

Forecasting Rainfall in Mauritius using Seasonal Autoregressive Integrated Moving Average and Artificial Neural Networks

Jayrani Cheeneebash (Corresponding author)

Dept. of Mathematics, Faculty of Science, University of Mauritius, Reduit, Mauritius

Tel: 230-403-7496 E-mail: jayrani@uom.ac.mu

Ashvin Harradon

Dept. of Mathematics, Faculty of Science, University of Mauritius, Reduit, Mauritius

Ashvin Gopaul

Dept. of Mathematics, Faculty of Science, University of Mauritius, Reduit, Mauritius

Tel: 230-403-7512 E-mail: a.gopaul@uom.ac.mu

Received: September 28, 2017 Accepted: October 11, 2017

doi:10.5296/emsd.v7i1.12566 URL: <https://doi.org/10.5296/emsd.v7i1.12566>

Abstract

In this paper, two forecasting methods namely, the autoregressive integrated moving average (ARIMA) and the artificial neural network (ANN) are studied to forecast the amount of rainfall in Mauritius. Indeed due to the geographical location of Mauritius, the rainfall pattern is deeply affected by the season prevailing whereby the period of summer receives a relatively high amount of rainfall when compared to winter. As such, forecasting rainfall can help the local authorities to manage the distribution of water in the country especially during droughts. The results obtained from both methods are compared in terms of their mean square error, mean absolute difference and mean absolute percentage difference. It is then seen that artificial neural network is a much better model as it is more accurate. This is due to its nonlinearity characteristic and ability to learn and train itself.

Keywords: Forecasting rainfall, Artificial neural networks, Seasonal ARIMA

1. Introduction

Mauritius is a small island of volcanic origin situated between the latitudes $19^{\circ}59$ S and $20^{\circ}32$ S and longitudes $57^{\circ}18$ E and $57^{\circ}49$ E (Meteorological Services. National Climate Committee, 1999). It has a central plateau surrounded by mountain ranges. Because of its geographical location, Mauritius has two seasons, summer and winter. The months of December to March which form part of the summer season (November-April) are the months where most rain falls as where the pattern for summer and winter is quite contrasting due to the seasonal limitations (Meteorological Services. National Climate Committee, 1999). However, in both cases, the central plateau (found in the middle) receives the most amount of rain because of its height above sea level. Another factor that affects the amount of rainfall is the number of cyclones that hit our island in summer.

Rainfall in Mauritius has always followed a random pattern despite having a seasonal trend. As such, it is quite challenging to predict rainfall accurately in Mauritius. Rainfall data is collected by stations set up across the island of Mauritius that operates throughout the year (Dhurmea, Boojhawon, & Rughooputh, 2010). Rainfall impacts on each and every activity of mankind such as crop production, fishing, transportation by land, sea among others. Mauritius has faced serious drought problems in the past years and has been affected in various sectors of the economy. Difficulties faced in the agricultural sector often lead to high prices of the related products and these have a negative impact on the standard of living of people. It is thus important for the relevant stakeholders to get an accurate model to forecast rainfall for the island.

The study of forecasting models for rainfall in Mauritius is quite seldom in the literature and thus it difficult to compare our study with other findings. A recent work by (Fowdar et al., 2014) was done on the classification of rainfall patterns in Mauritius using principal component analysis. However the development of forecasting techniques for rainfall prediction has attracted much interest across the globe. Venkata Ramana et al. (2013) have used wavelet neural network analysis for monthly rainfall prediction. Chang'alet al. (2010) have shown how farmers in the South-western Highland of Tanzania predict rainfall using local environmental indicators and astronomical factors. Kyada, P. M. & Kumar (2015) have used the a adaptive neuro-fuzzy inference system for forecasting rainfall on a daily basis.

The paper is organised as follows. First we study the Autoregressive Integrated Moving Average (ARIMA) and in the next section, the Artificial Neural Network (ANN) is considered. Then we test the models and compare the results so as to determine which method gives better output.

2. Methodology

A time series is defined as a collection of random variables over a regular period: hourly, daily, weekly, monthly or on a yearly basis. The key purpose of these series is to formulate mathematical representations whose aim is to aid us to describe the fluctuation of the data over a time t .

