Grid-Search Criterion Technique for ARMA Model Optimization for Rain Rate Prediction in a Tropical Location

Akoma D. B[1*], Adeyeye S. D[2] , Opeyemi O[3], Ayoola O. T[4], Gbalaja M[5] and Ojo J. S[6]

[1,2,6]Federal University of Technology, Akure, Nigeria
[3]Bamidele Olumilua University of Education Science and Technology Ikere, Ekiti, Nigeria
[4]Ekiti State University, Ado-Ekiti, Nigeria
[5]Federal College of Education, Okene, Kogi State
*Corresponding author email: dbakoma@futa.edu.ng, akoma.blessing.d@gmail.com

Abstract. Intense rainfall is a significant factor affecting communication systems as radio signals interact with raindrops along their propagation path in the troposphere. However, the spatial and temporal variability of rainfall makes its measurement and prediction challenging. This study endeavors to investigate statistical methodologies for predicting time series values of rain rate using rainfall data collected for seven months in the year 2014's rainy season. The collection was gathered by the Communication Physics Research Group using the Micro Rain Radar installed within the Department of Physics. The study focuses on understanding rain rate patterns across various heights, with particular attention to the 320m height due to the susceptibility of the 160m height to near-field effects. After outlier removal, statistical distributions and monthly trends were analyzed, revealing distinct peaks and clustering of rain occurrences post-August. The First Order, regressive model, auto regressive model and Auto Regressive Moving Average (ARMA)models has been used in predicting time series values of rain rate. Ultimately, the ARMA model emerged as the most effective, leveraging both autoregressive and moving average components to enhance predictive accuracy by capturing complex data dependencies.

Keywords: MRR, ARMA, Grid-search criterion, Rain rate

1 Introduction

Rainfall is essential for agriculture, ecosystems, and water supplies, [1], [2] noted that it is Africa's most prominent climatic feature and a key marker of climate change. It is a crucial component of the climate and necessary for the moderate survival of life. But excess can cause flooding, and deficiency can cause different degrees of drought, both of which have detrimental effects on human activity. Excessive rainfall intensity can adversely affect communication systems, internet connectivity, and radio broadcasting. This is mainly because of the high-frequency waves' engagement with the atmosphere's components such as raindrops. This interaction causes signal degradation along the transmission path, especially under various meteorological conditions within the troposphere. Rain attenuation becomes particularly pronounced in frequency bands surpassing 10 GHz, posing a significant obstacle to the establishment of radio communication

I. Adimula et al. (eds.), *Proceedings of the 8th URSI-NG Annual Conference (URSI-NG 2024)*, Advances in Physics Research 12,
https://doi.org/10.2991/978-94-6463-644-4_25

systems [3]. However, rainfall exhibits considerable spatial and temporal variability, rendering measurement and prediction challenging [1]. The complexity of rainstorm events, limited historical samples, and the involvement of multiple meteorological factors make it difficult to accurately predict rare and extreme rainstorms [4]. Additionally, the high volatility and complicated nature of atmospheric data further complicate accurate rainfall prediction [5]. Rainfall prediction involves forecasting rainfall patterns over a specific region for an extended period. Moreover, rainfall prediction also assists communication engineers in efficiently designing communication infrastructure, ensuring optimal allocation of resources for communication networks [6]. The objective of this research is to assess the performance of Auto Regressive Moving Average model using walk forward validation technique in predicting rain rate in a tropical location

2 The Research Site

The communication laboratory under the Communication Research Group (CRG) of the Department of Physics at the Federal University of Technology, Akure (FUTA), located in Ondo State, Nigeria, has been collecting rainfall data using rain gauges and a vertically-pointing Micro Rain Radar (VPMRR) for over a decade. The study area spans from the southern coastal region (15 m above sea level) through the rocky hills of the northeast of Nigeria which is presented in figure 1a. The area is characterized by a tropical rainforest climate, with distinct dry and wet seasons. In the southern region, the average monthly temperature is approximately 27°C, with an average relative humidity above 75% [3]. Rainfall is experienced throughout the year, with a possible exception from November to January. The average annual rainfall in the southern region is around 2000 mm. In contrast, the northern region experiences a mean monthly temperature of roughly 30°C, with a relative humidity slightly below 70%. It has a distinct dry season from November to March, and the annual mean rainfall decreases to approximately 1800 mm [8].

