Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6, 520-537
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i6.2117

Performance of advanced-deep learning algorithm for modeling and prediction of rain height over some selected locations in the tropics and sub-tropics zone for radio propagation applications

Yusuf Babatunde Lawal[1*], Pius Adewale Owolawi[2], Chunling Tu[3], Etienne Van Wyk[4], Joseph Sunday Ojo[5]
[1,2,3]Department of Computer Systems Engineering, Tshwane University of Technology, South Africa; lawalyb@tut.ac.za (Y. B. L.), owolawipa@tut.ac.za (P. A. O.), duc@tut.ac.za (C. T.)
[4]Faculty of Information and Communications Technology, Tshwane University of Technology, South Africa; vanwykea@tut.ac.za (E.V. W.)
[5]Department of Physics, Federal University of Technology, Akure, Nigeria; ojojs_74@futa.edu.ng (J. S. O.)

Abstract: As demand for high-frequency broadband communication services keeps rising, rain-induced attenuation remains the predominant threat to radiowave propagation. Accurate prediction of attenuation requires continuous measurement and monitoring of rain-induced meteorological parameters, specifically rain rate and rain height, due to their spatio-temporal variations. Rain height is an upper dataset mostly computed from Zero- degree Isotherm Heights (ZDIH) measured by radar. This research proposes a novel approach for predicting rain height from earth surface data such as surface temperature, pressure, total cloud cover, dew point temperature, surface solar radiation, water vapor amount in the air, and humidity. This research investigates the relationship between meteorological surface data and rain height. Subsequently, six machine learning models were employed for predicting rain height using the ten years surface data as input variables. The models were applied to six sub-tropical (Polokwane, Pretoria. and Cape Town) and tropical (Sokoto, Akure, and Port Harcourt (PH)) stations in South Africa and Nigeria, respectively. Analysis of the results shows that the Gradient Boosting Algorithm (GBA) performed best with determination coefficients greater than 0.80 and RMSE less than 350 in all three stations in South Africa. However, all the models failed to produce good result for the Nigeria stations. The Random Forest model has the fairest performance metrics with r2 of 0.40, 0.46 and 0.46 in Sokoto, Akure and PH. respectively. GBA is recommended for predicting rain height in South Africa. The research outcome would assist radio engineers in improving the prediction of rain-induced attenuation and determining appropriate fade mitigation techniques.

Keywords: *Advanced deep learning, Gradient boosting model, Rain height, Rain-induced attenuation, Spatio-temporal variation, ZDIH.*

1. Introduction

Rain height, the distance from the Earth's surface up to the top of the precipitation column, is one of the most important parameters in meteorology and telecommunications. It is thus necessary to know the height of the rain for predicting the rain-induced attenuation on satellite communication systems operating above 10 GHz. Especially in the tropical and sub-tropical regions where severe and frequent rainfalls happen. Rain attenuation to estimate the necessary fade margin and recommend appropriate fade mitigation techniques to alleviate the rain effect on communication networks. The traditional methods of measuring rain height from freezing level heights are mainly through satellite radar or ground-based radar such as MRR. Several studies have revealed that the satellite method is not ideal for

* Correspondence: lawalyb@tut.ac.za

the measurement of localized high-resolution lower atmospheric data due to the variability of atmospheric parameters, satellite distance to target variables, comparatively low spatio-temporal resolution, etc [1-3]. The ground-based radar method is popularly preferred because for localized measurement [4]. However, there are few ground-based radar observatory stations across the world due to procurement, installation costs, and complex maintenance [5]. Most weather observatory stations are equipped with multiple sensors that can measure atmospheric variables such as temperature, pressure, humidity, solar radiation, wind speed, precipitation, etc at the earth surface and near-earth surface.

