2022 Cherry Blossom Prediction Competition

When Does Cherry Blossom Bloom?

February 28, 2022

Author:
Feng Yu
Lezheng Fang
Xinyu Zhang
Yuxin Wang

Contact Email:
fyu9@ucmerced.edu
lezheng.fang@gatech.edu
xzhang97@ncsu.edu
yuxinwang2023@u.northwestern.edu

1 Introduction

A cherry blossom is the flower of many trees of the *Prunus* subg. *Cerasus*. Since ancient times, humans have been observing, documenting, and recognizing it as a significant agricultural indicator as well as a cultural symbol. In recent centuries, due to the high ornamental and cultural values, cherry trees travel overseas as ambassadors of friendship and peace, bringing along traditional activities, cherry blossom viewing (hanami). In addition to appreciating the flowers themselves, scientists and researchers have found the cherry blossom, among many other phenological phenomena, a key indicator for both local and global climate changes. An accurate prediction of cherry blossom plays a remarkable role in advancing our understanding of phenology, climate change, and its challenges. In this report, we implement a data-driven method, namely the logistic phenological cherry blossom (LPCB) model, to predict the cherry blossom peak bloom dates in Kyoto, Washington DC, Liestal-Weideli, and Vancouver from 2022 to 2031.

2 Methodology

In this section, we briefly review the general phenological modeling for plants blossom prediction in Section 2.1 and introduce the customization with the data-driven logistic regression methods in Section 2.2. Temperature forecasting and data manipulation including missing data imputations are presented in Section 2.3 and Section 2.4 respectively.

2.1 Phenological Model

In general, cherry trees require a low-temperature period in fall and winter to break dormancy [1] and a warm-temperature period to accumulate heat, resume the growth and develop the bud swelling and flower blossom [2]. A commonly used model to quantify the required heat leverages the Cumulative Degree-Days (CDD) measure as shown in Equation (1) with corresponding threshold temperature defined accordingly [3],

$$CDD_{t,th} = \sum_{i=t_0}^{t} (T_i - T_0), \tag{1}$$

where t is the current time point (day of the year), t_0 is the starting date of the CDD accumulation, T_i is the temperature at the heat accumulation day, and T_0 is the threshold temperature at which cherry starts to accumulate heat. However, t_0 and T_0 vary for different locations and cherry species, which can be obtained by observing field experiments. Therefore, we propose a more generalized data-driven method to conduct the analysis without assuming predefined t_0 and T_0 values using logistic regression.

We note that, the CDD concept has also been linked with citrate metabolism to explain the underlying molecular biology mechanism [4]. Hence, we use CDD as a key variable in our prediction model, and consider it across different stages of cherry blossom starting from the swallow bud stage to full blossom, which helps establish and assess our statistical model [5].

2.2 Logistic Regression

In this section, we introduce how to construct the logistic phenological cherry blossom (LPCB) model based on observed cherry blossom data without defining threshold temperature or aforementioned phenological dates. To begin with, the observed cherry blossom data has given the annual peak blossom day, $t_b(l, y)$, for each location l and year y. If we define Z_t as the indicator of whether the cherry trees have fully blossomed on t^{th} day of this year, then

$$Z_t = \begin{cases} 0 & \text{if } t < t_b; \\ 1 & \text{if } t \geq t_b. \end{cases} \tag{2}$$

can be observed from data. Define $p_t = P(Z_t = 1)$ as the probability of day t reaching the peak bloom day t_b. Note that p_t is approximately monotone increasing as the day t increases. Then, the LPCB model is defined as:

$$logit(p_t) = log\left(\frac{p_t}{1-p_t}\right) = \beta_0 + \beta_1 t + \beta_2 \sum_{i=1}^{t} T_i + \varepsilon_t, \quad \varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2), \tag{3}$$

which is equal to model in Equation (4) by doing the sigmoid transformation:

$$p_t = \left\{1 + exp\left[-\left(\beta_0 + \beta_1 t + \beta_2 \sum_{i=1}^{t} T_i + \varepsilon_t\right)\right]\right\}^{-1}, \tag{4}$$

where β_0 describes the intercept of the logistic model, β_1 the slope associated with time t, and β_2 the slope with accumulative temperature (cumulative degree day CDD_t without the threshold temperature). By fitting the model, Eq.(4), to data, we could obtain estimated coefficients, and thus the experience function of cherry blossom date for different locations.

