UK Weather Data Analysis

Kieran Billingham

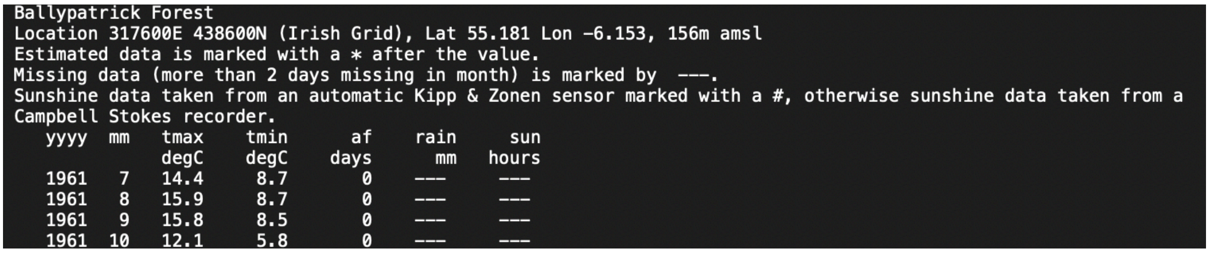
12 May 2019

Overview

This R script and summary paper aims for perform time series analysis on UK weather data. The data will be explored for information and the time series built in a cumulative manner. This will provide a good understanding of trends and seasonality for the forecasting of future weather. All graphical outputs are saved in the working directory of the R script, whilst some plots of interest will be included in this summary.

Reading and Formatting

Historic UK weather data is available from the Met Office. 33 historic station files were provided in .txt format. Within these files was the station name and meta data along with some basic weather information. The data is formatted as in Figure 1.



```
Ballypatrick Forest
Location 317600E 438600N (Irish Grid), Lat 55.181 Lon -6.153, 156m amsl
Estimated data is marked with a * after the value.
Missing data (more than 2 days missing in month) is marked by  ---.
Sunshine data taken from an automatic Kipp & Zonen sensor marked with a #, otherwise sunshine data taken from a
Campbell Stokes recorder.
   yyyy  mm   tmax    tmin      af    rain     sun
              degC    degC    days      mm   hours
   1961   7   14.4     8.7       0    ---     ---
   1961   8   15.9     8.7       0    ---     ---
   1961   9   15.8     8.5       0    ---     ---
   1961  10   12.1     5.8       0    ---     ---
```

Figure 1, the raw data format for Ballypatrick Forest station

For these tasks only the ***yyyy, mm, tmax, tmin*** fields are required but all the data is read and formatted so that further analysis could be performed in the future.

Each station is loaded and saved as a data frame within a list named **data**. Keeping the data from each station separate is good practice and prevents any leakage, whilst storing the files within a list allows them to be actioned on within a single process rather than individually (line 54). This will come in useful later when applying the same functions to all stations. All meta data characters from the 7 variables are removed at this stage, while NA's are recoded to NaN's.

Stationed have opened and closed at different times but in order to directly compare the time series for temperature, all stations needed the same date range. Stations in Ballypatrick, Dunstaffnage, Lowestoft, Manston and Nairn were closed prior to the end of 2018 and so were removed from the data list. The first complete year that all stations were open was 1979, and so this was selected as the first date. This left 480 observations for 27 open stations.

Exploratory Analysis

To explore the temperature at stations, maximum temperature and minimum temperature were extracted from the data and stored as time series objects. Summary information, as well as max, min, ranges and means could then be calculated. They are stored in the **Range** and **Average** lists.

Trend and Seasonality

Trends

A time series object can be seen as a signal composed of an additive combination of simpler signals. The trend of the series is a signal that is not cyclic and therefore can be described by a parametric function. In this section, the trend of Max and Min temperatures is estimated using a linear regression, a global polynomial, a local polynomial and a generalized additive model. The moving average is included as a base line of the general fit, but it does include a seasonality error.

