IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 13, Issue 6 Ver. III (Nov. - Dec. 2017), PP 57-63
www.iosrjournals.org

Rainfall Data Forecasting By SARIMA and BPNN Models

[1]Seneviratna,DMKN, [2]Rathnayaka,RMKT
[1]*Department of Interdisciplinary Studies, Faculty of Engineering, University of Ruhuna, Sri Lanka*
[2]*Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka*
Corresponding Author: [1]seneviratna, DMKN

Date of Submission: 12-12-2017 Date of acceptance: 22-12-2017

I. Introduction

Forecasting is very appropriate and important component in different ways due to the competitive atmospheres around the world. It is a significant methodfor inferring the future behavior of a particular field under the uncertainty conditions with the limited known behaviors. This scenario is well suited for many practical applications such as rainfall data, because of the importance of knowing about weather for agricultural purposes as well as for preventing natural disasters.Practically, rainfall data forecasting is a difficult task due to its' nonlinear behavior with high volatility and complex nature. Furthermore, rainfall is considered as the most important climatic element that influences agriculture. Therefore monthly rainfall forecasting plays an important role in the planning and management of agricultural scheme and management of water resource systems.

There are different kind of forecasting methods can be seen in the literature ([3], [5], [9], [10]). Basically, classical time series models such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) are mainly used in time series forecasting. Meher et al. has developed ARIMA model for simulating and forecasting monthly mean rainfall data in 38 rain-guage stations in India [6]. Based on the trend and seasonal behaviours of the data, they have identified ARIMA $(1,0,0)(0,1,1)^{12}$ as the most suitable model for simulating and forecasting. Also, the appropriateness and the accuracy of the model have been shown by considering 12 months ahead forecasting. ARIMA models are very useful especially the researcher has only the past data related to the particular variable other than the related factors. Also the importance of rainfall data forecasting is enormous as it is valid for decision makers, people who are involving in the agricultural fields and all other to avoid from natural disasters. Therefore, many researchers attempt investigate suitable models and accurate forecasting. Among the literature, SARIMA $(0,0,0)(0,1,1)_{12}$ was used by Mohamed et al. [7] to simulate monthly rainfall in Nyala station, Sudan.Soltaniet al. [12] and Etuk et al. [4] also used SARIMA model to fit the monthly rainfall data in Iran.

Nnaji [8] analyzed monthly rainfall data in 21 years for the six regions in Nigeria. Here, rescaled range statistic, the standard fluctuation analysis and the de-trended fluctuation analysis were applied and then pointed out that Nigeria exhibit a tendency to Self-Organized Criticality (SOC) in the short range. Also, he showed that the used standard fluctuation analysis method is not enough to have a good result compared to the other two. Feng etal. [5] studied and presented a method to model and investigate monthly, seasonal and annual distribution patterns, their trends and variability in rainfall.For this study, they have used trend analysis and ARIMA modelling. Based on these results, they demonstrated that how to analyze trend in a time series and how to find the suitable fitted model for the data.

Currently, a new trend can be seen in modeling and forecasting rather using only classical time series models due to the complexity with the trend and seasonal patterns of the rainfall data.Artificial Intelligence methods, Machine Learning methods are such an emerging fields in modelling and forecasting time series data. Among these different approaches, Sumiet al. [13] used machine learning models to forecast an average daily and monthly rainfall. In this study, they concentrated on three aspects such as modelling inputs, modeling methods and pre-processing techniques. Through this approach, a new hybrid multi-model was proposed and its performances were compared with its constituent models.

This study proposes anovelmethodology to forecast monthly rainfall data under the different scenario. The methodology consists of a hybrid model that is a combination of Three months Moving Averages (MA), Seasonal Autoregressive Integrated Moving Average (SARIMA) and Neural Network model (NN). Here SARIMAmodel is used to identify the future direction of the rainfall by considering the three months moving averages of monthly rainfall data. Then, the results are combined to the back propagation neural network (BPNN) modelling to forecast one step ahead monthly rainfall. Accordingly, the new hybrid model consists of two components such as direction and value forecasting. Here, the specialty of the proposed model is the usage of three months MA that smooth the data well and BPNN component that captures both linear and nonlinear patterns in the data. The empirical investigation of the proposed new hybrid method is conducted by using monthly rainfall data in two areas namely, Alupola and Ginigathhena, Sri Lanka over the period from January 2006 to December 2015 and January 2009 to December 2015 respectively. The forecasting was made to obtain one-step-ahead out sample value and evaluated by error measures including the mean absolute error (MAE), mean squared error (RMSE) and mean absolute percentage error (MAPE). Also, the performance of the proposed model is compared with the constituent SARIMA model.

