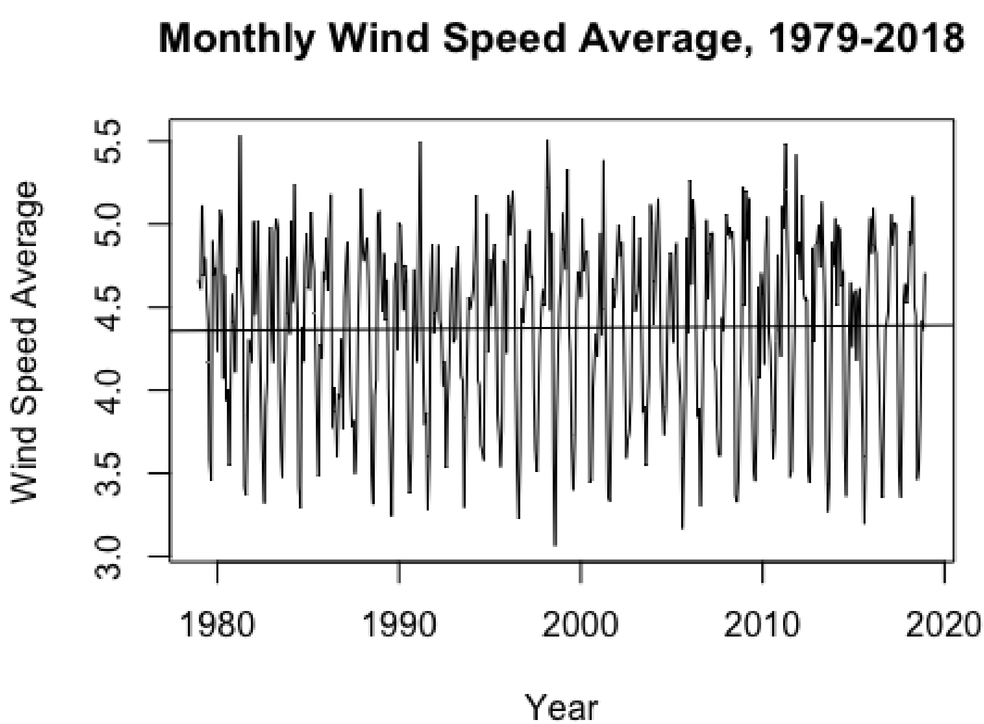
STAT GU4221/GR5221 Project: Forecasting Spatial-Temporal Climate Data
Christopher Jason Halim

I. Introduction

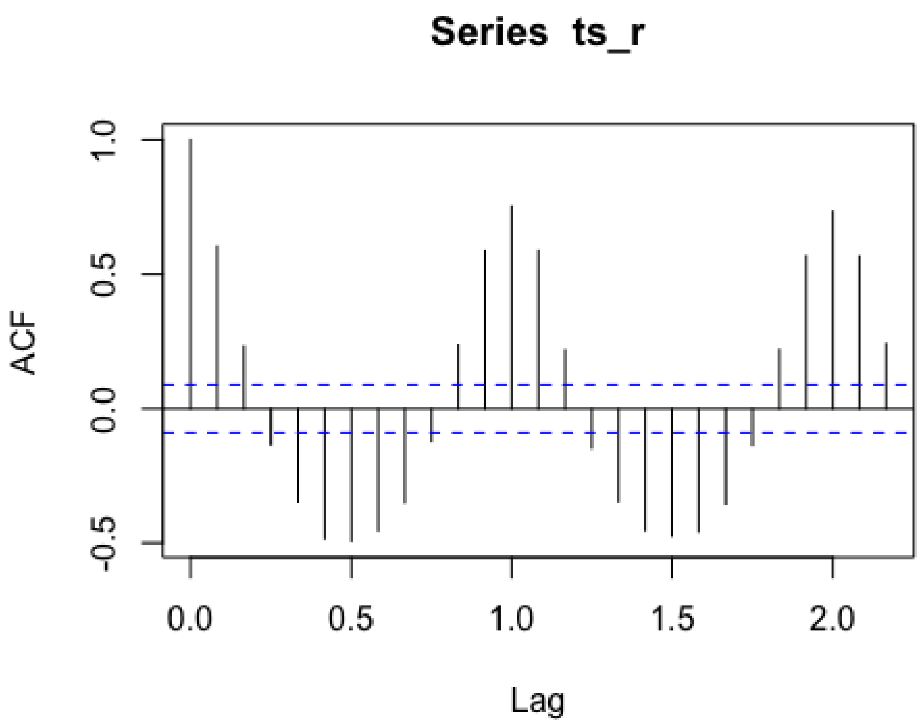
The goal of the project with the data set of interest (*WindSpeed_Month_Ave.csv*) is to forecast the average of windspeed months ahead by using techniques from Time Series Analysis. In this project, I calculated the average monthly windspeed based on the windspeed timestamp from 918 different locations for any given month. Based on that average monthly windspeed, we can then use time series models to predict future windspeed for any given region. Since we have the average windspeed for 918 different locations, we can then take this average to forecast average windspeed in the future. Here is the overview of the modified dataset:

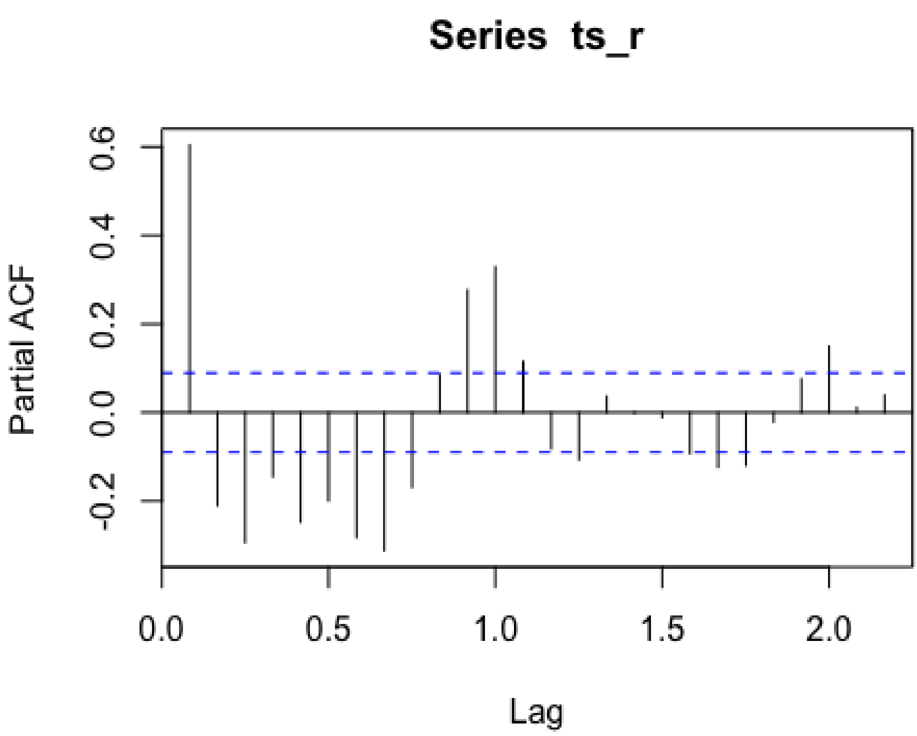
```
##       date      avg
## 1 Jan 1979 4.665798
## 2 Feb 1979 4.608589
## 3 Mar 1979 5.113349
## 4 Apr 1979 4.692132
## 5 May 1979 4.794925
## 6 Jun 1979 4.582700
```

After modifying the data, I then converted the data frame to a time series object. Here is a brief analysis of the converted time series dataset along with a graph to illustrate the data:



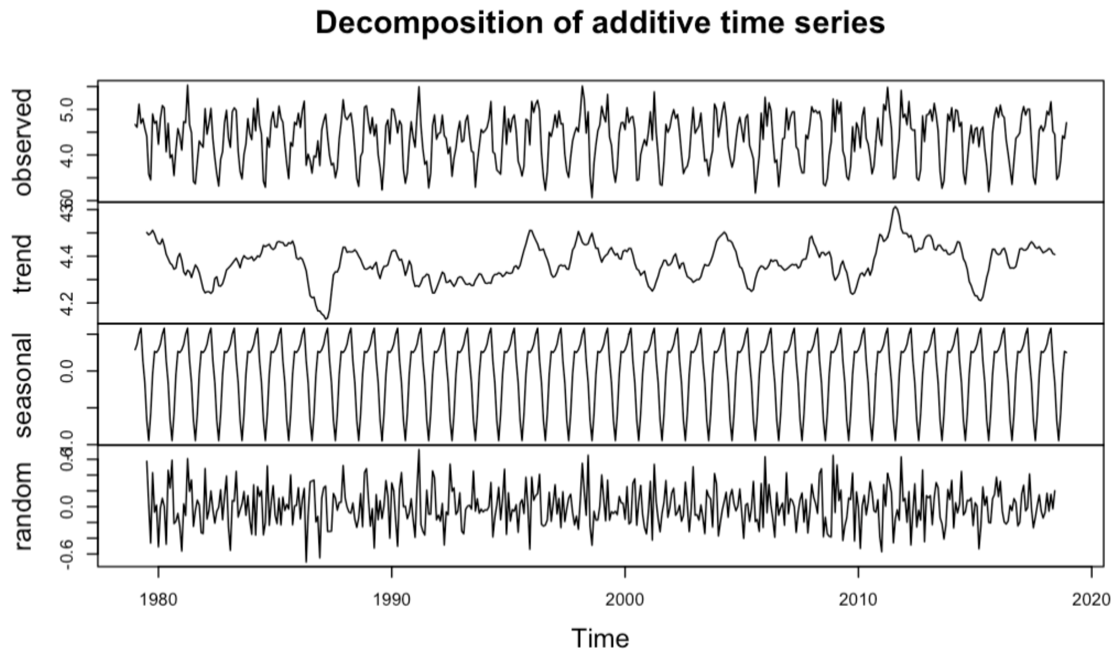
I am also considering the autocorrelation function (ACF) and the partial autocorrelation function (PACF) graphs for exploratory data analysis:





If we take a look at the ACF and PACF plot above, we can see the seasonality and non-seasonality trend, which is fitting with the model that we are going to use in the project that will be discussed in the next part.

I also do the decomposition of the time series by using decomposition function along with the component of the time series data:



II. Statistical Models

For the purpose of this project, I will consider the seasonal ARIMA model. I am using the auto.arima function from the forecast library. This enables the function to search for the best fit of the seasonal ARIMA model. As we are using the seasonal ARIMA model for this data, we should make an assumption that the data is non-stationary.

In addition, we also should assume that the data should be univariate because ARIMA works on a single variable. Auto-regression works with past values, in this case the past values of the average windspeed from 918 sites.

A seasonal ARIMA model is formed by including additional seasonal terms in the regular ARIMA models, where it is written as: ARIMA (p,d,q) $(P,D,Q)_m$, where (p,d,q) is the non-seasonal part of the model and $(P,D,Q)_m$ is the seasonal part of the model. M is equal to the number of observations per year, which in our case is twelve since we are observing the average of windspeed per month.

In seasonal ARIMA, the difference between the seasonal and non-seasonal components of the model is the non-seasonal components of the model utilize the backshift operator from the seasonal period. The additional seasonal terms are then multiplied by the non-seasonal terms.

III. Results

For this seasonal ARIMA model, I found that in order to get the best model, the non-seasonal order of the model is (p,d,q = 1,0,0) and the seasonal order of the model is (P,D,Q = 1,1,0) with frequency(m) equals to 12. Here is a brief results of the seasonal ARIMA model with auto.arima function:

```
## Series: ts_r
## ARIMA(1,0,0)(1,1,0)[12] with drift
##
## Coefficients:
##          ar1      sar1   drift
##       0.0412   -0.4976  -1e-04
## s.e.  0.0463    0.0401   9e-04
##
## sigma^2 = 0.1011:  log likelihood = -127.91
## AIC=263.81   AICc=263.9   BIC=280.41
##
## Training set error measures:
##                        ME      RMSE       MAE        MPE     MAPE      MASE
## Training set -0.001111316 0.3128792 0.2421642 -0.3613827 5.534605 0.8400243
##                     ACF1
## Training set 0.001935232
```

From the result above, we can observe that the RMSE value is **0.3128792**, which is quite good for the model as it can predict the data relatively accurate. In addition, if we take a look at the MAPE value, which consider actual values fed into the model and fitted values from the model, the value of **5.5346%** means that this model is about 95% accurate.

Here is the result of the forecast in the next twelve months from the ARIMA model along with estimated standard errors:

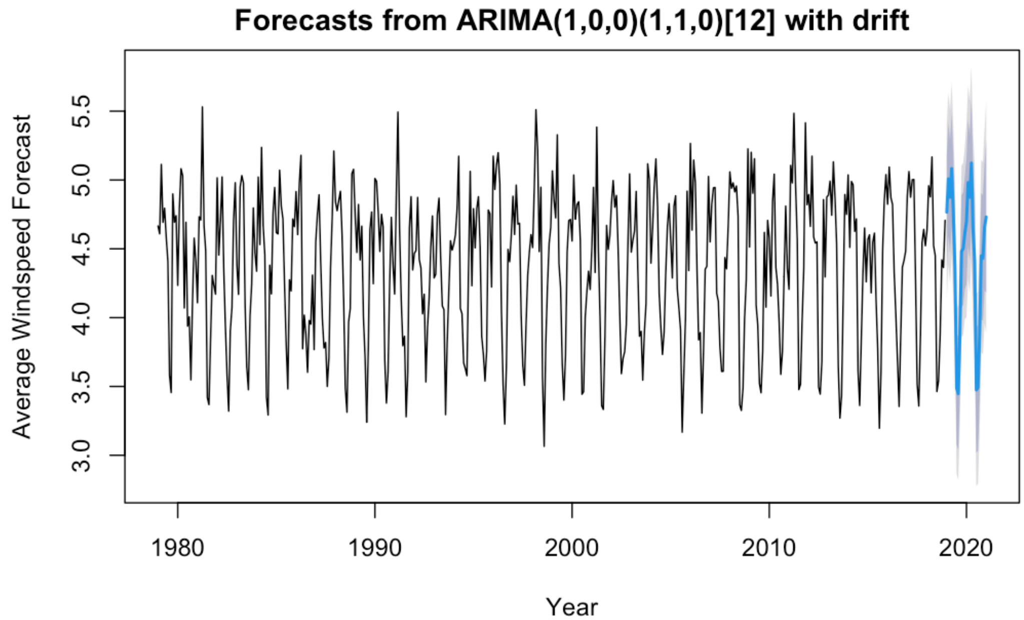
```
## $pred
##           Jan      Feb      Mar      Apr      May      Jun      Jul      Aug
## 2019 4.767738 5.009782 4.875964 5.084834 4.761593 4.366283 3.489992 3.449068
##           Sep      Oct      Nov      Dec
## 2019 3.873284 4.483225 4.501888 4.614804

## $se
##            Jan       Feb       Mar       Apr       May       Jun       Jul
## 2019 0.3168678 0.3171371 0.3171375 0.3171375 0.3171375 0.3171375 0.3171375
##            Aug       Sep       Oct       Nov       Dec
## 2019 0.3171375 0.3171375 0.3171375 0.3171375 0.3171375
```

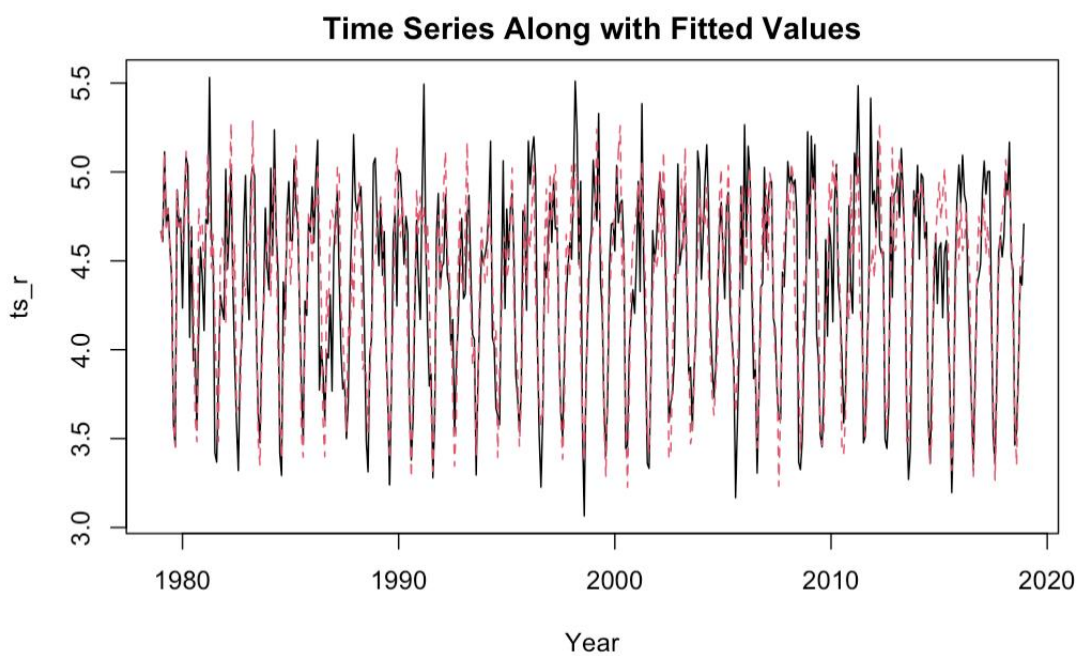
I also calculated the residual sum of squares of the seasonal ARIMA model:

```
##      ARIMA
## 1 1.569876
```

Here is the time series plot of the next twelve months forecast from the ARIMA model:



Here is the time series plot along with the fitted values:



Here is the fitted vs. residual values plot: