Real-TimeSeries-Markdown

Danial

3/8/2021

Task 1: Getting the Data

All data for this assessment were obtained by downloading the Temperature data for 10 districts which each districts having 3 parameters (Tmin,Tmax, Tmean). The chosen districts is as below:

- Northern_Ireland
- Scotland_N
- Scotland_E
- Scotland_W
- England_E_and_NE
- England_NW_and_N_Wales
- Midlands
- East_Anglia
- England_SW_and_S_Wales
- England_SE_and_Central_S

This will results in having 30 Time-series and to ease the amount of pages for this report, only the plot/results for Northern-Ireland Time Series would be presented unless stated otherwise. However, The code will include all time-series using lapply() and Map() functions.

Task 2: R-programming

```
##      Tmin Tmax Tmean
## [1,] -7.4 -0.6  -3.5
## [2,] 14.7 26.7  20.4
```

Task 3:EDA (Exploratory Data Analysis)

Using Tmean time-series, the coldest/warmest region could be obtained by averaging each districts time-series. The Widest/lowest range could also be obtained and the results is as below:

- Coldest: Scotland_N with an average mean temperature of 6.84 degree celcius
- Warmest: England_SE_and_Central_S with an average mean temperature of 9.76 degree celcius
- Widest range: East_Anglia with a range of 22.9 degree celcius
- lowest range: Scotland_N with a range of 17.4 degree celcius

Figure 1 shows the barplot for each districts range.

With the given Four Seasons data, a plot for summer and winter could be obtained to see the trend over the year and it could be concluded that both seasons increases in temperature as seen in Figure 2 and Figure 3. This could be linked with the emergence of global warming

Figure 1: Widest range in ascending order

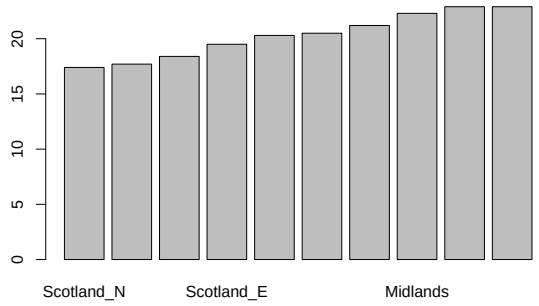


Figure 2: Northern Ireland:Winter

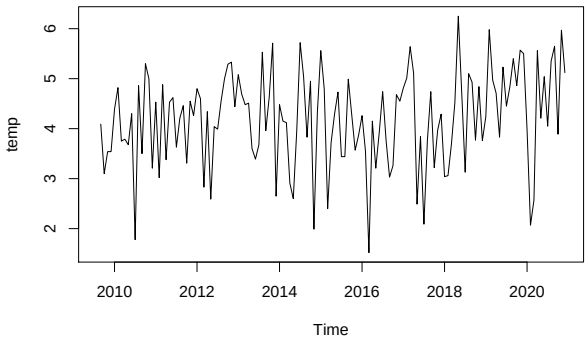
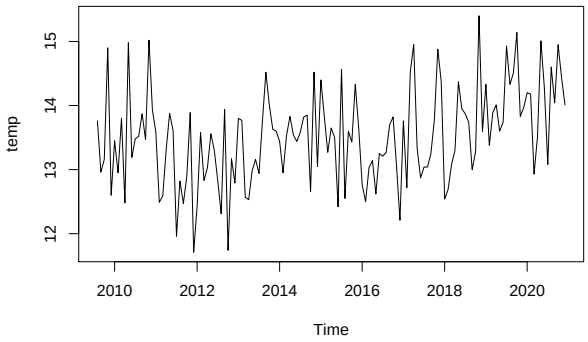


Figure 3: Northern Ireland:Summer



Task 4: Trend and Seasonality

In this task, the time series was subsetted until the year 2019 and would undergo prediction using Regression lines, Seasonality and both combined. The chosen model would be chosed based on the Akaike information criterion (AIC) where it would gives the number of prediction error and thus lower value means better model.

