Stochastic Modeling and Applications
Vol. 25 No. 2 (July-December, 2021)
ISSN: 0972-3641

Received: 27th July 2021 Revised: 25th October 2021 Selected: 30th November 2021

FORECASTING THE RAINFALL USING TIME SERIES MODEL

M.SARANYADEVI[1.], AND A. KACHI MOHIDEEN[2].

Abstract

Forecasting measure, which is used to predict the future as accurately as possible. In forecasting, the accuracy of results attains when all the information about the data set is available, including historical data and knowledge of any future events that might impact the forecasts. The appropriate forecasting methods which are depends on the available data sets. Most of the quantitative prediction problems using the time series models which is used to estimate how the sequence of observations will continue into the future. ARIMA models and Exponential Smoothing models are the most comprehensive methods of time series forecasting and which is also giving the corresponding methodology to the problems. While Exponential smoothing models are results which is based on the trend and seasonality of the given data set and ARIMA models are used to describes the auto-correlation in the data.

In this paper, the rainfall prediction has been attempted for Tamil Nadu state using 14 years of rainfall. The data are taken month wise for the period of 2006-2019 and also used 5 years of testing data (2015-2019). For this study, the time series models Auto Regressive Integrated Moving Average (ARIMA) and Exponential Smoothing are used to predict the data and the performance of the models compared using mean square error and correlation coefficient.

Keywords: Time series modeling, Autoregressive Integrated Moving Average, Rainfall, and simple exponential smoothing.

INTRODUCTION

Tamil Nadu is the state which is based on Agriculture. For agricultural activities water is the main source for its success. There are many water resources available such as Bore wells, canals and rivers which are used for agricultural purposes. The rainfall is major resource for all other resources. Hence, the rainfall is the greatest asset and is most essential one for agriculture. The prediction of rainfall is vital risk due to inception and non-uniformityand its accuracy will provide benefits to various fields. Many of the researchers study the prediction of rainfall for the state Tamil Nadu. Time Series Analysis is a specific way of analyzing a sequence of data set collected over an interval of time. Time Series models are appropriate models to predict the metrological phenomenon such as Rainfall, Humidity, Temperature, Drought etc., and environmental management field.

This study aimed to forecast Rainfall of Tamil Nadu state. Monthly rainfall forecasting plays an important role in the planning and management of the agricultural scheme and water resource management. In this paper, the most appropriate approaches such as ARIMA (Auto Regressive Integrated Moving Average) and Exponential Smoothing methods are used to predict the rainfall data and also find the best model to predict the rainfall data set with maximum accuracy and which is minimizing the error value. For this study, 14 years of Rainfall data of Tamil Nadu (2006-2019) is considered.

In this work, classical statistical seasonal ARIMA model and Exponential Smoothing methods are used. Those model analyses the data and forecast the rainfall for future years in Tamil Nadu. Both the models are compared and results the best model found with accuracy which minimizes the error. This study may helpful for future planning of agriculture and predict the food production of the state.

REVIEW OF LITERATURE:

In this review study we discussed about previous works done in the topic time series models for forecasting, especially ARIMA and Exponential Smoothing and thereby, it helps the researchers for identifying an actual theoretical and methodological contribution to the topic. It helps to carry out the appropriate research work and enables them to conclude. As a part of the investigation, the research must experience the research work already done on similar problems by going through professional journals, research articles, and other literature.

In this section, we will discuss some related studies and recall the most important findings of these studies. Mushin et al., (2012) used the Box-Jenkins methodology to build a seasonal ARIMA model for monthly rainfall data taken for Dhaka Station for the period from 1981-2010 (June). The selected ARIMA (0,0,1) (0,1,1)12 model gives us two years of predicted monthly rainfall along with their 95% confidence interval that can help decision- makers to establish strategies, priorities, and proper use of water resources in Dhaka.

