Time Series and Forecasting: Mean Temperature in the city of Delhi

Bruno Vaz[1]
[1]Department of Computer Science

Student ID:201705247

Keywords: R, Holt-Winters, TBATS, NNAR

24th October, 2021

Abstract

Predicting temperatures can be a powerful tool in different areas of science. In this report, the characteristics of the time series were firstly explored, followed by the use of three types of methods to fit the time series data and obtain forecasts: Holt-Winters' seasonal method, TBATS and Artificial Neural Networks. Lastly, the models were compared and it was assessed which ones could be more useful.

1 Introduction

In the last few years the concerns regarding global warming, climate change and the increase of the mean temperature of the planet have been growing stronger. In this report, the daily mean temperature in the city of Delhi, India*, from the 1^{st} of January 2013 to the 24^{th} of April 2017, was studied.

Firstly, it was performed exploratory data analysis, so as to have an idea of how the series behaved and to see its main characteristics. Next, three types of methods were used - Holt-Winters, TBATS and Artificial Neural Networks - to model the time series and make forecasts. Lastly, some comments were made regarding each model and they were compared with each other to decide which was better at capturing the time series patterns and making forecasts.

The analysis performed† was done using the software *R*, along with some of its packages, namely *fpp2*, *tseries* and *astsa*.

2 Exploratory Data Analysis

2.1 Plotting the time series

First and foremost, it was important to have an idea of how the time series behaved over time. Figure 1 shows the values of temperature from the beginning of the year 2013 up to the end of April, 2017.

By analyzing Figure 1, the annual seasonal patterns were evident - the temperatures vary with the seasons. The peaks ought to correspond to the warmer months and the troughs to the colder months. It was interesting to note that, in all years, the temperatures rapidly increased from the period of January to mid-May (where they reached the peaks). Right after

*https://www.kaggle.com/sumanthvrao/daily-climate-time-series-data

†The code used to perform the analysis can be found at: https://github.com/2010b9/Forecasting-Mean-Temperature-in-the-city-of-Delhi

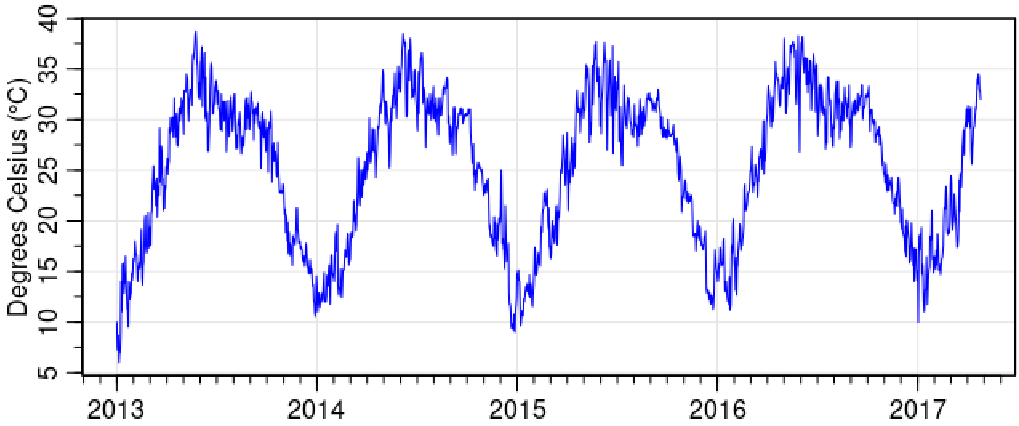
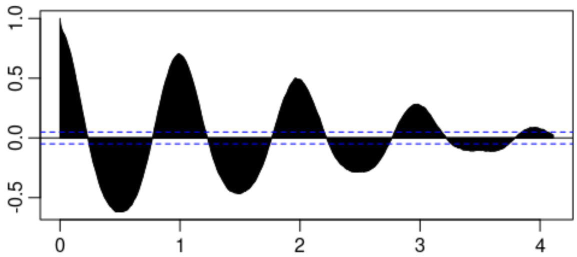


Fig. 1. Mean Temperature in the city of Delhi, India.

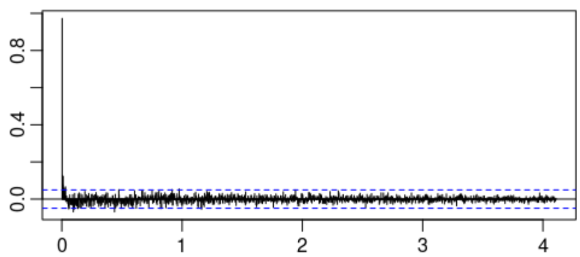
that, there were periods of roughly 3 months in which the values of temperature decreased very slowly, after which they felt abruptly until January. Besides the seasonality, it seemed to be present a slight upward trend, meaning that the mean temperature had been increasing. This may be caused by climate change and global warming.

2.2 Sample ACF and PACF

Another relevant features of a time series are its ACF and PACF, which are represented in Figures 2a and 2b, respectively.



(a) Sample ACF.



(b) Sample PACF.

Fig. 2. Sample ACF and PACF of the mean temperature at the city of Delhi, India.

From Figure 2a, it was possible to see that the correlation between observations decreased very slowly as the lag increased. It was also possible to observe the cycles very clearly - a consequence of the seasonality observed in the time series plot. The amplitude of these cycles decreased as the lag increased. Figure 2b showed that, despite the sample PACF being very high at the first lag, it rapidly tended to 0 and remained near 0, meaning there was no correlation between observations which were a certain number of lags apart, if we neglected the effects of the observations between them.

Note that, a stationary time series is one whose properties do not depend on the time at which the series was observed. Thus, since the time series presents seasonality, it is non-stationary: the seasonality would affect the value of the time series at different times.

3 Holdout

Having a general overview of the whole time series, and since one of the purposes of this project is to perform forecasting, the time series was divided in train and test sets. The former consisted of the observations from the 1^{st} of January 2013 to the 25^{th} of April 2016. The latter contained the data from the 26^{th} of April 2016 to the 24^{th} of April 2017. In the following sections, only the train set was used to derive models and make forecasts. The test set was used to evaluate the quality of the predictions, so as to compare different models and discuss results.

The term "time series" or just "series" may be used to refer to the train set. If a reference is made to the test set, it will be carefully pointed out.