II. Methodology

2.1Data

For this study, monthly rainfall data in two areas Alupola and Ginigathhena, Sri Lanka was considered over the period from January 2006 to December 2015 and January 2009 to December 2015 respectively. The two data samples contain 237 and 78 observations respectively. The data series are considered as raw data series and their 3-months moving average data series. Accordingly, this study consists of four data series as Series I to IV and they are raw data series of monthly rainfall in Ginigathhena area, 3-months moving average data series in Ginigathhena area, monthly rainfall in Alupolla area, and 3-months moving average data series in Alupolla area respectively.

2.2Modelling by SARIMA

If the time series does not show any other seasonal or cyclical behaviors, the stationary series can be used to determine the order of the autoregressive terms (AR) and moving average terms (MA). For this, we can use both autocorrelation function (acf) and partial autocorrelation function (pacf) by judging the corresponding acf and pacf plots.

By examining the Preliminary analysis, if we select a model as ARIMA (p,d,q); where p is the number of autoregressive terms, d is the number of differences of the original series and q is the number of moving average terms, the typical equation could be represented as,

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

([1], [2], [4]).

If the time series $\{X_t\}$ is seasonal of period s, then the series can be modelled as,

$$A(L)\Phi(L)\nabla^d \nabla^D_s X_t = B(L)\Theta(L^s)\varepsilon_t \quad (2)$$

Where $\Phi(L)$ and $\Theta(L)$ are called the seasonal AR and MA operators respectively. If they are polynomials of order P and Q respectively, then the time series $\{X_t\}$ is said to follow a seasonal autoregressive integrated moving average (SARIMA) of orders p, d, q, P, D, Q and s designated $SARIMA \ (p,d,q) \times (P,D,Q)_S$ model. Here, the operator ∇_s is the seasonal difference operator and D is the seasonal differencing order [4].

2.3Modelling by BPNN

In the feed forward neural network architecture, the artificial neurons are organized as layers.And here, the information strictly flows forwardfrom input layer to output layer. But, the errors of the network are propagated backwards and therefore the network structure is known as feed forward back propagation network (BPNN). The architecture of this network consists ofinput layer, one or more hidden layers and output layers. But, in this study, a three-layer consisting input, one hidden and output layers is used as widely used network model in the literature. The supervised learning method is used in this back propagation process to compare the networks outputs with the actual outputs on basis of least mean square error performance and then the randomly started weights are adjusted to minimize the error. The general structure of the three layer back propagation neural network is shown in Figure 1 [11].

Where, $X_i ; i = 1, 2, \dots, m$ represents the inputs, $R_j ; j = 1, 2, \dots, n$ represents the outputs of the hidden layer; $Y_k ; k = 1, 2, \dots, z$ represents the outputs of the network; W_{ij} and W_{jk} represent the connection weights. Net input (net_j), the output R_j of the i^{th} node in the hidden layer and Y_k can be calculated using the equations $(3), (4)$ and (5).

$$net_j = \sum_{i=1}^{m} w_{ij} x_i \ ; \ j = 1, 2, \dots, n \quad (3)$$

$$R_j = f_{hidden}\left(net_j\right) = f\left(\sum_{i=1}^{m} w_{ij} x_i\right) \quad (4)$$

$$Y_k = f_{output}\left(\sum_{j=1}^{n} w_{jk} R_i\right); \; k = 1, 2, ..., z \quad (5)$$

f_{hidden} and f_{output} are transfer functions of hidden layer and the output layer which are sigmoid functions such as log-sigmoid, in-sigmoid, pure-sigmoid, …etc. [11].

