Ha Dao

Dec 14th, 2022

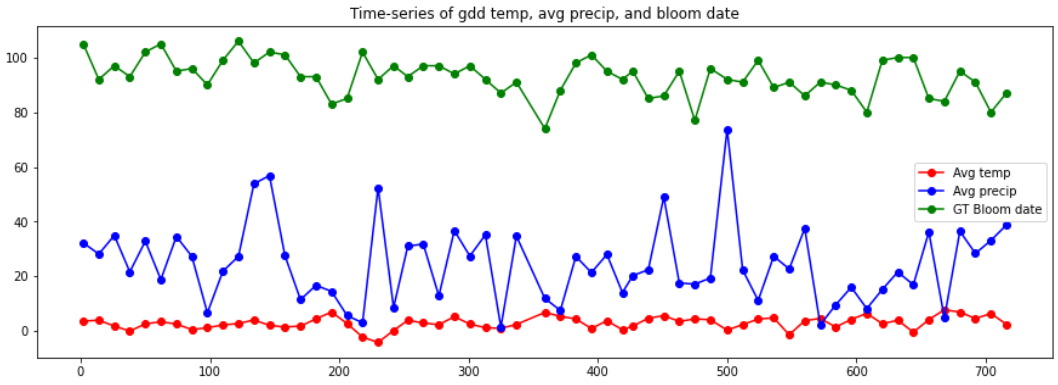
Hdao5

Cherry Blossom Prediction

Plant phenology, including blooming, leaf growth, leaf coloration, and defoliation, is strongly influenced by the environment and climate (Masago et al., 2022). Cherry trees, in particular, have experienced earlier blooming in the past decade, making them a sensitive indicator of climate change. Given its enormous environmental, economic, and cultural significance, accurately determining the peak blooming date of cherry trees can provide a theoretical foundation for tourism administrators and travelers to plan their activities. Using extreme gradient boosting (XGBoost), our main goal is to accurately estimate the peak blossoming date of cherry trees using historical sequential temperature information.

Cherry Blossom - the subject of our project, is found throughout the northern hemisphere. Our study gathered peak bloom date information from four locations: Kyoto, Washington D.C, Vancouver, Liestal and obtained climate data from the National Oceanic and Atmospheric Administration via weather stations located near each city. We calculated the accumulated growing degree days (GDD) by analyzing temperature data from the months of December of the previous year, as well as January and February of the current year using the formula from USA-NPN:

