Course: Time Series

Secession: 2010 Winter

Student: Wen-Kai Chen

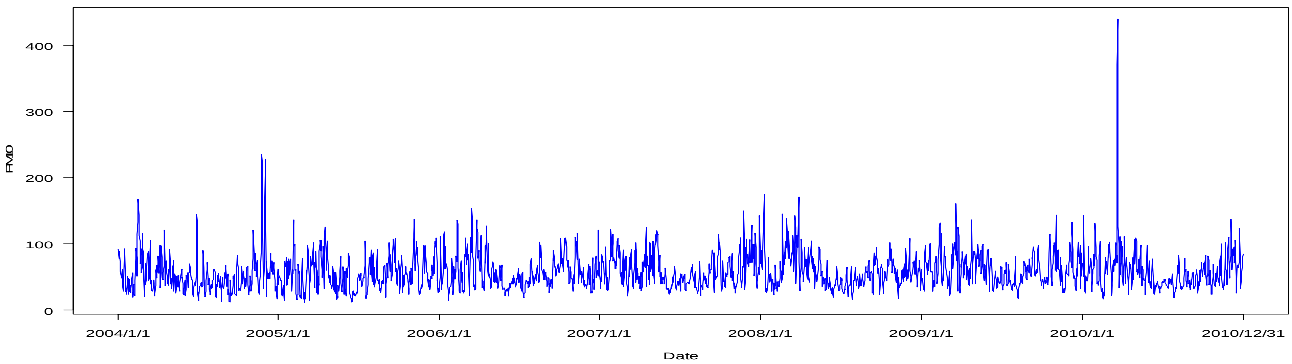
Introduction

Air pollution is one my interested top since lots of scooters emit the exhaust on the streets in Taiwan and sometimes sandstorm bellowed from mainland China also affects the air quality of Taiwan. One of the major pollutants monitored in Taiwan is PM10 which is the particulate matter smaller than 10μm. The reason PM10 is identified as the key index is that the particle is light and tiny enough to float in the air and can easily enter inside of our respiratory system. If people exposed to the environment with high density of PM10, the respiratory system might be damaged.

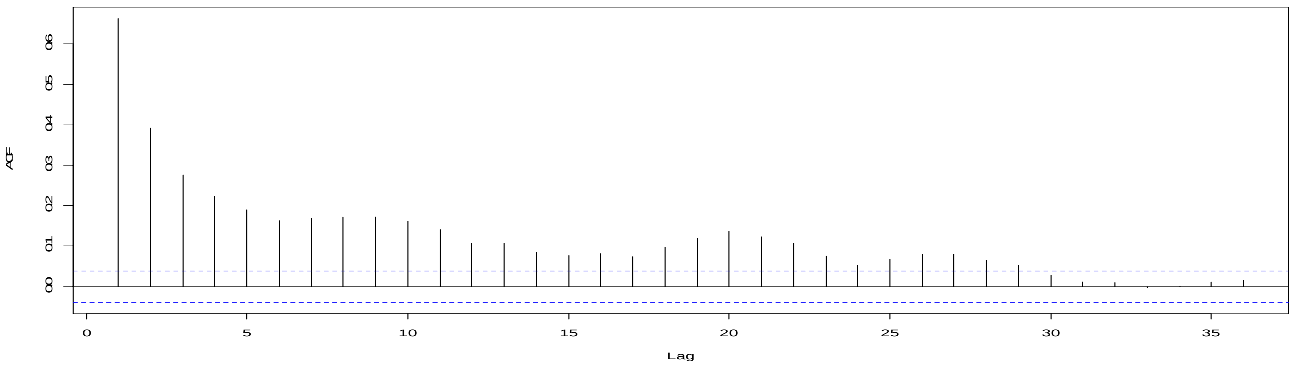
This study applied air quality monitoring data of my hometown FengYuan, in middle Taiwan, from website: http://taqm.epa.gov.tw/taqm/zh-tw/default.aspx. The downloaded data is hourly observation from Jan 1st 2004 to Dec 31st 2010, but daily average is adapted in the study.

Analysis

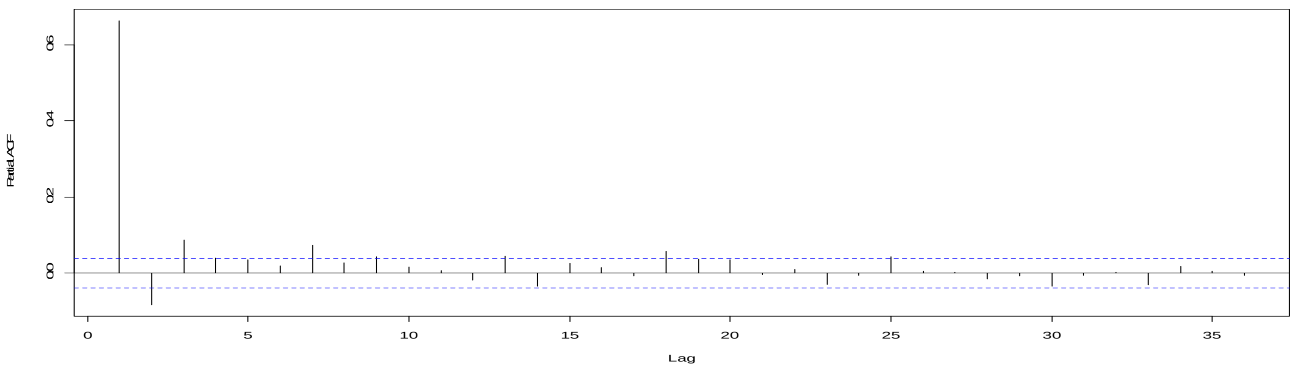
The ACF plot present there might be seasonal autocorrelation at the first place. The EACF contains a triangular of zeros at (1,6), thereby suggesting an ARMA(1,6) model for PM10.



