**University Of Connecticut**

**OPIM 5671 – Data Mining and Business Intelligence**

Time Series Forecasting Project Proposal

# RainFall Prediction in Mumbai

By

Group 5

Lahari Maddula

Pradeepti Dokka

Sai Deepika Bandari

Sanchita Godse

Shuang Ma

**Table of Contents**

1. Executive Summary
2. Introduction

   2.1 Problem Statement

   2.2 Data Description

3. Data Exploration
4. Time Series Exploration

   3.2. Series Value Graph

   3.3. Seasonal Cycle Graph

   3.4. Seasonal Component Graph

   3.5. Cross Correlation Graphs

   3.6. Trend Component Graph

   3.7. Augmented Dickey-Fuller Unit Root Test

5. Data Models & Forecasting
6. Model Comparison

   5.1. AIC and SBC Values

   5.2. AIC and SBC Comparison for All Models

7. Findings and Recommendations
8. Conclusion
9. References
10. Appendix

## 1. Executive Summary:

In today's era of climate change and sustainable development, the renewable energy sector plays a pivotal role in mitigating environmental impacts and ensuring a sustainable future. Accurate weather forecasting, particularly precipitation forecasting, is crucial for optimizing renewable energy generation, agricultural practices, and building energy efficiency. The Prediction of Worldwide Energy Resources (POWER) project, which utilizes NASA's meteorological data, offers a unique opportunity to enhance our forecasting capabilities. This business case explores the potential benefits of leveraging POWER data for precipitation forecasting in the city of Mumbai from 2000 to 2020.

## 2. Introduction:

The POWER project, supported by NASA, provides valuable meteorological data for renewable energy planning, building efficiency, and agriculture. One of the key variables in this dataset is precipitation, which plays a significant role in multiple sectors. Accurate precipitation forecasting is essential for:

Renewable Energy: Precipitation forecasts can help energy operators anticipate changes in solar and wind power generation due to weather conditions, enabling efficient energy resource management.

Precipitation: forecasts assist in optimizing heating and cooling systems in buildings, reducing energy consumption and costs.

Agriculture: Farmers can use precipitation forecasts to plan irrigation and planting schedules, improving crop yields and resource utilization.

### Problem Statement:

Mumbai, a densely populated coastal city, experiences significant rainfall variations throughout the year. Accurate precipitation forecasting is vital to ensure sustainable energy generation, building energy efficiency, and agricultural productivity. However, existing forecasting models may not capture the nuances of Mumbai's unique climate.

## 3. Data Exploration:

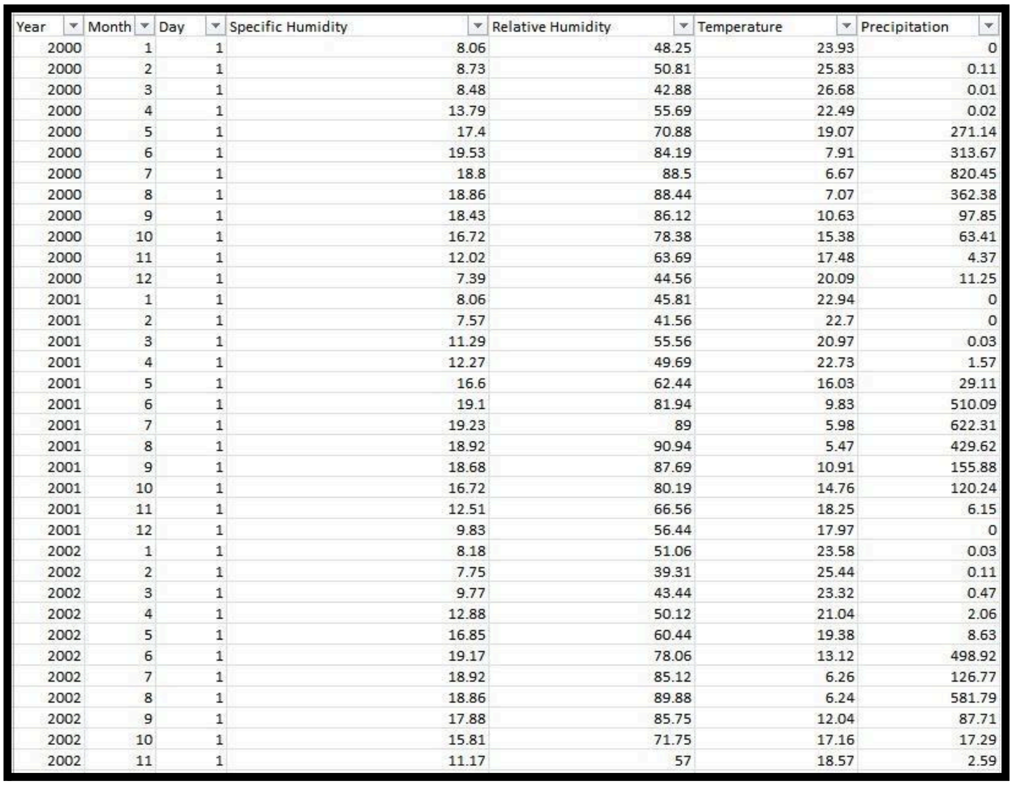
### Data Description:

The Data Set used is sourced from the Kaggle. It has information on the duration of rainfall for the first day of every month from 2000 to 2020. It has 252 observations.

The dataset explored in this report includes the following meteorological variables: specific humidity, relative humidity, temperature, and precipitation.

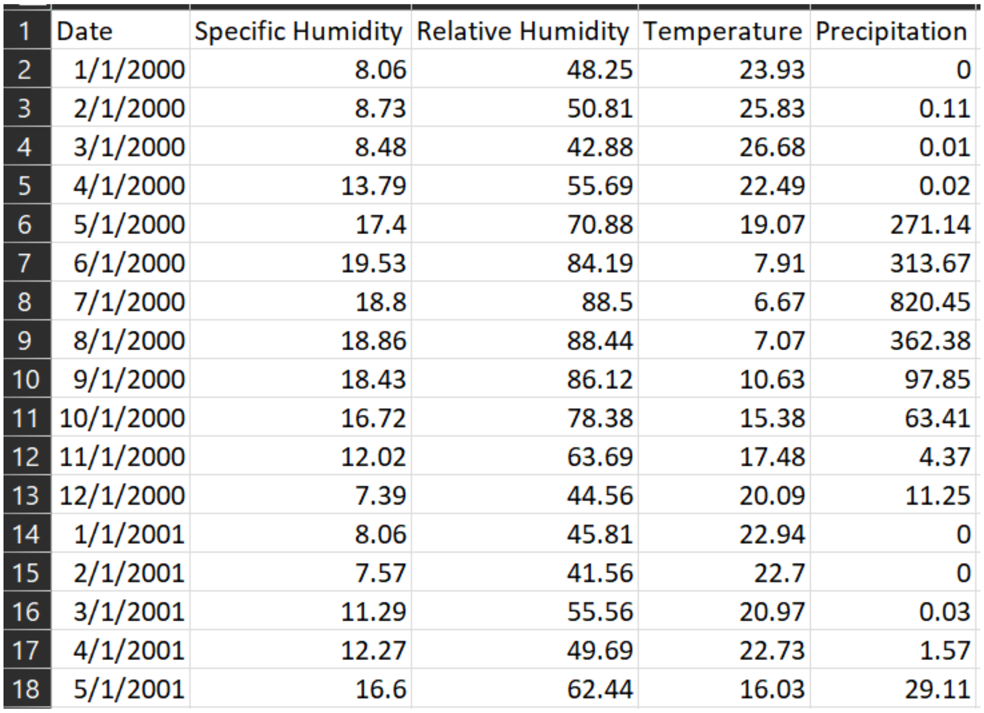
Meteorological data plays a pivotal role in understanding weather patterns and their impact on various sectors, including renewable energy, agriculture, and building management. The dataset under examination contains a diverse array of meteorological variables, each contributing unique insights into our analysis. In this section, we will delve into the specifics of these variables and the steps taken to prepare the dataset for comprehensive analysis.

Snapshot of the dataset before cleaning:



| Year | Month | Day | Specific Humidity | Relative Humidity | Temperature | Precipitation |
|---|---|---|---|---|---|---|
| 2000 | 1 | 1 | 8.06 | 48.25 | 23.93 | 0 |
| 2000 | 2 | 1 | 8.73 | 50.81 | 25.83 | 0.11 |
| 2000 | 3 | 1 | 8.48 | 42.88 | 26.68 | 0.01 |
| 2000 | 4 | 1 | 13.79 | 55.69 | 22.49 | 0.02 |
| 2000 | 5 | 1 | 17.4 | 70.88 | 19.07 | 271.14 |
| 2000 | 6 | 1 | 19.53 | 84.19 | 7.91 | 313.67 |
| 2000 | 7 | 1 | 18.8 | 88.5 | 6.67 | 820.45 |
| 2000 | 8 | 1 | 18.86 | 88.44 | 7.07 | 362.38 |
| 2000 | 9 | 1 | 18.43 | 86.12 | 10.63 | 97.85 |
| 2000 | 10 | 1 | 16.72 | 78.38 | 15.38 | 63.41 |
| 2000 | 11 | 1 | 12.02 | 63.69 | 17.48 | 4.37 |
| 2000 | 12 | 1 | 7.39 | 44.56 | 20.09 | 11.25 |
| 2001 | 1 | 1 | 8.06 | 45.81 | 22.94 | 0 |
| 2001 | 2 | 1 | 7.57 | 41.56 | 22.7 | 0 |
| 2001 | 3 | 1 | 11.29 | 55.56 | 20.97 | 0.03 |
| 2001 | 4 | 1 | 12.27 | 49.69 | 22.73 | 1.57 |
| 2001 | 5 | 1 | 16.6 | 62.44 | 16.03 | 29.11 |
| 2001 | 6 | 1 | 19.1 | 81.94 | 9.83 | 510.09 |
| 2001 | 7 | 1 | 19.23 | 89 | 5.98 | 622.31 |
| 2001 | 8 | 1 | 18.92 | 90.94 | 5.47 | 429.62 |
| 2001 | 9 | 1 | 18.68 | 87.69 | 10.91 | 155.88 |
| 2001 | 10 | 1 | 16.72 | 80.19 | 14.76 | 120.24 |
| 2001 | 11 | 1 | 12.51 | 66.56 | 18.25 | 6.15 |
| 2001 | 12 | 1 | 9.83 | 56.44 | 17.97 | 0 |
| 2002 | 1 | 1 | 8.18 | 51.06 | 23.58 | 0.03 |
| 2002 | 2 | 1 | 7.75 | 39.31 | 25.44 | 0.11 |
| 2002 | 3 | 1 | 9.77 | 43.44 | 23.32 | 0.47 |
| 2002 | 4 | 1 | 12.88 | 50.12 | 21.04 | 2.06 |
| 2002 | 5 | 1 | 16.85 | 60.44 | 19.38 | 8.63 |
| 2002 | 6 | 1 | 19.17 | 78.06 | 13.12 | 498.92 |
| 2002 | 7 | 1 | 18.92 | 85.12 | 6.26 | 126.77 |
| 2002 | 8 | 1 | 18.86 | 89.88 | 6.24 | 581.79 |
| 2002 | 9 | 1 | 17.88 | 85.75 | 12.04 | 87.71 |
| 2002 | 10 | 1 | 15.81 | 71.75 | 17.16 | 17.29 |
| 2002 | 11 | 1 | 11.17 | 57 | 18.57 | 2.59 |

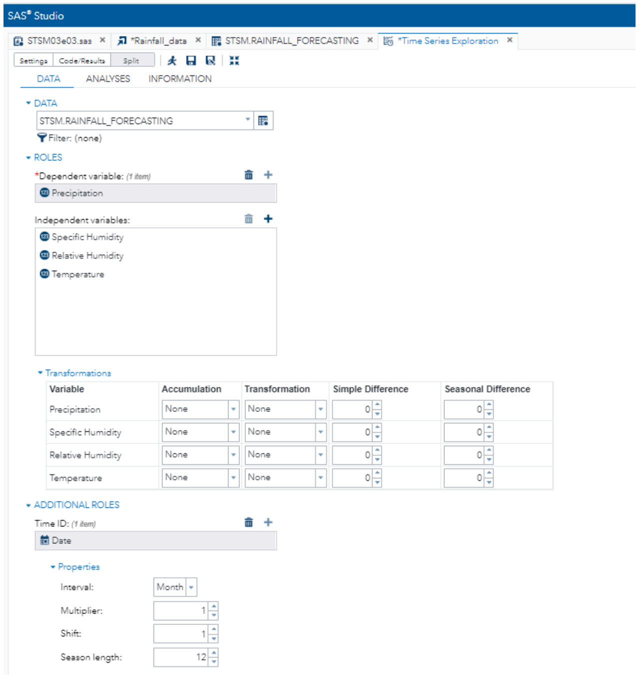
Cleaned Dataset:



| 1 | Date | Specific Humidity | Relative Humidity | Temperature | Precipitation |
|---|---|---|---|---|---|
| 2 | 1/1/2000 | 8.06 | 48.25 | 23.93 | 0 |
| 3 | 2/1/2000 | 8.73 | 50.81 | 25.83 | 0.11 |
| 4 | 3/1/2000 | 8.48 | 42.88 | 26.68 | 0.01 |
| 5 | 4/1/2000 | 13.79 | 55.69 | 22.49 | 0.02 |
| 6 | 5/1/2000 | 17.4 | 70.88 | 19.07 | 271.14 |
| 7 | 6/1/2000 | 19.53 | 84.19 | 7.91 | 313.67 |
| 8 | 7/1/2000 | 18.8 | 88.5 | 6.67 | 820.45 |
| 9 | 8/1/2000 | 18.86 | 88.44 | 7.07 | 362.38 |
| 10 | 9/1/2000 | 18.43 | 86.12 | 10.63 | 97.85 |
| 11 | 10/1/2000 | 16.72 | 78.38 | 15.38 | 63.41 |
| 12 | 11/1/2000 | 12.02 | 63.69 | 17.48 | 4.37 |
| 13 | 12/1/2000 | 7.39 | 44.56 | 20.09 | 11.25 |
| 14 | 1/1/2001 | 8.06 | 45.81 | 22.94 | 0 |
| 15 | 2/1/2001 | 7.57 | 41.56 | 22.7 | 0 |
| 16 | 3/1/2001 | 11.29 | 55.56 | 20.97 | 0.03 |
| 17 | 4/1/2001 | 12.27 | 49.69 | 22.73 | 1.57 |
| 18 | 5/1/2001 | 16.6 | 62.44 | 16.03 | 29.11 |

