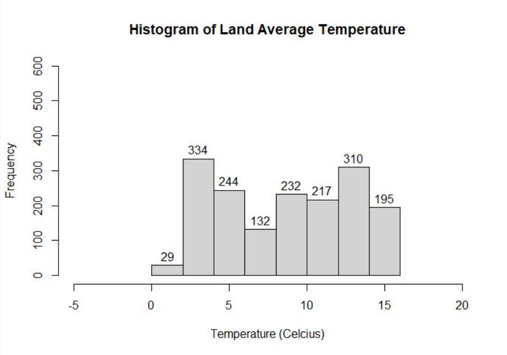
Time Series Analysis of the Climate

SF2943

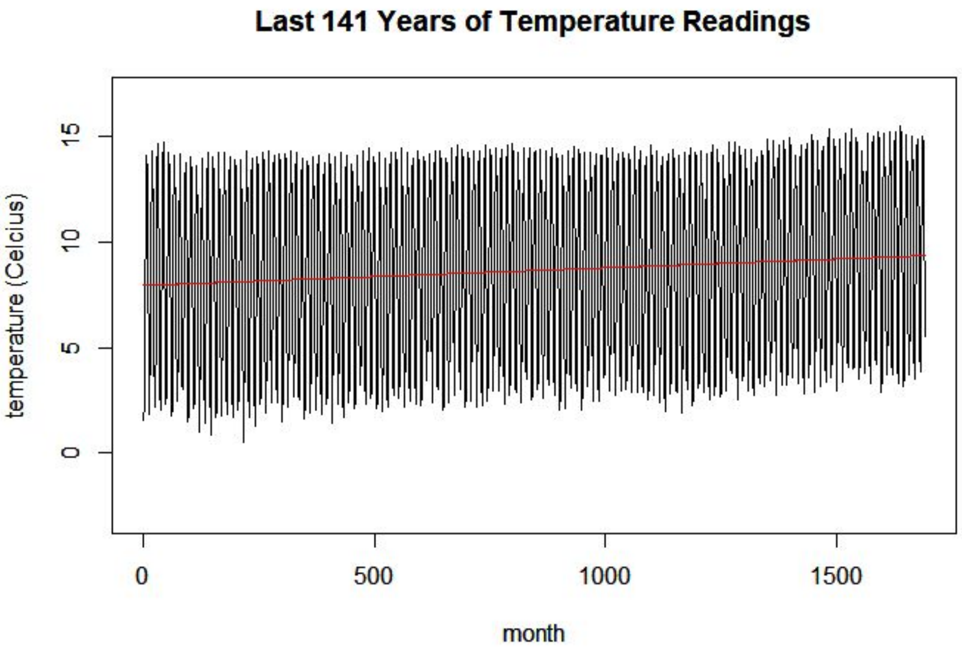
Yage Hao | Ellinor Kalderén | Thomas Lundqvist | Ioannis Athanasiadis

