Project – Cherry Blossom Prediction

Authors: Shuangyan Wu, Bingnan Li, Cong Cheng

Date: August 3, 2022

UNIVERSITY OF GEORGIA

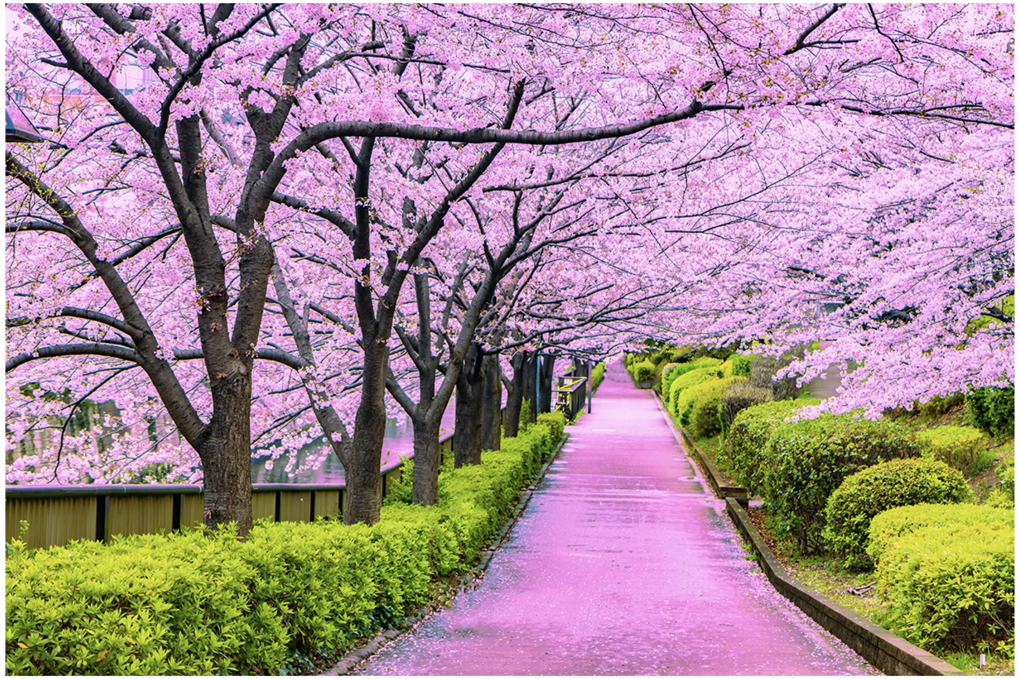
Agenda

- Background
- Data Summary
- Data Analysis
- Summary

Background

- **In Japan, one of the big cultural events is watching Cherry Trees (Sakura) blossom. It is a national cultural event and festival (Hanami) in Japan. Historically, this occurred sometime in April, but it appears to be occurring earlier, possibly because of global warming.**

- **Our goal is to use historical cherry blossom data from Japan to make predictions of the date of Peak Cherry Blossom Boom as a function of latitude and year**



Adopted from:
https://www.kcpinternational.com/2011/07/cherish-the-beauty-of-japans-cherry-blossoms/

Data Summary

Two datasets:

<table>
<tr><td>Name</td><td>HistKACFB (data1)</td><td>ModJapanFB (data2)</td></tr>
<tr><td>Description</td><td>Data contains only one city Kyoto/Arashiyama from 801-2021.</td><td>Data contains 103 cities from 1953-2021.</td></tr>
<tr><td>Variables</td><td>Year
Full-flowering date (DOY)
Full-flowering date
Source code
Data type code
Reference Name</td><td>Location (103 cities)
Latitude (111 categories)
Longitude
Altitude (in meters)
Year
Fbloom date
Fbloom doy</td></tr>
<tr><td>Size</td><td>1221 by 6</td><td>6573 by 7</td></tr>
</table>

Data Summary

Data1:

- **388** rows with missing data (total **1221** rows) (missing 31.8%)
- Complete data from 1946 to 2021

Data2:

- **45** locations without missing data
- **118** duplicated rows
- **8** locations with **2** latitudes: Japan/Izuhara, Japan/Kochi, Japan/Kushiro, Japan/Muroran, Japan/Nagoya, Japan/Sendai, Japan/Tottori and Japan/Yakushima. (**466** rows)
- Each location should have **69** years but there are **4** cities with missing data for more than **45** years (65%).
- Using **5924** rows to do the analysis. (69 (year) * 99 (latitude) = 6831, 86.7% complete)

Data Analysis (content)

