An Analytics of short-term rainfall Prediction model for Delhi, India: Seasonal-ARIMA vs Holt-Winters Exponential Smoothing method

Overview

- Introduction
- Literature review
- Data and Methodology
- Experimental Results
- Conclusion
- Future work

Introduction

- **Predictive Analytics** is the branch of the advanced analytics which is used to make predictions about the future events.
- Method for extraction of hidden predictive information from large database.
- Computational process that helps identifying interesting patterns and forecasting using the past information.

Introduction

Main objective of this paper is as follow:

To identify the most apt model using time series analysis`s non-invasive techniques and obtain forecast of the rainfall pattern from the model selected on the basis of historical rainfall data of New Delhi, India.

Literature Review

	Title	Authors	Year	Journal
[1]	Statistical modeling of monthly rainfall in selected stations in forest and savannah eco-climatic regions of Nigeria, Journal of Climatology and Weather Forecasting	Akinbobola A.	2018	Journal of Climatology and Weather Forecasting
[2]	Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method	Mahmud I.	2016	Environmental Engineering Research
[3]	Machine Learning Techniques for Short-Term Rain Forecasting System in the North-eastern Part of Thailand	Ingsrisawang L.	2012	International Journal of Computer and Information Engineering
[4]	A Seasonal ARIMA model for forecasting monthly rainfall in Tamil Nadu	Nirmala M.	2011	International Journal of Computer Engineering and Technology

