I have left a lot of comments here. Please don’t take it personally, I’m just trying to help you out. If you take it with an open mind, your paper and your writing will truly benefit. Best, your classmate.

Predict “peak bloom date” for cherry trees in 4 different locations

# Summary

Because cherry blossoms don't last long, forecasting the number of cherry blossom days is very important for a country to attract tourists. most of them propose that temperature is the main effect of cherry blossom bloom. Other studies have suggested that other climatic variables such as [blah, blah, blah] may play a role in the timing of cherry tree blossoming.[citation] It is of interest to investigate whether the timing of peak bloom can be accurately predicted from weather and climatic variables. Data on weather and the timing of peak bloom was collected for three different sites: Kyoto, Japan; Liestal, Switzerland; and Washington, DC, USA. To predict the timing of peak bloom for the three locations, non-seasonal autoregressive integrated moving average (ARIMA) and time series linear models (TSLM) were fit to the data. Forecasts for peak bloom dates for 2022-2032 were extrapolated from the best fitting models using weather data forecasted from seasonal ARIMA models (SARIMA) fitted to the historical data for our covariates. Predictions for the day of peak bloom were also made for a fourth location (Vancouver, BC, Canada) where historic data for peak bloom dates were not available. Forecasts for Vancouver were made by extrapolating a naïve ARIMA model from historic data from the other three locations.

To prevent our model from overfitting the data, , we split into 90-10 training-testing dataset for weather and bloom day of the year. seasonal autoregressive integrated moving average models, we can forecast the bloom day of the year for Washington DC, Kyoto, and Liestal by non-seasonal autoregressive integrated moving average (ARIMA) models or time series linear models (TSLM). For Vancouver, historic data for cherry blossom peak bloom days were not available., Forecasted bloom dates for Vancouver were extrapolated from an ARIMA model and a naïve model fit to the data from the other three locations.

# Introduction

The impact of climate change has attracted the attention of academics worldwide. Climate-related changes in the sequence of environmental events, such as the timing of snowmelt at high altitudes , have had major implications for some annual herb flower species. There is evidence that climate change has affected the timing of flowering phenology. 78 percent of all studies in 21 European countries indicated earlier flowering, with an overall phenological activity movement of 2.5 days per decade. For flowering cherry trees in Kyoto, Japan, the peak date has been gradually shifting from mid-April to early April. Cherry trees in Washington, DC are blooming six days earlier than they were a century ago. Temperature collected from weather stations in Washington, DC have risen by 1.6 degrees Celsius (2.8 °F). In Liestal, Switzerland, the timing of cherry tree flowering occurred between early and mid-April fairly consistently until 1990. But since the 1990s, the timing of flowering has been occurring earlier and earlier. The timing of cherry tree flowering is highly sensitive to [changes in / rise in] temperature, especially during the winter and early spring. The timing of cherry tree blossoming is a useful indicator of the effects of climate change on flowering phenology. From a cultural standpoint, being able to predict the precise timing of cherry blossoming is critical, as many festivals across the world revolve around the specific timing of this phenological event, and flower watching tours are in high demand. In Japan, there is a saying "seven days of cherry blossoms," which is to say that cherry blossoms take only about 7 days from the time it opens to the time it fades.

Accurate prediction of the flowering period for ornamental plants can provide guidance for public travel arrangements.

Presently, predictions are made for the timing of cherry tree blossoming through observations blah, blah, blah. Prediction for the timing of cherry blossoming could perhaps be improved with statistical modeling. In this paper, we have proposed both site specific and naïve statistical models for predicting the timing of peak bloom for cherry trees in four locations: Kyoto, Japan; Liestal, Switzerland; Washington, DC, USA; and Vancouver, BC, Canada.

**Data**

Peak bloom dates from previous years have been provided for three locations: Kyoto, Liestal, and Washington. In our dataset, historic peak bloom dates range from 1921-2021for Washington, from 1888- 2021 for Kyoto, and from 1894-2021 for Liestal. For Vancouver, no historic data was available. We used the bloom day of the year as the dependent variable for our model. Data for covariates were collected through the National Oceanic and Atmospheric Administration (NOAA) website (http[s://w](http://www.noaa.gov/))ww.n[oaa.gov/)](http://www.noaa.gov/))  from weather stations with similar latitude, longitude, and altitude as the precise location of the cherry trees in each of the cities observed in our study. The data set we found has 26 variables, including temperature, extreme temperature (), (), and precipitation ().

The main goal of this study was to forecast the peak bloom days of the year from 2022 to 2033 in four locations: Washington, DC; Kyoto, Japan; Liestal, Switzerland; and Vancouver, Canada.

CovariatesBecause the timing of cherry blossom blooming is highly sensitive to temperature, we chose several temperature- related covariates. Due to the high number of missing values for mean temperature (TVAG), minimum temperature (TMIN) and extreme minimum temperature (EMNT) in some locations, we were not able to use these variables in our analysis. We instead chose extreme maximum temperatures (EMXT) and maximum temperatures (TMAX) as temperature covariates as the data was more complete for these variables for the four locations of interest.

We split the data into roughly three groups based on historical peak bloom dates (i.e., early bloom date, regular bloom date & late bloom date). The average monthly TMAX for each of the three groups was calculated. Fig. 1 shows monthly maximum temperature over time for each of the three locations. Overall, the trend was similar in Washington DC, Liestal, and Kyoto. The lowest TMAX occurs in January, followed by a slight increase in February. There is a jump in TMAX between February and March. Of the three bloom date groups, the early bloom date group has the highest TMAX in February and March. As a result, we believe that TMAX is correlated with peak bloom date. We conducted the same analysis on extreme temperatures and found that- while the overall difference between the three bloom date groups was not as significant as the average temperature, extreme temperatures have an impact on blooming time. Differences between monthly extreme temperature for Kyoto, Washington, and Liestal are shown in Fig. 2.

Figs. 3-5 show correlation matrices for the dependent variable and covariates included in models for each of the three bloom locations. The plots show March maximum temperature (TMAX\_3) is negatively correlated with peak bloom date in Washington,DC (r = -0.74), Kyoto (r = - 0.77), and Liestal (r = -0.7), while in Washington, DC December extreme temperature (TMAX\_12) is positively correlated with peak bloom date (r = 0.2). Because March extreme temperatures and December temperatures were found to be the two variables that accounted for the maximum amount of variability in bloom times, these two variables were selected as covariates for our study.

# Preprocessing

Data for covariates were obtained for Washington DC from 1945–2021, for Kyoto from 1951–2021, and for Liestal from 1901–2021. Historic peak bloom data were used only for years for which covariate data was available. For Washington, DC, there were missing values for TMAX and EMXT for January and February 2022, so data for these two months were dropped from the analysis for DC. For Liestal there were missing values for January 2022 and data for this month was dropped from the analysis. There were also missing values for January 2017 in the Liestal dataset but these missing values were handled using the tsclean() function in the forecast package in R. For Kyoto missing values for 1978-06, 2004-07, 2004-08, 2011-01, 2012-03, 2012-11, and 2012-12 were similarly handled using tsclean(). tsclean() is a function that identifies and replaces outliers and missing values in a time series with linearly interpolated values. It does allow us to utilize forecasting models that are sensitive to outliers or do not handle missing values.

The datasets for each of the three bloom locations were split into 90% training /10% testing datasets. The only covariates used in the final analysis were TMAX from March and EMXT from December .

# Methods

**Autoregressive integrated moving average models (ARIMA)**

ARIMA is a statistical model for time series data. ARIMA models can be used either to better understand data through inference or to predict future trends (forecasting). ARIMA with p, d, and q is a standard notation for ARIMA models, where integer values replace the parameters to indicate the kind of ARIMA model utilized. Here p is number of autoregressive terms (AR order), d is number of non-seasonal differences, and q is number of moving-average terms (MA order). An ARIMA model is denoted by the following notation: ARIMA(p,d,q). The auto.arima() function in R uses a variation of the Hyndman-Khandakar algorithm (Hyndman & Khandakar, 2008), which combines unit root tests, minimisation of the AICc and MLE to obtain an ARIMA model. The arguments to auto.arima() provide for many variations on the algorithm.

**Seasonal Autoregressive Integrated Moving Average (SARIMA)**

SARIMA models are an extension to ARIMA for univariate time series data with an explicit seasonal component. SARIMA models have up to three additional hyperparameters for modeling the seasonal component of the time series. Seasonal components in SARIMA models can be either autoregressve (AR), differencing (I), or moving average (MA), as well as an additional parameter for the period of the seasonality. There are four seasonal elements that must be setup that are not part of ARIMA: the seasonal autoregressive order (P),the seasonal difference order (D), the seasonal moving average order (Q), and the period (m). A SARIMA model is denoted by the following notation: SARIMA(p,d,q)(P,D,Q).

**Time series linear model (TSLM)**\

TSLMs are used to fit linear models to time series data and include components for both trend and seasonality. TSLMs are similar to the commonly used lm() function for linear models, but tslm() has more features for dealing with time series.