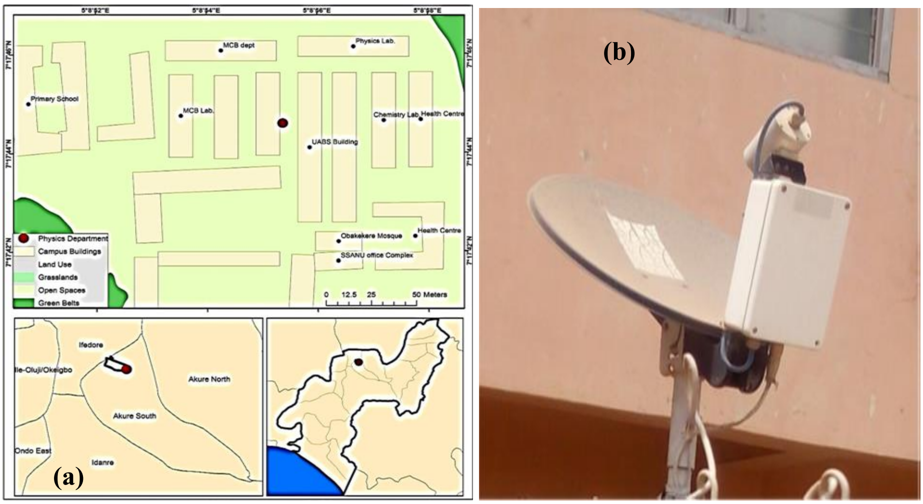

Fig. 1. (a) Map of Ondo State, Nigeria, indicating the research site, (b) Micro Rain Radar (MRR) Outdoor Unit at the Physics Department

In the range in the northern part of the state, the mean monthly temperature remains around 30°C during the warmer months but can drop significantly to around 6°C during the colder months. The relative humidity stays slightly under 70% on average. Overall, the southern region receives rainfall throughout the year, while the northern region experiences a distinct dry season. These climatic differences contribute to variations in temperature, humidity, and rainfall patterns across the study area, which are essential considerations for understanding the dynamics of precipitation and its impacts on various sectors in the region.

2.1 VPMRR Principle

The data collected was measured by a 24 GHz Vertical Pointing Micro Rain Radar (VPMRR), presented in Figure 1b, installed at the research site. It measures rainfall parameters by vertically transmitting an electromagnetic wave into the atmosphere. The transmitted wave is sent back to the parabolic antenna when it encounters raindrops along its transmission path.

Along vertical profiles that span several kilometers above the radar, the Micro Rain Radar (MRR) simultaneously retrieves a variety of rain data, such as rain rates (RR), liquid water content (LWC), number of drops (N), fall velocity (W), and radar reflectivity (Z) [3][9][10].

2.2 Data and Methodology

Seven months (April- October) of data for the year 2014 were retrieved from the Communication Physics Research Group measured by the MRR of the Department of Physics at the Federal University of Technology Akure (FUTA). The rain rate (RR), and the time stamps dates which are some of the parameters required for this investigation, were sorted out of the raw data collected. The raw data is charcterised by other parameters like the Liquid Water Content (LWC), Fall velocity (W), Number of drops (N), [9] noted that an overestimation of the measurement of rain parameters using VPMRR may occur at heights near the zero-degree isotherm height because water is observed to transform from one form to another at these heights. For this reason, the height considered for this study was limited to 2880 m. The details of this overestimation and its correction is presented in [11] and [3]. The retrieved data only focused on the wet season months which are known for a high amount of rainfall. More so, in cleaning the data set, [12] noted that the lowermost range gates up to 160 m are affected by near-field effects. Hence this study limited the vertical profile from 320m to 2880 m and sampled over 1 hour interval.

