**Forecasting Peak Bloom Dates for Cherry Blossom Trees**

Jonathan Abbamonte, Priyanjali Bukke

STAT 634 - Case Study 1

March 22, 2022

**Introduction**

The blooming of cherry trees is one of the most celebrated phenological events. Every year, millions of people in Japan take part in *hanami* (“flower viewing”) festivals, parties, and picnics to view and enjoy the beauty of the blossoms of ornamental cherry trees (*sakura*) during the short 1–2 week period in early-Spring that they are in bloom. The tradition of *hanami* in Japan dates back to at least the Heian period (8th to 12th centuries)[[1]](#endnote-1) and is a deeply-entrenched part of Japanese culture. Outside of Japan, the blooming of cherry trees also attracts thousands of visitors annually in cherry tree hotspots where *hanami*-inspired festivals are heldsuch as Washington, D.C.; Vancouver, Canada; Macon, Georgia; Stockholm, Sweden; and the Brooklyn Botanic Garden in New York City.

As cherry tree blossoms are in bloom for only a short period of time, being able to anticipate the precise timing of blossom blooming is important for planning *hanami*, travel, and cherry-blossom-related events. Of particular interest is being able to predict the timing of peak bloom, when most of the blossoms have fully bloomed, as this is when the trees typically appear most aesthetically pleasing to view.

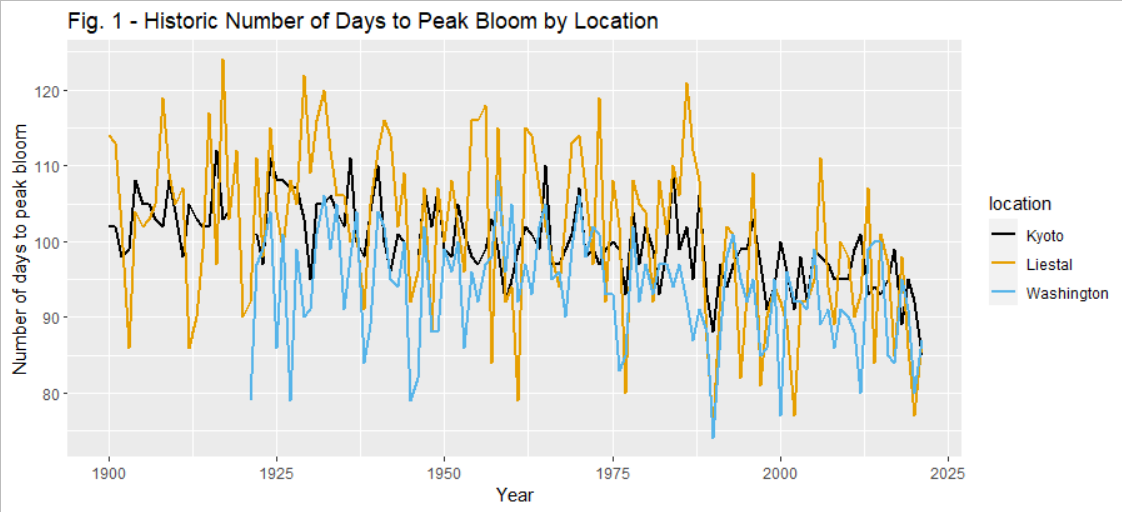
It is widely believed that the timing of cherry blossom blooming is closely tied to climatic factors, particularly temperature. Many studies have shown that pre-bloom temperature is significantly associated with cherry blossom blooming, especially during February and March.[[2]](#endnote-2),[[3]](#endnote-3),7,10 However, temperature alone does not fully account for the timing of cherry tree blooming.[[4]](#endnote-4),[[5]](#endnote-5),[[6]](#endnote-6),7 Research on cherry blossom bloom timing has suggested that other potentially influential factors may include the length of photoperiod (daylight) the trees are exposed to,[[7]](#endnote-7),[[8]](#endnote-8),[[9]](#endnote-9) windspeed,[[10]](#endnote-10) precipitation,[[11]](#endnote-11) urban heat island effect,10 and solar radiation.10 In extension, these factors and their interactions may be important predictors for the timing of peak bloom.

In this report, we have used statistical modeling to predict the timing of peak bloom in four locations: Washington, D.C. (USA), Kyoto (Japan), Liestal-Weideli (Switzerland), and Vancouver, B.C. (Canada). Multiple linear regression and multiple linear regression with seasonally-adjusted autoregressive integrated moving average (SARIMA) errors were used to model historic peak bloom dates for each of the four locations. Forecasted peak bloom dates for 2022-2032 were additionally extrapolated from these models. Time will tell how accurate our models are able to forecast peak bloom dates.

**Data**

Data for modeling the peak bloom date were obtained through a variety of sources. The covariates for which data described in this section were collected for are explained in further detail in the Methodology section. Historic data for peak bloom dates, the response variable for our analysis, were obtained through a GitHub repository for the Cherry Blossom Prediction Competition hosted by George Mason University. Historic peak bloom data dates back to A.D. 812 for Kyoto, 1894 for Liestal, and 1921 for Washington, D.C. No historical data for the peak bloom date were available for Vancouver, B.C. The definition of peak bloom differed for each site. For Kyoto, peak bloom is defined as the day by which 80% of the cherry blossoms are in full bloom. In D.C. and Vancouver, the definition of peak bloom is 70% of the cherry trees and in Liestal, 25% of the trees. In statistical modeling, peak bloom dates were weighted, with 70% bloom set as the baseline. Past data indicates that the time from first bloom to peak bloom for cherry blossoms typically lasts about one week.[[12]](#endnote-12),[[13]](#endnote-13) We assumed that cherry blossoms progressed from first bloom to peak bloom linearly, so weights were assigned to historic bloom dates based the percentage of linear progression to (or from) 70% bloom. Table 1 shows the cherry blossom cultivar observed in each of the four locations. Note that the cultivar for which peak bloom was observed was different for each site. Due to a lack of replicate data on bloom times for cherry cultivars in different locations, we made the assumption in our models that there was no variation in bloom times (all other factors being held equal) between the different cherry cultivars shown in Table 1. Fig. 1 shows a time series of historic data for number of days of the year to peak bloom for each of the three locations in our analysis where PBD data was available.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1**. Locations and Cultivars of Cherry Trees | | | | |
| Location | Latitude (°) | Longitude (°) | Altitude (m) | Cherry Blossom Cultivar |
| Kyoto, Japan | 35.0120 | 135.6761 | 44 | *Prunus jamasakura* |
| Liestal-Weideli, Switzerland | 47.4814 | 7.730519 | 350 | *Prunus avium* |
| Washington, D.C., USA | 38.8853 | -77.0386 | 0 | *Prunus* × *yedoensis* ‘Somei-yoshino’ |
| Vancouver, B.C., Canada | 49.2237 | -123.1636 | 24 | *Prunus* × *yedoensis* ‘Akebono’ |



Data for all covariates was collected from their sources at either daily sampling intervals or monthly sampling intervals. Daily/monthly data was subsequently translated into annual covariate data. It was necessary to transform daily/monthly data into annual data because the response variable for our analysis (annual peak bloom date) is an annual metric.

