

# Time series analysis of rainfall for the state of Odisha

Cite as: AIP Conference Proceedings **2435**, 020051 (2022); <https://doi.org/10.1063/5.0083522>  
Published Online: 18 March 2022

Rajni and Sudip Patra



View Online



Export Citation

## ARTICLES YOU MAY BE INTERESTED IN

### 2-Dimensional study of ion size effects on nanoflows

AIP Conference Proceedings **2435**, 020026 (2022); <https://doi.org/10.1063/5.0083519>

### Linear shoaling and linear dispersion using Boussinesq type equations for water wave propagation in nearshore

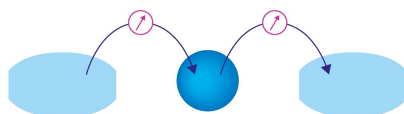
AIP Conference Proceedings **2435**, 020032 (2022); <https://doi.org/10.1063/5.0083751>

### A method for singular weakly linear Volterra-Integro-differential equations by Euler polynomials

AIP Conference Proceedings **2435**, 020048 (2022); <https://doi.org/10.1063/5.0083523>

Webinar

Interfaces: how they make  
or break a nanodevice



March 29th – Register now



Zurich  
Instruments



# Time Series Analysis of Rainfall for the state of Odisha

Rajni<sup>a)</sup> and Sudip Patra<sup>b)</sup>

*Jindal Global Business School, O. P. Jindal Global University, Sonapat, Haryana, 131001, India,*

<sup>a)</sup> *Corresponding author: rajni@jgu.edu.in*

<sup>b)</sup> *Email: spatra@jgu.edu.in*

**Abstract.** The rainfall data analysis and its timely prediction is very important for agricultural practices planning. In case of heavy/scanty rainfall prediction, early forecasting helps to plan for disaster management in areas of high risk. In this article, time series analysis of rainfall for the state of Odisha is carried out. Odisha is a state which is constantly being hit by waves of droughts, floods and cyclones. Thus, it becomes extremely important to analyze and forecast for rainfall in this region so that proper measures to avert disasters can be undertaken. For time series modeling, rainfall data for last 50 years has been analyzed using Auto-Regressive Integrated Moving Average (ARIMA) method. Several suitable ARIMA models are identified based on significant parameters. The variability in prediction and actual provides for the consistency of model. The months of highest and lowest precipitation have been identified. Later on the efforts to identify regions in Odisha for scanty and heavy rainfall has been made and predictions are made for 10 years into the future time period. It can be seen that accurate predictions help for crop planning and adopting better agronomic practices for crops, proper disaster management in risk zones and providing timely relief to those in affected areas.

## INTRODUCTION

Climate change is occurring worldwide, and fluctuations are happening in different regions of the country, India. The change in pattern of climatic conditions has vast effect on agriculture. India is a country which depends on rainfall for most of its agricultural requirement of water. Odisha is one such state heavily relying on rainfall for irrigation of its crops. It is the 8<sup>th</sup> largest state in India by area. Odisha lies between the latitudes 17.780N and 22.730N, and between longitudes 81.37E and 87.53E. The state has an area of 155,707 km<sup>2</sup>, which is 4.87% of total area of India, and a coastline of 450 km. Apart from industry, agriculture forms a major part of its economy. Odisha also battles many natural calamities almost every year. Almost every few years the state is affected by floods, droughts, and cyclones in different regions. Due to changes in climate pattern, the intensity of the natural calamities has increased which needs to be investigated. Cyclones, floods, and droughts are ultimately linked to rainfall in the region which needs a thorough investigation. The essential source of water supply is the Southwest monsoon which enters this region in second week of June and continues till first week of October [1]. The state receives 80% of its rainfall from middle of June to end of September. But the pattern of distribution of this rainfall is unpredictable. It been studied by few researchers that the number of days of rainfall also might reduce together with a marginal increase of 7–10% in its annual rainfall by 2050 for India. The number of seasons has decreased from six to two in the region. Also, the number of days of rainfall has reduced by 30 i.e., from 120 to 90 days [2]. If a model is developed for efficient prediction of rainfall in the region, it would lead to better water resource management. Deviating away from it may lead to crop failures, or else lead to severe damage. The current study investigates the time series modeling of Odisha for the whole year from 1960–2015. Seasonal ARIMA modeling is used for this time series analysis. An effort to forecast for next ten years has been done and presented. The ARIMA method is used because of the characteristic of the variable considered in this research, i.e., rainfall. Rainfall dataset is stationary. Several studies have been conducted using ARIMA for forecasting tourism, COVID-19 pandemic, yield production of crops, climate variability [3, 4] using ARIMA modeling. Box and Jenkins [5] first used ARIMA modeling in their research work. Haines et al. used it to model birth data [6]. Several researchers have recently undertaken investigation in this emerging area. Praveen et al. [7] analyzed the rainfall changes occurring in India with respect to trend and made forecasting using parametric and machine learning approaches. Dimri et al.