Meenakshi Sundaram (2014) aimed to predict the future of rainfall. Many of the researchers focused on rainfall prediction used many of time series models, they have considered the models for different countries, states and districts. O.N. Dhār et.al. in this contribution the authors revealed the average rainfall of northeast monsoon using standard methods and also evaluated the trend of the data set, periodicities, and variability for the actual rainfall. O.N. Dhār and Rekecha, correlated two different season rainfall which are southwest and northeast. The study reveals that the time series model SARIMA is a very useful tool for the prediction of rainfall. And this researcher, consider the MAPE value which is very less. And then fitted the SARIMA model (0,1,1) (0,1,1)12.

Yamoah et al., (2016) study the rainfall pattern of Brong Ahaf (BA) region of Ghana and found the best models for the given set of data. In their study, they have considered the rainfall data from 1975-2009 (Source: Department of Meteorology and Climatology, BA region). The study resulted that the region getting maximum rainfall during the months September and October and getting rain fall which is less than normal during January, December, and February months. They have evaluated the model SARIMA (0,0,0) (1,1,1)12 which is given minimum AIC score of 8.985894 and they have stated that the SARIMA (0,0,0) (1,1,1)12 models is an appropriate model to predict the monthly average rainfall of the region Brong Ahafo Region of Ghana.

Ramesh Reddy et al., (2017) are used Seasonal Autoregressive Integrated Moving Average (SARIMA) model to analyze the monthly average rainfall Rayalaseema (India) using R language. In this study, they disclosed the results which is based on AIC and BIC measures and the model which is giving minimum measures that is the best model for the given data set. Finally, they fitted best model using ARIMA (5,0,1) (2,0,0)12 and the model resulted the data seasonality.

Sidiq (2018) aimed to forecast rainfall using the Time Series model. Monthly rainfall of 48 data got from Badan Metrologic Dan Geophysical (BMG) Bandung from January 2011 to December 2013 is processed by a computer program to look at the pattern in Minitab model ARIMA. In ARIMA model, Box-Jenkins method used to study the rainfall data which are taken from the year 2011-2013. By using the measures MAE, MAPE and ARIMA, the authors concluded that the ARIMA models are resulted the accurate prediction of rainfall for future years. ARIMA models are resulting the accurate prediction of rainfall, so it can be suggested for planning agriculture and plantations which are based on rainfall data.

Kamath and Kamat (2018), this study exhibited the performance various Time-series analysis models and compared its forecasting accuracy. Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Network (ANN), and Exponential Smoothing State Space (ETS) models are compared. The time series models are compared and fit the best model to the rainfall data which are taken from the year 2006 to 2016 for the district Idukki on monthly basis. This paper resulted the best time series model for the given data set is ARIMA model and the model outperformed with minimum Root Mean Squared Error (RMSE) with maximum accuracy.

Dwivedi et al., (2019), in the work ARIMA models are used to simulate and forecast the rainfall using its linear approach and it was compared with ANN. In this study, considered the rainfall for 32 years which are from 1980-2011 for Junagadh and also takenthe testing data for 5 years which are from 2012-2016.

Refonaa et al., (2019) proposed linear regression analysis to predict the monthly rainfall data. The prediction of Rainfall is used to plan the future economic of the state using the current atmospheric state. Various models are used to predict the rainfall data with greateraccuracy with minimum error values.

METRICAL METHOD

In this study, we considered secondary data which is taken from the official website indiastat.com (which contains the reliable data sources). The rainfall always measures in the unit millimeters. Month-wise rainfall considered from January 2006 to December 2019.

OBJECTIVES:

The objective of this study is given below:

- To find the best time series model which gives the maximum accuracy with minimalerror.
- To compare two-time series models for the rainfall data and compare the modelsnumerically and graphically.
- For both models, error values are calculated and compared.

EXPONENTIAL SMOOTHING METHOD:

Exponential smoothing method is a time series forecasting method which is calculated for univariate data which are having systematic trend or seasonal component (i.e.,) seasonality. Exponential smoothing methods are more effective methods which is an alternative model for Box-Jenkins ARIMA models.

In time series methods, Box-Jenkins ARIMA methods are used to improves the models, which are followed the linear sum of past observations or in the intervals (which are in missing values in intermediate level). Precisely, the observations are in decreasing nature geometrically.

There are three main types of exponential smoothing time series forecasting methods, that are single, double and triple exponential smoothing methods.