## 4. Time Series Exploration:

### Input:



DATA

STSM.RAINFALL_FORECASTING

Filter: (none)

ROLES

*Dependent variable: *(1 item)*

Precipitation

Independent variables:

Specific Humidity
Relative Humidity
Temperature

Transformations

| Variable | Accumulation | Transformation | Simple Difference | Seasonal Difference |
|---|---|---|---|---|
| Precipitation | None | None | 0 | 0 |
| Specific Humidity | None | None | 0 | 0 |
| Relative Humidity | None | None | 0 | 0 |
| Temperature | None | None | 0 | 0 |

ADDITIONAL ROLES

Time ID: *(1 item)*

Date

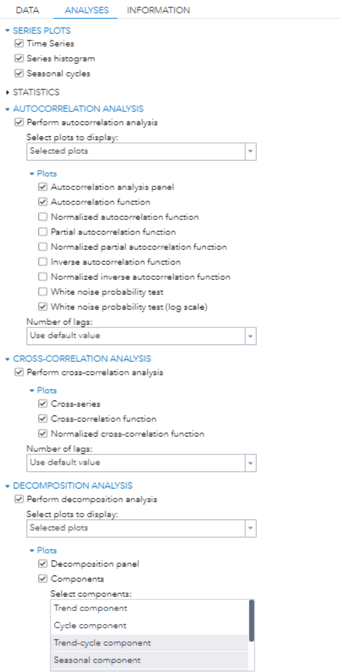
Properties

Interval: Month

Multiplier: 1

Shift: 1

Season length: 12



SERIES PLOTS

☑ Time Series
☑ Series histogram
☑ Seasonal cycles

STATISTICS

AUTOCORRELATION ANALYSIS

☑ Perform autocorrelation analysis

Select plots to display: Selected plots

Plots

☑ Autocorrelation analysis panel
☑ Autocorrelation function
☐ Normalized autocorrelation function
☐ Partial autocorrelation function
☐ Normalized partial autocorrelation function
☐ Inverse autocorrelation function
☐ Normalized inverse autocorrelation function
☐ White noise probability test
☑ White noise probability test (log scale)

Number of lags: Use default value

CROSS-CORRELATION ANALYSIS

☑ Perform cross-correlation analysis

Plots

☑ Cross-series
☑ Cross-correlation function
☑ Normalized cross-correlation function

Number of lags: Use default value

DECOMPOSITION ANALYSIS

☑ Perform decomposition analysis

Select plots to display: Selected plots

Plots

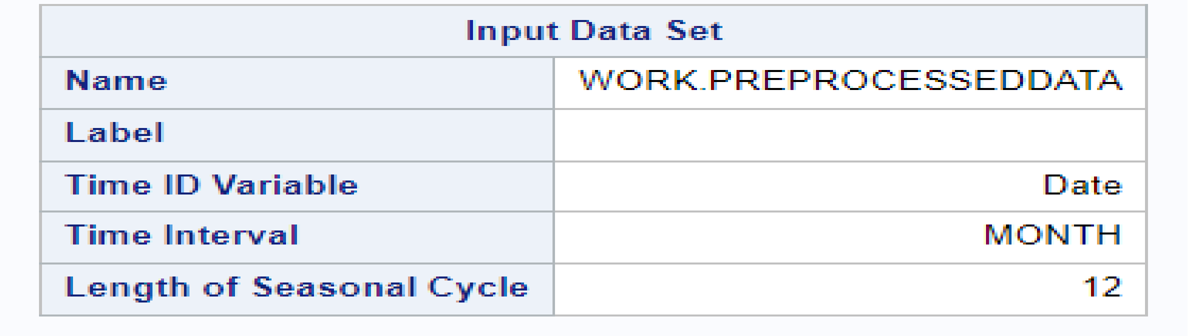
☑ Decomposition panel
☑ Components

Select components:

Trend component
Cycle component
Trend-cycle component
Seasonal component

Precipitation is used as the dependent variable for the time series analysis. The Time Id is Date and the independent variables are Specific Humidity, Relative Humidity, Temperature.

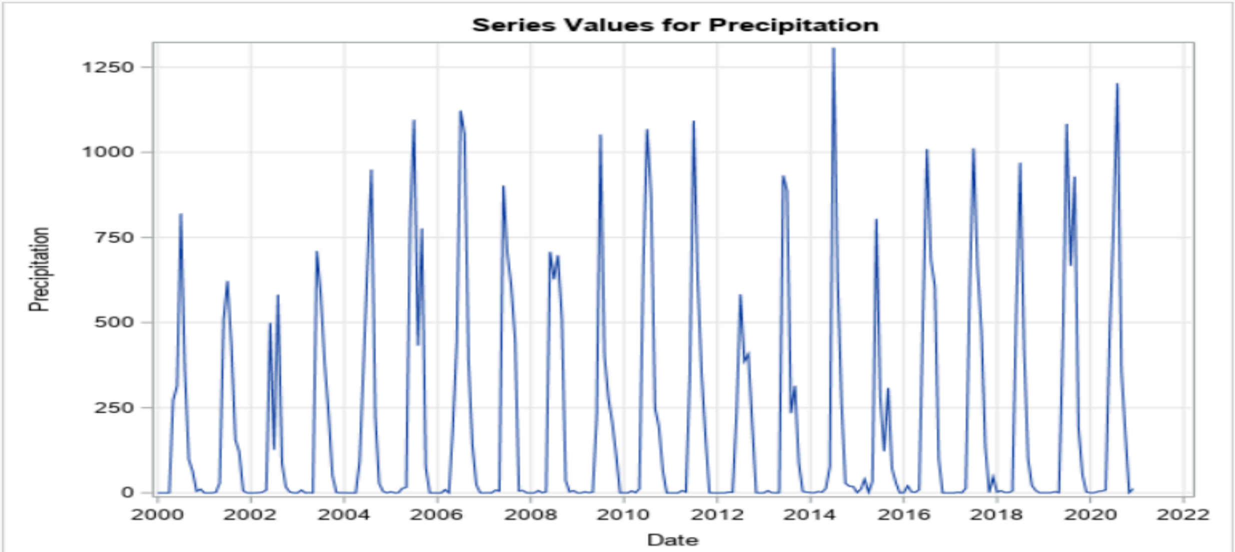
**Output:**

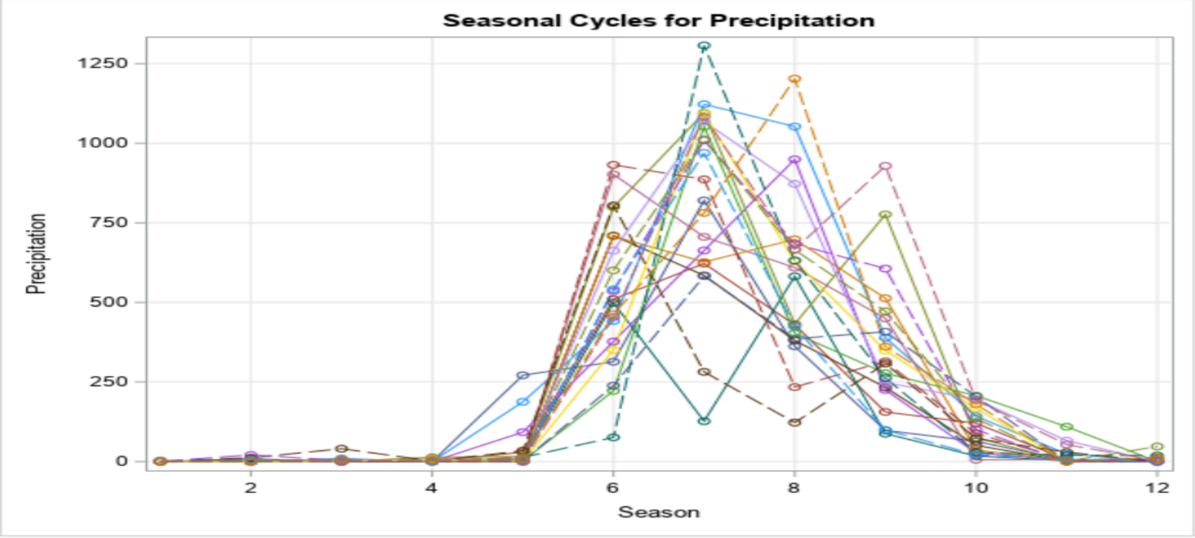


| Input Data Set | |
|---|---|
| Name | WORK.PREPROCESSEDDATA |
| Label | |
| Time ID Variable | Date |
| Time Interval | MONTH |
| Length of Seasonal Cycle | 12 |

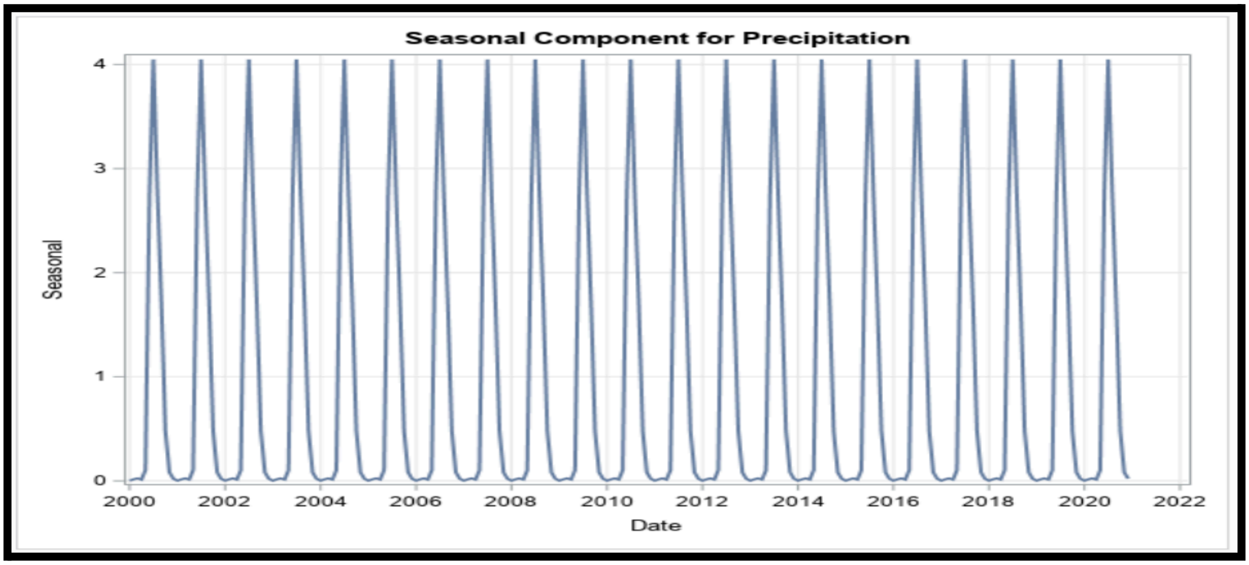
Length of our Seasonal Cycle is 12.

**Series Value for Precipitation:**



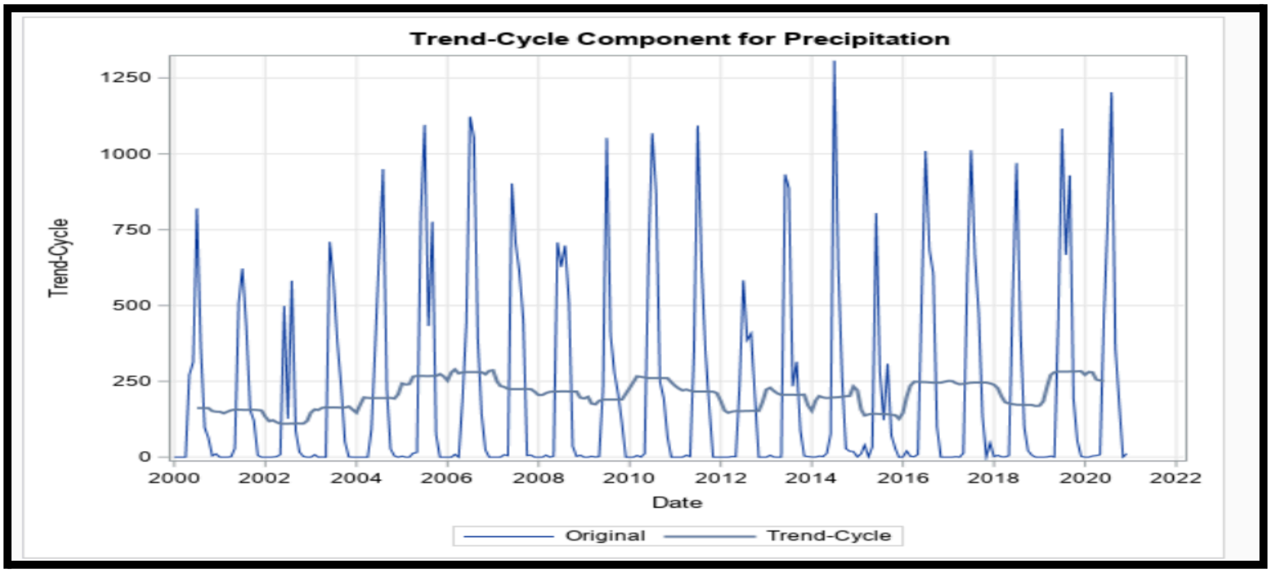


**Seasonal Component Graph:**



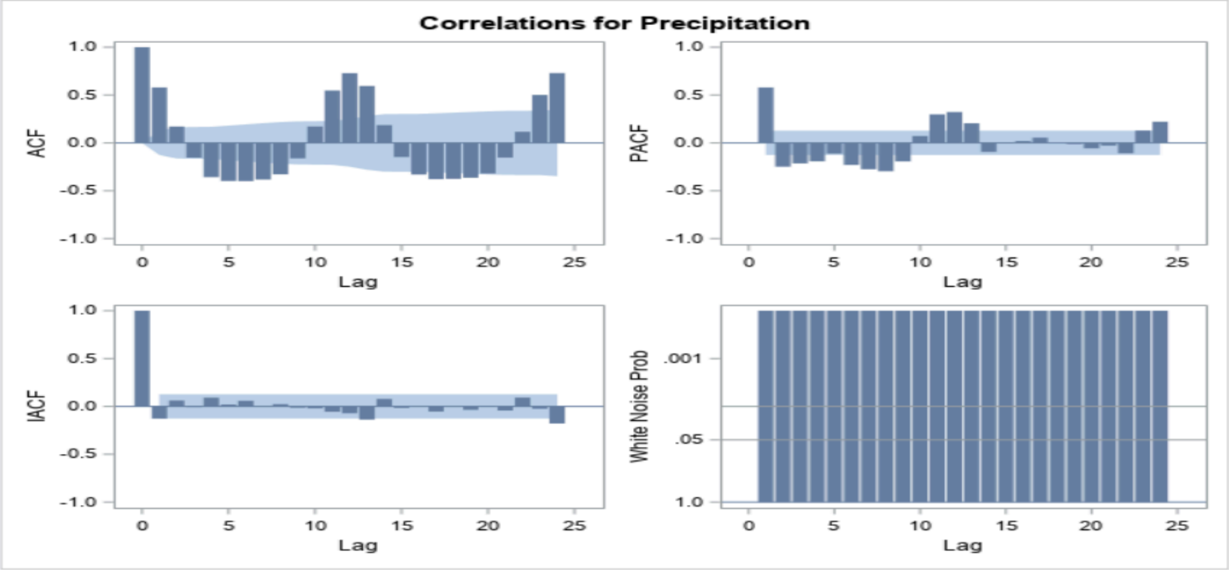
The seasonal plot and seasonal cycle plot clearly reveal seasonality in the series.