[8] studied the seasonal analysis of the monthly mean minimum and maximum temperatures and the precipitation for the Bhagirathi river basin in India.

## METHODOLOGY

**Data Collection:** The rainfall dataset (January-December) starting from 1960 to 2015 was obtained from the portal of data.gov.in. The data has been compiled by Indian Meteorological Department (IMD). Figure 1 shows the rainfall data for the state of Odisha from 1960-2015.

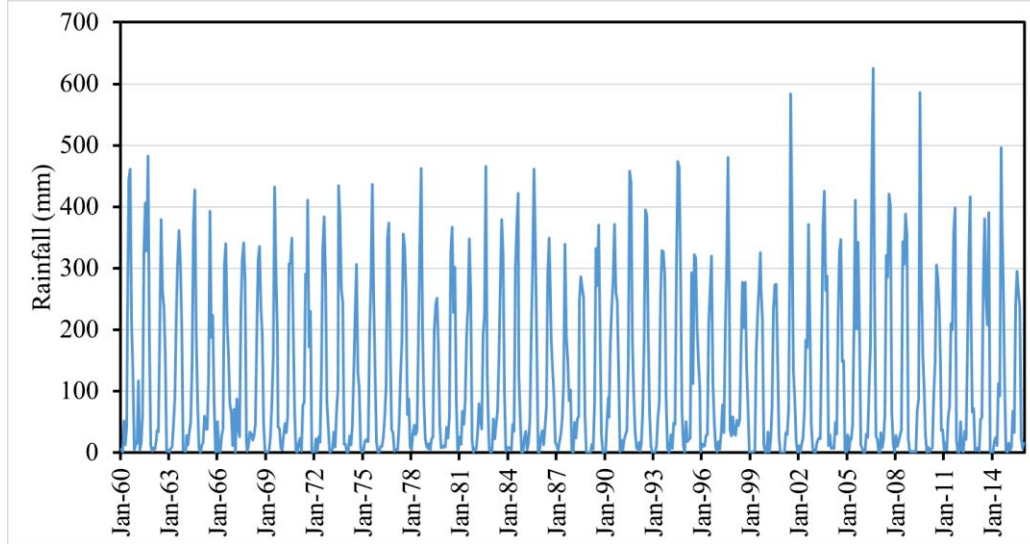


FIGURE 1. Rainfall in the state of Odisha from 1960-2015.

An Autoregressive Integrated Moving Average (ARIMA) model is used for this time series analysis. Box and Jenkins introduced ARIMA modeling in their research work in 1976 [5]. It works on time series which are stationary, i.e., the mean, variance and covariance are constant over time. While working with a non-stationary time series, it first needs to be converted to a stationary time series to apply ARIMA. This is because the classical regression results are invalid in non-stationary time series. The Augmented Dickey-Fuller (ADF) unit-root test is used to identify whether a time series is stationary. ARIMA, AR refers to Autoregressive models, MA refers to Moving average models and I refers to the number of lags used in differencing the data (applicable for non-stationary time series). The mathematical form of ARIMA model is shown in Equation. (1), respectively.

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} - \dots - \omega_q \varepsilon_{t-q} \quad (1)$$

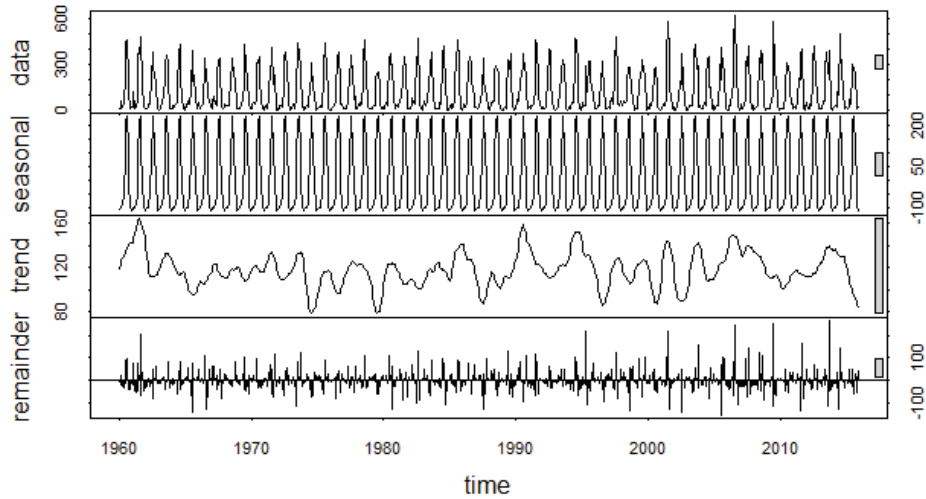
where  $\phi_t$  are auto-regressive coefficients to be estimated,  $p$  Number of lags. “ $p$ ” is a parameter for the model,  $\omega_q$  are moving average coefficients which are to be estimated and  $\varepsilon_{t-q}$  are the error terms. “ $q$ ” represents the number of error terms used in model.

