

# When will the Cherry Trees bloom?



*Done by*  
**Leung Hoi Kit Alvin**  
**Aaron Lee Wei Qi**

*From the National University of Singapore*

## *Introduction*

In order to answer the question of, “When will the Cherry Trees bloom?”, we decided to approach the problem from a time series approach. A time series is a series of data points indexed based on linear time order. Alternatively, a time series is a sequence taken at successive spaced points in time. Given this definition, our cherry blossom bloom timing data can be considered a time series and time series forecasting can be used to try and forecast future values.

## *Methodology (SARIMA)*

Autoregressive Integrated Moving Average, or ARIMA, is one of the most widely used forecasting methods for univariate time series data forecasting. It belongs to a family of models that aim to predict future values based on previously observed values while taking into account both overall trends in the data and seasonalities present.

In order to configure an ARIMA model, we have to find out what hyperparameters are suitable to model both the trend elements of the underlying time series, which in this case is the number of days after the start of the year we would predict the cherry blossoms to bloom in the different locations.

We choose 3 hyperparameters for the trend elements. The first of them being the trend autoregressive order, which is the number of immediately preceding values in the time series that are used to predict the value at the present time. The second hyperparameter is the trend difference order. This is the number of differencing procedures required to get a stationary series from the original time series data, where differencing is the transformation of the series to a new time series where the values are the differences between consecutive values of the series. Finally, the moving average order is the series of averages of different subsets of the full data set in order to smooth out the influence of outliers.

To build our models for the 3 locations (Washington, Kyoto, and Liestal), we decided to train an ARIMA model individually for each, plotting the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of their cherry blossom bloom timing to find the trend autoregression order ( $p$ ) and trend moving average order ( $q$ ). Meanwhile, we used differencing to find the appropriate number of unit roots to see if a trend was present in the data.

This allowed us to build a base model for each location which could be used to forecast future predicted blossom timings.

## *Introducing Exogenous Variables (Temperature)*

To attempt to better account for some residuals and further improve our model, we decided to also include temperature data over time from weather stations near the chosen locations in order to see if such information can help explain the variance observed between our predictions and what is actually observed better.

Conceptually we would expect that this variance could be caused by differences in temperature, as according to Naoko Abe, author of *The Sakura Obsession*, cherry trees need a full month of chilly weather below 41 degrees to properly blossom when it gets warmer (Burnett, 2021). Therefore, we decided to extract the max temperatures of each location, and obtain the average temperature in each season, given that it is the variable with the highest likelihood of having an impact on the predictions.

After extracting the information regarding max temperatures for each location, the data has been split into the average max temperatures of each season. We only include the winter season in our analysis, as we realize that cherry trees tend to blossom at the start of spring. To account for the earlier blooming, we choose the winter period as it corresponds to the time right before the blossoms bloom. We then include the maximum average temperature in spring as exogenous variables in our SARIMA model, improving it to a SARIMAX model to predict the blooming date at each location for the following 10 years.

## *Extrapolating to Vancouver*

For our predictions for Vancouver, we chose to simply duplicate the values for Washington DC as opposed to averaging the values for the 3 locations. This is because based on our research we believe that the blossom timings are impacted by location, and in the absence of external information regarding the local conditions at Vancouver we decided that a good approximation would be the predicted timings of the closest ascertainable location. Furthermore, both locations have similar species of Cherry trees, both being of the *Prunus × yedoensis* cultivar.

With this in mind choosing the values for Washington makes more sense than using information from Kyoto and Liestal as well since Washington is much closer to Vancouver than the other 2 locations.

## *Final Predictions*

Our final predictions for the predicted blooming date at each location for the following 10 years can be found in the table below:

| Year | Kyoto | Liestal | Washington DC | Vancouver |
|------|-------|---------|---------------|-----------|
| 2022 | 97    | 95      | 89            | 89        |
| 2023 | 91    | 93      | 99            | 99        |
| 2024 | 98    | 91      | 95            | 95        |
| 2025 | 96    | 93      | 92            | 92        |
| 2026 | 90    | 92      | 97            | 97        |
| 2027 | 95    | 92      | 92            | 92        |
| 2028 | 95    | 92      | 92            | 92        |
| 2029 | 94    | 92      | 91            | 91        |
| 2030 | 92    | 92      | 85            | 85        |
| 2031 | 93    | 92      | 93            | 93        |

## *Citations (APA)*

- paylessimages. (2022). Cherry Blossom [Photograph]. [www.Gardenia.Net](https://www.gardenia.net/storage/app/public/guides/detail/40380018_mOptimized.jpg).  
[https://www.gardenia.net/storage/app/public/guides/detail/40380018\\_mOptimized.jpg](https://www.gardenia.net/storage/app/public/guides/detail/40380018_mOptimized.jpg)
- Burnett, E. (2021). What The Cherry Blossom Bloom Can Tell Us About Climate Change. *National Public Radio*. Retrieved from <https://www.npr.org/>