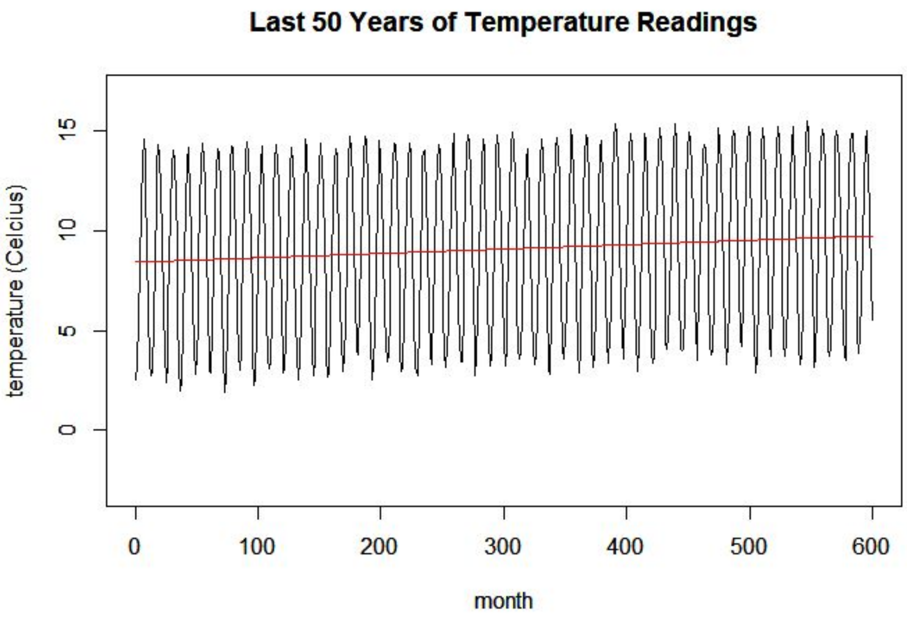
Dataset Description

- Climate change: Earth Surface Temperature from kaggle
- 12 measurements a year from 1750 - 2015 including the 95 % interval
- Due to missing values and inaccurate measurements, 1875-2009 was used for training and 2010-2015 for testing

