Time Series Study of

Maximum Wind Speed In New York

STA 9701 Time Series Course Project 1

Xiaoying Zhao

November 12, 2019

Abstract

The objective of this project was to study a time series model representative of maximum monthly wind speed in New York, and then to use this time series model to predict the maximum monthly wind speed. The data was collected monthly from January 2006 to December 2018, totally 156 analytical observations. This time series generated by seasonal auto regressive integrated moving average (ARIMA) model was used to predict the maximum monthly wind speed of 2019.

Table of Contents

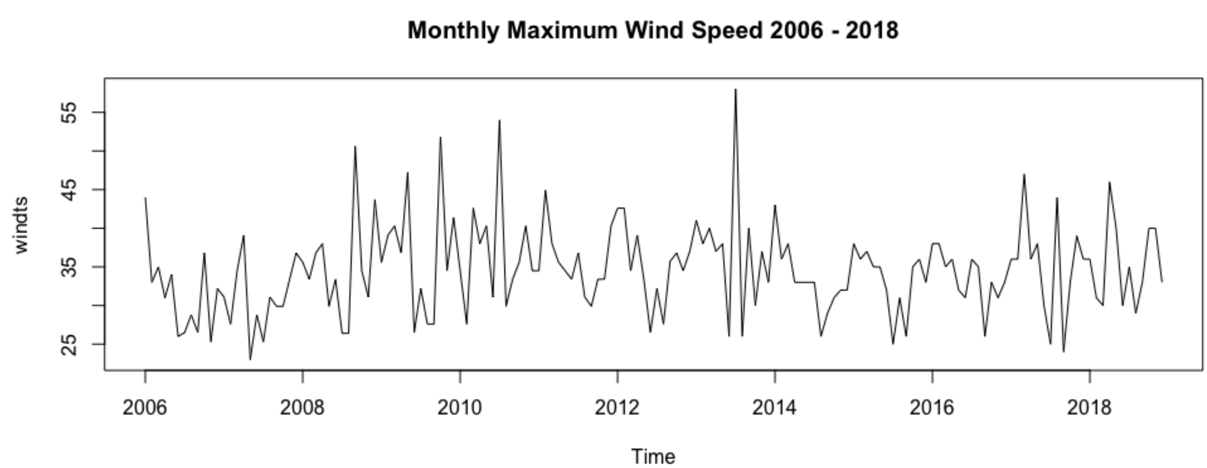
1 Introduction

The wind of New York was never friendly. On November 2017, powerful wind in New York City caused scaffolding to collapse and injured at least five people[1]. On October 2019, the wind brought down loads of trees and knocked out power to thousands on Long Island[2]. This project was interested in the powerful wind in New York. It focused on presenting a time series model of monthly maximum wind speed. It firstly studies the characteristics of this time series, including stationarity, seasonality, and decomposition. Then it fitted the data into corresponding model based on the characters. It finally adopted model to forecast monthly maximum speed in the future. The prediction on wind speed eventually can be transferred into valuable application of reusable resources.

2 Exploratory Analysis

2.1 Data Exploration

Data adopted for this time series was from a website *Weather For Central New York*[3]. Since the project was interested in the wind of New York, it focused on the maximum wind speed of each month from January 2006 to December 2018. The unit of wind speed was miles per hour (mph).