$$\sum \frac{T_{max}+T_{min}}{2} - T_{base\ (0)}$$



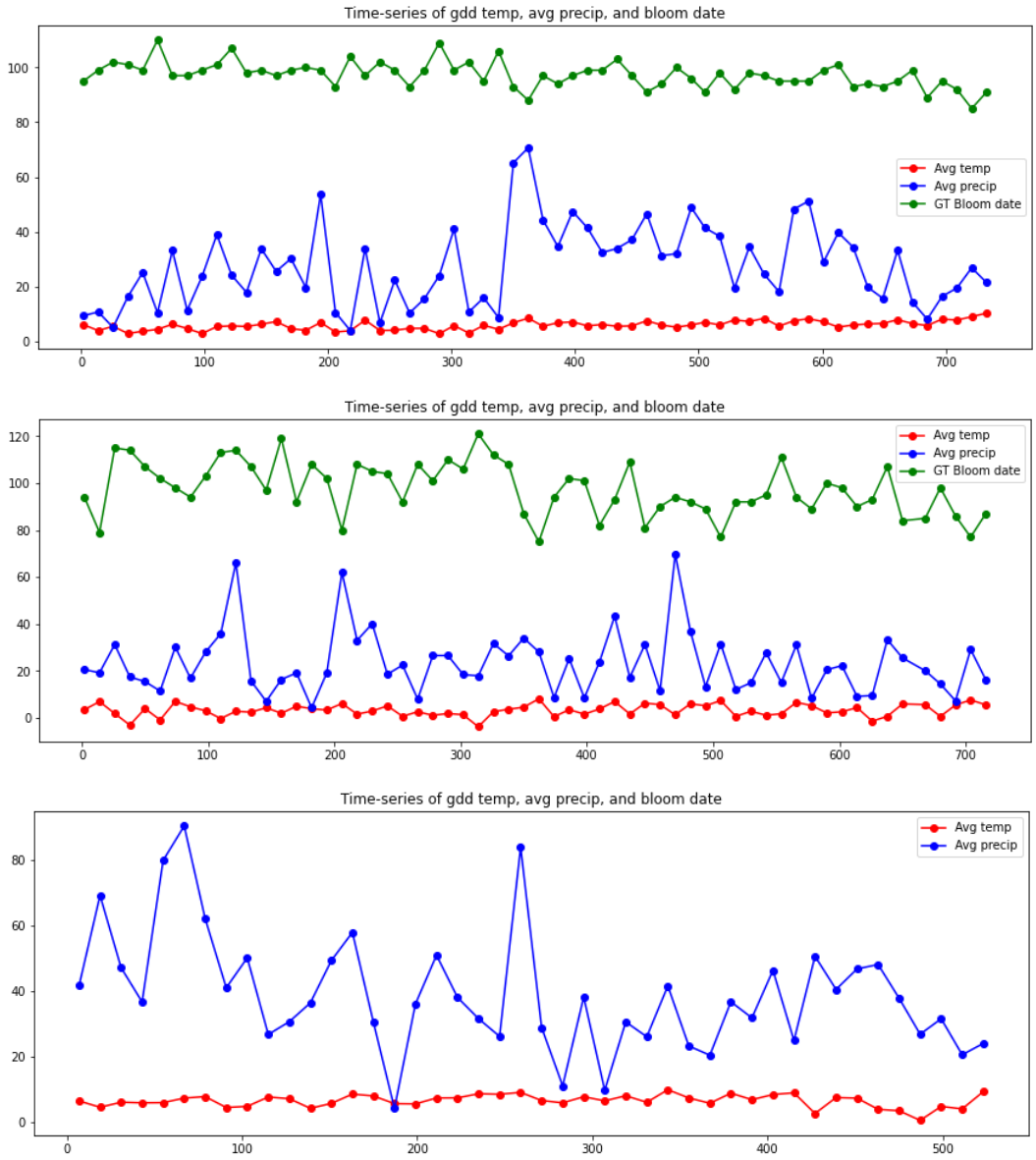


Figure 1: Time series of gdd_temp, avg_precip, and bloom_date of Washington D.C, Kyoto, Liestal-Weideli, and Vancouver

To account for the unpredictable nature of weather data, we made the assumption that our predicted bloom date could deviate by as much as plus or minus 4 days. With this in mind, we divided the bloom_doy into categories of equal size and generated two new categorical features. This enables us to observe how the bloom date remains relatively stable from year to year, despite potential variations.

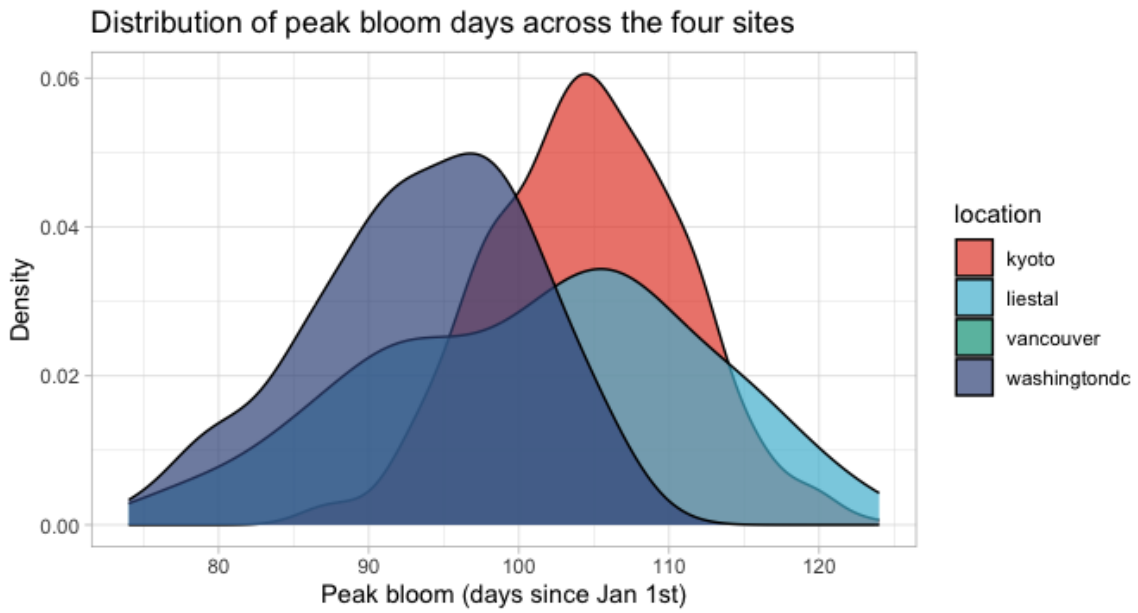


Figure 2: Distribution of peak bloom days across the four sites

Based on Figure 2, it is evident that the distribution of peak bloom days varies considerably among the four locations. As a result, a general model to quantify peak bloom days may not be practical, and individual models should be developed to characterize this variation. Peak bloom days in Kyoto exhibit a higher degree of concentration while comparing to the other three locations, the days are more widely dispersed. The observed difference may be attributed to dissimilarities in location and climate characteristics. Missing data is imputed using the K-Nearest Neighbors (KNN) algorithm, identifying rows in the dataset that were similar for each missing data point and treating them as neighbors to impute missing values. This technique allowed us to create a complete dataset and reduce the number of missing values. We then used the Extreme Gradient Boosting (XGBoost) algorithm to predict the peak bloom date of cherry trees.

The results showed that our best-performing model, XGBoost, achieved an MAE of 7.7 across the four sites on the test data, with an average MAE for each location of 5.422 (Kyoto), 10.527 (Liestal), 7.151 (Washington DC), and 9.23 (Vancouver). The accuracy of our predictions varied across the four sites, with the Kyoto dataset achieving the highest accuracy and the Liestal dataset achieving the lowest accuracy. One of the key findings of our study is that the bloom date does not change drastically year by year, which allowed us to include bloom date categories from the last 2 years as good features for prediction. This is an important insight for future research on cherry blossom peak bloom prediction.

In conclusion, our study provides a practical and accurate method for predicting cherry blossom peak bloom dates using historical temperature data and machine learning algorithms. This information can be useful for tourism administrators and travelers, as well as for understanding the effects of climate change on plant phenology. Nonetheless, our projections were constrained as we lacked additional relevant features, such as measurements of relative humidity, solar radiation, and wind speed. Incorporating these variables may enhance the performance of our model. Furthermore, employing more advanced neural network models and other supervised machine learning techniques could lead to further improvements in the accuracy of our forecasts.

Reference:

Masago, Y., & Lian, M. (2022). Estimating the first flowering and full blossom dates of Yoshino Cherry (Cerasus × yedoensis ‘Somei-Yoshino’) in Japan using machine learning algorithms. *Ecological Informatics*, *71*, 101835. https://doi.org/10.1016/j.ecoinf.2022.101835