2.1 ARIMA

A time series $\{a_t : t = 0, 1, 2, 3, \dots\}$ is an autoregressive moving average model, ARMA (p, q), if it is stationary

$$(AR): a_t = c + \sum_{i=1}^p \mathcal{G}_i a_{t-i} + w_t,$$

where c is a constant, \mathcal{G}_i are the autoregressive, AR, parameter and w_t is the white noise

and equation (MA): $a_t = \mu + w_t + \sum_{j=1}^q \theta_j^* w_{t-j}$ where μ is the expectation, which is often zero in

moving average, MA, part. w_t is the white noise, θ_j^* are the parameters of the MA and

where $w_t \sim W_N(0, \sigma^2)$, we have the ARMA as

$$(ARMA): a_t = c + w_t + \sum_{i=1}^p \mathcal{G}_i a_{t-i} + \sum_{j=1}^q \theta_j^* w_{t-j}$$

where $\mathcal{G}_p \neq 0$, $\theta_q^* \neq 0$ and $\sigma^2 > 0$.

Also it is to be noted that $\mathcal{G}(B)a_t = \theta^*(B)a_t$. The general form of autoregressive integrated

moving average model, ARIMA, is ARIMA (p, d, q) or $\mathcal{G}(B)\nabla^d a_t = \theta^*(B)w_t$, where $\mathcal{G}(B)$

is the AR operator having order p , $\theta^*(B)$ is the MA operator of order q , the value at time t of

our series is a_t , $w_t \sim W_N(0, \sigma^2)$ and d is the differencing order.

To fit a time series to the ARIMA structure, we proceed through three steps according to Box and Jenkins.

Step 1 Identify possible models to be used; perform stationary test.

Step 2 Specify model; determine the parameters.

Step 3 Forecast future values.

Finally the general equation is given by

$$\mathcal{G}(B)\nabla^d a_t = \theta^*(B)w_t \quad \text{or or} \quad \mathcal{G}(B)(1-B)^d a_t = \theta^*(B)w_t, \text{ where}$$

$$\mathcal{G}(B) = 1 - \mathcal{G}_1 B - \mathcal{G}_2 B^2 - \dots - \mathcal{G}_p B^p \quad \text{and} \quad \theta^*(B) = 1 + \theta_1^*(B) + \theta_2^*(B^2) + \dots + \theta_q^*(B^q) \quad \text{and}$$

$$w_t \sim W_N(0, \sigma^2).$$

2.2 Seasonal Autoregressive Integrated Moving Average (SARIMA)

In time series analysis, not all series are always fully random. Certain series tend to follow a seasonal pattern that is a particular pattern tends to repeat itself after a particular time lapse. Such series are known as SARIMA (Seasonal Autoregressive Integrated Moving Average) and the general form is denoted by SARIMA(p, d, q)*(P, D, Q) where, p , d , and q are the AR, differencing and MA order of non seasonal differencing and P , D , and Q are the order of seasonal component (Chang, Gao, Wang, & Hou, 2013).

The seasonal ARIMA is given by

$$\mathcal{G}_p(B^S) \mathcal{G}(B) \nabla_s^D \nabla^d a_t = \delta + \theta_Q^*(B^S) \theta^*(B) w_t$$

where w_t is the white noise and $w_t \sim W_N(0, \sigma^2)$,

$\mathcal{G}(B)$: Autoregressive model,

$\theta^*(B)$: Moving average model,

$\mathcal{G}(B^S)$: Seasonal autoregressive model and

$\theta_Q^*(B)$: Seasonal moving average model,

having order P and Q and where S is the cycle lag which is usually 12 for monthly and 365 for daily seasonal data.

2.2 Artificial Neural Networks

The artificial neural network is a computational model inspired from the elementary neural system of the animal's brain. The architecture of ANN consists of three layers. These are the input layer, the hidden layer and the output layer. The input layer is meant for users to input their data in the network. The responsibility of the hidden layer is to execute all calculations and processing needed whereas the output layer only displays the final results.

The back propagation algorithm was first introduced in the late 1986 (Rumelhart & McClelland, 1986) and has since then ascertained itself as the most valuable and user friendly algorithm to solve inexplicit problems as well as non-linear ones.