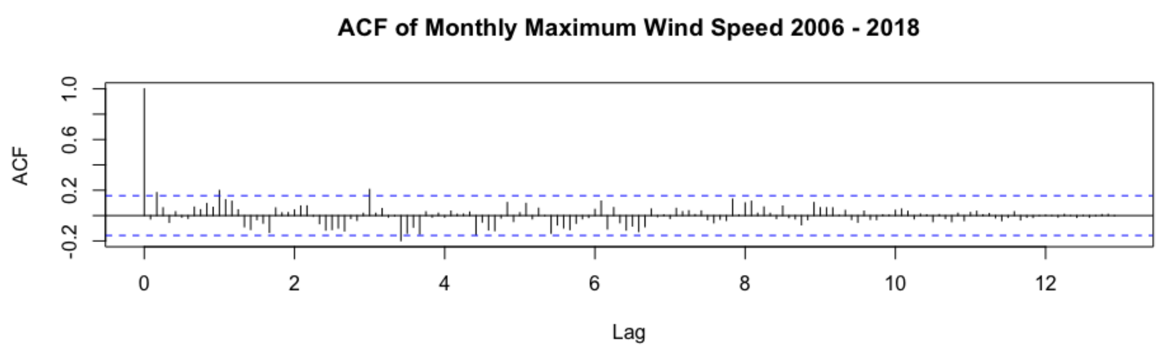
---

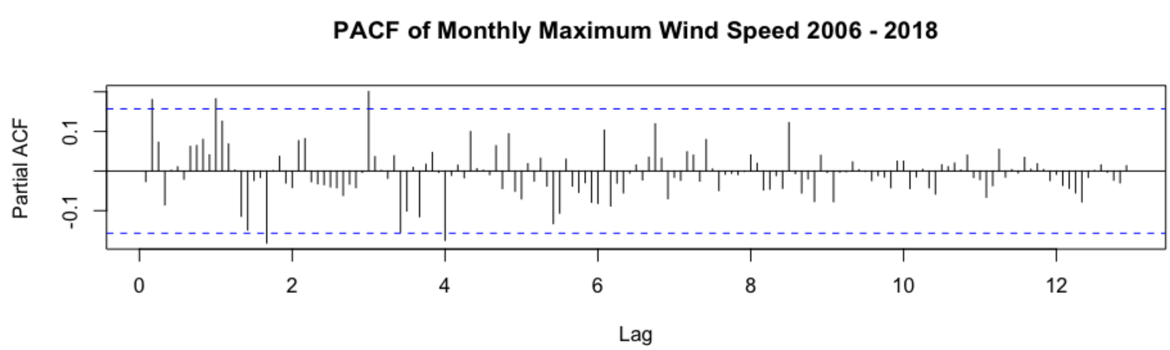
[1] https://weather.com/news/news/2017-11-19-strong-wind-nyc-northeast-impacts
[2] https://www.nbcnewyork.com/on-air/as-seen-on/Wind-Damage-Cleanup-on-Long-Island-After-Storm_New-York-563328242.html
[3] Data Source: http://www.cnyweather.com/wxwindsummary.php?

Let us look at the plot of this time series. The plot below is the time series of monthly maximum speed of 156 observations in New York. The plot illustrated that the random fluctuations were roughly constant at the speed of 35 mph, meaning this time series might be described by an additive model. It also seemed to reveal the seasonal variation, showing some peaks occasionally, but the interval between peaks were not stable.

It also witnessed the smallest maximum wind speed, approximately 25 mph, during every summer. But from January 2011 to January 2012, the plot did not present any bottom wind speed. It turned that Hurricane Irene[4] made landfall on Coney Island and Long Beach during summer 2011, as a tropical storm with winds of 70 mph. Hudson River also flooded, and five people died in the storm.

There was no clear decreasing or increasing trend in overall. Consequently, it seemed this time series was stationary, suggesting that this time series did not require a transformation. The trend and seasonality can be truly revealed by decomposition of time series later in this report.





[4] https://en.wikipedia.org/wiki/Hurricane_Irene

As for auto-correlation function (ACF) plot, which presented auto-correlation value with its lagged value, it cut off lag 3 and tailed off. As for partial auto-correlation function (PACF) plot, which presented correlation of the residuals, it also cut off lag 2 and tailed off. ACF and PACF confirmed the stationarity of maximum wind speed data. Both ACF and PACF eventually tailed off, meaning the data was supposed to fit into an auto regressive moving average model.