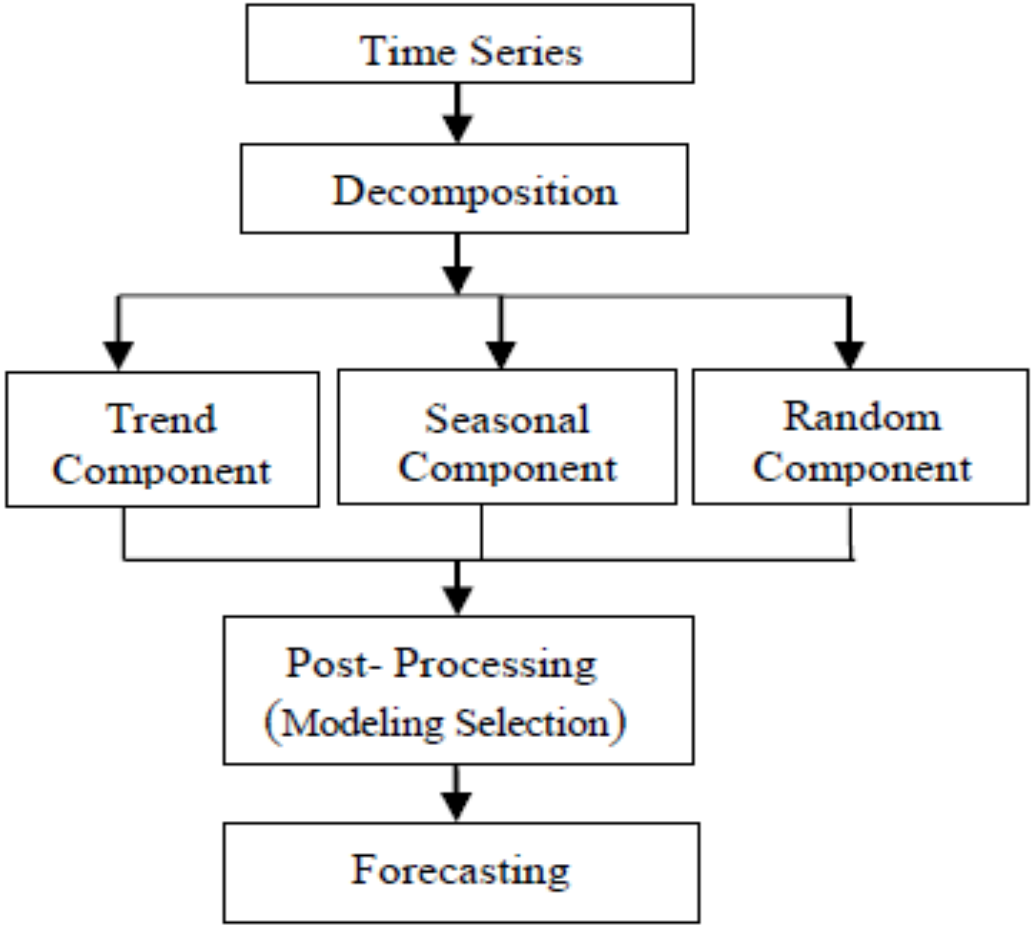
Data Description

- **Tools Used:** R software
- **Dataset Used:** Historical Rainfall data of Delhi (monthly from January`06 – July`18). 152 observations with measures in mm.

Time Series Analysis

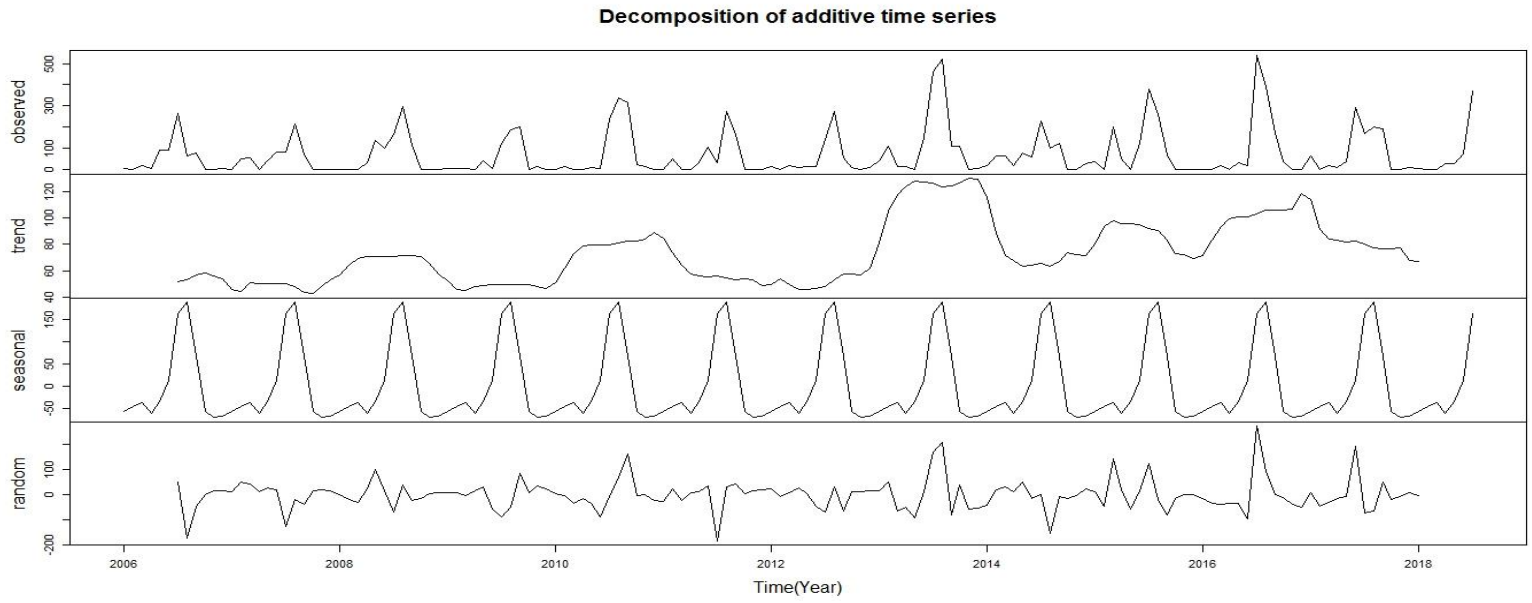
- It is a platform that lets you explore, analyze, and forecast a time series.
- A time series is a sequence of data points, measured at uniform time intervals.
- The main objective is to model a process, and to provide its compact description.
- Methods used for such analyses:
  - ARIMA(Auto Regressive Integrated Moving Average)
  - Exponential smoothing models

Process for performing such a analysis



Decomposition of Time Series

- A time series can be broken down into various individual components.
- The decomposition of time series is a statistical method that breaks a time series down into its various components .