Climatic variable data for potential covariates were acquired for all four locations from publicly accessible data from the U.S. National Oceanic and Atmospheric Administration’s (NOAA) Global Historical Climatology Network – Daily (GHCN-Daily), Version 3,[[14]](#endnote-14) and the NOAA’s Global Surface Summary of the Day (GSOD).[[15]](#endnote-15) Data was acquired from land-based weather stations that were closest to the approximate coordinates for where peak bloom date data was collected. The approximate latitude, longitude, and altitude for where peak bloom date data was measured is shown in Table 1. The weather stations from which climatic data were acquired are: Kyoto, JA (GHCN:JA000047759, GSOD: 47759099999) for Kyoto, Japan; Reagan National Airport (GHCN:USW00013743, GSOD:72405013743) for Washington, D.C.; and Basal-Binningen, SZ (GHCN:SZ000001940, GSOD:6601099999) for Liestal-Weideli, Switzerland. For Vancouver, due to inconsistent data collection from weather stations and excessive missing values, climatic data was acquired from three different weather stations within two miles of one another: Vancouver International Airport (GHCN station: CA001108473) for data from 1937-2013, Vancouver International Airport (GHCN station: CA001108395) for data from 2014-2015, and Vancouver Sea Island (GHCN station: CA001108380) for data from 2016-2021.

Daily data for precipitation, snow depth, minimum temperature, maximum temperature, average temperature, and percent of possible sunshine were acquired from the GHCN-Daily (hereinafter “GHCN”). Daily data for wind speed and sea-level air pressure were acquired from GSOD. These data were subsequently converted into usable potential covariates for our models. The covariates derived from GHCN data that were used in model selection were: cumulative precipitation (sum of daily precipitation from Jan. 1st – peak bloom date (PBD)), snow depth (sum of daily snow depth from Jan. 1st – PBD), cumulative minimum temperature (TMIN) (sum of daily minimum temperature from Jan. 1st – PBD), cumulative maximum temperature (TMAX) (sum of daily maximum temperature from Jan. 1st – PBD), average of cumulative minimum and maximum temperature (TAVG) (mean of annual TMIN and annual TMAX), percent of possible sunshine (sum of daily percent of possible sunshine from Jan. 1st – PBD), and cumulative temperature difference (). The covariates derived from GSOD data that were used in model selection were: average windspeed (average daily windspeed from Jan. 1st – PBD) and average sea-level air pressure (SLP) (average daily sea-level pressure from Jan. 1st – PBD).

Additional climatic variable data for Kyoto, Japan were acquired from the Japan Meteorological Agency, Tables of Monthly Climate Statistics (WMO station ID: 47759). Data taken from this source included monthly data on sea-level air pressure, total precipitation, mean percentage of possible sunshine, and total number of hours of sunshine.

Data was additionally collected on the total number of minutes of sunlight from the National Oceanic Atmospheric Administration’s (NOAA) solar calculator.[[16]](#endnote-16) This data was converted into average number of hours of sunlight from January 1st – April 1st from the year corresponding to the year of PBD, and from May 1st – September 30th of the year prior to the corresponding PBD year.

Data was collected on plant hardiness zones for each of the four locations using the Plantmaps website.[[17]](#endnote-17)

Data for growing degree days (GDD) were calculated from GHCN daily minimum temperatures and maximum temperatures. GHCN daily minimum and maximum temperatures were converted into hourly temperatures using the make\_hourly\_temps() function in the chillR package in R. GDD was subsequently calculated from hourly temperatures using the GDD() function in the chillR package.

For estimating number of chilling days and heating days (see Covariates section for further explanation of these terms), daily temperature data from GSOD was used.

Missing data was handled in a variety of different ways depending on the length of missing value runs, the number of missing values, and the nature of the distribution of the known data. Missing data was also handled at two levels of the data processing: at the daily/monthly level and at the annual level.

At the daily sampling level, where there were only a relatively limited number of missing values (for each year, less than 25% of the data values from Jan 1st to PBD), linear interpolation was used to fill in missing values. Linear interpolation of missing values was used particularly for imputing missing values for daily minimum temperature, daily maximum temperature, daily average temperature, daily sea-level air pressure, daily precipitation, daily wind speed, daily snow depth, and daily percent of possible sunshine. In cases where there were longer stretches of missing data (i.e. months or years of missing daily data) or in cases where the number of missing values exceeded more than 25% of data values for a given year, the annual covariate data for that year was set to missing at the annual data processing level. For estimating the number of chilling days and heating days, missing daily temperature data from GSOD was imputed with data from nearby weather stations. Data from nearby weather stations was obtained from the Ogimet webpage (obtained through the climate package in R).[[18]](#endnote-18)

At the annual data processing level, attempts were made to impute data for years with missing values where it was reasonable to do so. If the trend of annual data for a covariate approximated a stationary time series with constant random variation and if the distribution of the data was approximately normally distributed, we took the empirical mean and standard deviation of the data and imputed missing values by randomly sampling from a normal distribution with the same mean and standard deviation as the data. When the trend of the covariate time series was relatively flat and missing value runs were short but the distribution of the data was not normal, we randomly sampled data for missing values from the distribution of the known data. When a time series of the annual covariate data displayed a clear trend or clear moving average and when there did not appear to be a change in variation about the series mean over time, missing annual values were imputed either only over sections of the series that were approximately stationary or imputation was made through forecasted values from a SARIMA model fit to a time series of the covariate data.

When gaps in missing data were serious and lengthy (i.e. missing value runs of several years at a time) and a time series of the covariate data was highly stochastic or showed signs of unpredictable random walks, moving averages, or nonconstant variance, missing values were left in the annual data. If there were only a couple of years of missing data for the covariate, a model with casewise deletion was sometime run with the variable if the variable was deemed highly important to prediction. More often, however, variables with missing annual data had to be dropped from the analysis entirely due to poor data quality.