2.3 Temperature Forecasting

To obtain temperature data for the prediction, We conduct time series analyses of temperature forecasts. Since temperatures have obvious year-round circulations with the temperature high in summer and low in winter, we conduct the Seasonal and Trend decomposition using Loess (STL, [6]) to remove the seasonal effect and fit the Auto-Regressive Integrated Moving Average (ARIMA) model to predict the future for the future twenty years.

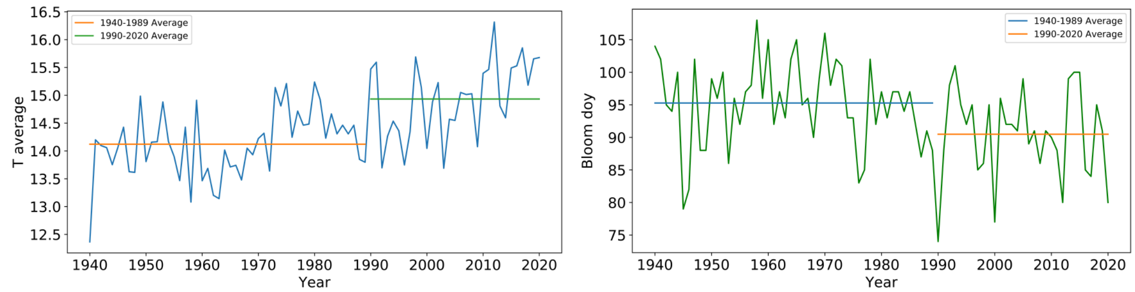
2.4 Data Collection and Imputation

In this report, annual cherry blossom dates and daily temperature history are collected for four considered locations: Kyoto, Washington DC, Liestal-Weideli, and Vancouver. Detailed blossom records for Kyoto (1953-2021), Washington DC (1936-2021), and Liestal-Weideli (1901-2021) have been provided. The cherry blossom data in Seattle/Vancouver (2014-2021) is individually sourced from the University of Washinton Cherry Blossoms(@uwcherryblossom). The temperature data for Vancouver is obtained from the Weather Dashboard (https://vancouver.weatherstats.ca/). The temperature data for all other three locations are obtained from the National Centers for Environmental Information (https://www.ncei.noaa.gov/).

Note that the collected temperature data suffers from minor data missing ($< 10\%$) and different methods of temperature documentation – certain years with either daily average temperature documented or high-and-low temperatures documented but not both. Hence, we use the daily average temperature as T_t, and impute the missing T_t via the Expectation-Maximization with bootstrapping algorithm with **Amelia** package in R [7] by considering the average of 10 imputed data-sets to enhance stability.

3 Result Analysis

After training the model on the imputed dataset for the 4 locations, we further adjust the estimated coefficients based on the chilling requirement and then obtain the predictions for 4 locations over the future ten years.



(a) (b)

Figure 1: (a) Annual average temperature in Washington, D.C., 1940-2020. The y-axis represents the average temperature of each year in Celsius. (b) Full bloom day (shown in day-of-year) in Washington, D.C., 1940-2020.

3.1 Exploratory Data Analysis

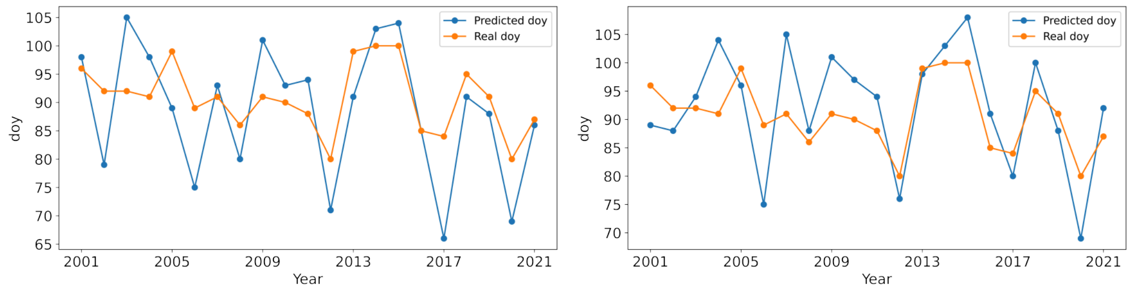
In Fig. 1, we observe a significant shift in average temperature between 1990-2021 in Washington D.C. The average temperature increased by about $0.8°C$, as depicted in Fig. 1a. Meanwhile, Fig. 1b illustrates that the average full blossom day advances by 5 days during the same period. We observed several extremely early blossom days during recent years. To test whether our model can catch these abnormal, we selected the temperature profile in Washington, D.C. (from 1937 to 2000) as the training set to initialize our model and predict the full blossom day (in the unit of day-of-year, DOY) between 2001-2021. We can validate our predictions by comparing them with the real bloom DOY. As illustrated in Fig. 2a, our model has successfully captured the trend from the year 2001 to 2021. By comparing $\beta_0, \beta_1, \beta_2$ within the validation set (2001 - 2021), and compute the mean squared error (MSE), we locate the best p value for the time series predicting the β parameters. For Washington, D.C., $p = 9, 3, 13$ for $\beta_0, \beta_1, \beta_2$, respectively. The p value of other locations are shown in the table below.

