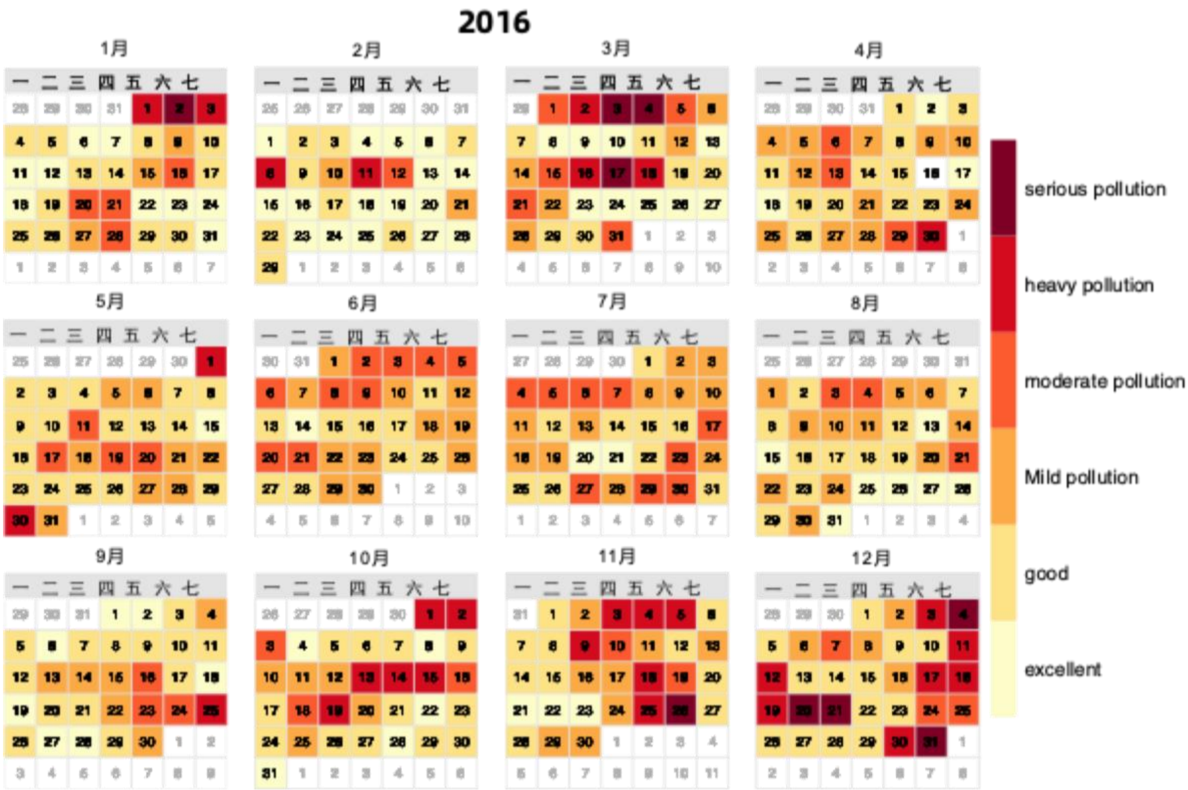
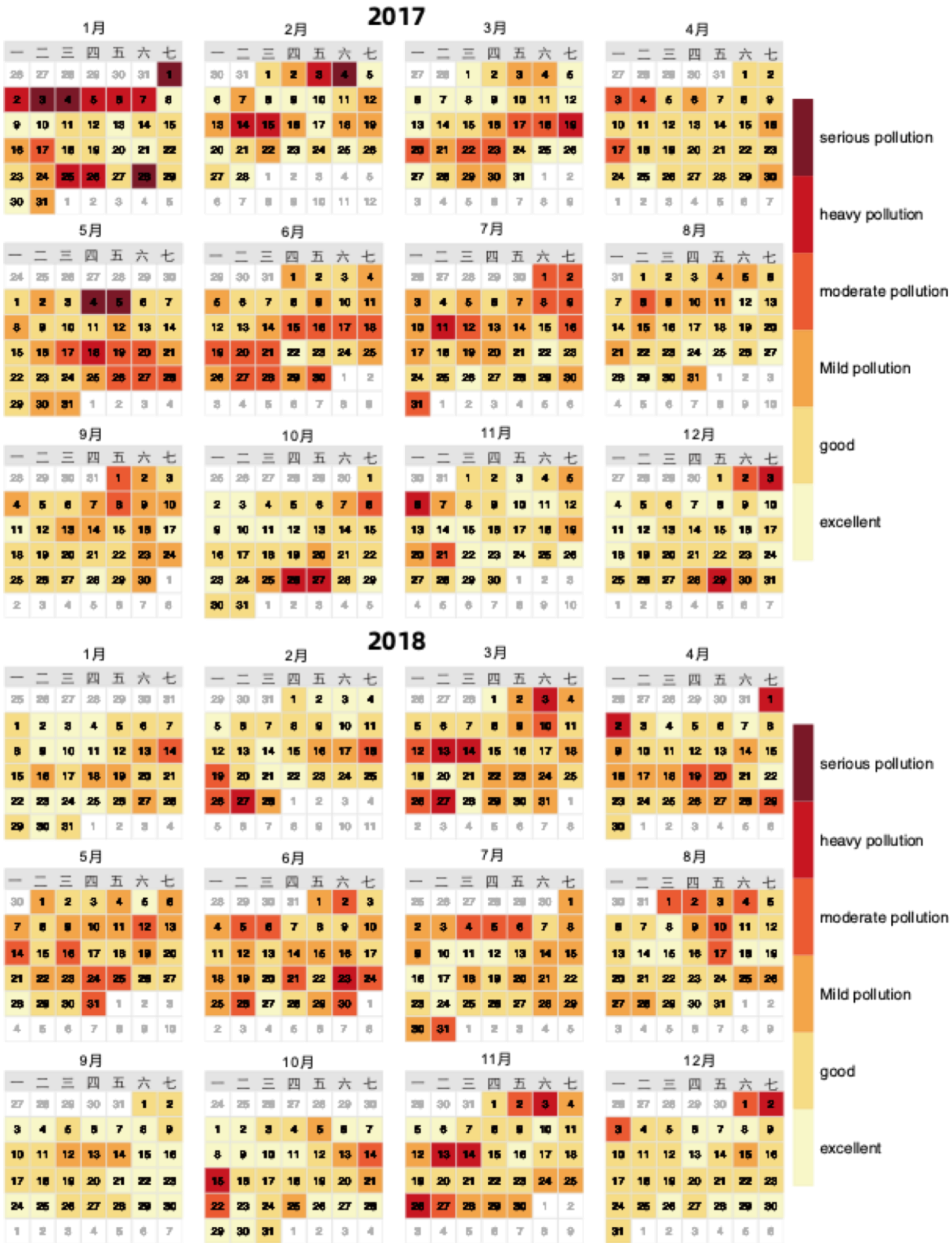
Time Series analysis of AQI using machine learning