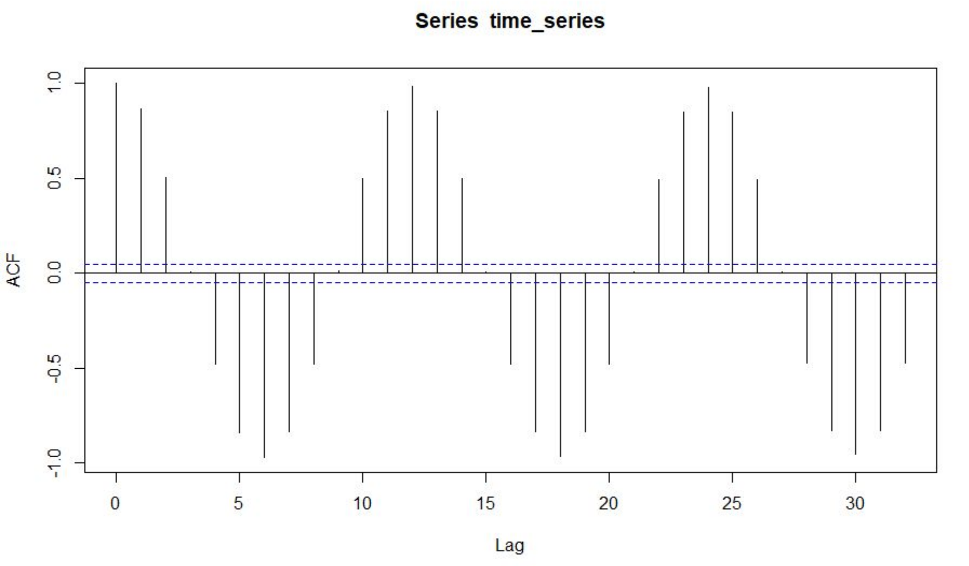


Stationarity analysis



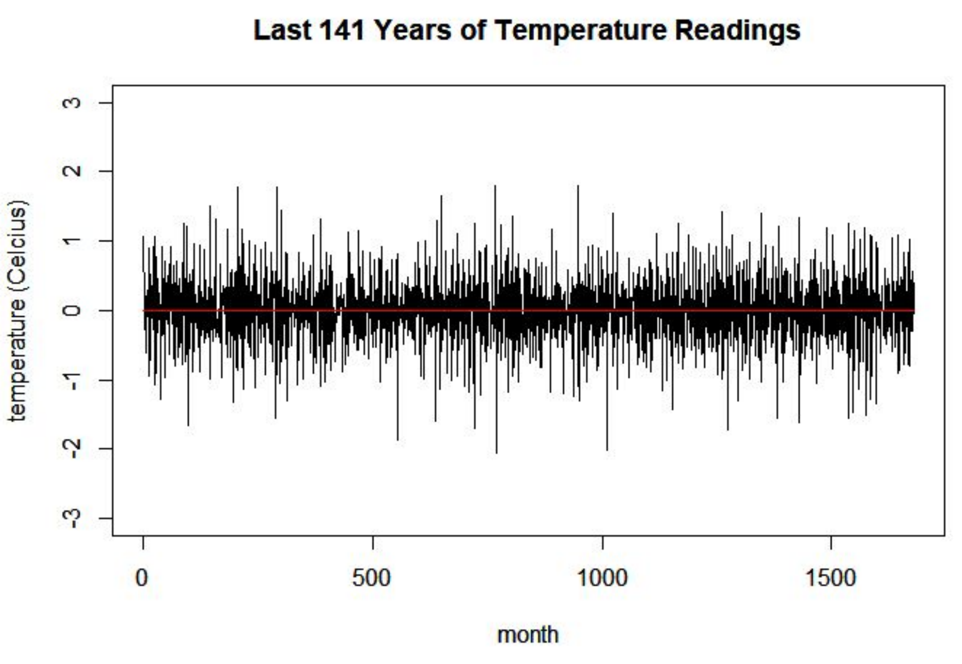


Stationarity analysis



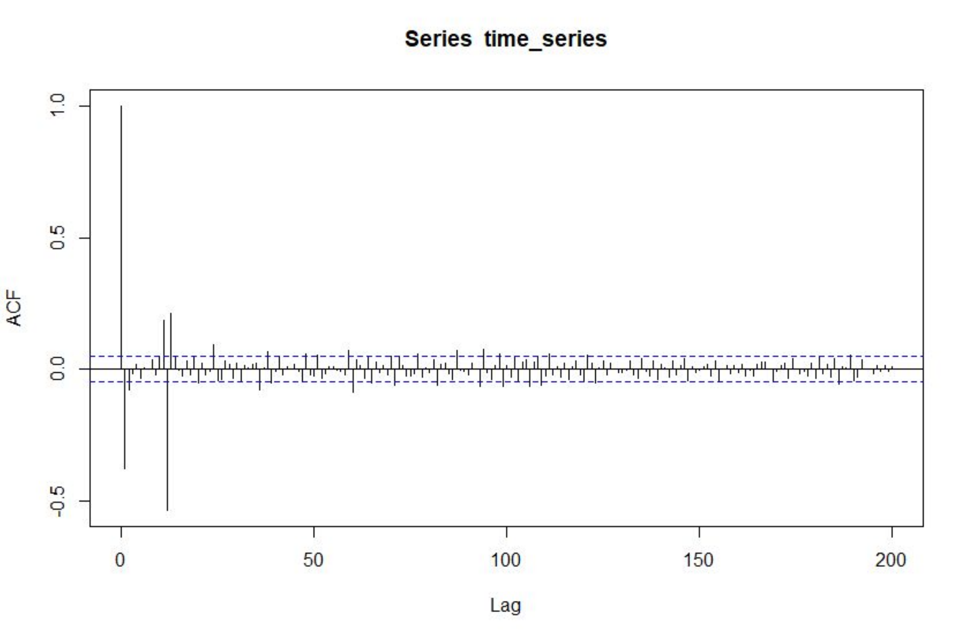
- There seems to exist a positive trend
- Both the data and the ACF shows a seasonality with period 12 (months)
- Seems reasonable since the data is by month.

Stationarity analysis



- By differencing the data with a shift of 12, and then by 1 the data looks more stationary.
- The trend is also removed with this method.

Stationarity analysis



- Here we can see the ACF and that the correlation goes away after just a few values, indicating that the periodicity is gone as well.
- Thus this transformed data is stationary.