2.2 Stationarity

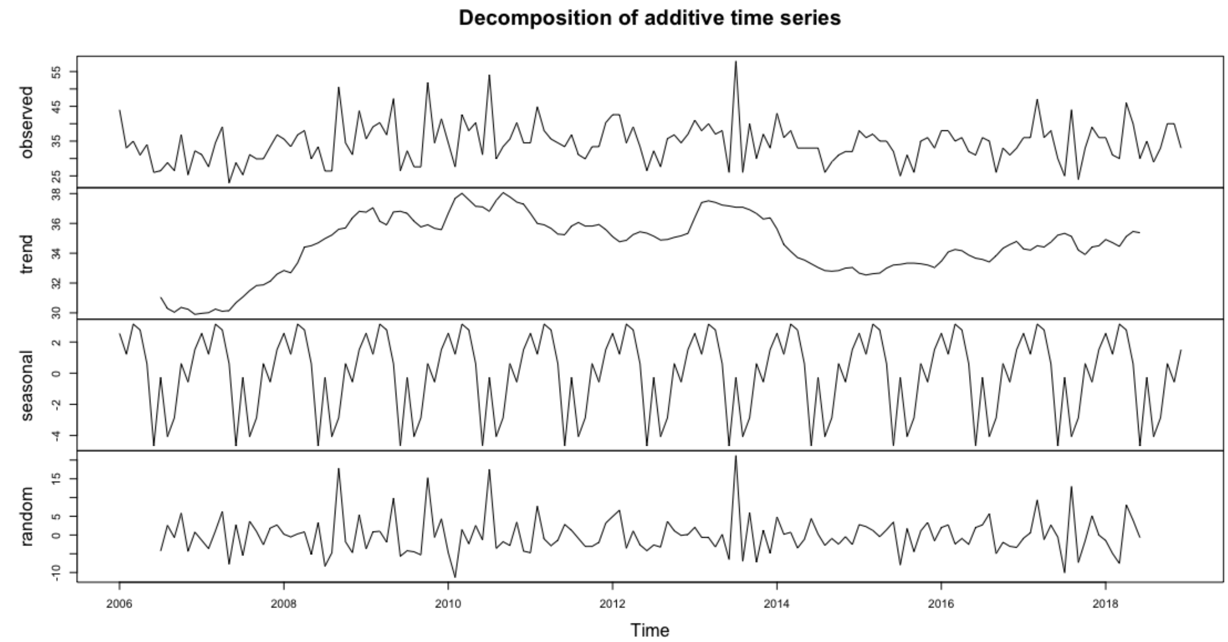
```
        Augmented Dickey-Fuller Test

data:  windts
Dickey-Fuller = -4.6199, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

To confirm the stationarity of maximum wind speed time series, this research applied augmented Dickey-Fuller test to the data. The Dickey-Fuller test was examining whether $\emptyset = 0$ in the model of $MaxSpeed_t = \alpha + \beta t + \emptyset MaxSpeed_{t-1} + e_t$. The augmented Dickey-Fuller test considered the higher-order autoregressive processes. The null hypothesis for these two tests were same: the time series were non-stationary. The output was presented above, showing that the probability that the null hypothesis was true was 0.01. Actually, in R, p-value was smaller than printed p-value. It meant that the null hypothesis was rejected, and that maximum wind speed time series was stationary.

2.3 Decomposition

Time series decomposition can deconstruct a time series into several components, which usually include a trend component, a seasonal component, and an irregular component. The plot below showed the decomposition of maximum wind speed time series.



The plot above represented the original time series (first), the estimated trend component (second), the estimated seasonal component (third), and the estimated irregular component (last) respectively. The estimated trend plot showed that the monthly maximum wind speed increased to 36 mph from January 2006 to January 2009 and remained there to January 2014, then it dropped to 32mph and followed a slightly increasing fluctuation. The estimated seasonal component plot revealed the seasonality of this time series.

2.4 Data Character Conclusion

Combing all the results revealed by data exploration on this section, the project concluded that maximum wind speed time series was a seasonal, stationary time series. Since the data was stationary originally, the project would not conduct any data transformation procedures.

3 ARIMA Modeling

3.1 The Matrix of Sample Extended ACF

The matrix of sample extended ACF is used to present possible candidate models. In this case, the matrix provide 13 candidate models. Then based on the results of matrix, the corresponding Akaike information criterion (AIC) were also presented below.

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

Model	AIC	Model	AIC
ARMA(0,1)	1000.23	**ARMA(4,2)**	997.99
ARMA(0,11)	1008.1	**ARMA(5,0)**	1000.96
ARMA(1,1)	1001.36	**ARMA(6,0)**	1002.93
ARMA(2,0)	996.98	**ARMA(6,1)**	1004.62
ARMA(2,1)	998.55	**ARMA(7,0)**	1004.85
ARMA(3,0)	998.13	**ARMA(7,1)**	1006.56
ARMA(3,1)	999.49		

The yellow highlighted model, ARMA(2,0) had the smallest AIC value. This model was temporally considered as the best fit. Then the research tried other methods to explore the possible candidate models.

3.2 Auto ARIMA

Based on the previous research of maximum wind speed time series data, the study assumed that the data could fit into Autoregressive moving-average (ARMA) model. Autoregressive integrated moving average (ARIMA) model can be also applied here. If an ARIMA model has same MA and AR order of an ARMA model and no differencing, then this ARIMA model is equivalent to an ARMA model.

The data also satisfied two assumptions that it should be stationary, and that it should be univariate. Also, data showed the seasonality, therefore, the maximum wind speed data should be fitted into seasonal ARIMA model.

The research fitted the data by using auto ARIMA function in R, based on Akaike information criterion (AIC), and received the result below.