Jinhang Yang

1 Data

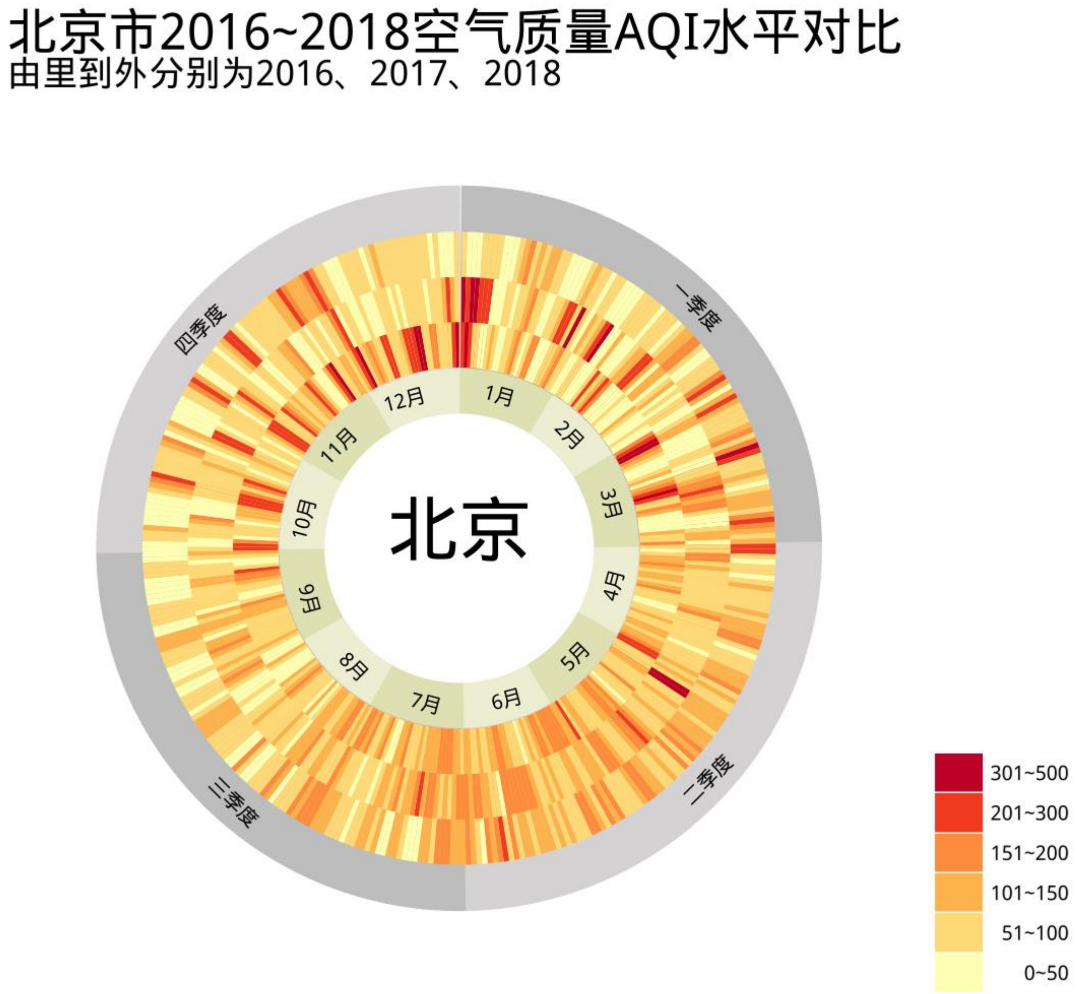
1.1 Annual Calendar Thermal Map of AQI in Beijing from 2016 to 2018