Parameter selection and modeling

- We considered our time series to be a realization of a linear process - ARIMA(p,d,q) model
  future values : linear combination of the past values

- Our time series has already been transformed into stationary one - ARMA(p,q)

- Both ACF and PACF are indicative of an ARMA's order
  - We computed the sample ACF and PACF of our time series

Parameter selection and modeling

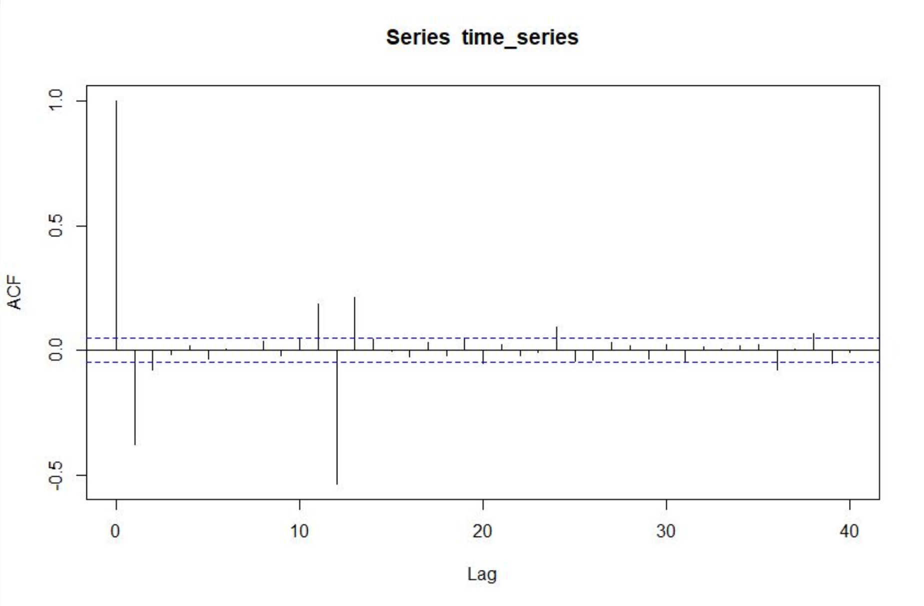


Figure 1. sample ACF function

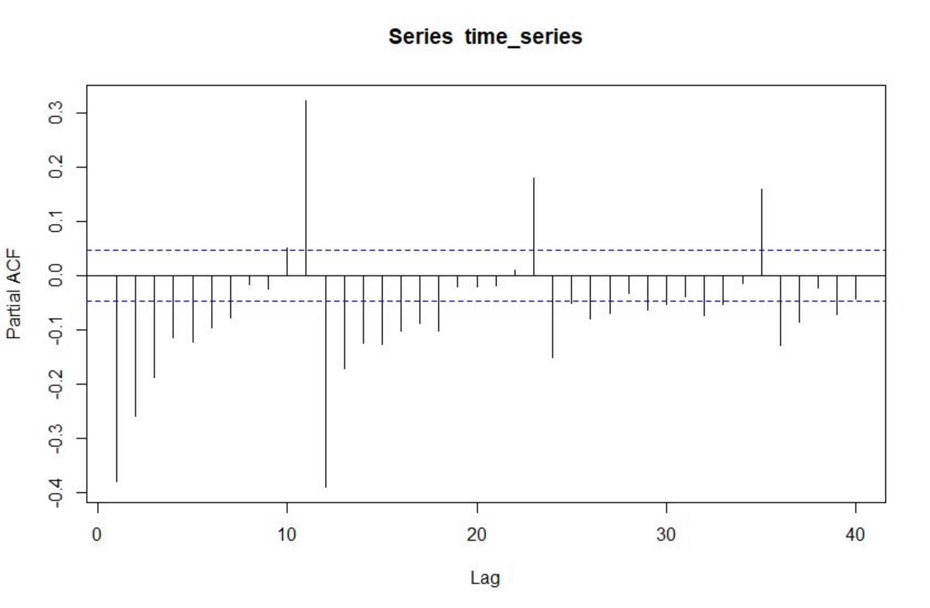


Figure 2. sample PACF function

Parameter selection and modeling

Theory:

- AR(p) process
  - ACF significant values even for lags larger than p
  - PACF insignificant for lags larger than p

- MA(q) process
  - ACF insignificant for lags larger than p
  - PACF significant values even for lags larger than p

- ARMA(p,q) process
  - ACF significant values even for lags larger than p
  - PACF significant values even for lags larger than p