**Methodology**

**Covariates**

Several covariates were considered for statistical modeling of the day of peak bloom. To begin, we attempted to explore temperature as a predictor based on the chill days model developed by Cesaraccario, et al. (2004)5 and modifications to the model by Chung, et al.(2011).4 This two-step phenology model proposes that buds enter a dormancy phase typically on a set date in the fall, when the temperature goes below a threshold temperature (Tc), and accumulate chill days (Dc) until a chill requirement (Rc) is satisfied. Once this chill requirement is satisfied, the buds begin to develop with accumulation of anti-chill days (Dh) and may reach peak bloom once a heat requirement (Rh) is met. The daily chill days or anti-chill days values were calculated based on daily maximum, minimum, and average temperatures as described by Cesaraccario, et al.(2004).5 The optimal dormancy initiation dates and parameters to use to predict peak bloom dates for each location (with historical peak bloom date data available) were identified as those which minimized root mean square error between observed and predicted peak bloom dates. The sum of chilling days and heating days coincided with a prediction for the day of peak bloom, a potential covariate we are calling the ‘optimized estimate for peak bloom date based on the sum of heating days and chilling days.’

As an alternate metric for temperature accumulation, growing degree days (GDD) were computed using the GDD() function from the chillR package in R which uses the model for calculating GDD from Luedeling et al. (2013).[[19]](#endnote-19) The function uses the average hourly temperatures to compute daily GDD values as follows

where refers to the sum of the hourly temperatures per day, refers to the base temperature above which growing degree days accrue, and is an indicator function.[[20]](#endnote-20) The GDD for a given day was zero when the base temperature was greater than the average of the hourly temperatures. Values for were computed using the make\_hourly\_temps() function in the chillR package in R from hourly temperatures from daily maximum and minimum temperatures. This function calculates hourly temperatures following a theoretical temperature curve that interpolates hourly temperatures using a sine function from sunrise until a critical temperature is reached and a logarithmic decay function for cooling overnightas described in Linvill (1990).[[21]](#endnote-21) In computing GDD for cherry trees in this report, as this is the minimum temperature for plant physiology.9,10

In addition to GDD, the number of chilling days, and the number of heating days, other temperature-related and temperature proxy variables were tested for model fit including cumulative minimum temperature (TMIN), cumulative maximum temperature (TMIN), average cumulative temperature (TAVG), cumulative temperature difference (TDIFF), plant hardiness zone, and the optimized estimate of peak bloom day based on the sum of heating days and chill days. Plant hardiness zones are climatic zones as defined by the United States Department of Agriculture (USDA) that indicate the expected annual minimum temperature for the location.

We chose to use cumulative daily measures of temperature rather than average or median measures of temperature because research on plant phenology suggests that the timing of flower blooming is likely due to an accumulation of temperature rather than daily average temperature. It is believed that plant species exhibit minimum/maximum temperature thresholds after which there is no further daily accumulation towards the chilling/heating requirement.[[22]](#endnote-22),9,10 As a result, measures of temperature that reflect some accumulation of temperature over time are more likely to have predictive power for bloom time than average or median temperatures.

Of all the temperature-related variables considered, average cumulative temperature, cumulative temperature difference, GDD, heating days, chilling days, and optimized peak bloom date were all selected as predictors in modeling. However, not all these covariates were used in the final models. Heating days and optimized peak bloom day, for instance, both displayed high collinearity with GDD. But since GDD was a much stronger predictor for PBD, heating days and optimized peak bloom were not used in modeling. Cumulative minimum and cumulative maximum temperatures showed predictive power in some models, but not all. Cumulative minimum and maximum temperatures were abandoned in favor of stronger covariates such as average cumulative temperature and cumulative temperature difference. GDD, and TAVG were selected for all models used in this study and were significant predictors in almost every case (TAVG was not a significant predictor for Liestal). GDD was found to correlate highly with the peak bloom dates (Liestal: ; Kyoto: ; Washington: ) and proved to be the strongest predictor in our models. Chilling days was selected for predicting peak bloom day in Washington, Kyoto, and Liestal, but not in Vancouver. TDIFF was selected for Washington and Liestal, but not the other cities. Plant hardiness zone (an ordinal variable) was found to be a significant predictor, but there were issues with incorporating an ordinal variable into time series modeling so the variable was ultimately not included in any of the models.

Several sunlight-related variables were also tested for model fit including average number of sunlight hours between January 1st and April 1st, average number of hours of sunlight between May 1st and September 30th from the prior summer, and mean percentage of possible sunshine, and total hours of sunlight from Jan 1st to March 31st. The average number of sunlight hours from May-September was selected in models for Washington, Liestal and Vancouver, but for Kyoto, total number of sunlight hours (Jan-Mar) had more predictive power. Total hours of sunlight was only fit as a covariate for Kyoto so it is possible that this predictor has more predictive power for other sites as well. Total hours of sunlight was not included in models for the other sites because this variable came from data provided by the Japan Meteorological Agency for which no other site had data from. Average number of hours of sunlight from Jan-Mar was also selected for model fitting for Washington, D.C.

Sea-level air pressure (0.1\*mbar) was selected for every model where there was data available for this metric. Unfortunately, sea-level air pressure (SLP) was not available for a long enough series for Liestal or Vancouver so the covariate could not be incorporated into models for those sites.

A study has suggested that wind speed may be a significant predictor for bloom time.10 However, due to the paucity of data on wind speed, we were only able to include wind speed as a covariate in the model for Washington, D.C.

Variable selection was performed using backward stepwise regression on linear models. Covariates present in the models with a balance of the lowest AIC and highest adjusted  were selected for modeling. The model with the lowest AIC was used unless the adjusted of a model could be greatly improved by using a model with a slightly higher AIC (subjective judgement). If variables displayed high variance inflation factor (VIF), these covariates were removed from the model and backward stepwise regression was reimplemented with the offending variable left out. The process was iterated until a model of best fit had no covariates displaying high VIF.

**Statistical Analysis**

Basic linear models were first fit to the data for Washington, D.C., Kyoto, Japan, and Liestal, Switzerland (Model 1). Models of best fit (Model 1) were selected for each site using backward stepwise regression. As described in the previous section, the model with the best balance between the lowest AIC and the highest adjusted while displaying no high collinearity between covariates was selected as Model 1 for each site.

Subsequently, data for each of the covariates included in Model 1 for each site were forecasted to 2032. Multiple linear regression models with SARIMA adjusted errors were fitted with the covariate treated as the response variable and all other reasonably associated variables for which historic climate data was available in our dataset were used as predictors for modeling the covariate. Forecasts to the year 2032 for the covariate were then extrapolated from the SARIMA adjusted linear models. These forecasted covariates were then used as predictors in Model 1 to forecast values for the day of peak bloom from 2022-2032. Unfortunately, due to time constraints with completing the analysis, we were unable to compute forecasted PBDs for Kyoto.

To improve prediction accuracy, multiple linear regression models with SARIMA adjusted errors were each fit to the data for Washington, Kyoto, and Liestal (Model 2). In Model 2, the peak bloom date is the response variable and the predictors are same covariates used for each site as in Model 1. Model 1 fits a linear regression model to the peak bloom date and forecasts PBD using covariates forecasted with separate linear models with SARIMA adjusted errors for each covariate. By contrast, Model 2 fits a single linear model with SARIMA adjusted errors to the peak bloom date and forecasts the peak bloom date by extrapolating the seasonally-adjusted trend of PBD given the historic data for the covariates.