The motivation for this study is the scarcity of accurate localized real-time rain height data for estimation of rain-induced attenuation and fade margin especially in the subtropical region. The research seek to investigate whether it is possible to process lower atmospheric data with the help of machine learning algorithms to estimate the rain height more accurately. These models will deploy input parameters such as surface temperature, surface pressure, total cloud cover, dew point temperature, surface solar radiation, the amount of water vapor in the air, and humidity to enhance the estimation. The aim of the research is to apply machine learn models to predict rain heights from the earth surface atmospheric variables for tropical and sub-tropical regions. Variables of interest that influence ZDIH and rain height shall be identified while ML models such as Surface Vector Regressor, Random Forest, Gradient Boosting, XGboost, and Neural Network shall be deployed to predict rain heights

Numerous researches have been conducted to provide various methods of rain height estimation since it cannot be measured directly. Conventional techniques involve the use of precipitation data obtained from radar measurements for direct estimates of rain profiles. The two common types of radar observation are satellite-borne and ground-based radars. For instance, the satellite-borne Tropical Rainfall Measuring Mission (TRMM) precipitation radar, which provides freezing level height useful in international rainfall height imagery [6-7]. The Dual Precipitation Radar (DPR) of the Global Precipitation Measurement (GPM) mission, which is the successor of the TRMM, is capable of providing detailed vertical profiles of precipitation, which aid in the measurement of the bright band, the zero-degree isotherm height, and the rain height [8]. The ground-based radar uses Doppler radar to remotely measure the vertical profile of the troposphere up to about 10 km above ground level. These radars detect the speed and intensity of the precipitation particles needed to estimate the position of the melting layer, which is often referred to as the bright band, and freezing level heights from radar reflection. The bright band is a region with increased radar reflectivity as a result of melting snowfall, and it is normally situated close to the zero isotherm height. Rain height estimation could be enhanced using considerable information such as zero isotherm height, bright band, shape, and orientation of hydrometeors from dual-polarisation radars [9]. However, these techniques are based on remote sensing of upper-air atmospheric variables through radar echo. The radiosonde method provides a more accurate approach for measuring upper data due to its in situ nature. Although this method provides localised data, which is preferred to satellite-borne radar, there is insufficient global data due to the limited number of radio-sounding stations across the world. Algorithms and models, such as artificial neural networks and regression trees, have also been developed to estimate rain height from radar data. For instance, Meneghini et al applied the Surface Reference Technique (SRT) to the radar signal's interaction with the Earth's surface to calibrate the radar and improve rain height estimates in [10]. Upper air atmospheric parameters and 0°C isotherm levels were integrated into a single unique model through the use of machine learning by Mandeep (2008) [11] to enhance rain height estimation in Malaysia. Lawal et al recommended a latitude-dependent equation for the computation of rain heights in Nigeria in [12]. The equation has a determination coefficient of 0.8, which implies that more research could be carried out to improve the estimation accuracy. Nalinggam et al created rain attenuation models in [13] for Southeast Asia and stressed how machine learning could be useful in studying rain in tropical regions.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

However, none of these investigations utilized machine learning to estimate rain heights from the earth's surface or near-surface atmospheric data at the regional level, particularly in the areas where climate change is evident. Akure is one of the study stations located in Nigeria, and it experiences a tropical climate. Similarly, Pretoria is located in South Africa, a subtropical region where rainfall also has devastating effects on radio signals operating above 10 GHz. Therefore, these locations are obviously convenient to study, model, and develop machine learning techniques for rain height estimation from atmospheric variables at the earth surface. This study intends to address this challenge by using data of the local atmosphere and ML models to enhance the accuracy of rain height estimations for tropical and subtropical zones.

1.1. Climatology of Research Locations and Data Acquisition

The research locations are Polokwane, Pretoria, and Cape Town in South Africa, and Sokoto, Akure, and Port Harcourt in Nigeria. According to the climatological classifications of African countries, the former country lies in the sub-tropical region, while the latter is categorised under the tropical region. The coordinates and elevations of the study locations are presented in Table 1. The locations were selected based on latitudinal distribution across the country since some previous works reported latitudinal dependence of rain heights [14-17] and [12]. South Africa is a sub-tropical region located in the southernmost part of the African continent. It experiences four seasons of weather annually, namely: summer (December–February), autumn (March–May), winter (June–August), and spring (September–November). It is a subtropical highland climate, characterized by warm, rainy summers and mild, dry winters.