PM10



Series PM10



AR/MA

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	x	x	o	o	x	o	o	o	o	o	x	o	o
2	x	x	o	o	o	x	o	o	o	o	o	o	o	o
3	x	x	x	o	o	x	o	o	o	o	o	o	o	o
4	x	x	o	o	x	x	o	o	o	o	o	o	o	o
5	x	x	x	x	x	x	o	o	o	o	o	o	o	o
6	x	x	x	x	x	x	o	x	o	o	o	o	o	o
7	x	x	x	x	x	x	x	x	o	o	o	o	o	o

The result turns out the estimates of MA coefficients *ma5* is not significant. Hence, a model fixing *ma5* to be zero was subsequently fitted as followed, which has smaller AIC.

R output:

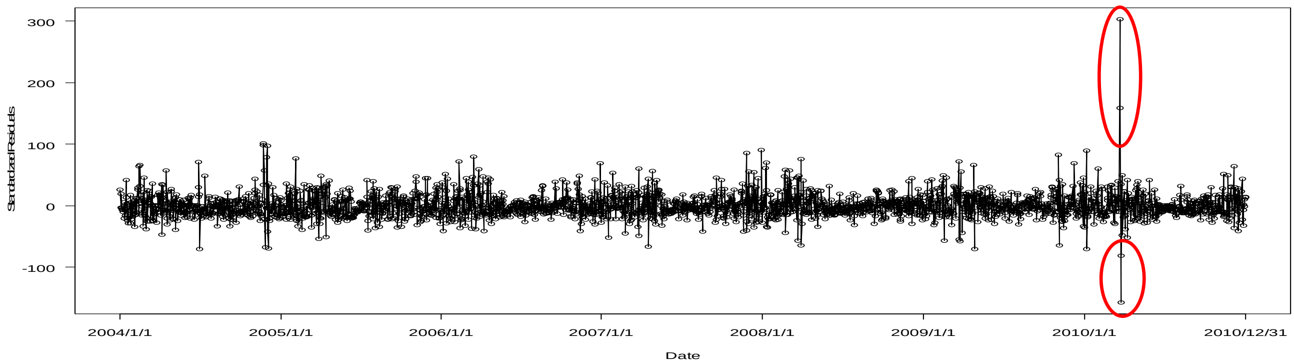
```
arima(x = PM10, order = c(1, 0, 6))
Coefficients:
         ar1      ma1      ma2      ma3      ma4      ma5      ma6  intercept
      0.9560  -0.2401  -0.3191  -0.1309  -0.0503  -0.0037  -0.0500    57.3273
s.e.  0.0172   0.0263   0.0239   0.0224   0.0219   0.0216   0.0219     1.8445
sigma^2 estimated as 400.4:  log likelihood = -11254.61,  aic = 22525.21
```

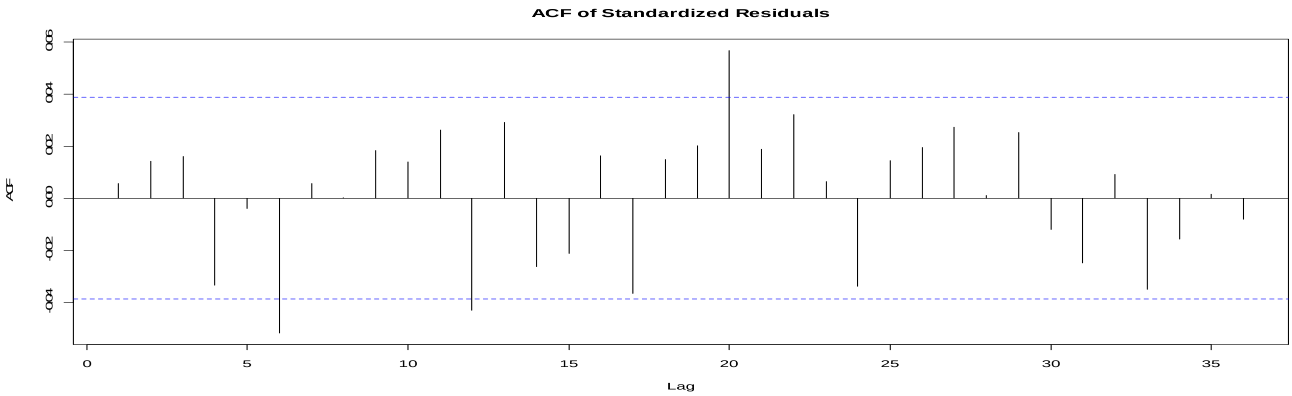
The result fixed ma5=0 shown below has slightly smaller AIC.