For each station a plot as in Figure 2 is saved in the **images/trend** subdirectory. It is clear here that the polynomial functions do not perform well, they overfit the data which makes then unsuitable for forecasting. Linear regression is a good fit for this data, but it is known that this isn't a very accurate method. A spliced method is much 'smarter' than a GLM, whilst being resistant to over fitting.

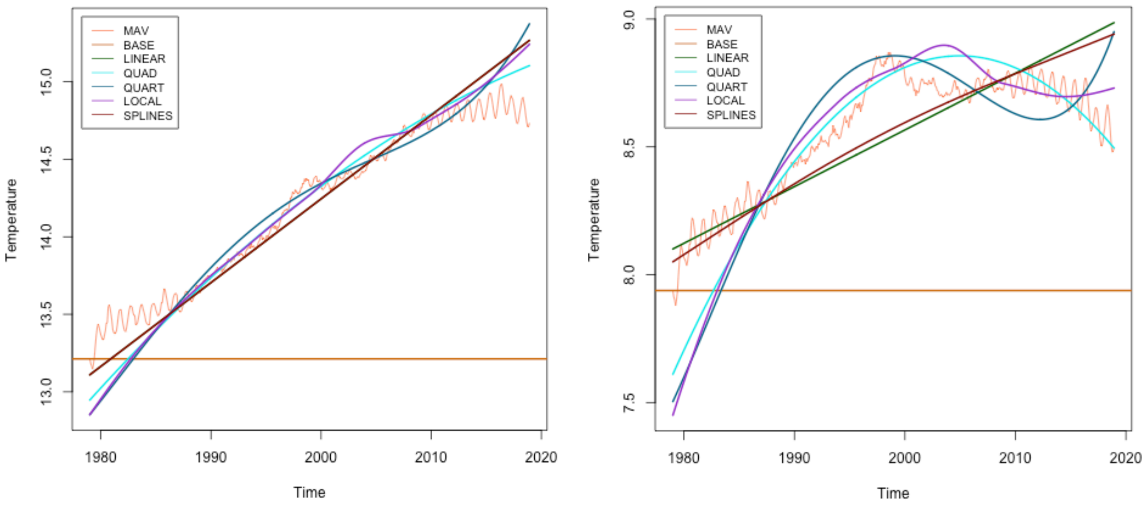


Figure 2, Temperature trends for Eastbourne maximum and minimum temperatures.

Each station has its own trend for maximum and minimum temperature. This is shown in Figure 4 using the Spliced trends. The range in temperatures is due to the geographic spread of the stations and this is good evidence for continuing to tread each station as a completely isolated series.

One time series has a GAM trend that is noticeably non-linear. Figure 3, shows Braemar's minimum temperature trend. The large spike in minimum temperature here has been considered by the GAM but ignored by the GLM.

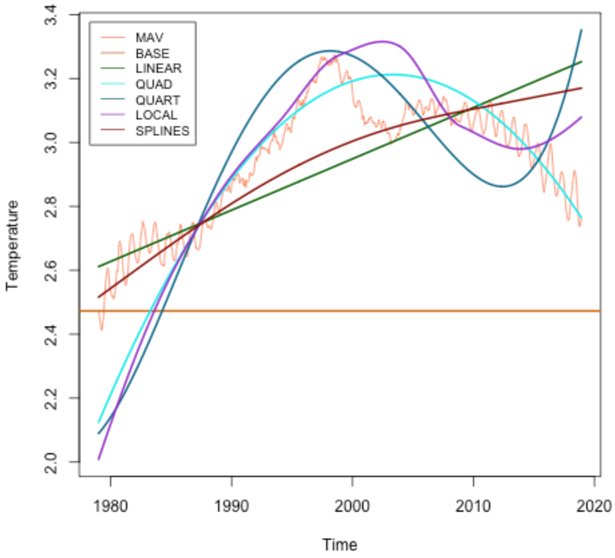


Figure 3, Braemar minimum temperature trends

Figure 4 does confirm that no station is seeing a negative trend. The national average temperature is certainly rising. It is also interesting to note that the minimum and maximum trends are not overlapping, indicating that the warming is more general than dependent on more sunny days.

