Boston Weekly Temperature Time Series Analysis

Course: MA 585 Time Series
Author: Ziran Min（U59274427)

Abstract

The goal of this project is to fit Boston weekly temperature data into Seasonal Autoregressive Integrated Moving Average (SARIMA) time series models and make prediction. The Boston daily maximum temperature data from 2008 to 2018 are collected from National Oceanic and Atmospheric Administration website, and the weekly temperature is computed by averaging every 7 days of daily observations. Through procedures of data transformation, model selection, model diagnostics, and forecasting, I find that SARIMA $(1,0,1)*(0,1,1)_{52}$ is the model with lowest AICc, but more comprehensive criteria are needed for selecting model that can forecast better.

Keywords: Time Series, Temperature, Boston, SARIMA, Model

1. Introduction and Data Selection

The weather in Boston is changeable every year. This is what I have learned since I came to Boston as a college freshman. You could wear a shirt one day and wear Canadian Goose the other day in one week, so the changeable temperature in Boston encourages me to study temperature data and conduct time series analysis.

The first idea comes into my mind is to find Boston daily temperature data and make prediction for the temperature on a specific day. I find Boston daily temperature data from the National Oceanic and Atmospheric Administration (NOAA) website. There is a Climate Record Station at Boston Logan Airport and I download the daily maximum temperature data from 2008 to 2018 from this Station Data Inventory. (Note: The link to Logan Airport Data Inventory is https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00014739/detail)

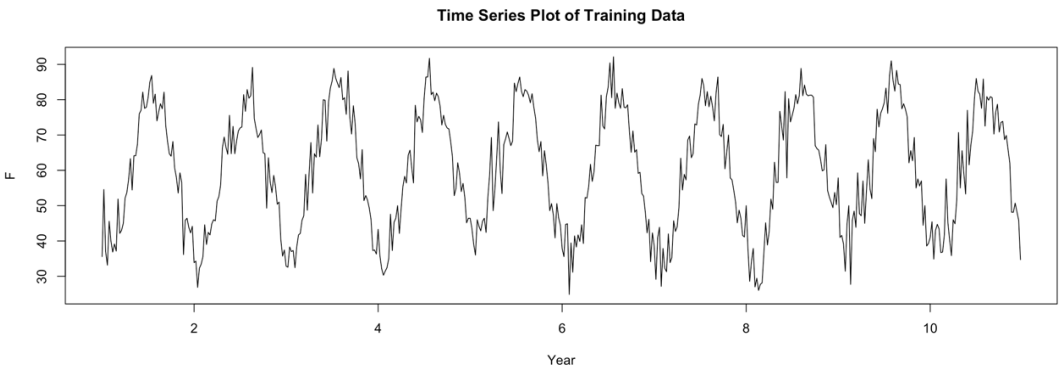
After downloading the data, I plan to fit model based on the first 10 years (2008 - 2017) of the data and test on the last year (2018). The period of my data is a year, so I think 10 years is enough for me to catch the seasonal pattern of the data. However, the daily temperature data has period of 365, and it is very time-consuming to fit these daily data into time series models in R. Therefore, I decide to convert daily data into weekly data by calculating the average of daily maximum temperature every 7 days.

Because 12/31/2007 and 12/31/2018 are both Mondays, I include the temperature on 12/31/2007 and exclude the temperature on 12/31/2018. In this way, every week starts on Monday and ends on Sunday.