4 Modelling and Forecasting

Since the time series has a high frequency (daily observations), and due to the seasonal complexities, SARIMA models were not able to model and make forecasts of the time series accurately. Hence, more appropriate methods were used. In Sub-Section 4.1 the Holt-Winters' seasonal method was used, followed by the TBATS method in Sub-Section 4.2 and Artificial Neural Networks in Sub-Section 4.3.

4.1 Holt-Winters

The Holt-Winters' seasonal method extends simple exponential smoothing to allow the forecasting of data with trend and to capture seasonality. This method involves a forecast equation and three smoothing equations: one for the level, one for the trend, and one for the seasonal component.

There are two variations to this method that differ in the nature of the seasonal component. The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportionally to the level of the series (see Sections 7.2 and 7.3 of [1])[‡].

By the analysis conducted in Sub-Section 2.1, the additive method seems more appropriate. However, both additive and multiplicative methods were studied.

[‡]R automatically calculates three parameters - α (concerning the level), β (concerning the trend) and γ (concerning the seasonality) - by minimizing the squared error of the one-step ahead forecast. Three initial conditions (values) are also automatically estimated.

4.1.1 Holt-Winters' additive method

First, an additive model was used to fit the time series. Figure 3 shows the observations of the time series (that is, the mean temperature in the city of Delhi), in blue, while the red line represents the fitted values using the Holt-Winters' additive model.

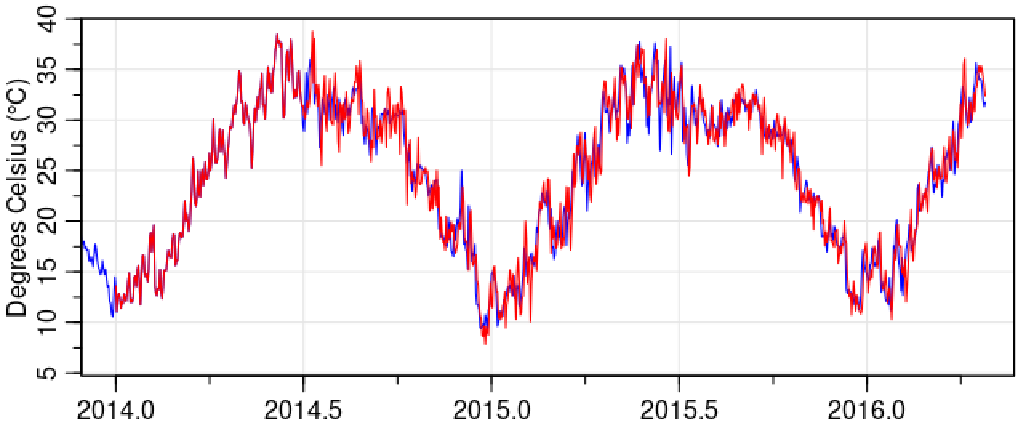


Fig. 3. Holt-Winters' additive model. The blue line represents the original time series and the red line the fitted values given by the Holt-Winter's additive model.

It is possible to see that the model seems to be fitting the time series very accurately. In fact, too much accurately in the first half-year, which may indicate over-fitting. To confirm this suspicion, the residuals were analysed. Figure 4 shows a plot of the residuals and Figure 5 a histogram of the residuals along with a normal distribution of mean zero (red line).

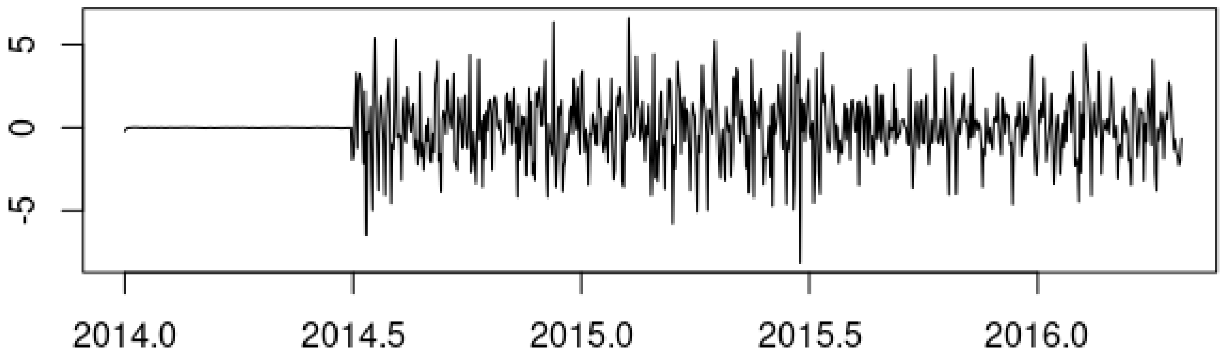


Fig. 4. Plot of the residuals (Holt-Winters' additive model).

From Figure 4 it can be seen the almost constant line that goes from the beginning of 2014 up to the middle of the same year (around June or July), which indicates that, indeed, it may have been over-fitting of the series in this period. This is reflected in the histogram (Figure 5), where there are two noticeable peaks around zero. Nevertheless, if we ignore those peaks, the histogram seems to follow a similar distribution to the one represented by the red line (a Gaussian distribution).

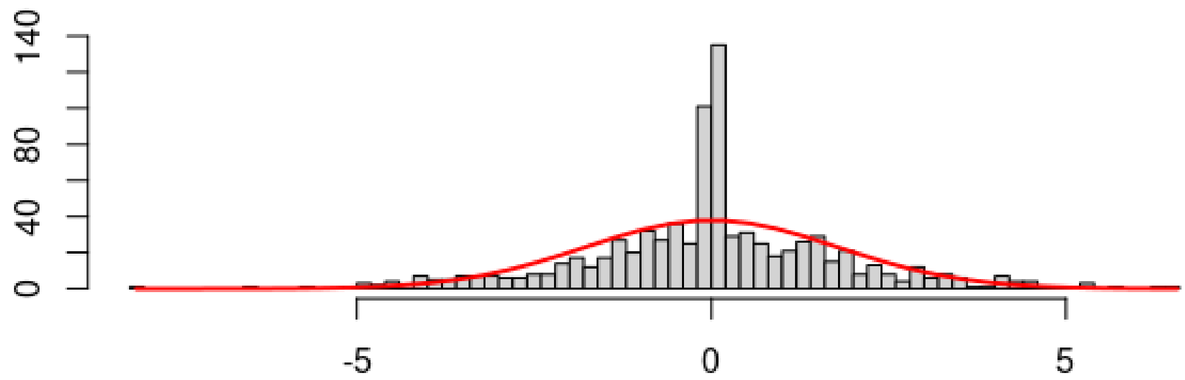
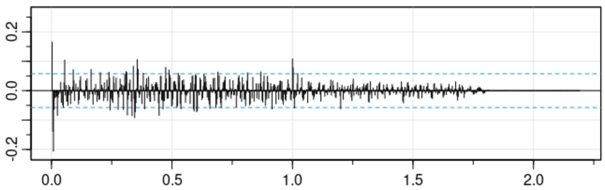
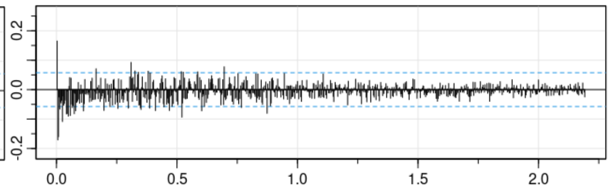