Location	β_0	β_1	β_2
DC	9	3	13
Kyoto	1	10	7
Liestal	10	18	5

However, one may notice that some of our predicted full blossom DOYs are much earlier than the observed real full blossom DOYs. Such shift lies in the chilling unit (CU) accumulation in fall and winter, and the chilling process has been taken into account by adjusting the estimated coefficients of the model in Equation (3) with more details in the following section.

3.2 Adjusted Coefficients

In Fig. 2a, some predictions tend to forecast an early blossom date than the actual date as described, which might result from the shift of heat accumulation starting points, and in fact, a warmer winter could delay the bloom date. According to the chilling requirement of cherry blossom [1], a higher winter temperature may delay the CDD accumulation. Thus, we introduced an adjustment term based on the average winter temperature on the β_0 factor to improve our prediction result. We consider the average temperature in the last 70 days of every year as the chilling unit (CU) for the next year. A warm winter (high CU) would delay the bloom DOY such that one can reduce the value of β_0, and vice versa for a small CU. For Washington, D.C., if CU < 6.5, $\beta_0 \to \beta_0 + 50$; if CU > 8.0, $\beta_0 \to \beta_0 - 40$. The adjusted predictions are demonstrated by Fig.(2b). The adjustment could reduce the shift caused by CU in winter. For Kyoto, $\beta_0 \to \beta_0 + 50$ if CU < 9.0, while for Liestal $\beta_0 \to \beta_0 - 50$ when CU > 6.5.



(a) (b)

Figure 2: (a) The validation for DOY from 2001 to 2021 for Washington DC. The training set is the bloom DOY, and the temperature profile from 1937 - 2000. β_i is predicted by AR(P) models, for $\beta_0, \beta_1, \beta_2$, $P = 9, 3, 13$, respectively. (b) The validation for DOY from 2001 to 2021. The training set is from 1937 - 2000. CU is given by the average daily temperature during the last 71 days in the previous year. If CU < 6.5, $\beta_0 \rightarrow \beta_0 + 50$; if CU > 8.0, $\beta_0 \rightarrow \beta_0 - 40$.