R output:

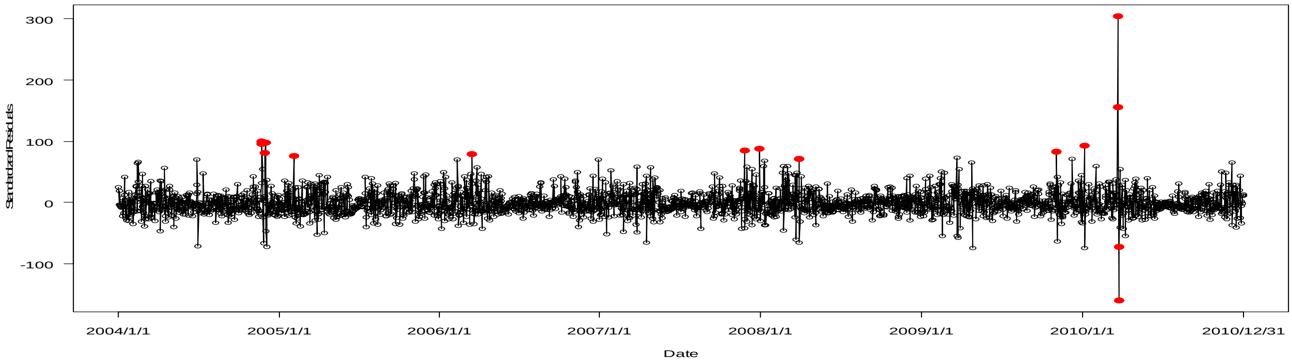
```
arima(x = PM10, order = c(1, 0, 6), fixed = c(NA, NA, NA, NA, NA, 0, NA, NA))
Coefficients:
         ar1      ma1      ma2      ma3      ma4  ma5      ma6  intercept
      0.9553  -0.2395  -0.3192  -0.1315  -0.0506    0  -0.0505    57.3018
s.e.  0.0169   0.0262   0.0240   0.0222   0.0218    0   0.0218     1.8395
sigma^2 estimated as 400.4:  log likelihood = -11254.62,  aic = 22523.24
```

The standardized residual plot shows that there are some odd points having relative large residuals, while ACF shows most of autocorrelation are not significant except at lag6, lag12, and lag20 though they are still samll. The seasonal autocorrelation concern looks subtle in ACF.





If we take the outliers into consideration of model by detecting innovative outlier of R, where they are marked as red dots below, the result shows a smaller AIC of 201901.16 than original AIC of 22523.24. To check the fitted model, residuals spray within in a band and the sample ACF of residuals improved a little, only significant correlations at lag15 and lag 20.



R output:

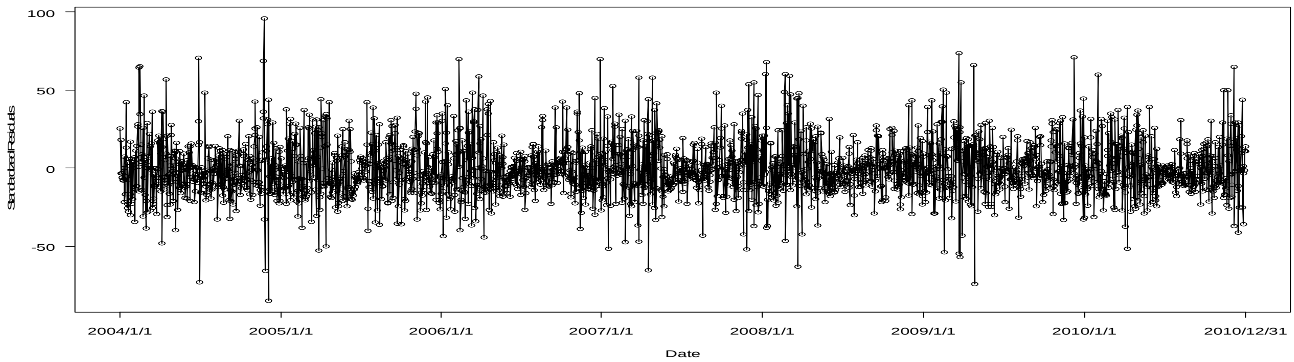
arima(x = PM10, order = c(1, 0, 6), fixed = c(NA, NA, NA, NA, NA, 0, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA), io = c(325, 326, 333, 334, 398, 802, 1418, 1453, 1542, 2125, 2188, 2264, 2265, 2266, 2268))

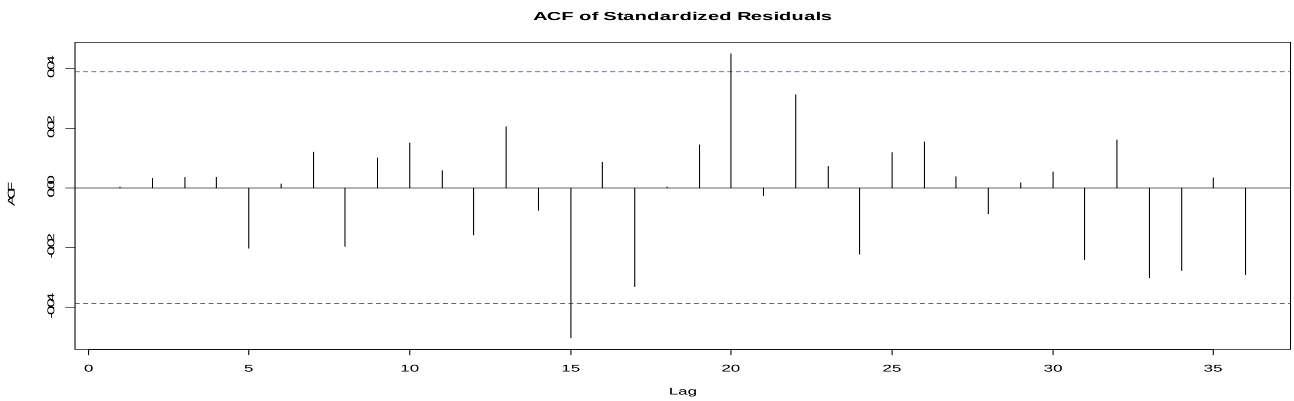
Coefficients:

	ar1	ma1	ma2	ma3	ma4	ma5	ma6	intercept	IO-325	IO-326	IO-333	IO-334	IO-398	IO-802	IO-1418	IO-1453
	0.9528	-0.2431	-0.2849	-0.1188	-0.0679	0	-0.0533	56.7297	63.3712	71.5329	70.0097	117.3659	61.3475	48.1101	67.445	69.6042
s.e.	0.018	0.0277	0.0249	0.0228	0.0227	0	0.0219	1.7068	16.9535	17.0032	16.9995	16.9389	14.2571	14.2655	14.3363	14.2615

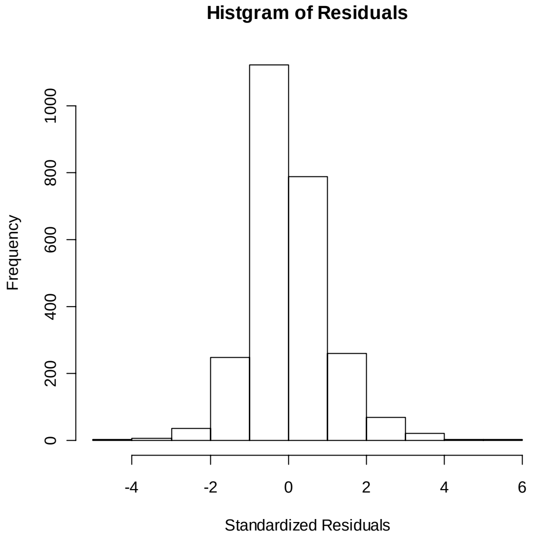
	IO-1542	IO-2125	IO-2188	IO-2264	IO-2265	IO-2266	IO-2268
	76.1	84.7926	94.014	296.9269	360.122	39.9985	-72.5966
s.e.	14.2953	14.2985	14.2749	17.2783	20.1712	17.319	14.3608

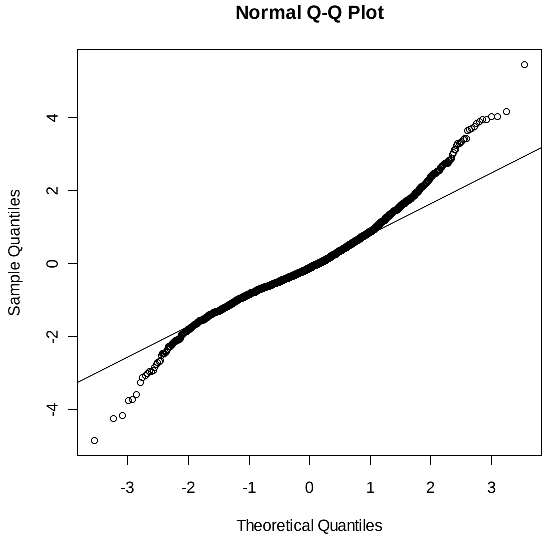
sigma^2 estimated as 310: log likelihood = -10928.58, aic = 21901.16





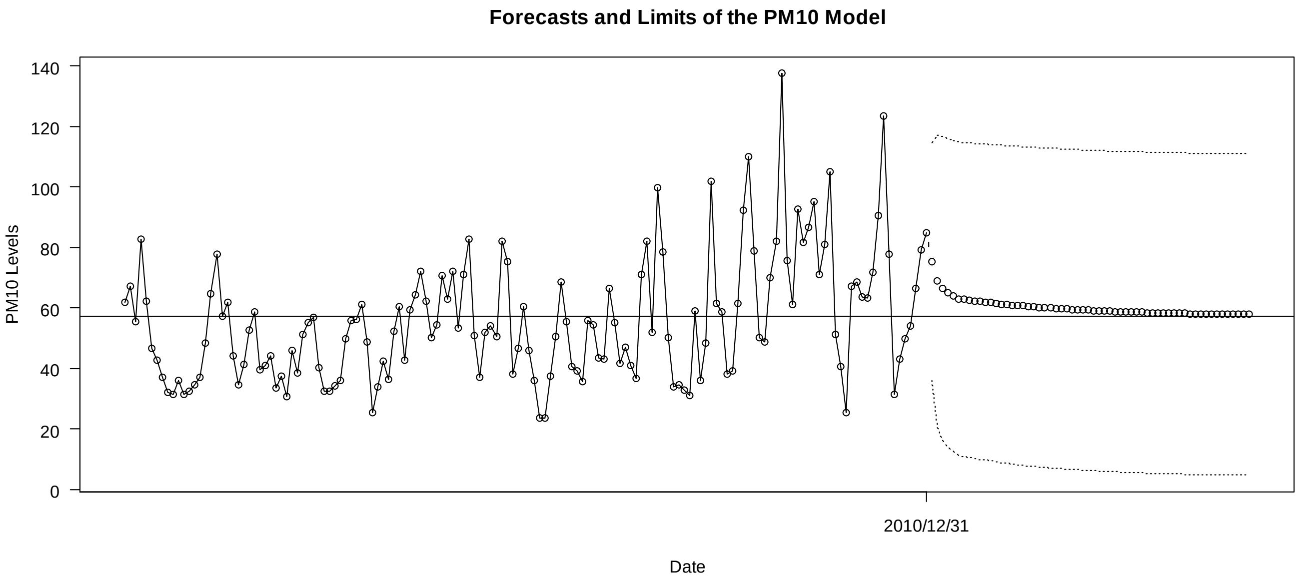
To look further, we examine the normality of the error term via quantile-quantile plot and histogram of the residuals. Although the shape of histogram is bell-shaped, it seems to have a higher and narrower peak than normal distribution. It is confirmed by Normal Q-Q plot that the residuals distribution has a heavy tail.