Time series plot for Rainfall data

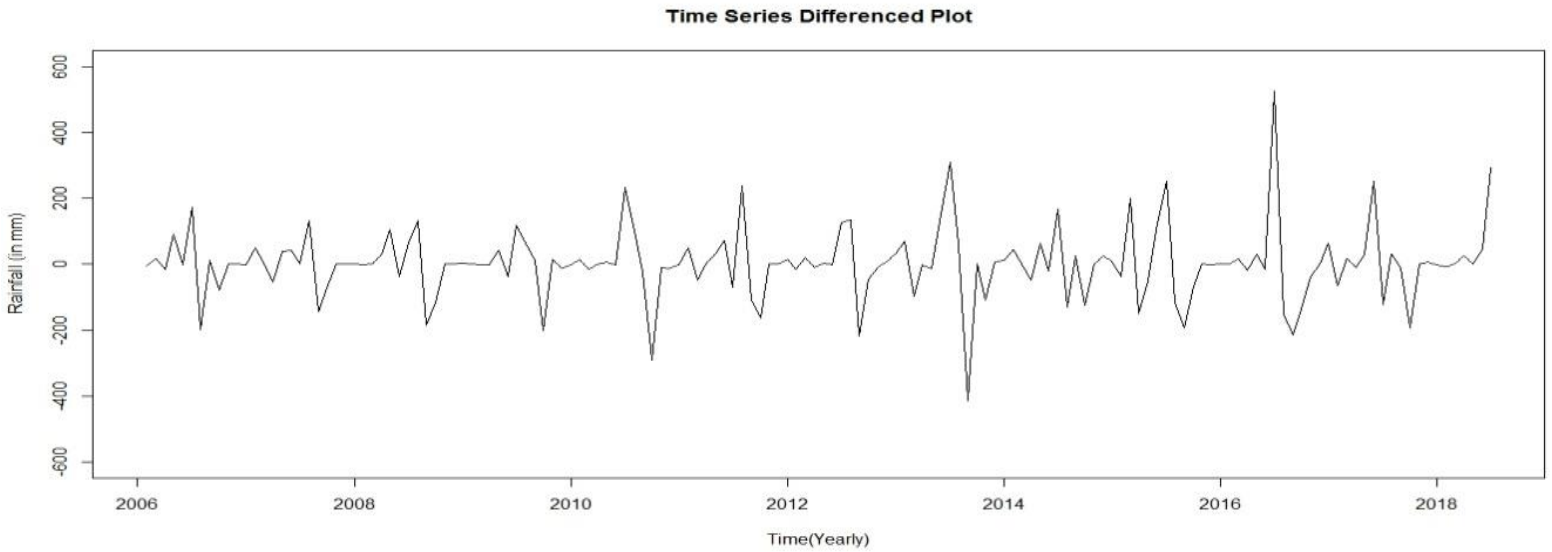


Fig. Stationary time series plot after second order seasonal differencing for rainfall time series data.

The model is stationary as its mean and variance is constant over time.

Model Selection

- *SARIMA Model:* $(p,d,q)(P,D,Q)_s$

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cut off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

- *Holt-Winters Model:*

$$y_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)}$$

Estimation of three components:

l_t for the level component,

b_t for the trend,

s_t for capturing the seasonal component on the basis of the weights given to the observation.