**Trend Component Graph:**



From the above graph we can observe that there is no particular trend followed for the data.

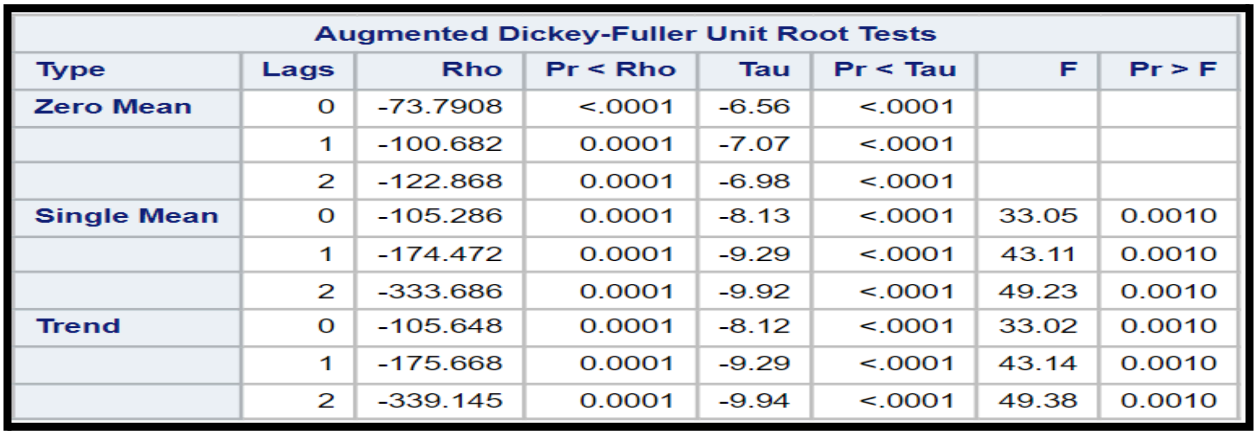
**Correlation Graphs:**



Auto Correlation(ACF) graph shows a strong seasonality with the precipitation and we can observe the significant lags at 2,11 in both PACF and IACF graphs. White Noise Probability graphs clearly lacks the white noise and has significant signal in it which can be retrieved from modeling.

Since our data set shows strong seasonality and no trend which can be used in the model determination.

**Augmented Dickey-Fuller Test:**



Augmented Dickey-Fuller Unit Root Tests

| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
|---|---|---|---|---|---|---|---|
| Zero Mean | 0 | -73.7908 | <.0001 | -6.56 | <.0001 | | |
| | 1 | -100.682 | 0.0001 | -7.07 | <.0001 | | |
| | 2 | -122.868 | 0.0001 | -6.98 | <.0001 | | |
| Single Mean | 0 | -105.286 | 0.0001 | -8.13 | <.0001 | 33.05 | 0.0010 |
| | 1 | -174.472 | 0.0001 | -9.29 | <.0001 | 43.11 | 0.0010 |
| | 2 | -333.686 | 0.0001 | -9.92 | <.0001 | 49.23 | 0.0010 |
| Trend | 0 | -105.648 | 0.0001 | -8.12 | <.0001 | 33.02 | 0.0010 |
| | 1 | -175.668 | 0.0001 | -9.29 | <.0001 | 43.14 | 0.0010 |
| | 2 | -339.145 | 0.0001 | -9.94 | <.0001 | 49.38 | 0.0010 |

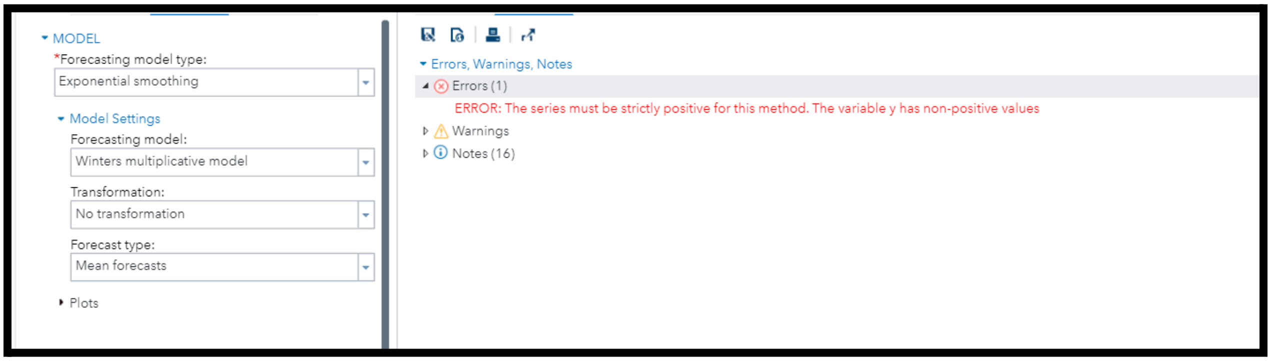
Based on the Augmented Dickey-Fuller Unit Root Test analysis results, we observe that both the Rho and Tau values for Precipitation are less than 0.05. Since our series exhibits a mean value, we employ a single mean check. Consequently, we have sufficient evidence to reject the null hypothesis that our data is non-stationary. As a result, we can confidently conclude that the data is stationary, and there's no necessity to perform differencing on the series.

## 5. Modeling and Forecasting:

### 5.1 Exponential Smoothing Models:

### 5.1.1 Winters Multiplicative Model:

Since the data is stationary and has seasonality and no trend, we can start modeling with Winters Multiplicative. But as the data has "0" values in the Precipitation column, this model does not work.



MODEL

*Forecasting model type: Exponential smoothing

Model Settings

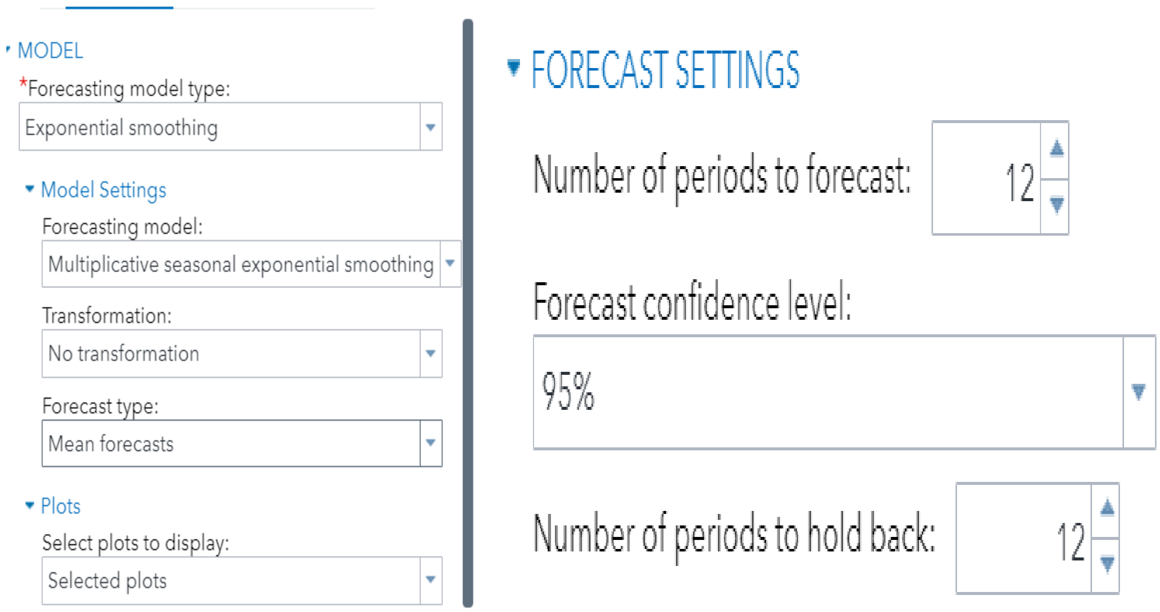
Forecasting model: Winters multiplicative model

Transformation: No transformation

Forecast type: Mean forecasts

ERROR: The series must be strictly positive for this method. The variable y has non-positive values