SARIMA: ARIMA (p, d, q)×(P, D, Q): it represents a p-order autoregressive model with q-order moving average having d times differenced time series, P represents autoregressive order for seasonal, D represents seasonal differencing and Q represents seasonal moving average order. ARIMA modeling involves identification of the suitable models, then estimating their AIC values and finally selecting and fitting the best model after diagnostic checking and use it for forecasting.

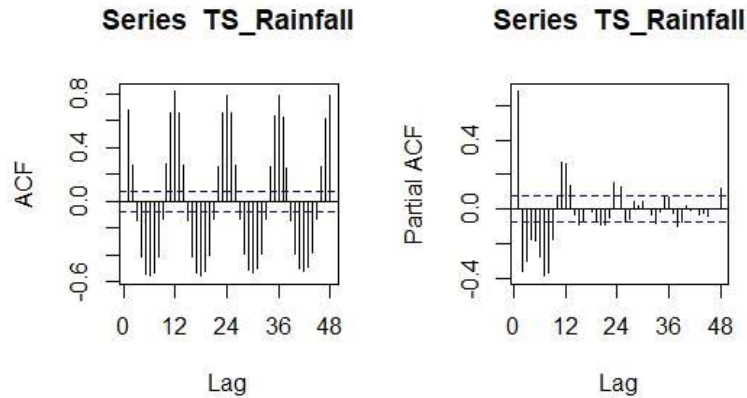
In the first step, the dataset is checked for stationarity first and then autocorrelation (ACF) and partial autocorrelation function (PACF) are examined. The differencing order ‘d’ is obtained from stationarity test, Auto-Regression and Moving Average orders are identified from the plots of ACF and PACF. Moving to second step, suitable models are examined by estimating a suitable criterion parameter. Then the identified models are compared with their suitable criterion parameter, in this case Akaike information criterion (AIC) values. After selecting a best fitting model, it is checked for autocorrelation, homoscedasticity, and normality. Finally, the model is ready to forecast future time periods.

## RESULTS AND DISCUSSION

The time series analysis is conducted on the R platform. Apart from R package, “tseries”, “urca”, “forecast” and “fpp2” was also used simultaneously to obtain the output for ARIMA [9,10]. The package has been developed for conducting forecasting analysis of time series data. All the tests and processes required for conducting the analysis is contained in this package. The rainfall data is analyzed for 55 years from 1960-2015. The data from 2016-2017 is used as the test set for checking the validity of the model. The rainfall data was disintegrated into its components to better understand the seasonal effects. The decomposition of rainfall data into its components is shown in Figure 2.



**FIGURE 2.** Components of rainfall data of Odisha.



**FIGURE 3.** ACF and PACF plots for the rainfall dataset.

The time series is stationary as confirmed by ADF test and KPSS test. Then moving onto the identification of the model, it was done using ACF and PACF plots (as shown in Figure 3) of the rainfall data. It is observed here that at lag 12, the strongest positive correlation comes, which is after negatively correlated lags (3 to 9). Thus, 12 is the appropriate seasonal parameter for the model. The model follows MA(0) process. From the plot of partial autocorrelation, it is observed that a strong cutoff occurs in the correlation at lag 1. It means the rainfall dataset follows an AR(1) process and thus  $p=1$ .

Also seasonal differencing of order 1 was applied to the model. The Ljung-Box test is also performed for Lags at 4, 8 and 12 as the data is seasonal. It is found that the model is free of autocorrelation. The best fit model out of the

numerous models identified is ARIMA(1,0,0)×(1,1,0) [12]. This was identified using the parsimony principle and minimum AIC values, and which passed the test of randomness. The residual plot of the ARIMA model is shown in Figure 4. This model is then used for forecasting rainfall until 2026 which is shown in Figure 5.