**Ljung-Box Test**

The Ljung-Box Test is used to check if autocorrelation exists in a time series. The Ljung-Box Test statistic is

"

𝑞 = 𝑛(𝑛 + 2) ∙ ) 𝜌+(𝑗)!/(𝑛 − 𝑗)

#$%

where 𝑛 is the number of observations and 𝜌+(𝑗) the autocorrelation coefficient in the sample with a lag of 𝑗. LSTS\_lbtp computes 𝑞 and returns the p-values graph with lag 𝑗.

**Forecasting Covariates for the Fitted Models**

A SARIMA model was utilized to forecast TMAX for March and EMXT for December to 2032 using historic data for total annual TMAX and EMXT. Two models were fit to the data. The first model (Covariate Model 1) was fitted using the auto.arima() function in R. Model 2 was fitted by checking the ACF plot. The Covariate Model 2 was judged to produce better results, so fitted values from Covariate Model 2 were used to predict weather data for TMAX and EMXT.

Forecasted values of TMAX and EMXT were subsequently applied to six different models to predict the date of peak bloom for cherry blossoms in Washington DC, Kyoto, and Liestal. The first model that we fit (Forecast Model 1) was an ARIMA without any covariates. The second model (Forecast Model 2) was a model fitted with TMAX in March and EMXT in December as covariates. The third model fit was an ARIMA with TMAX 3 alone as the covariate (Forecast Model 3). Forecast Model 4 was fit to an ARIMA model with EMXT alone as the covariate. The fifth model (Forecast Model 5) was a TSLM with covariates for March TMAX, December EMXT, and a covariate for trend. Lastly, Forecast Model 6 is a TSLM with March TMAX and a covariate for trend.

For Vancouver, since historic peak bloom data were not available for this location, a different methodology had to be employed to predict the timing of cherry blossom bloom. To predict peak bloom dates for Vancouver, a linear model was fit to a combined dataset from all bloom dates and covariates from Washington, DC, Kyoto, and Liestal. We see that both predictors TMAX\_3 and EMXT\_12 are significant. After fitting a linear model to the combined dataset, a time series model was used to forecast the value of these predictors for Vancouver.

# Results

The ARIMA\_auto model had wide confidence intervals for future times after predicting and forecasting TMAX and EMXT for DC, as shown in Fig. 6 and 7 respectively. By contrast, the ARIMA\_manual model had much narrower confidence intervals, but as shown in Fig. 6(b), some of the historic data values for TMAX (black line)—such as TMAX for 2017—did not fall within the predicted confidence interval. Despite the occasional deviations from the predicted confidence intervals, we chose to use outcomes from the ARIMA\_manual model for our predicted weather data, because it did not have wide confidence intervals in the future. We also used the same methods for Kyoto and Liestal to obtain forecasted values for TMAX and EMXT.

In Table 1, we can see that the TSLM models (AIC: 206.83 and 209.44) outperformed the ARIMA models for Washington, DC. However, the residuals from the TSLM models reject the null hypothesis of the Ljung-Box test in Table 2, implying the residuals from these models are serially correlated. As a result, it was only possible to use the ARIMA models for the Washington DC dataset. The best fitting ARIMA model for DC was ARIMA\_with\_covariates (AIC: 388.29), so this was the model selected for forecasting peak bloom dates for DC.

For Kyoto and Liestal, TSLMs also outperformed the ARIMA models.. The Ljung-Box test statistic was not significant at the 0.1 level for both Kyoto TSLM models as well as the Liestal TSLM model with covariates. As a result, we used TSLMs for forecasting for Kyoto and Liestal. . On the contrary, the ARIMA model with EMXT covariates (BIC: 367.91; 820.68) for the other two location datasets are only slightly worse than the ARIMA models without any covariates (BIC: 364.16; 818.76), we also remove the models for these two locations.

Using the predicted weather data as the covariates, we predicted the peak bloom date and compared the results. A comparison of the results from each of the selected models is shown in Table 3. For DC, ARIMA with TMAX and EMXT covariates had the lowest RMSE (5.02) and MAPE (4.27) for predicting the date of peak bloom for cherry trees. For Kyoto, TSLM with the TMAX covariate has the smallest RMSE (5.02) and MAPE (4.27) when we predict blooming days during the year from 2014 to 2021 as shown in Table 3. Lastly, ARIMA with TMAX covariate fits the Liestal testing dataset well, with a RMSE of 8.40 and a MAPE of 8.09. In Fig. 8-10, the ARIMA models with no covariates have wider confidence intervals than the other models for the three locations. We find that our model does not capture some historic peak bloom times where flowering was earlier in the year than our model predicted. As can be seen from Fig. 9, peak bloom dates for Kyoto in 2018 and 2020 are not within the 95% confidence interval for all of the models in Fig. 9. Using the best model for each location, peak bloom day was forecasted from 2022 to 2033 (Table 4).

For Vancouver, Fig. 14shows that a as time passes, TMAX\_3 increases. In Fig. 15, here is no apparent trend that could help predict the value of TMAX\_3. We build two models, one Naive model, and one ARIMA model. Fig, 16 and 17 show that the ARIMA model had both a lower standard deviation and a better residual plot compared to the Naive model.Using the ARIMA model, we predicted TMAX\_3 to be 50.

Similarly Fig. 18shows a positive trend between EMXT\_12 and time and Fig. 19 shows not trend in EMXT\_12 over time. We compared two-time series models to predict EMXT\_12: one Naive model and one ARIMA model. We see that the ARIMA model had both a lower standard deviation and a better residual plot (Fig. 20 and 21). So, using the ARIMA model, we predicted the EMXT\_12 to be 51.

# Conclusion

Table 4 shows our forecasts for the number of days into the year the cherry blossoms are expected to reach full bloom according to our models. For DC, the average peak bloom day is around 91 days which is earlier than the other locations. Kyoto has the narrowest confidence interval (about 11 days for 95% CI), and Liestal has the widest confidence interval (about 32 days for 95% CI). For Vancouver, our model forecasts the peak bloom as 104 days into the year.

The days of peak bloom are difficult to predict. This is because weather is unpredictable. For our models, the temperature we predict is not high enough for warmer years, which makes our prediction for flowering days inaccurate.

Table 5 shows that when we utilize the same models with real weather datasets, all models give much lower RMSE and MAPE for prediction than models with predicted weather covariates. We found that ARIMA with TMAX and EMXT covariates was the best model for DC, Kyoto, and Liestal, because MAPEs are the lowest across all models for those cities in Table 5, and the 80% confidence intervals including most real test data are shown in Fig. 11-13. If we can predict the weather well in the future, then ARIMA models with TMAX and EMXT covariates will be proper for predicting of the peak bloom date for cherry blossoms.

Several methods for forecasting the weather dataset were also tested, including Holt-Winters, Double Seasonal Holt-Winters and Anomaly Detection, but these models did not produce good results. A vector autoregressive model was also fitted to the data, but the covariates were not stationary, so this method could not be applied in our analysis. Even though we find the temperature dataset from the Copernicus website (https://cds.climate.copernicus.eu/). The temperature dataset from 1950 to 2100 by using the climate models for Washington DC. However, there is a weak correlation between temperature and the bloom day of year.

Hence, only if we can find a good forecast climate method to forecast temperature, we can easily forecast the cherry blossom days of the year using an ARIMA model with monthly maximum temperature from March and extreme maximum temperature from December.

# Reference:

David W. Inouye. Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. 01 February 2008, doi: 10.1890/06-2128.1.

P. Q. Craufurd., T. R. Wheeler. Climate change and the flowering time of annual crops. Journal of Experimental Botany. Volume 60, Issue 9, July 2009, Pages 2529–2539.

NBC Washington Staff (March 28, 2021). Cherry Blossom Trees Reach Peak Bloom Early Amid Warm Temps.

Impact of global warming on a group of related species and their hybrids: cherry tree (Rosaceae) flowering at Mt. Takao, Japan. American Journal of Botany 94: 1470–1478.

Autoregressive Integrated Moving Average (ARIMA)., By ADAM HAYES., Updated October 12, 2021., https://[www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp.](http://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp)

Fit a linear model with time series components — tslm • forecast., https://pkg.robjhyndman.com/forecast/reference/tslm.html.

Ljung-Box Test Plot in LSTS: Locally Stationary Time Series., https://rdrr.io/cran/LSTS/man/Box.Ljung.Test.html.

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2.