SINGLE EXPONENTIAL SMOOTHING:

Single Exponential Smoothing (SES) is a time series forecasting method for univariate data which are neither follows the trend nor seasonality. It has need of a single parameter, which is used notated that alpha (a) also called smoothing factor or smoothing coefficient.

SES calculates the smoothed series as a restraining coefficient period for the actual values. The extrapolated smoothed series is a constant, equal to the last value of the smoothedseries during the period when actual data on the underlying series are available. The simple moving average method is a special case of exponential smoothing; the exponential smoothing is more in its data usage. In exponential smoothing, a new estimate is the addition of the estimate for the present time and a portion of the error $(x_t - \hat{x}_t)$ generated in the present time, that is

$$\bar{X}_{t+1} = \hat{x}_t + \alpha(e_t) \qquad (1)$$

This equation is usually written as

$$s_t = s_{t-1} + \alpha(x_t - s_{t-1}) \qquad (2)$$

where s_t = the new estimated value or forecasted value for the next time period (made in thepresent time period);

s_{t-1} = the estimated or forecasted value for the present time period (made in the last time period);

x_t= the actual data value in the present time period; $x_t - s_{t-1}$ = estimating or forecasting error for the present time period;

α = a weight value or discount ranges between 0.01 and 1.00. The smoothing equation can be written as

$$\bar{X}_{t+1} = s_t + \alpha x_1 + (1-\alpha)s_{t-1} \tag{3}$$

Or in another way of smoothing equation can be written as follows:

Next period forecast= weight (present period observations) + (1−weight) (present period forecast).

The smoothing equation is constructed based on averaging (smoothing) of past values of a series in decreasing (exponential) manner. The observations are weighted, with the more recent observations are given more weight. The weights are α used for the most recent observation, α (1-α) for the next most recent observation, $\alpha(1-\alpha)^2$ for the next, and soon. A teach time for producing a new forecasted value, the weighted observation along with the weighted estimate for the present period is combined.

$$\bar{X}_{t+1} = \bar{\$} = \alpha x_1 + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \cdots + \alpha(1-\alpha)^{t-1}x_1 \tag{4}$$

or

$$\bar{X}_{t+1} = \bar{\$} = \alpha \sum_{j=0}^{t} (1-\alpha)^j . x_{t-1} \tag{5}$$

Since the double exponential smoothing can evaluate in linear trends with no seasonal pattern. The triple exponential smoothing can handle both trend and seasonal patterns in time series data. Figure shows the selection procedure of different exponential smoothing methods.

ARIMA Model

An Auto Regressive Integrated Moving Average (ARIMA) model is a standard statistical model for time series forecast and analysis. The two authors Box and Jenkins are proposed the model ARIMA in their seminal 1970 text book Time series analysis: Forecasting and Control. The ARIMA model consists of three steps such as Identification, Estimation and Diagnostic checking. It is an iterative process model. In the stage of Identification, assess whether the time series data is stationary, and if not, how many differences are required to make it stationary and also identify the parameters of ARMA model such as p, d and q (p-the number of lag observations included in the model called lag order, d-number of times that the raw observations are differenced is called degree of differencing and q- size of moving average window called order of moving average). In the second stage Estimation, it involves using numerical methods to minimize a loss or error terms. In the third stage Diagnostic checking, it checks whether the model is fit for the given data or not. There are two processes involved in diagnostic checking that are Overfitting and Residual Errors.

ARIMA model was introduced by Box and Jenkins [1], and they recommend differencing non-stationary series one or more times to obtain stationary. The term integrates disused because the differencing process can be reversed to obtain the original series. When the explanatory variables in a regression model are time-lagged values of the forecast variable, then the model is called an autoregressive (AR) model. The general form of an autoregressive model of order p denoted as AR (p) is

$$y_t = c + b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \cdots b_p y_{t-p} + e_t \qquad (6)$$

where at is the error or residual term and p is an integer denoting the order in which the observations in the time series are correlated. When a time series is analyzed using its dependence relation with the past error terms, a moving average (MA) model is applied.