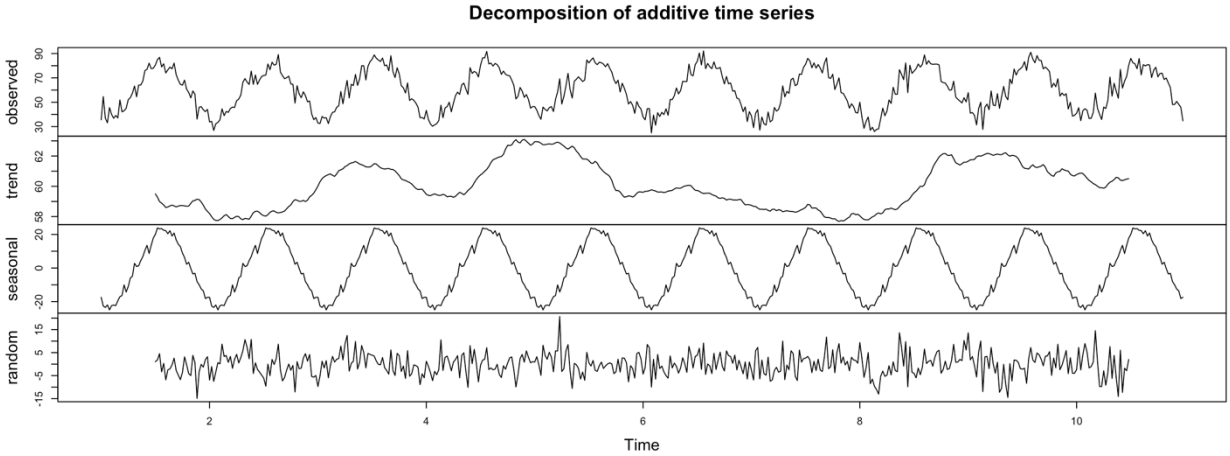
In the end, I have 574 weeks of Week Average Temperature (in °F). The first 520 weeks is the training set (when I use term "the data" in later sections of the paper, I mean the training data by default) for fitting into models, and the last 54 weeks is the testing set for forecasting.

2. Data Description and Data Transformation

The time series plot and the additive decomposition graph of the data are shown in Figure 1 and Figure 2.



(Figure 1)



(Figure 2)

The data seem have no trend pattern but do have a clear seasonal pattern. The variance doesn't very a lot as time *t* changes, so it is not necessary to carry out log-transformation.

To further determine whether we need order 1 differencing to remove potential trend component, I conduct augmented Dickey-Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. The output of two test are shown in Figure 3 and Figure 4.

```
> adf.test(train1)

        Augmented Dickey-Fuller Test

data:  train1
Dickey-Fuller = -7.7081, Lag order = 8, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(train1) : p-value smaller than printed p-value
```

(Figure 3)

```
> kpss.test(train1,null="Trend")

        KPSS Test for Trend Stationarity

data:  train1
KPSS Trend = 0.02725, Truncation lag parameter = 6, p-value = 0.1

Warning message:
In kpss.test(train1, null = "Trend") : p-value greater than printed p-value
```

(Figure 4)

In ADF test, p-value is smaller than 0.01, so we accept the alternative hypothesis that the data are stationary. In KPSS test, p-value is greater than 0.1, so we fail to reject the null hypothesis that the data are stationary. Therefore, both tests show that there is no unit root and I don't need order 1 differencing.

Because there are approximately 52 weeks a year, I remove seasonal component by differencing the data with lag = 52 (order 52).