2.3 Time Series Model

1. First Order Auto-Regressive Model

A time series refers to a sequence of observations of one or more variables recorded over a period. Typically, these observations are taken at regular intervals such as hourly, daily, monthly, or yearly. A first-order autoregressive model involves regressing a value in a time series against its immediately preceding values within the same time series [7]. Where X_t represents the value of the time series at time t, φ_n represents the autoregressive coefficients, which determine the influence of previous time steps on the current value [14].

$$X_t = \varphi_1 X_{t-1} + Z_t \quad (1)$$

2. Autoregressive model

A time series X_t is considered an autoregressive process of order p, denoted as AR(p), if it can be expressed as a linear combination of its previous p values along with a random disturbance term.

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + Z_t \quad (2)$$

Where Z_t denotes a purely random process with zero mean and variance σ^2 [7].

3. Auto regressive moving average (ARMA) model

A mixed autoregressive moving average model with p autoregressive terms and q moving average terms is abbreviated ARMA (p, q) and may be written as:

$$\varphi(B)X_t = \theta(B)Z_t \quad (3)$$

The terms φ(B) and θ(B) represent polynomials in B of finite orders p and q, respectively. ARMA processes are significant because they allow for a more efficient approximation of many real datasets using fewer parameters compared to pure AR or MA processes. Autoregressive-moving-average (ARMA) models are mathematical tools used to describe the persistence or autocorrelation observed in time series data. These models find applications in various fields including hydrology, dendrochronology, econometrics, and others [7].

3 Result and Discussion

This section presents the result of the analysis based on the extracted data. The vertical profile of rain rate, drop diameter and liquid water content observed for different rain rate clusters was also presented.

3.1 Data Preprocessing

3.1.2 Vertical Profile of the Distribution of Rain Rate.

Figures 2a and 2b depict the occurrence of rain rate values across different heights, the mean rain rate values across different heights. Despite the observation of high occurrences of rain rate values at 160 m, 1600 m, and 1760 m heights, respectively, as illustrated in Figure 2a, the mean rain rate values across different heights shown in figure 2b indicate that 160 m experiences the highest, followed by 320 m, and then 640 m in that sequence. From the mean rain rate values, it appears that the 160 m height is associated with a significant number of high rain rate values, followed by the 320 m height.

Since 160 m has the tendency to be affected by near -field effects [12]. This study limited the height to only 320 m based on the mean rain rate values presented in Figure 2b. Moreso, the height would also provide more information about the occurrence of heavy rainfall, which is one of the important factors to be considered when designing satellite and terrestrial microwave radio links [13]. After removing the outlier observed in the figure 4a which showed the box plot distribution of rain rate value at the 320 m height, further examination of the 320m height revealed a statistical distribution wherein 75% of the dataset exhibited heights below 3 mm/hr, while the remaining values ranged between 4 mm/hr to 30 mm/hr as seen in figure 4b. The monthly average plot shown in figure 5b was to showcase any notable trend of the behavior of rai across the months considered for this study. From the plot, two prominent peaks were noticed in July and September, with a distinct break observed in August.