### 5.1.2 Multiplicative Seasonal Exponential Smoothing Model:



MODEL

*Forecasting model type: Exponential smoothing

Model Settings

Forecasting model: Multiplicative seasonal exponential smoothing

Transformation: No transformation

Forecast type: Mean forecasts

Plots

Select plots to display: Selected plots

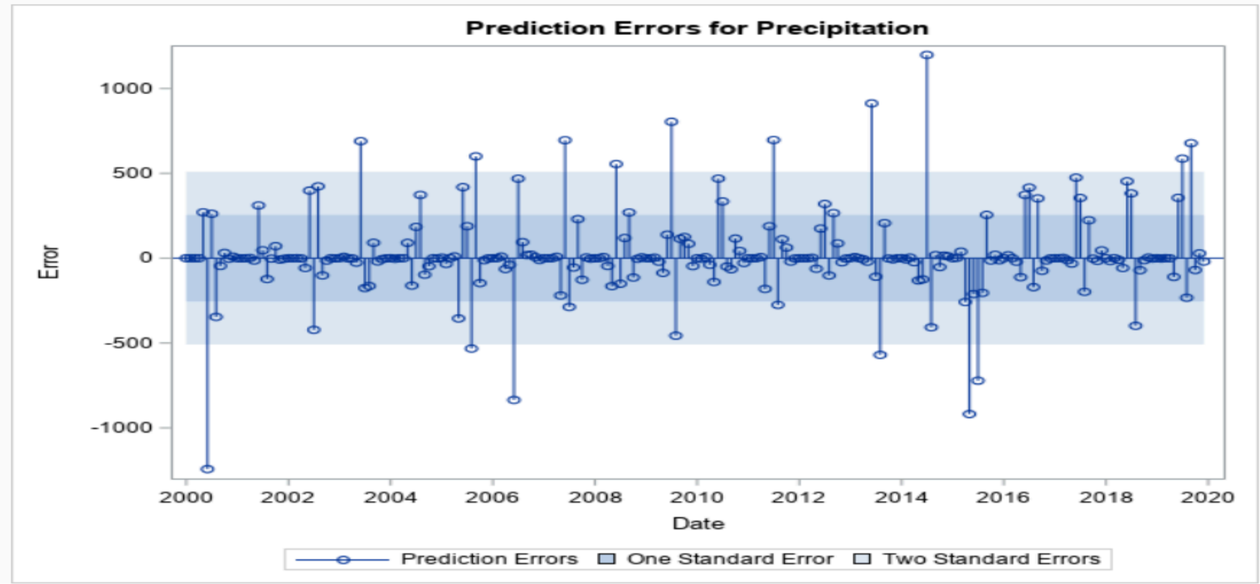
FORECAST SETTINGS

Number of periods to forecast: 12

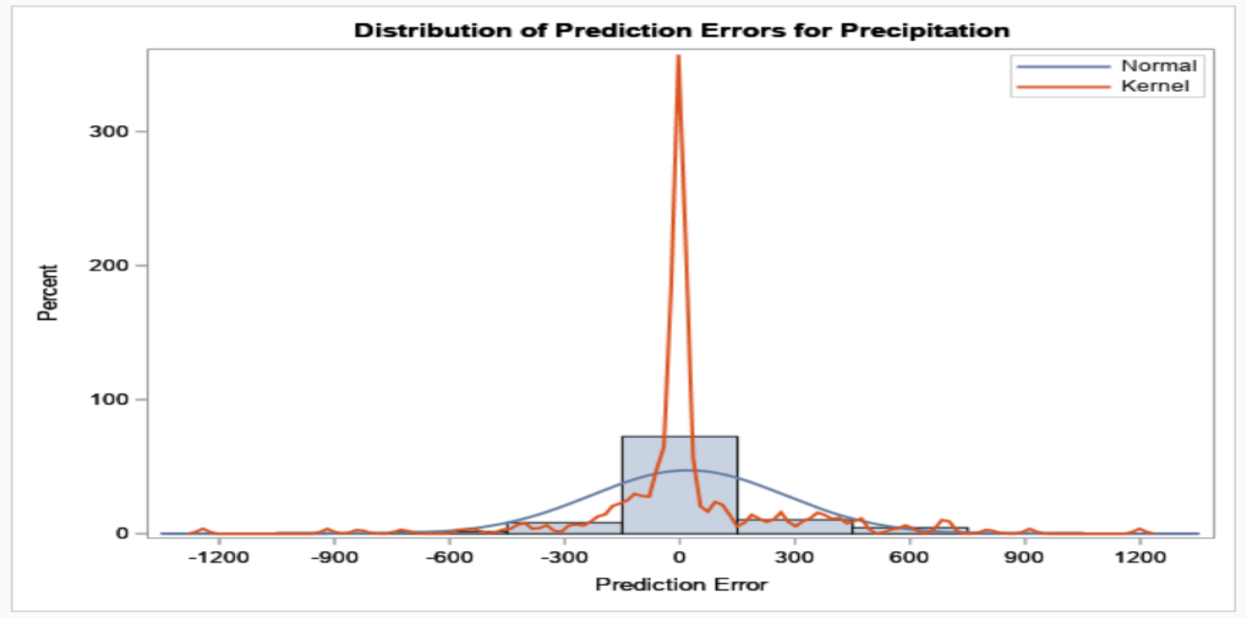
Forecast confidence level: 95%

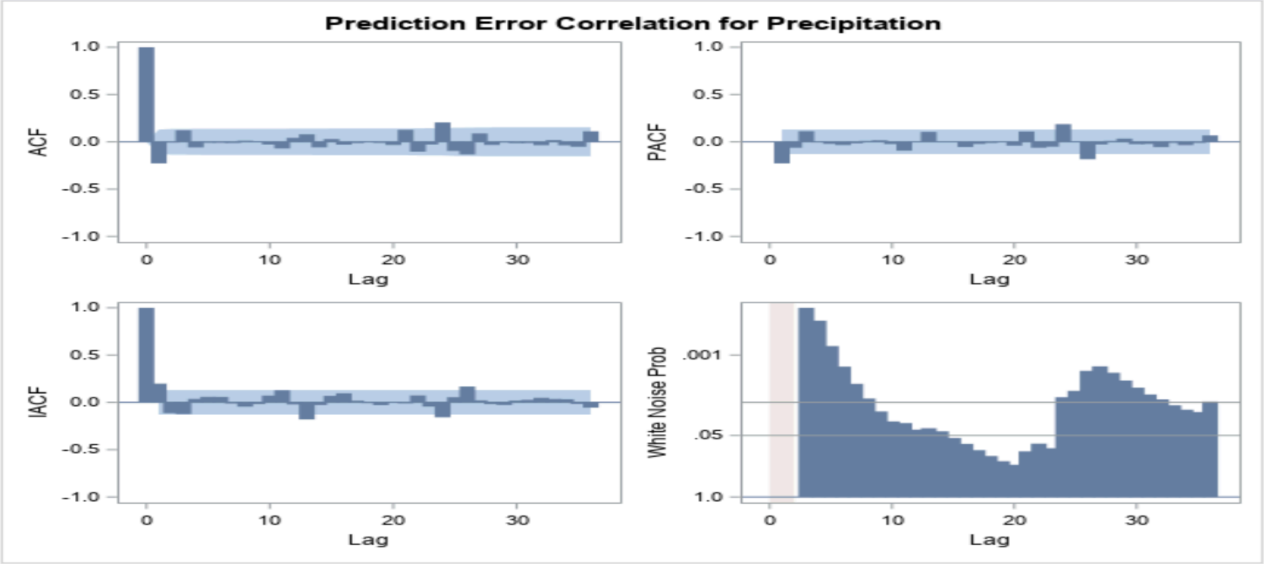
Number of periods to hold back: 12

Here we have Precipitation as the dependent variable, Date as Time ID and we forecasted a one full season and hold back one full season.



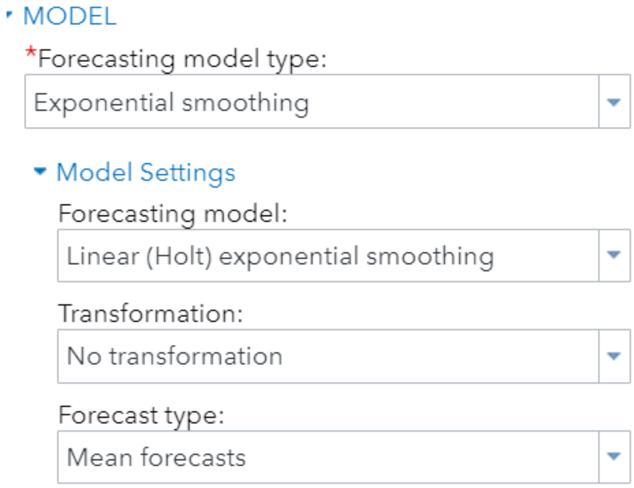
We see that there are errors that stand outside the one standard and two standard errors.




Prediction Error Correlation for Precipitation

From the above ACF and PACF graphs we can see that there are significant lags at lag 2, lag 22, lag 25 which says that prediction errors are correlated with past values of the predictor variables and variable which is being predicted. Also we have observed that there is still signal in the residuals from the white Noise Probability Test.

### 5.1.3 Linear (Holt) Exponential Smoothing :


MODEL

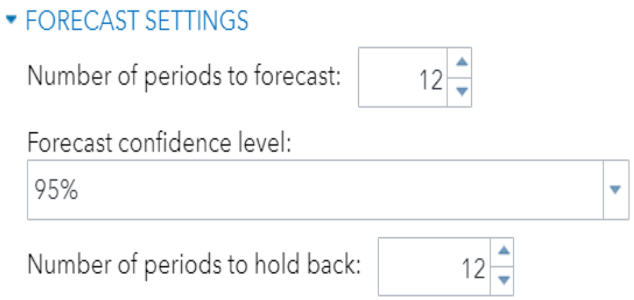
*Forecasting model type:
Exponential smoothing

Model Settings

Forecasting model:
Linear (Holt) exponential smoothing

Transformation:
No transformation

Forecast type:
Mean forecasts

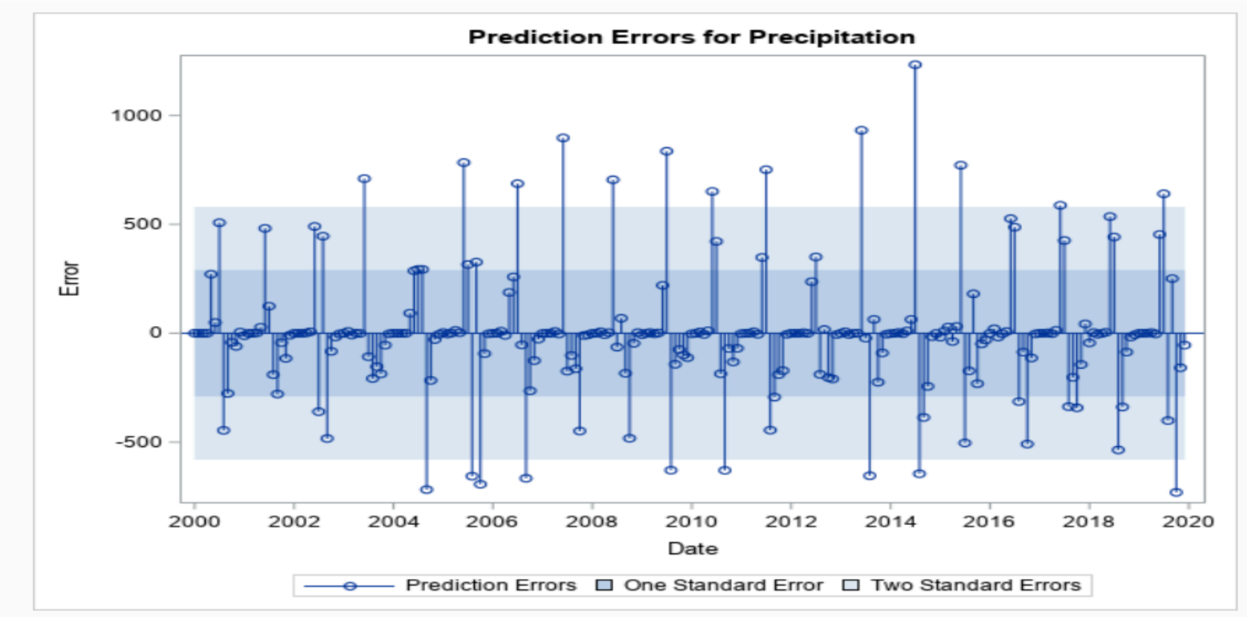


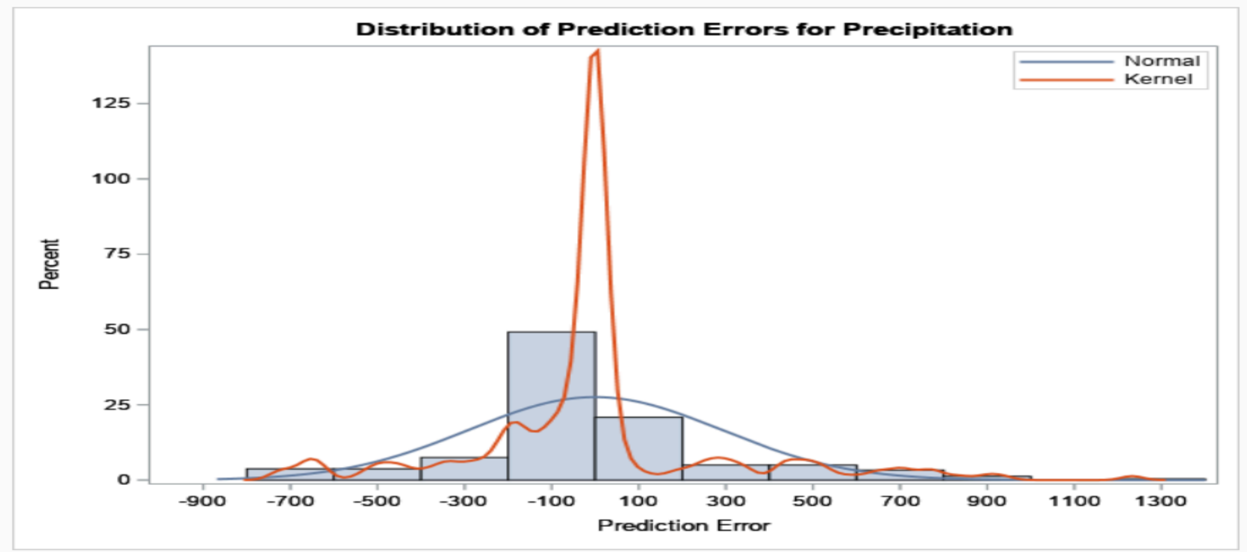
▾ FORECAST SETTINGS

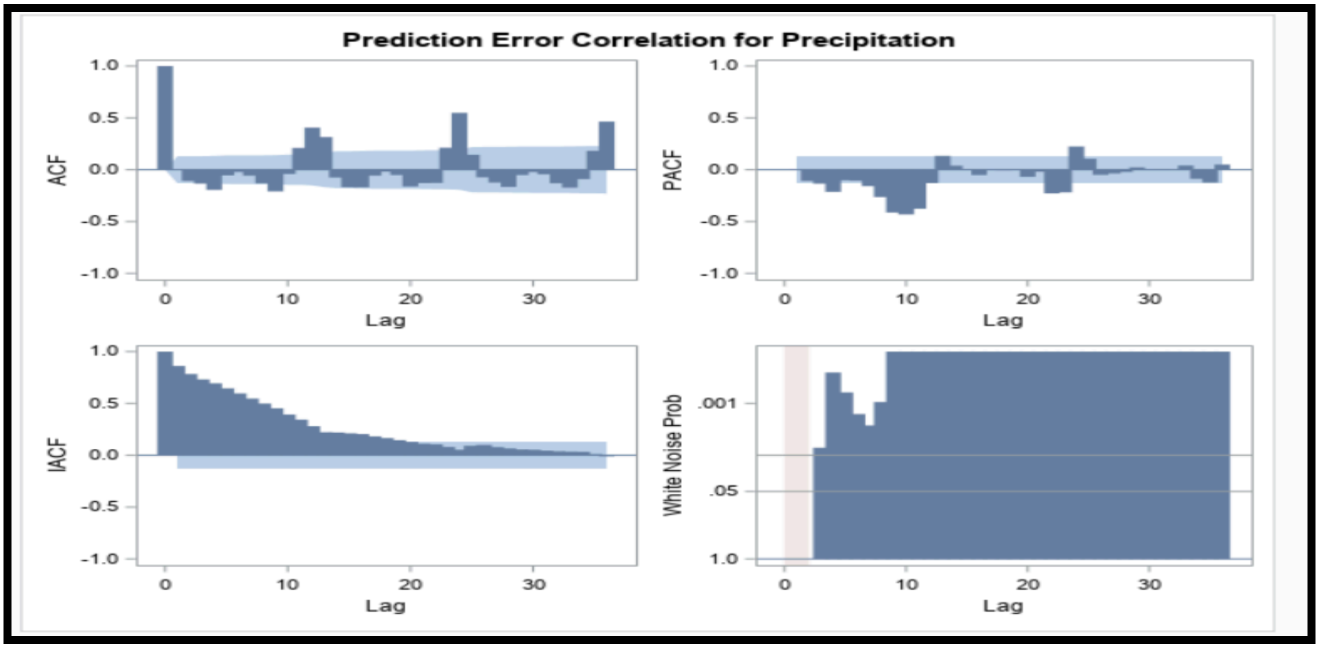
Number of periods to forecast: 12

Forecast confidence level: 95%

Number of periods to hold back: 12



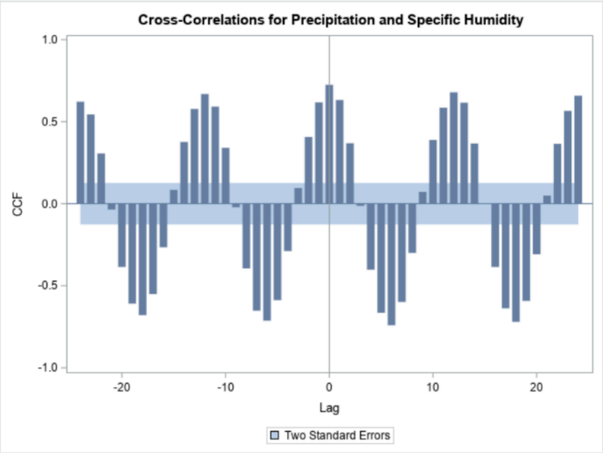


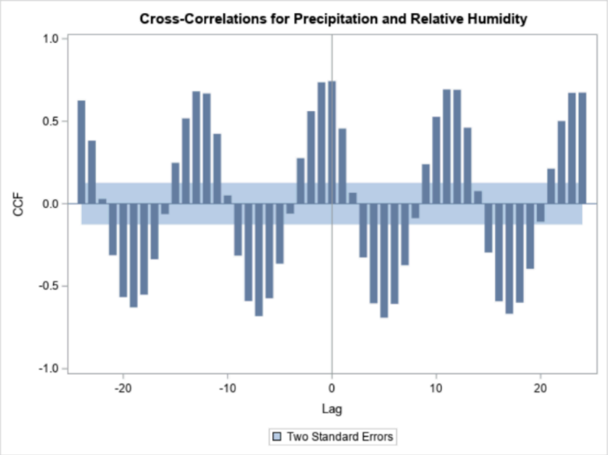


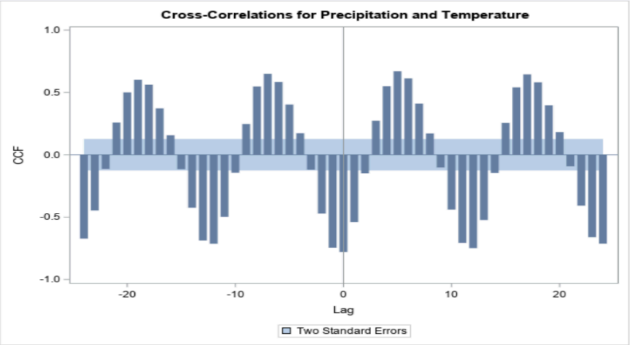
From the above results we can say that ACF, PACF, IACF show significant lags and hence prediction errors are clearly dependent on the past values and also there is a lot more signal that needs to be retrieved from the residuals. We can clearly conclude that is not the better model.