Fig. 5. Histogram of the residuals (Holt-Winters' additive model). The red line represents the probability density function of a normal distribution with mean 0.

To assert if the residuals could be considered white noise, their auto-correlation and partial auto-correlation functions were analysed. Figure 6 shows the ACF and PACF of the residuals.



(a) ACF of the residuals.



(b) PACF of the residuals.

Fig. 6. ACF and PACF of the residuals.

On the one hand, the ACF decreases as the lag increases and by the end of the second year it is almost zero, which means that the residuals become less correlated as they become further apart. On the other hand, there are a lot of values outside the significance threshold during the first year. This indicates that some correlation in the series was not explained. There is also a spike exactly a year after the first lag: this is common because of the seasonality of the series.

The analysis of the PACF is quite similar. The values decrease and tend to zero as the lag increases, meaning that the correlation between the residuals which are a certain number of lags apart diminishes, if we neglect the effects of the residuals between them. Like in the ACF, there are some spikes outside the significance threshold during the first year: some partial correlation was not explained.

Next, it was performed forecasting, and the predicted values were compared with the real ones (in the test set). Figure 7 shows a plot of the real values (red line) and the predicted values (blue line) during the period between the 25^{th} of April 2016 and the 24^{th} of April 2017, along with an 80% and a 95% confidence intervals.

Although the analysis of the residuals showed some correlation, the predictions appear to be very precise, since the real values (red) and the predicted ones (blue) are very close.

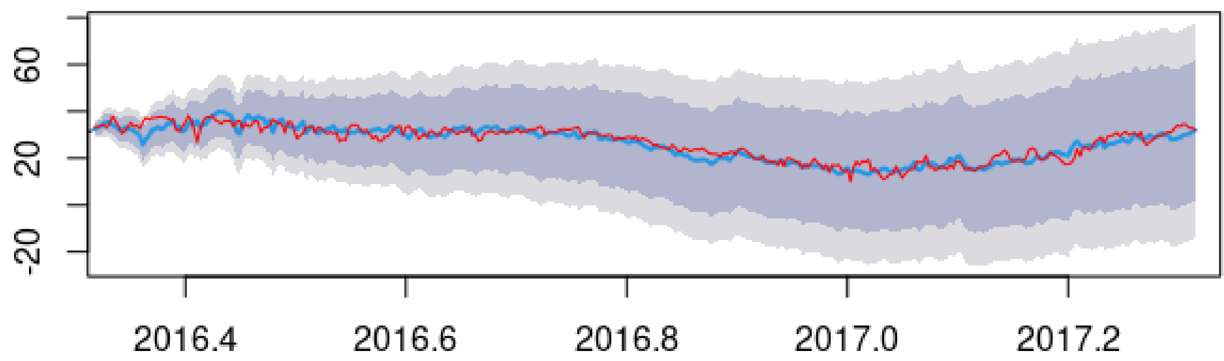


Fig. 7. Forecasting results in the test set (Holt-Winters' additive model). The red line represents the test set values, the blue line the predicted values and the lighter blue lines an 80% and 95% confidence intervals.

4.1.2 Holt-Winters' multiplicative method

To start, a multiplicative model was used to fit the time series. Figure 8 shows a plot of the time series (blue line) and the fitted values using a multiplicative Holt-Winters' model (red line).

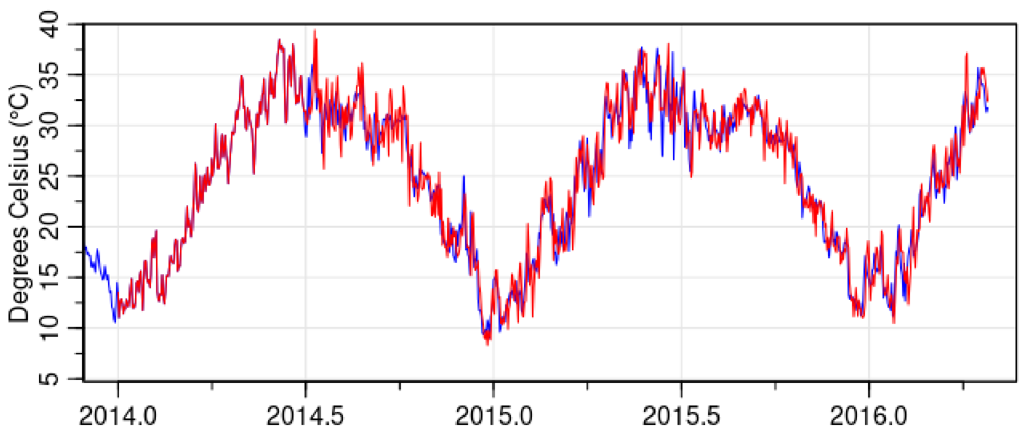


Fig. 8. Holt-Winters' multiplicative model. The blue line represents the original time series and the red line the fitted values given by the Holt-Winter's multiplicative model.

Similarly to the additive model, the multiplicative one also manages to model the series very accurately. In fact, in the first half-year, the real values and the fitted ones seem to overlap - there may have been some over-fitting. To analyse this further, an inspection of the residuals was performed, as previously. Figure 9 shows the plot of the residuals and Figure 10 a histogram of the residuals along with a normal distribution of mean zero (red line).