Data1

- Effects of recent global warming

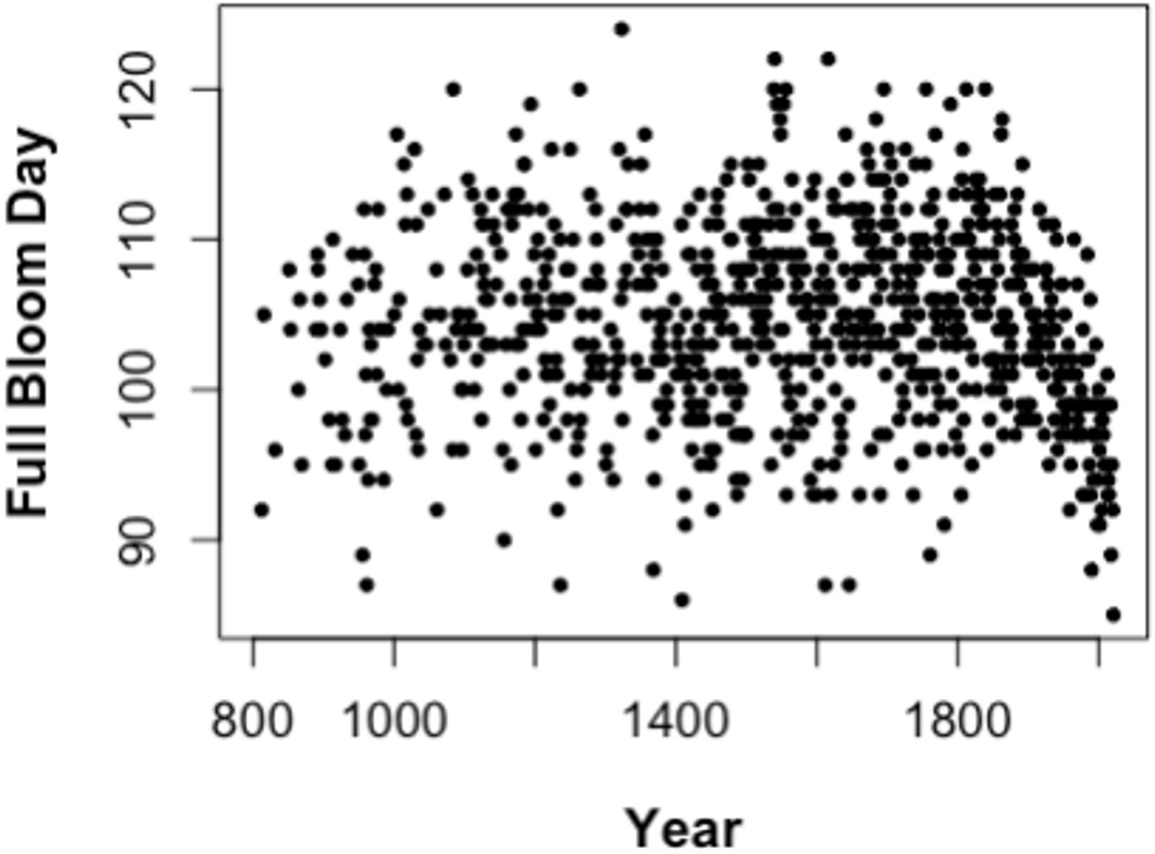
Data2

- Visualization: Fbloom DOY with year and Latitude
- Additive model: fits year and location (1-to-1 corresponds to latitude)
- Model assessment
- Time series model: predicts future

Data 1: Effects of recent global warming

Kyoto/Arashiyama

Adopted from google map.



Data 1: Effects of recent global warming

Cubic spline: Piecewise polynomial function + continuous first two derivatives

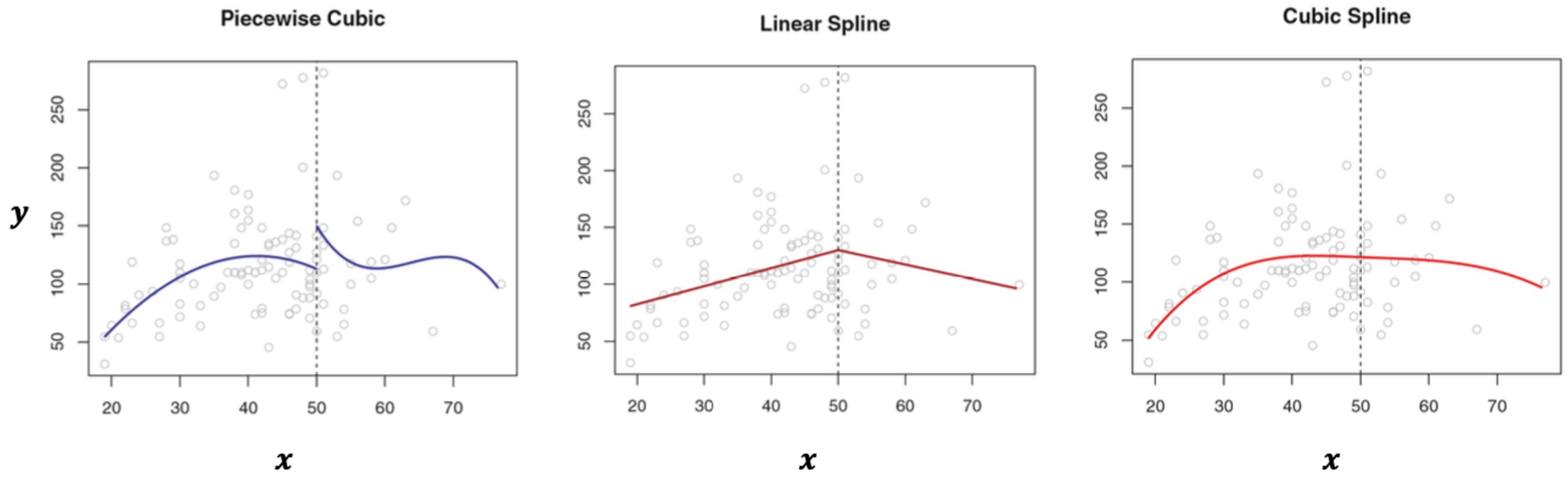
Knots: the points where third derivative might not exist.

Consider 2 knots with 3 regions:

$$c(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3, \quad x \in [0, \tau_1]$$