Table 1.
The six study locations and their geographical coordinates.

Zone	Station	Lat (°)	Long (°)	Elevation (m)
South Africa -Sub-tropical Zone	Polokwane	-23.905	29.467	1315
	Pretoria	-25.733	28.183	1332
	Cape Town	-33.917	18.425	25
Nigeria -Tropical Zone	Sokoto	13.023	5.245	296
	Akure	7.255	5.206	353
	Port Harcourt	4.078	7.005	16

Nigeria has diverse geoclimatic characteristics, ranging from the Sahel region at the north to the Savannah at the Center to the coastal region at the southernmost part of the country. The average annual temperature is approximately between 26 and 28 °C, while the average annual rainfall is about 2000 m, especially in the coastal region. Nigeria experiences rainy and dry seasons only. The rainy season runs approximately between May and September, while the dry season reigns between November and March. High humidity is dominant and longer rainy seasons are dominant in the south, while intense heat and longer dry seasons reign in the north [18-20].

2. Overview of the ML Models

2.1. Random Forest Model

The Random Forest Model is a machine learning algorithm that deploys a combination of several decision trees to produce a unique result. It is an ensemble of learning that is suitable for classification and regression analysis. Its working principle involves building several decision trees during the training phase and then providing its final decision in the form of the class modes for classification analysis or the mean of the individual trees' predictions for regression analysis.

The model integrates multiple decision trees to improve predictive performance and control overfitting through the processes of bootstrap sampling, tree construction, and aggregation. The bootstrap stage randomly generates a bootstrap sample size and replaces it with a training set k from the original data. At the tree construction stage, the bootstrap sample grows a decision tree T_b using a

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

subset of features, $m << p$, *where* p is the total number of features. Equations (1) and (2) are used at the prediction stage for classification and regression analysis, respectively.

$$\hat{Y} = mode\{T_i^b(X)\}_{b=1}^{B} \quad (1)$$

$$\hat{Y}_i = \frac{1}{B}\sum_{b=1}^{B} T_i^b(X) \quad (2)$$

where $\hat{Y}_i$ is the final prediction for input sample X. B is the total trees in the forest. T_i^b is the prediction from the b^{th} tree for the i^{th} sample X. The model leverages the multiple decision trees in the forest to improve prediction accuracy.

2.2. Gradient Boosting Model

Gradient Boosting is a meta-algorithm ensemble learning method where weak models like decision trees are built stage wise consecutively in a sequential fashion to reduce the magnitude of a given loss function to the barest minimum. This method is used to break down a strong predictor into multiple weak predictors in order to deal with a complex data set [21]. The aim of Gradient Boosting is to produce a model F(x) that predicts the target variable y. It builds the model incrementally by adding weak learners $h_m(x)$, usually decision trees, in such a way that each new model fixes the flaws in the previous ones. The algorithm starts with an initial model which is usually the mean value of the target variable for regression problems as defined in equation (3) [22-23].

$$F_0(x) = arg\,min \sum_{i=1}^{n} L(y_i, c) \quad (3)$$

where L is the loss function, y_i and c are the true values and constant respectively.

The succeeding model is an additive model built in a stage-wise manner using equation (4). The loss function at each stage m is minimized by adding a base learner $h_m(x)$ defined by equation (5). This is accomplished by fitting $h_m(x)$ to the loss function's negative gradient.

$$F_m(x) = F_{m-1}(x) + vh_m(x) \quad (4)$$

where $h_m(x)$ is the new base learner added at stage m and v is the learning rate which controls the impart of each base learner.

$$h_m(x) = arg\,min_h \sum_{i=1}^{n} \left[-\frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}\right]^2 \quad (5)$$

Equation (4) is updated continuously by adding new base learners scaled by the learning rate.

