Want to attend a cherry blossom festival? Check the temperature

Jiyeon Park, Lee Park, Tori Stanton
University of Kentucky

Introduction

Enjoying cherry trees at full bloom is an event looked forward to worldwide. Enthusiasts in many countries wait patiently for the beautiful trees to unveil their colorful flowers as well as related local festivals. Unfortunately this phenomenon is not consistent over the years, leaving those hopeful to see the blooms unable to plan trips to witness them. Local authorities that plan the festivals struggle scheduling the mass event due to the inconsistency of the blooming. Accurately predicting when these trees will bloom would give time to plan to those wanting to experience their beauty and those organizing events.

Many plant species have cues that trigger certain physiological processes. Some plant species use cues based on temperature, moisture, or photoperiod. While cherry trees are not very sensitive to precipitation or photoperiod like some other species (Heide, 2007), temperature is a known factor that influences when cherry trees bloom (Miller-Rushing et al., 2007). In particular low temperatures in February and March are good predictors of bloom timing (Chung et al., 2011). Temperatures will be a function of geography, such as altitude, latitude, and longitude, as well as global climate trends. As the climate warms, particularly as winters become more mild, peak cherry bloom timing may shift earlier in the year (Chung et al., 2011).

The objective of this project was to create a model to predict the day of year (DOY) when cherry trees would bloom in four different locations (Kyoto, Japan; Liestal-Weideli, Switzerland; Vancouver, Canada; Washington, D.C., USA). We explored several model techniques to gain insights into the data and to assess which type of model had the best prediction ability.

Methods

Prior to collecting outside data, we chose five cities from South Korea, Japan, Switzerland, United States ensuring the city that we need to predict for was included. We selected cities to ensure a range of different latitudes and longitudes were represented in a given country.

Temperature data collected included maximum, minimum, and average daily temperature and daily precipitation when available for all selected cities. This weather data was collected from the rnoaa R package (Chamberlain, 2021) for all cities included in the model during which the closest weather station has been in operation. In cases of missing data we used three methods. First, if an entire month was missing data at a given location, this was left as missing data. Second, if for a given day at a given location, only one temperature (minimum, maximum, average) value was missing a linear regression was used to estimate the missing value. This linear regression was created using all temperature data collected from rnoaa for a given country. In the third case, where we were missing two or all three of the temperature values, a moving average 4 was used with exponential weights to impute the missing values. The missing precipitation values are treated as 0.

Model for 2022 prediction

Once we had a complete set for the weather data, we created new variables based on this data. A list of created variables is given in github repository (https://github.com/leeparkuky/peak-bloom-prediction/blob/main/data/README.md) .

After creating these variables, we used the given data (including latitude, longitude, and altitude) and the weather variables to create a model to predict bloom day of year. The feasible solution algorithm (Lambert et al., 2020), was used on all variables to search for interaction terms. The feasible solutions algorithm is a fast, flexible algorithm that stochastically searches for variable combinations that will improve the predictive quality of a specified model, if possible. With the interaction terms, we plotted the data to understand and visually ensure the effectiveness of interaction terms.

Once we had a complete set of variables and interactions, we explored different types of models, including simple linear regression with backward elimination, Lasso, Ridge regression, linear regression with PCA, tree regression with and without bagging or boosting, MARS (Multivariate Additive Regression Splines) and MARS with adaptive boosting. Linear models use the standardized data while other methods do not.

After testing all these types of models, we chose to use Ridge regression for our final prediction with an alpha value of 0.251. After selecting a model we cross-validated with 100 replications; each replication randomly selected 70% of the data for the training set and 30% of the data for the test set. The mean and variance of RMSE can be found at https://github.com/leeparkuky/peak-bloom-prediction/tree/main/models.

Model for 2023-2031

Given many of the variables included in the 2022 prediction model were based on temperature and long term temperatures are difficult to predict, we chose to use different but simpler models to predict bloom DOY from 2023-2031 (Figure 1). For Kyoto, Liestal, and Washington D.C. we fit the quadratic regression model with year and square of year for each individual location. For Vancouver, since their historical data are not provided, we used year, latitude, longitude, and altitude and fitted them into a Generalized Additive Model (GAM). The Generalized Additive Model performed better than simple linear model including features, year, square of year, latitude, longitude and altitude.

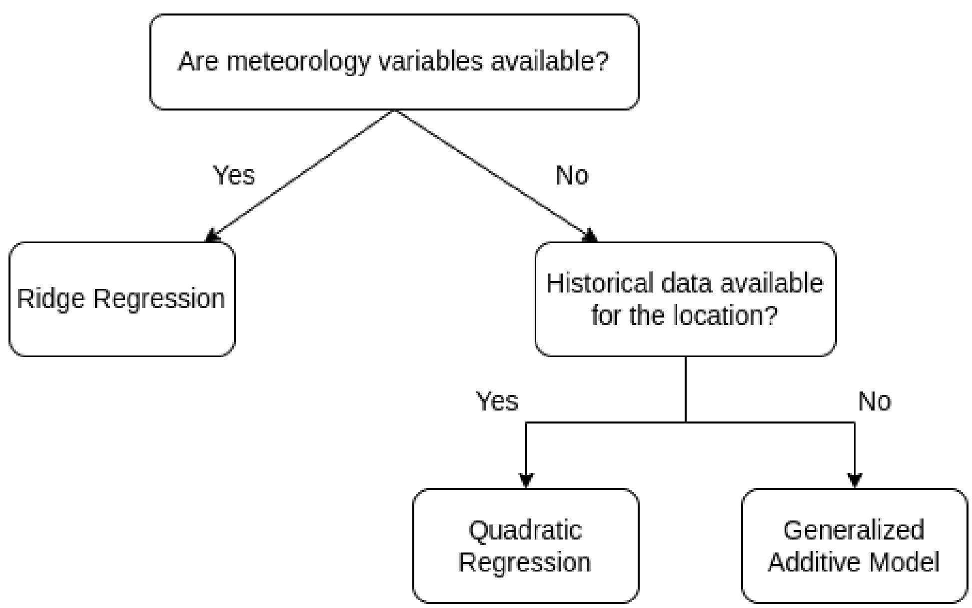

Figure 1: Model selection criteria based on availability of data

Results and Discussion

Model for 2022 prediction

We found a number of interaction effects that could be researched further. One noticeable interaction term is an interaction between the number of days with temperatures below five degrees Celsius and the average temperature in January. When the average temperature in January is low, increasing the number of days below five degrees Celsius, causes a small decrease in bloom DOY. However when the average temperature in January is higher, increasing the number of days below five degrees Celsius, causes a much steeper decline in bloom DOY when all other conditions are held constant.

Another important interaction term in the model was the number of days with an average temperature below five and precipitation amount in February. If there is low precipitation in February, as the number of days with temperatures below five Celsius increases, the bloom DOY increases sharply. If there is high precipitation in February, then the effect of the number of days with temperatures below five Celsius influencing bloom DOY becomes less, when all other factors are held constant.

Model for 2023-2031 prediction

When predicting bloom DOY for the years after 2022, the biggest challenge we faced is absence of meteorology data. General trends in meteorology are available, but they are not stable and accurate when it comes to monthly temperatures and precipitation. As climate change continues, winters are expected to become warmer in many regions, but it is not always the case as Korea has been suffering from colder winters. Therefore, we decided not to predict weather but rely on the annual trend of each location. We have historical data for Washington DC, Kyoto, and Liestal. For those cities, we fit a quadratic regression between year and bloom DOY to predict the remaining nine years. Table 1 below lists the coefficients of the linear model.

Table 1. Coefficients of quadratic regression between year, squared year, and bloom day of year

	Kyoto	**Liestal**	**Washington DC**
$Intercept$	-81,311.561	-123,821.679	-63,796.316
$Year$	81.295	123.597	63.749
$Year^2$	-0.020	-0.031	-0.016

However, in the case of Vancouver, historical data was not available, so we used all the historical data for all the cities with variables, including latitude, longitude, altitude and year, to fit a GAM. The predicted values and its confidence intervals when each feature are fit to GAM individually are shown in Figure 2.