Martynas Rakickis, Time Series Forecasting with R Using Belgium Flights Data https://rpubs.com/marakis/belgium\_flights\_forecasting

Ubiqum Code Academy (2019), Forecasting non-seasonal time series with ARIMA models, https://rpubs.com/Mentors\_Ubiqum/Tutorial\_ARIMA\_models

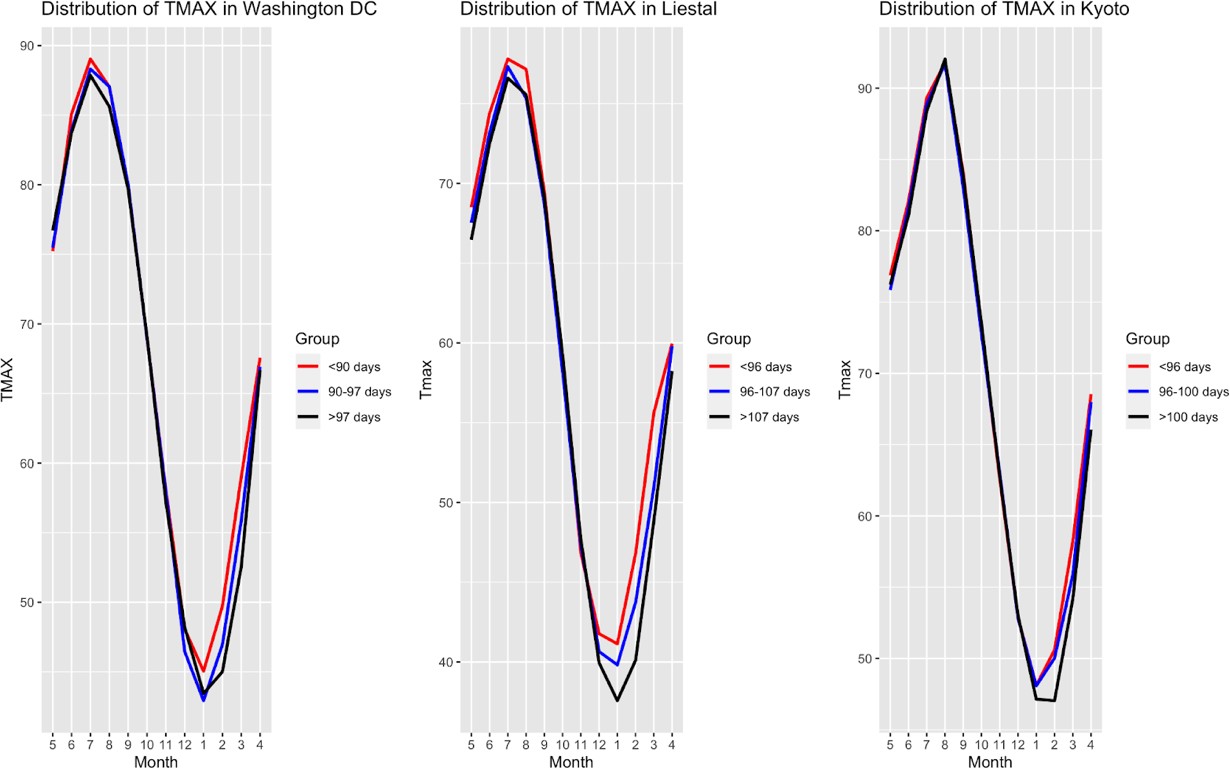


Fig. 1: Monthly Maximum Temperature in Washington DC, Liestal, and Kyoto.

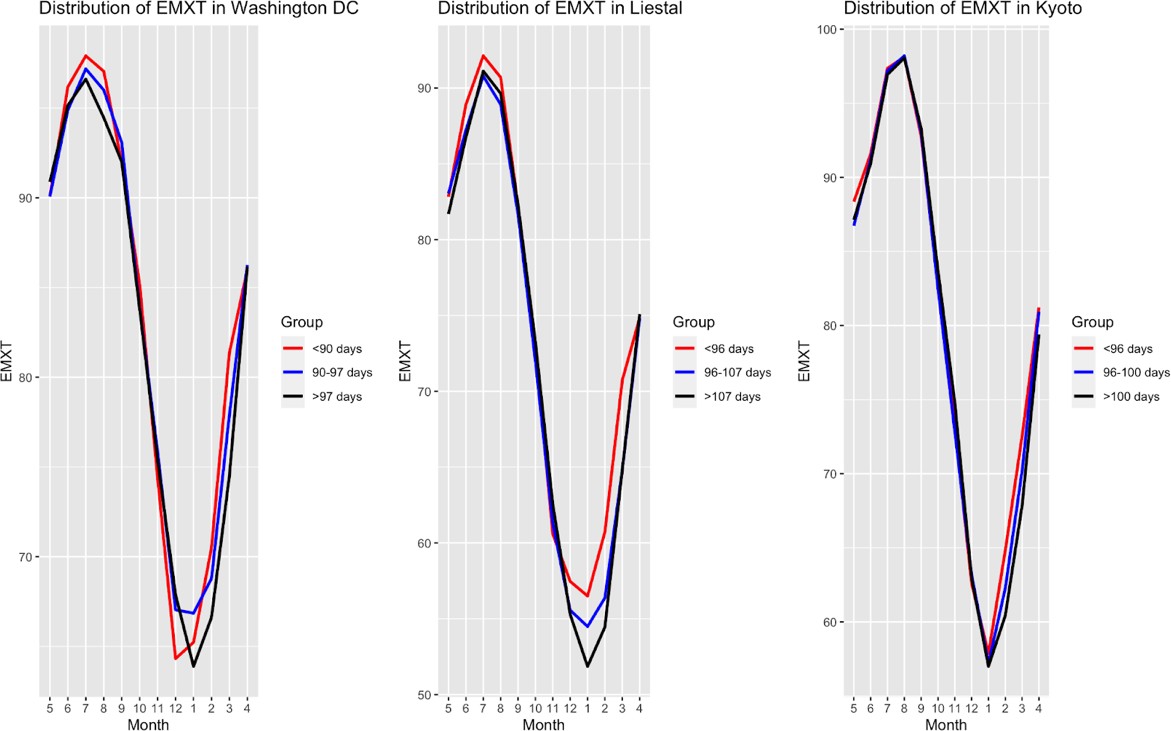
m

Fig. 2: Monthly Extreme Temperature in Washington DC, Liestal, and Kyoto.

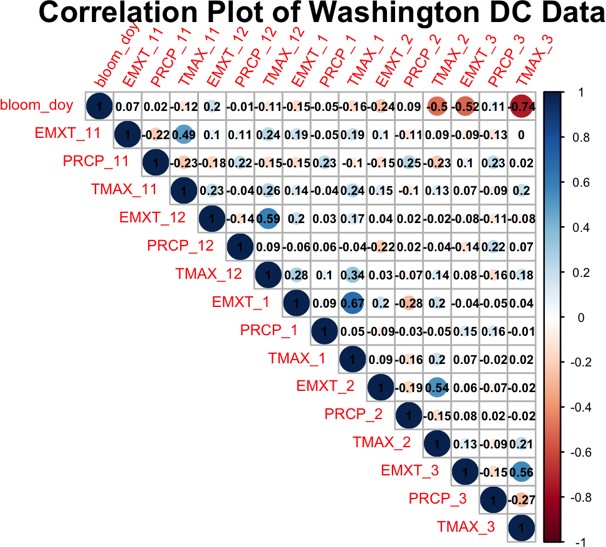


Fig. 3: Correlation between bloom date and covariates in Washington DC, USA.

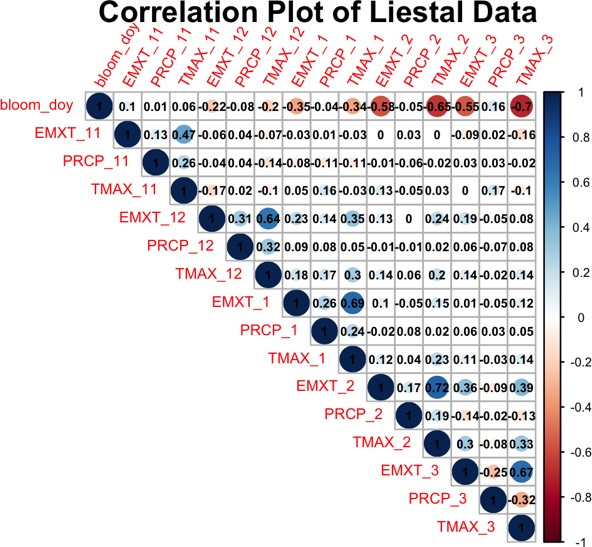


Fig. 4: Correlation between bloom date and covariates in Liestal, Switzerland.

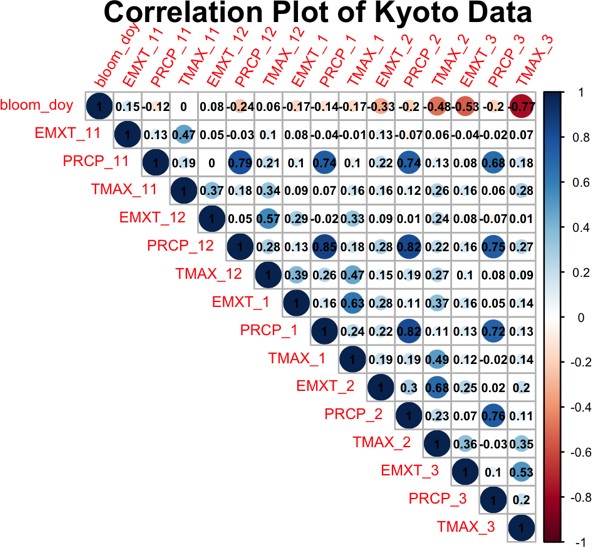


Fig. 5: Correlation between bloom date and covariates in Kyoto, Japan.