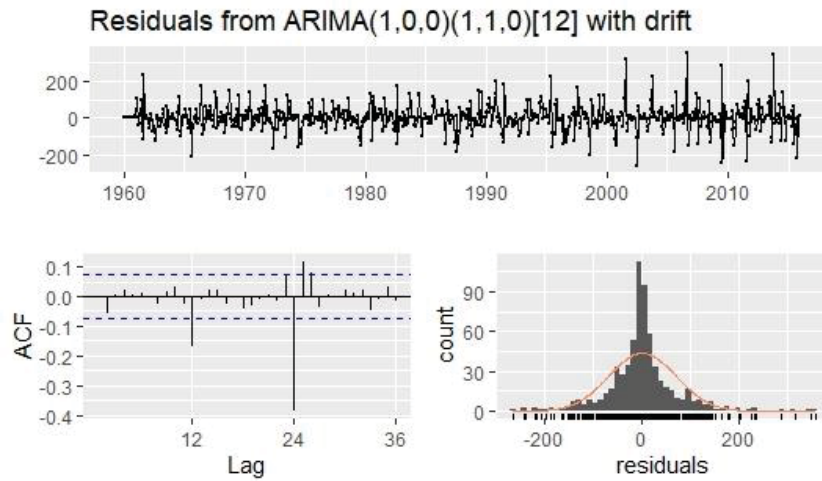


FIGURE 4. Residual plot for diagnostic checking

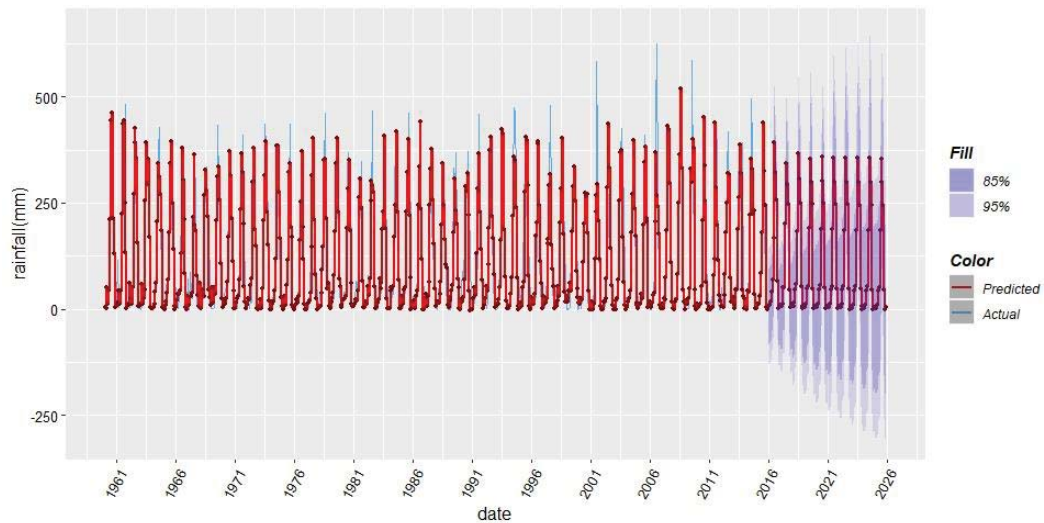


FIGURE 5. Forecasting of rainfall for 10 years from 2016-2025.

July receives the highest average rainfall with 358.89 mm. August receives second highest rainfall with average of 301.54 mm. The month of November receives the lowest average rainfall, i.e., 1.36 mm. Figure 6 shows the plot of actual and predicted values of rainfall from 2016-2017.