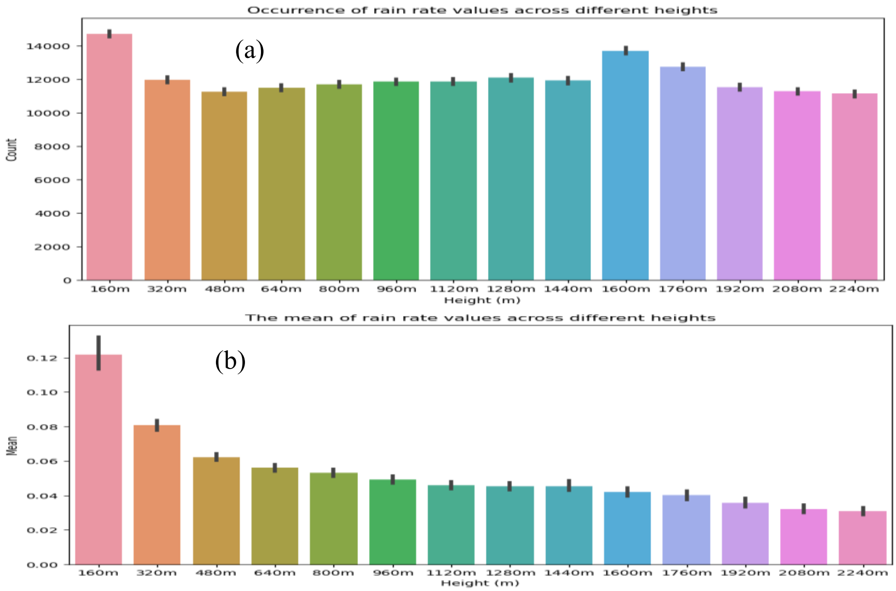

Fig. 2. (a) Occurrence of rain rate values across different height (b) The mean rain rate values across different height

Following this break, the rain episodes displayed higher rates compared to those before August, potentially indicating convective precipitation. Additionally, the time series plot presented in figure 5a highlights a clustering of rain occurrences post-August, suggesting a more regular pattern of rain events compared to the period preceding the August break.

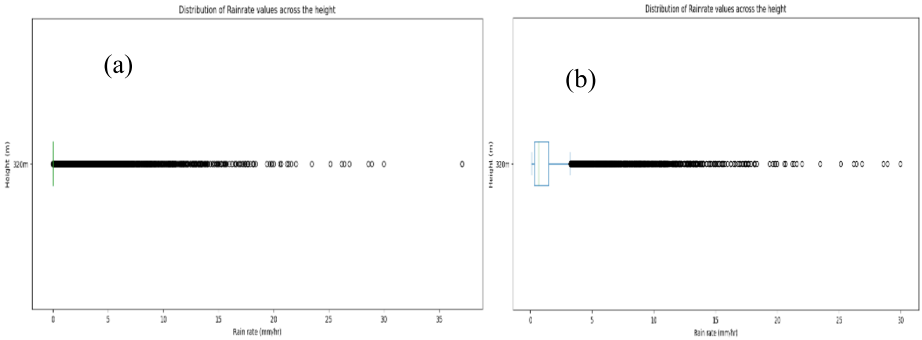

Fig. 4. Distribution of rain rate values across (a) 320m with an outlier (b) after filtering the outlier

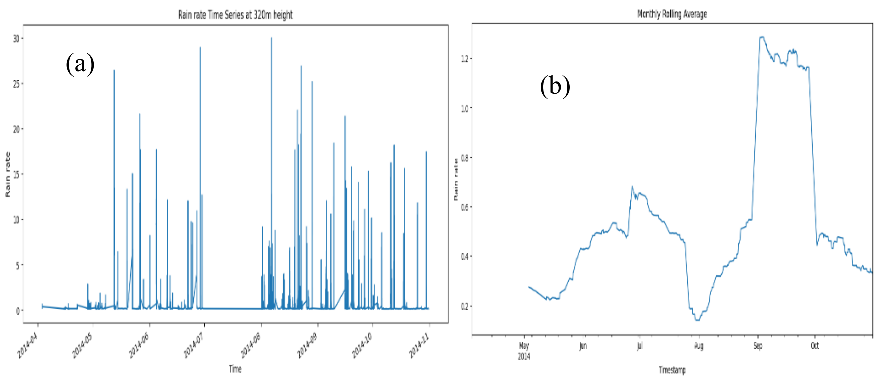

Fig. 5. (a) Time series plot of rain rate values at 320m height (b) Monthly moving average of rain rate values

3.2 Simulation of Results on Python

First Order Auto Regressive Model.