We did not have better results from the exponential models where we did not have the chance to add independent variables. Our data set consists of independent variables. In some cases independent variables make a significant impact on the prediction of the dependent variable. We check the cross correlation if independent variables have an impact on the dependent variable.

**Cross Correlation Graphs:**







Above correlations graphs indicate that the precipitation has significant dependence on all the variables i.e., Specific Humidity, Relative Humidity, Temperature at lag 0, lag 0 and lag 5 respectively.

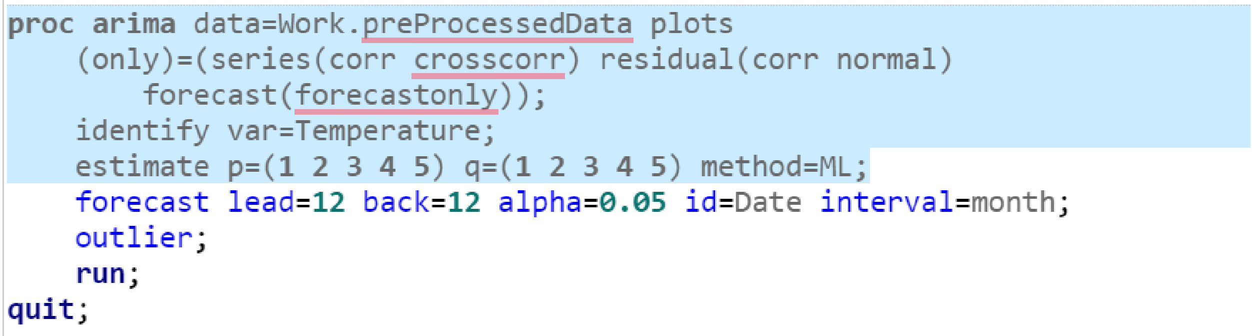
### 5.2 Pre-Whitening:

As part of our analysis, we are performing prewhitening, an essential preliminary step, especially in the context of ARMA (AutoRegressive Moving Average) modeling, before delving into the ARMA model development. Its primary purpose is to eliminate the autocorrelation structure present within a time series. Autocorrelation, where data points at different time steps are correlated, can complicate the analysis and modeling process. By prewhitening the data, this correlation is effectively removed, rendering the data more amenable to modeling. The transformed data becomes akin to white noise, a stationary series with no autocorrelation, simplifying the modeling process.

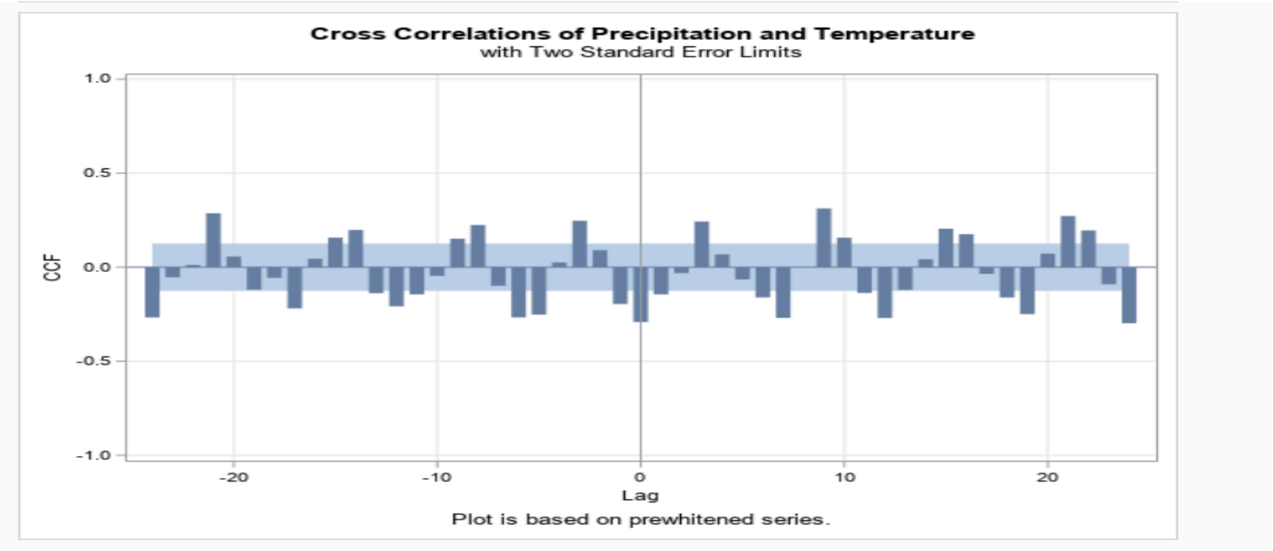
Prewhitening also enhances the stability of parameter estimation, improves model performance, and aids in diagnostic checks by ensuring that residuals exhibit white noise properties, aligning with the assumptions underlying ARMA modeling. In essence, prewhitening is a critical step that paves the way for more accurate, reliable, and interpretable ARMA model results and predictions, and we perform it diligently before embarking on ARMA model development

### h5.2.1 Prewhitening (Temperature ARMA 5,5):

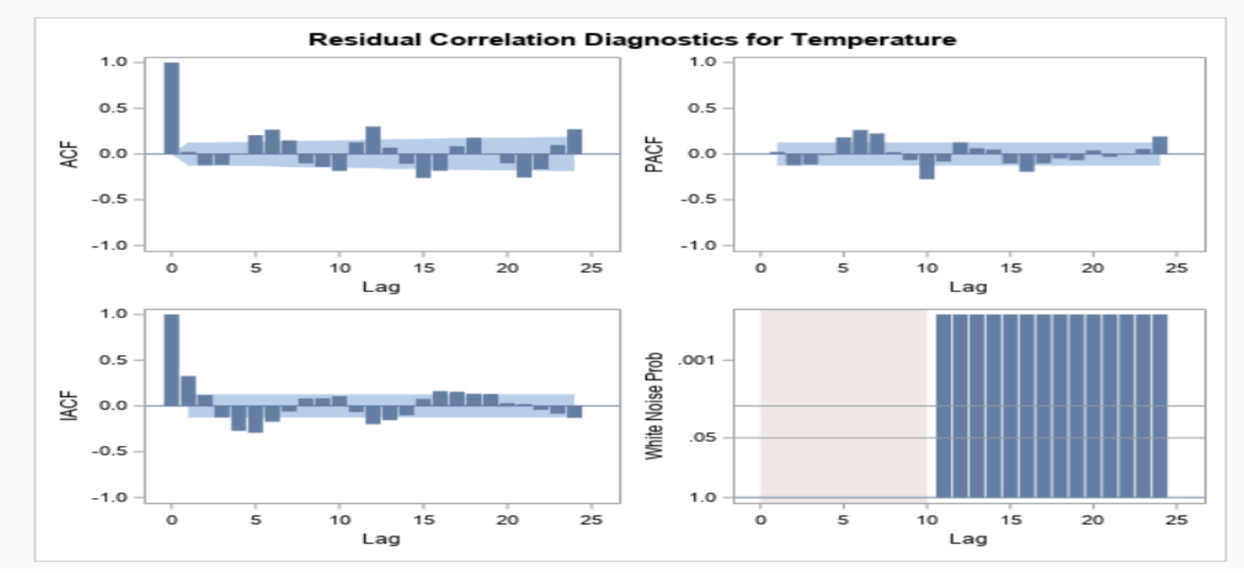
In this process, first we run the ARMA model by taking the temperature as the dependent variable. Once we have the better output of less significant lags at ACF and PACF, we consider the p,q values of that model and we apply the same model (ARIMAX 5,5) to the original dependent variable(Precipitation) by taking temperature as the independent variable. Here we modify the code as shown below.

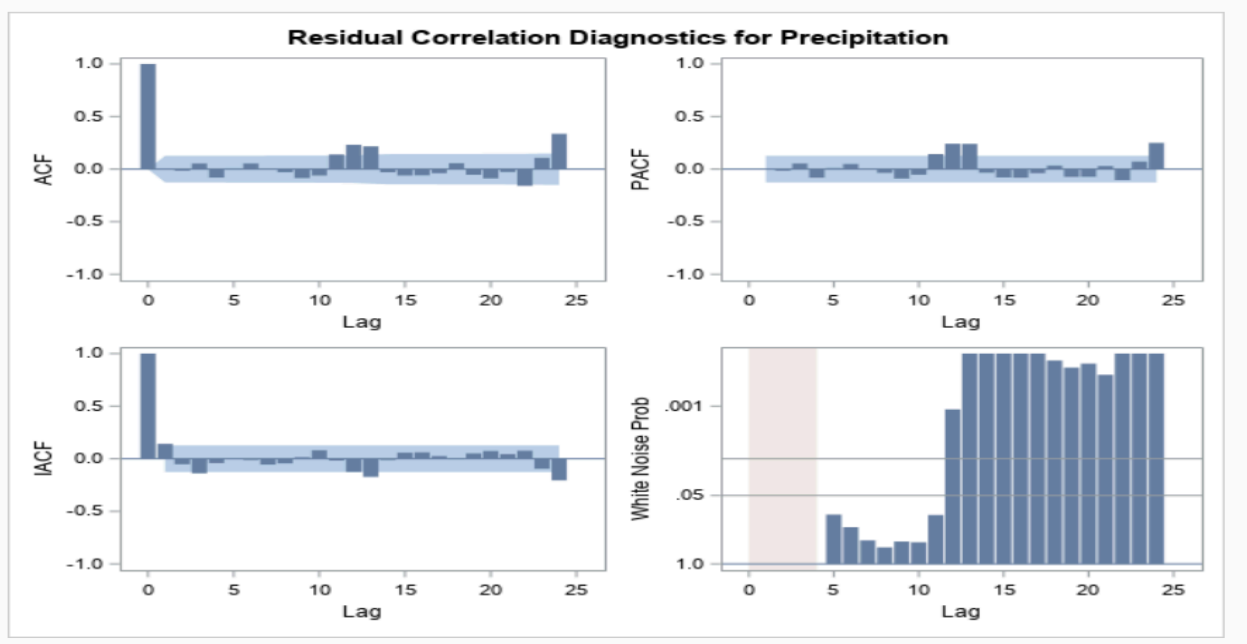


```
proc arima data=Work.preProcessedData plots
    (only)=(series(corr crosscorr) residual(corr normal)
        forecast(forecastonly));
    identify var=Temperature;
    estimate p=(1 2 3 4 5) q=(1 2 3 4 5) method=ML;
    forecast lead=12 back=12 alpha=0.05 id=Date interval=month;
    outlier;
    run;
quit;
```



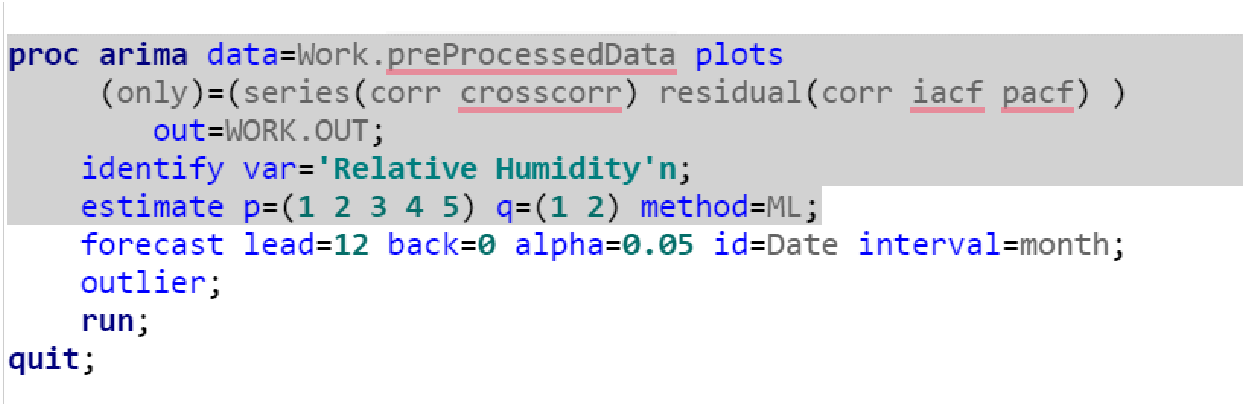
After prewhitening from the above graph, we say that the prewhitening reduced the dependency of the past values of temperature on the precipitation.