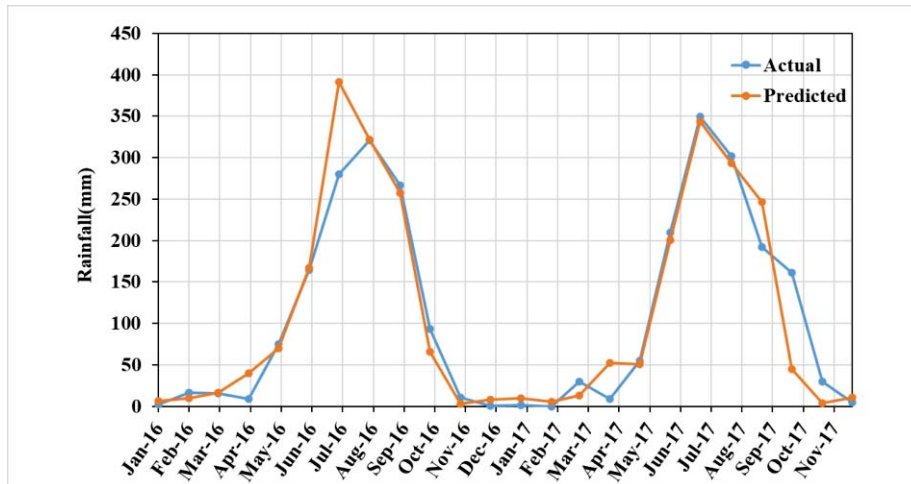


FIGURE 6. Comparison between actual and predicted values from 2016 to 2017.

## CONCLUSIONS

An ARIMA model is applied for analysis and forecasting of rainfall dataset for the state of Odisha. The model is identified, tested for validity, and then used to predict for next 10 years rainfall. It is observed that ARIMA is one of the best methodologies to model rainfall forecasting. More investigation in this regard needs to be conducted, such as dividing the state into regions, or district wise, or grid wise and studying rainfall pattern to have more better understanding with respect to region-wise. The current analysis will serve as a foundation for future investigation in this area.

## Appendix

The following section describes the set of codes used in R for conducting the time series analysis.

---

```
library(forecast)
```

```
library(tidyverse)
```

```
library(fpp2)
```

```
extrafont::loadfonts(device="win")
```

```
library(ggpubr)
```

```
library(dplyr)
```

```
library(hrbrthemes)
```

```
library(tseries)
```

```
library(urca)
```

```
library(ggfortify)
```

```
library(ggthemes)
```

```

library(zoo)

options(scipen = 999)

##Reading the CSV file

Rainfall_Raw <- read.csv(file.choose(), stringsAsFactors=FALSE)

TS_Rainfall=ts(Rainfall_Raw$Rainfall..mm., frequency=12, start=1960)

Rainfall_Test<-read.csv(file.choose(), stringsAsFactors=FALSE)

TS_Rainfall_Test=ts(Rainfall_Test$Rainfall..mm., frequency=12, start=2016)

##stationarity test

adf.test(TS_Rainfall)

kpss.test(TS_Rainfall,null="Trend")

## Stationarity Test

ndiffs(TS_Rainfall)

## Identification of Model by plotting ACF and PACF plots

par(mfrow=c(1,2))

Acf(TS_Rainfall, 48)

Pacf(TS_Rainfall,48)

After identification of models, the residuals are checked.

Model1 = arima(TS_Rainfall, order =c(1,0,0,1,1,0) [12])

Model_1_residuals=residuals(Model1)

Box.test(Model_1_residuals, lag = 48, type = "Ljung-Box")

checkresiduals(Model1)

adf.test(Model_1_residuals)

Model_test1=arima(training,order =c(1,0,0,1,1,0) [12])

forecast_1 = forecast(Model_test1,h=120)

accuracy(forecast_1,test)

##After checking residuals, the plots were generated.

```

## ACKNOWLEDGEMENTS

The research work was funded by Jindal Global University Research Grant: JGU/RGP/2019/001 and received support from Jindal Global Business School.

## REFERENCES

1. Central Soil and Water Conservation Research and Training Institute, CSWCRTI Vision 2030 (2011).
2. S. Mishra, D. K. Swain, D. Mishra, and G. H. Santra, (2021), Intelligent and Cloud Computing: Smart Innovation, Systems and Technologies, vol 153. Springer, Singapore.
3. D. R. Kothawale, J. V. Revadekar, & R. Kumar, *J Earth Syst Sci* **119**, 51–65 (2010).
4. P. Narayanan, A. Basistha, S. Sarkar, S. Kamna, *Comptes Rendus Geoscience*, **345** (1), 22-27 (2013).
5. G. E. P. Box, G. M. Jenkins, Time Series Analysis, Forecasting and Control, San Francisco: Holden-Day (1976).
6. L.M. Haines, W.P. Munoz, C.J. Van Gelderen, *J. Appl. Stat.* **16**, 55–67 (1989).
7. B. Praveen, S. Talukdar, Shahfahad *et al.* *Scientific Reports* **10**, 10342 (2020).
8. T. Dimri, S. Ahmad, and M. Sharif, *J. Earth Syst. Sci.* **129**, 149 (2020).
9. R. H. Shumway, D. S. Stoffer, Time Series Analysis and Its Applications with R Examples, Fourth ed, Springer (2017).
10. R. J. Hyndman, G. Athanasopoulos, *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2 (2018).