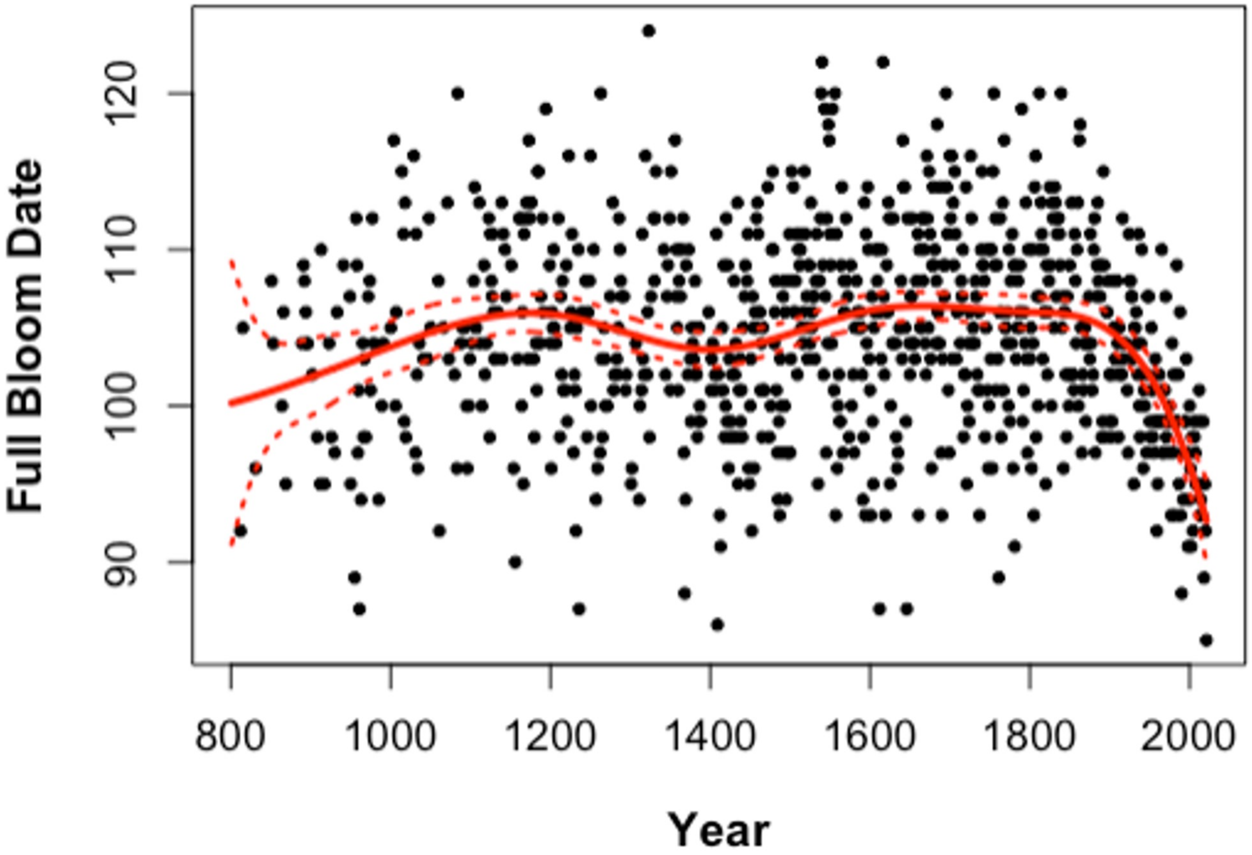
$$c(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - \tau_1)_+^3, \quad x \in [\tau_1, \tau_2]$$

$$c(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - \tau_1)_+^3 + \beta_5 (x - \tau_2)_+^3, \quad x \in [\tau_2, 1].$$



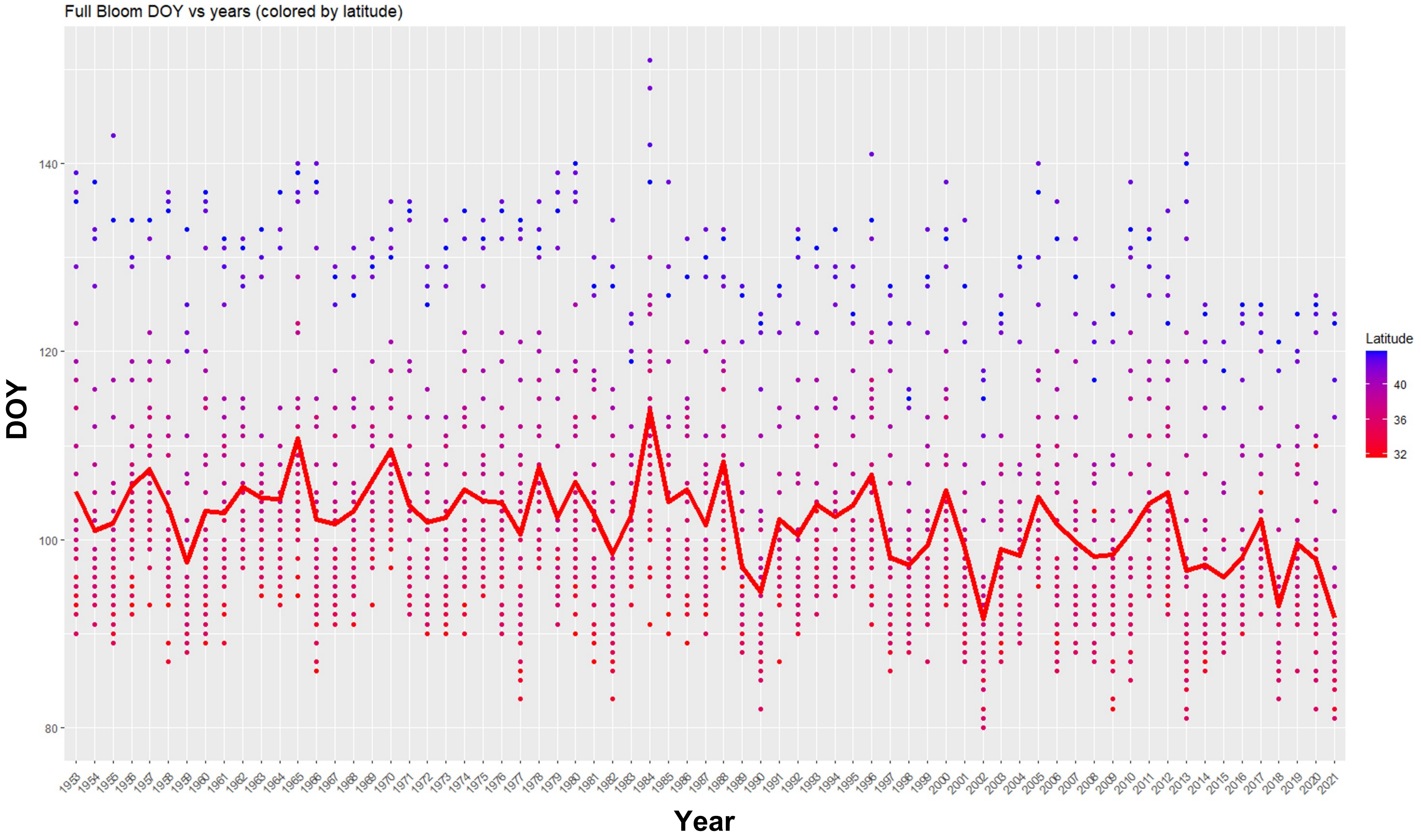
Data 1: Effects of recent global warming