```
Series: windts
ARIMA(1,0,0)(1,0,1)[12] with non-zero mean

Coefficients:
          ar1    sar1     sma1     mean
      -0.0788  0.7745  -0.6111  34.4144
s.e.   0.0819  0.3251   0.4179   0.6482

sigma^2 estimated as 33.46:  log likelihood=-493.75
AIC=997.5   AICc=997.9   BIC=1012.75
```

```
z test of coefficients:

           Estimate Std. Error z value  Pr(>|z|)
ar1       -0.077196   0.082746 -0.9329  0.350855
sar1      -0.452591   0.171509 -2.6389  0.008318 **
sma1       0.723301   0.140038  5.1650 2.404e-07 ***
intercept 34.548629   0.494179 69.9112 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The output shown above was a seasonal ARIMA model classified as SARIMA(p,d,q)(P,D,Q)$_m$. (p,d,q) presented the non-seasonal part of the model. (P,D,Q) presented the seasonal component. In this case, the model was SARIMA(1,0,0)(1,0,1)$_{12}$. The whole seasonal period was 12. ARIMA(1,0,0) was actually AR(1). The z-test of coefficients confirmed the significance of each parameters in the model.

```
Series: windts
ARIMA(0,0,0)(1,0,0)[12] with non-zero mean

Coefficients:
        sar1     mean
      0.2116  34.5256
s.e.  0.0805   0.5737

sigma^2 estimated as 33.62:  log likelihood=-494.8
AIC=995.61   AICc=995.77   BIC=1004.76
```

```
z test of coefficients:

           Estimate Std. Error z value  Pr(>|z|)
sar1       0.211560   0.080474  2.6289  0.008566 **
intercept 34.525512   0.573699 60.1805 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

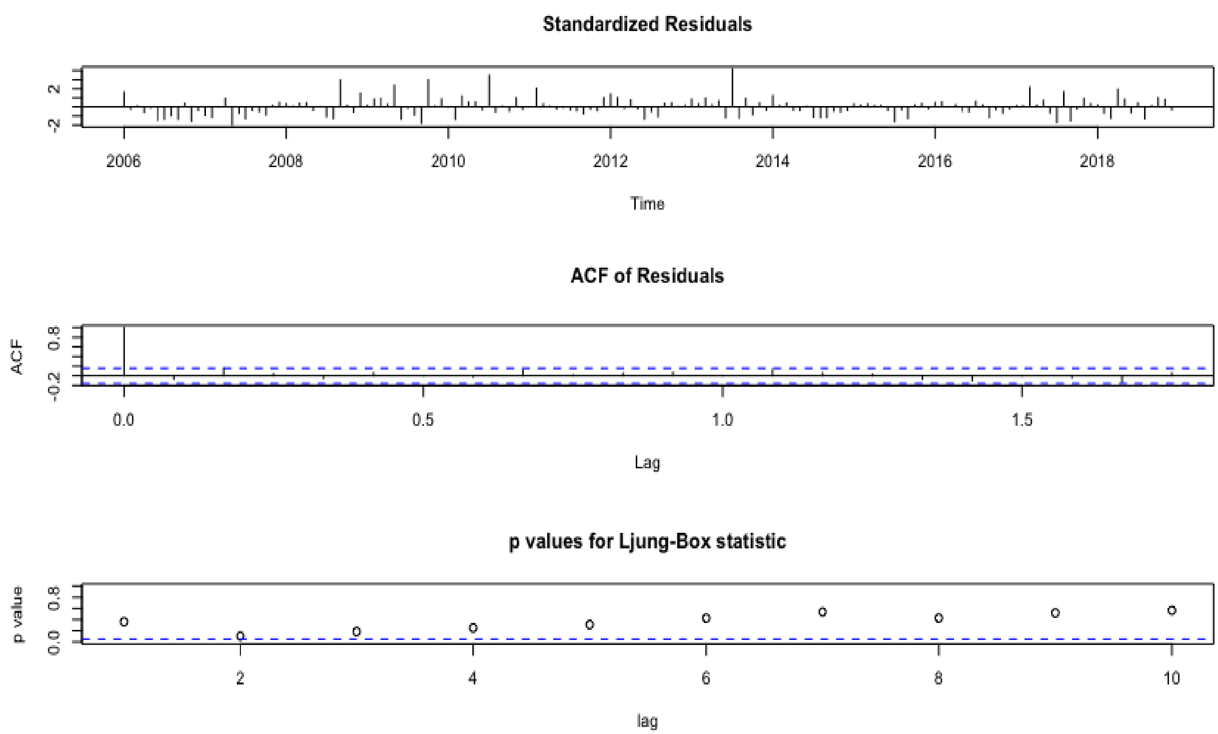
