Group Project

Manqiu Liu

(916767589)

Wenxin Ma

(916767550)

1 | INTRODUCTION

Based on the data of annual temperature anomalies (1850–2018) for the northern hemisphere, we are going to analyse the time series data of northern hemisphere annual temperature from 1850 to 2018.We use a loess model to fit the trend and also analyse the rough part with ARIMA model based on AICC selection method. For the hidden periodogram part, spectral density analysis are used to help in diagnosing the residuals of the final model.

2 | DATA ANALYSIS AND MODEL SELECTION

To begin with, from the time series plot of the raw data we can see that the temperature throughout the years changes a lot. There seems to be a trend in the data and the variation is not consistent at first sight. Since the mean of temperature is not constant, the series is not stationary. We can also find that in the first 25 years, the fluctuation of the data is larger than that in the following years, although we made an attempt to transform the data but get λ=0.94, which is too big, so there is no need to transform it.

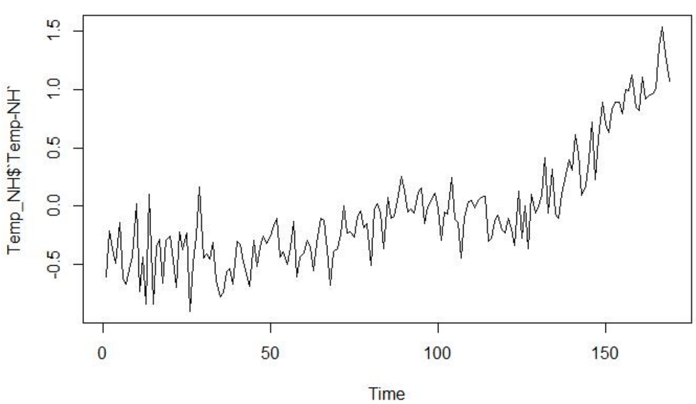


Figure 1 Time series plot of raw data

We first use the loess method to detrend, and the plots below is the loess estimation and the rough part left.The following analysis will focus on the rough part.

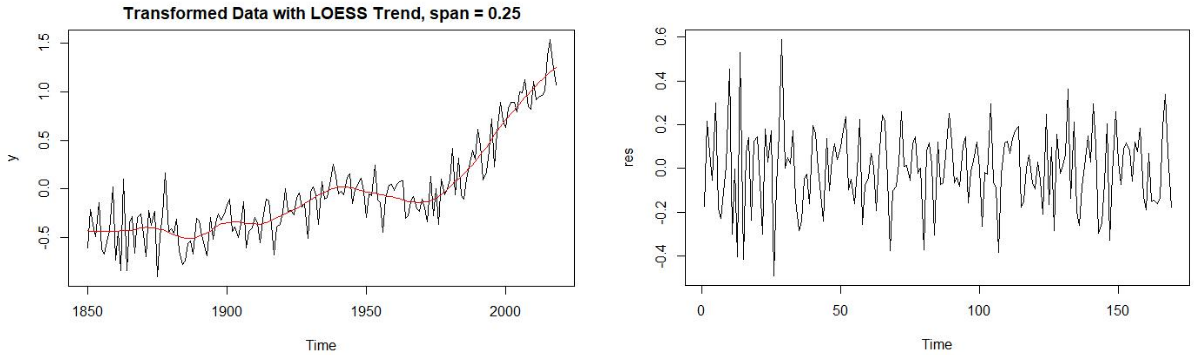


Figure 2 Loess trend and rough part

On the one hand,the ACF plot cuts off after lag 3. The PACF plot seems to tail off after 3.

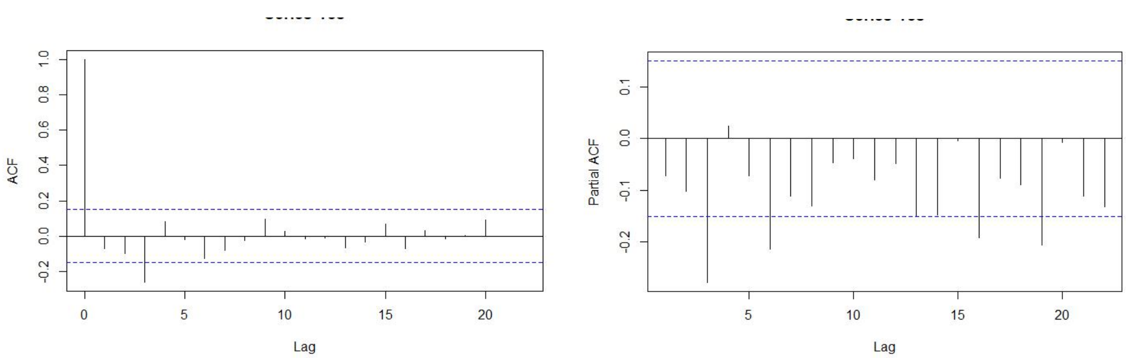


Figure 3 ACF and PACF of rough part

On the other hand,the spectral density of the raw data and the smoothed periodogram (spans=15) are given below.The differences between smoothed periodogram and the spectral densities are not prominent. They are similar for frequencies most of the part with a high peak at 0.22 and a low peak at frequency 0.35.