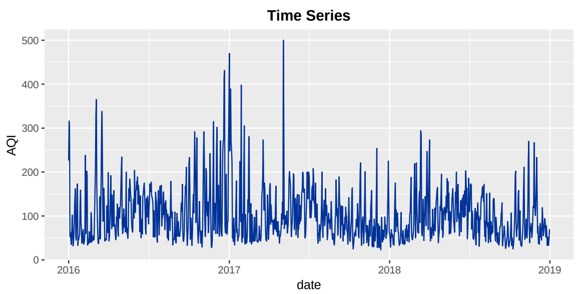

2017
1月
2月
3月
4月
5月
6月
7月
8月
9月
10月
11月
12月
一 二 三 四 五 六 七
serious pollution
heavy pollution
moderate pollution
Mild pollution
good
excellent
2018
1月
2月
3月
4月
5月
6月
7月
8月
9月
10月
11月
12月
一 二 三 四 五 六 七
serious pollution
heavy pollution
moderate pollution
Mild pollution
good
excellent

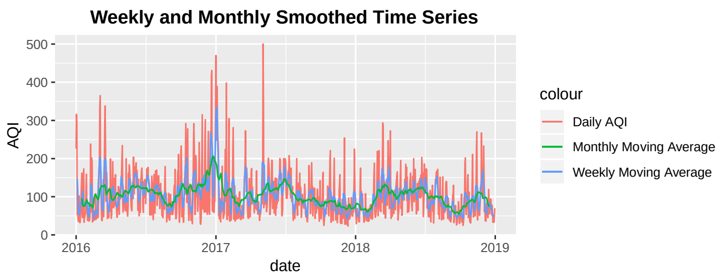
1.2 Cross-sectional monthly and quarterly charts from 2016 to 2018

Further, based on the polar coordinate transformation of ggplot2, the following figure shows the AQI value of Beijing's air quality index for three years, totaling 1095 data points on a chart.[1]