3. Model Selection and Parameter Estimation

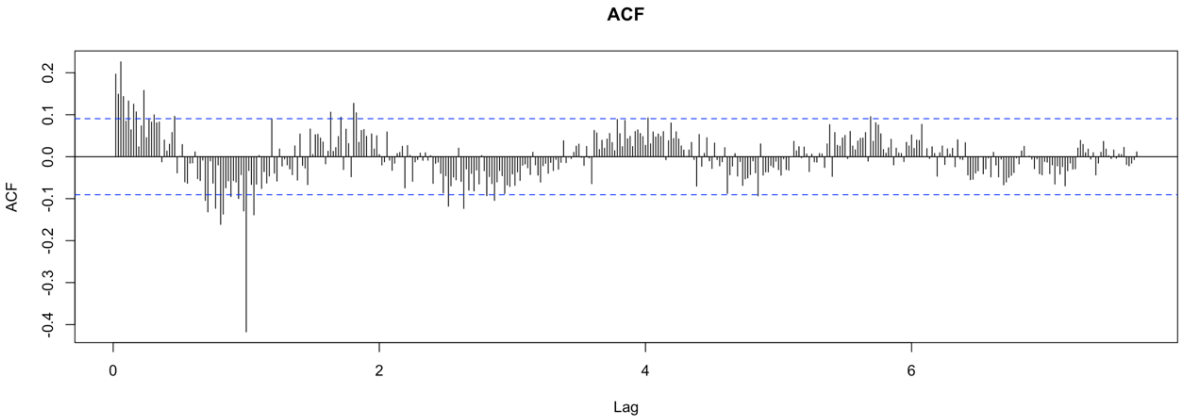
To select best model that fits the data, I first use the *auto.arima* function in R to see which is the best model it generates. The following Figure 5 shows the output.

```
> fit_auto = auto.arima(train1)
> fit_auto
Series: train1
ARIMA(0,0,1)(2,1,0)[52]

Coefficients:
         ma1     sar1     sar2
      0.1835  -0.6371  -0.3148
s.e.  0.0425   0.0483   0.0506

sigma^2 estimated as 41.09:  log likelihood=-1544.45
AIC=3096.9   AICc=3096.99   BIC=3113.49
```

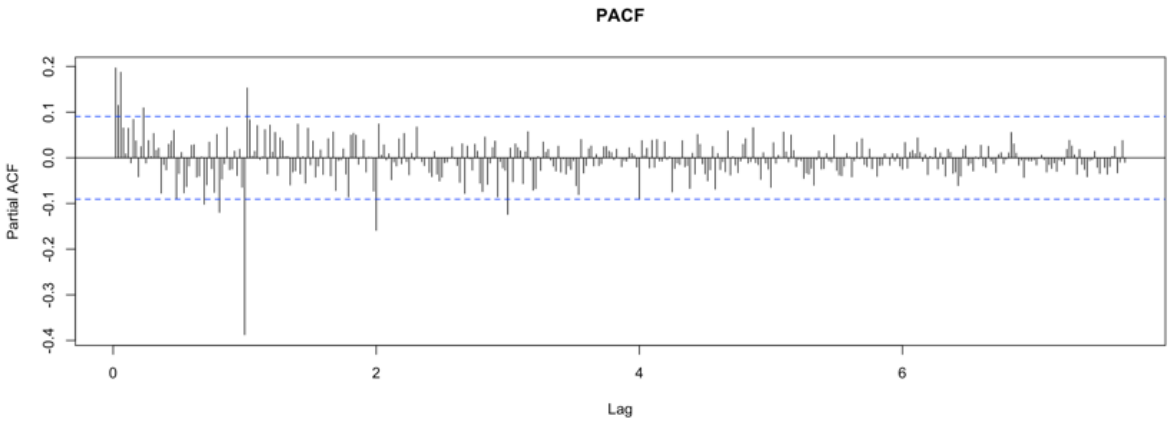
(Figure 5)

The *auto.arima* function selects SARIMA $(1,0,1) * (1,1,1)_{52}$ as the best model with AICc = 3096.99, but are there any other better models (with lower AICc)?

Then I compute the ACF and PACF graphs shown in Figure 6 and Figure 7.



(Figure 6)



(Figure 7)

To identify non-seasonal ARIMA(p, q) component, we can see that in the first period, at lag 1, 2, 3, and etc., both ACF (damped sine wave) and PACF decay to 0 exponentially, so a lower order ARIMA(p, q) model is applicable.

To identify seasonal SARIMA(P, Q) component, we can see that at the seasonal lags S, 2S, 3S, and etc., both ACF and PACF decay to 0 exponentially, so a lower order SARIMA(P, Q) model is applicable.

According to the above ACF and PACF analysis, I decide to set $p \leq 1$, $q \leq 1$, $P \leq 2$, and $Q \leq 2$. Furthermore, because I don't have trend differencing but a seasonal differencing of order 52, I set $d = 0$ and $D = 1$.