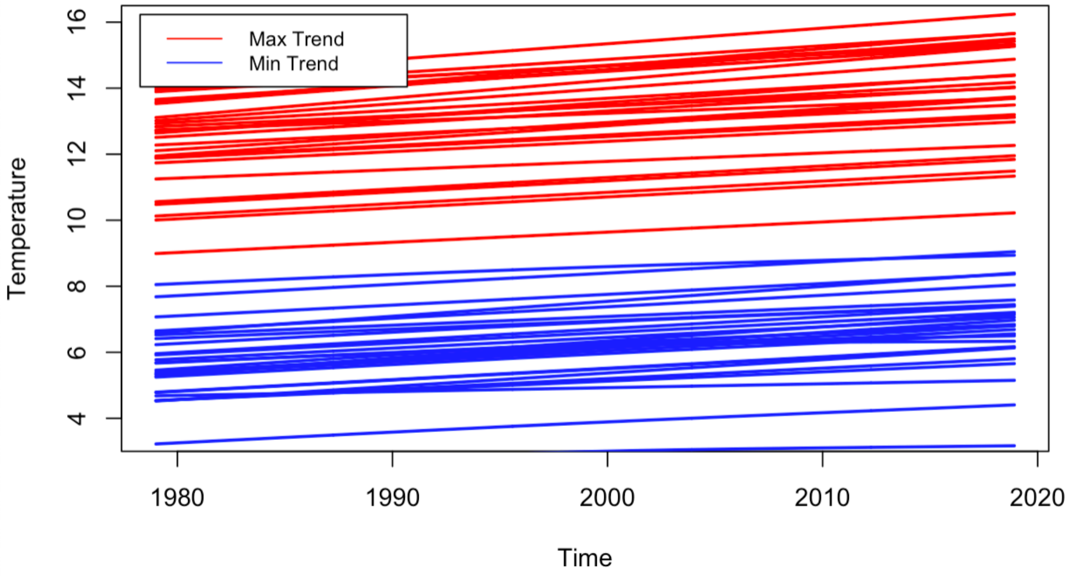


Figure 4, Maximum and minimum spliced temperature trend for all stations.

The temperature difference between max and min is not increasing nationally though. Some stations are seeing an increase (those red in Figure 5), whilst some are seeing a decrease (green in Figure 5).