This research also applied the auto ARIMA function based on Bayesian information criterion (BIC) and found a different model SARIMA(0,0,0)(1,0,0)$_{12}$. ARIMA(0,0,0) was actually a white noise model. The z-test of coefficients also confirmed the significance of parameter in this model.

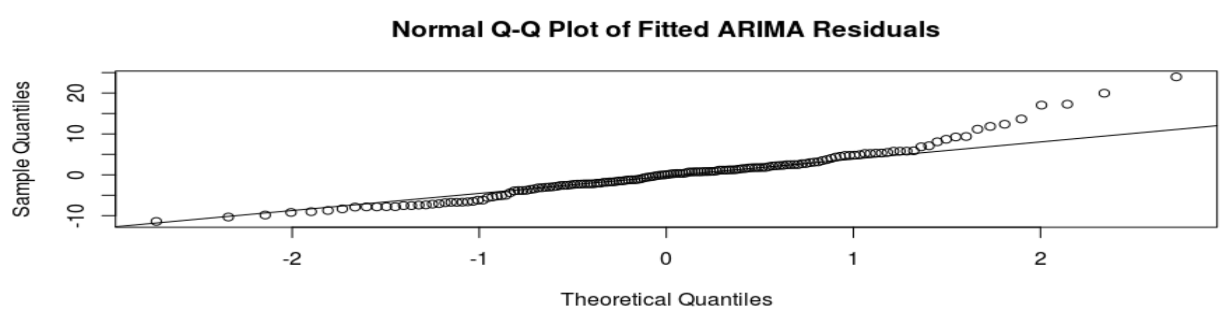
R used maximum likelihood estimation to find the best fit ARIMA model. Then this report considered the model with lowest AIC to be the best fit. These two models are summarized into the chart below.

Model	AICc	AIC	BIC
ARIMA(1,0,0)(1,0,1)$_{12}$	997.9	997.5	1012.75
ARIMA(0,0,0)(1,0,0)$_{12}$	995.61	995.77	1004.76

The AIC values of these two models were really close to each other. ARIMA(0,0,0)(1,0,0)$_{12}$ had a slightly smaller AIC. It also can be concluded that AIC value and its corresponding AICc value were close to each other. This is due to the large number of observations. When the number of observations was large enough, AICc converted to AIC. For the consistency, this research would continually adopt AICc as criterion. In this case, ARIMA(0,0,0)(1,0,0)$_{12}$ would be temporally considered as the best fit model.

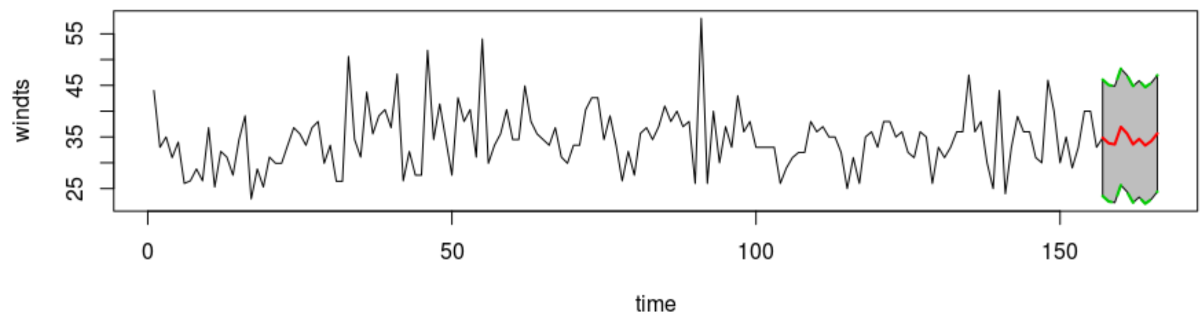
3.3 Diagnostic Measurements





The diagnostic of this model was mainly based on the ACF plot, Ljung-Box statistics p-value plot and Q-Q plot. ACF of the residuals did not show any significant autocorrelations. P-values of Ljung-Box statistics were also smaller than 0.05. As for the q-q plot, most values were normal as they rest on a line and were not all over the place. All the informative above assumed that there was no pattern in the residuals, and that this model cam be used to calculate the forecast.

4 Forecasting