The following Table 1 shows all possible combinations of SARIMA model $(p, 0, q) * (P, 1, Q)_{52}$ that satisfy:

(1) RSudio doesn't give me "initial value in 'vmmin' is not finite" error when I try to fit data into to the model;
(2) At least one of p and q is not 0;
(3) At least one of P and Q is not 0.

p,d,q	*P, D, Q*	*S*	*AICc*
(1, 0, 1)	(1, 1, 1)	52	3025.47
(1, 0, 1)	(1, 1, 2)	52	3027.52
(1, 0, 1)	(1, 1, 0)	52	3104.86
(1, 0, 1)	(2, 1, 0)	52	3066.70
(1, 0, 1)	(0, 1, 1)	52	3023.63
(1, 0, 1)	(0, 1, 2)	52	3025.47
(0, 0, 1)	(1, 1, 1)	52	3059.61
(0, 0, 1)	(1, 1, 0)	52	3130.01
(1, 0, 0)	(0, 1, 1)	52	3052.92
(0, 0, 1)	(0, 1, 1)	52	3057.92
(0, 0, 1)	(2, 1, 0)	52	3096.99

(Table 1)

I also include SARIMA $(0,0,1)*(2,1,0)_{52}$, which *auto.arima* function selects, in the last row of Table 1. Now we can see that it is not the best model because many other models have lower AICcs than it.

SARIMA models $(1,0,1)*(1,1,1)_{52}$, $(1,0,1)*(1,1,2)_{52}$, $(1,0,1)*(0,1,1)_{52}$, and $(1,0,1)*(0,1,2)_{52}$ give me the top 4 lowest AICcs that are around 3023 to 3028. I set them as my candidate models.

Among four of them, SARIMA $(1,0,1)*(0,1,1)_{52}$ has the lowest AICc, 3023.63. Furthermore, Figure 8 shows the coefficients and standard deviations of this model. The 95% confidence intervals of the 3 coefficients are:

AR1 = (0.8987 – 1.96 * 0.0387, 0.8987 + 1.96 * 0.0387) = (0.8228, 0.9746)
MA1 = (-0.7405 – 1.96 * 0.0577, -0.7405 + 1.96 * 0.0577) = (-0.8536, -0.6274)
SMA1 = (-0.9985 – 1.96 * 0.2556, -0.9985 + 1.96 * 0.2556) = (-1.4995, -0.4975)

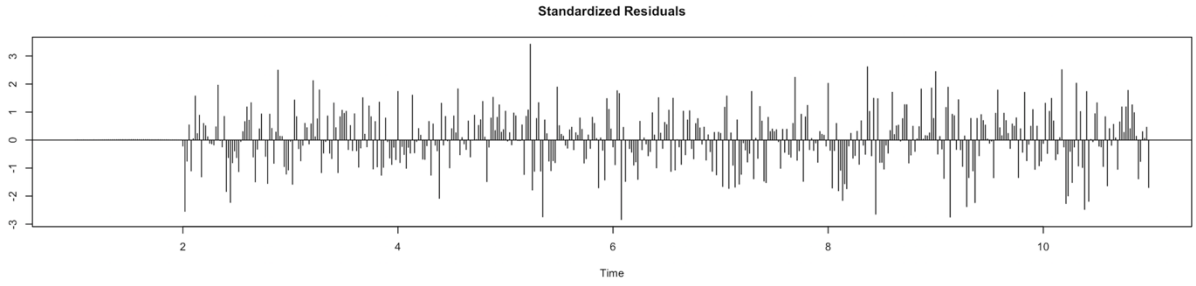
```
Series: train1
ARIMA(1,0,1)(0,1,1)[52]

Coefficients:
         ar1      ma1     sma1
      0.8987  -0.7405  -0.9985
s.e.  0.0387   0.0577   0.2556

sigma^2 estimated as 28.71:  log likelihood=-1507.77
AIC=3023.54   AICc=3023.63   BIC=3040.14
```