Forecast Plots

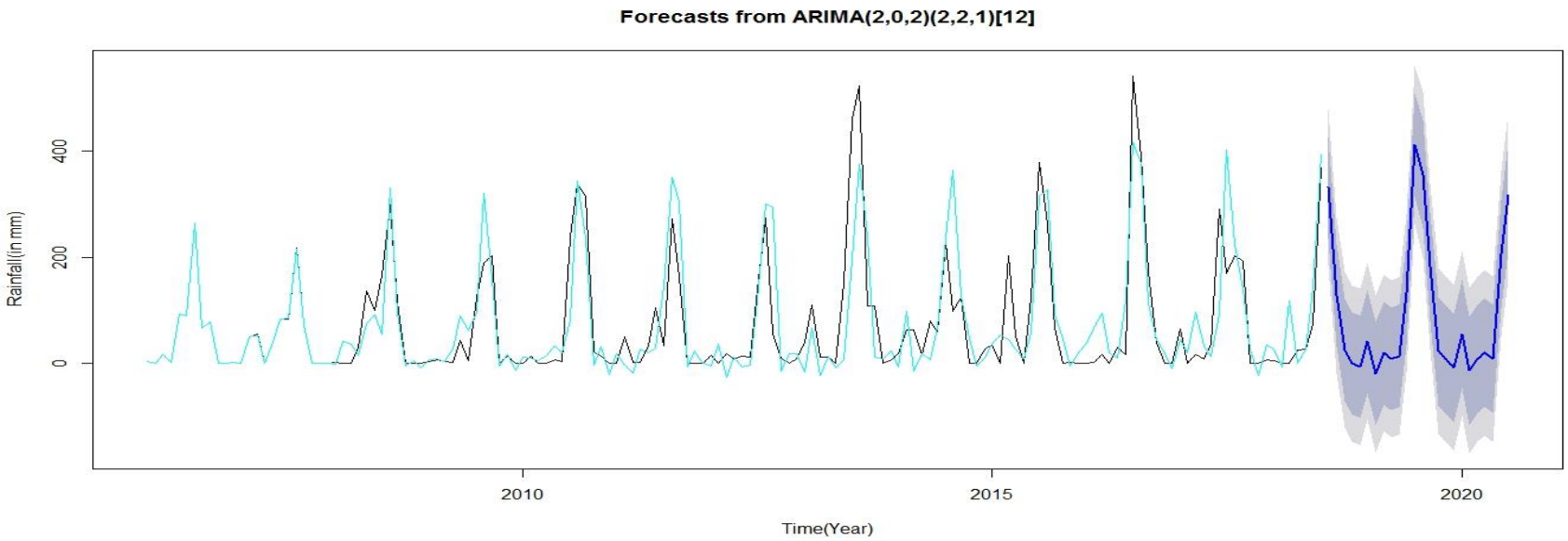


Fig. Forecast using SARIMA .

The prediction intervals are computed for 80% and 95% .
The model selected for the data is with:
$(2,0,2)(2,2,1)_{12}$

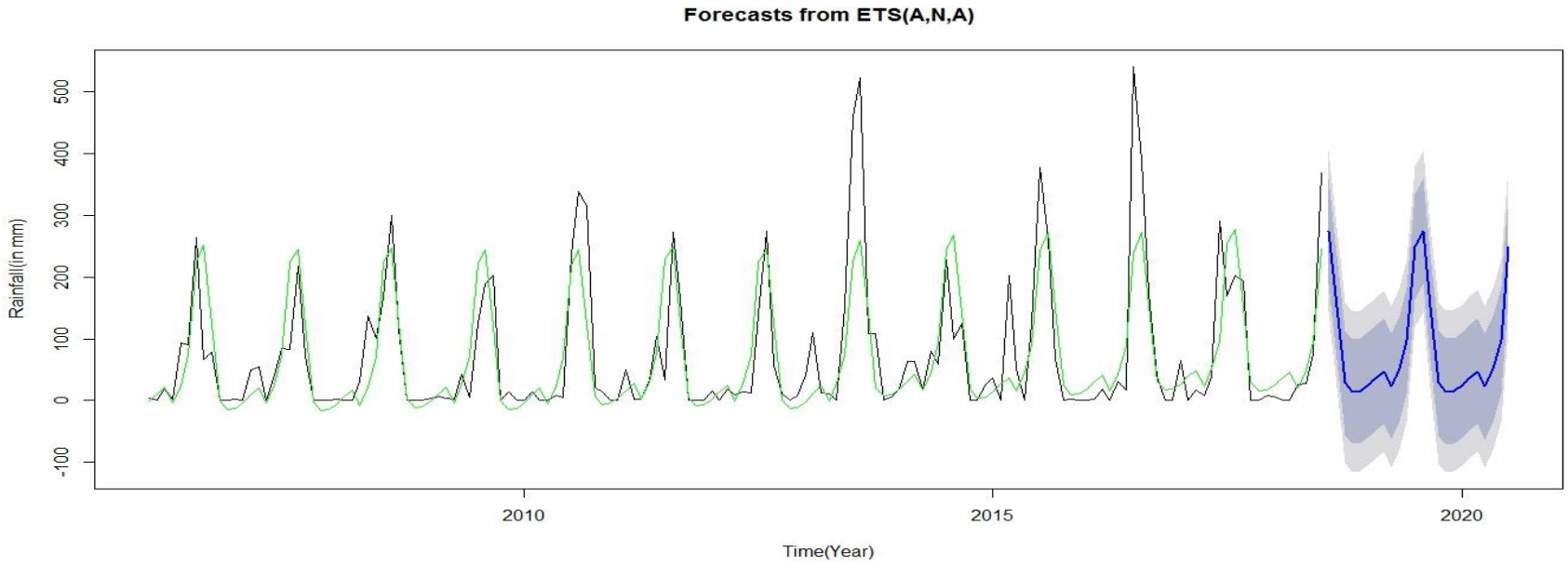


Fig. Forecast using Holt Winter`s method with the predictive interval of 80% and 90%.

The model selected for the data is :
$\alpha = 0.0285$, $\beta = 0$, $\gamma = 0.0001$