3.3 Validation of Temperature Forecasting

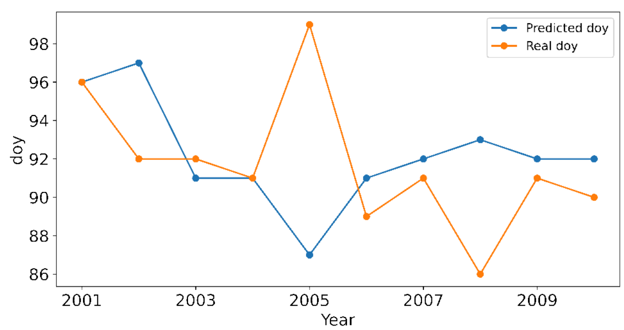
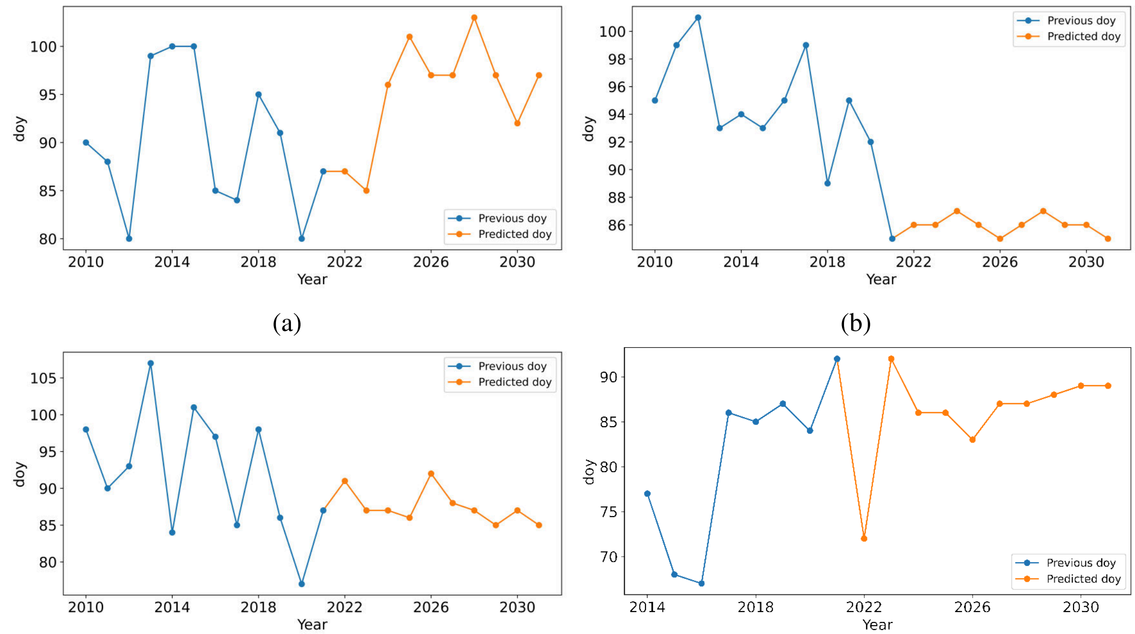


Figure 3: The validation for the temperature time series. The temperature and β_i parameters are all predicted by time series. The results match well with the real bloom doy.

To validate the temperature forecast, we test the predictions from 2001 to 2010 based on forecasted temperature estimated using methods in Section 2.3. As shown in Fig. 3, the predictions are still reasonable, and for which we can validate the robustness of the compartmental time-series temperature forecasting method.

4 Discussion and Conclusion

To address the competition requirement, here we demonstrate the future blossom trend in the 4 locations, as shown in Fig.(4). Our 2022 Washinton D.C prediction result is March 28th which is the same as 2021. Our 2022 Kyoto prediction result is March 27th which is 1 day later than 2021. Our 2022 Liestal-Weideli prediction result is April 1st which is 2 days later than 2021. Our 2022 Vancouver prediction result is March 13th which is 19 days earlier than 2021.



(a) (b) (c) (d)

Figure 4: (a) Prediction result of Washington D.C.. Y-axis represents the day of year. (b) Prediction result of Kyoto. (c) Prediction result of Liestal-Weideli. (d) Prediction result of Vancouver.

References

[1] K Mahmood, JG Carew, P Hadley, and NH Battey. Chill unit models for the sweet cherry cvs stella, sunburst and summit. *The Journal of Horticultural Science and Biotechnology*, 75(5):602–606, 2000.

[2] V Hochmaier. Chilling unit accumulation and degree-day requirements of four sweet cherry (prunus avium l.) cultivars. In *VI International Cherry Symposium 1020*, pages 203–207, 2009.

[3] Raymond Bonhomme. Bases and limits to using 'degree. day'units. *European journal of agronomy*, 13(1):1–10, 2000.

[4] Philippe Lobit, Michel Génard, BH Wu, P Soing, and Robert Habib. Modelling citrate metabolism in fruits: responses to growth and temperature. *Journal of Experimental Botany*, 54(392):2489–2501, 2003.

[5] Costanza Zavalloni, Jeffrey A Andresen, and JA Flore. Phenological models of flower bud stages and fruit growth ofmontmorency'sour cherry based on growing degree-day accumulation. *Journal of the American Society for Horticultural Science*, 131(5):601–607, 2006.

[6] Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning. Stl: A seasonal-trend decomposition. *J. Off. Stat*, 6(1):3–73, 1990.

[7] James Honaker, Gary King, and Matthew Blackwell. Amelia ii: A program for missing data. *Journal of statistical software*, 45:1–47, 2011.