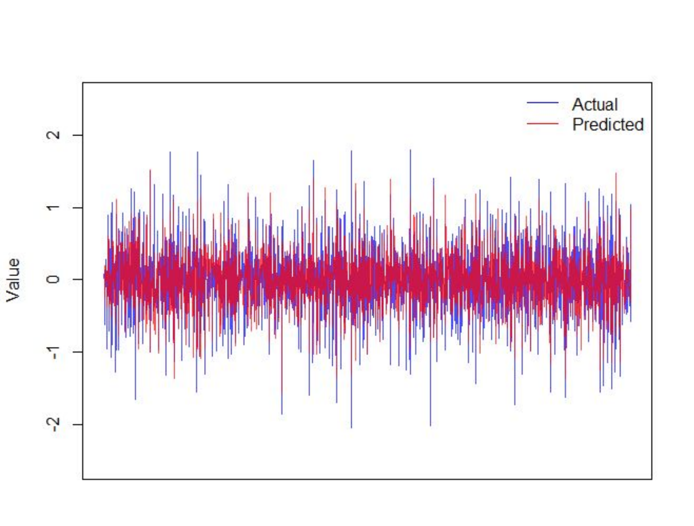
Based on these:

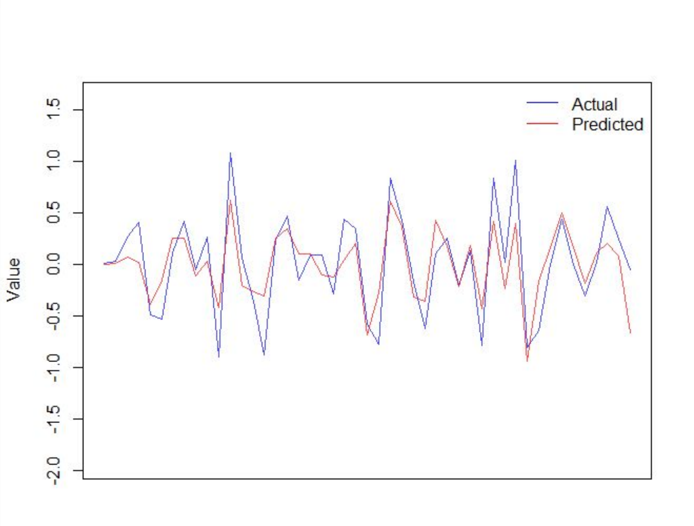
- The time series seems to have both AR and MA contribution
  - There is an evident repetition pattern in the ACF and PACF due to not having perfectly removed the seasonality

Parameter selection and modeling

- Parameter searching for p=[1,15] and q=[1,15]
- Evaluating the models using AIC metric
  - accounts both for model performance and complexity
- Resulted in p=4 and q=13
  - Future values rely heavily on the 4 previous ones
  - Future values rely on some white noise process of 13 lags apart