```
Forecast for univariate time series:
    Lead Forecast  S.E Lower Upper
157    1     34.8 5.76  23.5  46.1
158    2     33.8 5.76  22.5  45.1
159    3     33.6 5.76  22.3  44.9
160    4     37.0 5.76  25.7  48.2
161    5     35.7 5.76  24.4  47.0
162    6     33.6 5.76  22.3  44.9
163    7     34.6 5.76  23.3  45.9
164    8     33.4 5.76  22.1  44.6
165    9     34.2 5.76  22.9  45.5
166   10     35.7 5.76  24.4  47.0
```

The ARIMA(0,0,0)(1,0,0)$_{12}$ was adopted in this section to predict the maximum wind speed from January 2019 to October 2019. The forecasted results were presented in both time series plot above and the table on the left. The forecasts were shown in a red line, and the 95% prediction intervals as shaded area.

Since October 2019 passed already, the results can be compared with the actual records. The table below showed the monthly maximum wind speed from January 2019 to October 2019 and its corresponding prediction from the best fitted model. The green highlighted month were the actual maximum wind speed records that fall into the 95% prediction interval. The model seemed to work well.

Maximum Wind Speed	Jan	**Feb**	**Mar**	**April**	**May**	**Jun**	**Jul**	**Aug**	**Sept**	**Oct**
Forecast	34.8	33.8	33.6	37.0	35.7	33.6	34.6	33.4	34.2	35.7
Reality	49.0	45.0	36.0	38.0	41.0	32.0	40.0	33.0	26.0	44.0

5 Conclusion

This project provided a presentation of a time series model of monthly maximum wind speed in New York, and then fitted the data into to a seasonal auto regressive moving average model, and then used this time series model to predict the maximum monthly wind speed in the following months. We can know from the report that New York underwent powerful wind entire year. Although people suffered from the wind damage, strong wind can bring economic benefits. There were 13 wind farms in Upstate New York, which have massive windmills. Wind power, as a sustainable and renewable energy, has relatively smaller impact on the environment. This type of energy should be high promoted.

It was lucky to have originally stationary data without any transformation, which eased the workloads. But I would like to experience some possibility to do necessary transformation, such as log transformation or seasonality differencing. Hope data for next project require these procedures. From this project, I also found that my understanding of ACF and PACF was not as deep, or accurate as desired. Knowledge components related to ACF and PACF needs to be reviewed.

Appendix (R Code)