Predictions for Vancouver, B.C. were more difficult as there are presently no historic data on peak bloom dates for this site. We took two different approaches to forecast peak bloom times for Vancouver. The first method used employed a multiple linear regression model (Model 3). To forecast peak bloom days, a single multiple linear regression model of best fit was fitted to all peak bloom dates and a single design matrix including all covariate data from Washington, Kyoto, and Liestal. In Model 3, the location corresponding to each observation was ignored. A model of best fit was selected through backward stepwise regression and the model with the best balance between having the lowest AIC and the highest adjusted while displaying no high collinearity between covariates was selected as Model 3. Forecasted peak bloom dates to 2032 for Vancouver were extrapolated from Model 3 using forecasted covariates for Vancouver. Forecasted covariates for Vancouver used in Model 3 were derived in similar fashion as the forecasted covariates in Model 1. Linear models with SARIMA adjusted errors were fit to each covariate from Model 3, using all reasonably associated variables from available Vancouver historic data as predictors. Forecasted values for the covariates were extrapolated from these models and then used to extrapolate PBD for Vancouver from Model 3.

As a second method, a vector autoregressive (VAR) model was fit to historic data from Washington, Kyoto, and Liestal. PBD for Washington, Kyoto, and Liestal were treated as the endogenous variables in the model and so the model was a  dimensional VAR. Forecasted peak bloom dates for each of the three cities was extrapolated from the model. We then took the average of the peak bloom days over forecasted values for the three cities as our prediction for the peak bloom dates in Vancouver to 2032.

Due to missing values for many variables, the analysis and model fitting for each of the four locations needed to be truncated to years where there was sufficient data to fit models. Historic data was truncated to data since 1931 for Liestal, since 1953 for Kyoto, since 1973 for Washington, and since 1956 in Model 3.

Statistical analysis was performed in R, version 4.1.2 (release date: 2021-11-01).

**Results**

For three locations with historical peak bloom dates available, October 1st is found as the optimal dormancy initiation date and Table 2 summarizes the optimal parameters found. These parameters are subsequently used to calculate heating days and chill days from historic temperature data.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 2**. Estimated Parameters for the Cherry Blossoms per the Phenology Model | | | |
| Location | Tc  (°C) | Rc  (chill days) | Rh  (anti-chill days) |
| Kyoto | 6 | -125.0 | 235.0 |
| Liestal | 5 | -68.0 | 210.5 |
| Washington, D.C. | 5 | -13.5 | 264.6 |

Coefficients and standard errors for multiple linear regression models fit to Kyoto, Washington, and Liestal are shown in Tables 3, 4, and 5 respectively. Forecasting results from extrapolation of these models are shown in Table 6.  As can be seen from Table 3, GDD is highly significant at the 0.05 level to predicting peak bloom times in Kyoto. Average temperature also have a p-value less than 0.05. Unfortunately, due to data sparsity and missing values, the data available for useable covariates for Kyoto was limited.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 3.** Multiple Linear Regression Model: Kyoto, Japan | | | | | |
|  | β | s.e. | t | p |  |
| (Intercept) | 219.44 | 172.028 | 1.28 | 0.207 |  |
| Avg. Temperature | -0.01 | 0.003 | -2.51 | 0.015 | \* |
| Chill Days | 0.02 | 0.011 | 1.95 | 0.056 |  |
| GDD | 4.25 | 0.171 | 24.80 | < 2e-16 | \*\*\* |
| Total Hrs. Sunlight | -0.01 | 0.004 | -1.74 | 0.087 |  |
| SLP (Jan-Mar) | -2.00 | 1.681 | -1.19 | 0.238 |  |
|  |  |  |  |  |  |
| AIC | 230.8 |  |  |  |  |
| Adj. R | 0.917 |  |  |  |  |
|  |  |  |  |  |  |
| \*\*\* p < 0.001 \*\* p < 0.01 \* p < 0.05 | | | | | |

The model for Washington, D.C. is a more developed. GDD is still the most significant predictor by far for predicting the peak bloom date. But several other predictors are also significant to the model at the 0.05 level. Average temperature is significant at the 0.01 level. Sea-level pressure (SLP), wind speed, cumulative precipitation, and cumulative temperature difference are all significant at the 0.05 level. The covariates in this model match more closely with our expectations from research that has been done previously on the timing of cherry tree blooming. The ample availability and quality of the data for this site allowed us to build a fuller model that explains more of the variance we see in peak bloom dates.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 4.** Multiple Linear Regression Model: Washington, DC | | | | | |
|  | β | s.e. | t | p |  |
| (Intercept) | 442.30 | 2084.000 | 0.21 | 0.833 |  |
| SLP | -3.38 | 1.614 | -2.10 | 0.043 | \* |
| Cum. Precipitation | 0.01 | 0.004 | 2.13 | 0.039 | \* |
| Wind Speed | -0.95 | 0.395 | -2.42 | 0.021 | \* |
| Avg. Temperature | -0.01 | 0.003 | -3.09 | 0.004 | \*\* |
| Cum. Temp. Diff. | 0.12 | 0.054 | 2.22 | 0.032 | \* |
| Chill Days | -0.04 | 0.020 | -2.00 | 0.052 |  |
| GDD | 4.36 | 0.344 | 12.68 | < 0.001 | \*\*\* |
| Avg. Hrs. Sunlight (Jan-Mar) | 18.05 | 59.670 | 0.30 | 0.764 |  |
| Avg. Hrs. Sunlight (May-Sep) | -20.24 | 107.900 | -0.19 | 0.852 |  |
|  |  |  |  |  |  |
| AIC | 201.3 |  |  |  |  |
| Adj. R | 0.931 |  |  |  |  |
|  |  |  |  |  |  |
| \*\*\* p < 0.001 \*\* p < 0.01 \* p < 0.05 |  |  |  |  |  |

In the linear regression model for Liestal, only GDD and average hours of summer sunlight are significant. GDD is highly significant at the 0.001 level. Average sunlight hours is significant at the 0.01 level.

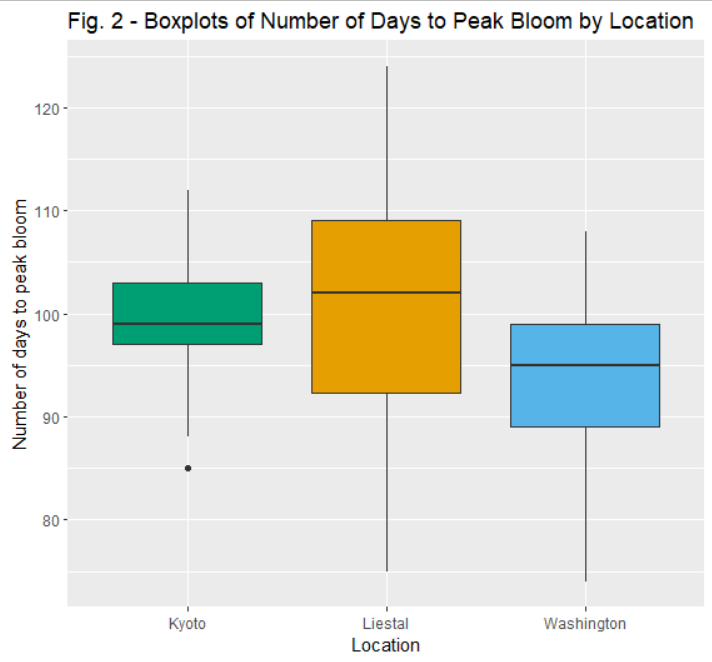
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 5.** Multiple Linear Regression Model: Liestal, Switzerland | | | | | |
|  | β | s.e | t | p |  |
| (Intercept) | -798.70 | 273.100 | -2.92 | 0.005 | \*\* |
| Avg. Hrs. Sunlight (May-Sep) | 54.77 | 18.650 | 2.94 | 0.005 | \*\* |
| Cum. Temp. Diff. | 0.02 | 0.012 | 1.38 | 0.173 |  |
| GDD | 4.47 | 0.064 | 70.34 | < 2e-16 | \*\*\* |
| Avg. Temperature | 0.00 | 0.001 | -1.54 | 0.128 |  |
| Chill Days | 0.01 | 0.007 | 1.29 | 0.202 |  |
|  |  |  |  |  |  |
| AIC | 162.6 |  |  |  |  |
| Adj. R | 0.996 |  |  |  |  |
|  |  |  |  |  |  |
| \*\*\* p < 0.001 \*\* p < 0.01 \* p < 0.05 | | | | | |

Table 6 shows the forecasted number of days into the year for peak bloom for 2022-2032 from the multiple linear regression models for Washington and Liestal. Figure 2 shows side by side boxplots for the number of days into the year for historic peak bloom dates since 1900 for Liestal and Kyoto, and since 1921 for Washington. The forecasted results for Liestal in Table 6 do not appear to be sensible as 6 out of the next 10 years from this model are forecasted to have PBDs later than 110 days into the year. The last year which had a PBD later than 110 days in Liestal was 2006 and before then, the next closest years were 1984, 1986, and 1987. From Figure 2, we can see that all 6 of these forecasts lie above the third quartile of number of days to peak bloom for Liestal. Forecasted dates for D.C. are closer to historic data for this site. For D.C., all forecasted PBDs except 2025, 2026, and 2027 lie within the interquartile range of the historic PBDs for this site as shown in Table 6 and Fig. 2.

Tables 7, 8, and 9 show the coefficients and standard errors for the multiple linear regression models with SARIMA adjusted errors.

For Kyoto, the model of best fit had 5 non-seasonal autoregressive lags, 4 non-seasonal moving average coefficients, 1 seasonal autoregressive lag and 2 seasonally adjusted moving average terms. The model showed seasonality with a period of 5 years. In Table 7, we note that the coefficient for GDD is considerably larger than its standard error and is therefore highly significant to the model. In fact, GDD dominates over all other covariates in the model (the covariates in the model were on a similar enough scale to be compared). All other factors held constant, an increase of one growing degree day is positively associated with a roughly 4 day delay in the peak bloom date. SLP is also significant in this model. All other factors held equal, an increase in average sea-level air pressure by 0.1\*mbar is associated with a peak bloom date that is about 4 days earlier. Most of the seasonal and nonseasonal autoregressive and moving average lags are also important for the Kyoto model. Of particular note, the 5th order nonseasonal autoregressive lag is highly important for this model.

|  |  |  |
| --- | --- | --- |
| **Table 6.** Forecasts for Cherry Blossom Peak Bloom (2022-2032): Basic Linear Regression Model (Model 1) | | |
| Year | Washington, DC | Liestal, Switzerland |
| 2022 | 99.8 | 110.3 |
| 2023 | 94.2 | 116.0 |
| 2024 | 93.0 | 113.3 |
| 2025 | 84.4 | 86.6 |
| 2026 | 86.9 | 115.8 |
| 2027 | 99.0 | 93.2 |
| 2028 | 93.9 | 96.0 |
| 2029 | 95.1 | 80.9 |
| 2030 | 96.9 | 115.4 |
| 2031 | 98.7 | 114.1 |
| 2032 | 93.3 | 104.9 |



|  |  |  |
| --- | --- | --- |
| **Table 7.** Coefficients for Multiple Linear Regression Model with SARIMA Adjusted Errors for Cherry Blossom Peak Bloom: Kyoto, Japan | | |
| Model: | SARIMA (5, 0, 4) (1, 0, 2) [5] | |
|  |  |  |
|  | β | s.e. |
| AR1 | 0.324 | 0.077 |
| AR2 | -0.424 | 0.075 |
| AR3 | -0.331 | 0.075 |
| AR4 | 0.352 | 0.066 |
| AR5 | -0.865 | 0.074 |
| MA1 | 0.164 | 0.144 |
| MA2 | 1.243 | 0.179 |
| MA3 | -0.133 | 0.169 |
| MA4 | 0.427 | 0.160 |
| SAR1 | 0.327 | 0.138 |
| SMA1 | 1.191 | 0.158 |
| SMA2 | 1.000 | 0.166 |
| Intercept | 427.477 | 17.613 |
| Drift | -0.027 | Inf |
| Avg. Temperature | -0.003 | 0.001 |
| Chill Days | 0.044 | 0.003 |
| GDD | 3.737 | 0.029 |
| Total Sunlight Hrs. (Jan-Mar) | -0.007 | 0.001 |
| SLP (Jan-Mar) | -3.975 | 0.186 |
|  |  |  |
|  | 0.3817 |  |
| AIC | 186.58 |  |
| BIC | 231.26 |  |
| Log lik | -73.29 |  |
| RMSE | 0.5229 |  |
| MAPE | 0.4424 |  |