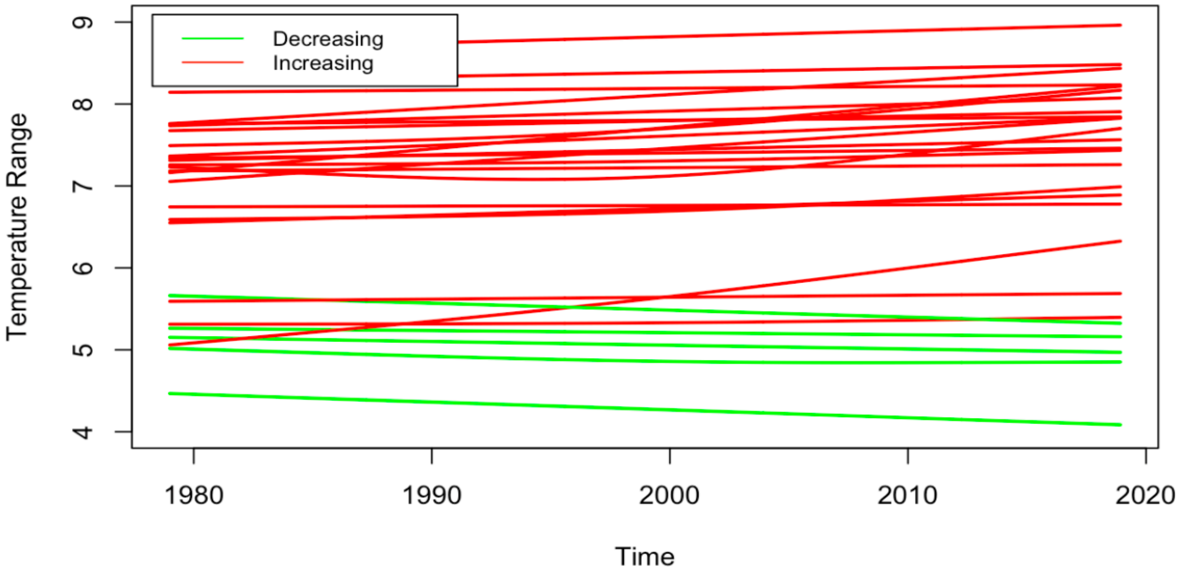


Figure 5, The temperature range for each stations trend fit.

Seasonality

The seasonality of the time series is the harmonic temperature change due to seasonal shift. It can be identified by a number of methods, and both the ANOVA and harmonic method were tested. It was decided that the harmonic method captured more of the peaks in temperature and so was used for the rest of the task. Otherwise the two methods perform very similar, as can be seen in Figure 6. The estimates for each month are also very similar and can be found in the object list as **Estimates.**

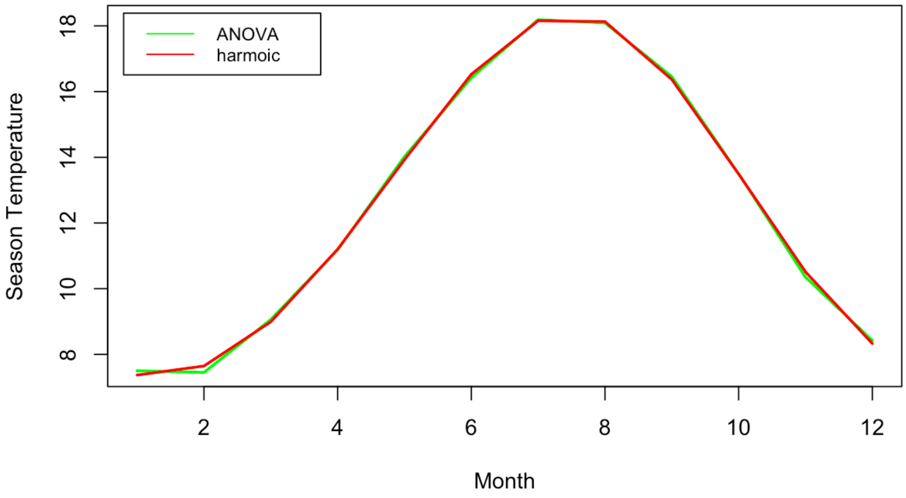


Figure 6, ANOVA and harmonic seasonality fits comparison.

It is important to add the trend of the data into the seasonality estimate; this should capture all the non-stationary components of the time series. The GAM from the past section was used for the Max and Min temperature estimates.

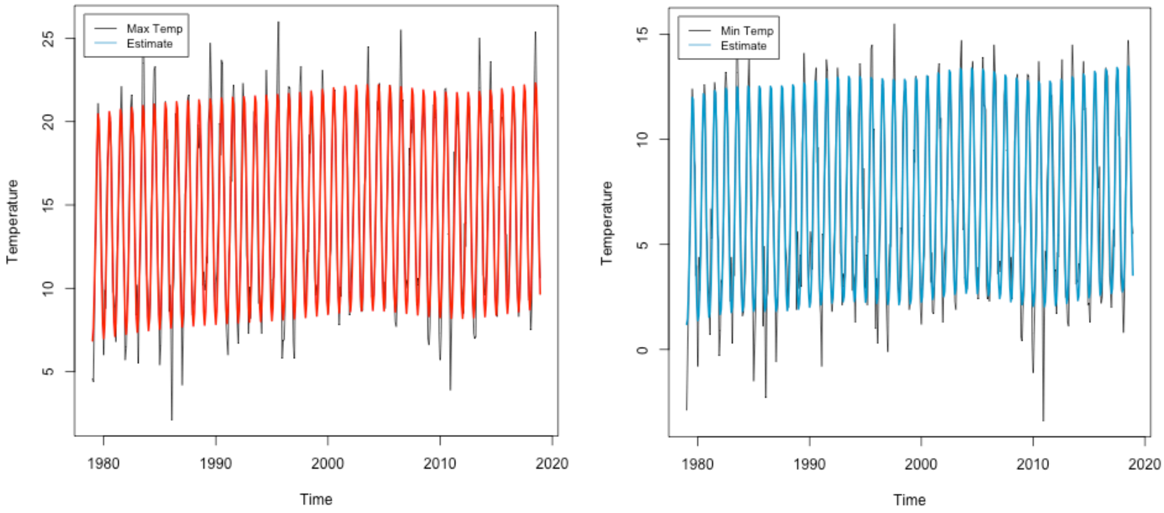


Figure 7, Seasonality estimates with GAM trend for Cardiff Station.

Figure 7, an example of the seasonality plots produced for each station max and min. The plots for the other stations can be found in the subdirectory **Images/Seasonality.** It can be seen here that estimate does a good job of capturing the overall trend but cannot account for the larger peaks and troughs.

Forecasting

Maximum and minimum temperatures for all stations is averaged to create a UK average temperature time series for max and min. In order to fit an ARMA forecasting model to the data, it needs to be stationary. Differencing the trend for max temperature once gives Figure 8.

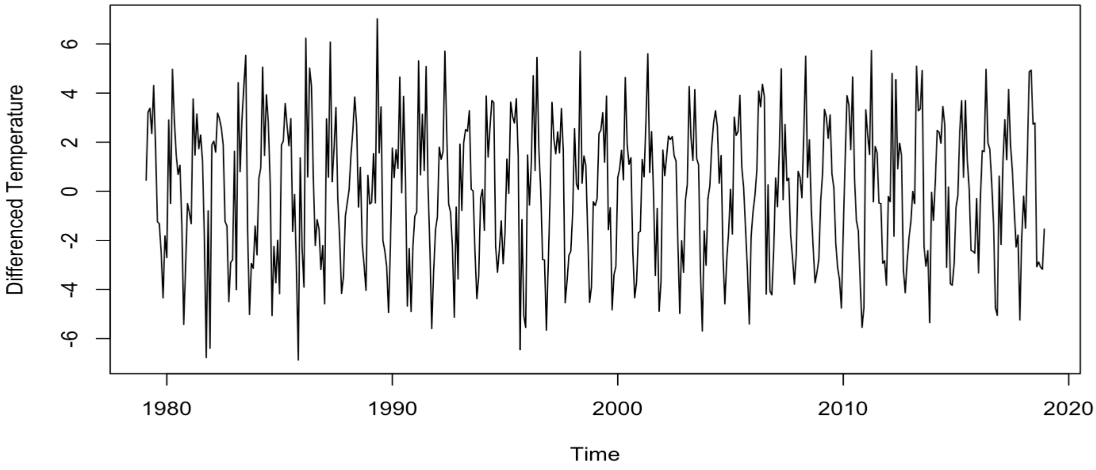


Figure 8, The differenced time series for max UK temperature.

This has just suppressed the trend and seasonality, to get the stationary residuals. A decomposition plot shows the additive nature of time series data (Figure 9), if the trend and seasonality can be modelled separately then an ARMA can be fit to the residuals.