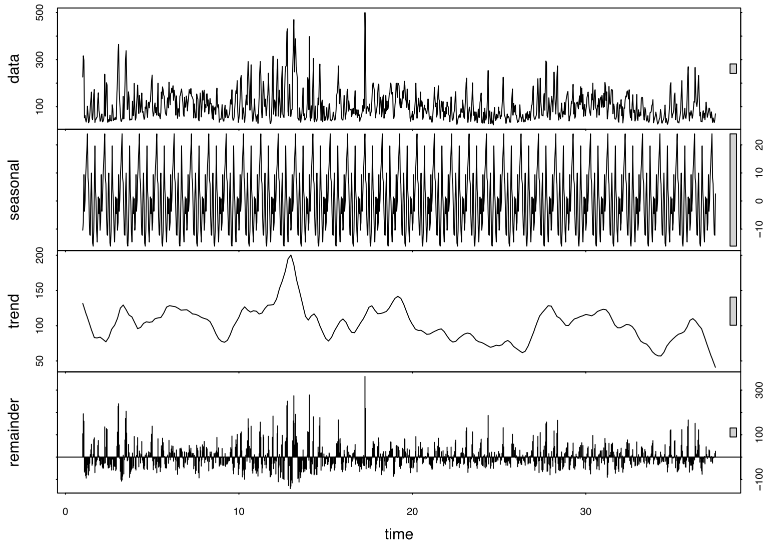


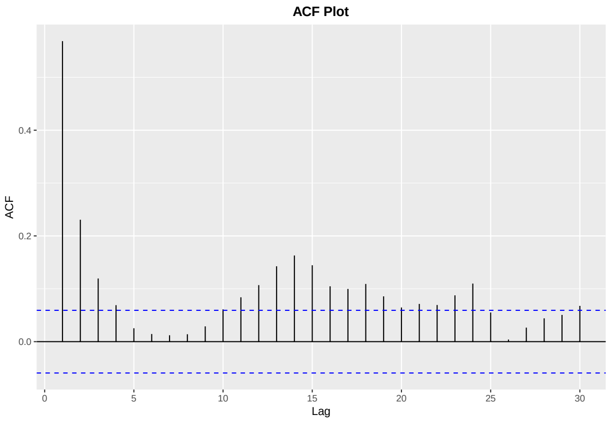
1.3 Time series plotting

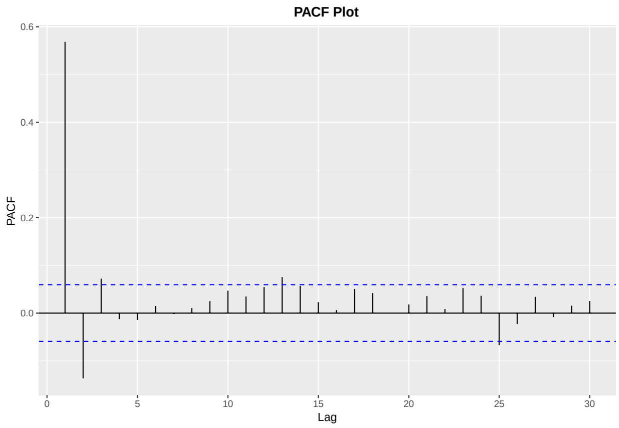




2 Model 1 : Autoregressive Integrated Moving Average model (ARIMA)





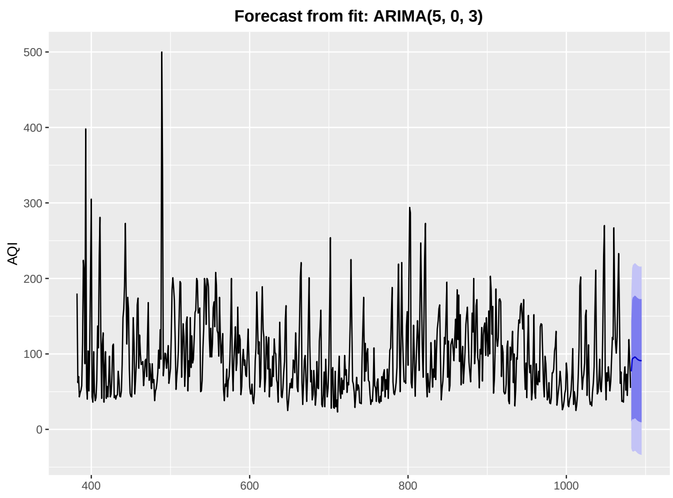


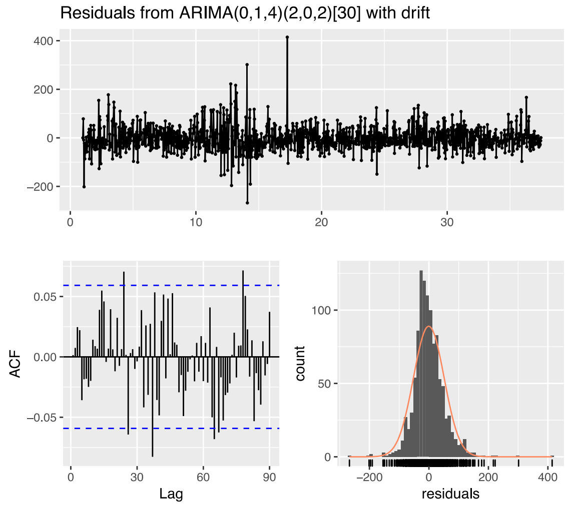
	ARIMA (2,0,0)	ARIMA (3,0,0)	**ARIMA (5,1,3)**	ARIMA(1,1,2)(1,0,1)[30]	ARIMA(3,1,1)(0,0,1)[90]
AIC	11620.03	11616.46	11606.86	11793.76	11792.91
Log likelihood	-5806.02	-5803.23	-5794.43	-5890.88	-5890.45

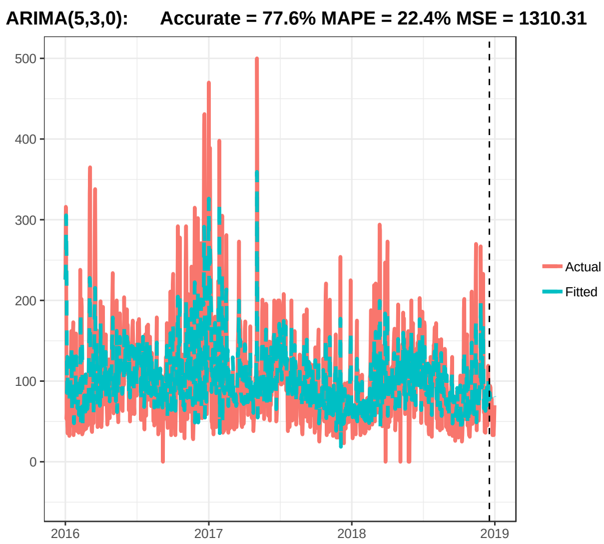
ARIMA (5,1,3) with minimum AIC value and maximum likelihood estimation is selected to establish the following model:

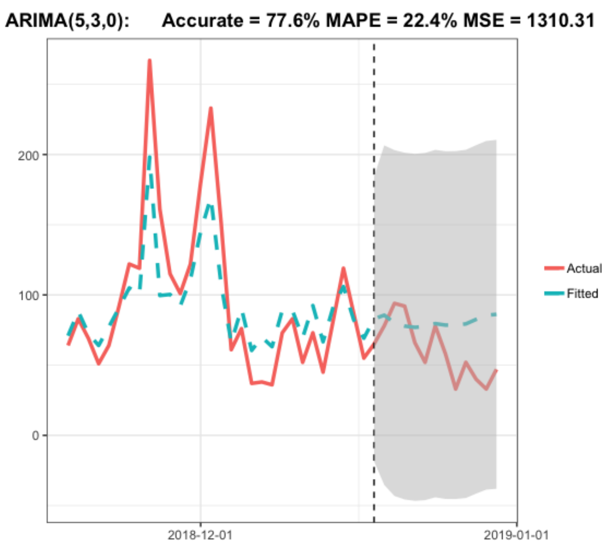
$$(1-\varphi_1 B^1 - \varphi_2 B^2 - \varphi_3 B^3 - \varphi_4 B^4 - \varphi_5 B^5)(1\text{-}B)y_t = c + (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)$$

