1 Part A - Analysis of time series

1.1 Description of the original Time Series

We wanted to work on climatic data for this part of the project, because we knew that it is a classic example of time series. We also think that it is a relevant one, since climate change is a really important topic and thus the analyze of the climate is more important than ever.

We have chosen to work on Stockholm Historical Weather Observations, and more especially on the dataset "Homogenized monthly mean temperatures" which contains the monthly mean temperatures in Stockholm in the last century (Source : https://bolin.su.se/data/stockholm). Here is what the original Time series looks like.

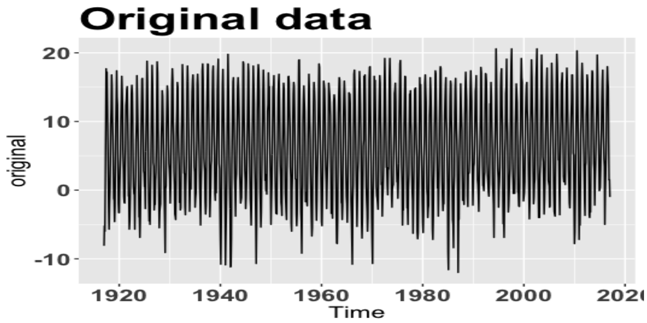
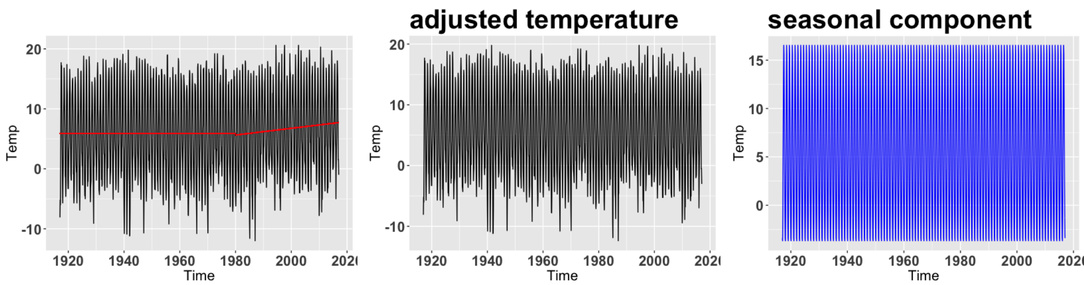


Figure 1: Original Time Series

As expected, there is a significant seasonal component in this plot. Thus, it is not a realization of a stationary time series, and we will need to apply some transformation to it.

1.2 Separation of the data in several components

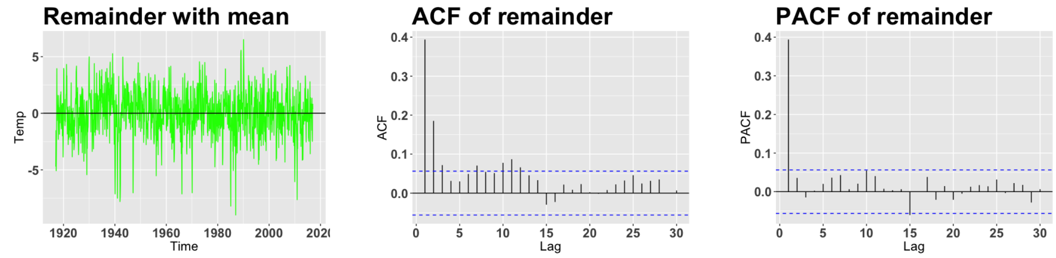
We used R packages("forecast") in order to separate the original time series in three different components : a seasonal, a trend and a remainder one.



[1] trend [2] adjusted data [3] seasonal

Figure 2: Decomposition

First of all, we tried to generate a trend component. As you can see, until 1980 it seems no trend but after this time the temperature goes bit higher and higher. It appears to follow a linear trend so that we separated the data at 1980 and applied linear regression to the latter data. As a result, the average temperature until 1980 is 5.89 and the straight line starts at 5.62 at 1980. Since our aim was not to obtain the exact breaking point, we conseidered that it was a satisfying assumption. Secondly, after eliminating the trend component from the original data we extracted the seasonal component by taking an average temperature in 12 months, since a temperature shall have a period of 12 months. The remained value is what we focused on afterward.



[1] remainder [2] ACF [3] PACF

Figure 3: Decomposition

To see if the remainder is stationary, we computed the ACF. It is reasonable to conclude that a sequence of a remainder Y_t and some lagged remainder is independent of other sequence from the auto-correlation function for the observed data as above.