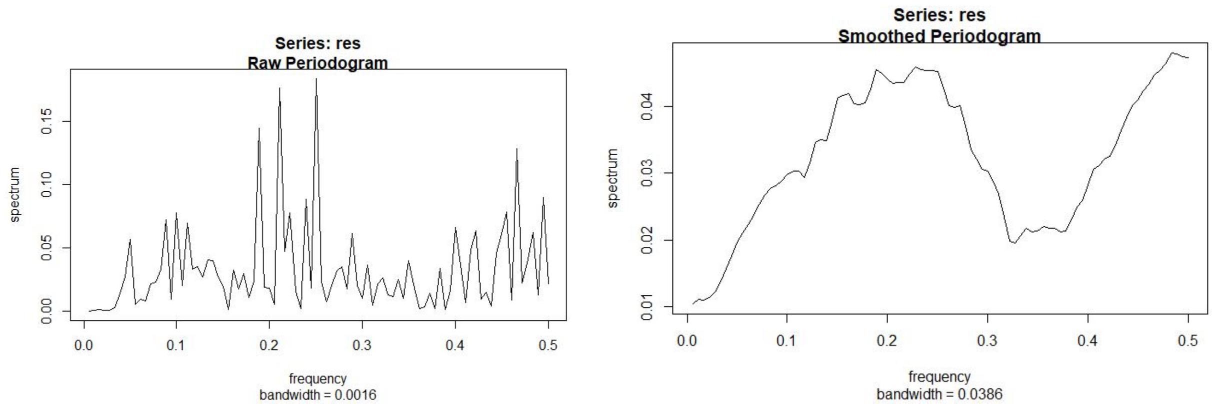
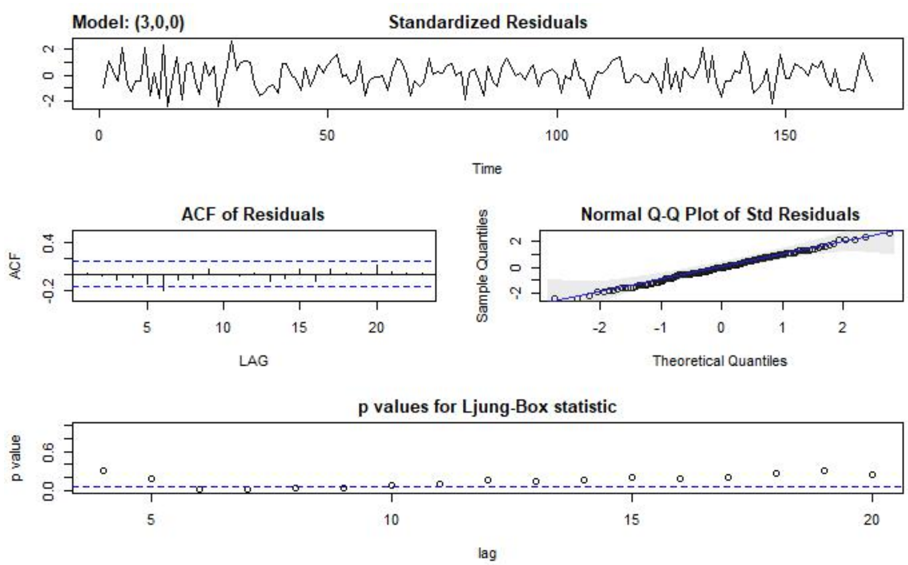


Figure 4 Raw periodogram and smoothed version

We want to use an ARMA(3,0) model as the preliminary identification for the rough part. The residuals of the ARMA(3,0) model looks like a white noise and the ACF and PACF plot of it both approves this property.



To choose the final model, we use AICC criterion to select the most appropriate one. Since both the ACF and PACF plots are negligible after lag 4, we only consider ARIMA models whose values of p and q up to 4. Using the auto.arima function in R, we choose the ARMA(3,0,0) model with predicted parameters -0.1049 , -0.1242, -0.2831, respectively. The AICC value of the final model is -105.59, which is exactly the same model we chose before and the properties of residuals are as above.

For the final model, we first use a method to calculate the critical value of selecting the neighbors in the process of smoothing the periodogram. The critical value plot is as follows, and we select the neighbor value with the smallest critical value, which is 12. And then, we plot the spectral density and the smoothed periodogram on the same graph.