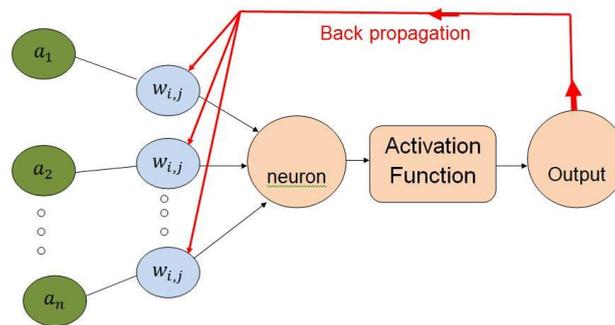


Figure 1. Back propagation method of ANN

The above mentioned process operates by making use of a technique known as supervised learning. The basic methodology is to insert our set of data in the network series and it generates weights between each node in such a way that we get the least possible error. The weights are altered and modified after each run and the error is calculated by the difference between the actual and the expected value. Once the error is calculated, the latter is sent back to the weight so that they are readjusted in such a way that the error becomes a minimum. One important criterion to take note of is that the weight can only change from -1 to 1. This algorithm works by repeating the whole concept until we get an acceptable error.

The process of training for ANN is as follows. We feed our original data in the network. The system starts by taking some random weights and determining its output. The errors are calculated at the end of each iteration and learning is achieved when these errors become sufficiently small. The network can then be used for prediction purposes.

3. Results and Discussions

In this section, we consider the amount of rainfall at Arnaud (in mm), for the period of July 2012 to July 2014 as this was the only data set available to us. Since the data was collected on a daily basis, we had sufficient data points to conduct the simulation. Furthermore our models can be applied to any set of rainfall data. This is the region where the largest reservoir of Mauritius, Mare aux Vacoas, is situated. The data was recorded on a daily basis.

We consider only 660 time step for the implementation purpose of both methodologies. We then use these data in both models to forecast 60 time step ahead and we then compare with the original recorded series to calculate the accuracy. Once the best model is found, the latter can be used for forecasting on a longer period of time.

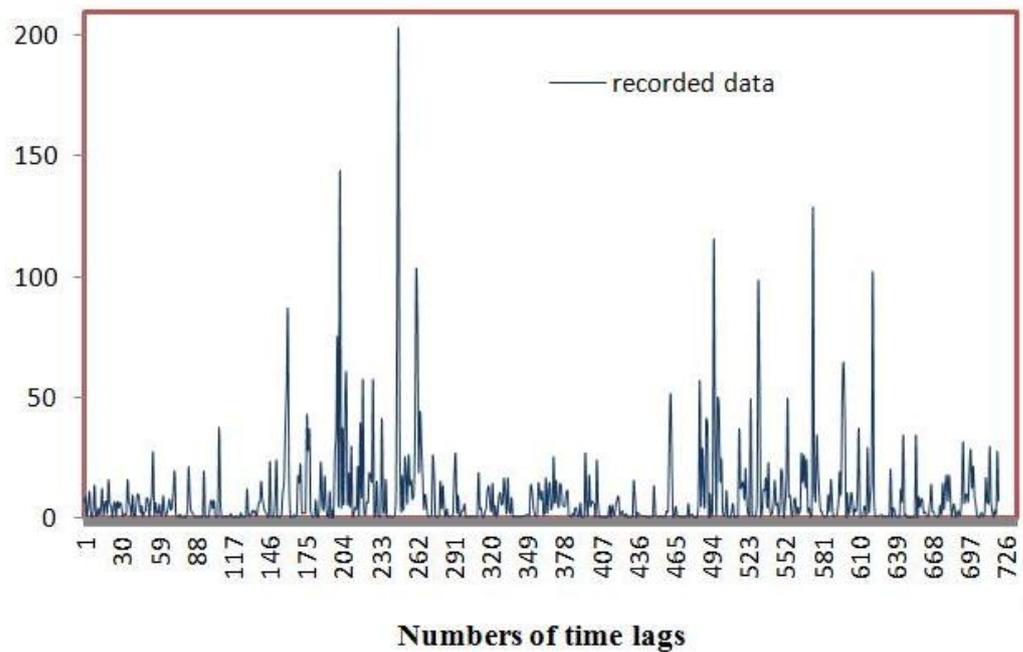


Figure 2. Daily rainfall recorded at Arnaud for the period July 2012 to July 2014

We normalise the raw data between the interval [1, 2] using

$$x_0 = a + \frac{(x - x_{\min})(b - a)}{(x_{\max} - x_{\min})}$$

where x_0 is the normalised value, x is the original value, x_{\min} and x_{\max} represent the minimum and maximum value of the series respectively and a and b are the limits between which we want to normalise the data. This normalisation best suits our data as there are some zero entries.