The general form of the MA(q) model of order q is

$$y_t = c + e_t - \varphi_t e_{t-1} - \varphi_t e_{t-2} - \cdots - \varphi_q e_{t-q} \qquad (7)$$

Autoregressive (AR) model can be efficiently coupled with the moving average (MA) model to form a general and useful class of time series models called autoregressive moving average ARMA (p, q) models. However, it can be used only when the time series is stationary. When a time series is studied based on the dependence relationship among the time-lagged values of the forecast variable and the past error terms, an autoregressive integrated moving average (ARIMA) model is more appropriate. It can be used when the time series is non-stationary. The general form of the ARIMA (p, d, q) model is

$$y_t = c + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_p y_{t-p} + e_t - \varphi_1 e_{t-1} - \varphi_2 e_{t-2} - \cdots - \varphi_q e_{t-q} \qquad (8)$$

where p, d, and q represent, respectively, the order of an autoregressive part, the degree of difference involved in the stationary time series which is usually 0,1 or at most 2, and the order of the moving average part. An ARIMA model can be obtained by estimating its parameters. The values of p and q can be determined from the patterns in the plotting of the values of ACF and PACF. The spikes falling above the time axis are used to estimate the value of p. The spikes falling below the time axis are used to estimate the value of q. For an AR (p) model, the spikes of ACF decay exponentially other is a sine wave pattern and the spikes of PACF are close to zero beyond the time lag q whereas the spikes of PACF decay exponentially or there is a sine wave pattern.

COMPUTATIONAL RESULTS

In this study, we computed the results for monthly rainfall data of Tamil Nadu. The effective results of this study calculated through the SPSS package. The data are examined using time series models (ARIMA and exponential smoothing models) and identified the appropriate model. This results may helpful for the decision makers to make policies.

Rainfall Rate in Tamil Nadu:

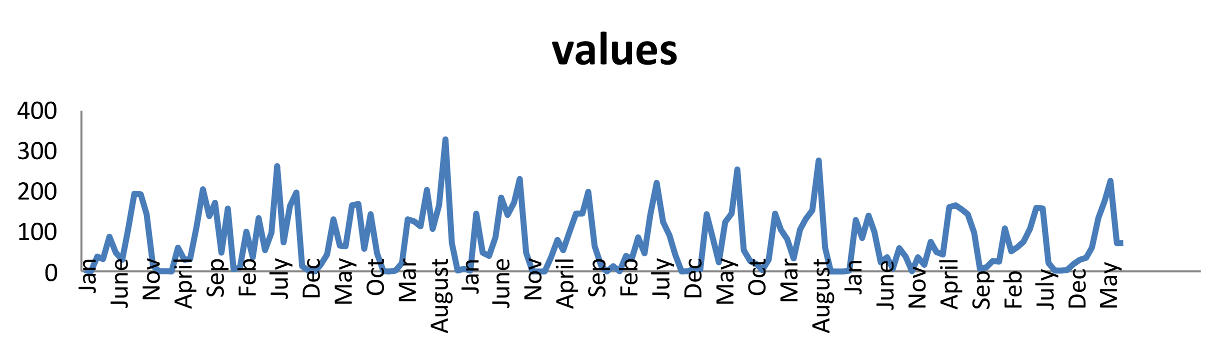


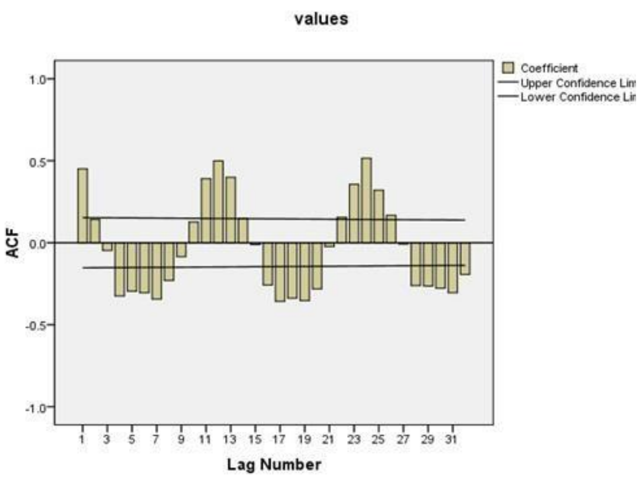
Fig1. Trend of actual rainfall data (in mm)

Fig.1 shows that the trend of rainfall data which is taken from 2006 to 2019.

Increasing and decreasing values are there over a period of time.

Identification of ARIMA Model

In ARIMA methods, in the stage of identification it checks the stationarity of the given data set. This process results that nature of the given data set is non-stationarity and in the first differencing process, the given rainfall data reaches the stationary and thus the values of were 1. The graphs of the sample ACF and PACF were plotted in figure 4 and Table-1



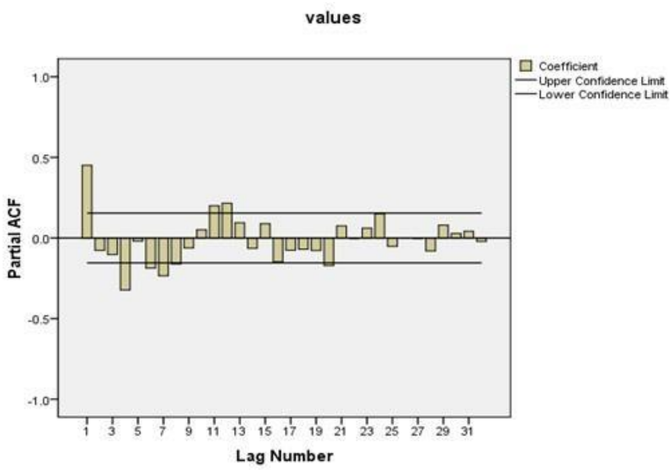


Fig. ACF of differenced data PACF of differenced data

Table 1. ACF and PACF of Rainfall

ations			Statistic		ocorrelation	
Lag	**ıe**	**ig.**	**ue**		**Value**	
	1	)76	'81	)	.451	
	2	)76	265	)	-.077	
	7	)76	550	)	-.103	
	5	)76	)94	)	-.322	
	)5	)76	359	)	-.020	
	)5	)75	'92	)	-.187	
	4	)75	794	)	-.235	
	0	)75	208	)	-.162	
	5	)75	500	)	-.061	
10	7	)74	414	)	.051	
11	1	)74	241	)	.199	
12	9	)74	912	)	.216	
13	9	)74	210	)	.094	
14	8	)73	283	)	-.064	
15	2	)73	308	)	.089	
16	8	)73	797	)	-.147	
17	8	)73	018	)	-.076	
18	8	)72	798	)	-.070	
19	3	)72	658	)	-.078	
20	2	)72	950	)	-.171	
21	3	)72	055	)	.075	
22	5	)71	756	)	-.005	
23	6	)71	772	)	.061	
24	5	)71	404	)	.150	
25	1	)71	964	)	-.052	
26	7	)71	605	)	.001	
27	)8	)70	618	)	-.003	
28	1	)70	524	)	-.080	
29	3	)70	781	)	.080	
30	7	)70	704	)	.028	
31	)5	)69	117	)	.042	
32	)3	)69	967	)	-.021	

In the above table, it shows that there is a trend of continuous decline in the autocorrelation coefficient and those values are significant which means the given time series data reaches the condition of stationarity. Based on the diagnostic test, the parameter values are identified that are 1,0 and 0 (i.e.) the range of ARIMA model is (1,0,0). Table.2 revealsthe diagnostic test of ARIMA (1,0,0).

Table 2: BIC values of ARIMA (p, d, q)

ARIMA (p, d, q)	IC Values
ARIMA (0,1,0)	8.630
ARIMA (0,0,1)	8.356
ARIMA (1,0,0)	**8.345**
ARIMA (0,1,1)	8.596
ARIMA (1,1,0)	8.617
ARIMA (1,0,1)	8.378
ARIMA (1,1,1)	8.410

Based on the above table, we decided that the parameters 1,0 and 0 are fit the bestmodel to given rainfall data (ARIMA (1,0,0)) which is having the minimum BIC value.

Table 3: Estimated ARIMA model of (1,0,0) for rainfall data

s	Estimate	E		-value
	77.340	799		.000
	0.451	069		.000

Table 4. Estimated ARIMA model fit statistics

Fitted Statistic	ARIMA (1,0,0)
R^2	0.203
R^2	0.203
RMSE	62.941
MAPE	739.769
MaxAPE	4.847E4
MAE	51.890
MaxAE	211.801
Normalized BIC	8.345

Table 5. Forecast the monthly rainfall data

onths	licted	CL	LCL
Jan	.3	98.5	-49.8
Feb	6	12.2	-60.2
Mar	5.7	15.3	-61.8
April	7.1	16.1	-62
May	7.2	16.3	-61.9
June	7.3	16.4	-61.9
July	7.3	16.5	-61.8
Aug	7.3	16.5	-61.8

	7.3	16.5	-61.8
	7.3	16.5	-61.8
	7.3	16.5	-61.8
	7.3	16.5	-61.8

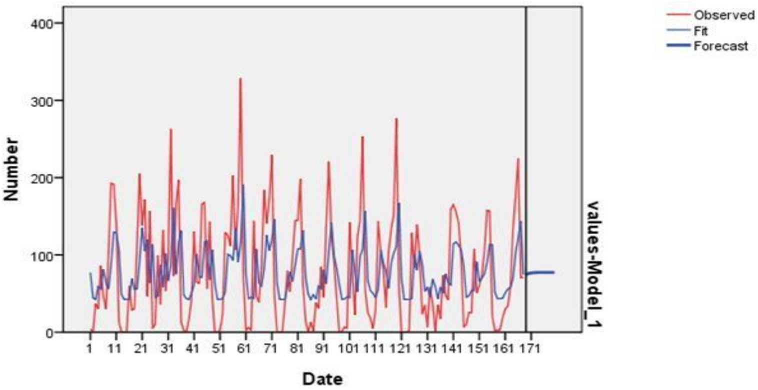


Fig. 2 Actual and forecast the rainfall for future years

In Table.5 the forecasted rainfall predicted using ARIMA(1,0,0) and Fig.2.explains the nature of the predicted rainfall values using the estimated rainfall with the observed values. From the above study, more fluctuations are reflecting in the actual rainfall data than predicted one. The predicted data reveals that Tamil Nadu state will more rainfall in future days.

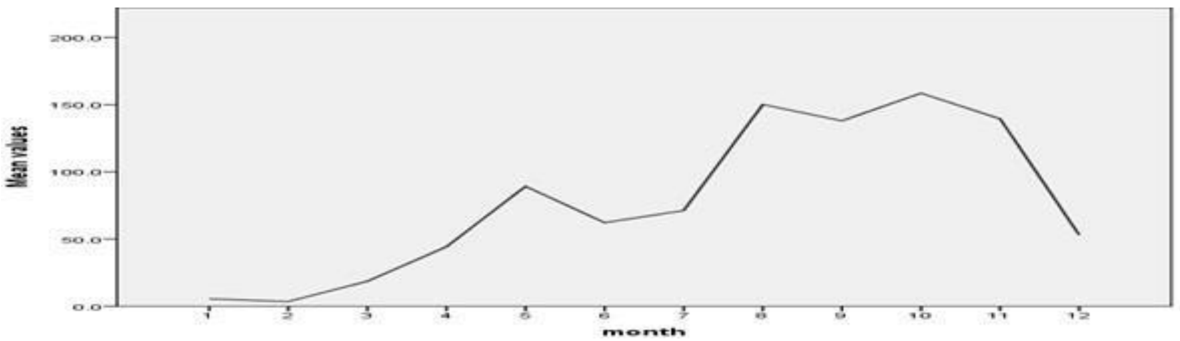


Fig 3. Rainfall prediction for actual rainfall data

In Fig. 3 the rainfall predicted using exponential smoothing method (in SPSS package). It exposed that the highest and lowest level of rainfall during the period 2006-2019. Using Table.6 it revealed that the lowest rainfall occurred in the month of February.

Table 6. Forecast Values in Exponential Smoothing

h			

Table 7. Model comparison

MODEL	RMSE
ARIMA	**62.941**
Exponential Smoothing	70.296

In Table.7 the two models of time series are compared using Root Mean Square Error (RMSE). The model which is having minimum RSME value it is the best fit model for the given rainfall data and also using the measure RSME performance of the model is evaluated. The RMSE value is minimum for the ARIMA(1,0,0) and it is giving the best fit for the given data set. So it is concluded that the ARIMA(1,0,0) model is outperforming when compared with exponential smoothing model.