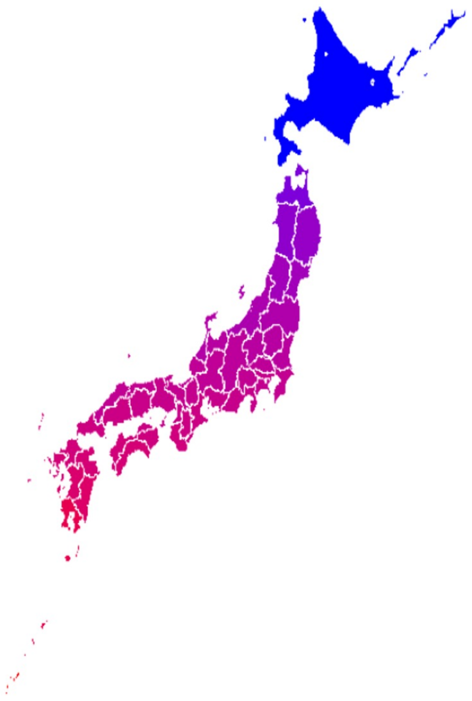
Cubic Spline



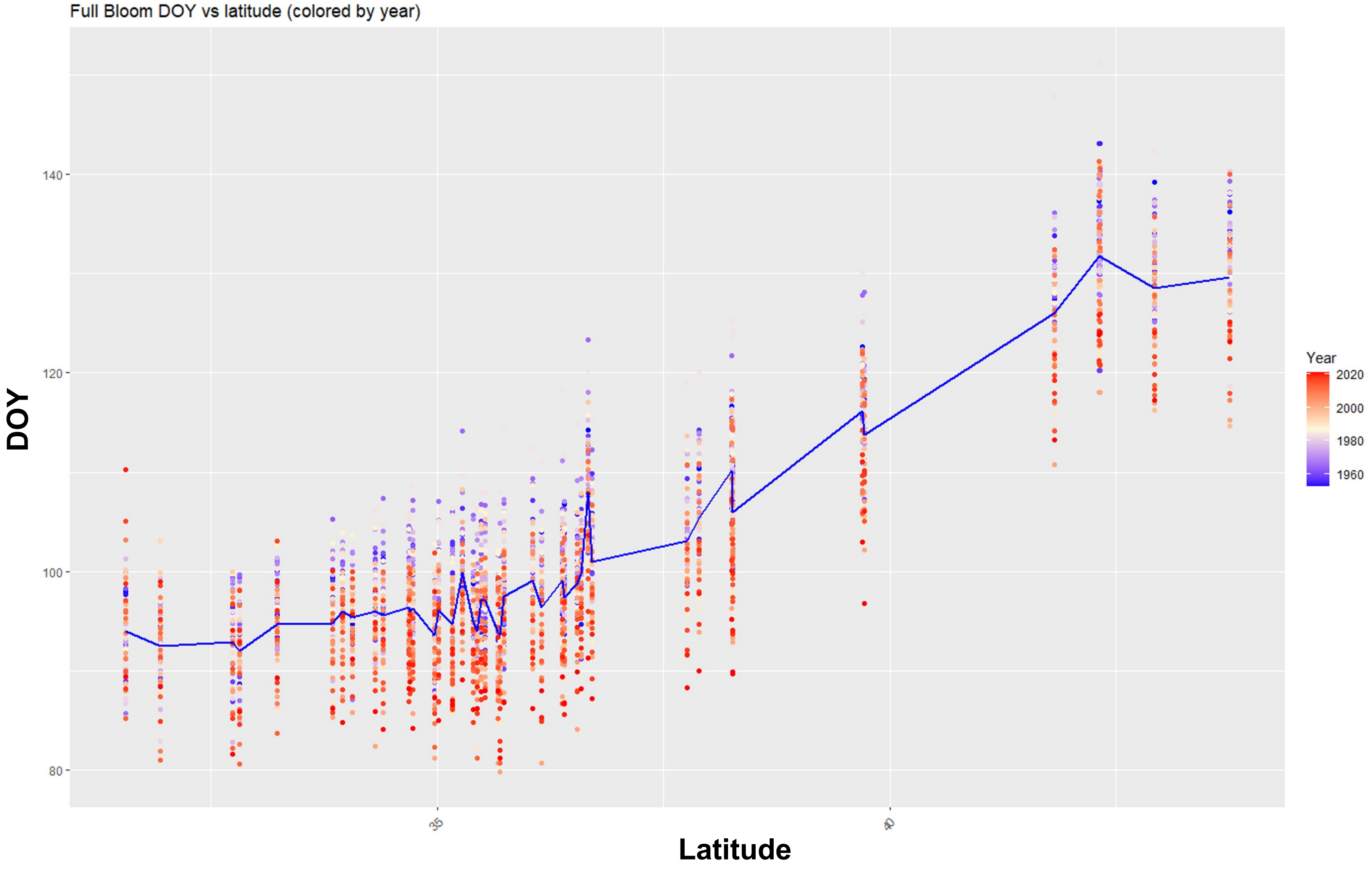
Knots:
(1000, 1200, 1400, 1600, 1800)

Data 2: Fbloom DOY vs. Year (Complete Data)





Fbloom DOY vs. Latitude (Complete Data)



Additive Model

(Fitted with all modern data)