2.3. XGBoost Model

The XGBoost is an optimized version of the gradient boosting model with focus on flexibility, higher efficiency, and portability. The Machine learning algorithm is initiated by the objective function which combines the loss function and a regularization term that helps to control the model complexity as stated in equation (6) Chen & Guestrin, 2016 [23]).

$$\ell(\theta) = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (6)$$

where L is the loss function such as mean squared error for regression), $\Omega(f_k)$ is the regularization term, θ represents the parameters of the model, $\hat{y}$ is the predicted value, and f_k are the individual trees in the ensemble. The objective function is optimized by applying the second-order Taylor expansion to approximate the loss function, hence the objective function is expressed by equation (7) as;

$$\ell^{(t)} = \sum_{i=1}^{n} [g_i f_i(x_i) + \frac{1}{2} h_i f_i(x_i)^2] + \Omega(f_t) \quad (7)$$

where $g_i = \partial_{\hat{y}(t-1)} L_{\left(y_i, \hat{y}_{i(t-1)}\right)}$ *and* $h_i = \partial^2{}_{\hat{y}(t-1)} L_{\left(y_i, \hat{y}_{i(t-1)}\right)}$ are the first and second derivatives of the loss function with respect to the prediction.

2.4. Support Vector Regressor Model

SVR is a type of support vector machine algorithm that attempts to find a function whose difference in value between that function and the observed value is not greater than a specified margin. The model

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

ensures that the function is as flat as possible. It is commonly used for regression analysis. Equation (8) is used to obtain the most deviated function f(x) within a specified margin, which is known as the objective function.

$$f(x) = \langle w, x \rangle + b \quad (8)$$

where $\langle w, x \rangle$ denotes the dot product of the weight vector w and input vector x, b is the bias term. The SVR loss function is epsilon-insensitive i.e. the errors within the margin of the $\in$ is ignored, thus the loss function is given by [24]. The optimal values of $wandx$ can be obtained using complex optimization procedure described in the work of [25] and [26]

$$L = \begin{cases} 0 & if\ |y-f(x) \leq \in \\ |y-f(x)|-\in & otherwise \end{cases} \quad (9)$$

The regression function for the non-linear regression after optimization is given by equation (10).

$$f(x) = \sum_{i=1}^{n} (\gamma_i - \gamma_i^*) K\langle x_i, x \rangle + b \quad (10)$$

where γ_i and γ_i^* are the Lagrange multipliers, K is the kernel function

2.5. Neural Network Model

It is an advanced machine learning model whose computing structure and function are similar to those of biological neural networks in animals. It consists of interconnected nodes (neurons) of layers that can learn to recognize patterns in data trends. It is made up of a minimum of one hidden layer and an output layer. Each layer contains several neuronees, and the neuronees of each layer are interconnected to every other neurons in other layers. The activation function in equation (11) is performed by each neurone in the network layer as the inputs weighted sum. For instance, the a_j for a neuron j in the network layer l is given by;

$$a_j^{(l)} = \phi\left(\sum_i w_{i.j}^{(l)} a_i^{(l)} + b_j^{(l)}\right) \quad (11)$$

Where ϕ is the activation function, $w_{i.j}^{(l)}$ is the weight connecting neuron i in layer $l-1$to neuron j in layer l, $l, a_i^{(l-1)}$ is the output of neuron i in the previous layer, and $b_j^{(l)}$ is the bias term. The loss function L which determines the performance of the neuron network is obtained by comparing the network output $\hat{y}$ with the true target$y$. Equation (12) is the loss function used for the regression analysis [27] and [28].

$$L(y, \hat{y}) = \frac{1}{2}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (12)$$

The loss function gradients with respect to the weights are computed using equation (13) and the result is used to update the weights to minimize the loss.

$$\frac{\delta L}{\delta \omega_{i,j}^{(l)}} = \delta_j^{(l)} a_i^{(l-1)} \quad (13)$$

where $\delta_j^{(l)}$ is the error term for neuron $j$in layer l

2.6. K-Nearest Neighbors (KNN) model

K-Nearest Neighbors (KNN) algorithm is a non-parametric, supervised machine learning model that stores all available data and makes predictions based on the most similar instances in the training data. It is commonly used for regression and classification [29] and [30]). The distance metrics such as Euclidean distance in equation (14) is used to find the k nearest neighbor.