Results :









3 Bayes

3.1 Function

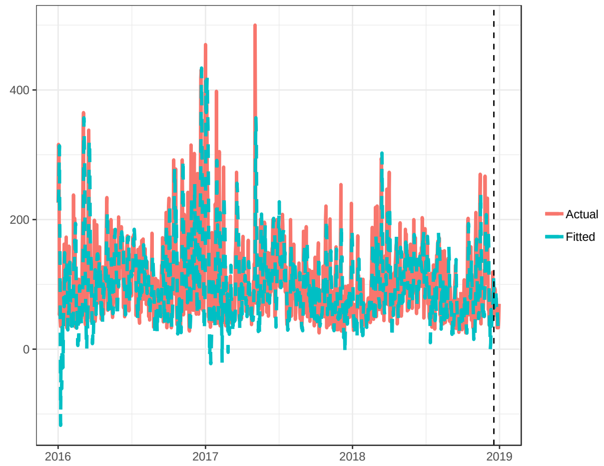
$$Y_t=\mu_t+x_t\beta+S_t+e_t, e_t\sim N(0,{\sigma_e}^2)$$

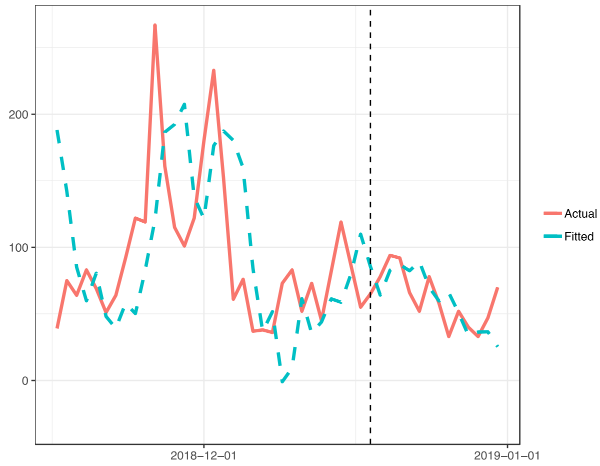
$$\mu_{t+1}=\mu_t+\nu_t, \nu_t\sim N(0, {\sigma_v}^2).$$

- x_t : represents a set of regression vectors
- S_t : seasonality
- μt : local horizontal quantities, which define the change of potential states over time, are expressed as observed trends.

3.2 Results:

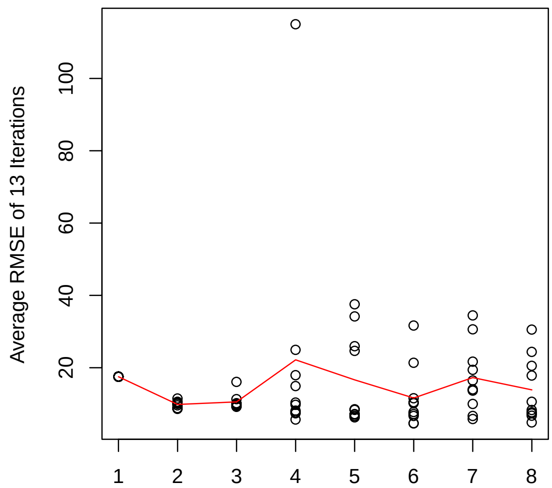
Bayesian: Accurate = 85.05% MAPE = 14.95% MSE = 76.16



