Cherry Blossom Prediction Competition

Douglas Nedza

March 1, 2022

The blossoming of cherry trees are archetypical of phenological study, and records of the date of blossoming are often studied in relation to the spring temperatures preceding. This relationship with temperature is bi-directional, with some studies using these blossoming dates as a proxy for reconstructing spring time temperatures in Japan as far back as the 9th century (Aono 1997). Other studies have focused on using observed or projected temperature to estimate the date of cherry tree blossoming(Beaubien & Hamann 2011, Chung et al 2011, Chung et al 2016).

For cherry trees, the phenology of spring blossoming is considered primarily driven by the aggregation of warming temperatures. This relationship can be estimated with the appropriately named thermal-time model. When the temperature rises above a species specific threshold, the thermal time begins to aggregate in proportion to amount the temperature exceeds the threshold. This model assumes that when the necessary thermal-time is achieved, the cherry tree will blossom. Previous studies employing this model demonstrate success at reconstructing observed cherry blossoming dates using observed temperatues (Chung et al 2011).

Due to the cherry blossoms sensitivity to temperature, particularly close to the date of blossoming, using the thermal-time model as a predictive tool would be difficult without accurate forecasts of temperature at least one month in advance. In contrast to the explicit thermal-time model, this study approaches predicting cherry blossom dates by using larger scale features, such as local seasonal and monthly average temperatures, as well as indices representing modes of climate variability such as El Nino and the North Atlantic Oscillation. While these predictors can not resolve temperature on the daily resolution required by the thermal-time approach, climate modes can exhibit a modulating influence on seasonal temperatures, thereby containing some information about how quickly heat will aggregate and cherry trees will produce and open their blossoms.

In addition to modulating effects of spring temperature, the possibility exists that climate conditions during one year could influence blossoming in the subsequent season. Phenological studies have highlighted the importance of sufficiently cold temperatures and the associated dormant state that occurs in association to be critical in the development of blossoms in the following spring. Information on seasonal temperatures and climate modes from the previous year are included in this study to assess any predictive potential.

Under historic climates, seasonal variability in temperature was the primary modulator of cherry blossom date. With today's changing climate, an increasing overall temperature is likely exhibiting influence on many species' phenology (Primack et al 2009, Beaubien & Hamann 2011). It is possible that the recent, unprecedented early cherry blossom dates were driven by underlying global warming related to anthropogenic emissions. The impact of an increasing trend in temperature, particularly when predicting beyond the current season, must be accounted for. To produce projections of seasonal and monthly temperature, a linear trend is fit to recent observations. This linear trend can then be extrapolated into the future providing a guiding reference temperature even a decade into the future.

Thus far, many potential predictors have been introduced that may contain predictive information about the blossoming date of cherry trees. Each of these predictors can be justified with a physical argument, and further obscuring the choice, many of these predictors are correlated. In addition, due to the limited number of observations and the large number of predictors, overfitting to the training data is a concern, potentially harming predictions made on a year outside of the training set. For this study, a regularized regression method, LASSO, is used to assist in the selection of predictors based on the data we do have. LASSO is an extension of standard linear regression that incorporates a regularization parameter that attempts to combat overfitting (Tibshirani 1996). One of the ways the LASSO algorithm is helpful in cases of many predictors is the ability for the algorithm to remove less useful predictors from the analysis by setting the associated coefficient to zero. As a result, only a few variables are used when making predictions.

When predicting within the season (2022 for example), the regression model selected still has several predictors, including seasonal and monthly temperatures and climate indices. However, as the lead time of the prediction is increased, the predictor set becomes smaller as the last observed data and climate modes become irrelevant to cherry blossoms. For example, the winter temperature of 2022 may be important when estimating the cherry bloom date in 2022, but is far less important when predicting the cherry bloom date in 2023 or later. Even climate modes, which have the potential to endure for months or years, quickly become irrelevant when prediction future cherry blossom dates. In contrast, the linear fit to the observed temperatures continues to be included in the predictor set even when predicting 9 years into the future.