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^{p} (x_{im}, x_{jm})^2} \quad (14)$$

where $x_i andx_j$are two points in p-dimensional feature space. The algorithm computes the distance between a test point, say x and all the points in the training set and picks the k point with the smallest distances for regression analysis. The target value y_i is predicted by averaging the target values of the $i-th$ k nearest neighbor, thus the predicted value $\hat{y}_i$ is given by equation (15) [30].

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

$$\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i \quad (15)$$

3. Methodology

Ten years of earth surface atmospheric daily data for the two stations were downloaded from the Era-5 Climate Data Store of the European Centre for Medium-Range Weather Forecast (ECMWF). The atmospheric variables deployed from the dataset are: earth surface temperature in Kelvin, earth's surface pressure in Pascal, earth's surface due point temperature in kelvin, wind speed magnitude in meters per second, total cloud cover in 0-1, total precipitation in meters, earth's surface solar radiation in joules per square meter, total precipitation in meters, and relative humidity in percentage. The corresponding daily mean rain heights for the study periods and stations were also computed from zero-degree isotherm heights using the ITU-R P.839-3 procedure in [31] with the aid of the Python programming language. The atmospheric variables were correlated with rain height to study the relationship between each of the atmospheric surface variables and rain height. The scatter plots were also generated for visualization. It was preprocessed by handling missing values and outliers. Rain heights were modelled using six deep-learning models by applying the surface data as the independent input variables. The advanced machine learning models applied are Random Forest, Gradient Boosting, XGBoost, Support Vector Regressor, Neural Network, and K-Nearest Neighbors. The parameters applied to each model are presented in Table 2. The data was split into train and test sections. The models were trained using different subsets of features variables. The performance of each model was evaluated and compared using performance metrics such as R^2, RMSE, MAE, and cross-validation.

Table 2.
The machine learning models and the applied parameter values.

	Models	Parameters	Applied values
1	Random Forest	n_estimators, max_depth, min_sample_list, min_sample_leaf	200, 20, 2, 1
2	Gradient Boosting	n_estimators, learning_rate, max_depth, subsample	200, 0.1, 5, 0.9
3	XGBoost	n_estimators, learning_rate, max_depth, subsample,	200, 0.1, 5, 0.9
4	Neural Network	hidden_layer_sizes, activation, solver, alpha	(100, 50), relu, adam, 0.0001
5	SVR	C, gamma, kernel	1, scale, rbf
6	KNN	n_neigbors, weights, metric	5, distance, euclidean

The performance of the best model among the rest was improved by gridsearch hyper parameter tuning and feature engineering through date inclusion. The influence and contribution of each input variables was identified and analyze feature importance analysis for the most preferred model

4. Results

4.1. Seasonal Trend of Rain Heights

The time series plots of the rain height data used for the research are presented in Figures 1–6. It was observed that the South African stations exhibited a sinusoidal trend pattern over the years, while the Nigerian stations produced apparently irregular patterns. This could be attributed to the frequent rainfall associated with the tropical regions, as reported by [20]. Although heavy precipitations are usually experienced during the rainy season, occasional rainfalls occur in almost all the dry season months, especially in the coastal stations. The dominant influence of temperature on rain height, which was discussed in subsequent sections, is also responsible for the clear crest and trough pattern in South

Africa. Extreme minimum and maximum temperatures are usually experienced during the winter and summer, respectively. For instance, Pretoria has a minimum and maximum annual temperature of 5

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

°C and 30 °C, respectively, which implies a difference of 25 °C. Polokwane and Cape Town exhibit similar differences [32-34]. On the contrary, the stations in Nigeria possess lower differences. Akure has a maximum and minimum annual temperature of about 22 °C and 33 °C, respectively. The difference between the two extremes annually in Port Harcourt is about 8 °C. Sokoto has similar differences [36] and [37].