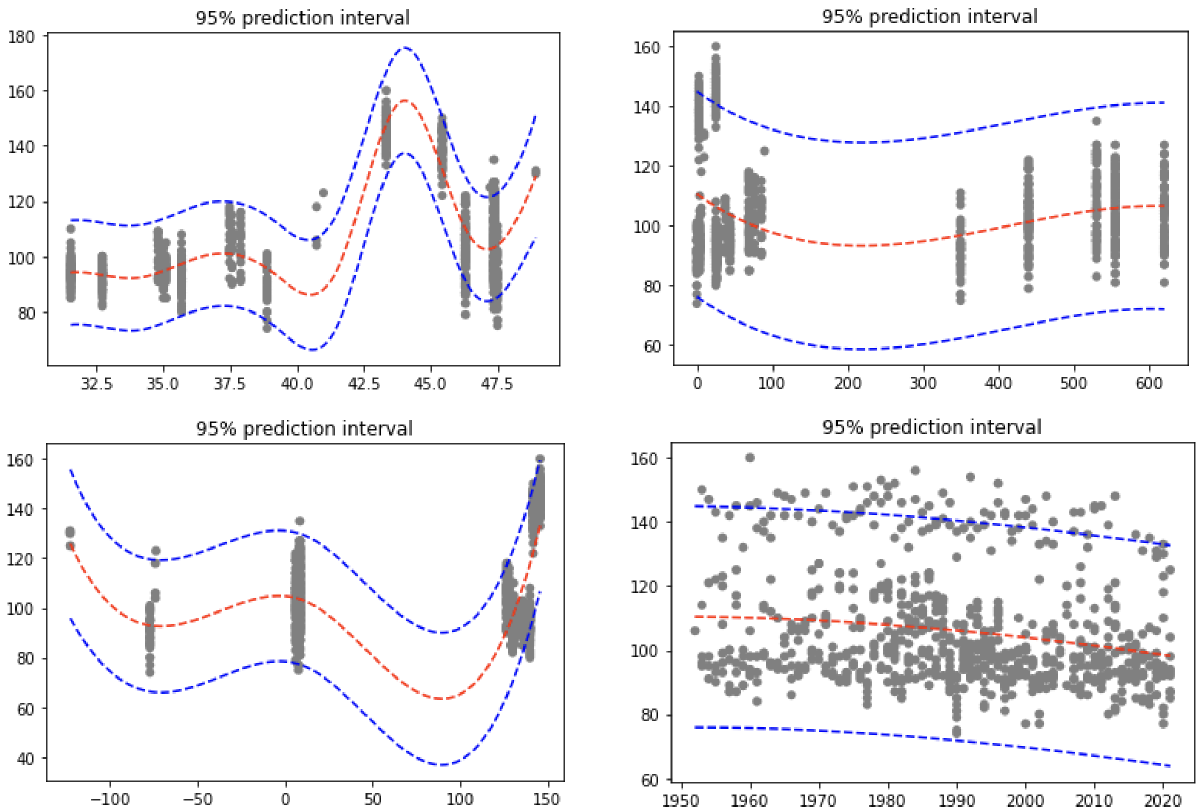

Figure 2: GAM fit between bloom DOY and latitude (top left), altitude (top right), longitude (bottom left) and year (bottom right).

Conclusion

We used Ridge regression to predict cherry tree bloom DOY. We found several factors and interactions related to winter and spring temperatures that are related to bloom DOY. Precipitation was found important, but this may be a lurking variable for cloud cover or temperatures. Geography is also important in when cherry trees bloom. We found that cherry tree blossom timing relies on many different factors and may become more challenging as we face more volatile and unpredictable meteorological conditions going forward.

References

Chamberlain, S. 2021. rnoaa: 'NOAA' Weather Data from R. R package version 1.3.8. https://CRAN.R-project.org/package=rnoaa

Chung, U., Mack,, L., Yun, J.I., and Kim, S.H. 2011. Predicting the Timing of Cherry Blossoms in Washington, DC and Mid-Atlantic States in Response to Climate Change. PLoS ONE, vol. 6, no. 11, p. e27439.

Heide, O.M. 2007. Interaction of photoperiod and temperature in the control of growth and dormancy of Prunus species. Scientia Horticulturae 115 309314.

Lambert J., Gong, L., and Elliott, C. 2020. rFSA: Feasible Solution Algorithm for Finding Best Subsets and Interactions. R package version 0.9.6. https://CRAN.R-project.org/package=rFSA

Miller-Rushing, A.J., Katsuki, T., Primack, R.B., Ishii, Y., Lee, S.D. and Higuchi, H. 2007. Impact of global warming on a group of related species and their hybrids: cherry tree (Rosaceae) flowering at Mt. Takao, Japan. Am. J. Bot., 94: 1470-1478.