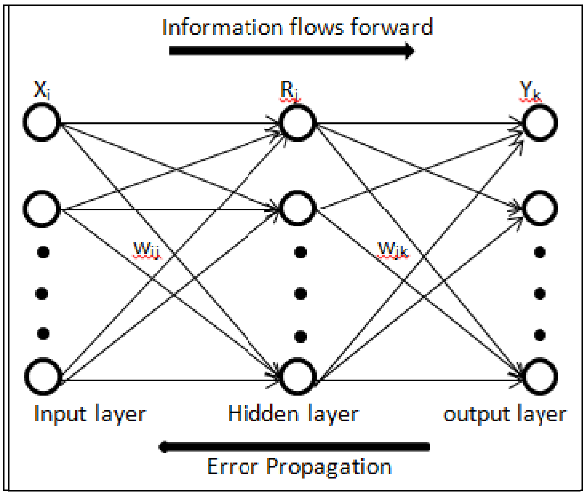

Figure 1: A three layer feed forward back propagation neural network

III. The Proposed Hybrid ARIMA-BPNN Methodology

The proposed new hybrid model consists of two forecasting components such as direction and value forecasting. For the direction forecasting process, raw data series convert to its 3-months moving averages and then SARIMA is employed to identify the immediate future direction by forecasting one step ahead monthly rainfall.Back-propagation neural network model is used to obtain the one step ahead value forecasting such that the identified directions of the 3-monthsmoving averages are combined to the BPNN forecasting process to find the forecast values from raw data series.

Proposed Hybrid ARIMA-BPNN Algorithm

Step 1: Input data

This step considers forms of raw data series.

$Y_t = \{y\}_t$: Raw data series ;where $t = 1,2, \ldots, n$.

Step 2: 3-months moving average conversion

The second step is allocated to identify the direction of the next value of the data series by considering the moving average conversion. Therefore, the whole data series is converted into its 3-months moving average series using two forms of actual and predicted moving average series M_A and M_p respectively.

For the data series $\{Y_t\}$; $t = 1,2, \ldots$ then M_A and M_p are calculated as follows.

$$(MY_A)_n = \frac{1}{5}(y_n + y_{n-1} + y_{n-2} + y_{n-3} + y_{n-4}) = \frac{1}{5}\sum_{i=n}^{n-4} y_i \quad (6)$$

$$\left(MY_p\right)_n = \frac{1}{5}(y_{n+1} + y_{n-1} + y_{n-2} + y_{n-3} + y_{n-4}) = \frac{1}{5}\sum_{\substack{i=n+1 \\ i\neq n}}^{n-4} y_i \quad (7)$$

The series M_A can be normally calculated for whole samplen. As equation shows, the immediate future value y_{n+1} of the raw data series is required to calculate the n^{th} observation of the seriesM_p and hence the corresponding calculations are done up to $(n-1)^{th}$ observation excepting the last value of the original series.

Step 3:SARIMA forecasting for M_p series

In this step, Box-Jenkins methodology ([11],[4]) is employed on the calculated M_p series to find the best SARIMA model that fits the moving average values well and forecast the next value$\left(\widehat{M}_p\right)_n$.

Step 4: Decision over the direction of next value

The direction can be examined by comparing the forecast value of the predicted moving average series with the n^{th} value of the actual moving average series. If the forecast value is less than the corresponding value of the other series, then it can be able to suggest that the next value of the original series will decrease with respect to its n^{th} value and otherwise it will increase. The relevant two decisions relevant to the data series Y_t are displayed as follows.

If $(MY_P)_n < (MY_A)_n$ then $y_{n+1} < y_n$ or

If $(MY_P)_n > (MY_A)_n$ then $y_{n+1} > y_n$

Step 5: BPNN forecasting

The identified BPNN network [11]is used here to forecast $(n+1)^{th}$ rainfall data value. The network is run 1000 times and observed all the forecast values and extract the particular forecast values that satisfy the direction predicted by **Step 4**. Finally, the average value of selected forecasts is got into the final forecast of the experiment.

IV. Forecasts Evaluation And Comparison

The accuracy level of the forecast could be measured by residual by obtaining the difference of actual and forecast values. More precisely, residuals are converted to various forms to understand the behavior of the forecast results as a concise summary. In this study, MAE and MAPE are utilized to evaluate the accuracy of in-sample forecasts. This study only considers one-step ahead forecast, the calculation of the error measures are as follows.

$$MAE = \left| Y_i - \hat{Y}_i \right|$$

$$MAPE = \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

Where Y_i and $\hat{Y}_i$ are the actual value of the original series and predicted value from the proposed hybrid model respectively. The smaller values of these error measures are considered to find the more accurate forecast result among the focused models.