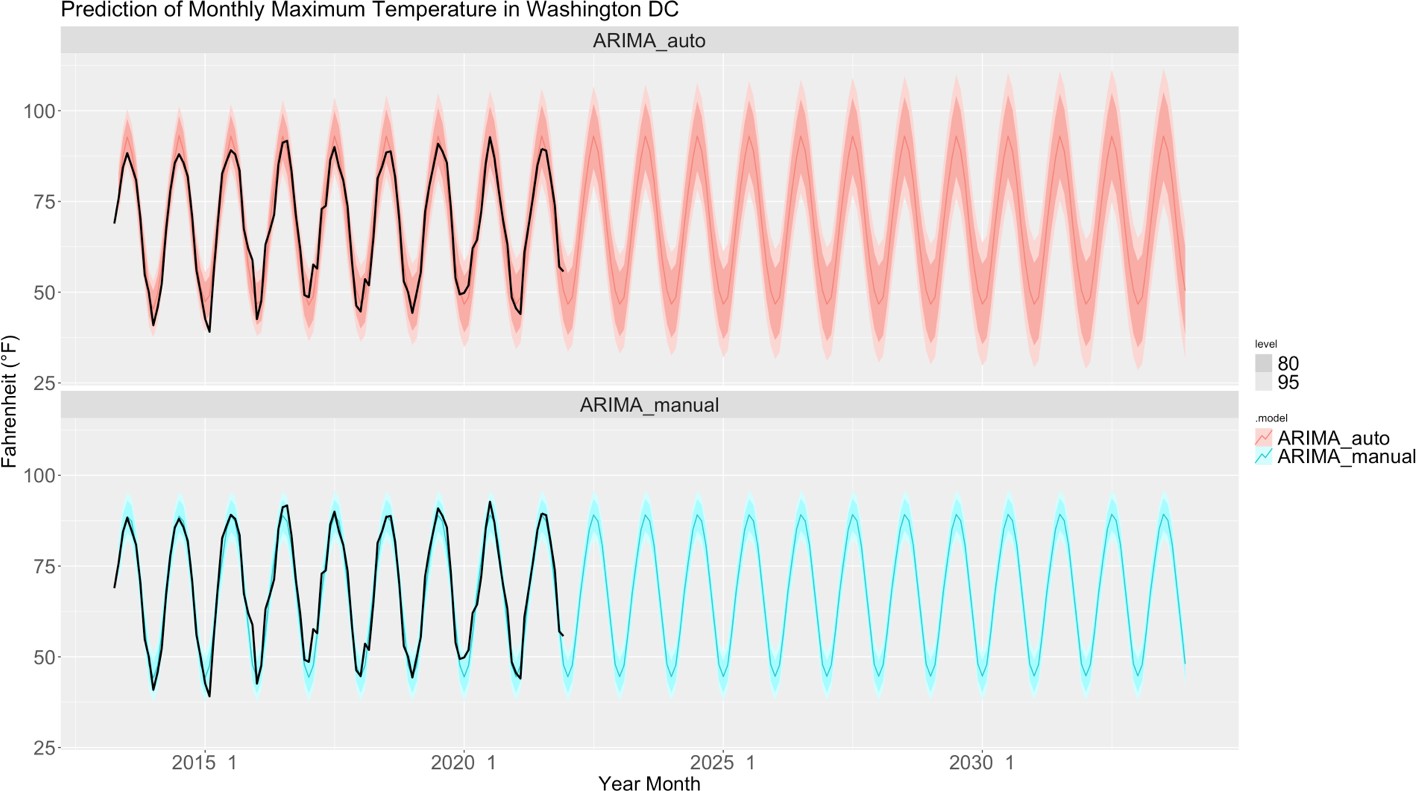


Fig. 6: Prediction and Forecast monthly maximum temperature (TMAX) in Washington DC, from May 2013 to December 2033. (Black bold line is testing data.)

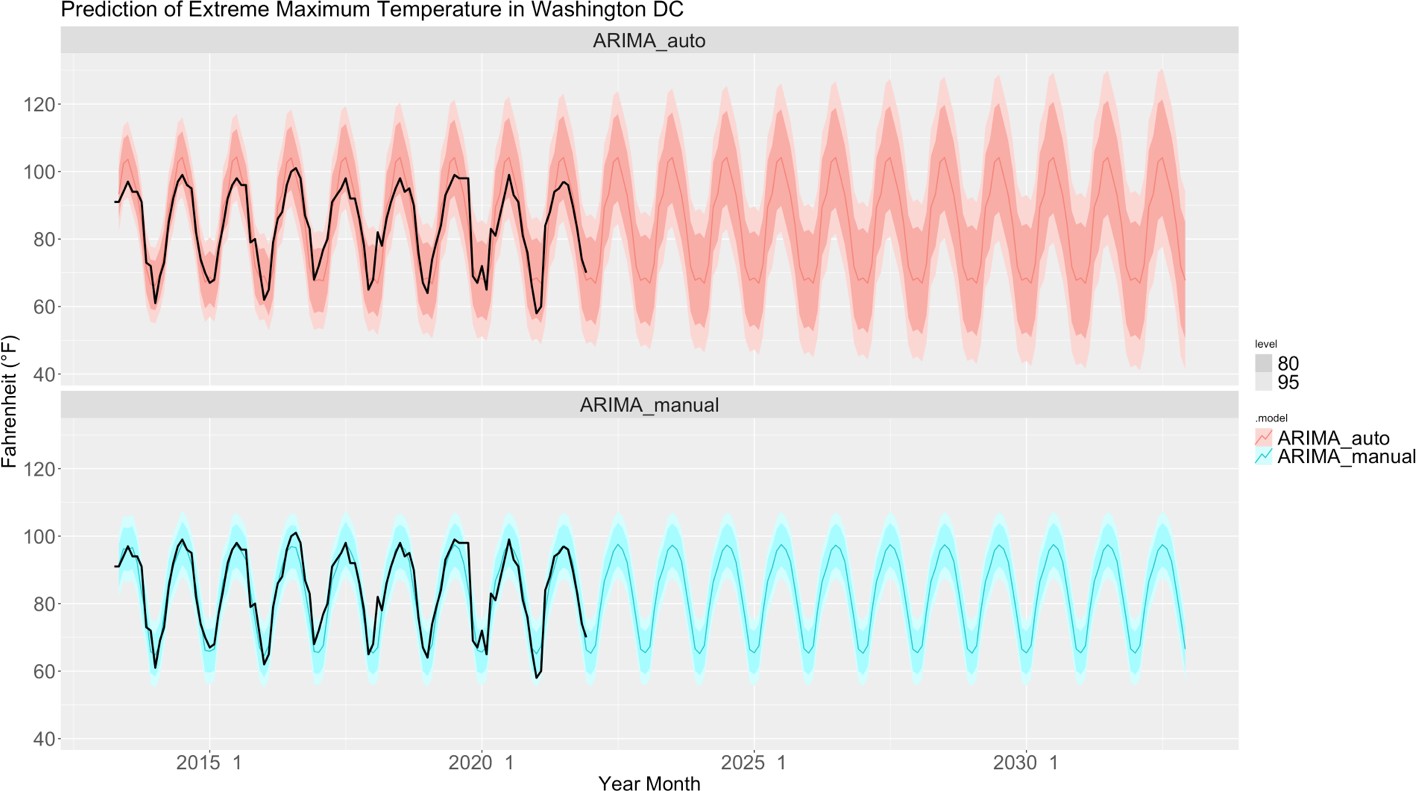


Fig. 7: Prediction and Forecast monthly extreme maximum temperature (EMXT) in Washington DC, from May 2013 to December 2032. (Black bold line is testing data.)