The residuals obtained using the multiplicative method are quite similar to the ones obtained using the additive method (Figure 4). There is an almost straight line in the period between the start of 2014 and the half of 2014, meaning that, indeed, it may have been over-fitting of the data in this period. The effects of the over-fitting in the first half-year are visible in the histogram, where two peaks around zero are shown. Note that, the normal distribution seems

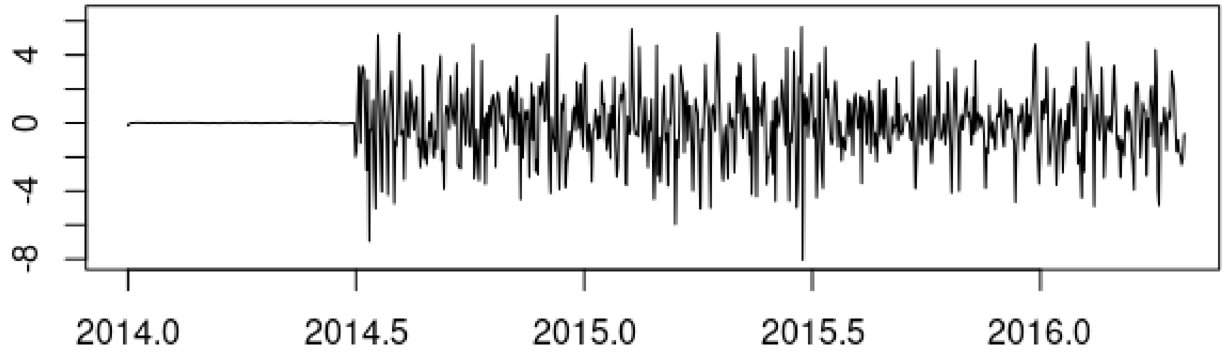


Fig. 9. Plot of the residuals (Holt-Winters' multiplicative model).

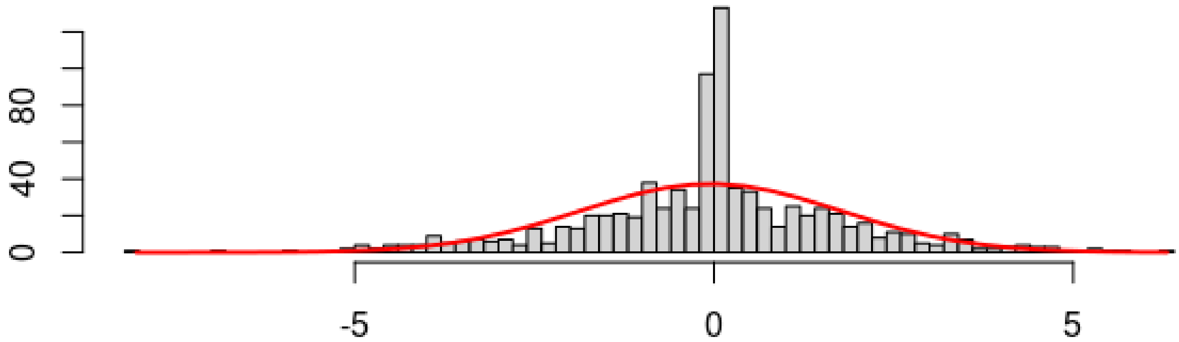
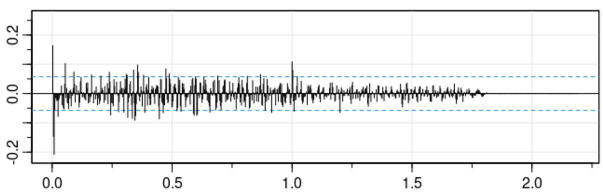


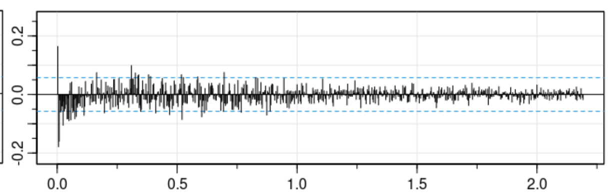
Fig. 10. Histogram of the residuals (Holt-Winters' multiplicative model). The red line represents the probability density function of a normal distribution with mean 0.

to approximate the distribution given by the histogram if we don't consider the two mentioned peaks.

A more thorough analysis of the residuals was conducted, so as to assert if there was unexplained correlation. Figure 11 shows the auto-correlation and partial auto-correlation functions of the residuals.



(a) ACF of the residuals.



(b) PACF of the residuals.

Fig. 11. ACF and PACF of the residuals.

Since the plot of the residuals was similar to the one regarding the additive model, it was expected that both ACF and PACF were also similar. Indeed, this is what happens. Both decrease to zero as the lag increases and, in the first year there are some values outside the significance threshold, which indicates that some correlation and partial correlation was not well explained. Once again, there is a spike in the ACF plot exactly a year after the first lag, which is due to the seasonality of the series.

After the analysis of the residuals, forecasting was performed. Figure 12 shows a plot of the real values (red line) and the predicted values (blue line) during the period between the 25^{th} of April 2016 and the 24^{th} of April 2017, along with an 80% and a 95% confidence intervals.

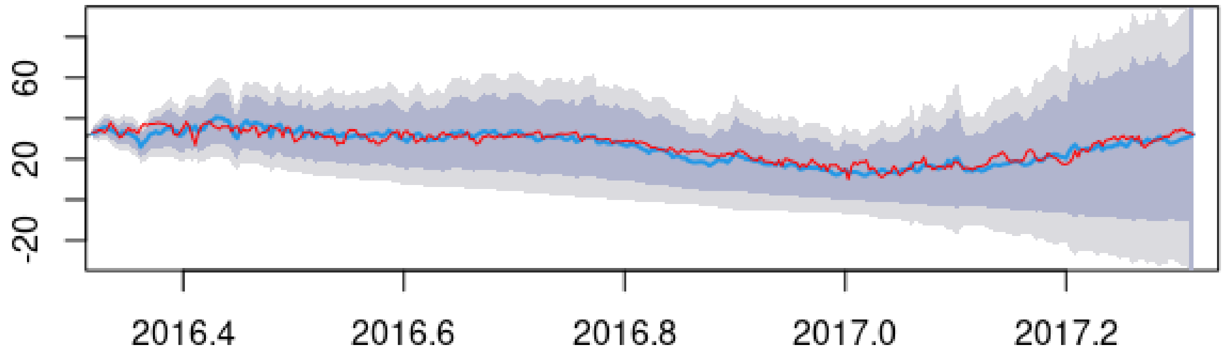


Fig. 12. Forecasting results in the test set (Holt-Winters' multiplicative model). The red line represents the test set values, the blue line the predicted values and the lighter blue lines an 80% and 95% confidence intervals.