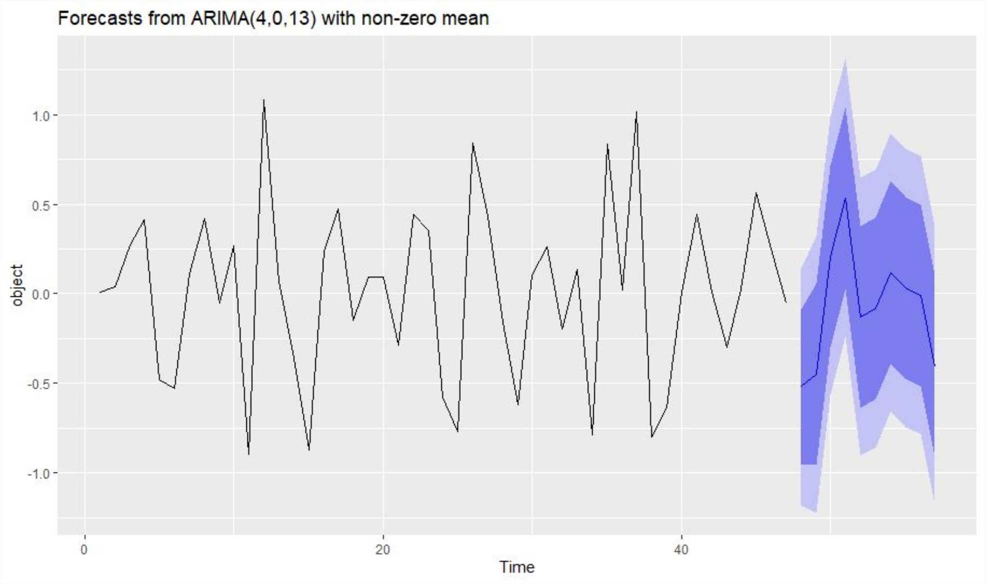
Results





ARMA(4, 13)	on **train**-split	on **test**-split
MSRE	0.09067194	0.07901927

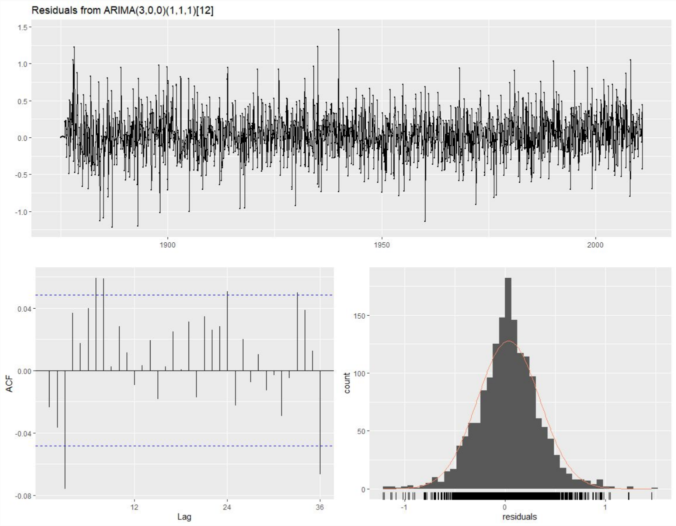
Forecasts

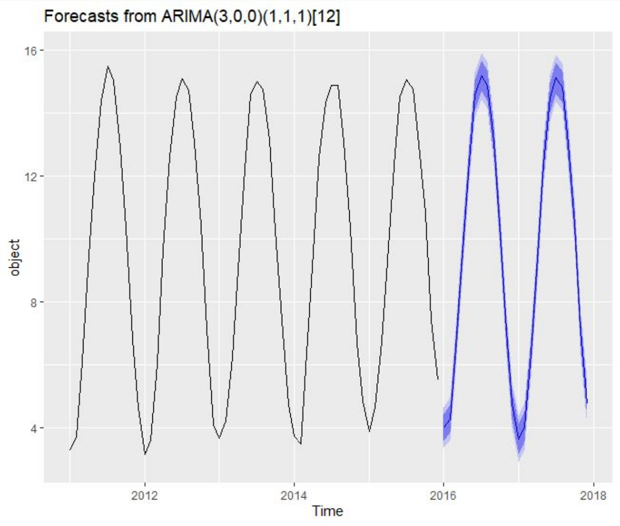


Forecasts from ARIMA(4,0,13) with non-zero mean
object
1.0
0.5
0.0
-0.5
-1.0
0
20
40
Time

Alternative Approach

- Artificial Neural Networks(ANN)-based approach
- auto.arima() function fits ARIMA(3, 0, 0)(1, 1, 1)[12]





Difficulties and Improvements

- auto.arima() does not give the ‘best’ model but the model with lowest AIC and BIC.
- Box-Ljung test results (ACF plots) indicate residuals are not perfect white noise.
- Consider improving our preprocessing steps besides simply removing trend and seasonality.
- Consider ARCH effect.

Conclusion

- We have removed trends and seasonality in order to obtain stationarity
- The series was determined to be stationary according to the ACF
- Different ARMA(p, q) models were compared with AIC
- The models with the lowest AIC was ARMA(4, 13) and was determined to be the optimal model
- MSRE for both sets were low and similar
- The predictions made seemed reasonable comparing to previous values