Table 1: Compare Models for Training data (left: Washington DC, middle: Kyoto, right: Listal).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DC Models Compare** | | | | **Kyoto Models Compare** | | | | **Liestal Models Compare** | | | |
| **model** | **AIC** | **AICc** | **BIC** | **model** | **AIC** | **AICc** | **BIC** | **model** | **AIC** | **AICc** | **BIC** |
| ARIMA\_without\_covariate | 447.74 | 447.93 | 452.15 | ARIMA\_without\_covariate | 359.90 | 360.10 | 364.16 | ARIMA\_without\_covariate | 813.41 | 813.53 | 818.76 |
| ARIMA\_with\_covariates | 388.29 | 389.27 | 399.31 | ARIMA\_with\_covariates | 314.39 | 315.09 | 322.90 | ARIMA\_with\_covariates | 737.08 | 737.47 | 747.77 |
| ARIMA\_with\_TMAX\_covariate | 395.84 | 396.48 | 404.72 | ARIMA\_with\_TMAX\_covariate | 312.95 | 313.36 | 319.33 | ARIMA\_with\_TMAX\_covariate | 737.06 | 737.29 | 745.08 |
| ARIMA\_with\_EMXT\_covariate | 443.54 | 443.92 | 450.15 | ARIMA\_with\_EMXT\_covariate | 361.53 | 361.95 | 367.91 | ARIMA\_with\_EMXT\_covariate | 812.66 | 812.89 | 820.68 |
| TSLM\_with\_covariates | 206.83 | 207.80 | 217.93 | TSLM\_with\_covariates | 134.89 | 135.94 | 145.60 | TSLM\_with\_covariates | 440.93 | 441.52 | 454.34 |
| TSLM\_with\_TMAX\_covariate | 209.44 | 210.07 | 218.32 | TSLM\_with\_TMAX\_covariate | 133.50 | 134.19 | 142.07 | TSLM\_with\_TMAX\_covariate | 442.20 | 442.59 | 452.93 |

Table 2: Ljung–Box Test of the Models’ Residuals (left: Washington DC, middle: Kyoto, right: Listal).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ljung–Box test for DC** | | | **Ljung–Box test for Kyoto** | | | **Ljung–Box test for Liestal** | | |
| **model** | **lb\_stat** | **lb\_pvalue** | **model** | **lb\_stat** | **lb\_pvalue** | **model** | **lb\_stat** | **lb\_pvalue** |
| ARIMA\_without\_covariate | 0 | 0.99 | ARIMA\_without\_covariate | 2.72 | 0.10 | ARIMA\_without\_covariate | 0.02 | 0.88 |
| ARIMA\_with\_covariates | 0.11 | 0.74 | ARIMA\_with\_covariates | 0 | 0.98 | ARIMA\_with\_covariates | 0.37 | 0.54 |
| ARIMA\_with\_TMAX\_covariate | 0.02 | 0.89 | ARIMA\_with\_TMAX\_covariate | 2.71 | 0.10 | ARIMA\_with\_TMAX\_covariate | 0.03 | 0.85 |
| ARIMA\_with\_EMXT\_covariate | 0.24 | 0.62 | ARIMA\_with\_EMXT\_covariate | 0 | 0.96 | ARIMA\_with\_EMXT\_covariate | 0.77 | 0.38 |
| TSLM\_with\_covariates | 9.63 | 0 | TSLM\_with\_covariates | 2.09 | 0.15 | TSLM\_with\_covariates | 2.73 | 0.10 |
| TSLM\_with\_TMAX\_covariate | 8.16 | 0 | TSLM\_with\_TMAX\_covariate | 1.99 | 0.16 | TSLM\_with\_TMAX\_covariate | 3.55 | 0.06 |

Table 3: Compare Models for Testing data with predicted weather data (left: Washington DC, middle: Kyoto, right: Listal).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DC models** | **ME** | **RMSE** | **MAE** | **MPE** | **MAPE** | **Kyoto model** | **ME** | **RMSE** | **MAE** | **MPE** | **MAPE** | **Liestal model** | **ME** | **RMSE** | **MAE** | **MPE** | **MAPE** |
| ARIMA\_with\_covariates | -0.61 | 6.87 | 6.12 | -1.27 | 6.85 | ARIMA\_with\_covariates | -3.95 | 5.72 | 4.57 | -4.48 | 5.11 | ARIMA\_with\_covariates | -2.08 | 8.47 | 7.26 | -3.09 | 8.16 |
| ARIMA\_with\_TMAX\_covariate | -2.40 | 7.28 | 6.51 | -3.26 | 7.41 | ARIMA\_with\_TMAX\_covariate | -3.85 | 5.65 | 4.49 | -4.37 | 5.03 | ARIMA\_with\_TMAX\_covariate | -1.89 | 8.40 | 7.21 | -2.87 | 8.09 |
| ARIMA\_without\_covariate | -0.12 | 7.03 | 6.25 | -0.74 | 6.95 | ARIMA\_without\_covariate | -3.89 | 5.73 | 4.62 | -4.42 | 5.16 | ARIMA\_without\_covariate | -3.24 | 8.82 | 7.44 | -4.36 | 8.45 |
|  | | | | | | TSLM\_with\_TMAX\_covariate | -2.91 | 5.02 | 3.81 | -3.35 | 4.27 | TSLM\_with\_covariates | -4.22 | 9.14 | 7.56 | -5.42 | 8.66 |

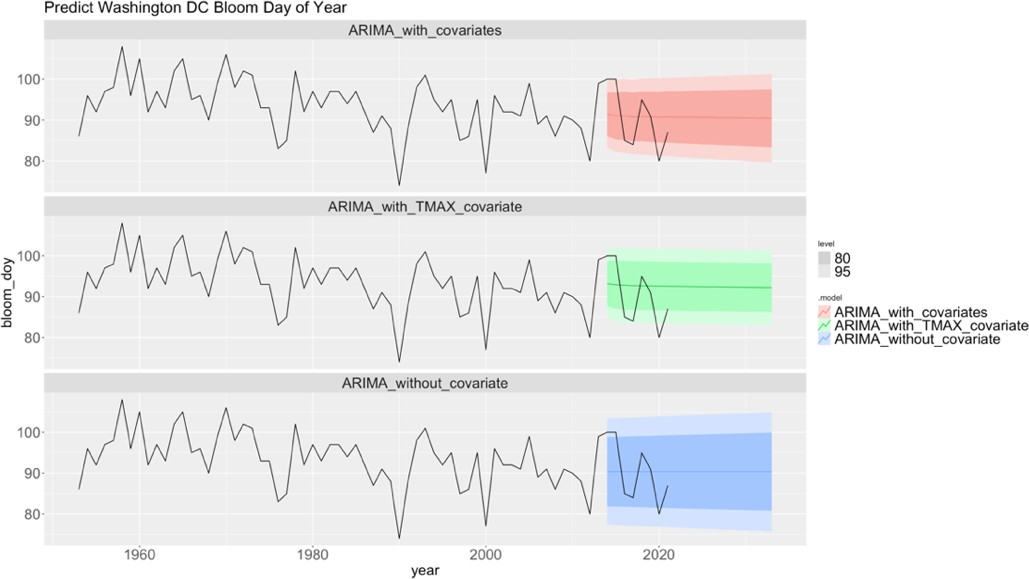


Fig. 8: Prediction and Forecast Bloom data of year in Washington DC with predicted weather data. (Black bold line is real data.)

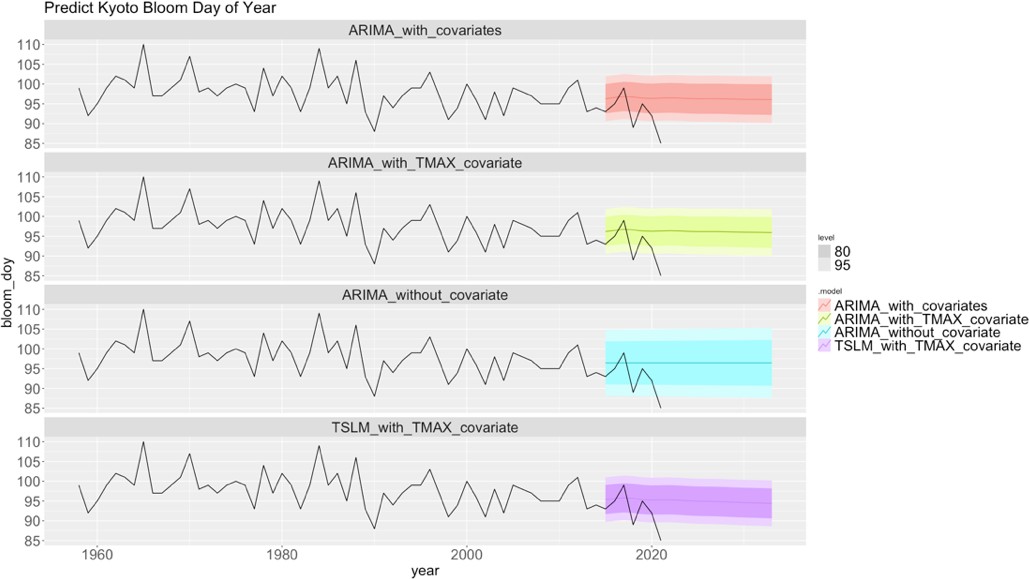


Fig. 9: Prediction and Forecast Bloom data of year in Kyoto with predicted weather data. (Black bold line is real data.)