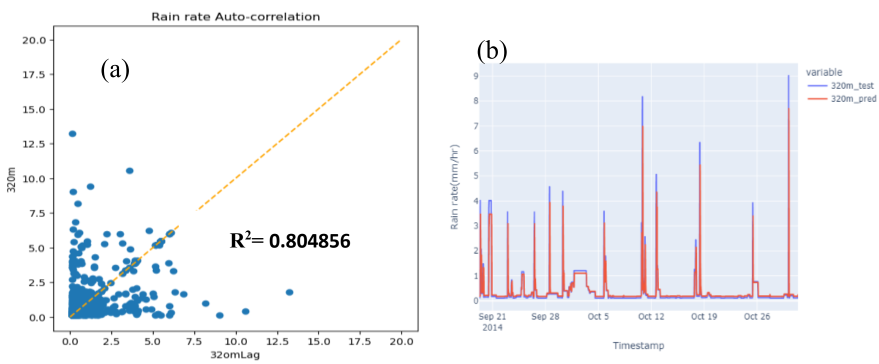

Fig. 6. (a) Rain rate auto-correlation with 320m lag and 320m height (b) The plot of the actual data and Predicted data for the first Order autoregressive model

Autoregressive Model.

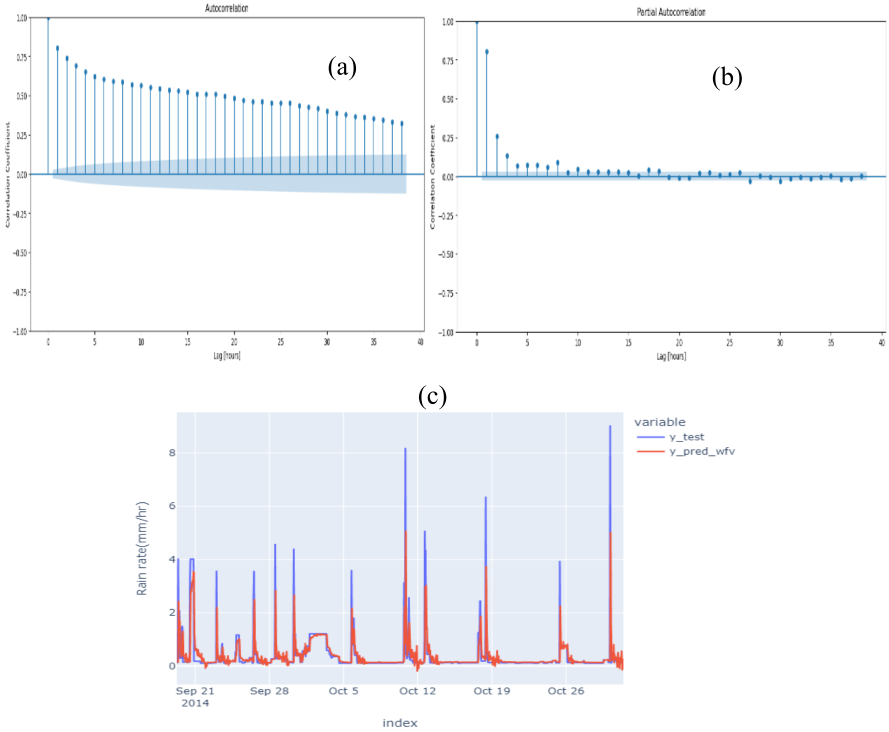

Fig. 7. (a) Autocorrelation function plot for the actual data (b) Partial Autocorrelation function plot for the actual data (c) The plot of the actual and the predicted data for the autoregressive model