Once again, the predictions appear to be very accurate, despite the analysis of the residuals having shown that there was some correlation that was not explained.

4.2 TBATS

TBATS models use a combination of Fourier terms with an exponential smoothing state space model and a Box-Cox transformation in a completely automated manner. They also allow for the seasonality to change slowly over time (see [2] and/or Section 11.1 of [1])[§].

First, a TBATS model was fitted into the time series. Figure 13 shows both the time series (blue line) and the fitted values of the TBATS model (red line).

Visually, it appears that the TBATS model did a good job fitting the series. So next, an inspection of the residuals produced by the model was made. Figure 14 show the plot of the resulting residuals and Figure 15 a histogram of the residuals along with a normal distribution of mean zero (red line).

By analysing the plot of the residuals, some spikes at the beginning of 2013, the end of 2014 and the middle of 2015 were visible. Nevertheless, the residuals appeared not to have any trend nor seasonality. Regarding the histogram, the distribution of the residuals seemed to be a little skewed to the left.

[§]To choose the final model, TBATS will consider various alternatives and fit quite a few models. It will consider models with a Box-Cox transformation (and without it), with and without trend, with and without an ARMA(p,q) process to model the residuals, with various amounts of harmonics used to model seasonal effects, etc. The final model is chosen using Akaike information criterion (AIC)

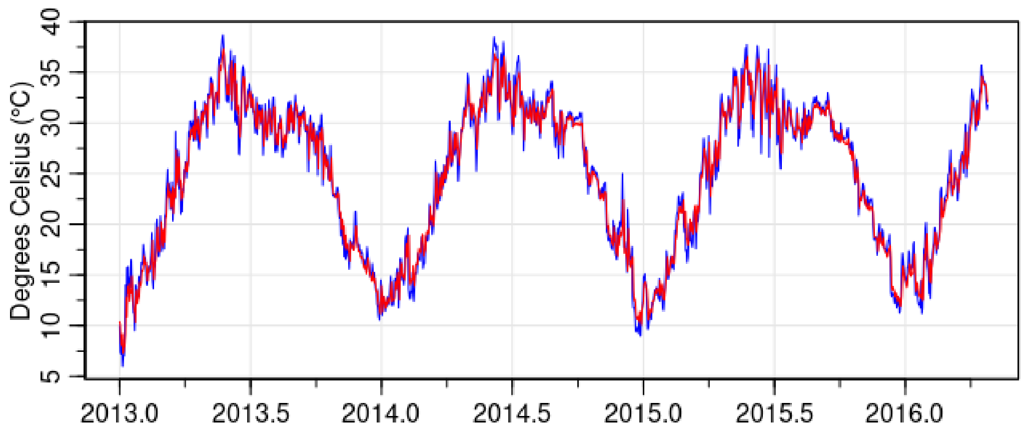


Fig. 13. TBATS model. The blue line represent the original time series and the red line the fitted values of the TBATS method.

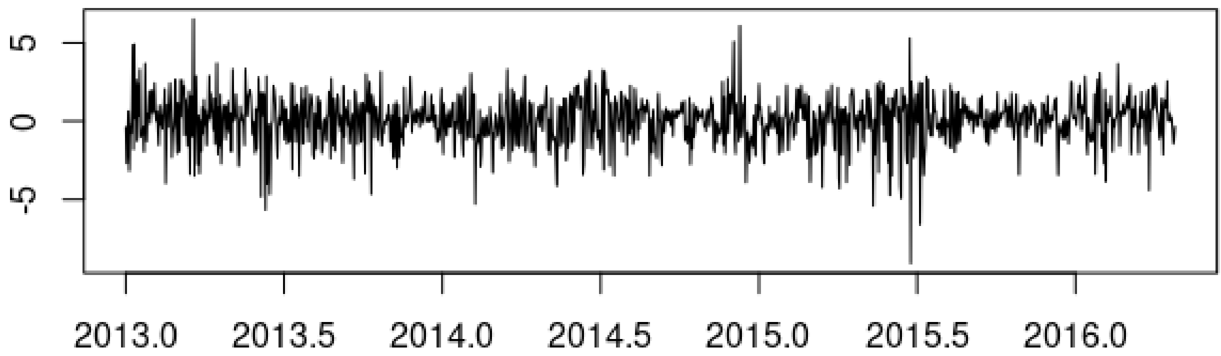


Fig. 14. Plot of the residuals (TBATS model).

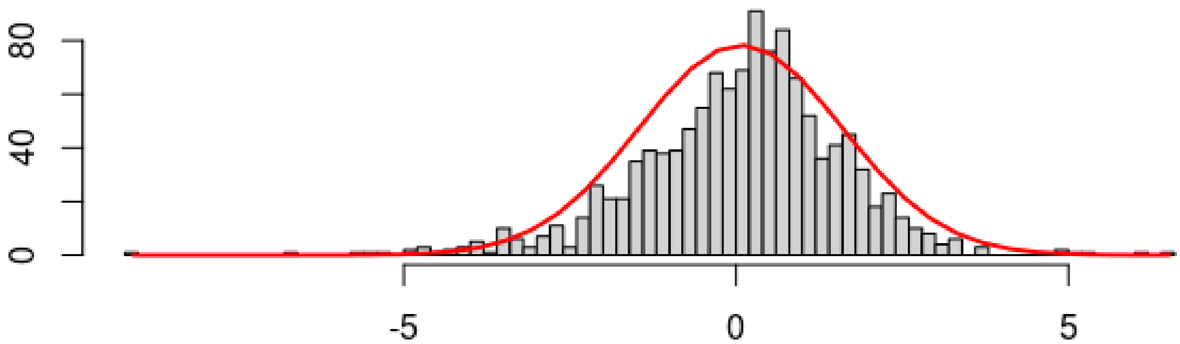
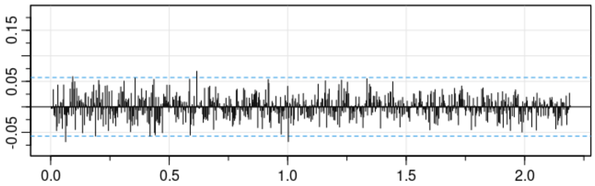


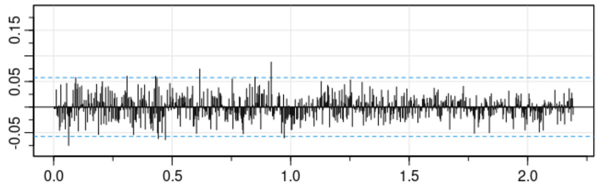
Fig. 15. Histogram of the residuals (TBATS model). The red line represents the probability density function of a normal distribution with mean 0.