Use additive model to fit **Fbloom DOY** by variables **Location** and **Year**

$$DOY_{ij} = \beta_0 + \alpha_i + \beta_j + e_{ij},$$

where:

$$i = 1, 2, \ldots, 98,$$
$$j = 1, 2, \ldots, 68.$$

i is for the location, j is for the year;

Location baseline: Japan/Maebashi
Year baseline: 1961

Additive Model Estimates

g = lm (DOY ~ factor (Location) + factor (Year))

Location	Estimate	Location	Estimate	Location	Estimate	Location	Estimate	Location	Estimate	Location	Estimate	Location	Estimate
Nemuro	**46.98**	Esashi	29.11	Matsumoto	8.81	Mito	1.65	Kobe	-1.38	Nagoya	-2.74	Nagasaki	-4.52
Kushiro	43.41	Hakodate	28.62	Sendai	8.61	Utsunomiya	1.36	Shimonoseki	-1.38	Hamamatsu	-2.92	Fukuoka	-4.83
Wakkanai	41.25	Aomori	22.42	Niigata	8.01	Maizuru	1.18	Nara	-1.40	Oshima	-3.17	Miyazaki	-4.86
Abashiri	37.15	Hachinohe	20.27	Wajima	6.64	Tsuruga	0.94	Takamatsu	-1.41	Fukue	-3.26	Uwajima	-5.08
Kutchan	36.91	Shinjo	19.12	Takada	6.23	Toyooka	0.81	Sumoto	-1.54	Miyakejima	-3.31	Kumamoto	-5.35
Mombetsu	36.32	Morioka	18.72	Fukushima	5.72	Matsue	0.36	Osaka	-1.69	Kagoshima	-3.38	Nobeoka	-5.90
Rumoi	36.30	Akita	16.35	Onahama	5.10	Choshi	0.12	Hiroshima	-1.80	Yokohama	-3.41	Kochi	-7.26
Urakawa	35.87	Miyako	16.18	Toyama	3.61	Yonago	-0.22	Hamada	-1.91	Kofu	-3.45	Miyakojima	-56.96
Muroran	34.30	Yamagata	12.81	Iida	3.40	Tottori	-0.33	Tokushima	-1.96	Owase	-3.57	Ishigakijima	-59.30
Hiroo	33.63	Takayama	12.34	Saigo	3.10	Tateyama	-0.96	Gifu	-2.43	Wakayama	-3.62	Naha	-61.93
Asahikawa	32.23	Sakata	12.01	Hikone	2.48	Okayama	-0.99	Tanegashima	-2.50	Izuhara	-3.71	Minamidaitoji ma	-62.57
Iwamizawa	31.55	Shirakawa	11.94	Kanazawa	2.38	Kumagaya	-1.03	Matsuyama	-2.58	Tokyo	-3.77	Kumejima	-64.75
Obihiro	31.16	Nagano	10.59	Hachijojima	2.36	Tsu	-1.13	Oita	-2.65	Shizuoka	-3.80	Naze	-65.38
Sapporo	30.30	Aikawa	10.07	Fukui	1.68	**Kyoto**	**-1.26**	Shionomisaki	-2.73	Saga	-4.27	**Funchatoge**	**-65.38**

Ref = Maebashi; Sorted

Additive Model Estimates

g = lm (DOY ~ factor (Location) + factor (Year))

Ref = 1961

Year	Estimate	Year	Estimate	Year	Estimate	Year	Estimate	Year	Estimate
1953	2.82	1967	-1.00	1981	0.18	1995	1.58	2009	-3.72
1954	-1.07	1968	0.52	1982	-3.48	1996	3.60	2010	-1.55
1955	-0.93	1969	3.71	1983	-0.27	1997	-3.21	2011	1.50
1956	2.49	1970	6.29	1984	9.95	1998	-4.94	2012	1.52
1957	4.36	1971	1.36	1985	1.02	1999	-1.77	2013	-5.14
1958	0.98	1972	-0.38	1986	1.77	2000	2.32	2014	-5.12
1959	-4.42	1973	-0.69	1987	-0.70	2001	-2.67	2015	-6.27
1960	1.10	1974	1.99	1988	4.93	**2002**	**-10.25**	2016	-2.29
1961	0.00	1975	1.76	1989	-5.46	2003	-2.77	2017	0.91
1962	2.41	1976	1.24	1990	-7.31	2004	-3.38	**2018**	**-8.24**
1963	1.50	1977	-1.71	1991	0.04	2005	3.04	2019	-2.46
1964	1.59	1978	5.15	1992	-1.35	2006	-1.30	2020	-3.68
1965	**8.22**	1979	1.09	1993	1.72	2007	-1.02	**2021**	**-9.59**
1966	-0.24	1980	4.29	1994	0.45	2008	-3.34		

Model Performance

g = lm (DOY ~ factor (Location) + factor (Year))