Trend: Linear,Quadratic and Cubic Models

The differences between the 3 models can be seen on Figure 4. By using the AIC() function, the linear model was chosen as it gives the lowest value for all districts and all parameters. The pmin() function was used to obtained the lowest value. In comparison, The AIC values of each model for the Northern_Ireland Time-Series at Tmin is :

- Linear model AIC: 8750.06
- Quadratic model AIC: 8751.68
- Cubic model AIC: 8754.24

Thus, the trend prediction model used the linear model with the predict() function and can be seen on Figure 5 with the red line being the prediction.

Figure 4: Regression Lines

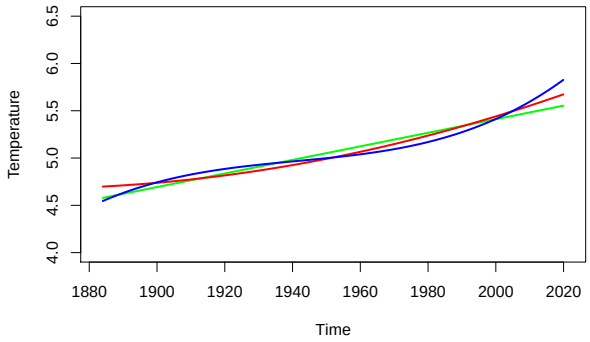
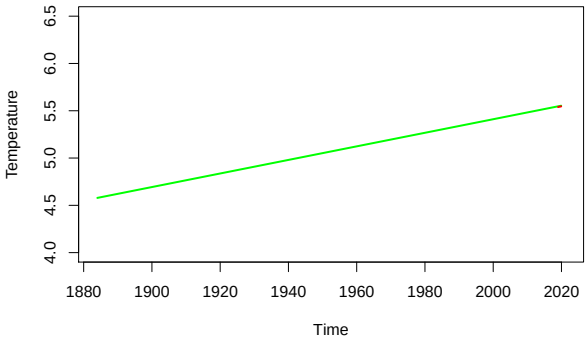


Figure 5:Linear Prediction



Seasonality: Average Model and Harmonic Model

A time-series would have a characteristic of seasonality if there is a pattern that exist over a one year period. Thus allowing seasonal to act like a trend but differs from regression lines. Prior to analyzing the seasonal trend, the linear-trend was removed to ensure the integrity of the seasonality in each time series.

The seasonal means/Average model would be applied first and the result can be seen on figure 6 with the black lines being the no-trend time-series while the green line would be the seasonal data using an Average model. Then, using the seasonal means, a prediction was made and can be seen as the red line on figure 6.

The harmonic model would then be applied by integrating trigonometric functions onto the time-series. Each data would then be tested until the fourth order of the harmonic model and the coefficients would then be analyzed to determine an appropriate model. Majority of the data in the fourth and the third model does not fall under the significance level and thus rejected from being an appropriate model. As an example using the Northern Ireland with the parameters Tmin, Image 1 shows the p-value for order 4:

	Estimate	Std. Error	t value	Pr(>|t|)
SIN.1	-1.19924034	0.04043705	-29.6569706	7.823210e-155
SIN.2	0.32006361	0.04043705	7.9151083	4.536970e-15
SIN.3	0.04618587	0.04043705	1.1421672	2.535529e-01
SIN.4	-0.07755479	0.04043705	-1.9179142	5.529709e-02
COS.1	-4.50652389	0.04043705	-111.4454230	0.000000e+00
COS.2	0.58442117	0.04043705	14.4526170	1.251769e-44
COS.3	-0.01092197	0.04043705	-0.2700982	7.871191e-01
COS.4	0.05084273	0.04043705	1.2573305	2.088148e-01

Figure 1: Image 1: Summary Report on Harmonic Order = 4.