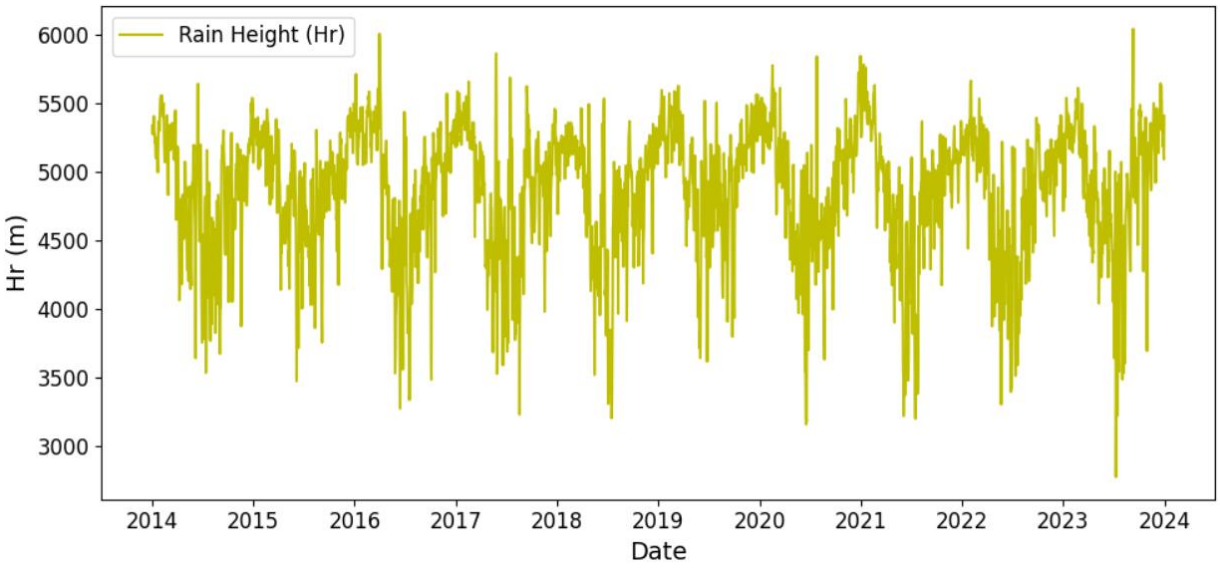


Figure 1.
Seasonal trend of rain heights in Polokwane.

