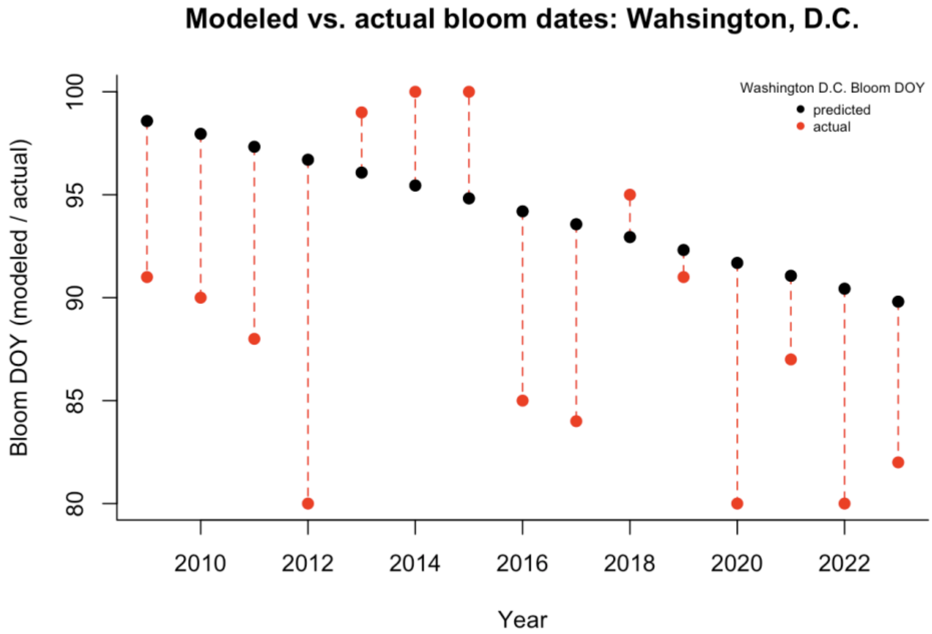
To calculate this year's DRD (which is a key value for this year's prediction), I needed more current temperature data which I retrieved from https://www.visualcrossing.com. However, it became clear that the weather data until late February would not be enough to adequately estimate this year's DRD. Therefore, I modeled 2024 temperature data between March 1st and June 1st by taking the mean temperatures in this time span from the past 10 years and included them into the model as well. To ensure that the quadratic regression does not depend too strongly on these imputed dates, I weighted them less than the actual temperature data.

USA NPN Dataset: Estimating Vancouver and New York City Bloom Dates

Estimating the Vancouver and New York City bloom date was more challenging since there are no (or only very little) records of past bloom dates. I thus used data from the USA NPN dataset to predict historic Vancouver bloom dates, using year, latitude, longitude, elevation, average maximum winter temperature, and species as predictorsI evaluated this procedure by reproducing past Washington bloom dates (of which I have the actual bloom dates). This is illustrated in figure 3.

Figure 3

Actual vs. predicted bloom dates in Washington, D.C. as a function of year



On average, the estimation was off by around seven days which is not ideal. However, in my opinion this was still the best way to go. I again compared the performance to a random forest model, which showed worse fit indices. I therefore predicted historic bloom dates in Vancouver and New York City using this method and then used the same procedure for this year's prediction as I used for the other locations.

Prediction Results

Liestal: March 31 (91)
Washington: March, 26 (86)
Kyoto: April, 2 (93)
New York City: April, 2 (93
Vancouver: April, 6 (97)

Literature

Chung, U., Mack, L., Yun, J. I., & Kim, S. H. (2011). Predicting the timing of cherry blossoms in Washington, DC and mid-Atlantic states in response to climate change. *PloS one*, *6*(11), e27439.

McDonald, R. I., Chai, H. Y., & Newell, B. R. (2015). Personal experience and the ‘psychological distance’of climate change: An integrative review. *Journal of environmental psychology*, *44*, 109-118.