3.1 Fitting the Rainfall Data Using the SARIMA Model

From Figure 2, we can see that our data follows seasonal trend that is there is a pattern that repeats itself over a certain time. For SARIMA, we carry out the implementation using 4 distinct steps: specification, estimation, simulation and forecasting the multiplicative ARIMA model.

First, we need to find the parameter d , which is differencing order. In the first step, we have to determine whether differencing is needed. This is achieved by plotting the graph together with the mean. If the latter is not zero as in our case, we carry out seasonal differencing with seasonality $d=365$ as our rainfall data has seasonality of 1 year.

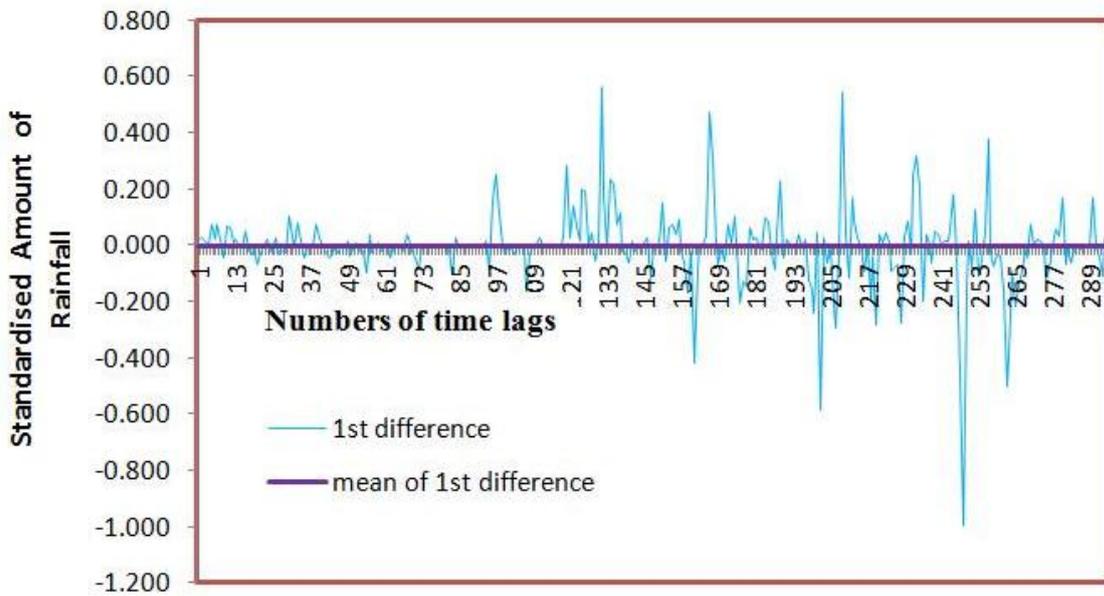
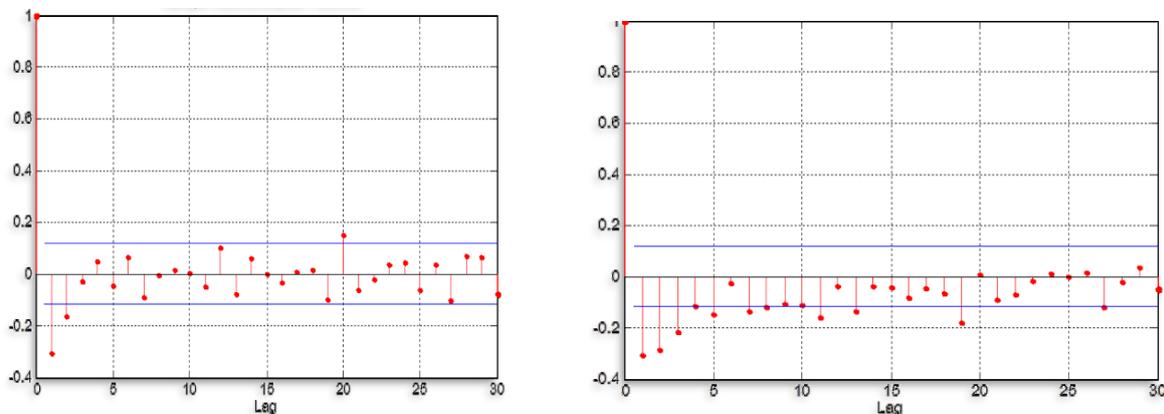


Figure 3. Differenced series together with the mean

We then plot the differenced series together with the mean to see if further differencing is needed. From Figure 3, the mean becomes zero after first differencing and the series is thus stationary. And we conclude that parameter d is 1. We now proceed with plotting the ACF and PACF of the series to figure out the parameters p and q respectively.