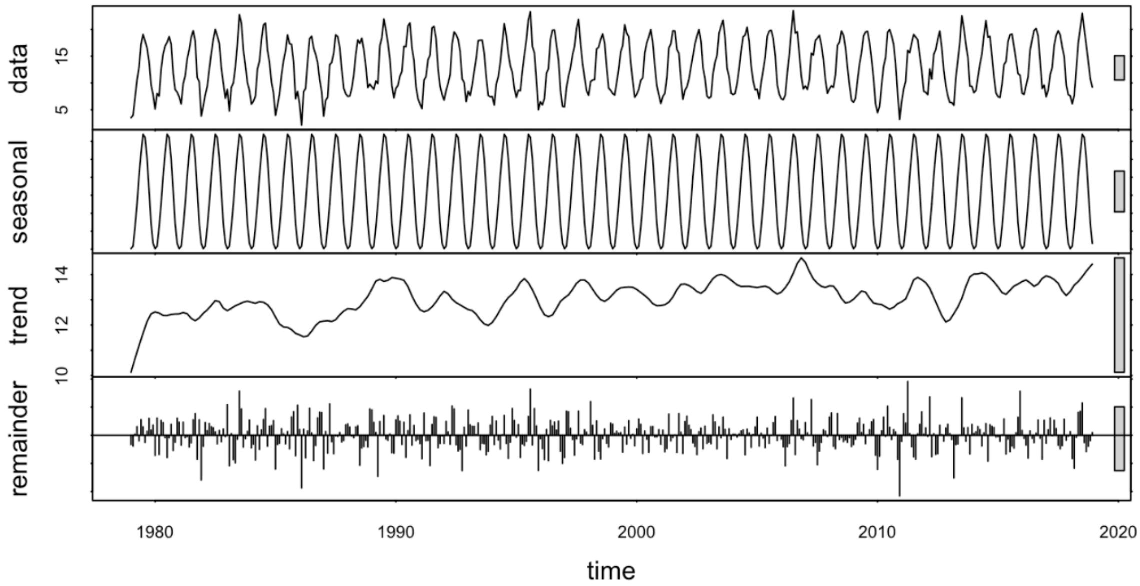


Figure 9, A decomposition diagram for the average maximum UK temperature

To remove the trend and seasonality, a linear model is estimated. The model has the shape as in Figure 10. The blue marked line represents the predicted trend for 2019.

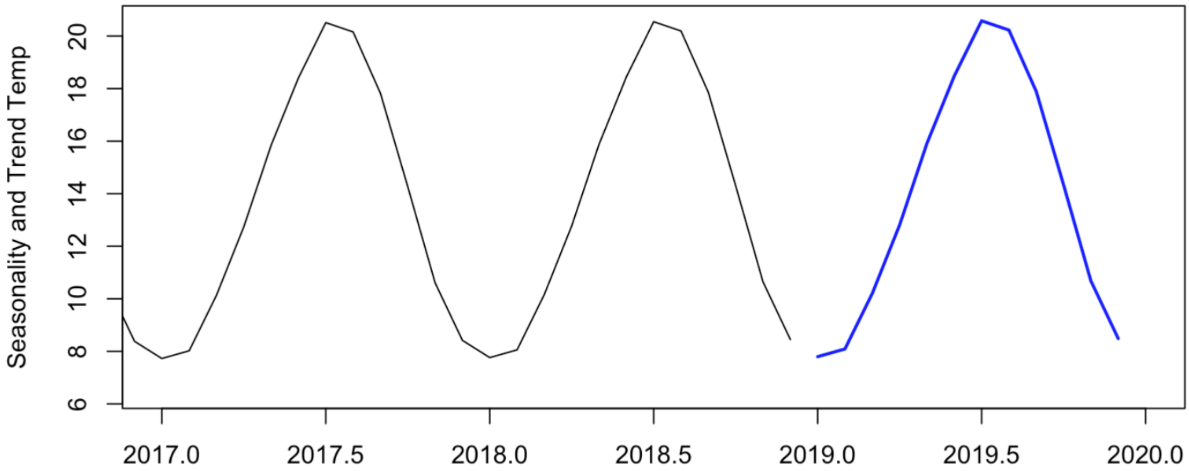


Figure 10, The linear prediction of the trend and seasonality

Using the auto.arima functions with differencing set to 0, an ARMA model can be applied to the residuals. The resulting model is shown in Figure 11, the prediction values are quite small, showing that average fluctuations in temperature are generally small, and unpredictable. Most of the change in temperature comes from the trend and seasonality of the system.