In order to continue the analysis of the residuals, a plot of its auto-correlation and partial auto-correlation functions was done. Figure 16 shows the ACF and PACF of the residuals.

Regarding the ACF, it was noticeable that almost all of its values were inside the significance threshold. For the few values which were not inside the significance threshold, their distance to the closest boundary (dotted blue lines) was very small. However, unlike the ACFs from



(a) ACF of the residuals.



(b) PACF of the residuals.

Fig. 16. ACF and PACF of the residuals.

both Holt-Winters' residuals, the ACF values of the residuals produced by the TBATS' model decay very slowly. It was possible to assert that there was very little correlation between the residuals at different lags.

The PACF behaved in a similar manner. The majority of the values were inside the significance threshold and the ones who were not, it was only by a tiny distance. This meant that residuals in any two given lags were not correlated if we ignored the effects of the residuals in between.

The p-values from the Ljung-Box test statistic were also used to see if there was no correlation between the residuals. Figure 17 shows the scatter plot of the p-values given by the Ljung-Box test statistic along various lags.

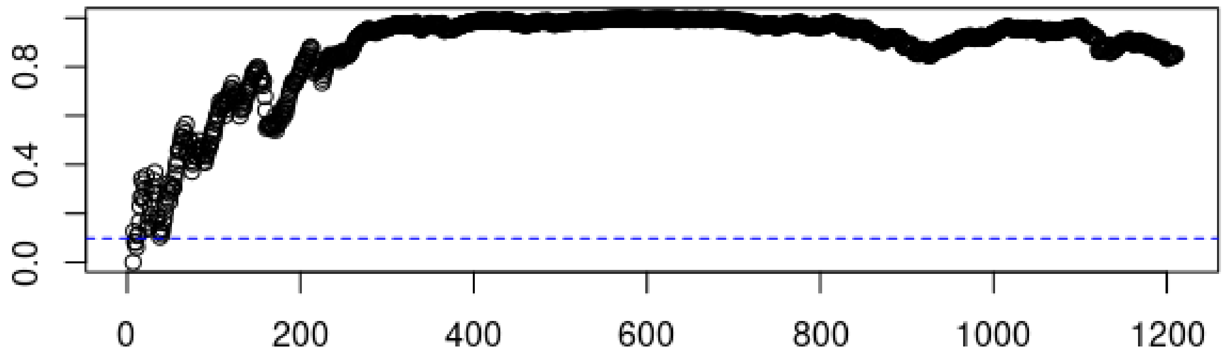


Fig. 17. Scatter plot of the p-values given by the Ljung-Box test statistic along various lags (TBATS model).

At the first few lags, some of the p-values were below the significance threshold. However, as the number of lags increased so did the p-values. Thus, the null hypothesis that the residuals are correlated was rejected and it was assumed that the residuals could be regarded as white noise (with some degree of certainty).

Since the model was fitting the series well, forecasting was performed. Figure 18 shows the real values of the test set (red line), the predicted values (blue line) and an 80% and a 95% confidence intervals (light blue lines).

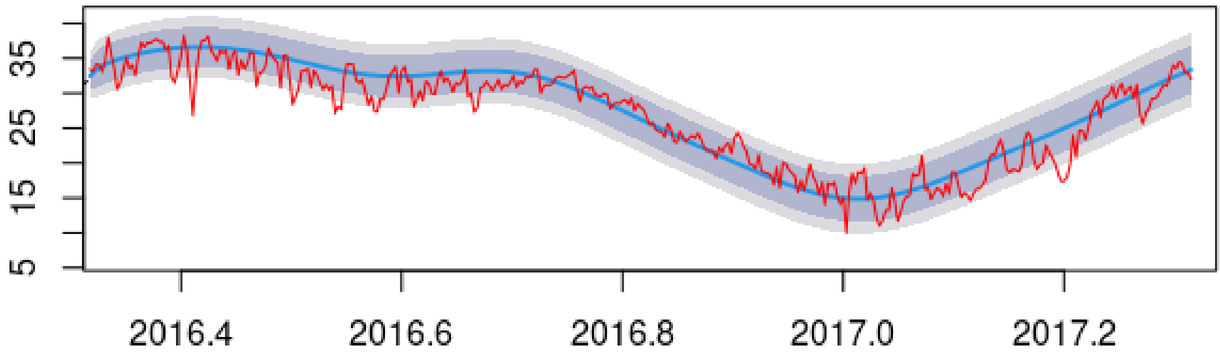


Fig. 18. Forecasting results in the test set (TBATS model). The red line represents the test set values, the blue line the predicted values and the lighter blue lines an 80% and 95% confidence intervals.

The TBATS model seemed to forecast very accurately. Also, it is noteworthy to mention that the blue line (predicted values) is much smoother than the ones obtained using both the additive (Figure 7) and multiplicative (Figure 12) Holt-Winters' models.

4.3 Artificial Neural Networks

A neural network can be thought of as a network of "neurons" which are organised in layers. The predictors (or inputs) form the bottom layer, and the forecasts (or outputs) form the top layer. There may also be intermediate layers containing "hidden neurons". (See Section 11.3 of [1])

To start, an artificial neural network[¶] model was fitted to the time series. Figure 19 shows the original time series (blue line) and the fitted values given by the neural network (red line).

As with the previous models, the NNAR seems to be fitting the series (train set) very accurately. Thus, the analysis of the residuals followed. Figure 20 shows a plot of the resulting residuals and Figure 21 a histogram of the residuals along with a normal distribution of mean zero (red line).

From the plot of the residuals, some high values were visible: a positive one almost at the end of 2014, and two negative ones around the half of 2015. Nonetheless, there was no trend nor seasonality in the residuals, which may be an indication that the series was well fitted. The histogram appeared to be very symmetric around zero and to approximate the normal distribution - not skewed, unlike the one from TBATS, nor with high peaks, unlike the ones from Holt-Winters.