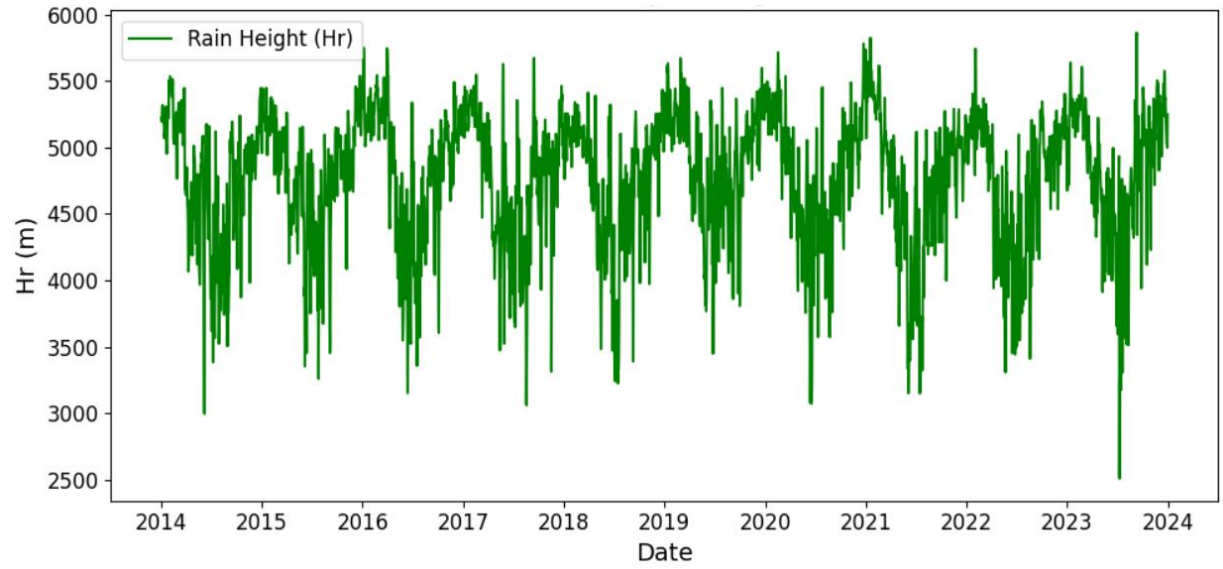


Figure 2.
Seasonal trend of rain heights in Pretoria.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

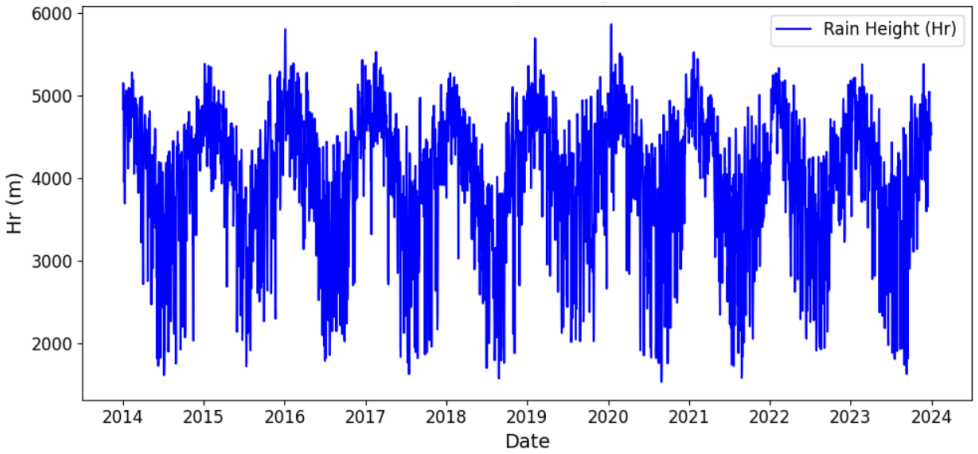


Figure 3.
Seasonal trend of rain heights in Cape Town.

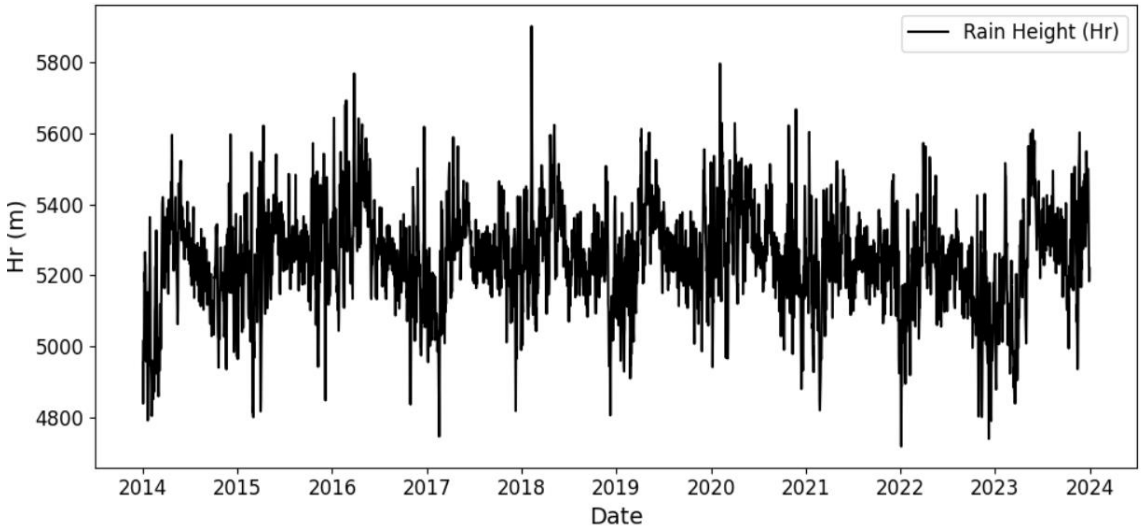


Figure 4.
Seasonal trend of rain heights in Sokoto.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