(a) ACF plot of rainfall data

(b) PACF plot of rainfall data

Figure 4. Autocorrelation plots of rainfall data.

From Figure 4, we can see that the highest peaks occur at lag 0, which indicates that p and q might both possibly be zero. The other significant peaks in the ACF plot occur at lag 1 and 2 while for the PACF, they occur at 1, 2 and 3. We then carry out the Ljung-Box test with the residuals (difference between estimated and original values) of the model to assess whether it is a good one or not.

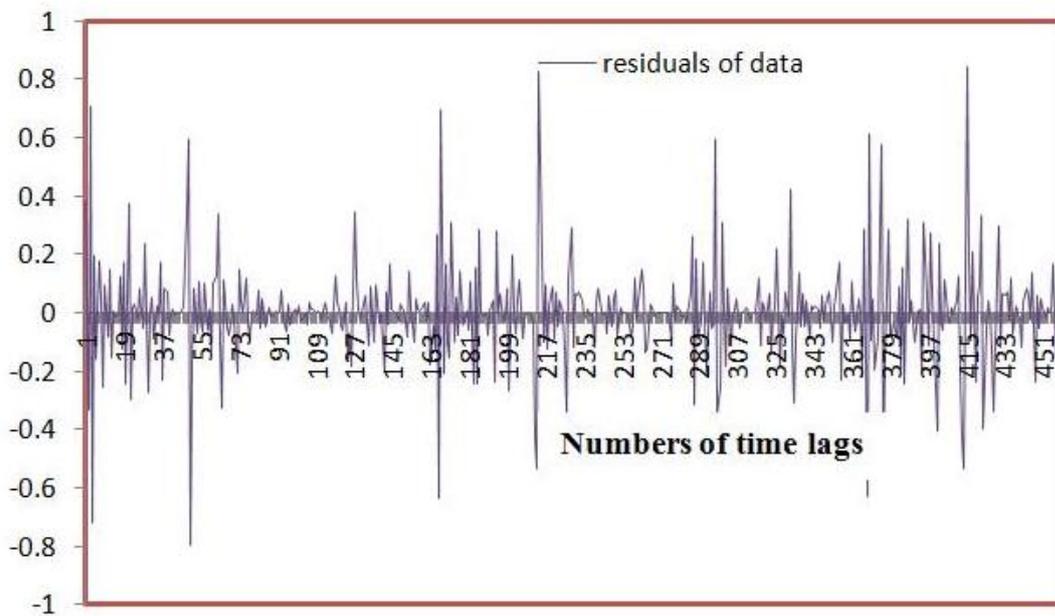


Figure 5. Residuals of the rainfall data

The Ljung Box test is a statistical test to check whether correlations exist between the residuals of a series. This test has been chosen as from literature it has been found as the most popular and efficient for this type of study. If the entries are independent, then we obtain a value of 0, else we obtain a 1 as in our case (Ljung-Box test, 2015). Our experiments show that the optimal solution for SARIMA is given as in Figure 6.