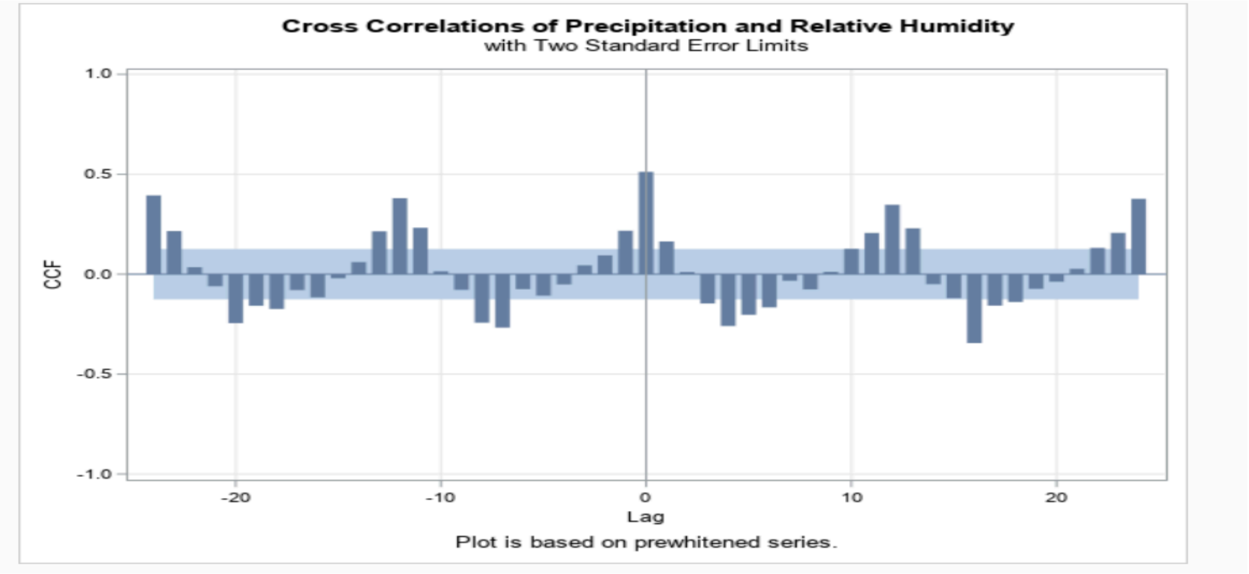


## 5.2.2 Prewhitening (Relative Humidity 5,2):

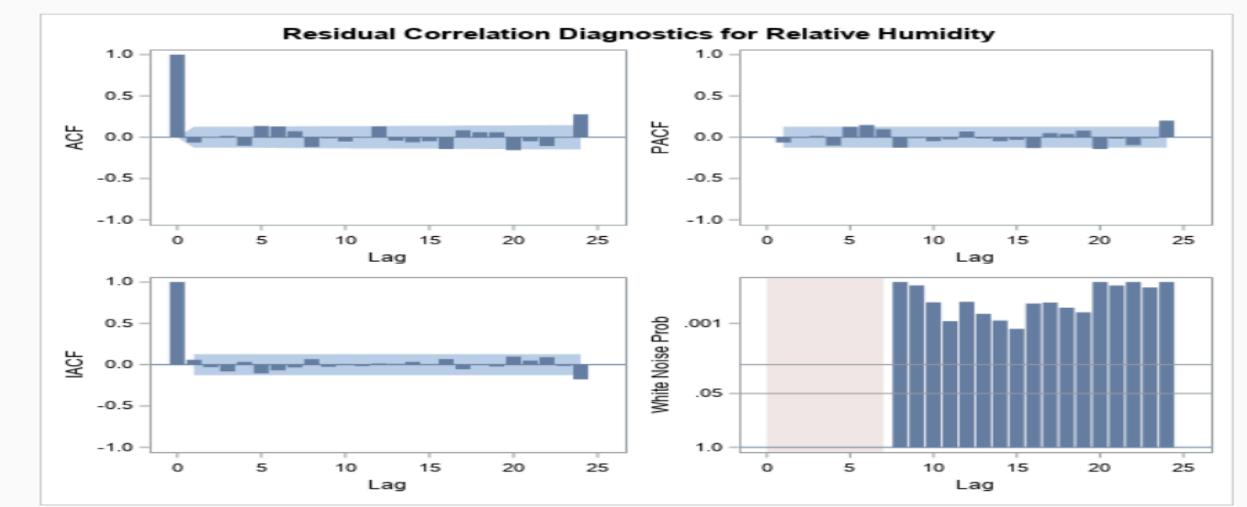
In this process, first we run the ARMA model by taking the Relative Humidity as the dependent variable. Once we have the better output of less significant lags at ACF and PACF, we consider the p,q values of that model and we apply the same model (ARIMAX 5,2) to the original dependent variable (Precipitation) by taking Relative Humidity as the independent variable. Here we modify the code as shown below.

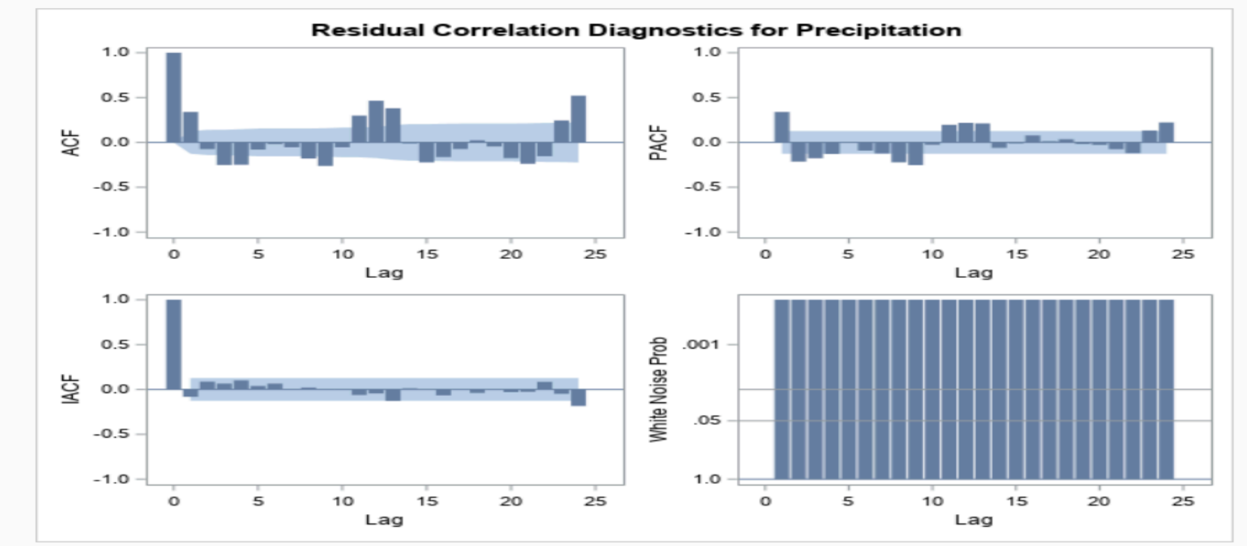


```
proc arima data=Work.preProcessedData plots
     (only)=(series(corr crosscorr) residual(corr iacf pacf) )
        out=WORK.OUT;
    identify var='Relative Humidity'n;
    estimate p=(1 2 3 4 5) q=(1 2) method=ML;
    forecast lead=12 back=0 alpha=0.05 id=Date interval=month;
    outlier;
    run;
quit;
```



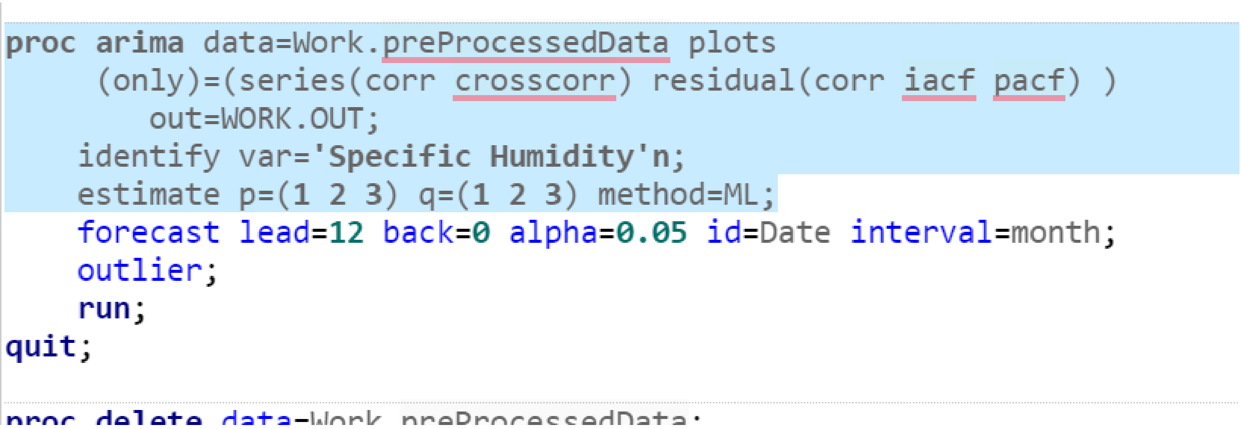
After prewhitening from the above graph, we say that the prewhitening reduced the dependency of the past values of Relative Humidity on the precipitation.



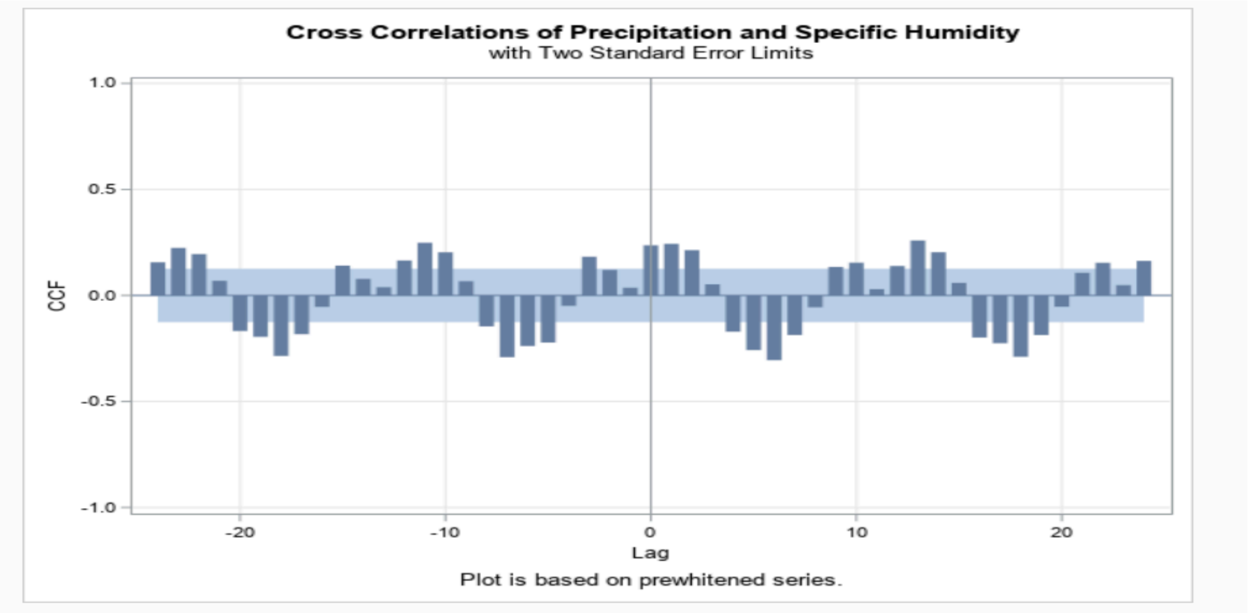


### 5.2.3 Prewhitening (Specific Humidity 3,3):

In this process, first we run the ARMA model by taking the Specific Humidity as the dependent variable. Once we have the better output of less significant lags at ACF and PACF, we consider the p,q values of that model and we apply the same model (ARIMAX 3,3) to the original dependent variable (Precipitation) by taking Specific Humidity as the independent variable. Here we modify the code as shown below.



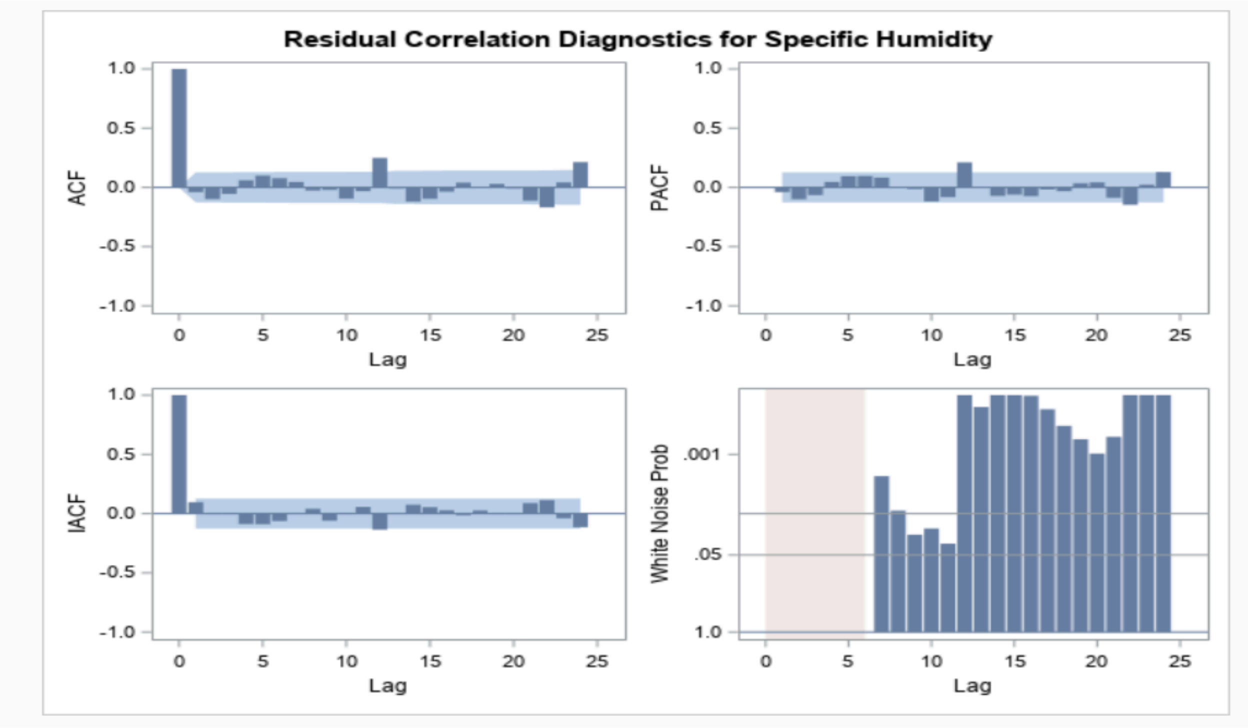
```
proc arima data=Work.preProcessedData plots
     (only)=(series(corr crosscorr) residual(corr iacf pacf) )
        out=WORK.OUT;
    identify var='Specific Humidity'n;
    estimate p=(1 2 3) q=(1 2 3) method=ML;
    forecast lead=12 back=0 alpha=0.05 id=Date interval=month;
    outlier;
    run;
quit;

proc delete data=Work.preProcessedData;
```