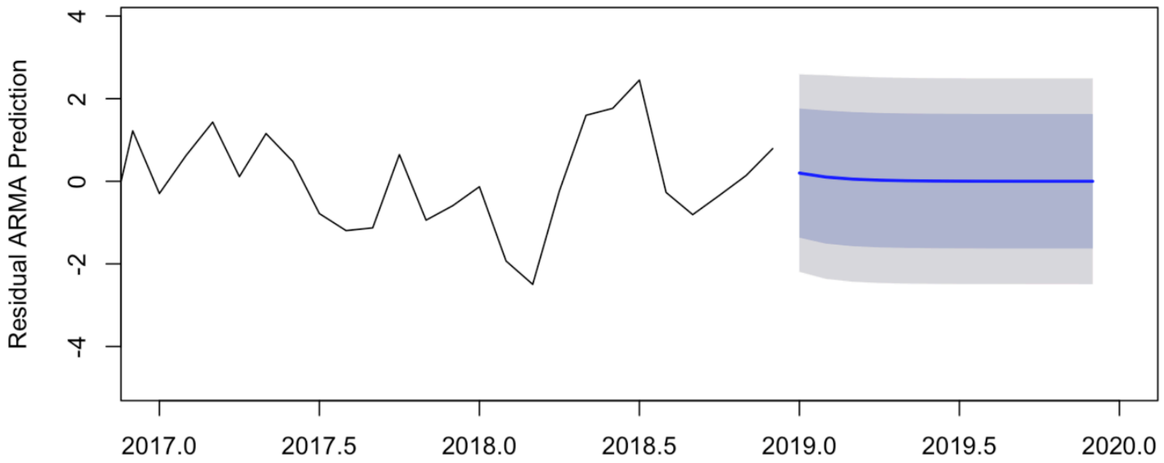


Figure 11, ARMA prediction for Maximum temperatures. Dark blue = predicted value, light blue area is the confidence band

To get complete predictions for the maximum and minimum temperatures for 2019, the ARMA and linear predictions were added. These are shown in Figure 12.

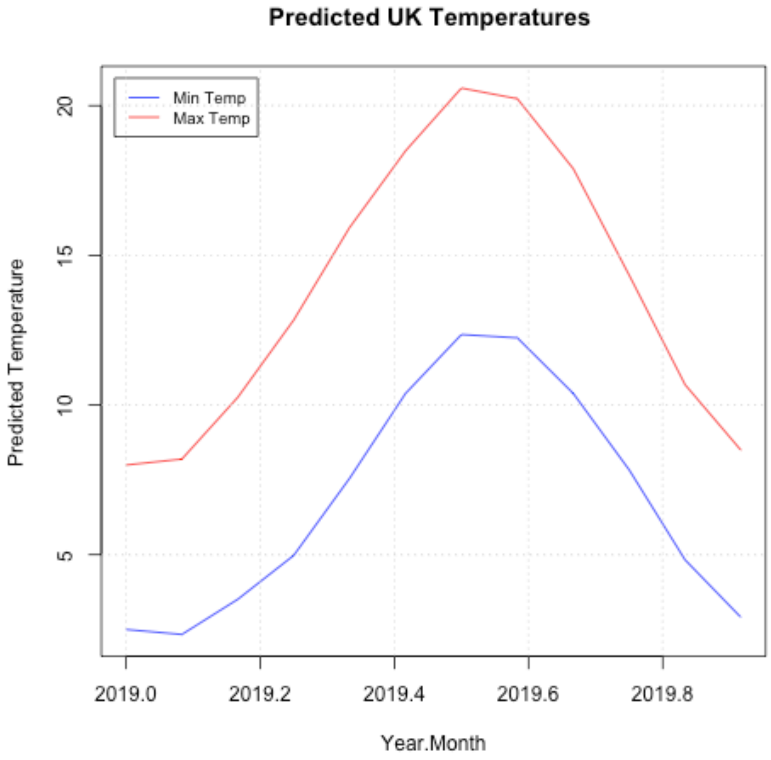


Figure 12, The predicted temperatures for max and min temp for 2019 using ARMA and linear

It can be seen that a high maximum temperature should not be expected in 2019 but an autumn with warmer evenings can be expected.