Image 1 shows that the order 4 and 3 is not appropriate and thus rejected. In choosing the appropriate model, the AIC function was used on Harmonic (1,2) and seasonal means and the lowest values would be on harmonic model with order 2 for all time series except for 3 districts in Tmax. The 3 districts would be Scotland_N, Scotland_W and England_NW. Despite that, the chosen model would still be Harmonic model with order 2

Figure 6: Northern Ireland: Seasonal Means Model

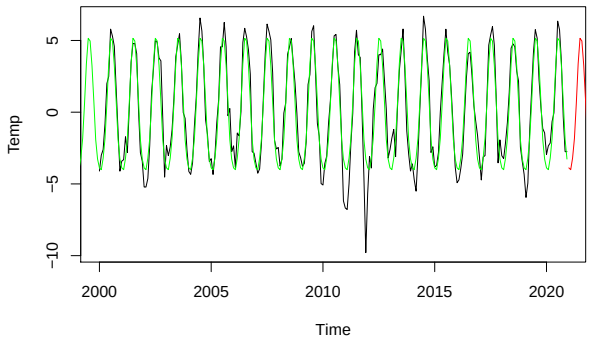
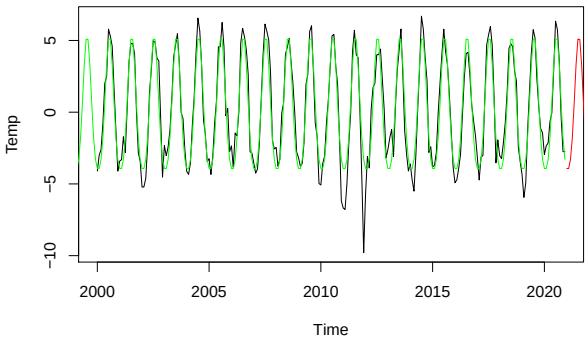


Figure 7: Northern Ireland: Harmonic model with order 2



Combined model (Linear and Harmonic Model (order=2))

Since the appropriate models for regression line and seasonality were determined, A list called final would be created by applying the combined model(Linear model and Harmonic model:order=2) on the time-series (1884-2019). On top of that, another list called test would also apply a quadratic model and harmonic model:order=2 on the time-series(1884-2019). Figure 8 and 9 would present the trend model and prediction for final and test datasets respectively.

Figure 8: Northern Ireland: Final Prediction

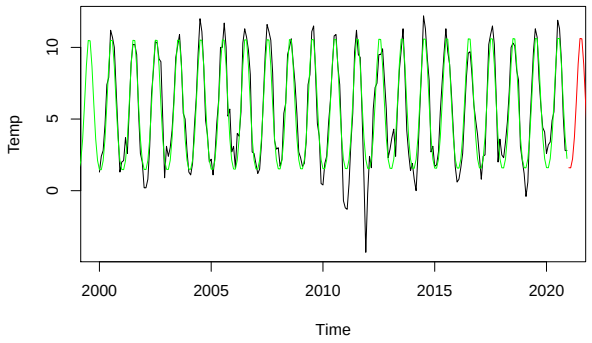
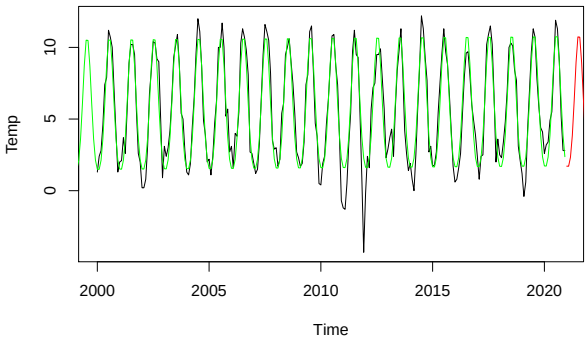


Figure 9: Northern Ireland: Test Prediction



Task 5:ARMA and Forecasting

Stationarity

In order to deploy an ARMA model, the time-series need to be stationary and this can be done by removing all the trends and seasonality obtained from the previous task. Once the trends are removed, an Augmented Dickey-Fuller test will be applied on the data in order to test the stationarity of the time-series. All time-series in both final and test model are stationary with having a p-value of less than 0.05 and reject the null hypothesis. Image 2 shows an ADF test done on the final-data on districts Northern Ireland with parameter Tmin.

```
        Augmented Dickey-Fuller Test

data:  dots[[1L]][[1L]]
Dickey-Fuller = -9.5604, Lag order = 11, p-value = 0.01
alternative hypothesis: stationary
```

##

ARMA Model ARMA model is a prediction model that uses autoregression and moving average prediction technique. **Autoregressive model** or AR uses previous values of the time-series value/dependant values to make prediction while **Moving Average model** or MA uses the series mean and previous error to make predictions. However, AR and MA model required appropriate order to create a prediction model with the less error and this could be done by looking at the ACF and PACF plots. The ACF plots will determine the MA order and the PACF plot will determine the AR model. Image 3 shows the ACF plot where it is significant at lag 4 before going down and Image 4 shows PACF plot where it is only significant at lag 1. This would propose that the appropriate model would be ARMA(1,4). However, the arima() function would have a built in AIC values and thus propose a more efficient approach in determining the ARMA model order.