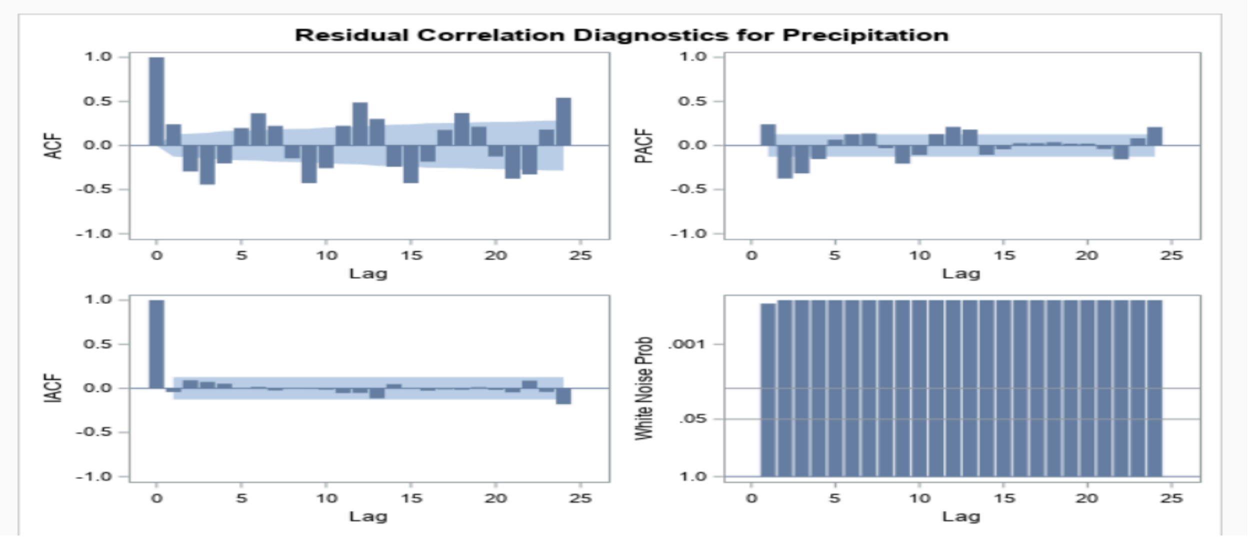


Cross Correlations of Precipitation and Specific Humidity
with Two Standard Error Limits

Plot is based on prewhitened series.

After prewhitening from the above graph, we say that the prewhitening reduced the dependency of the past values of Specific Humidity on the precipitation.





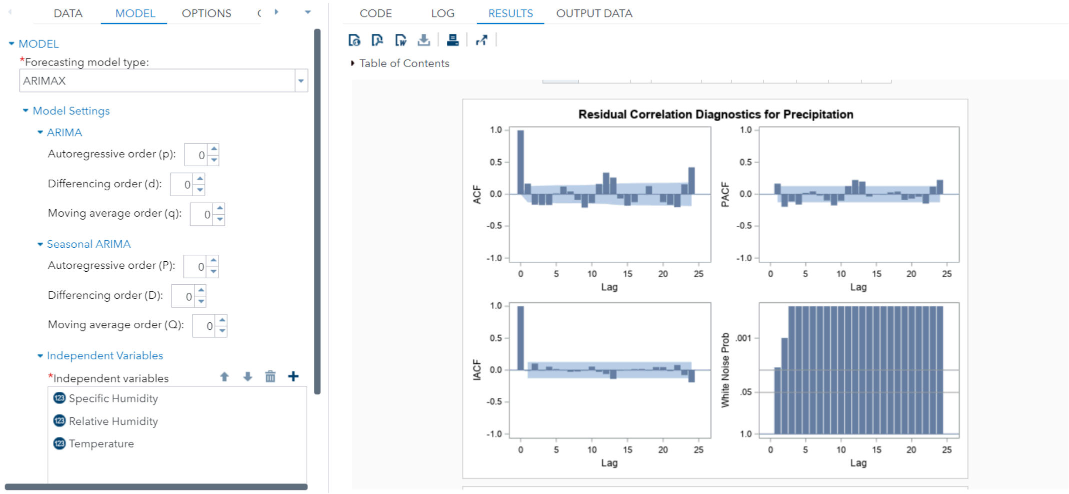
From Cross correlation graphs, we can observe that all independent variables have a significant impact.

## 5.3 ARIMAX Modeling:

Number of periods to forecast - 12 and holdback - 12

### 5.3.1 Basic ARIMAX model (0,0):

In the ARIMAX model, initially p and q were taken as 0 and 0 respectively. Precipitation is the dependent variable, Date is the Time Id and all 3 independent variables are added.



MODEL

*Forecasting model type: ARIMAX

Model Settings

ARIMA

Autoregressive order (p): 0

Differencing order (d): 0

Moving average order (q): 0

Seasonal ARIMA

Autoregressive order (P): 0

Differencing order (D): 0

Moving average order (Q): 0

Independent Variables

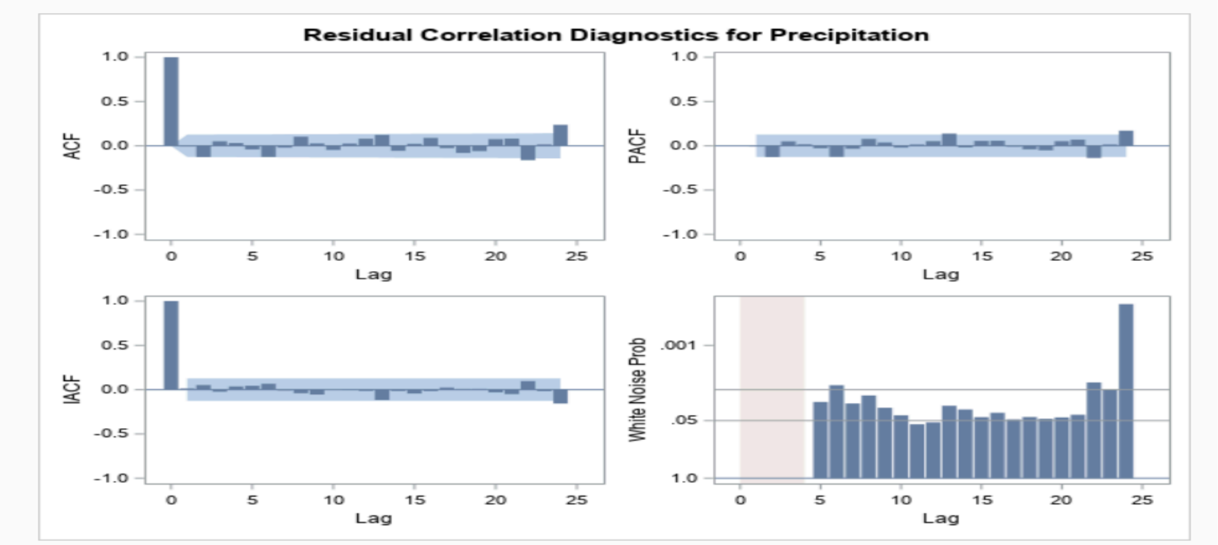
*Independent variables

- Specific Humidity
- Relative Humidity
- Temperature

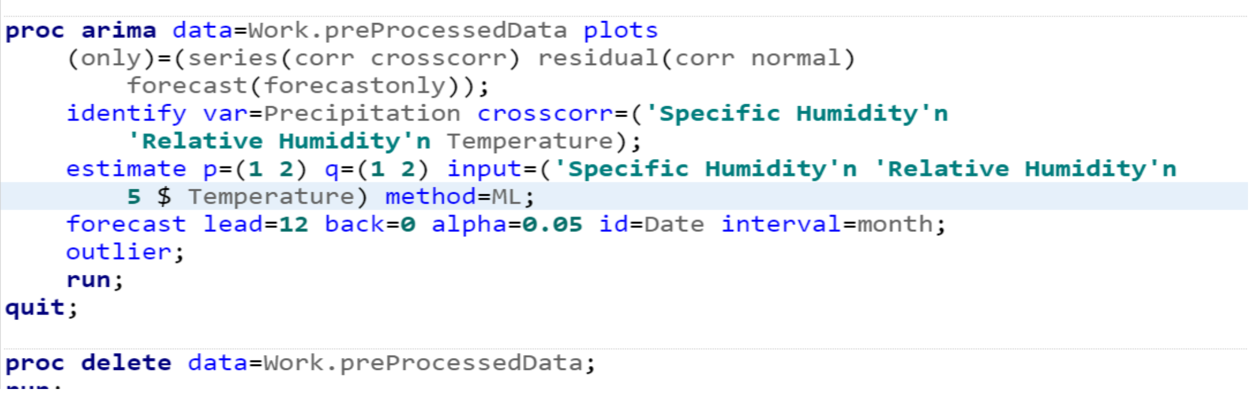
After running the model, we still have a signal in the residuals which needs to be extracted and hence we decided to go with ARIMAX(2,2).

### 5.3.2 ARIMAX(2,2) Model:

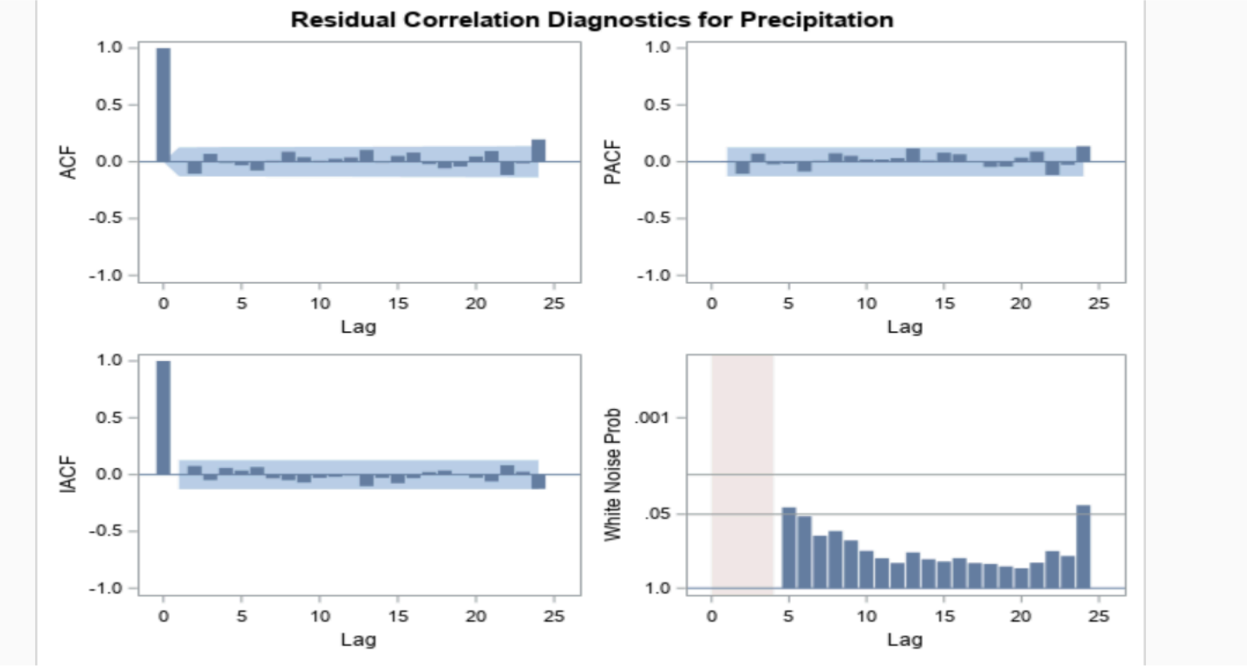
In the ARIMAX model, now p and q were taken as 2 and 2 respectively. Precipitation is the dependent variable, Date is the Time Id and all 3 independent variables are added.



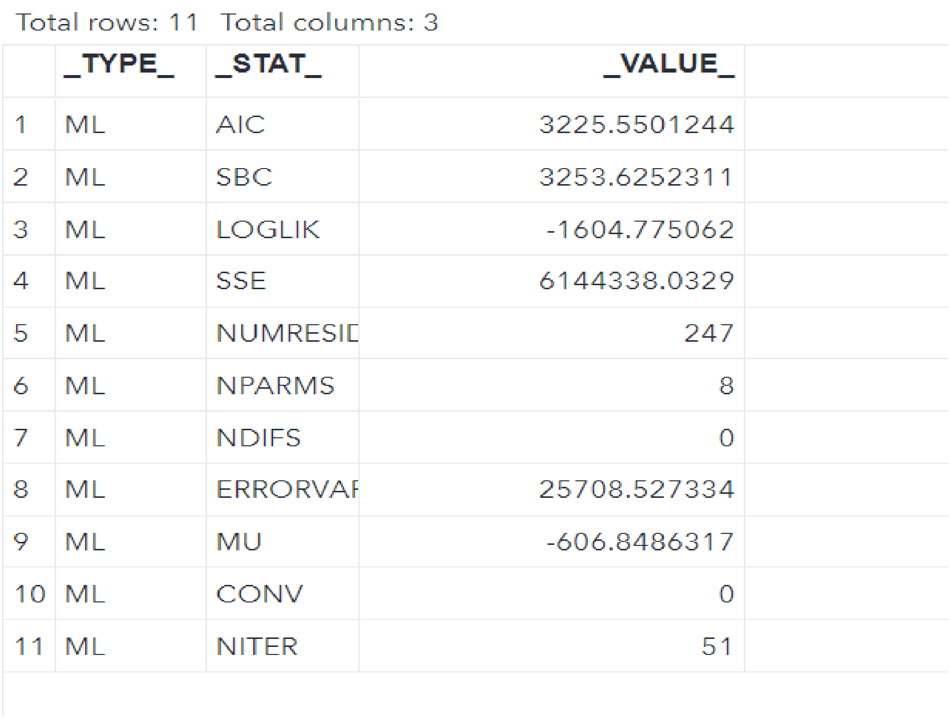
Increasing the p,q values gives us better results compared to the basic ARIMAX but there is significant signal in the residuals and hence we should continue our modeling. Earlier we observed that Temperature has a lag 5 effect on the precipitation. After adding lagged effect ‘5’ for the Temperature variable in the code:



```
proc arima data=Work.preProcessedData plots
    (only)=(series(corr crosscorr) residual(corr normal)
        forecast(forecastonly));
    identify var=Precipitation crosscorr=('Specific Humidity'n
        'Relative Humidity'n Temperature);
    estimate p=(1 2) q=(1 2) input=('Specific Humidity'n 'Relative Humidity'n
        5 $ Temperature) method=ML;
    forecast lead=12 back=0 alpha=0.05 id=Date interval=month;
    outlier;
    run;
quit;

proc delete data=Work.preProcessedData;
```