The difficulties that we encountered in this part were mainly about extracting the trend. When we just look at the temperature data from 1917 up to 2017, a trend does not look smooth. With the initial parameters, it was impossible to distinguish any trend. We had to investigate on the trend estimator's parameters to get a decent understanding of the trend, and then model it.

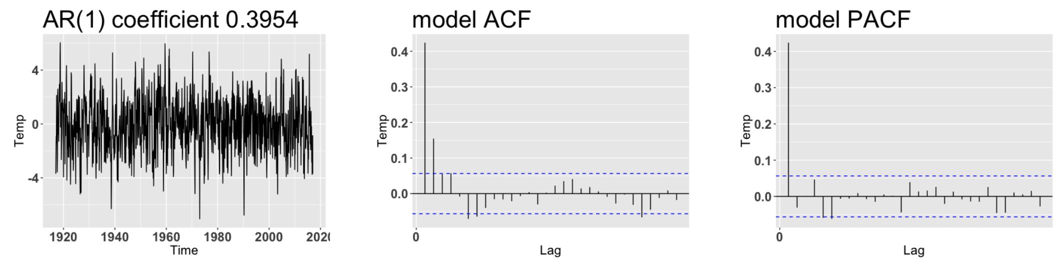
1.3 Modelling of the remainder

1.3.1 Selection of the model

Since Y_t has an auto-correlation of slight lagged other values so that it should be at least AR(p). Then, we assume that it can be written either $AR(p)$ or $ARMA(p,q)$, based on the results we saw in class.

1.3.2 Model fitting

We used a R package({forecast}(auto.arima) to fit an ARIMA model to our remainder. After running the function, we got that the best model is AR(1) with the coefficient 0.3954.



[1] remainder [2] model ACF [3] model PACF

Figure 4: model

1.3.3 Validation of the model

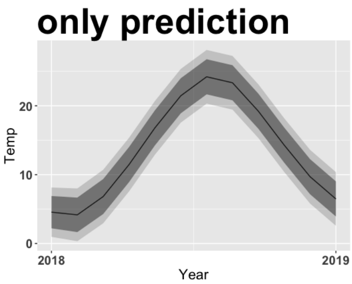
Once we have a model, one needs to validate its goodness of fit. We chose to plot a realization of our model and to compute Akaike Information Criterion (AIC) with the ”true” time series.

We got AIC=4896.51 with the AR(1)model for a 1200 length dataset. Still it is hard to interpret the criteron even if we had another AIC for another model. Instead, we generated ACF and PACF for model and compare them to see if it is reasonable. Neglecting the values inside the confident interval, both of ACF and PACF are quite similar and we can conclude at least the model we have is acceptable.

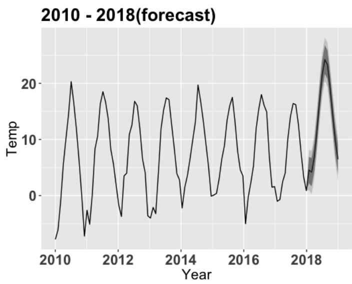
This model can be used to predict the evolution of the mean temperature in the near future. For example, we used it to predict data for the next year.

1.4 Forecasting

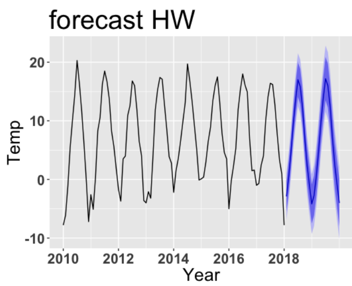
Since the forecast of a remainder composition quickly converges to 0, the trend and seasonal factor are dominating in the whole forecasting model as below. This might be a disadvantage in our case. Indeed, we are predicting small values using our AR model, and thus the result is almost negligible compare to the other components. It leads to a prediction which is standing out unnaturally. Therefore, we are a bit skeptical about the result.
A method called ”Holt-Winter method” we found may gives us more meaningful forecasting. However, the comprehension of this method is currently beyond our scope.



[1] forecast 2018



[2] merge forecast and original data



[2] Holt-Winter forecast

Figure 5: Forecast with confident interval

1.5 Conclusion for this Time series

This Time Series was really interesting to study. It is compose with a significant seasonal component, a trend which can be decompose into two main linear ones, and a stationary remainder. This remainder can be model using several techniques, among those saw in class or others. We got decent enough results using an AR(1) model, but surely a lot of investigations can still be added.