Besides, the robustness of the models is evaluated by changing the sample size. Thus, the aforementioned methodology is applied for several samples to demonstrate the effectiveness and accuracy of the proposed hybrid model. Moreover, another three hybrid models used in the literature are analyzed using the same data samples to examine the accuracy of the new hybrid model. Hence, the comparisons of forecasting performances of four hybrid models, the average and weighted combining of the proposed individual models and their constituent's models are provided and compared them with their forecasting error measures.

V. Empirical results

Based on the time plots and ADF tests, it is clearly says that the two raw rainfall data seriesare stationary. But, Correlograms of raw data series and their 3-months moving averages in Figure 1 shows that the series have oscillatory movements of period 12 months.Figure 2 shows the correlograms of monthly rainfall data and their 3-months moving averages in Ginigathhena and Alupolla areas.Accordingly, several models are considered and the most suitable models is selected based on the minimum values of three information criteria such as Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (SBIC), and Hannan-Quinn criterion (HQIC). Table 1 shows the selected SARIMA models for data series I, II, III and IV.

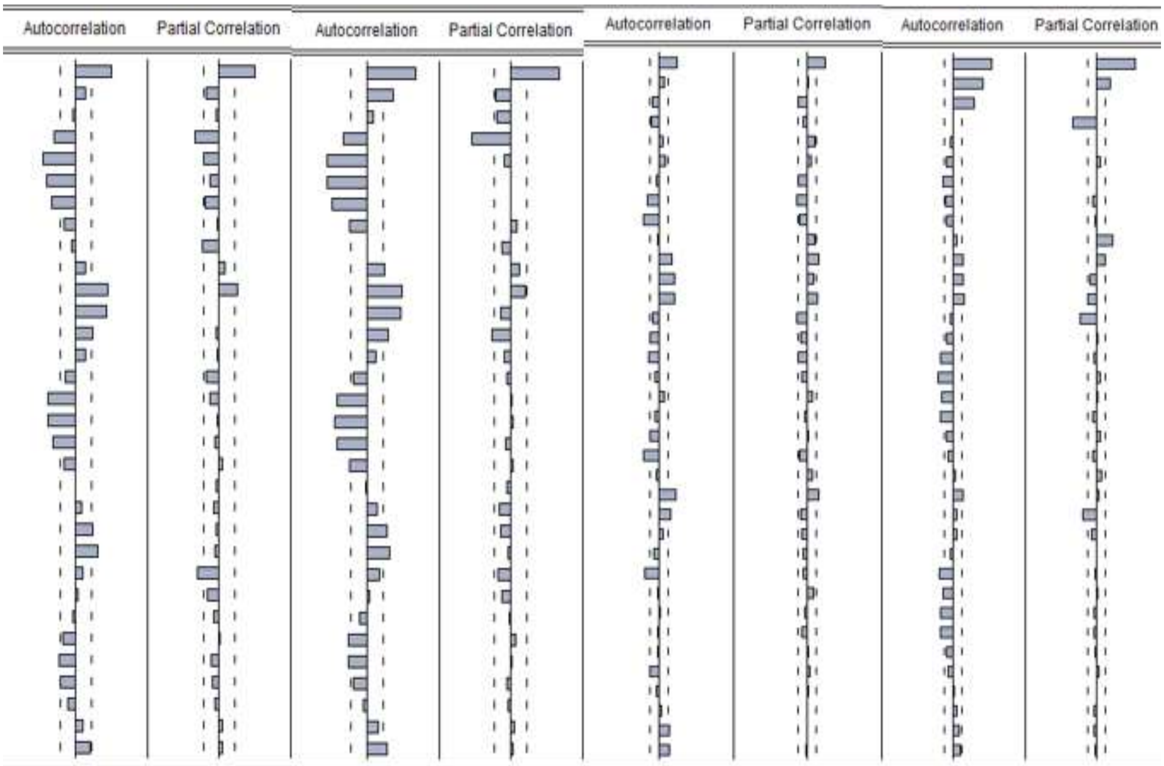

Figure 2: Correlograms

Table 1: Selected SARIMA Models

Data Series	Model
Series I (Ginigathhena monthly data series)	SARIMA $(0,0,1)\times(0,0,2)_{12}$
Series II (Ginigathhena 3-months moving average data series)	SARIMA $(0,0,3)\times(0,0,3)_{12}$
Series III (Alupolla monthly data series)	SARIMA $(0,0,1)\times(1,0,1)_{12}$
Series IV(Alupolla3-months moving average data series)	SARIMA $(1,0,2)\times(1,0,1)_{12}$

The proposed hybrid model mainly consists of two stages of direction and value prediction. The first stage was focused on identifying the direction of immediate next month rainfall value. The direction of next rainfall value is able to find through the comparison of the value of the series M_A and the forecasting value of the series M_P and all the information relevant to two samples I and II are summarized in Table 2.

Table 2SARIMA Forecasting Results for M_P Series II and IV

Series	Model	n^{th} value of M_A series	Forecast Value of M_P series	Comparison	Prediction of the direction of next month rainfall	Actual direction of $(n+1)^{th}$ rainfall
Series II	SARIMA $(0,0,3)\times(0,0,3)_{12}$	536.23	437.78	decrease	Downward ↓	Downward ↓
Series IV	SARIMA $(1,0,2)\times(1,0,1)_{12}$	725.83	480.27	decrease	Downward ↓	Downward ↓

As displayed results in Table 2, it can be concluded that the next future observation of the corresponding originalseries go upward or downward with respect to the comparison between n^{th} values of moving average series M_A andM_P. Accordingly, Table 2 shows that, the forecast value of M_P series II and IVare less than the n^{th} value ofM_A series II and IV respectively. According to the equations for calculating M_A and M_P series, it can be depicted that the immediate future value is less than the previous value of the original series I and III. Thus, it is able to identify the accurate direction of the next observation of rainfall data series I and III and process them for further analysis.