Model: R^2 = 0.97
Residual standard error: 3.82

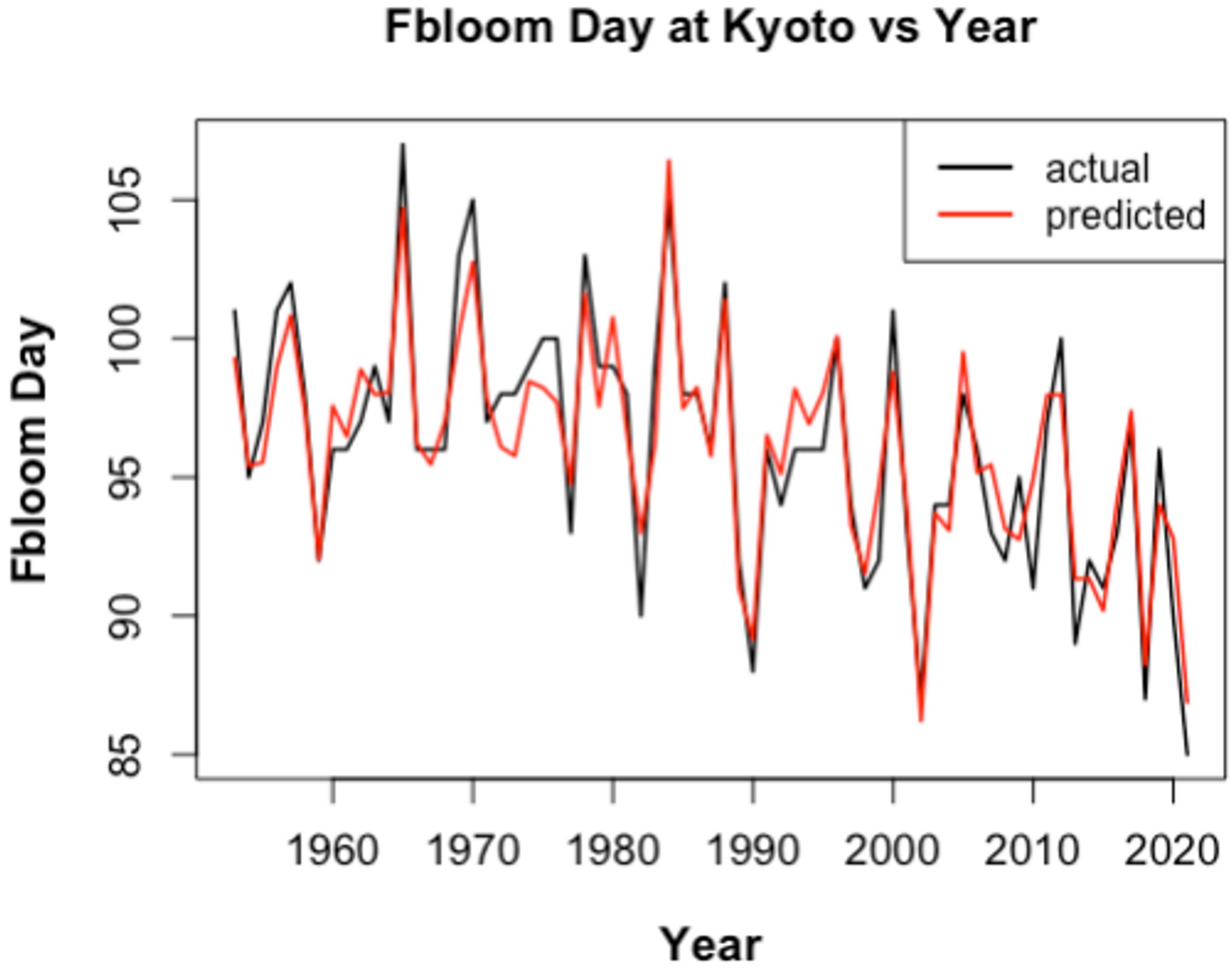
g = lm (DOY ~ factor (Location))

Model: R^2 = 0.94
Residual standard error: 5.33

g = lm (DOY ~ factor (Year))

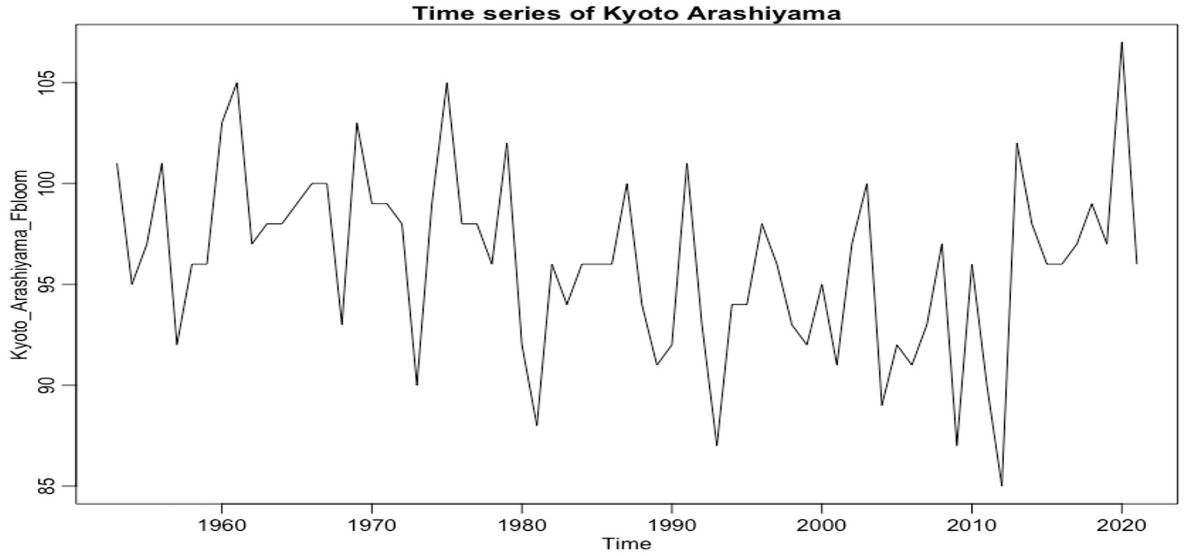
Model: R^2 = 0.06
Residual standard error: 21.31