Residual Analysis

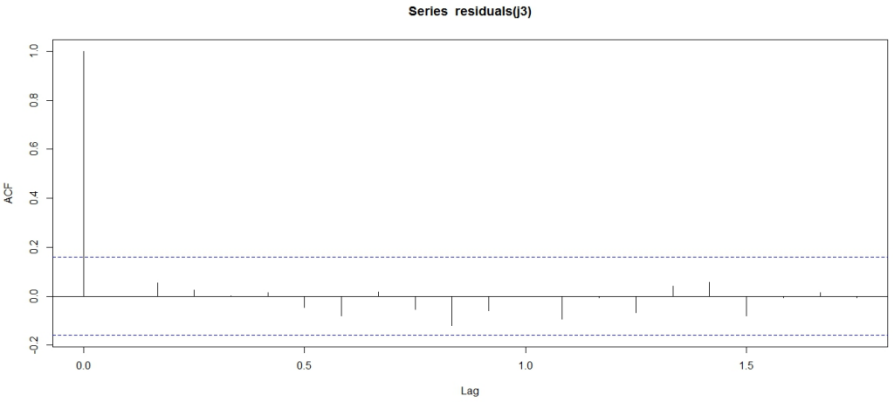


Fig. Autocorrelation plot for residuals obtained from SARIMA model

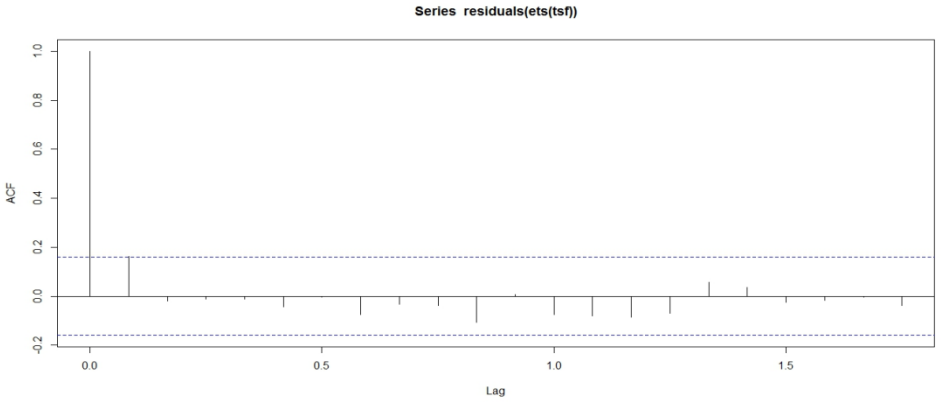
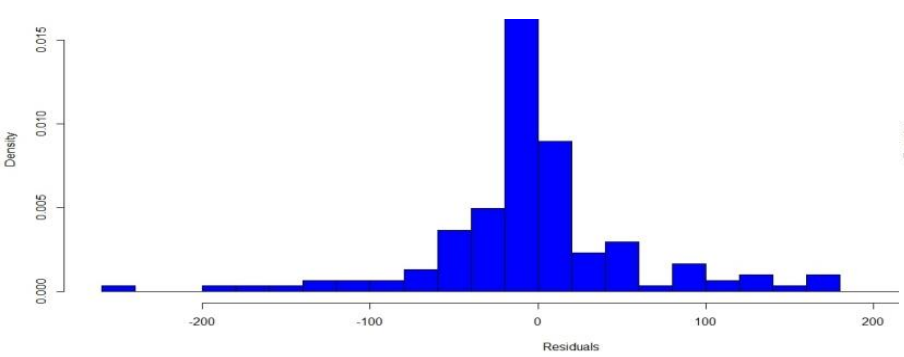


Fig. Autocorrelation plot for residuals obtained from HW model



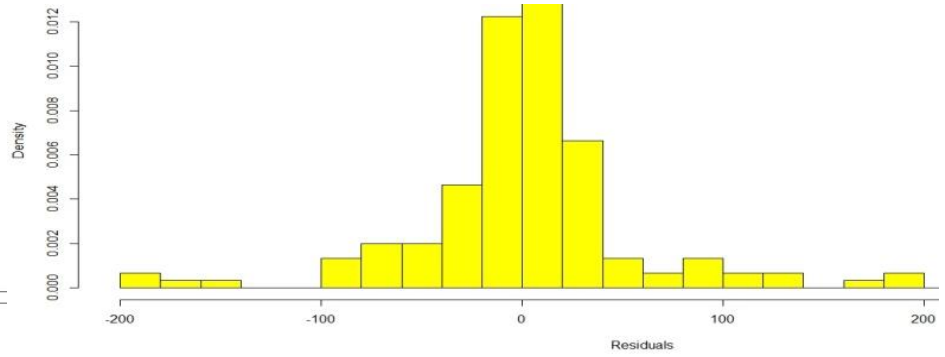


Fig. Histogram plot of residuals obtained from SARIMA and the HW models

- The ACF plot for residuals shows that the correlation is insignificant amongst the residuals indicating that there is no information left in the residuals that could be used for improving the model generated from the SARIMA and the HW method. Also, the histogram suggests that the residuals follow a normal distribution withapproximately zero mean.

Dickey-Fuller Test

Table 3. Summary of the ADF test.

SARIMA	Holt-Winters
Dickey-Fuller Test	Dickey-Fuller Test
Dickey-Fuller test value = -4.797 P-value = 0.01 Alternative hypothesis: stationary	Dickey-Fuller test value = -4.909 P-value = 0.01 Alternative hypothesis: stationary

Results of KPSS test for stationarity

KPSS Test for Level Stationarity

Data: Delhi Rainfall

KPSS Level = 0.0134, p-value = 0.1

KPSS Trend = 0.0125, p-value = 0.1

Here we have that the p-value is greater than 5% level of significance indicating that the series is stationary.

Model Selection Criteria

- Akaike Information Criterion(AIC) offers a relative estimate of the information lost when a given model is used to represent the process that generates the data.

Method Selection Criteria

• One can make a comparison between the model by comparing the MAE values of the model.

SARIMA	HW
MAE	**MAE**
38.018	40.215

SARIMA Model

- A Seasonal ARIMA(SARIMA) model is formed by including additional seasonal terms in the ARIMA models.
- Does not assume any particular pattern in the historical data of the series to be forecast.
- Formulation is : **ARIMA $(p, d, q)(P, D, Q)_s$**

Holt-Winters Forecasting

- It is further extension to time series which involves both trend and seasonal variations.
- comprises the forecast equation and three smoothing equations — one for the level l_t , one for trend b_t, and one for the seasonal component denoted by s_t .