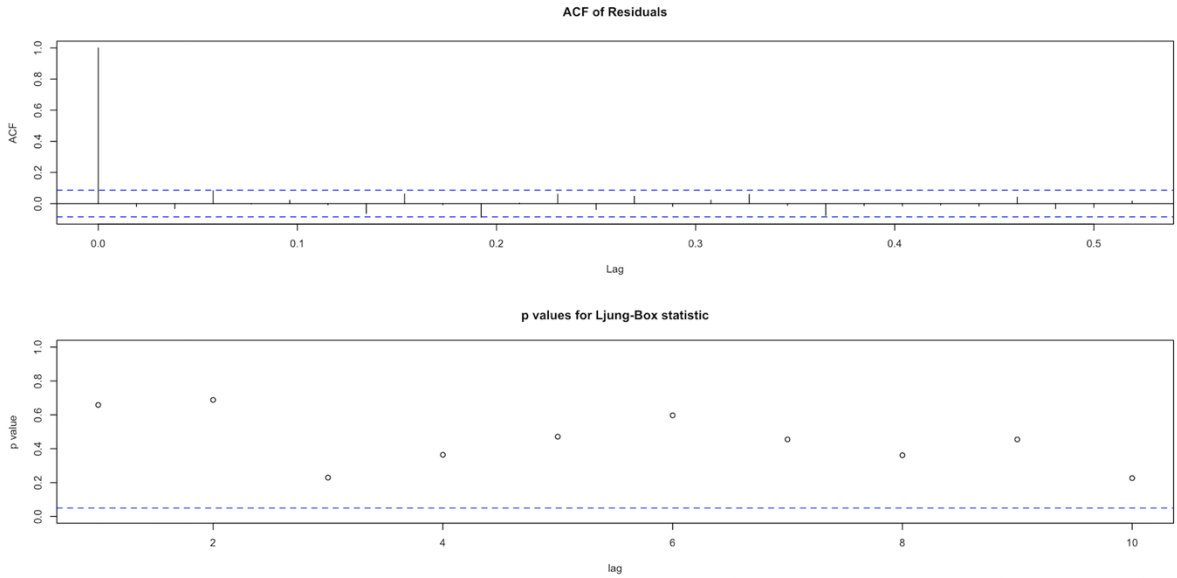
(Figure 8)

None of the 3 confidence intervals contains 0, so all coefficients are significantly different from 0. Now, SARIMA $(1,0,1)*(0,1,1)_{52}$ could potentially be my best final model. I will implement model diagnostics on it in next section, but use all 4 candidate models for forecasting.

4. Model Diagnostics

In order to determine whether SARIMA $(1,0,1)*(0,1,1)_{52}$ is a valid best fitted model, I need to analyze its residuals. The following Figure 9 shows the diagnostics plots of model SARIMA $(1,0,1)*(0,1,1)_{52}$.





(Figure 9)

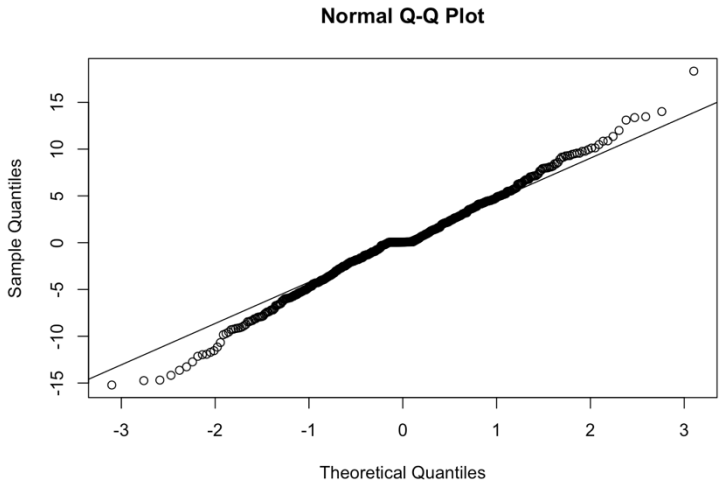
We can know that:

(1) In the first plot, standardized residuals are random distributed and don't have any special pattern;

(2) In the second plot, for lag ≥ 0, ACF are all insignificant and lay inside the confidence interval (i.e. equal to 0);

(3) In the third plot, p-values in all lags are significantly above the confidence interval, so we fail to reject the null hypothesis that all ACFs equal to 0 and conclude that all ACFs equal to 0 (i.e. residuals are independent)

From the Q-Q Plot in Figure 10, we also know that residuals are normally distributed.