The model for Washington, D.C. has 6 non-seasonal and 3 seasonal autoregressive terms, and 4 non-seasonal and 5 seasonal moving average terms. Like the Kyoto model, the model for Washington also has a seasonal period of 5 years. As can be seen from Table 8, the coefficients for GDD, wind speed, and SLP are all highly significant to predicting the peak bloom date. Similar to what we observed in the Kyoto SARIMA model, in the D.C. model, every additional growing degree day is associated with a roughly 4 day delay in PBD. The effect of SLP on PBD is a little stronger for D.C. than it was for Kyoto as an additional increase of 0.1\*mbar is associated with a PBD 6 days earlier. As for the SARIMA coefficients, it is clear that all the seasonally adjusted autoregressive terms, particularly SAR1, SAR2, SAR4, and SAR5, are highly important to the model’s predictive power. Nonseasonal autoregressive lag 2 and moving average lag 2 are also worth noting. The Washington model further displays a clear drift in trend as the drift coefficient for this model was highly significant.

|  |  |  |
| --- | --- | --- |
| **Table 8.** Coefficients for Multiple Linear Regression Model with SARIMA Adjusted Errors for Cherry Blossom Peak Bloom: Washington, DC | | |
| Model: | SARIMA (6, 0, 4) (3, 0, 5) [5] | |
|  |  |  |
|  | β | s.e. |
| AR1 | -1.7654 | Inf |
| AR2 | -1.331 | Inf |
| AR3 | -1.3839 | 0.0679 |
| AR4 | -1.8411 | Inf |
| AR5 | -1.2227 | Inf |
| AR6 | -0.3924 | 0.09 |
| MA1 | 1.0597 | 0.1076 |
| MA2 | 1.0297 | 0.0913 |
| MA3 | 0.826 | 0.1189 |
| MA4 | -0.1358 | 0.0794 |
| SAR1 | -1.2412 | 0.0125 |
| SAR2 | -0.9233 | 0.0068 |
| SAR3 | -0.1529 | 0.0123 |
| SAR4 | -0.9245 | 0.0069 |
| SAR5 | -1.241 | 0.0125 |
| SAR6 | -0.999 | Inf |
| SMA1 | 0.9304 | Inf |
| SMA2 | -0.6906 | Inf |
| SMA3 | -0.6742 | Inf |
| Intercept | 4842.508 | Inf |
| Drift | 0.034 | 0.001 |
| SLP (Jan-PBD) | -6.635 | 0.043 |
| Cum. Precipitation | 0.015 | Inf |
| Wind Speed | -0.983 | 0.004 |
| Avg. Temperature | -0.011 | Inf |
| Cum. Temp. Diff. | 0.081 | 0.002 |
| Chill Days | 0.006 | 0.001 |
| GDD | 4.264 | 0.019 |
| Avg. Hrs. Sunlight (Jan-Mar) | -97.944 | Inf |
| Avg. Hrs. Sunlight (May-Sep) | -221.542 | Inf |
|  |  |  |
|  | 0.0017 |  |
| AIC | 94.28 |  |
| BIC | 152.92 |  |
| Log lik | -16.14 |  |
| RMSE | 0.0257 |  |
| MAPE | 0.0228 |  |

The model for Liestal has 1 non-seasonal and 2 seasonal autoregressive lags, and 0 non-seasonal and 5 seasonal moving average coefficients. The period for seasonality is only 1 year but the model coefficients and fit differ from a non-seasonal autoregressive moving average model without a seasonal component (ARMA(3, 5)). Like the Kyoto and Washington linear models with SARIMA adjusted errors, GDD is highly significant in the Liestal model and the slope for this covariate is similar to the other two sites. In Liestal, an additional GDD is associated with a 4.5 day delay in PBD. Average number of hours of summer sunlight (May-Sep) was also a significant predictor in this model. The seasonally adjusted autoregressive order 1 and 2 lags are highly significant for the Liestal model.

|  |  |  |
| --- | --- | --- |
| **Table 9.** Coefficients for Multiple Linear Regression Model with SARIMA Adjusted Errors for Cherry Blossom Peak Bloom: Liestal, Switzerland | | |
| Model: | SARIMA (1, 0, 0) (2, 0, 5) [1] | |
|  |  |  |
|  | β | s.e. |
| AR1 | 0.323 | 0.172 |
| SAR1 | -0.374 | 0.016 |
| SAR2 | -1.000 | 0.004 |
| SMA1 | -0.198 | 0.230 |
| SMA2 | 1.081 | 0.958 |
| SMA3 | -1.173 | 0.477 |
| SMA4 | 0.021 | 0.968 |
| SMA5 | -0.723 | 0.681 |
| Intercept | -930.457 | 199.144 |
| Avg. Hrs. Sunlight (May-Sep) | 63.711 | 13.546 |
| Cum. Temp. Diff. | 0.002 | 0.013 |
| GDD | 4.533 | 0.073 |
| Avg. Temperature | -0.001 | 0.002 |
| Chill Days | 0.011 | 0.005 |
|  |  |  |
|  | 0.3848 |  |
| Log lik | -60.14 |  |
| AIC | 150.27 |  |
| BIC | 183.56 |  |
| RMSE | 0.5506 |  |
| MAPE | 0.4049 |  |

Table 10 shows a comparison of the predictive power of the linear models and the linear models with SARIMA adjusted errors and shows the root mean squared error (RMSE) of the actual peak bloom dates compared to the PBDs predicted by the model. The RMSE is much lower for the SARIMA adjusted models, indicating that these models have a much better fit to the historical data.

Table 11 shows the predictive power of the basic multiple linear regression models for Liestal and Washington as well as the general multiple linear regression model used to forecast peak bloom dates for Vancouver (Model 3). The table shows the value for the linear models along with their predicted . Predictive is a measure of how well a model is able to predict the response for new observations over the range of data fitted in the model. If the model is able to predict new observations with a small margin of error, the predicted should be very close to the for the model. As can be seen from Table 11, the predictive is extremely close to the value for the Liestal linear model, indicating that this model has very good predictive power within the range of the historic data. The difference between and predictive for Washington and Vancouver (Model 3) is fairly small as well indicating that these models also have good predictive power over the range of the historic data. The methods used to compute the predictive for these models are borrowed from Hopper (2014).[[23]](#endnote-23)

|  |  |  |
| --- | --- | --- |
| **Table 10.** Comparison of Predictive Power with Linear Regression Model and Linear Regression Model with SARIMA Adjusted Errors: RMSE | | |
|  | RMSE | RMSE |
|  | (Linear Model) | (SARIMA Model) |
| Kyoto |  | 0.5229 |
| Liestal | 0.7462 | 0.5506 |
| Washington | 1.5085 | 0.0257 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 11.** Comparison of Predictive Power with Linear Regression Model: Predictive | | | |
|  |  | Predicted | Difference |
| Liestal | 0.996 | 0.995 | 0.001 |
| Washington | 0.944 | 0.912 | 0.031 |
| Vancouver | 0.591 | 0.565 | 0.026 |