In the second stage, BPNN (1-2-1) model was employed 1000 times on series I and III separately and made its one-step-ahead forecasts. Then extracted all the forecast values those were consistent with the direction analysed in the previous stage. Then the average forecast of the filtered data was considered as the next month predicted rainfall value. The forecasting performances of the two series for SARIMA and Proposed hybrid models and their error measures MAE and MAPE are summarized in Table 3 and 4.

Table 3 Forecasting Results of Series I and III by Proposed Hybrid Model

Series	**Forecast Value**	**Actual Value**	**MAE**	**MAPE**
Series-I	328.2823	204.6999	123.5824	0.6037
Series-III	388.6921	243.1	145.5921	0.5988

Table 4 Forecasting Results of Series I and III by SARIMA Model

Series	**Forecast Value**	**Actual Value**	**MAE**	**MAPE**
Series-I	438.3560	204.6999	233.6561	1.1414
Series-III	459.3927	243.1	210.2927	0.8897

3.3Comparison of Forecasting Performance

The performance of two models at two different samples on monthly rainfall data under MAE and MAPE are reported in Tables 5.

Table 5: Forecasting performances of two models for series I and III

Series	**Model**	**SARIMA**	**Hybrid Model**
I	MAE	233.6561	123.5824
	MAPE	1.1414	0.6037
III	MAE	210.2927	145.5921
	MAPE	0.8897	0.5988

From Table 5, we can see that the proposed new SARIMA-BPNN hybrid model always gives the best result and each series it is significant and outstanding with respect to other results.Accordingly, the overall results show that the proposed new combining model is the best compared to the conventional SARIMA model alone.

VI. Conclusion

Theproposed methodology consists of SARIMA,and BPNN, and their combining models and One-step-ahead forecasts were obtained for monthly rainfall data series. Inthis study, monthly rainfall data from two areas namely Ginigathhena and Alupolla were used to demonstrate the effectiveness and the consistency of the proposed model comparative to the SARIMA model.Two error measures MAE and MAPE were used to compare the forecasting performances of the aforesaid models.In the distinct angle, we demonstrated that ARIMA model can be used for accurately predict the direction of the immediate future value of the series based on a particular form of 3-months moving averages. Based on the results, it can be seen that the proposed method has an ability to enhance the accuracy of BPNN forecasting as well as avoid some shortcomings, over fitting and under fitting problems appeared in the previous studies. The comparison results indicate that the proposed hybrid approach significantly outperformed than the conventional SARIMA forecasts. Therefore, the proposed model is suitable for modeling and forecasting rainfall data under the complex behaviors such as uncertainty, seasonality and high fluctuations.