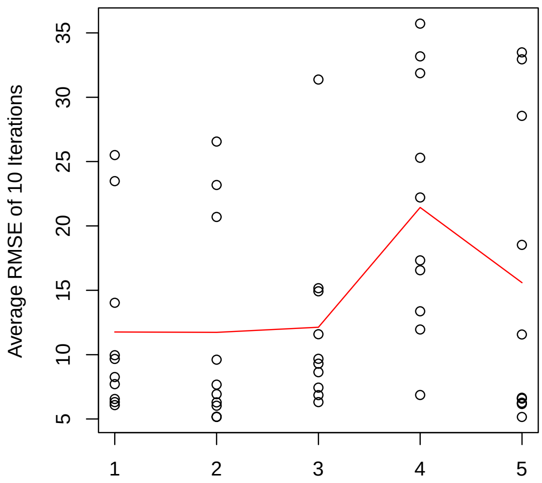


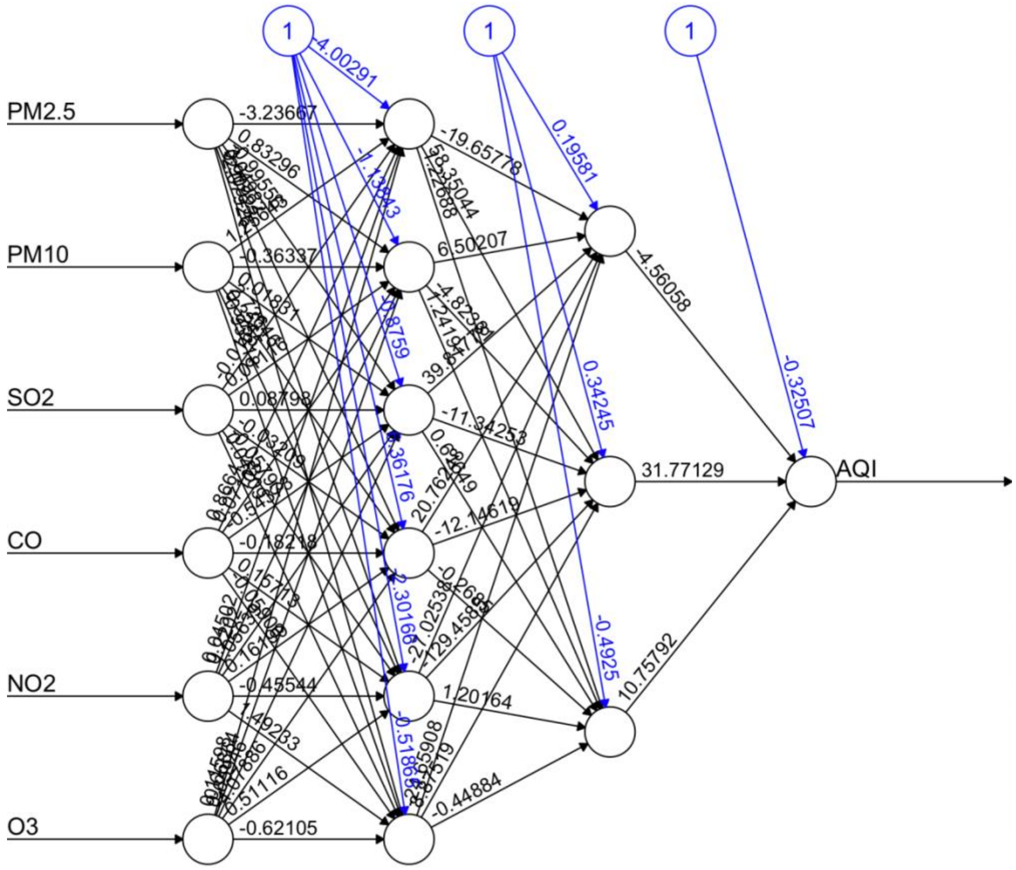
4 Neural Network

First Layer

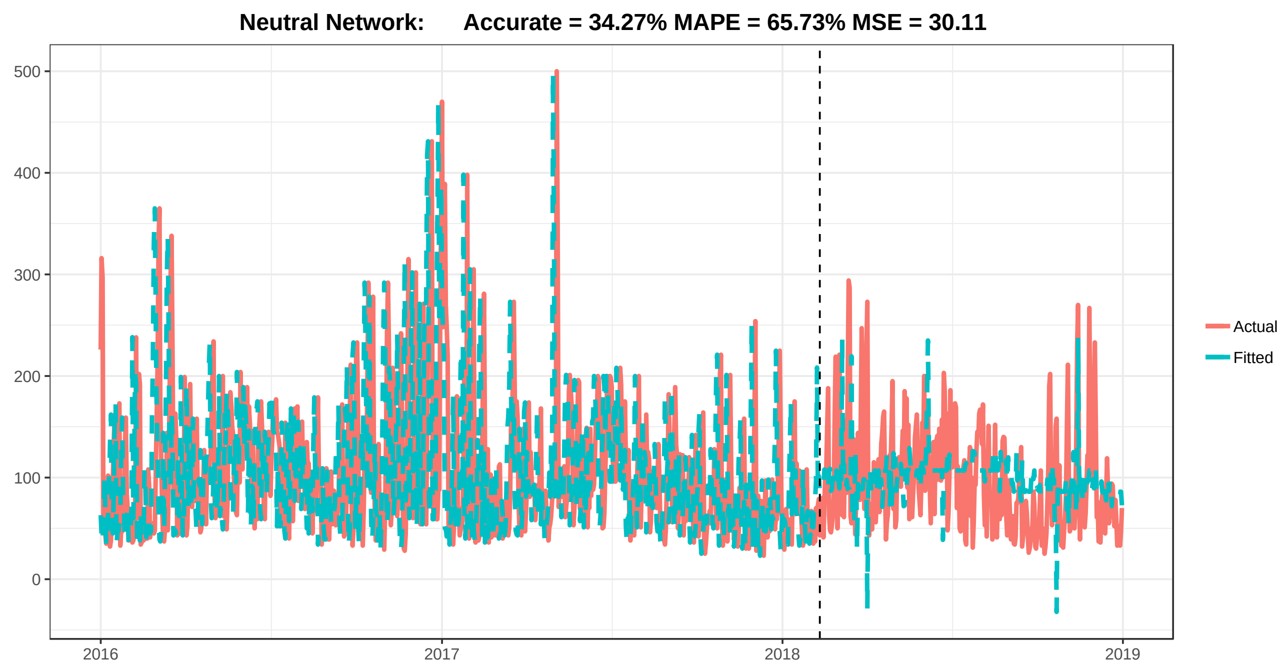


Second Layer




PM2.5
PM10
SO2
CO
NO2
O3
AQI
1
1
1
-4.00291
-3.23667
0.83296
-0.36337
0.08798
-0.18218
-0.45544
-0.62105
-19.65778
6.50207
0.19581
-4.56058
31.77129
10.75792
-0.32507
0.34245
-0.4925
-11.34253
-12.14619
1.20164
-0.44884