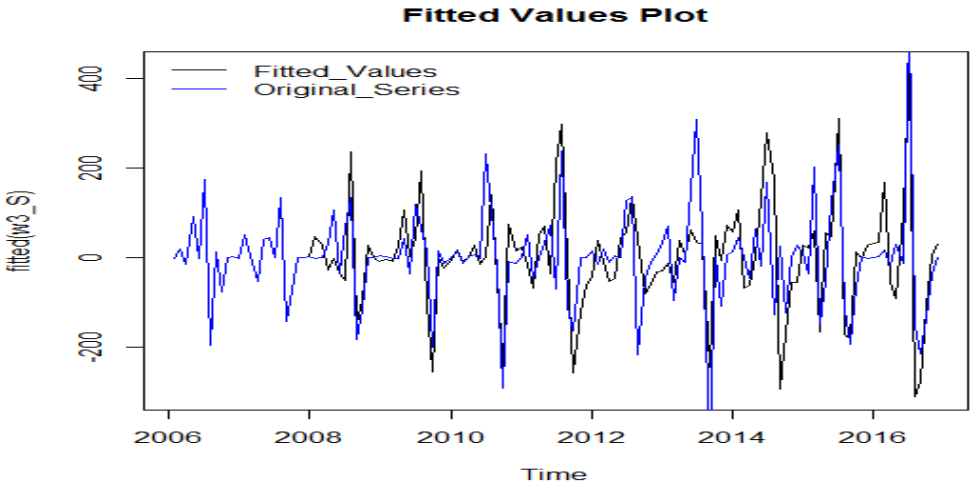


Fig. Plot of fitted values of SARIMA model over the original time series

The model selected for the data is :
$(2,0,2)(2,2,1)_{12}$

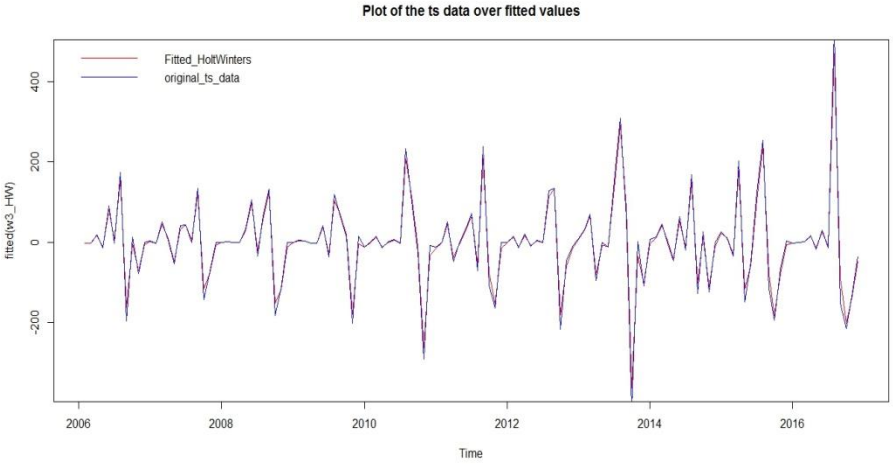


Fig. Plot of fitted values of Holt Winter`s model over the original time series.

The model selected for the data is :
$\alpha=0.0001$, $\beta=0.0001$, $\gamma=0.0001$

Conclusion

- The report includes two methods of time series analysis – SARIMA and Holt Winters for analyzing the variations in the historical rainfall data of New Delhi, India using R software.
- Further a comparison has been drawn between the results obtained from these two methods.
- Since the SARIMA model has MAE less than that obtained from Holt-Winters model. We concluded that the SARIMA model fits well for our dataset.

Future work

The aim would be to:

- Improve the maximum accuracy of the prediction.
- Investigate which measures are the best in reflecting the performance of existing models.
- Taking other factors into consideration while modeling the precipitation dataset.

THANK YOU

harpreetkaurguglani@gmail.com