Conclusion

The study successfully concluded that the prediction the level of rainfall for the future years, which results that the state Tamil Nadu will get increasing level of rainfall in future years. The two-time series models, ARIMA and exponential smoothing methods are compared and the best model found using the error value. Finally it is concluded that the ARIMA(1,0,0) which is giving the best fit for the given data set and also which is having minimum Root Mean Square Error (RSME) value when compared with exponential smoothing method with maximum accuracy.

Reference

1. Ebenezer Afrifa-Yamoah, Bashiru I.I. Saeed, Azumah Karim (2016) : "Sarima Modelling and Forecasting of Monthly Rainfall in the Brong Ahafo Region of Ghana". World Environment 2016, 6(1): 1-9 DOI: 10.5923/j.env.20160601.01
2. SK Khandar Babu, K. Karthikeyan, M.V. Ramanaiah and D. Ramanah (2011): "Prediction of Rain-fall Time series using Auto-Regressive Models". Available online at www.pelagiaresearchlibrary.com ISSN: 0976-8610 CODEN (USA): AASRFC
3. J. C. Ramesh Reddy, T. Ganesh, M. Venkateswaran, PRS Reddy (2017): "Forecasting of Monthly Mean Rainfall in Rayalaseema". International Journal of Current Research and Review DOI: 10.7324/IJCRR.2017.9244 ISSN: 0975-5241 (Online)
4. M. sidiq (2018): "Forecasting Rainfall with Time Series Model". IOP Conf. Series: Materials Science and Engineering 407 (2018) 012154 doi: 10.1088/1757-899X/407/1/012154
5. A. Geetha, G.M. Nasira (2016): "Time-series modelling and forecasting: modelling of rainfall prediction using ARIMA model. International Journal of Society Systems Science,2016 Vol.8 No.4, pp.361-372
6. Alivia Chowdhury and Amit Biswas (2016): "Development of a Monthly Rainfall Prediction Model Using Arima Techniques in Krishnanagar Sub-Division, Nadia District, West Bengal". International Journal of Engineering Studies and Technical Approach,Volume 02, No. 2, February 2016, ISSN No. 2395 - 0900
7. K. Senthamarai kannan and E. Sakthivel (2014) Fuzzy Time Series Model and ARIMA Model –A Comparative Study, Indian Journal of Applied Research, Volume: 4, Issue: 8, ISSN - 2249-555X
8. S. Meenakshi Sundaram under the guidance of M. Lakshmi (2014): "RainfallPrediction using Seasonal Auto Regressive Integrated Moving Average Model". PARIPEX- INDIAN JOURNAL OF RESEARCH Volume: 3, Issue: 4, April 2014ISSN - 2250 - 1991
9. Md. Mahsin, Yesmin Akhter, Monira Begum (2012): "Modelling Rainfall inDhaka Division of Bangaladesh Using Time Series Analysis ", Journal of Mathematical Modelling and Application 2012, Vol. 1. No.5, 67-73ISSN: 2178-2423
10. E Sakthivel and K Senthamarai Kannan (2014) Fuzzy Time Series Forecasting on Intuitionistic Fuzzy Sets, International Journal of Advanced Research in Computer Science and Software Engineering, volume 4, Issue 2.
11. P.E. Naill M. Momani (2009): "Time Series Analysis Model for Rainfall Data inJordan: Case Study for Using Time Series Analysis". American Journal of Environmental Sciences 5 (5): 599-604, 2009 ISSN 1553-35X@2009 Science Publications

M.SARANYADEVI: [1]Assistant Professor (GUEST), Department of Statistics, Government Arts College (Autonomous), Kumbakonam, Affiliated to Bharathidasan University, Tiruchirappalli – 620024, Tamilnadu, India.

A. KACHI MOHIDEEN: Assistant Professor, Department of Statistics, Periyar EVR College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli– 620024, Tamilnadu, India.