VII. Future Research

In this study, only the rainfall data were used to identify the model and their one-step ahead forecasting. But, based on the other research works, it can be seen that the rainfall data are highly fluctuates due to the external effects. Therefore, it is better to model time series data with the support of external factors. Also, In the BPNN modeling, it could improve the individual forecasting ability by reducing the variability of outputs. For this, one-sided (upper or lower) boundary method was used. For gaining more accurate outputs, we can find a procedure to establish two-sided boundaries that can rather reduce the variability of the forecast value and would be much closer to the actual value.

References

[1]. Box, G. & Jenkins, G., 1970. Time Series Analysis, Forecasting and Control. San Francisco and London: Holden-Day.

[2]. Box, G. & Pierce, D. A., 1970. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. Journal of the American Statistical Association, 65, pp. 1509-1526.

[3]. Diaz-Robles, L. et al., 2008. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. Atmospheric Environment, Volume 42, pp. 8331 - 8340.

[4]. Etuk, E. & Mohamed, T., 2014. Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by Sarima Methods. Intenat in Knowledgeional Journal of Scientific Research, 2(7), pp. 320 - 327.

[5]. Feng, G. et al., 2016. Trend Analysis and Forecast of Precipitation, Reference Evapotranspiration, and Rainfall Deficit in the Blackland Prairie of Eastern Mississippi. Journal of Applied Meteorology and Climatology, Volume 55, pp. 1425 - 1439.

[6]. Meher, J. & Jha, R., 2013. Time-series analysis of monthly rainfal data for the Mahanadi River Basin, India. Sciences in Cold and Arid Regions, 5(1), pp. 0073 - 0084.
[7]. Mohamed, t. M. & Ibrahim, A. A., 2016. Time Series Analysis of Nyala Rainfall Using Arima Method. SUST Journal of Engineering and Computer Science, 17(1), pp. 5 - 11.
[8]. Nnaji, C., 2011. Time Series Analysis of Monthly Rainfall in Nigeria with Emphasis on Self-Organized Criticality. Journal of Science and Technology, 31(1), pp. 139 - 151.
[9]. R.M. Kapila Tharanga Rathnayaka, D.M.K.N. Seneviratna, Wei Jianguo and Hasitha Indika Arumawadu, "An unbiased GM(1,1)-based new hybrid approach for time series forecasting", Grey Systems: Theory and Application, 6(3), 322-340, 2016, Emerald Group Publishing Limited; http://dx.doi.org/10.1108/GS-04-2016-0009.
[10]. R.M. Kapila Tharanga Rathnayaka, D.M.K.N Seneviratne, Wei Jianguo and Hasitha Indika Arumawadu, "A Hybrid Statistical Approach for Stock Market Forecasting Based on Artificial Neural Network and ARIMA Time Series Models", The 2nd International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC'2015-IEEE), Nanjing, China, 2015. http://besc2015.njue.edu.cn
[11]. Seneviratna, D. & Chen, D., 2014. Using Feed Forward BPNN for Forecasting All Share Price Index. Journal of Data Analysis and Information Processing, Volume 2, pp. 87 - 94.
[12]. Soltani, S., Modarres, R. & Eslamini, S. S., 2007. The use of time series odelling for the determination of rainfall climates of Iran. International Journal of Climatology, Volume 27, pp. 819 - 829.
[13]. Sumi, S. M., Zaman, M. & Hirose, H., 2012. A Rainfall Forecasting Method Using Machine Learning Models and its Application to the Fukuoka City Case. International Journal of Applied Mathematics and Computer Science, 22(4), pp. 841 - 854.

Seneviratna,Dmkn, . "Rainfall Data Forecasting By Sarima And Bpnn Models." IOSR Journal of Mathematics (IOSR-JM), vol. 13, no. 06, 2017, pp. 57-63.