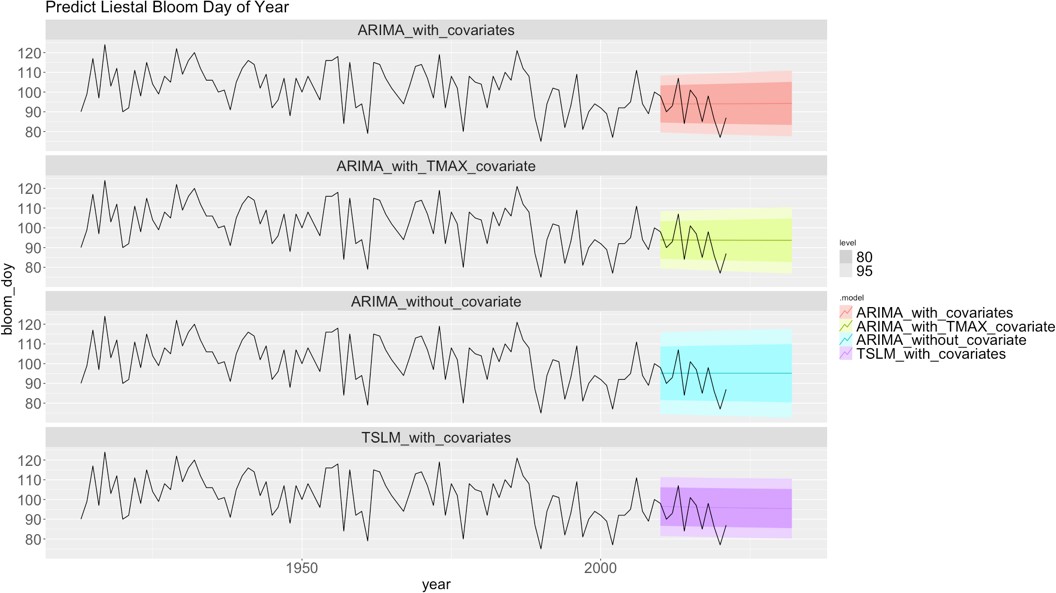


Fig. 10: Prediction and Forecast Bloom data of year in Liestal with predicted weather data. (Black bold line is real data.)

Table 4. Forecast bloom day of the year (left: Washington DC, middle: Kyoto, right: Listal).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DC** | mean | 80% CI | | 95% CI | | **Kyoto** | mean | 80% CI | | 95% CI | | **Liestal** | mean | 80% CI | | 95% CI | |
| 2022 | 91 | 84 | 97 | 81 | 100 | 2022 | 95 | 92 | 99 | 90 | 101 | 2022 | 94 | 83 | 104 | 78 | 110 |
| 2023 | 91 | 84 | 97 | 81 | 100 | 2023 | 95 | 91 | 99 | 90 | 101 | 2023 | 94 | 83 | 104 | 78 | 110 |
| 2024 | 91 | 84 | 97 | 81 | 101 | 2024 | 95 | 91 | 99 | 89 | 101 | 2024 | 94 | 83 | 104 | 78 | 110 |
| 2025 | 91 | 84 | 97 | 81 | 101 | 2025 | 95 | 91 | 99 | 89 | 101 | 2025 | 94 | 83 | 104 | 78 | 110 |
| 2026 | 91 | 84 | 97 | 81 | 101 | 2026 | 95 | 91 | 99 | 89 | 101 | 2026 | 94 | 83 | 104 | 77 | 110 |
| 2027 | 91 | 84 | 97 | 80 | 101 | 2027 | 95 | 91 | 99 | 89 | 101 | 2027 | 94 | 83 | 104 | 77 | 110 |
| 2028 | 91 | 84 | 97 | 80 | 101 | 2028 | 95 | 91 | 99 | 89 | 101 | 2028 | 94 | 83 | 104 | 77 | 110 |
| 2029 | 91 | 84 | 97 | 80 | 101 | 2029 | 95 | 91 | 98 | 89 | 100 | 2029 | 94 | 83 | 105 | 77 | 110 |
| 2030 | 91 | 84 | 97 | 80 | 101 | 2030 | 95 | 91 | 98 | 89 | 100 | 2030 | 94 | 83 | 105 | 77 | 110 |
| 2031 | 90 | 84 | 97 | 80 | 101 | 2031 | 95 | 91 | 98 | 89 | 100 | 2031 | 94 | 83 | 105 | 77 | 110 |
| 2032 | 90 | 83 | 97 | 80 | 101 | 2032 | 94 | 91 | 98 | 89 | 100 | 2032 | 94 | 83 | 105 | 77 | 111 |
| 2033 | 90 | 83 | 98 | 80 | 101 | 2033 | 94 | 91 | 98 | 89 | 100 | 2033 | 94 | 83 | 105 | 77 | 111 |

Table 5: Compare Models for Testing data with real weather data (left: Washington DC, middle: Kyoto, right: Listal).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DC models** | **ME** | **RMSE** | **MAE** | **MPE** | **MAPE** | **Kyoto model** | **ME** | **RMSE** | **MAE** | **MPE** | **MAPE** | **Liestal model** | **ME** | **RMSE** | **MAE** | **MPE** | **MAPE** |
| ARIMA\_with\_covariates | -0.19 | 3.93 | 3.05 | -0.47 | 3.42 | ARIMA\_with\_covariates | 0.29 | 1.49 | 1.23 | 0.26 | 1.35 | ARIMA\_with\_covariates | 0.85 | 6.15 | 4.11 | 0.64 | 4.61 |
| ARIMA\_with\_TMAX\_covariate | -1.62 | 4.42 | 3.49 | -2.08 | 3.95 | ARIMA\_with\_TMAX\_covariate | 0.43 | 1.52 | 1.29 | 0.41 | 1.39 | ARIMA\_with\_TMAX\_covariate | 0.5 | 6.46 | 4.39 | 0.26 | 4.91 |
| ARIMA\_without\_covariate | -0.12 | 7.03 | 6.25 | -0.74 | 6.95 | ARIMA\_without\_covariate | -3.89 | 5.73 | 4.62 | -4.42 | 5.16 | ARIMA\_without\_covariate | -3.24 | 8.82 | 7.44 | -4.36 | 8.45 |
|  | | | | | | TSLM\_with\_TMAX\_covariate | 1.2 | 1.92 | 1.66 | 1.24 | 1.77 | TSLM\_with\_covariates | -1.19 | 5.88 | 4.17 | -1.61 | 4.69 |

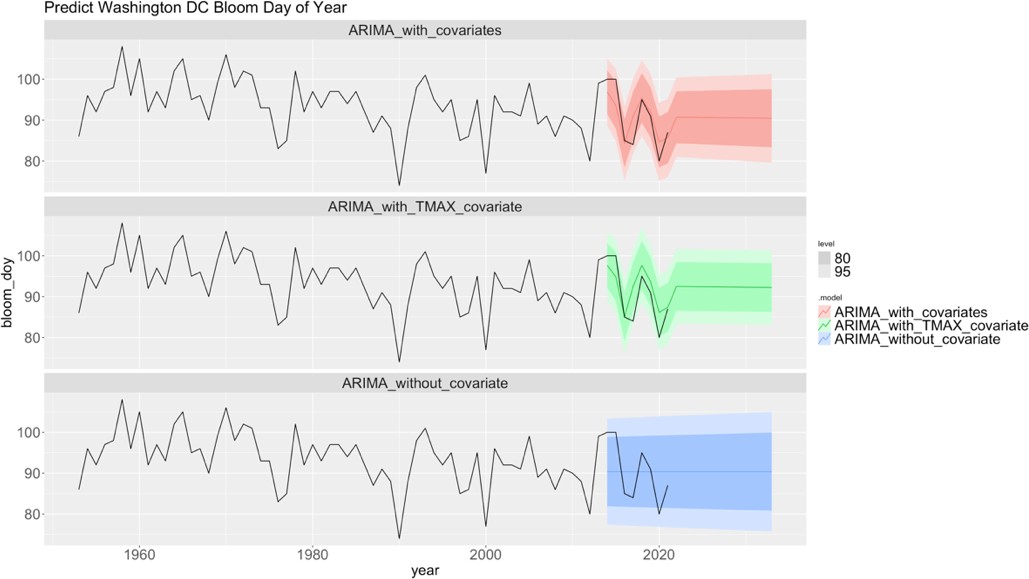


Fig. 11: Prediction and Forecast Bloom data of year in Washington DC with real weather data. (Black bold line is real data.)