Forecasts

The figure displays this series with forecasts out to lead time 365 with upper and lower 95% prediction limits for those forecasts. The forecasts approach to mean 57.3 exponentially and the prediction limits expend as the lead time increase. However, the prediction limits seems large.



Investigation of Outlier

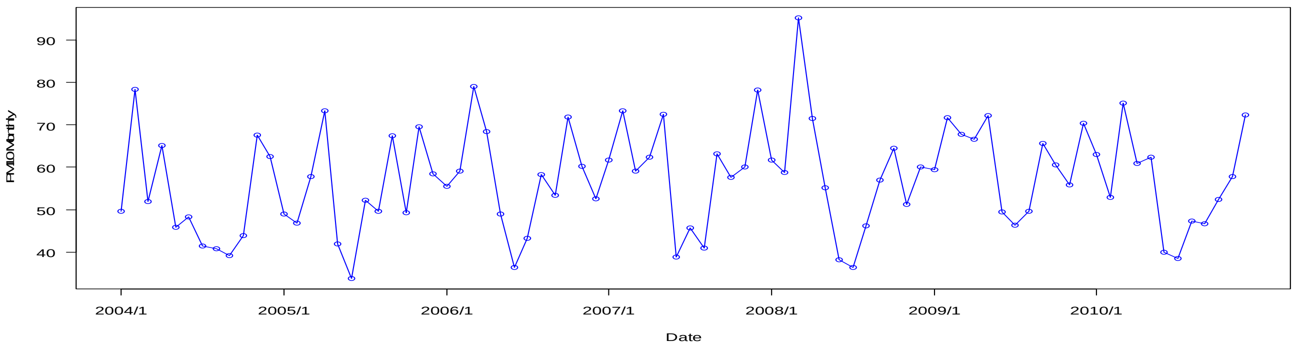
According to the sandstorm report on http://dust.epa.gov.tw/dust/zh-tw/Database.aspx , these dates were reported the air quality affected by sandstorm severely: 2005/11/29~2005/11/30, 2006/3/19~2006/3/20, 2006/3/29~2006/4/1, 2007/1/28~2007/1/29, 2007/12/30~2007/12/31, 2008/3/2~2008/3/4, 2009/4/25, 2009/12/26, 2010/3/21 ~ 2010/3/24, and 2010/4/29. In addition, during winter season, some abnormal PM10 readings might result from burning straws after rice harvest in an airless day, for example, 2004/11/23~2004/11/26 and 2004/12/1~2004/12/3. However, these dates only covered a part of the innovative outlier in the previous analysis.

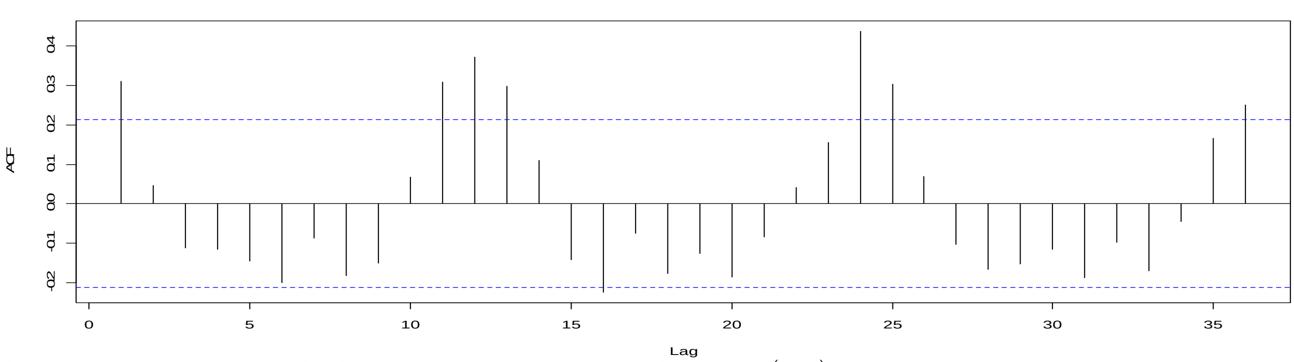
It seems that daily PM10 data would fluctuate wildly because of specific weather condition or human behavior. Meanwhile, those evens are anticipated to have seasonal pattern but not observed in the previous analysis.