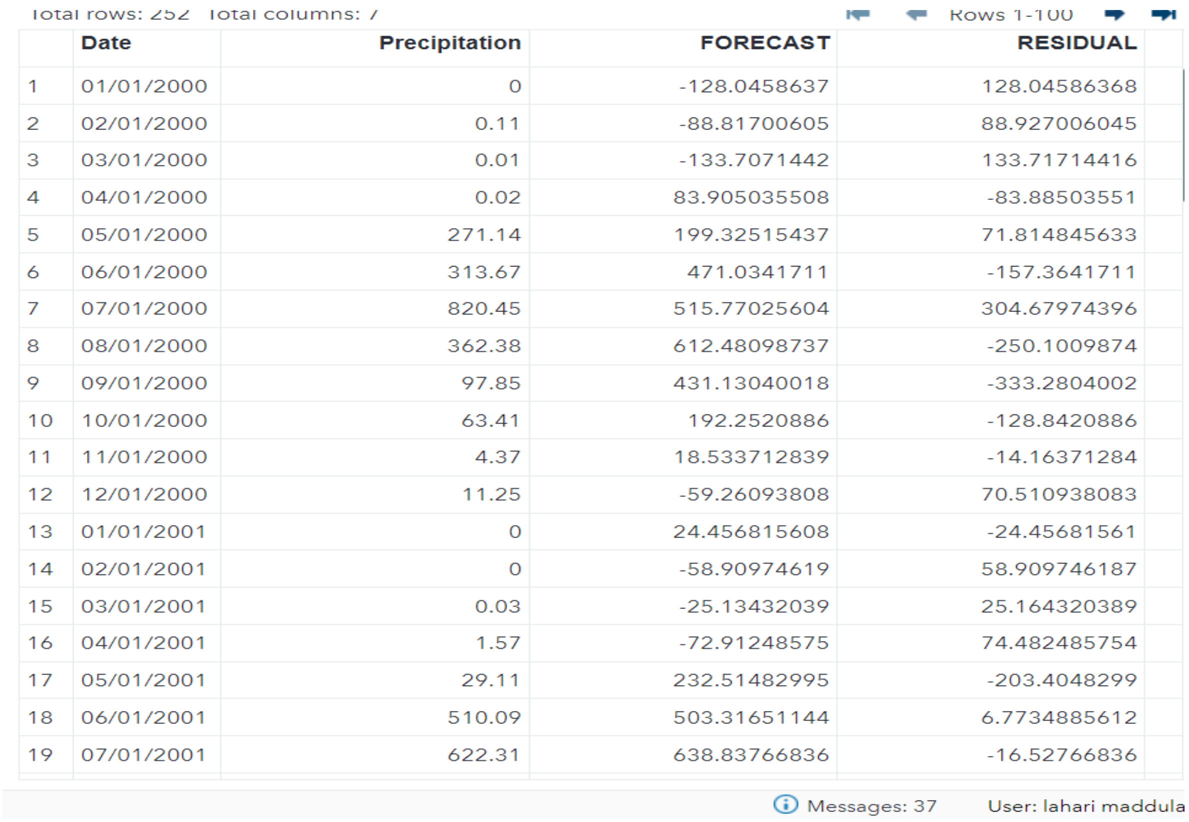
Adding the lag 5 effect of temperature gave us better results compared to the ARIMAX 2,2 but there is significant spikes at lag 5, lag 24 and hence we go with a two-step modeling process of applying modeling to the residual dataset from the above ARIMAX(2,2) model.



Total rows: 11 Total columns: 3

| | _TYPE_ | _STAT_ | _VALUE_ |
|---|---|---|---|
| 1 | ML | AIC | 3225.5501244 |
| 2 | ML | SBC | 3253.6252311 |
| 3 | ML | LOGLIK | -1604.775062 |
| 4 | ML | SSE | 6144338.0329 |
| 5 | ML | NUMRESID | 247 |
| 6 | ML | NPARMS | 8 |
| 7 | ML | NDIFS | 0 |
| 8 | ML | ERRORVAR | 25708.527334 |
| 9 | ML | MU | -606.8486317 |
| 10 | ML | CONV | 0 |
| 11 | ML | NITER | 51 |

## Snapshot of the Residual Dataset:



Total rows: 252 Total columns: 7

Rows 1-100

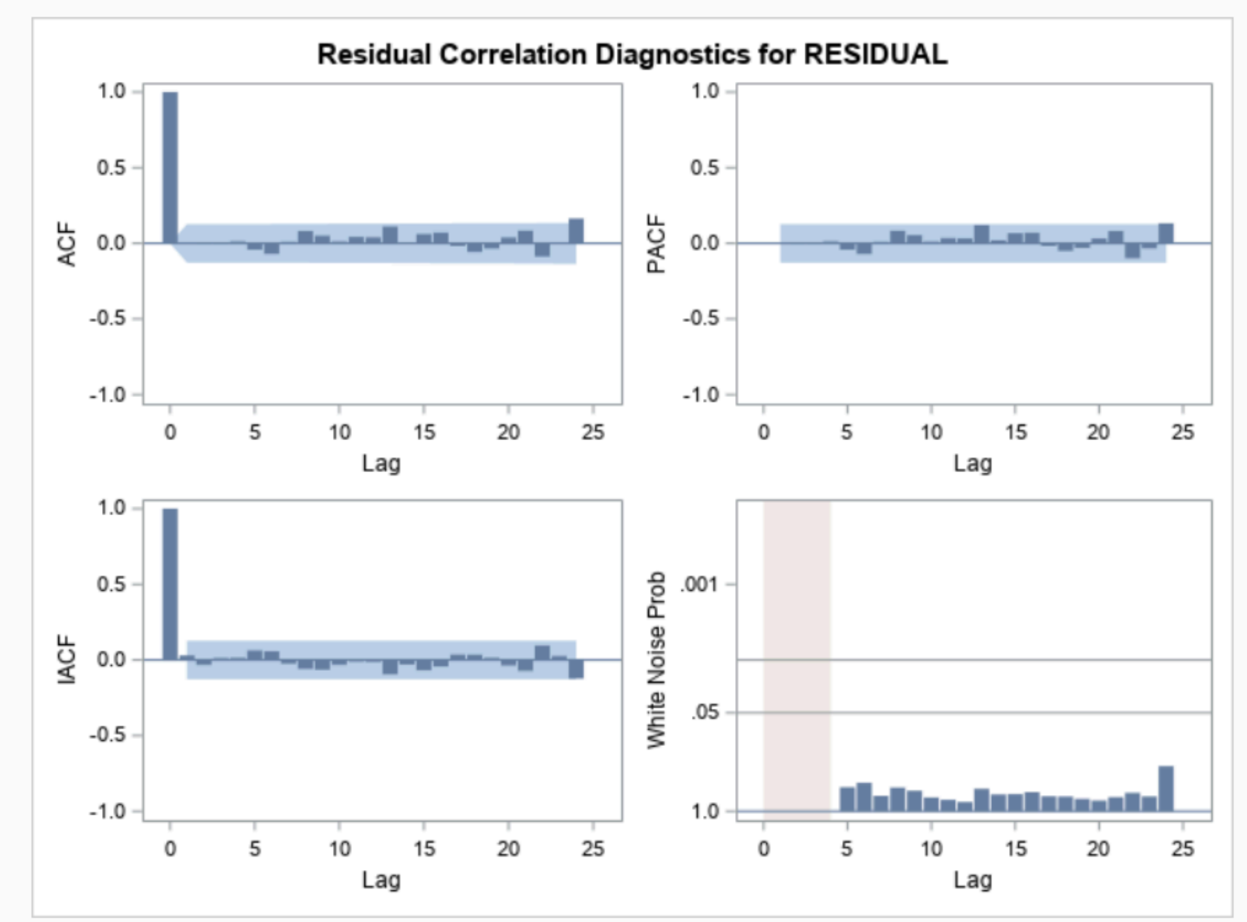
| | Date | Precipitation | FORECAST | RESIDUAL |
|---|---|---|---|---|
| 1 | 01/01/2000 | 0 | -128.0458637 | 128.04586368 |
| 2 | 02/01/2000 | 0.11 | -88.81700605 | 88.927006045 |
| 3 | 03/01/2000 | 0.01 | -133.7071442 | 133.71714416 |
| 4 | 04/01/2000 | 0.02 | 83.905035508 | -83.88503551 |
| 5 | 05/01/2000 | 271.14 | 199.32515437 | 71.814845633 |
| 6 | 06/01/2000 | 313.67 | 471.0341711 | -157.3641711 |
| 7 | 07/01/2000 | 820.45 | 515.77025604 | 304.67974396 |
| 8 | 08/01/2000 | 362.38 | 612.48098737 | -250.1009874 |
| 9 | 09/01/2000 | 97.85 | 431.13040018 | -333.2804002 |
| 10 | 10/01/2000 | 63.41 | 192.2520886 | -128.8420886 |
| 11 | 11/01/2000 | 4.37 | 18.533712839 | -14.16371284 |
| 12 | 12/01/2000 | 11.25 | -59.26093808 | 70.510938083 |
| 13 | 01/01/2001 | 0 | 24.456815608 | -24.45681561 |
| 14 | 02/01/2001 | 0 | -58.90974619 | 58.909746187 |
| 15 | 03/01/2001 | 0.03 | -25.13432039 | 25.164320389 |
| 16 | 04/01/2001 | 1.57 | -72.91248575 | 74.482485754 |
| 17 | 05/01/2001 | 29.11 | 232.51482995 | -203.4048299 |
| 18 | 06/01/2001 | 510.09 | 503.31651144 | 6.7734885612 |
| 19 | 07/01/2001 | 622.31 | 638.83766836 | -16.52766836 |

Messages: 37 User: lahari maddula

This is taken as the main dataset and ARMA(2,2) is applied. Now Residual is a dependent variable and there are no independent variables.

After Modeling ARMA(2,2):

Number of periods to forecast - 12 and holdback - 12



Here the residuals have very less spikes which means that there is very less signal in the residuals. The ACF, PACF and IACF graphs also do not contain any significant spikes which says that this model can be considered as the best model for forecasting.

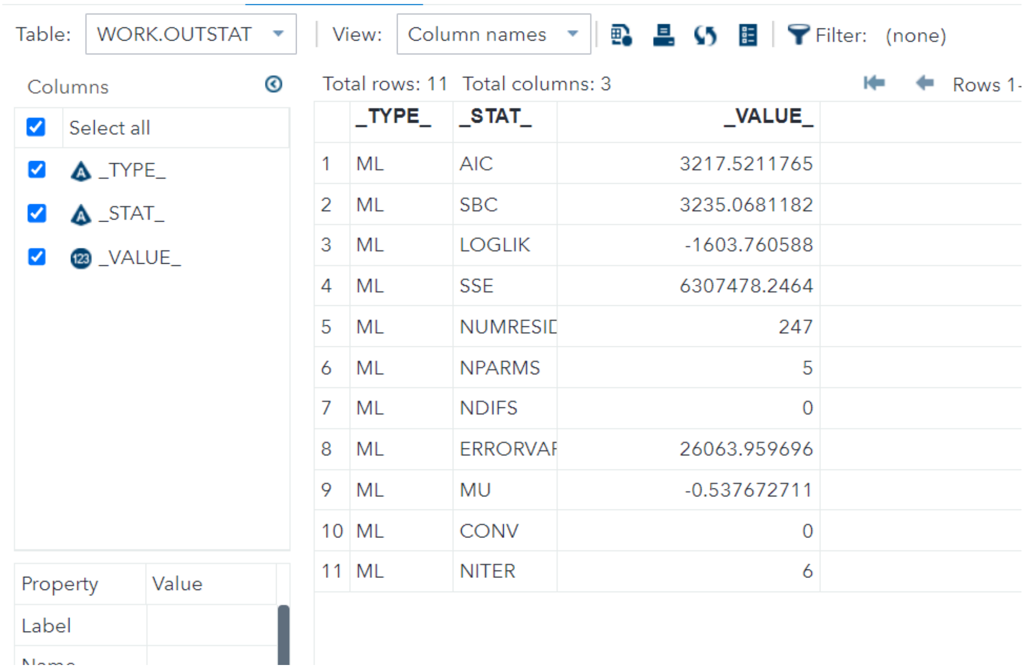


Table: WORK.OUTSTAT View: Column names Filter: (none)

Total rows: 11 Total columns: 3

| | _TYPE_ | _STAT_ | _VALUE_ |
|---|---|---|---|
| 1 | ML | AIC | 3217.5211765 |
| 2 | ML | SBC | 3235.0681182 |
| 3 | ML | LOGLIK | -1603.760588 |
| 4 | ML | SSE | 6307478.2464 |
| 5 | ML | NUMRESIC | 247 |
| 6 | ML | NPARMS | 5 |
| 7 | ML | NDIFS | 0 |
| 8 | ML | ERRORVAF | 26063.959696 |
| 9 | ML | MU | -0.537672711 |
| 10 | ML | CONV | 0 |
| 11 | ML | NITER | 6 |

## Conclusion:

The primary objective of this work was to create a reliable time series forecasting model for precipitation in Mumbai utilizing meteorological data from the POWER project. Prediction of precipitation is critical for a variety of industries, including renewable energy, building energy efficiency, and agriculture, particularly in a densely populated coastal city like Mumbai, which has a unique climate.

The initial stage of data exploration offered an overview of the dataset's structure as well as the correlations between meteorological variables. A thorough examination of the time series data revealed that precipitation in Mumbai indicated substantial seasonality but lacked a consistent pattern. The Augmented Dickey-Fuller Unit Root Test supported this finding, demonstrating the dataset's stationary identity.

Several models were taken into account throughout the forecasting phase. Winters Multiplicative and Multiplicative Seasonal Exponential Smoothing Model gave preliminary insights, but their results were unsatisfactory since they left key signals unmodeled. Moreover, the exponential models did not take use of the dataset's independent variables.