Table 12 shows the forecasts for the number of days into the year for cherry blossom peak bloom from linear regression models with SARIMA adjusted errors. Forecasts from the VAR model are shown in Table 13.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 12.** Forecasts for Cherry Blossom Peak Bloom (2022-2032): Linear Regression Model with SARIMA Adjusted Errors | | | | |
| Year | Kyoto, Japan | Washington, DC | Liestal, Switzerland | Vancouver, BC |
| 2022 | 97.2 | 98.3 | 114.7 | 91.6 |
| 2023 | 99.3 | 93.4 | 116.9 | 90.0 |
| 2024 | 94.2 | 93.4 | 117.3 | 89.2 |
| 2025 | 98.8 | 89.8 | 84.5 | 86.9 |
| 2026 | 100.5 | 89.3 | 114.4 | 88.5 |
| 2027 | 97.5 | 107.8 | 91.4 | 90.4 |
| 2028 | 91.0 | 94.2 | 94.3 | 90.2 |
| 2029 | 93.5 | 93.9 | 79.0 | 88.2 |
| 2030 | 98.0 | 96.9 | 116.8 | 88.9 |
| 2031 | 103.4 | 97.3 | 114.3 | 87.5 |
| 2032 | 99.8 | 101.8 | 107.4 | 88.5 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 13.** Forecasts for Cherry Blossom Peak Bloom (2022-2032): VAR Model | | | | |
| Year | Kyoto, Japan | Washington, DC | Liestal, Switzerland | Vancouver, BC |
| 2022 | 94.5 | 90.0 | 91.7 | 92.0 |
| 2023 | 95.0 | 92.1 | 99.5 | 95.5 |
| 2024 | 96.0 | 92.7 | 100.7 | 96.5 |
| 2025 | 96.2 | 93.0 | 101.5 | 96.9 |
| 2026 | 96.3 | 93.1 | 101.8 | 97.0 |
| 2027 | 96.3 | 93.1 | 101.9 | 97.1 |
| 2028 | 96.4 | 93.1 | 101.9 | 97.1 |
| 2029 | 96.4 | 93.1 | 101.9 | 97.1 |
| 2030 | 96.4 | 93.1 | 101.9 | 97.1 |
| 2031 | 96.4 | 93.1 | 101.9 | 97.1 |
| 2032 | 96.4 | 93.1 | 101.9 | 97.1 |

**Discussion**

The fitted basic linear models for Washington, Liestal, and Kyoto have a good fit to the historic data and the linear models for Washington and the general model for Vancouver (Model 3) have small differences between the predictive and indicating that these models are likely to predict well compared to the fitted model to new observations over the range of the historic data. It is then assumed that forecasted dates from Model 3 and from the basic multiple linear regression model for Washington are likely to be close to the actual PBDs for 2022-2032. While the predictive is very close to the value for the Liestal linear model, as shown above in Table 6 and in Fig. 1 and Fig. 2, the forecasted dates from this model do not appear likely given the recent trend in PBDs over the past 20 years and given the historic distribution of PBDs for that site. The basic linear models for Washington and Vancouver thus appear fair predictors of historic PBDs, though the Washington model likely predicts much better than the Vancouver model does as the model fit for Washington, D.C. is much better than the model fit for the Vancouver model. Due to their ability to predict over the range of the historic data with similar accuracy to their model fit, it is expected that these models will also produce fairly accurate forecasts to 2032.

The RMSE for the linear models with SARIMA adjusted errors for Washington and Liestal were notably narrower than for their corresponding basic linear models. This is an indication that the SARIMA adjusted models had a much better fit to the historic data than the basic linear models did. While the predictive ability of the SARIMA adjusted models was not able to be estimated directly due to time constraints with the writing of this report, the fact that these models, which use the same covariates, have lower RMSE is an indication that these models have better predictive power over the range of the historic data. We argue that the SARIMA adjusted models give more accurate forecasts due to the fact that these models are able to detect seasonal trends whereas simple linear models are not. Global weather patterns such as El Niño are well known to occur cyclically with quasi-predictive seasonality. We believe that modeling bloom dates—which are closely tied to cumulative heating and cooling temperatures—are likely to behave with quasi-regular seasonality as well. To fully assess the predictive power of these models, we would perform cross-validation on the data to see how well a model fitted to a random subset of the data (i.e. the training set) would perform in predicting the PBDs in a subset of years left out of the model used as a test set.

We note that GDD was a significant and important predictor of PBD in all of our models, both linear models and linear models with SARIMA adjusted errors. The effect of an additional growing degree day appears to delay the bloom date by about 4 days. The number of chilling days was selected for all Model 1 and Model 2 models and the covariate was significant in all the SARIMA adjusted models, but the effect size was small and not practically significant. Average number of hours of summer daylight (May-Sep) and total number of hours of sunlight (Jan-Mar) were significant predictors in some models, but not in others.

One covariate that surprised us was average sea-level air pressure. While average SLP was not a particularly important covariate in the basic linear models (Model 1), SLP was a highly significant predictor for both Washington and Kyoto in the SARIMA adjusted models (Model 2) and the magnitude and direction of this covariate’s coefficient was similar and practically significant in both cases. Unfortunately, there was not enough data for SLP to include this variable in the Liestal model but it would be interesting to see if we observe a similar association between SLP and seasonally adjusted PBD in other locations. We are not aware of any research to date on the effect atmospheric air pressure may have on bloom times for cherry trees, but the fact that these predictors were significant in modeling is perhaps indicative that some association exists, and it is perhaps an area that merits more research.

Average wind speed was also an important predictor for both the basic linear model and the linear model with SARIMA adjusted errors for Washington, D.C. Unfortunately, only Washington had sufficient data on wind speed to include in modeling. But given that past research on bloom dates for cherry blossom trees has also found wind speed to be a significant predictor for the timing of blooms,10 we believe there is good evidence to suggest that wind speed is an important factor in predicting PBDs. More research is needed to pinpoint the effect this variable may have on the timing of blooms.

As for the forecasts for Vancouver, the model used to fit the forecasted covariates for Vancouver did not have a very good value. This is due not only to the lack of adequate predictors available to choose from for modeling Model 3, but also to the heterogeneity of locations for data included in the model. We believe there is still much variation in PBDs to be accounted for, even our best models, but due to a lack of data for covariates and poor data quality for covariates that we did have access to, it was impossible to model PBDs more accurately. Some of the most promising predictors like average sea-level air pressure and wind speed were not available for all locations so these covariates could not be incorporated into Model 3. Other possible covariates which could hold promise in improving predictions such as plant hardiness zones, altitude, and latitude could not be incorporated into our models due to issues incorporating discrete data into our models or due to lack of replicate observations.