To reveal the pattern of data, below the Monthly average PM10 is used to mitigate the effect of extreme daily observations.

Monthly Data analysis:

Based on the sample autocorrelation plot, strong correlations are found at lag12, 24 and 36 and they are not significant at others if 2 lags from lag12 and 24. It looks like a typical ARMA(0,1)x(0,1,1)$_{12}$ model.





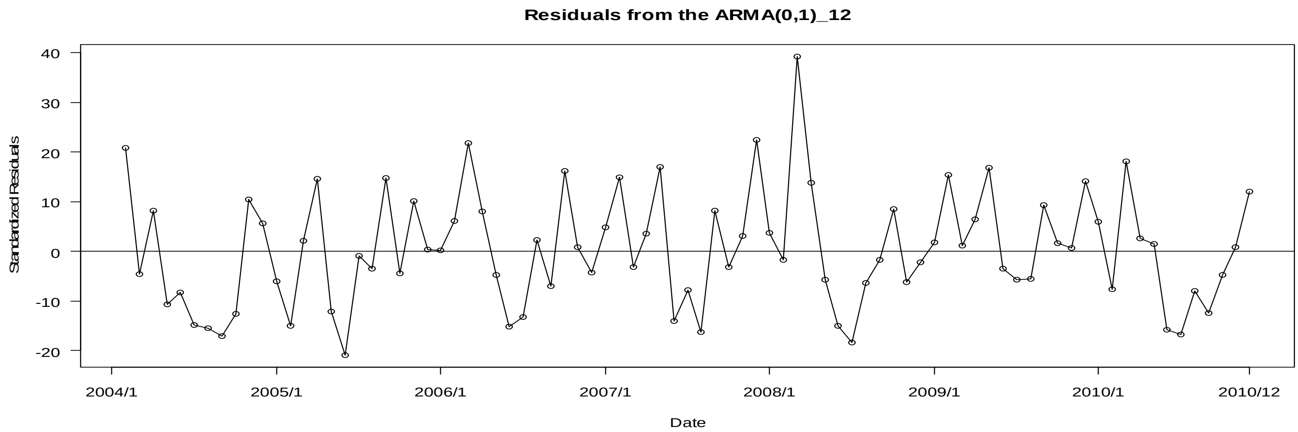
The coefficient of ma1 looks not significant and then ARMA(0,1)$_{12}$ is fitted.

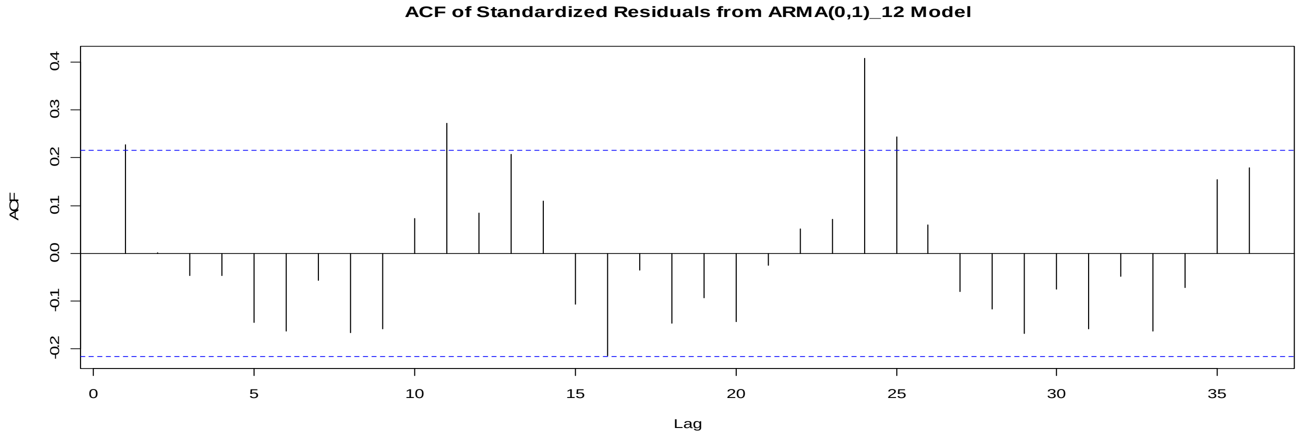
R output:

```
Coefficients:
         ma1    sma1  intercept
      0.2407   0.187    56.7984
s.e.  0.1076   0.092     1.7526
sigma^2 estimated as 124.7:  log likelihood = -322.12,  aic = 650.23
```