[¶] R automatically calculates the needed parameters. In this case, a NNAR$(4,1,3)_{365}$ model was obtained. It takes as input the last four observations and the last seasonal observation (one season corresponds to three hundred and sixty five days) and has three neurons in the hidden layer. The model consists of twenty networks with this structure (five neurons in the input layer, three in the hidden layer and one in the output layer) and computes the average of the twenty outputs.

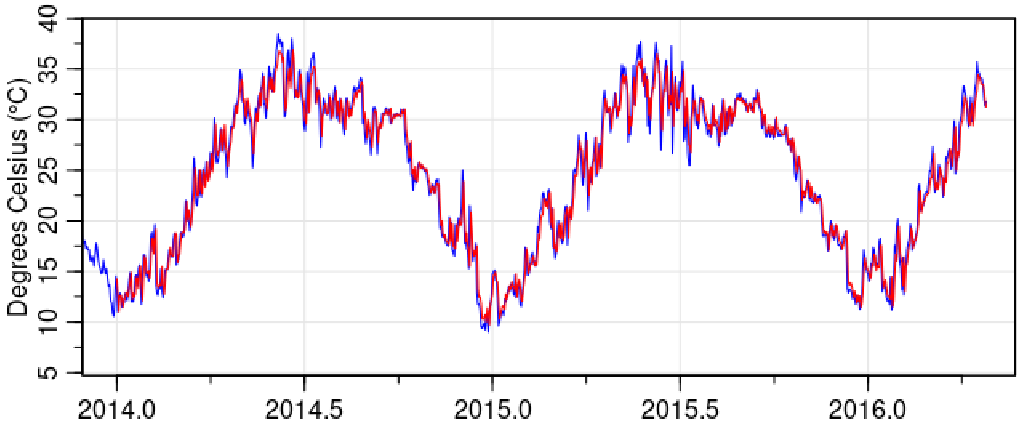


Fig. 19. NNAR model. The blue line represents the original time series and the red line the fitted values.

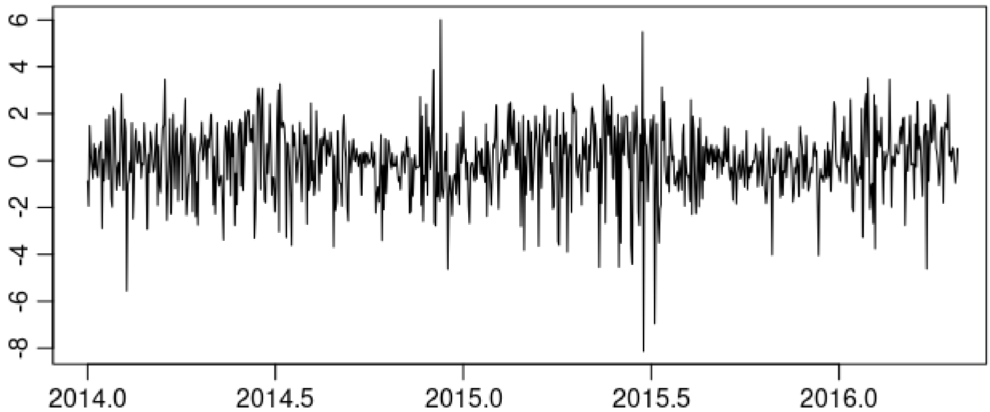


Fig. 20. Plot of the residuals (NNAR).

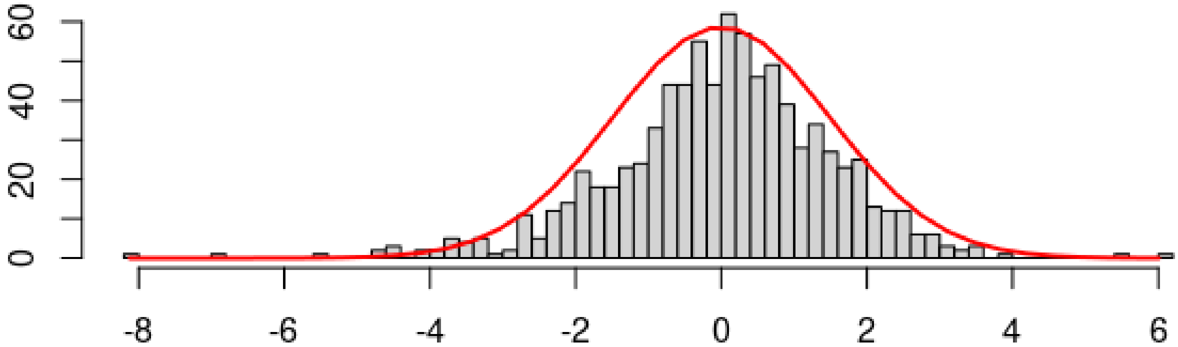
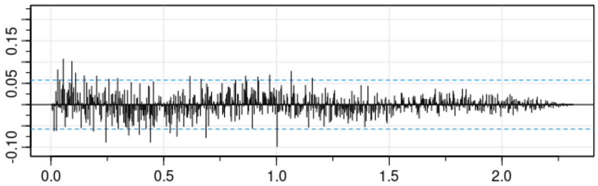
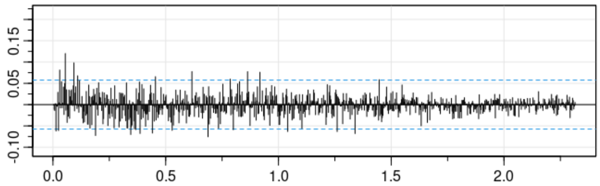


Fig. 21. Histogram of the residuals (NNAR). The red line represents the probability density function of a normal distribution with mean 0.

A further inspection of the residuals was performed by plotting their auto-correlation and partial auto-correlation functions. Figure 22 shows two plots: one for the ACF and other for the PACF of the residuals.



(a) ACF of the residuals.



(b) PACF of the residuals.

Fig. 22. ACF and PACF of the residuals.

Regarding the ACF, some high values (outside the significance threshold) in the first year and a high negative value one year after the first observation (this is because of the series' seasonality) were visible. This indicated that some correlation was not totally explained. However, the ACF values decayed to zero over time, meaning that the residuals became less correlated as the lag increased.

Regarding the PACF, the analysis was quite similar. There were some high values in the first year (some not explained partial correlation), but the PACF decayed to zero over time, which meant the residuals became less correlated as the lag increased, if we neglected the effect of the residuals in between.

