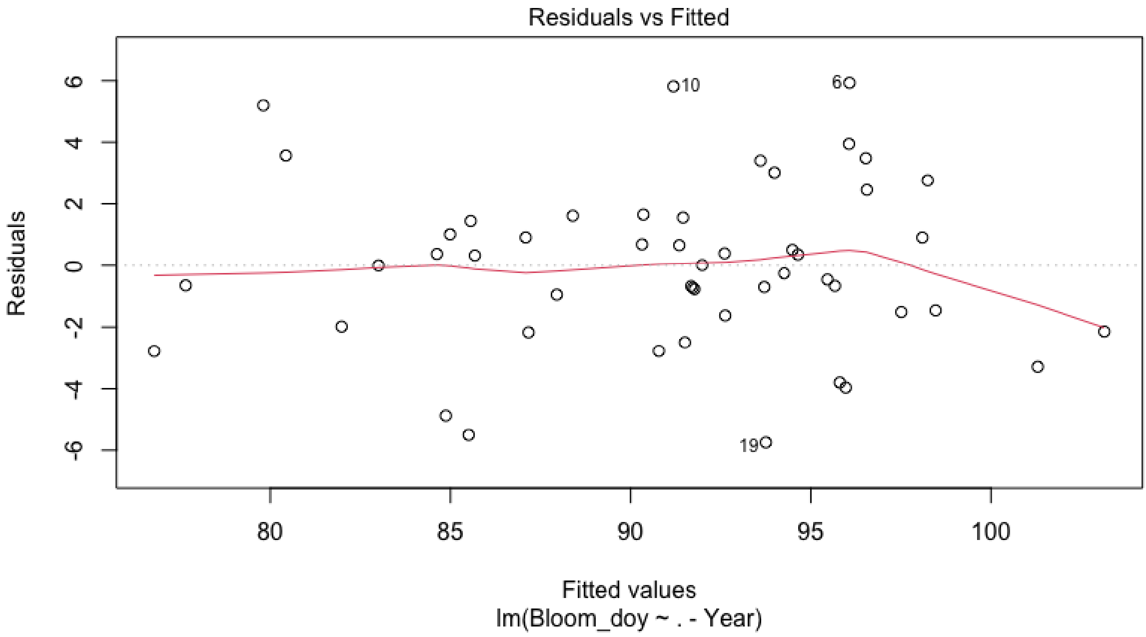
**Initial model assumptions**

I expected that natural cycles like the phenology of cherry trees would have a linear relationship with temperature or growing degree days. This turned out not to be the case when I tested for the conditions of linear regression in predicting the 2023 cherry blossom dates for Kyoto, Liestal, Vancouver, and Washington D.C. The residual plot below (for Washington D.C.) was enough evidence to look beyond linear regression to make predictions.




**Data**

I relied on two sources for my data: the data from the competition website, and `chillR`, an R package for phenology analysis, created by Eike Luedeling.

`chillR` allowed me to quickly search for the closest weather station to a given latitude and longitude, retrieve historical weather data for that location, and format the data for easy model ingestion.

I used the competition website data—more specifically, the bloom day-of-year for each location—as the endpoint for hourly weather data from January 1st (retrieved by `chillR`). So now I had weather data for all my locations, except for Vancouver. I had to get creative in finding historical bloom dates for Vancouver.

I went on the photo-sharing website, flicker.com, and searched for images of cherry blossoms in Vancouver as far back as possible up until 2022. I used the camera metadata (which is available for most of the uploaded images) instead of the upload date as the most accurate measure of what the cherry blossoms looked like at the time. (If the metadata weren't available, which was rarely the case, I would use the upload date.)

The earliest upload on flicker.com of cherry blossoms in Vancouver was on April 8th, 2002. For each year, from 2002 to 2022, I collected images—their corresponding dates and metadata—of cherry trees in full bloom.

vancouversun.com was another source I came across (through a Google search) to get images of historical cherry blossoms in Vancouver. I right-clicked on an image on vancouversun.com to bring up the web inspector toolbar to access a particular image's metadata (that included the date the image was taken as well as the image's source link). I found images that overlapped with the images on flicker.com. vancouversun.com extended my historical period by three years; I found images from 1999 to 2001. (I actually found images of cherry blossoms on April 6th, 1960 and April 10th, 1962.)