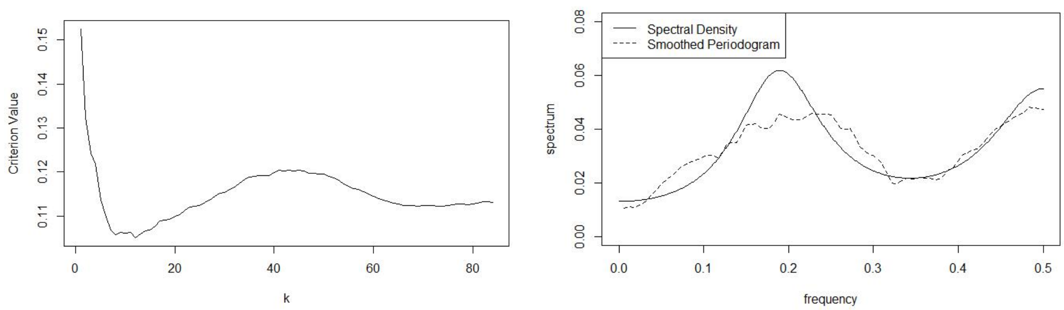


Figure 5 Critical value plot, spectral density and periodogram of final model

From the graph we can find that the spectral density and the smoothed periodogram are similar in variation. They both reach a high peak around 0.18 and a low peak around 0.35. Comparatively the smoothed periodogram is rougher than the spectral density plot.

With the final model, we plot the residuals and its ACF and PACF plot, and use the same method as before to choose the most appropriate neighbor value of the smoothed periodogram of the residuals, which is 39. We can see from the plot that the residuals after fitting the final model is approximately white noise, with the ACF and PACF almost cut off after lag 0. The smoothed periodogram of the residuals reaches high peaks at 0.08, 0.25 and 0.4 approximately, and reaches low peaks at 0.18, 0.35 respectively.