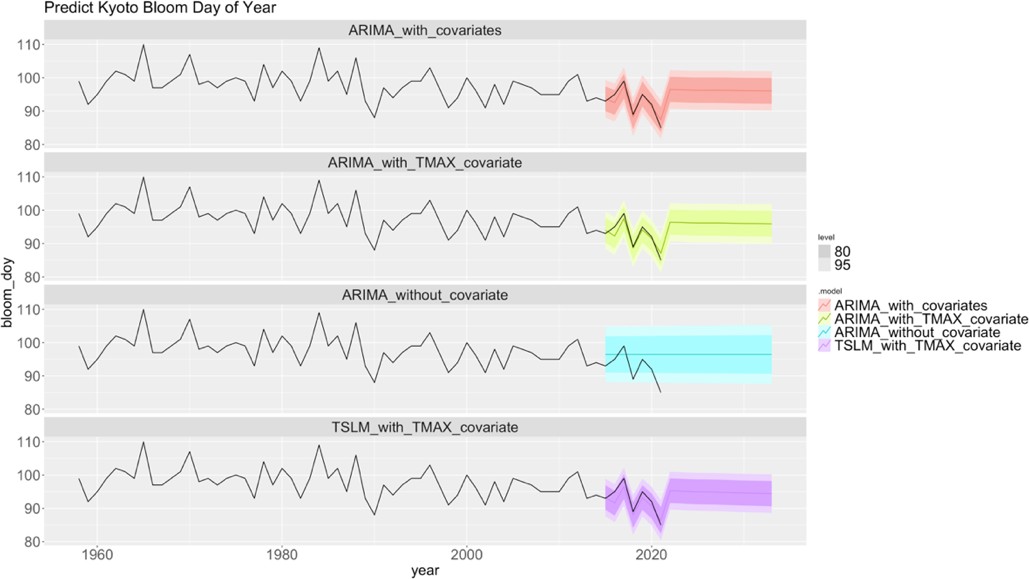


Fig. 12: Prediction and Forecast Bloom data of year in Kyoto with predicted weather data. (Black bold line is real data.)

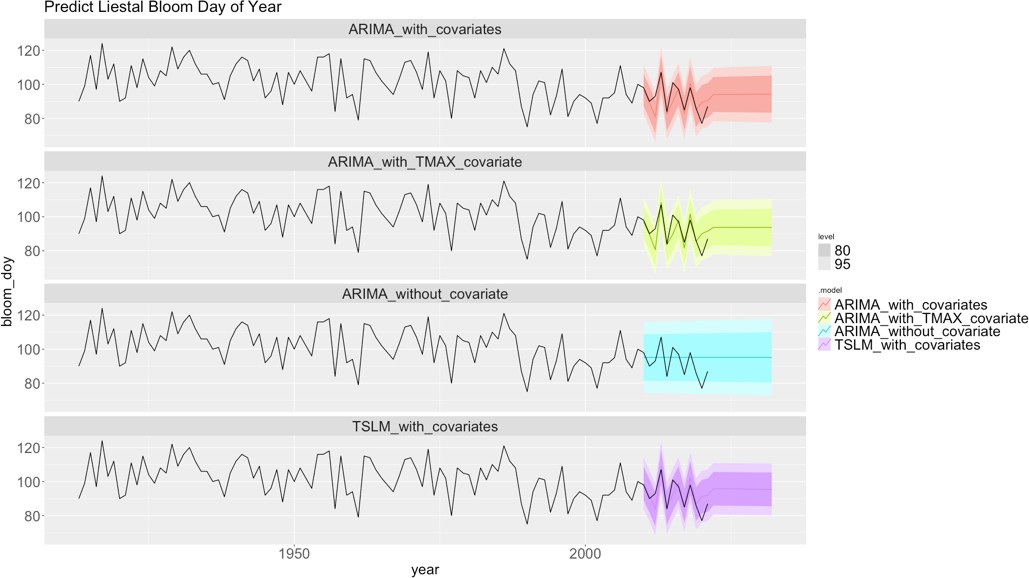


Fig. 13: Prediction and Forecast Bloom data of year in Liestal with predicted weather data. (Black bold line is real data.)

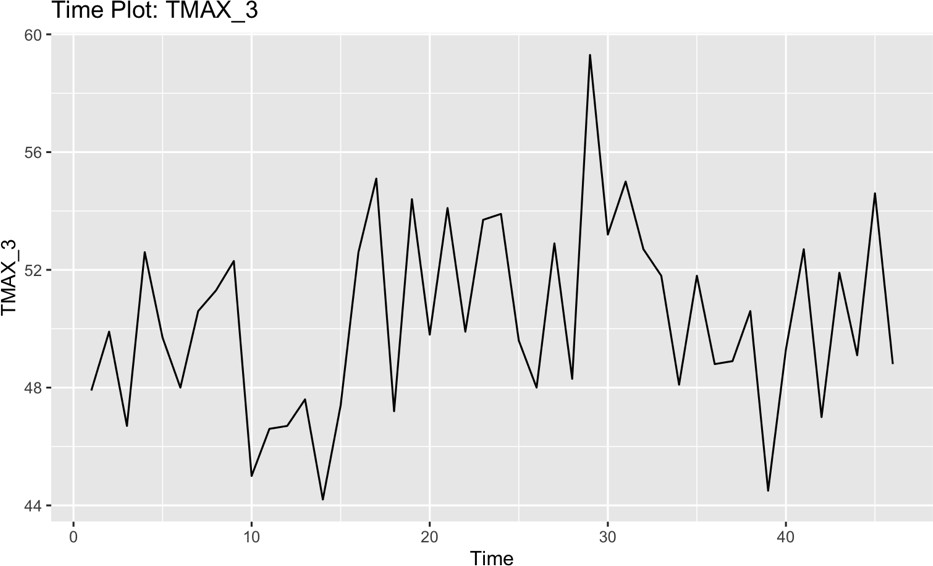


Fig. 14: Time plot of TMAX\_3 for Vancouver, Canada.

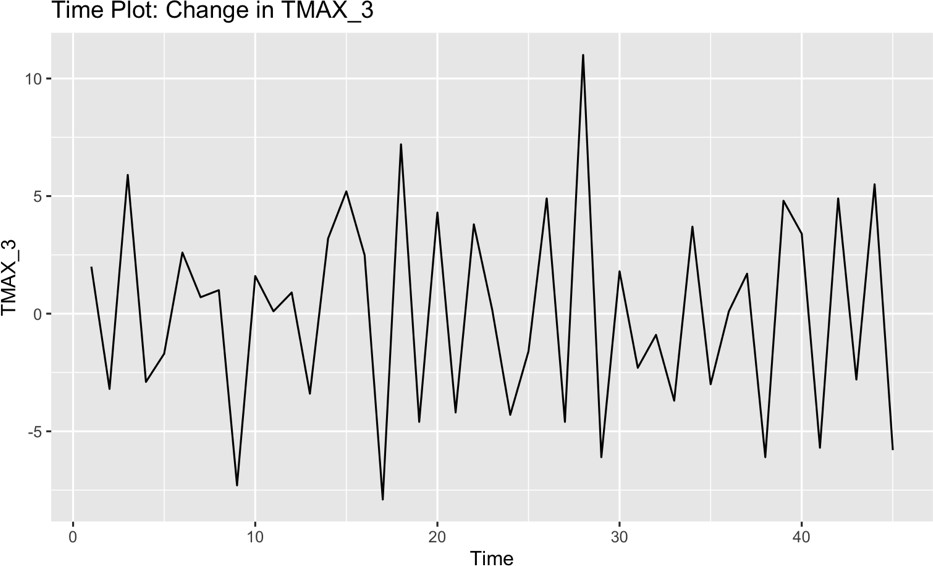


Fig. 15: Time plot of change in TMAX\_3 for Vancouver, Canada.

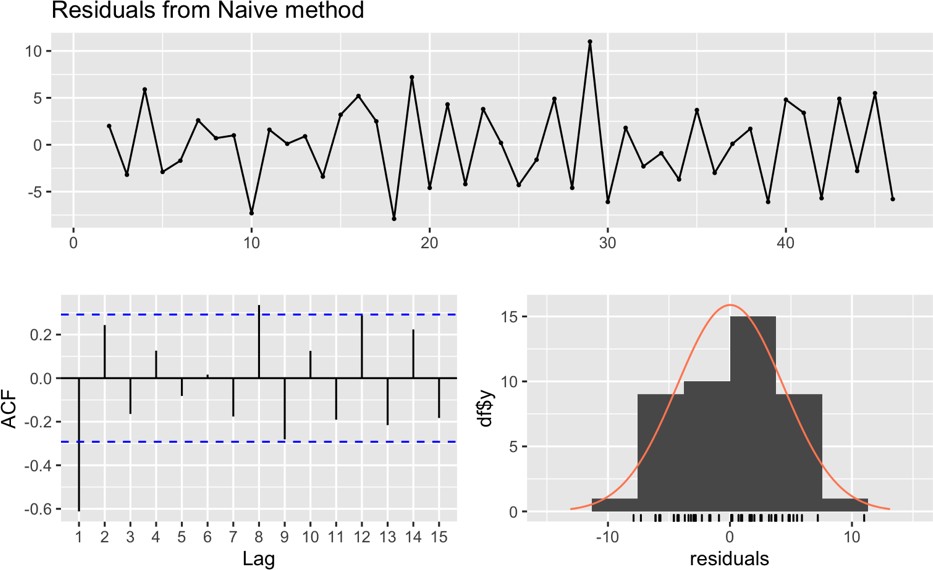


Fig. 16: Residual plot of Naïve model for TMAX\_3 for Vancouver, Canada.