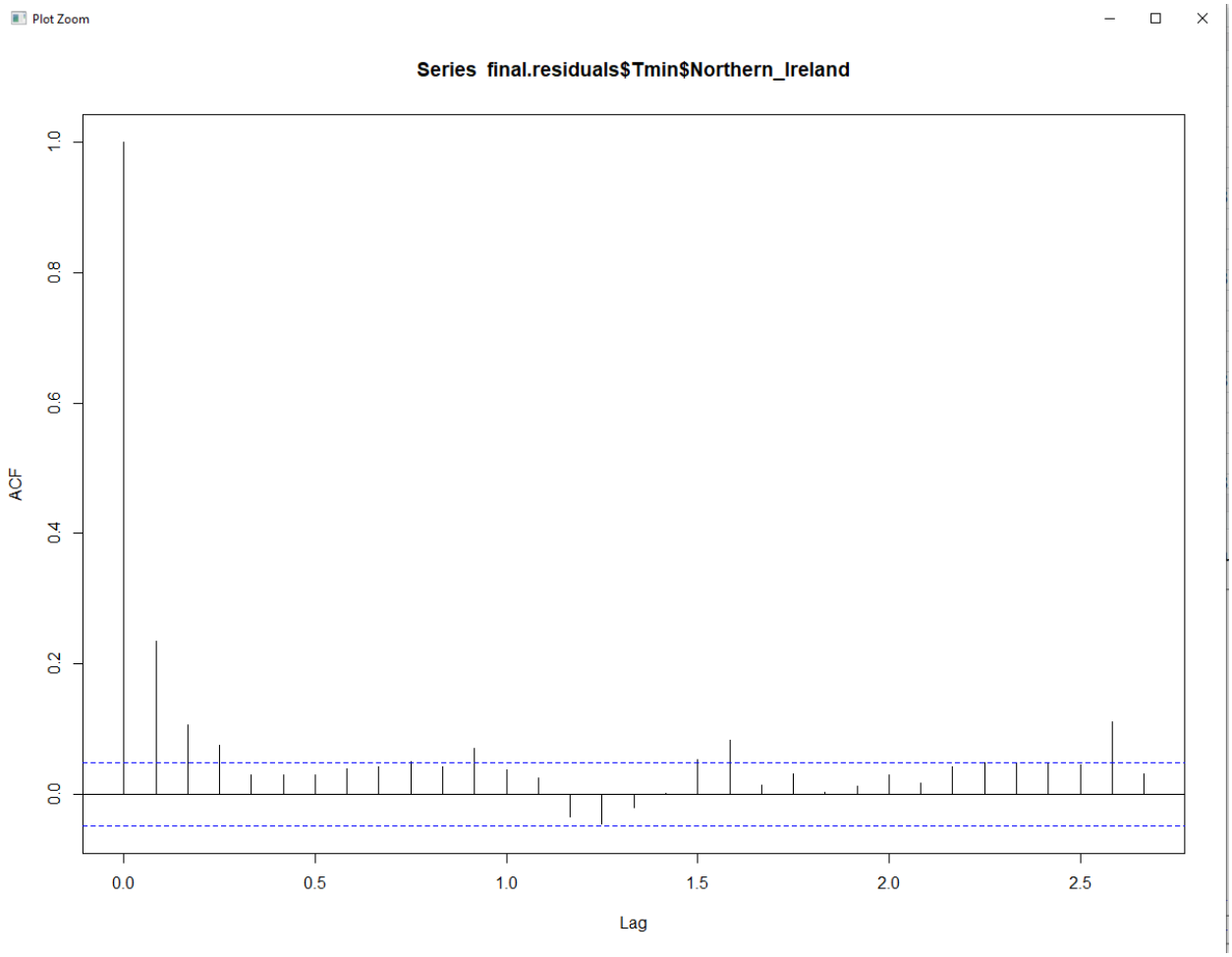


Figure 2: Image 3: ACF

With the chosen ARMA model, the prediction could be done and Figure 10 and 11 shows the prediction done on top of the actual value for the 2020 Temperature for both final and test time-series. Both Figure shows promising prediction but further forecast accuracy test would need to be done.

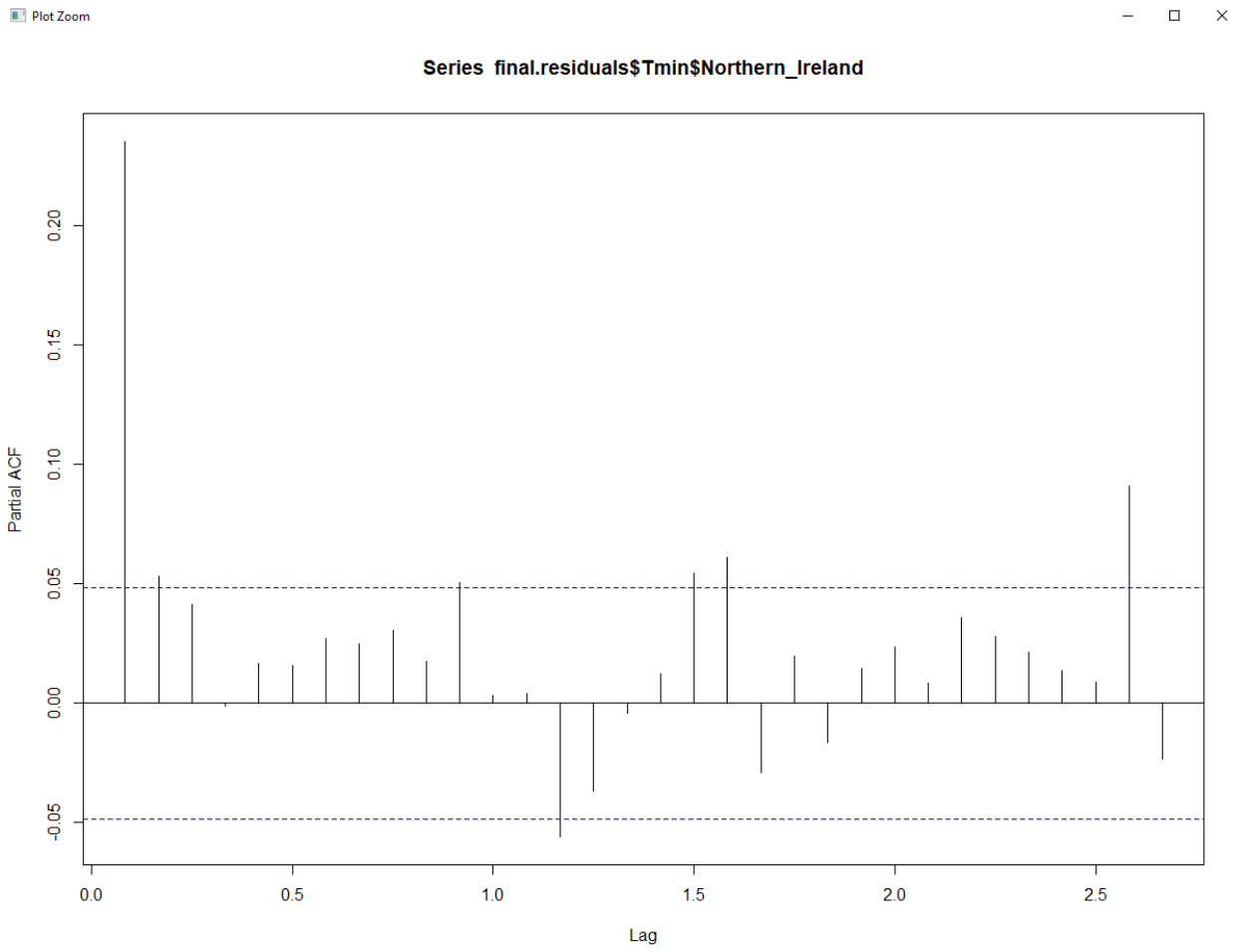


Figure 3: Image 4: PACF

Forecast Accuracy

Since all 30 time-series uses the same scale for the temperature data, A Scale-dependent errors measurement was used and the method used is the Mean Absolute error which is by averaging both the actual value means and the prediction value means and subtract both values.The value would then be compared and the threshold for a significant error would be determined before the accuracy test.

The significant error would be >=1

the result for all errors does not surpass 1 or = 1 and thus propose that the ARMA prediction model produces an accurate results. For example, using the Northen_Ireland accuracy results, the ARMA model produces an error of 0.11 for the final model and 0.06 for the test model

Figuure 10: Final:Tmin:Northern_Ireland prediction

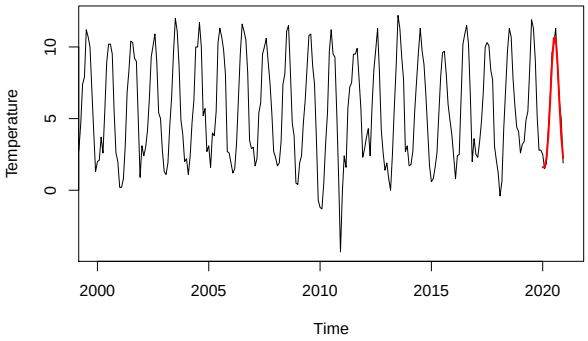
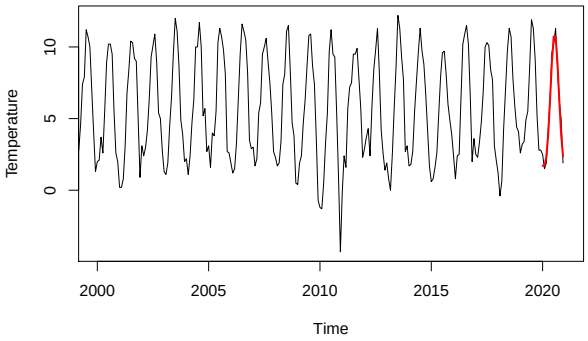


Figure 11: Test:Tmin:Northern_Ireland prediction



Conclusion

As conclusion, all 30 time-series were predicted using Regression Lines, Seasonality and ARMA model. All model would be integral in determining the pattern of the data. The Regression lines shows an increasing trend of temperature. The seasonality trend shows that there is a repeated patterns throughout the year and in this case, there is a four season each year in every region in UK. The ARMA model would create an accurate prediction under the appropriate AR and MA order.