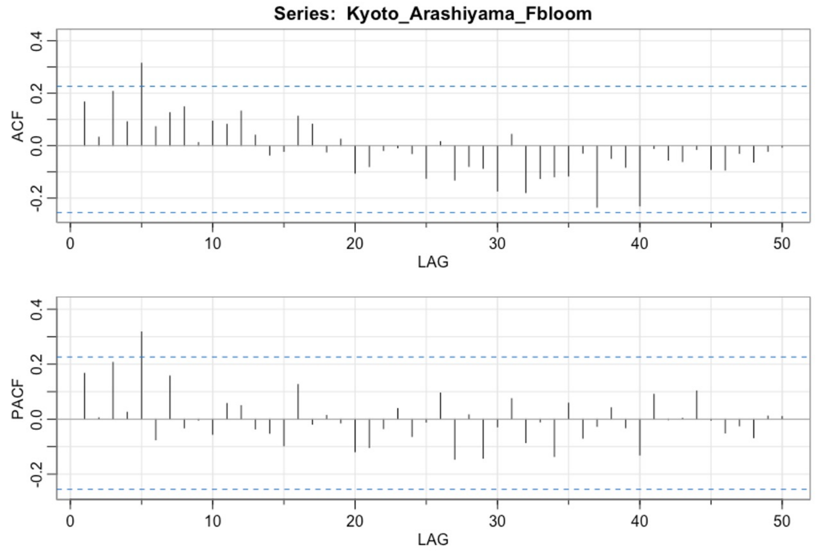
Model Fitting

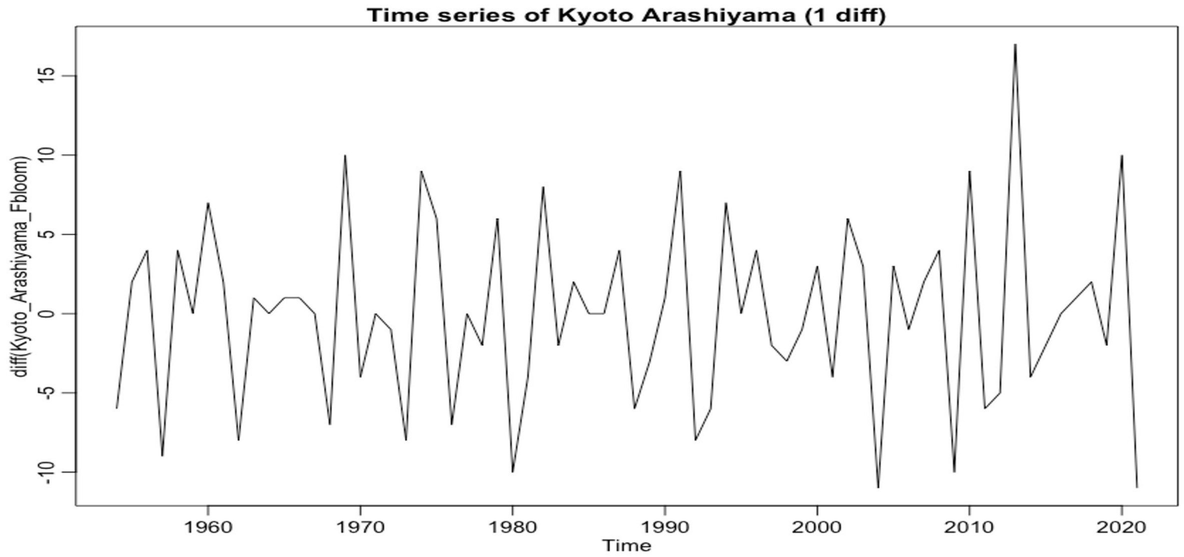


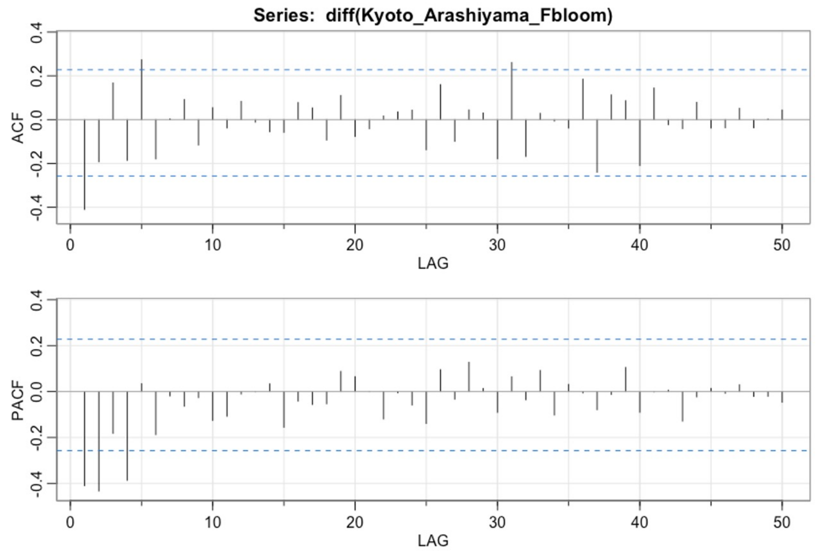
$$RMSE = \sqrt{\sum_{I=1}^{69} \frac{(\hat{y}_i - y_i)^2}{69}} = 1.61$$