(Figure 10)

The last step of model diagnostics is to fit residuals into an ARIMA model. If the best fitted model is an ARIMA(0, 0, 0) model with zero mean, then the residual is a White Noise Process and the original model is a valid model.

From Figure 11 (*final_try5* is the fitted SARIMA $(1,0,1)*(0,1,1)_{52}$ model), we can know that the residual of SARIMA $(1,0,1)*(0,1,1)_{52}$ model is indeed a White Noise Process.

```
> auto.arima(residuals(final_try5))
Series: residuals(final_try5)
ARIMA(0,0,0) with zero mean

sigma^2 estimated as 25.67:  log likelihood=-1581.65
AIC=3165.29   AICc=3165.3   BIC=3169.54
```

(Figure 11)

Therefore, SARIMA$(1,0,1)*(0,1,1)_{52}$ is a valid potential best model.

In Section 3, SARIMA models $(1,0,1)*(1,1,1)_{52}$, $(1,0,1)*(1,1,2)_{52}$, and $(1,0,1)*(0,1,2)_{52}$ give me AICcs that are little bit greater than the AICc of SARIMA $(1,0,1)*(0,1,1)_{52}$. I also carried out diagnostics for these three models, they all return the same good results as the SARIMA $(1,0,1)*(0,1,1)_{52}$ shows above.

In the next section, I will use all four of them to implement forecasting and compare the performance to furthermore determine whether SARIMA$(1,0,1)*(0,1,1)_{52}$ is my best model.

5. Forecasting

Forecast accuracy is an important aspect of judging a model efficiency in time series analysis. I use 10 years (2008 - 2017) of weekly temperature as training data to predict the weekly temperature of the next year (2018), by my potential best model SARIMA $(1,0,1)*(0,1,1)_{52}$, and other three candidate SARIMA models $(1,0,1)*(1,1,1)_{52}$, $(1,0,1)*(1,1,2)_{52}$, and $(1,0,1)*(0,1,2)_{52}$, and Holt Winters forecasting method.

To evaluate the accuracy of forecasting, I compute the measures of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAE) of each model or method. The results are shown in Tables 2.

	MAE	RMSE	MAPE
Holt Winters	6.2046	7.9557	13.2648
(1,0,1)*(0,1,1)_52	4.8402	6.2881	10.7176
(1,0,1)*(0,1,2)_52	4.8347	6.2779	10.7084
(1,0,1)*(1,1,1)_52	4.8350	6.2778	10.7095
(1,0,1)*(1,1,2)_52	4.8292	6.2799	10.6981

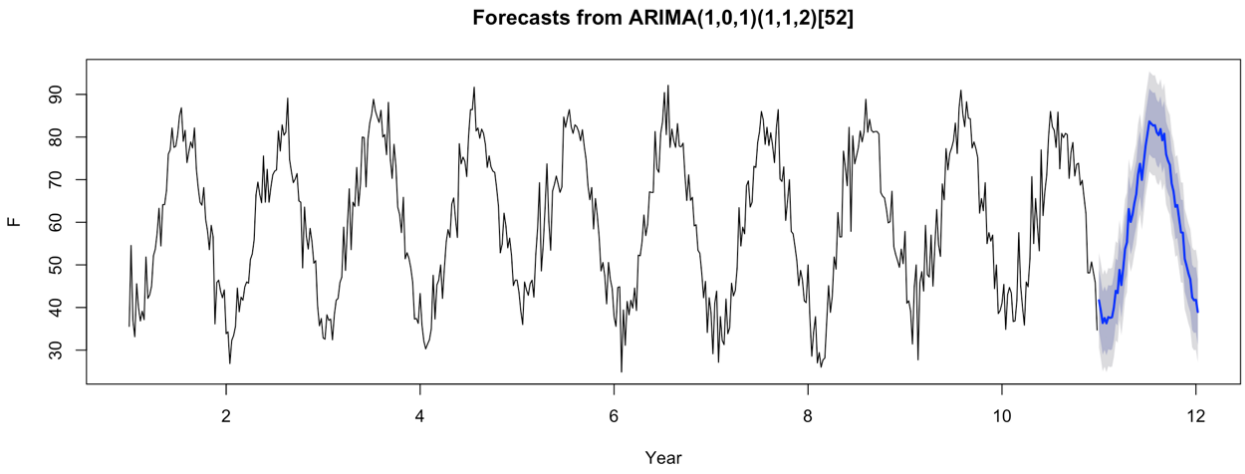
(Table 2)