```
#prepare

install.packages("ISLR")

install.packages("astsa ")

install.packages("forecast ")

install.packages("aTSA")

install.packages("xts")

install.packages('forecast', dependencies = TRUE)

install.packages("tseries")

install.packages("FitAR")

install.packages("TSA")

library(FitAR)

library(ISLR)

library(astsa)

library(forecast)

library(aTSA)

library(xts)

library(tseries)

library(TSA)


#importing and exploring

wind=read.csv("/Users/AmberZhao/Desktop/wind.csv")

head(wind)

dim(wind)

windts=ts(wind[,2],frequency=12,start=c(2006,1))

windts

plot.ts(windts, main="Monthly Maximum Wind Speed 2006 - 2018", ylab = "Wind Speed
mph", )

acf(windts, main="ACF of Monthly Maximum Wind Speed 2006 - 2018",lag=length(windts)-1,)
```

```
pacf(windts,main="PACF of Monthly Maximum Wind Speed 2006 - 2018",,lag=length(windts)-
1,)
acf(windts, main="ACF of Monthly Maximum Wind Speed 2006 - 2018, Lag =20",lag.max=20)
pacf(windts,main="PACF of Monthly Maximum Wind Speed 2006 - 2018, Lag = 20",lag.max=20)

#decomposition
windtscomponents = decompose(windts)
windtscomponents$seasonal
plot(windtscomponents)

#stationary test
adf_windts = adf.test(windts,alternative="stationary")
adf_windts

#seasonal adjustment
windts_seasonadjusted = windts - windtscomponents$seasonal
plot(windts_seasonadjusted, main="Seasonal Adjusted Monthly Maximum Wind Speed 2006 -
2018",
    ylab = "Wind Speed mph")

acf(windts_seasonadjusted, main="ACF of Seasonal Adjusted Monthly Maximum Wind Speed
2006 - 2018",lag=length(windts)-1,)
pacf(windts_seasonadjusted,main="PACF of Seasonal Adjusted Monthly Maximum Wind Speed
2006 - 2018",,lag=length(windts)-1,)

adf_windts_adjusted = adf.test(windts_seasonadjusted,alternative="stationary")
adf_windts_adjusted

#model fitting
```

```
eacf(windts)

BoxCox.lambda(windts)

auto.arima(windts,d=0)

#eacf model check

#diagnostic measurements
fitarima1=arima(windts,order=c(1,0,0),seasonal = list(order=c(1,0,1), period=12),method='ML')
library(lmtest)
fitarima1
coeftest(fitarima1)
checkresiduals(fitarima1)
tsdiag(fitarima1)

sarima(fitarima1,order=c(1,0,0),seasonal = list(order=c(1,0,1), period=12))

fitarima2=arima(windts,order=c(0,0,0),seasonal = list(order=c(1,0,0), period=12),method='ML')
coeftest(fitarima2)
acf(fitarima2$residuals, main="ACF of Fitted ARIMA Residuals")
qqnorm(fitarima2$residuals, main="Normal Q-Q Plot of Fitted ARIMA Residuals")
qqline(fitarima2$residuals)
tsdiag(fitarima2)
```

```
#forecasting
windfuture = forecast(fitarima1,10)
plot(forecast(fitarima1,10))

ARMA001=arima(windts,order=c(0,0,1),method='ML')
ARMA0011=arima(windts,order=c(0,0,11),method='ML')
ARMA101=arima(windts,order=c(1,0,1),method='ML')
ARMA200=arima(windts,order=c(2,0,0),method='ML')
ARMA201=arima(windts,order=c(2,0,1),method='ML')
ARMA300=arima(windts,order=c(3,0,0),method='ML')
ARMA301=arima(windts,order=c(3,0,1),method='ML')
ARMA402=arima(windts,order=c(4,0,2),method='ML')
ARMA500=arima(windts,order=c(5,0,0),method='ML')
ARMA600=arima(windts,order=c(6,0,0),method='ML')
ARMA601=arima(windts,order=c(6,0,1),method='ML')
ARMA700=arima(windts,order=c(7,0,0),method='ML')
ARMA701=arima(windts,order=c(7,0,1),method='ML')
```