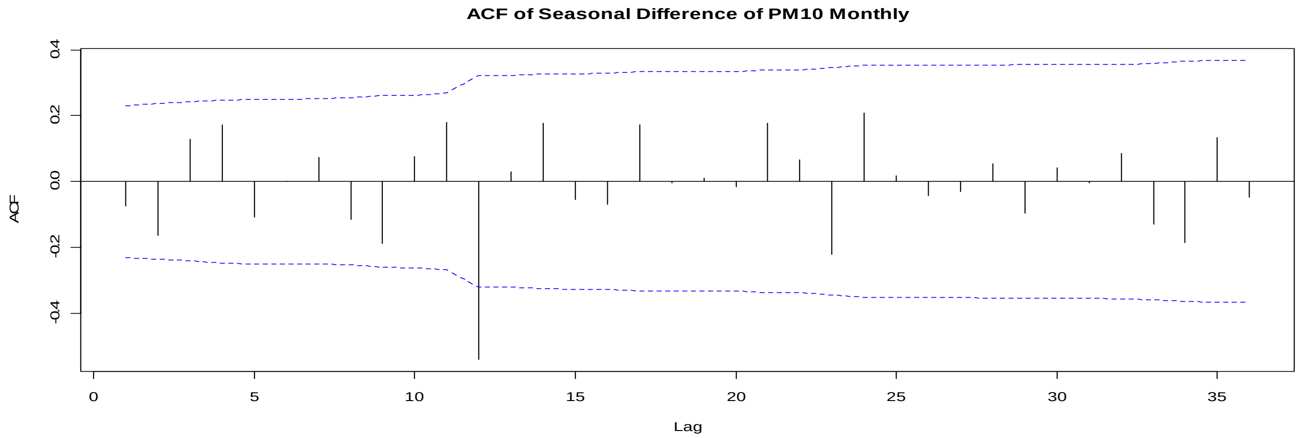
R output:

```
arima(x = PM10m, seasonal = list(order = c(0, 0, 1), period = 12))
Coefficients:
        sma1  intercept
      0.2510    56.7466
s.e.  0.0876     1.5210
sigma^2 estimated as 131.0:  log likelihood = -324.35,  aic = 652.7
```





The sample autocorrelation figure of residuals from ARMA(0,1)$_{12}$ indicates there are still seasonal correlation. If we look the sample autocorrelation function of seasonal difference,

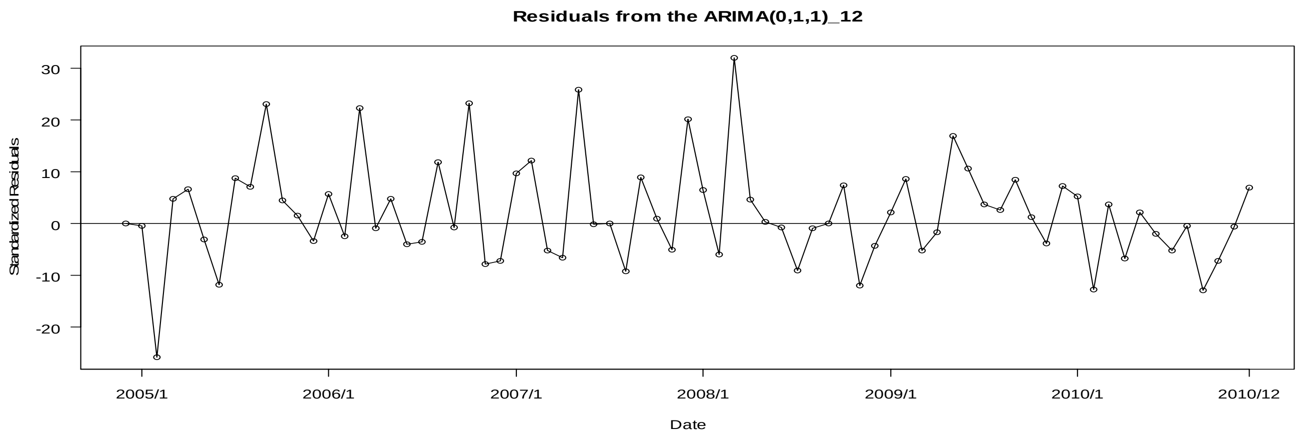


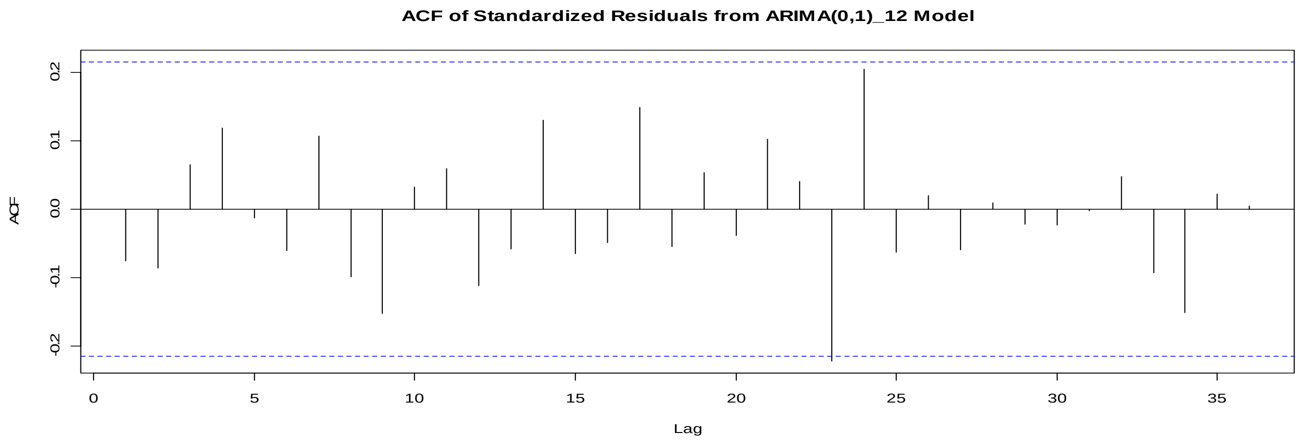
only correlation at lag 12 is significant. The result of ARIMA(0,1,1)$_{12}$ is as follows:

R output:

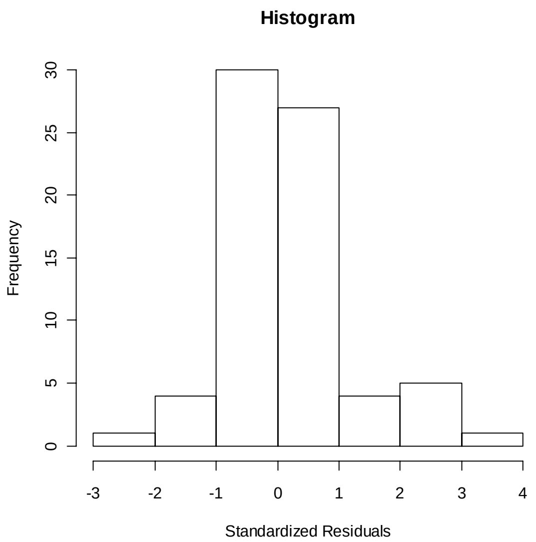
```
arima(x = PM10m, seasonal = list(order = c(0, 1, 1), period = 12))
Coefficients:
          sma1
       -0.7045
s.e.    0.1471
sigma^2 estimated as 99.7:  log likelihood = -271.91,   aic = 545.82
```

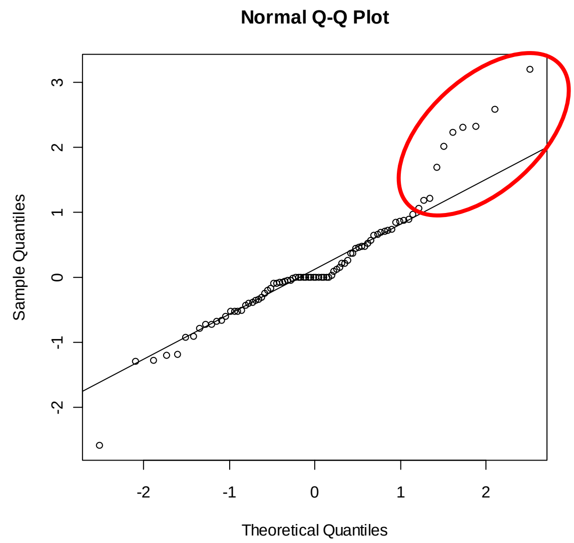
To look further, the ACF plot of ARIMA(0,1,1)$_{12}$ present no significant correlation except for marginal correlation at lag24. Overall, the model seems captured the essence of the dependence in the series. Furthermore, the Ljung-Box gives a chi-square value of 0.4883 with p-value 0.4847, which also indicates the model is appropriate.





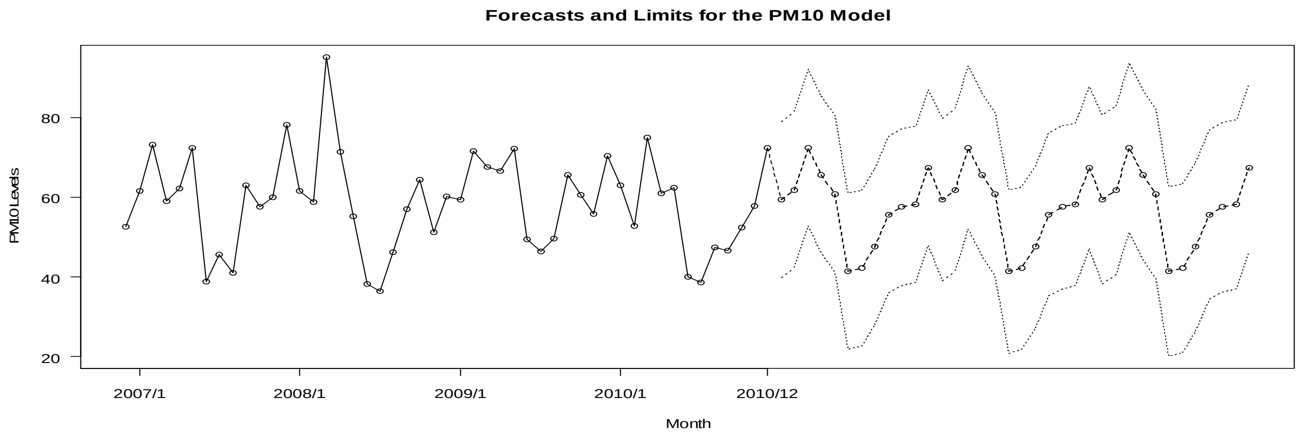
To check the normality of error term, the histogram and normal Q-Q plot of residuals both suggest it is non normal distribution.





Forecasting:

The plot displays 4 years of observed data and forecast out three years. The PM10 level has peak in winter (December and March), and valley in summer. The model capture the pattern as expected though the normality of residual is not statisfied.


Forecasts and Limits for the PM10 Model
80
60
40
20
PM10 Levels
2007/1
2008/1
2009/1
2010/1
2010/12
Month