Of the two models that we used to forecast PBDs for Vancouver, the basic linear regression model is more likely to yield more accurate results as this model was able to incorporate covariate data from Vancouver while the VAR model was not able to incorporate any data specific to Vancouver. The forecasts in Table 12 are thus more likely to be closer to the actual PBDs than the forecasts provided in Table 13. Moreover, the predicted for Model 3 was quite close to the value for the model, indicating that the fit we observe in the model is likely to be close to the values of PBD we would predict from this model. We conclude that the linear model (Model 3) is better for forecasting Vancouver PBDs than the VAR model.

**References**

1. Brooklyn Botanic Garden. Mizue Sawano: The art of the cherry tree. Brooklyn Botanic Garden; 2006. [↑](#endnote-ref-1)
2. Miller‐Rushing AJ, Katsuki T, Primack RB, Ishii Y, Lee SD, Higuchi H. Impact of global warming on a group of related species and their hybrids: cherry tree (Rosaceae) flowering at Mt. Takao, Japan. American Journal of Botany. 2007 Sep;94(9):1470-8. [↑](#endnote-ref-2)
3. Lakatos L, Szabó T, Szabó Z, Soltész M, Nyéki J. Relation of sour cherry blooming dynamics and meteorological variables. International Journal of Horticultural Science. 2009 Sep 2;15(4):17-23. [↑](#endnote-ref-3)
4. Chung U, Mack L, Yun JI, Kim SH. Predicting the timing of cherry blossoms in Washington, DC and mid-Atlantic states in response to climate change. PloS one. 2011 Nov 7;6(11):e27439. [↑](#endnote-ref-4)
5. Cesaraccio C, Spano D, Snyder RL, Duce P. Chilling and forcing model to predict bud-burst of crop and forest species. Agricultural and Forest Meteorology. 2004 Nov 20;126(1-2):1-3. [↑](#endnote-ref-5)
6. Jung JE, Kwon EY, Chung UR, Yun JI. Predicting Cherry Flowering Date Using a Plant Phonology Model. Korean Journal of Agricultural and Forest Meteorology. 2005;7(2):148-55.  [↑](#endnote-ref-6)
7. Shi P, Chen Z, Yang Q, Harris MK, Xiao M. Influence of air temperature on the first flowering date of Prunus yedoensis Matsum. Ecology and Evolution. 2014 Feb;4(3):292-9. [↑](#endnote-ref-7)
8. Palmer, B. What makes the cherry trees bloom when they do? Washington Post [Internet]. 2011 Apr 4 [cited 2022 Mar 17]. Available from: https://www.washingtonpost.com/national/what-makes-the-cherry-trees-bloom-when-they-do/2011/03/29/AFvRbRfC\_story.html. [↑](#endnote-ref-8)
9. Wang H, Wang H, Ge Q, Dai J. The interactive effects of chilling, photoperiod, and forcing temperature on flowering phenology of temperate woody plants. Frontiers in plant science. 2020 Apr 16;11:443. [↑](#endnote-ref-9)
10. Ohashi Y, Kawakami H, Shigeta Y, Ikeda H, Yamamoto N. The phenology of cherry blossom (Prunus yedoensis “Somei-yoshino”) and the geographic features contributing to its flowering. International Journal of Biometeorology. 2012 Sep;56(5):903-14. [↑](#endnote-ref-10)
11. Villar L, Lienqueo I, Llanes A, Rojas P, Perez J, Correa F, Sagredo B, Masciarelli O, Luna V, Almada R. Comparative transcriptomic analysis reveals novel roles of transcription factors and hormones during the flowering induction and floral bud differentiation in sweet cherry trees (Prunus avium L. cv. Bing). PloS one. 2020 Mar 12;15(3):e0230110. [↑](#endnote-ref-11)
12. Six stages of cherry blossom development from buds to flowers (infographics). Vancouver Cherry Blossom Festival [Internet]. 2020 Mar 9 [cited 2022 Mar 17]. Available from: https://www.vcbf.ca/six-stages-of-cherry-blossom-development-from-buds-to-flowers-infographics. [↑](#endnote-ref-12)
13. Bloom Watch. National Park Service [Internet]. [cited 2022 Mar 17]. Available from: https://www.nps.gov/subjects/cherryblossom/bloom-watch.htm. [↑](#endnote-ref-13)
14. Menne, Matthew J., Imke Durre, Bryant Korzeniewski, Shelley McNeill, Kristy Thomas, Xungang Yin, Steven Anthony, Ron Ray, Russell S. Vose, Byron E.Gleason, and Tamara G. Houston (2012): Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. NOAA National Climatic Data Center. doi:10.7289/V5D21VHZ [access date: Feb. 24, 2022]. [↑](#endnote-ref-14)
15. User Engagement and Services Branch, National Climatic Data Center, NESDIS, NOAA, U.S. Department of Commerce. Global Surface Summary of the Day [dataset]. [cited 2022 Mar 17]. Available from: https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-day. [↑](#endnote-ref-15)
16. Global Monitoring Laboratory. NOAA Solar Calculator [dataset]. Earth System Research Laboratories. [cited 2022 Mar 17]. Available from: https://gml.noaa.gov/grad/solcalc/index.html. [↑](#endnote-ref-16)
17. Hardiness Zone Map. Plant Maps [Internet]. [updated 2022; cited 2022 Mar 17]. Available from: https://www.plantmaps.com/. [↑](#endnote-ref-17)
18. Czernecki, B.; Glogowski, A.; Nowosad, J. Climate: An R Package to Access Free In-Situ Meteorological and Hydrological Datasets for Environmental Assessment. Sustainability 2020, 12, 394. https://doi.org/10.3390/su12010394. [↑](#endnote-ref-18)
19. Luedeling E, Kunz A, Blanke MM. Identification of chilling and heat requirements of cherry trees—a statistical approach. International Journal of Biometeorology. 2013 Sep;57(5):679-89. [↑](#endnote-ref-19)
20. Luedeling E, Fernandez E. Statistical methods for phenology analysis in temperate frui trees[r package chillr version 0.72.7] [Internet]. The Comprehensive R Archive Network. Comprehensive R Archive Network (CRAN); 2022 [cited 2022Mar7]. Available from: https://cran.r-project.org/web/packages/chillR/. [↑](#endnote-ref-20)
21. Linvill DE. Calculating chilling hours and chill units from daily maximum and minimum temperature observations. HortScience. 1990 Jan 1;25(1):14-6. [↑](#endnote-ref-21)
22. Cannell MG, Smith RI. Thermal time, chill days and prediction of budburst in Picea sitchensis. Journal of applied ecology. 1983 Dec 1:951-63. [↑](#endnote-ref-22)
23. Tom Hopper. Can we do better than r-squared? [Internet]. 2014 [cited 2022 Mar 21]. Available from: https://tomhopper.me/2014/05/16/can-we-do-better-than-r-squared/. [↑](#endnote-ref-23)