Results :

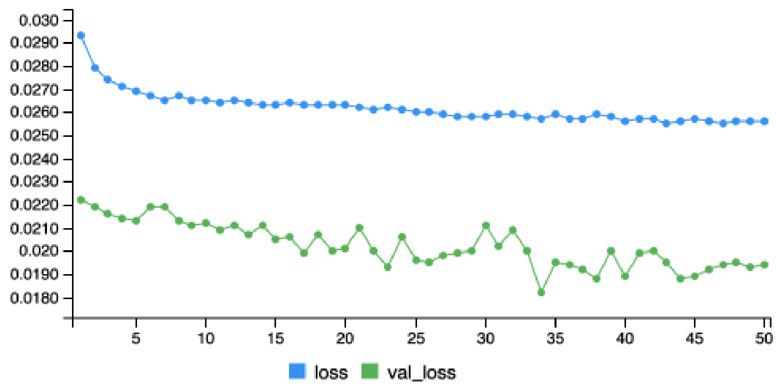

Neutral Network: Accurate = 34.27% MAPE = 65.73% MSE = 30.11
500
400
300
200
100
0
2016
2017
2018
2019
Actual
Fitted

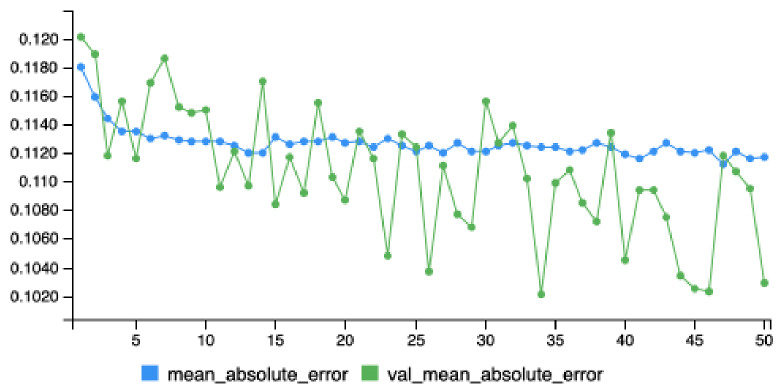
5 Long Short-term Memory (LSTM)

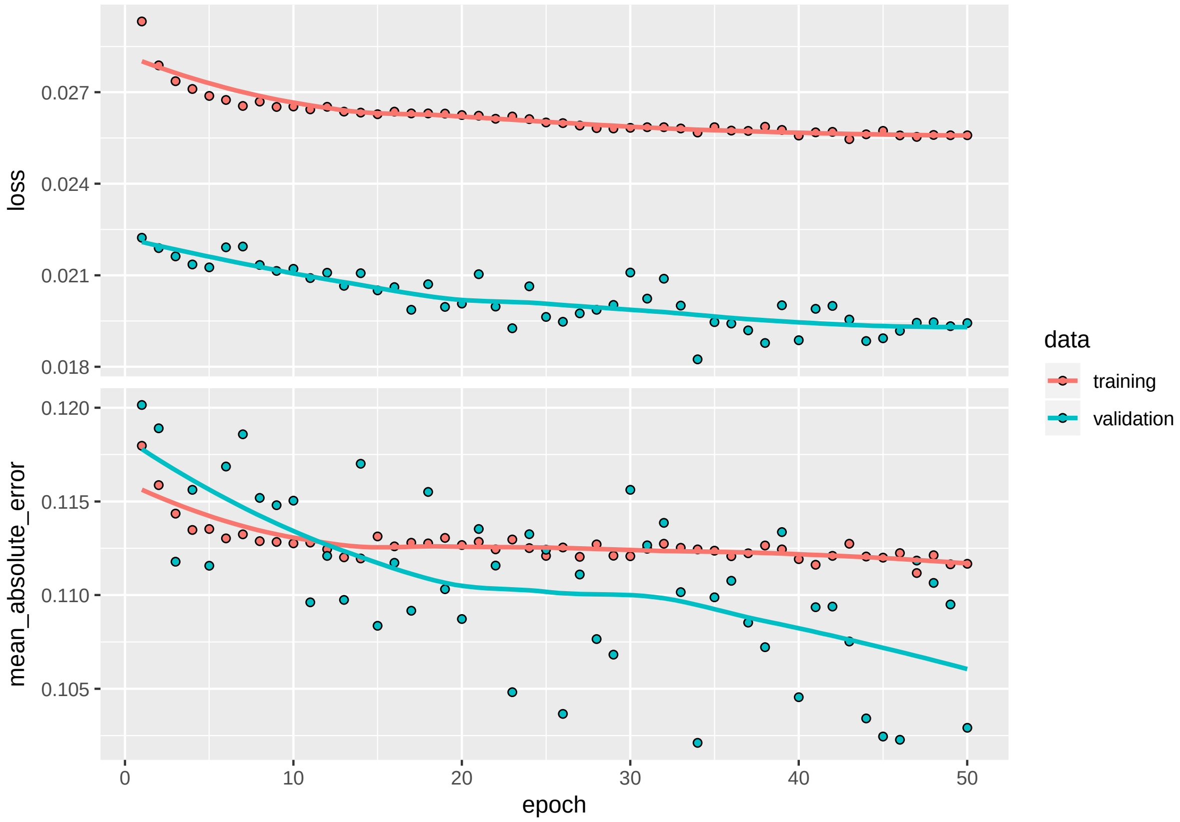
LSTM is an improvement on the recurrent neural network (RNN), which can largely avoid the gradient disappearance of conventional RNN. The results of model fitting using LSTM are as follows:

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_1 (LSTM)                (1, 1)                    12
_________________________________________________________________
dense_1 (Dense)              (1, 1)                    2
=================================================================
Total params: 14
Trainable params: 14
Non-trainable params: 0
```

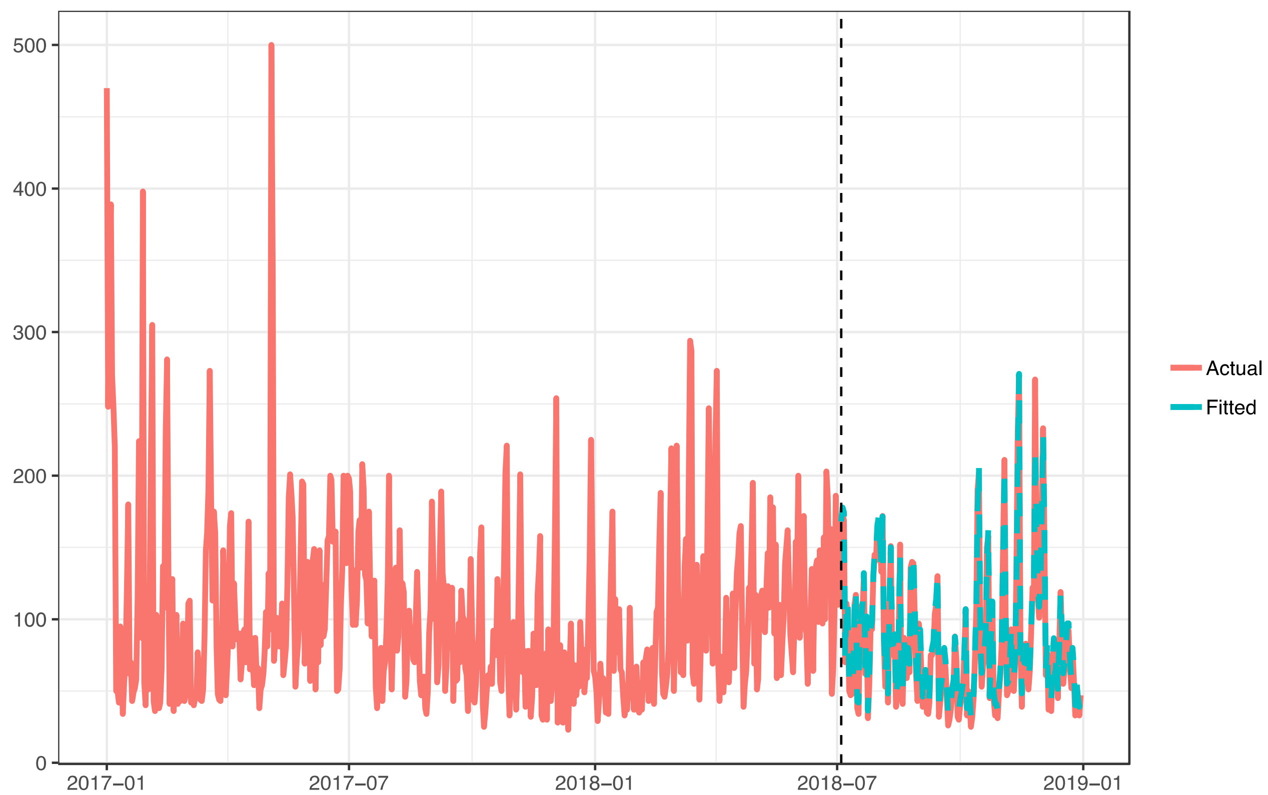
Based on this model, 50 times of fitting the original data, the loss value of the model converges gradually, while the MAPE and standard errors converge concussively. The overall MAPE value keeps around 11%, and the model is basically stable.






loss
0.027
0.024
0.021
0.018
mean_absolute_error
0.120
0.115
0.110
0.105
0
10
20
30
40
50
epoch
data
training
validation

Results:


500
400
300
200
100
0
2017–01
2017–07
2018–01
2018–07
2019–01
Actual
Fitted