3.3 Autoregressive Moving Average Model

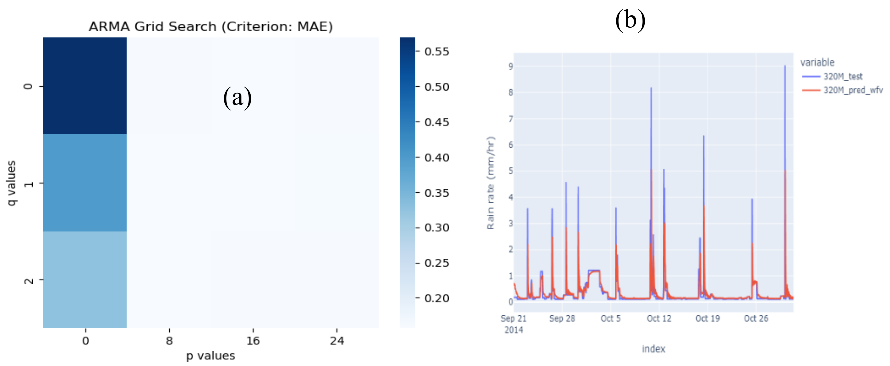

Fig. 8. (a) Auto Regressive Moving Average Grid search (b) The plot of the actual and the predicted data for ARMA model

Using equation 1, the predicted value is regressed on the immediate previous 1-hour rain rate value. In simulating this, the immediate previous 1-hour value and the current data were explored to see their correlation. Figure 6a showed that there is a strong correlation between the immediate previous 1-hour rain rate and the current rain rate. This was then simulated into the first-order regression model, and the plot is presented in figure 6b. Using equation 2 to simulate the autoregressive model, figure 7a presents the Autocorrelation Function (ACF) plot for the actual rain rate values at 320 m height. This plot explored the number of lags needed for the model to have good predictive power, and from the plot, a negative gradient was observed between the correlation coefficient and the number of lags. However, in determining the number of lags that are important for the predictive power of a model, the Partial Autocorrelation Function (PACF) was plotted and presented in figure 7b. From figure 7b, the points captured in the blue band have a correlation that is very small, hence it is not statistically significant for the predictive power of the model. Beyond 30 lag hours, the other points are noticed to be statistically insignificant. Hence for the autoregressive model, the number of lags is 30. The data was simulated to the autoregressive model with a walk-forward validation technique, and the plot was presented in figure 7c.

Using equation 3, the p-value which represents the autoregressive (AR) component of the model. It indicates the number of lagged values of the dependent variable to include in the model. In other words, it determines how many past observations of the time series are used to predict the current value while the q parameter, which represents the moving average (MA) component of the model, indicates the number of lagged forecast errors (residuals) to include in the model. The p parameter was ranged from 0 to 30 based on the PACF plot and 0 to 3 for the q parameter. The grid search for the p and q parameters with the criterion of the Minimum Absolute Error (MAE) was presented in figure 8a, and the corresponding matrix was presented in Table 1. From the table, the p and the corresponding q parameter with the lowest minimum absolute error was at P value of 8 and q values of 2, which were simulated into the ARMA model using a walk validation technique. The simulated plot was presented in figure 8b, and the metrics value of the models was presented in table 2. From the metrics values, the ARMA model performed better than the autoregressive model and the first-order model because it considers both the autoregressive and moving average components, allowing it to capture more complex patterns and dependencies in the data, leading to improved predictive accuracy. Additionally, the inclusion of both past observations and past forecast errors enhances the model's ability to capture and forecast the underlying behavior of the time series data.

Table 1. The grid search of the p and q parameter with a criterion of MAE

	P Values			
Q Values	**0**	**8**	**16**	**24**
0	0.569313	0.159113	0.158243	0.160437
1	0.398539	0.157908	0.158379	0.160469

2	0.326905	0.157218	0.159973	0.160013

Table 2. The Error matrix

MODEL	MAE
First-Order Autoregressive model	0.2
Auto-Regressive Model of many order	0.2
ARMA Model	0.17