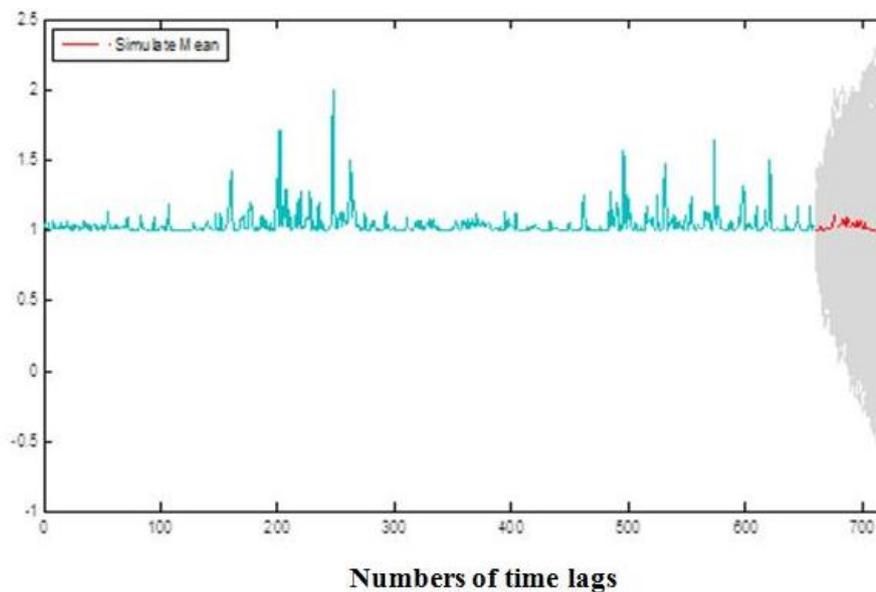


Figure 6. Forecast values using SARIMA

3.2 Fitting the Rainfall Data Using the ANN Model

We have used the Matlab toolbox for neural network to get our model. The normalised data is

split into two parts, one for input and one for output. We then choose the percentage of data we want to use as training, validation and testing. We then insert the number of hidden neurons and the appropriate time delays to train our network. The mean square error as well as the regression for the testing set are obtained and noted. However, since we cannot know beforehand which combination of time delay and hidden neurons will be the optimal solution, we vary both parameters from 1 to 10 and display the optimal result as follows in Figure 7.

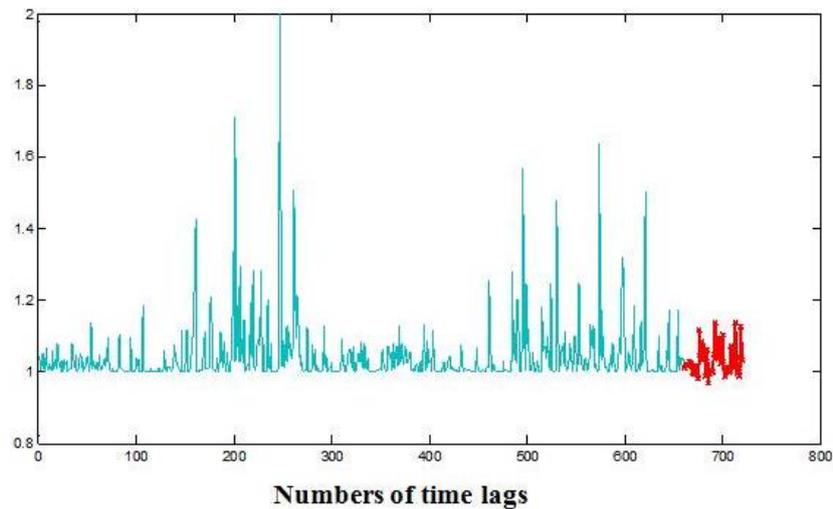


Figure 7. Forecasted series for the rainfall data

For the SARIMA model, it was observed that for this method, the best mean square error for the fitted part that is 60 time step was 0.0021 by using by using SARIMA(1,1,0) \times (P,D,Q) whereby the values for P , D , Q for which the error was minimised are 367, 1 and 0 respectively.

Secondly, for the ANN we fitted the data and forecasted the last 60 time step equally so that we can compare the results with the previous model. The parameters time delay and number of neurons were both varied from 1 to 10 and the results were recorded. The lowest mean square error of 7.9E-4 was at time delay to be 9 and number of neurons to be 7. The regression was noted to be 9.6E-1. The graph below shows the fitting of the forecasted series which can be visually seen to be better than SARIMA as it follows more or less the same pattern as the original series.