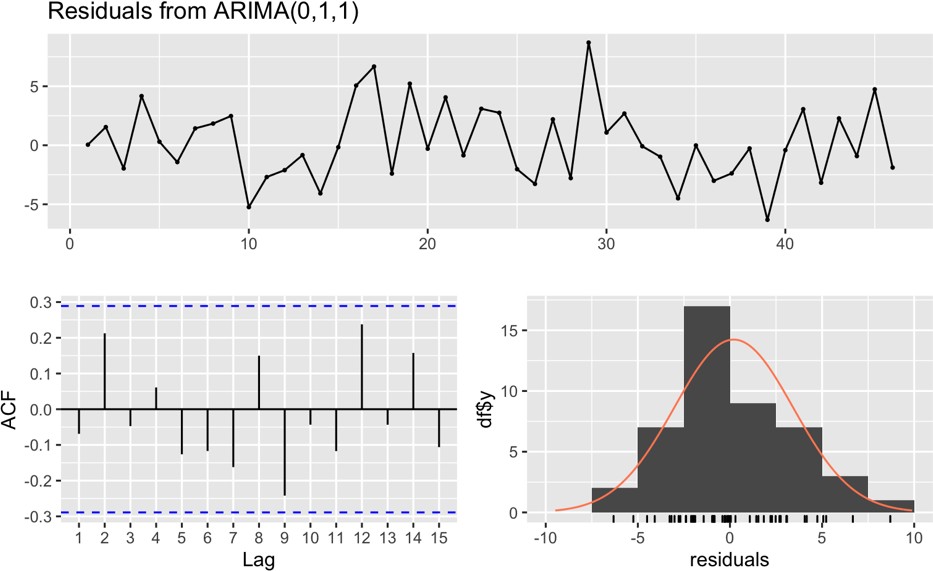


Fig. 17: Residual plot of Arima model for TMAX\_3 for Vancouver, Canada.

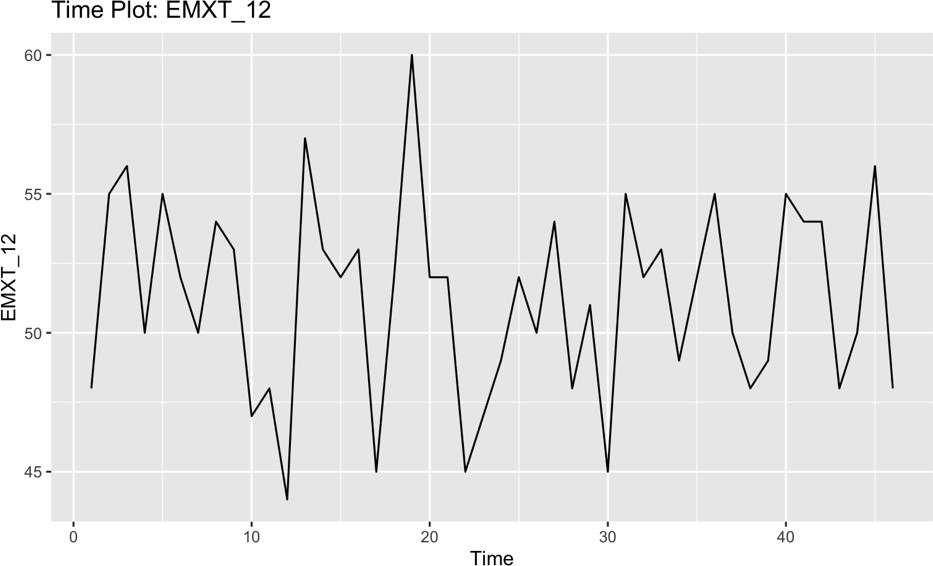


Fig. 18: Time plot EMXT\_12 for Vancouver, Canada.

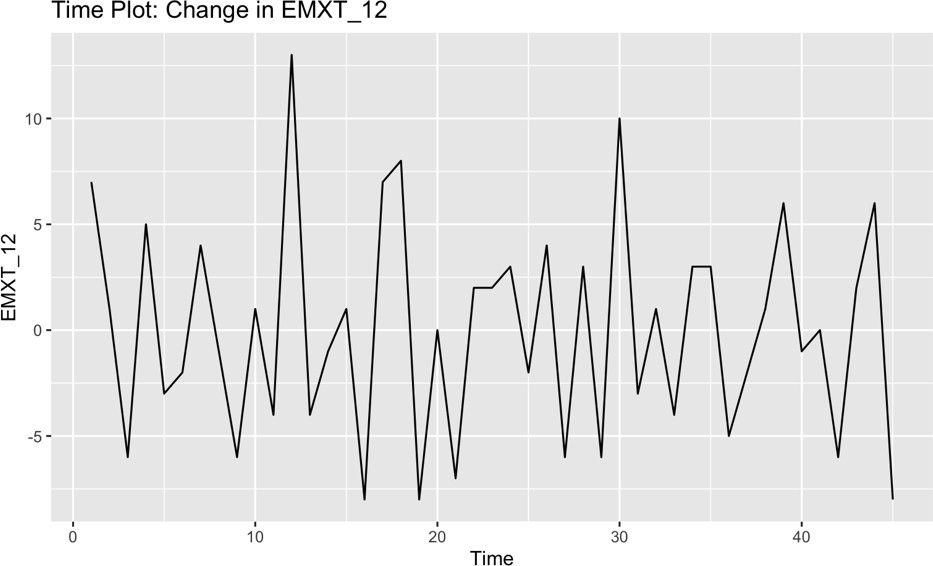


Fig. 19: The time plot of change in EMXT\_12 for Vancouver, Canada.

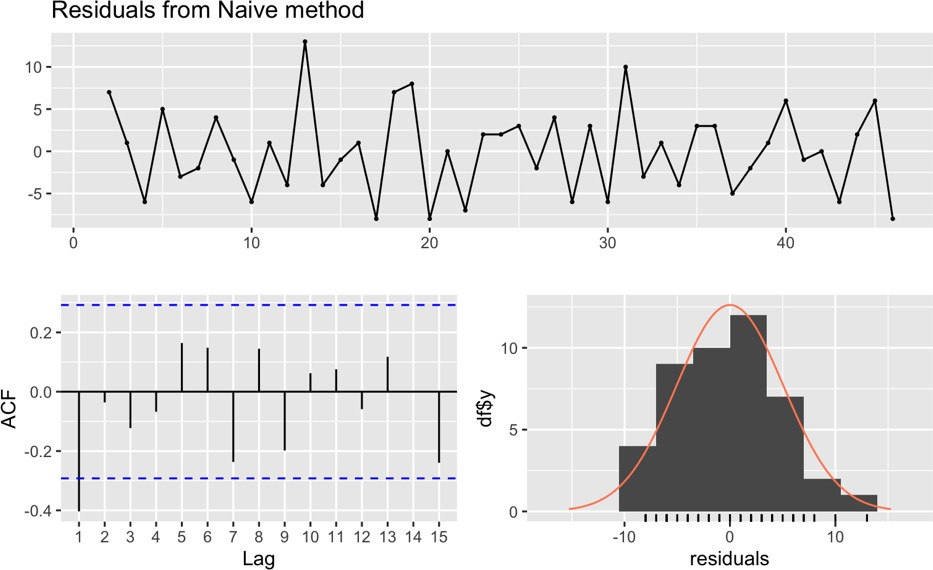


Fig. 20: The Residual plot of Naïve model for EMXT\_12 for Vancouver, Canada.

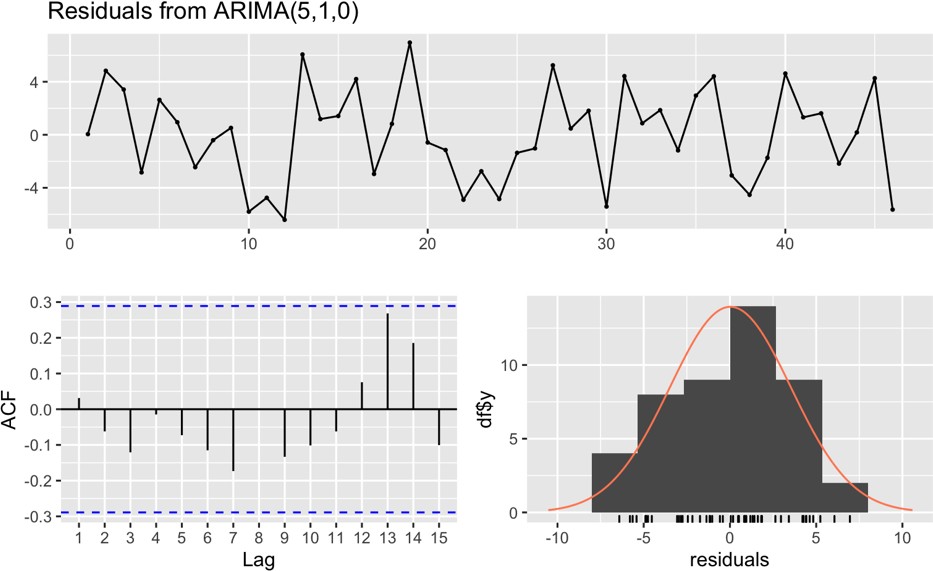


Fig. 21: The Residual plot of Arima model for EMXT\_12 for Vancouver, Canada.