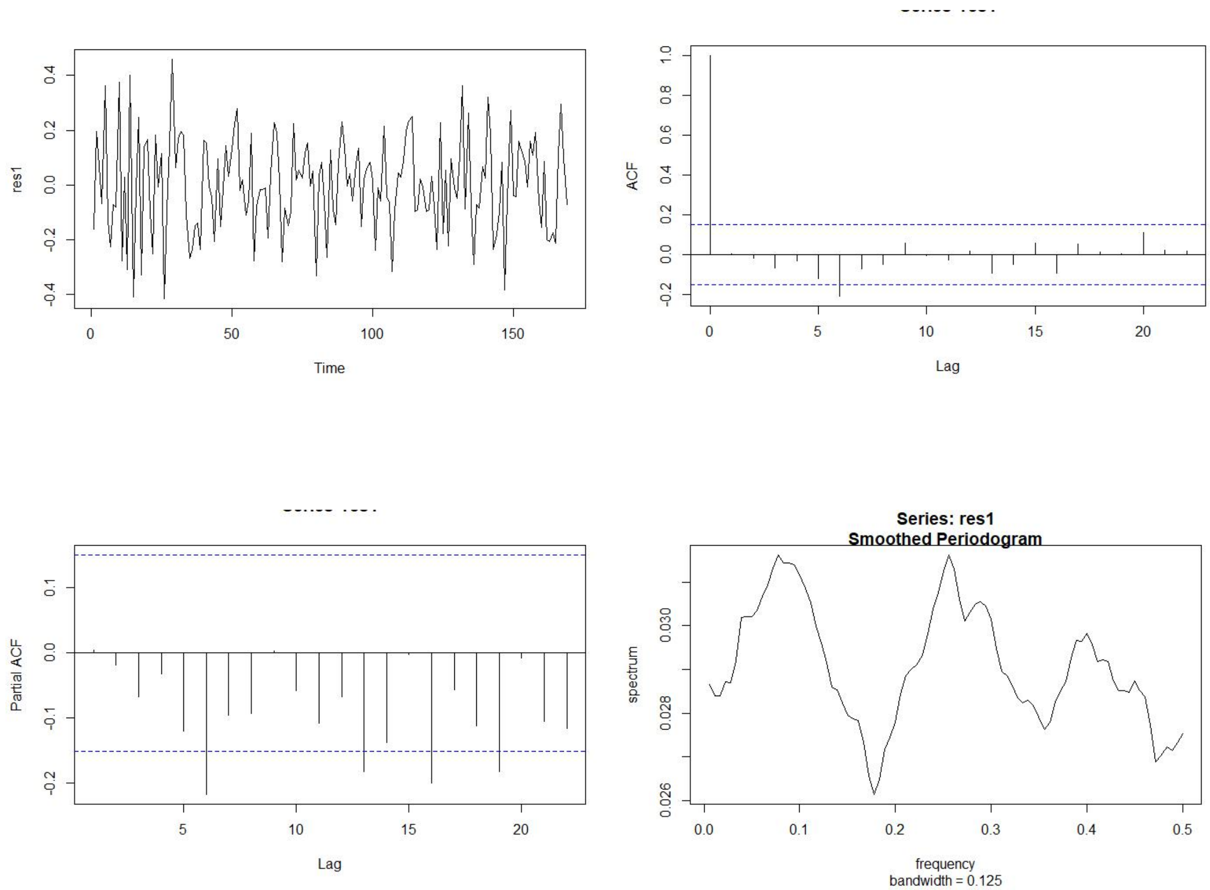


Figure 6 Diagnostics on residuals of final model

To summarize the final model we chose above, the parameters and corresponding standard errors are as follows:

Table 1 Estimated parameters and standard errors

	ar1	ar2	ar3
	-0.1049	-0.1242	-0.2831
s.e.	0.0739	0.0739	0.0745

Now we want to test the our fitted model's capability of forecasting, so we using all the rough part of the data except for the last 6 years to fit an ARMA(3,0,0) model and predict the last

6 years' rough part of the data, then we extrapolate the trend and add them up to predict the last 6 years' data. The prediction and the original data is as follows:

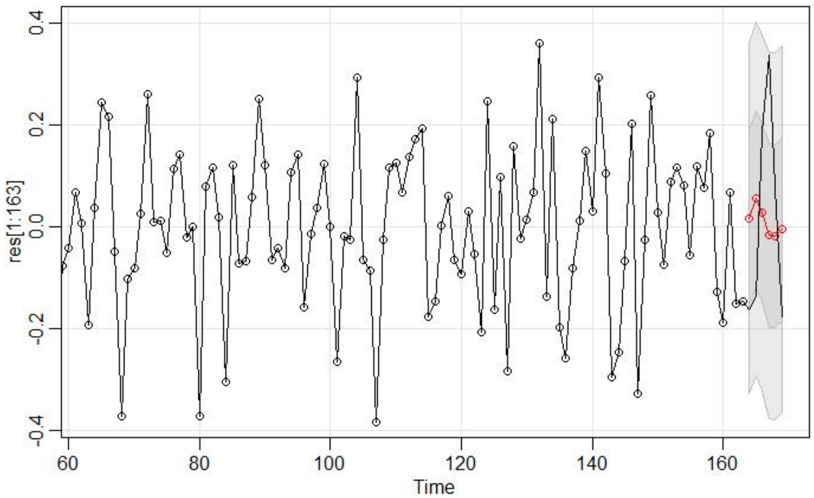


Figure 7 Rough part prediction

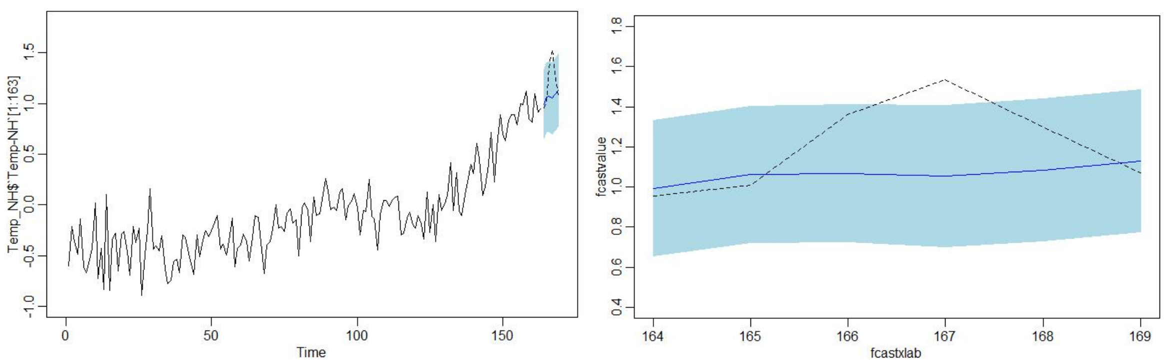


Figure 8 Forecasted values

We can find that the original data of the last 6 years is much rougher than our prediction but still approximately in the prediction intervals.

3 | CONCLUSION

In conclusion, we use a loess model without transformation to fit the trend and then an AR(3) model to predict the rough part. In general, the prediction based on this model seems reasonable with given data since they lie between the confidence interval of 95% significance. Nevertheless, it still cannot be ignored that the prediction of the last 6 years is not perfect.We find it difficult to predict the high peak which seems like seasonal component in the rough part thus leaving one point outside the interval.

Moreover, the spectral density analysis tells us that there is a lying periodogram in the data with a frequency of 0.24 or 0.5. Therefore, our analysis can still be improved by taking the frequency part into our model in the future.