I put the dates from the images I found on flicker.com and vancouversun.com into an Excel spreadsheet in the same format as the Vancouver competition data set. For each year, from 1999 to 2022, I then calculated:

- the difference (in number of days) between the dates of the first and last images uploaded, and
- the median day of the year of that period as the bloom day-of-year.

This became the Vancouver data set I ended up using for my predictions. Instead of a single data point (Vancouver's 2022 bloom day-of-year), I now had 23 data points to model my predictions.

## Models

### *Neural network*

I trained a deep neural network to predict bloom day-of-year with four inputs:

- *Chilling hours:* The cumulative hours since January 1st of a year up until the bloom day-of-year that the temperature was between 0 ºC and 7 ºC
- *Utah model:* Negative chill accumulation is where temperatures are greater than 16 ºC
- *Dynamic model*: Modeling internal chilling (chill portions), or endodormancy
- *Growing degree hours (GDH)*: Heat accumulation from an organism's base temperature until an optimal temperature. (In the context of this competition the optimal temperature would be when the cherry trees will bloom.)

Predicting future input values to use in predicting bloom day-of-year posed a challenge. I considered time series forecasting as an alternative because the data for all locations are at yearly intervals.

*ARIMA*

The Arima model predicted a flat bloom day-of-year across all locations for the next 10 years. This, I suspect, was due to the data being stationary/not seasonal; there was no trend for the Arima model. I wanted something between the limitations I faced with a neural network and the flatline prediction of the Arima model.

*General Regression Neural Network*

The time series forecasting with generalized neural network (tsfgnn) package by Francisco Martinez, Maria P. Frias, Antonio Conde, Ana M. Martinez allowed me to efficiently forecast evenly-spaced time interval data (time series) with a neural network. I got more nuanced predictions than with the Arima model and used only the competition data sets. This means that I made my final bloom day-of-year predictions by only using the historical bloom day-of-year data from the different locations as input into my general regression neural network (GRNN).

**Results**

| year | kyoto | liestal | washingtondc | vancouver |
|---|---|---|---|---|
| 2023 | 90 | 87 | 88 | 90 |
| 2024 | 90 | 86 | 90 | 91 |
| 2025 | 90 | 86 | 91 | 91 |
| 2026 | 90 | 86 | 87 | 91 |
| 2027 | 90 | 86 | 91 | 90 |
| 2028 | 90 | 85 | 88 | 90 |
| 2029 | 90 | 85 | 87 | 90 |
| 2030 | 89 | 84 | 88 | 90 |
| 2031 | 89 | 84 | 88 | 90 |
| 2032 | 88 | 83 | 87 | 90 |

The GRNN had the following mean absolute error (MAE) metrics for each location:

- Kyoto: 4.14
- Liestal: 8.09
- Washington D.C.: 6.55
- Vancouver: 5.69

**Further studies**

A further study could predict chilling hours, negative chill accumulation, chill portions, and growing degree hours, and then use these outcomes to predict future cherry bloom days.

**References**

Fearon, Dave. *263/365*. photo, 20 Sept. 2014. *Flickr*, https://www.flickr.com/photos/davefearon/15118079089/.

Luedeling, Eike. *Tree Phenology Analysis with R*. *inresgb-lehre.iaas.uni-bonn.de*, https://inresgb-lehre.iaas.uni-bonn.de/chillR_book/introduction.html. Accessed 23 Feb. 2023.

Luedeling, Eike, and Anja Gassner. 'Partial Least Squares Regression for Analyzing Walnut Phenology in California'. *Agricultural and Forest Meteorology*, vol. 158–159, June 2012, pp. 43–52. *DOI.org (Crossref)*, https://doi.org/10.1016/j.agrformet.2011.10.020.

Martínez, Francisco, et al. 'Strategies for Time Series Forecasting with Generalized Regression Neural Networks'. *Neurocomputing*, vol. 491, June 2022, pp. 509–21. *ScienceDirect*, https://doi.org/10.1016/j.neucom.2021.12.028.

*Multiple Input Multiple Output Strategy*. *R-Packages*, https://cran.r-project.org/web/packages/tsfgrnn/vignettes/tsfgrnn.html. Accessed 27 Feb. 2023.

'Photos: A History of Cherry Blossoms in Vancouver'. *Vancouversun*, https://vancouversun.com/gallery/photos-a-history-of-cherry-blossoms-in-vancouver. Accessed 30 Jan. 2023.

Río, Francisco Martínez del. *Tsfgrnn*. 2019. 14 Feb. 2023. *GitHub*, https://github.com/franciscomartinezdelrio/tsfgrnn.