Firstly, compared with SARIMA models, Holt Winters method gives me worst forecasting accuracy. This may imply that SARIMA model is a better way to forecast seasonal data than exponential smoothing method like Holt Winters.

Secondly, the potential best model SARIMA$(1,0,1)*(0,1,1)_{52}$, surprisingly doesn't have the lowest value in any one of the 3 measures. Instead, it has the highest MAE, RMSE, and MAPE among 4 candidate SARIMA models.

Thirdly, it's very interesting to notice that SARIMA $(1,0,1)*(1,1,2)_{52}$ has the lowest MAE, second lowest RMSE, and lowest MAPE, so it has the best overall forecasting performance. However, it has the highest AICc among 4 candidate models.

In Figure 12, we can see that SARIMA $(1,0,1)*(1,1,2)_{52}$ indeed forecasts the 11th year of weekly temperature in a pattern that looks very similar to the previous 10 years.



(Figure 12)

The rest 2 SARIMA models, $(1,0,1)*(0,1,2)_{52}$ and $(1,0,1)*(1,1,1)_{52}$, have the same AICc, 3025.47, and very similar accuracy measures, so it is worth to study more about their similarities for later study.

Therefore, the model with lowest AICc doesn't have the best forecasting accuracy. We shouldn't select the best model only based on one measure, AICc. Considering both AICc and 3 accuracy measures, I would rather to choose SARIMA $(1,0,1)*(0,1,2)_{52}$ or SARIMA $(1,0,1)*(1,1,1)_{52}$, which are ranked in the middle, than the other two as the best model.

6. Conclusion

With the goal of analyzing Boston Temperature Time Series data, I find 11 years (2008 to 2018) of Boston Daily data from NOAA website. After transforming daily data to weekly data and further differencing the seasonal component, I try to fit the first 10 years data into the best SARIMA model (with lowest AICc).

I find that *auto.arima* function fails to select the model with the lowest AICc. By looking the ACF and PACF graphs and comparing AICcs, I select the following 4 models, which have the lowest 4 AICcs, as candidates for the best model:

SARIMA $(1,0,1) * (0,1,1)_{52}$
SARIMA $(1,0,1) * (0,1,2)_{52}$
SARIMA $(1,0,1) * (1,1,1)_{52}$
SARIMA $(1,0,1) * (1,1,2)_{52}$

All of them show that residuals are White Noise Processes during model diagnostics. Then after implementing forecasting by these 4 models, I find that:

SARIMA models forecast my data better than Holt Winters method;
SARIMA $(1,0,1) * (0,1,1)_{52}$ has the lowest AICc, but worst forecasting accuracy;
SARIMA $(1,0,1) * (1,1,2)_{52}$ has the best forecasting accuracy, but highest AICc;
SARIMA $(1,0,1) * (0,1,2)_{52}$ and SARIMA $(1,0,1) * (1,1,1)_{52}$ are two similar models.

In the end, I conclude that if I only select best model based on AICc, SARIMA $(1,0,1) * (0,1,1)_{52}$ is my best model. However, in order to select a model that can perform better in forecasting, it is worthwhile for me to consider some more comprehensive model selection criteria rather just ranking models by one measure.