After analysing the residuals, forecasting was performed. Figure 23 shows the one-year forecasts computed by the NNAR model (blue line), the real values of the test set (red line) and an 80% and 95% confidence intervals (light blue lines).

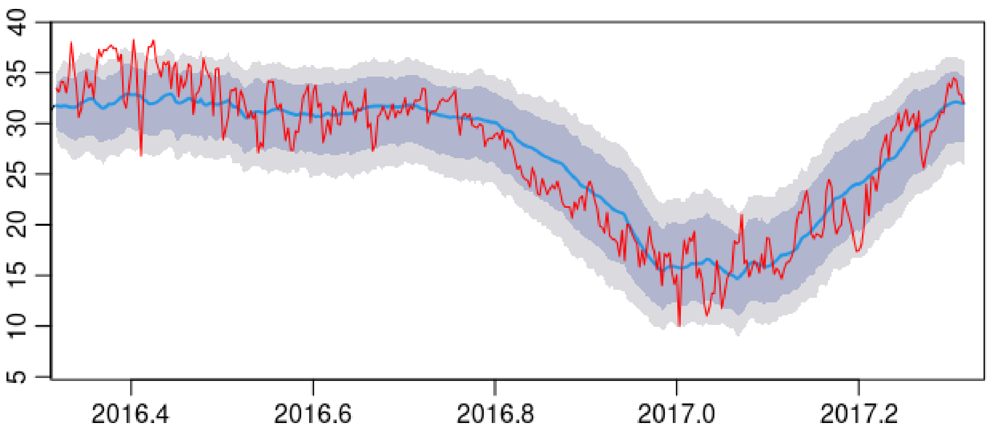


Fig. 23. Forecasting results in the test set (NNAR model). The red line represents the test set values, the blue line the predicted values and the lighter blue lines an 80% and 95% confidence intervals.

Like the TBATS model, the forecasts given by the NNAR model seemed to be smoother than the ones given by both the additive and multiplicative Holt-Winters' models. Nonetheless, they were not as smooth as the ones provided by the TBATS models.

It was also noticeable that the NNAR model seemed to capture the variations in the temperature, hence providing good predictions.

5 Comparing the models

In the previous sections, different models were fitted to the time series, their residuals were analysed and forecasts were made. Most of this process was performed in a visual way, with the support of graphics. In this section, however, a numerical approach was used to assert how well the models performed.

First, an analysis of the performance of the models in the train set was made, so as to assess how well they fitted the time series. Table 1[‖] shows six error measures for each of the four models created, in the train set.

	ME	RMSE	MAE	MPE	MAPE	MASE
HW (additive)	0.0015	1.7785	1.2186	-0.2110	5.4086	0.4779
HW (multiplicative)	-0.0477	1.8101	1.2445	-0.4472	5.4950	0.4881
TBATS	0.0697	1.5285	1.1527	-0.2091	5.1740	0.4525
NNAR	-0.0006	1.4896	1.1217	-0.4814	5.0053	0.4400

Table 1
Values of six error measures (train set).

All the models fitted the series very accurately, since the errors are small. Also, in every column, it is possible to see that the values are close to one another, meaning that the models' performance was similar.

By comparing the values across the different models, it was visible that the Artificial Neural Network model was the one with better results (less error) with respect to all the error measures but one, the MPE. Curiously enough, by focusing on the column regarding the MPE, it was seen that the worst value corresponded to the Artificial Neural Network model.

Hence, it was possible to state that the model which better fitted the series was the one using Artificial Neural Networks. This was very interesting, since the analysis conducted in the previous sections showed that the residuals resulting from the TBATS' model were the least correlated.

Now, that the errors regarding the train set were analysed, an analysis of the errors in the test set was performed, so as to assert which model performed better forecasts. Table 2 shows six error measures with respect to all the four models studied (as in Table 1), but now in the test set.

Regarding the test set, it is noticeable that the errors were not as small nor as close to each other as in the train set (which is normal). Nonetheless, the errors were low.

Despite providing the best results of all the models in the train set, the Artificial Neural Network model was not the best in the test set. The only measure of error in which this model was superior to all the others was the ME.

[‖] ME: mean error; RMSE: root mean squared error; MAE: mean absolute error; MPE: mean percentage error; MAPE: mean absolute percentage error; MASE: mean absolute scaled error

	ME	RMSE	MAE	MPE	MAPE	MASE
HW (additive)	0.4708	2.7300	2.1581	1.2607	8.5428	0.8464
HW (multiplicative)	0.7795	2.9198	2.3214	3.0739	9.1836	0.9105
TBATS	-0.7155	2.3903	1.8545	-3.2245	7.8926	0.7274
NNAR	-0.1591	2.5217	2.0147	-2.3089	8.4046	0.7902

Table 2
Values of six error measures (test set).

Indeed, the TBATS model provided better results than every other model for all the other five error measures, except the MPE. Strangely enough, by focusing on the MPE column, it was visible that the TBATS model was the one with the worse value.

6 Conclusion

Beforehand, it is not possible to know which approach will be better at modelling a given time series nor which one will produce better forecasts. Nonetheless, it is possible to have an idea of which methods to use. During the project, four models were used - Holt-Winters' additive model, Holt-Winters' multiplicative model, TBATS, and Artificial Neural Networks - all of which were very satisfactory at both modelling the time series and making forecasts. The analysis conducted allowed to state that the TBATS model produced better forecasts, whilst the Neural Network model better fitted the series.

References

[1] Hyndman, Rob J., and George Athanasopoulos. Forecasting: Principles and Practice. OTexts, 2018.

[2] De, Alysha M., et al. "Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing." JSTOR, 1 Dec. 2011, www.jstor.org/stable/10.2307/23239555?refreqid=search-gateway

[3] Date, Sachin. "Holt-Winters Exponential Smoothing." Medium, Towards Data Science, 21 Nov. 2020, www.towardsdatascience.com/holt-winters-exponential-smoothing-d703072c0572

[4] Skorupa, Grzegorz. "Forecasting Time Series with Multiple Seasonalities Using TBATS in Python." Medium, Intive Developers, 14 Jan. 2019, www.medium.com/intive-developers/forecasting-time-series-with-multiple-seasonalities-using-tbats-in-python-398a00ac0e8a

[5] Burba, Davide. "An Overview of Time Series Forecasting Models." Medium, Towards Data Science, 24 Nov. 2020, www.towardsdatascience.com/an-overview-of-time-series-forecasting-models-a2fa7a358fcb