Timeseries of Fbloom DOY of Arashiyama





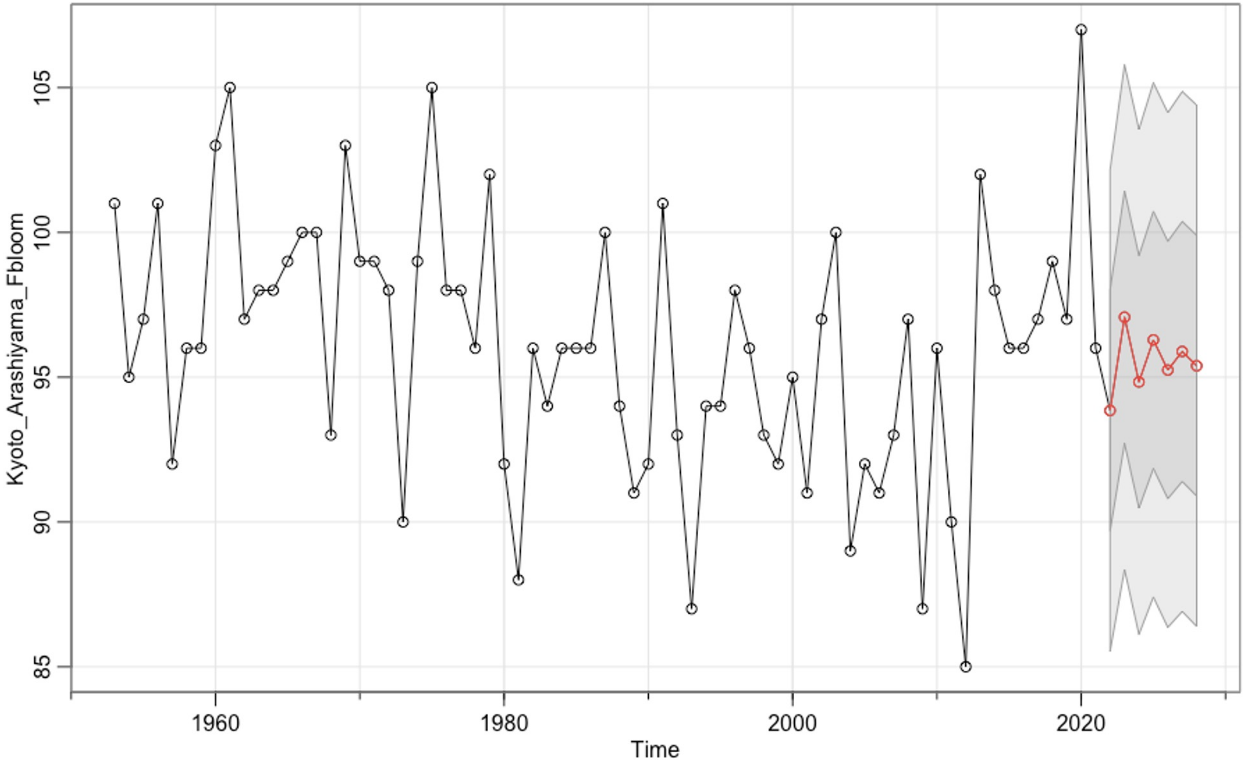


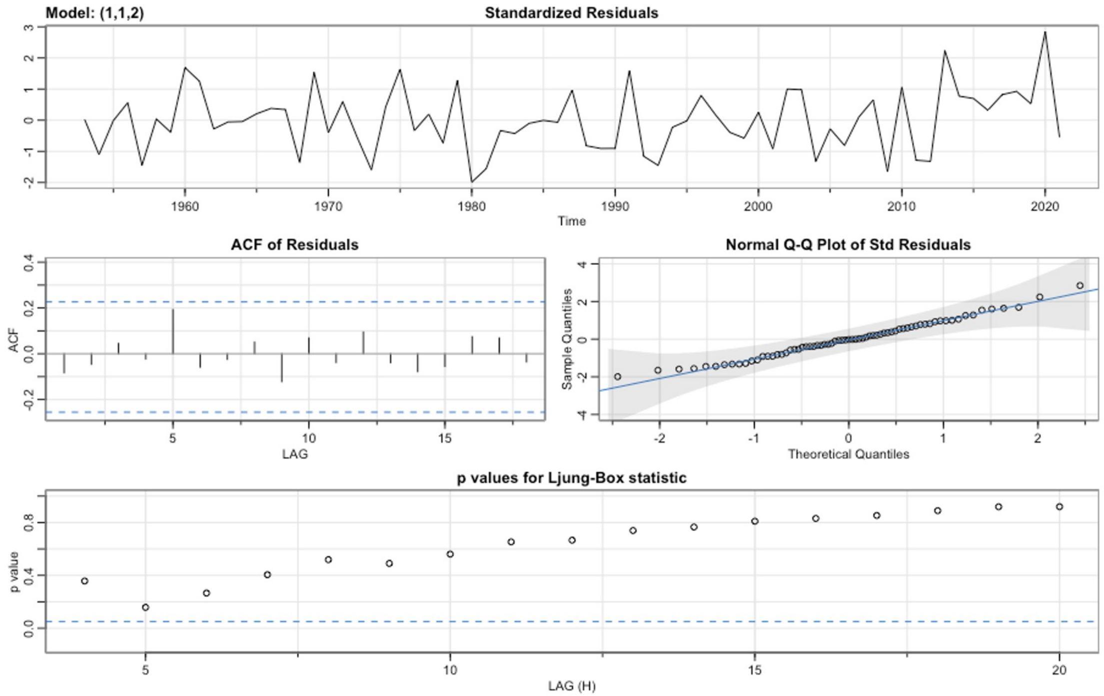


BIC table for model selection of ARIMA model

	q=0	q=1	q=2	q=3	q=4	q=5	q=6	q=7
p=0	438.465	406.189	410.395	410.06	412.496	415.838	419.024	423.224
p=1	429.327	410.401	**409.699**	411.726	417.5	419.7	423.238	423.238
p=2	418.892	412.338	411.818	415.905	415.745	416.07	426.816	429.345
p=3	420.866	416.532	416.012	419.984	418.26	420.243	423.651	427.765
p=4	412.046	415.296	416.724	416.351	420.275	422.66	426.487	429.988
p=5	416.194	418.627	420.52	420.395	424.07	427.798	432.012	436.147
p=6	416.636	420.64	424.566	424.044	428.065	430.199	434.356	439.334
p=7	420.762	424.03	427.738	431.956	434.422	434.408	437.775	442.613

Prediction of Full Bloom in Kyoto/Arashiyama for the incoming 7 years





AR1	MA1	MA2
-0.6851	0.0107	-0.7903

True 2022 Full Bloom DOY	Predicted 2022 Full Bloom DOY	Standard Error of prediction
93	93.85	4.16

Summary

- Climate change: from the historical full-flowering dates
- Data cleaning
- Data visualization: Fbloom day delays with higher latitude (temperature) in more recent years
- Additive model: Fbloom DOY ~ location + year
  R^2 = 0.97, Residual standard error = 3.81 (Good fitting)
- Prediction on future: Time Series ARIMA (1,1,2) model
  Fbloom day at Kyoto/Arashiyama in 2022: 94 ± 4

Thank you!

Figure above: adopted from https://theinvisibletourist.com/best-time-to-visit-japan-for-cherry-blossoms/