4 Conclusion

Analysis has been conducted on rain rate data during the wet season months of 2014. Despite frequent occurrences of rain rate values at 160m, 1600m, and 1760m heights, the mean rain rate values revealed 160m as the highest, followed by 320m and 640m. This study focused on the 320m height given the susceptibility of 160m to near-field effects. Statistical distribution explored on this height indicated that the majority of the rain rate values is below 3mm/hr. Simulation of the first-order regression model, autoregressive model, and ARMA model highlighted the superiority of the ARMA model, which incorporates both autoregressive and moving average components, thereby enhancing predictive accuracy by capturing intricate data dependencies. This study will assist communication engineers in efficiently designing communication infrastructure, ensuring optimal allocation of resources for communication networks.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Wang Jiayi, Raymond K W Wong, Mikyoung Jun, Courtney Schumacher, R Saravanan and Chunmei Sun : Statistical and machine learning methods applied to the prediction of different tropical rainfall types' Environ. Res. Commun. 3 111001 (2021).
2. Elisha Ademola A. and Abiodun Daniel O.:Statistical Analysis Of Rainfall Trend In Akure, Ondo State, Nigeria' Analele UniversităţiidinOradea,SeriaGeografie, no. 1, pp. 114-121 (2017).
3. Ojo J. S., Akoma D. B., Olurotimi E. O.:Dynamical Evolution of Vertical Profile of Rain Structures Observed Using Ground-based Radar over a Tropical Station Heliyon, Vol. 7(4), e06888 (2021).
4. Mahajan D. and Sharma, S. :Prediction Of Rainfall Using Machine Learning, Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, pp. 01-04 (2022).
5. Tong, Zhang., Jie, Liu., Jianlong, Wang: Rainstorm prediction via a deep spatio- temporal-attributed affinity network. Geocarto International, 37(26):13079-13097. doi: 10.1080/10106049.2022.2076914(2022).

6. Kundu Saranagata, Saroj Kr. Biswas, Deeksha Tripathi, Rahul Karmakar, Sounak Majumdar, Sudipta Mandal:A review on rainfall forecasting using ensemble learning techniques',e-Prime - Advances in Electrical Engineering, Electronics and Energy,Volume 6, 100296 (2023).
7. Dhaval P. Patel , Mitul M. Patel, Devendra R. Patel: Implementation of ARIMA model to predict Rain Attenuation for KU-band 12 Ghz Frequency' IOSR Journal of Electronics and Communication Engineering Volume 9, Issue 1, Ver. III , PP 83-87 (2014).
8. Emekwuru N, Ejohwomu O.:Temperature, Humidity and Air Pollution Relationships during a Period of Rainy and Dry Seasons in Lagos, West Africa'. Climate. 2023; 11(5):113 (2023).
9. Harikumar, R., Sampath, S., Sasi Kumar, V.,: Altitudinal and temporal evolution of raindrop size distribution observed over a tropical station using a K- band radar." Int. J. Rem. Sens. 33, 3286–3300 (2012).
10. Meteorologische Messtechnik GmbH :Micro rain radar user manual (2008)
11. Akoma D. B., Adetan O., Ojo J. S.: Dynamical assessment of altitudinal trend of drop size distributions in a tropical region of Nigeria'. J. Phys.: Conf. Ser. 2034012017(2021)
12. Andreas Foth , Janek Zimmer , Felix Lauermann , and Heike Kalesse-Lo.: Evaluation of micro rain radar-based precipitation classification algorithms to discriminate between stratiform and convective precipitation' Atmos. Meas. Tech., 14, 4565–4574 (2021)
13. John, D., Kanellopoulos., S.N., Livieratos.: A Modified Analysis for the Prediction of Multiple-Site Diversity Performance in Earth-Space Communication Including Rain Height Effects. Journal of Electromagnetic Waves and Applications, 11(4):485-513 (1997).
14. Shen Liu, James McGree, Zongyuan Ge, Yang Xie :Big data from mobile devices, Computational and Statistical Methods for Analysing ,Big Data with Applications, Academic Press, (2016)