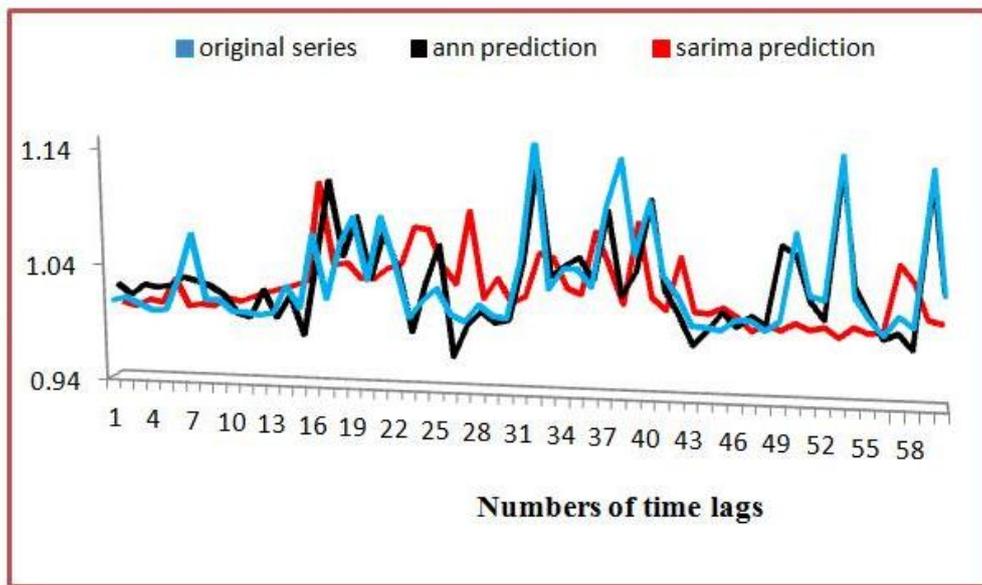


Figure 8. Comparison of forecasted values using ANN and SARIMA

Figure 8 shows that ANN model predicts of rainfall better than the SARIMA model. In Table 1, we provide the mean square error, root mean square error, mean absolute error and mean absolute percentage error of each model used. We find that ANN is clearly more accurate than the SARIMA model.

Table 1. Errors for SARIMA and ANN

Error	MSE	RMSE	MAD	MAPE
SARIMA	2.1E-3	4.6E-2	3.2E-2	3.02%
ANN	7.9E-4	8.9E-3	1.5E-2	1.46%

4. Conclusions

In this paper, the daily amount of rainfall recorded at the meteorological station in Arnaud was considered. We then tested the data for seasonality by differencing and for autocorrelation by the ACF and PACF plots. Then both SARIMA and ANN models were trained and their forecasted results were compared to the out-sample data. Our findings show that the ANN method is far more appropriate and accurate for forecasting rainfall since it fits better and produces an error which is less significant than that of SARIMA.

References

- Chang, X., Gao, M., Wang, Y., & Hou, X. (2013). Seasonal autoregressive integrated moving average model for precipitation series. *Journal of Mathematics and Statistics*, 500-505.
- Dhurmea, K., Boojhawon, R., & Rughooputh, S. D. (2010). Geostatistical approaches for estimating rainfall over Mauritius. *Research Week 2009-2010* (pp. 1-18). Reduit: University of Mauritius. Retrieved January 2015, from <http://vcampus.uom.ac.mu/researchweek/conference2009/uploads/paper1049595.pdf>

Fowdar, S. C., Rughooputh, S. D. D. V., Cheeneebash, J., Boojhawon, R., & Gopaul, A. (2014). Rainfall Analysis over Mauritius using Principal Component Analysis, *Environmental Management and Sustainable Development*, 3(2).

Meteorological Services. National Climate Committee. (1999). *Mauritius. Initial national communication of the Republic of Mauritius under the United Nations Framework Convention on Climate Change*. Vacoas (Mauritius): Meteorological Services.

Kyada, P. M., & Kumar, P. (2015). Daily rainfall forecasting using adaptive neuro-fuzzy inference system (Anfis) models. *Internation Journal of Science and Nature*, 6(3). 382-388.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing, Volume 1*. Cambridge: The MIT Press.

Venkata, Ramana, R., Krishna, B., Kumar, S. R., & Pandey, N. G. (2013). Monthly Rainfall Prediction Using Wavelet Neural Network Analysis, *Water Resources Management*, 27, 3697. <https://doi.org/10.1007/s11269-013-0374-4>

Chang'a, L. B., Yanda, P. Z., & Ngana, J. (2010). Indigenous knowledge in seasonal rainfall prediction in Tanzania: A case of the South-western Highland of Tanzania. *Journal of Geography and Regional Planning*, 3(4), 66-72.

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).