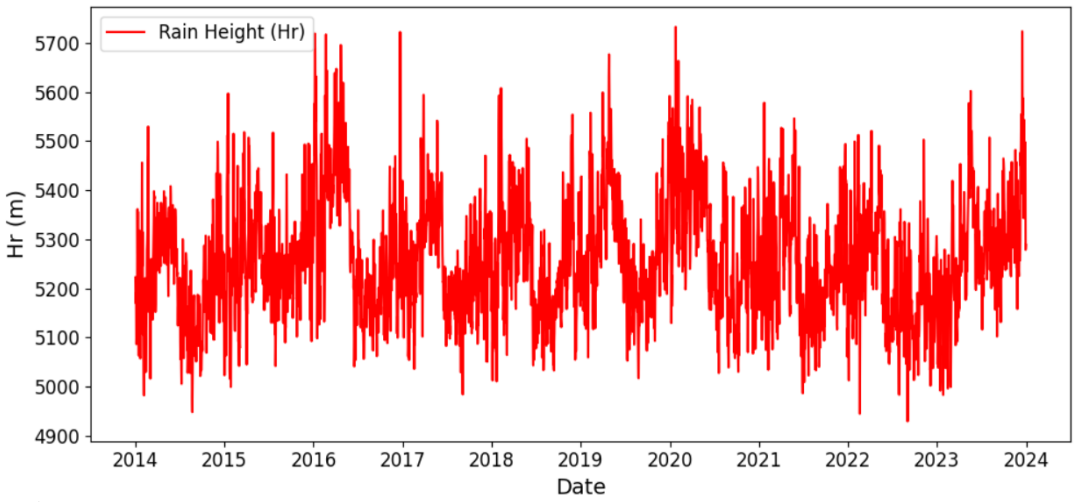


Figure 5.
Seasonal trend of rain heights in Akure.

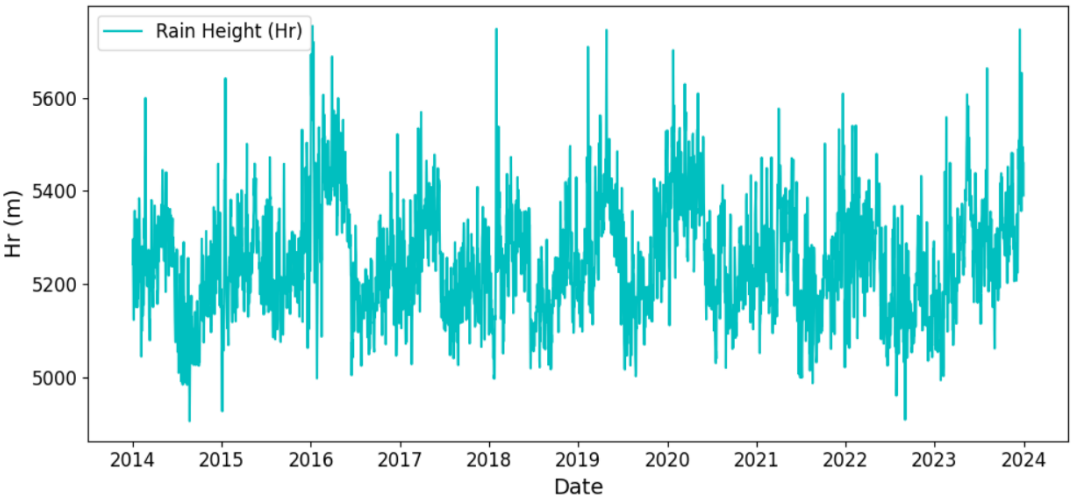


Figure 6.
Seasonal trend of rain heights in port Harcourt (PH).

4.2. Comparison of the Models

The six models described above were applied to all the stations and the results obtained are presented in Table 3-8 and Figures 7-12. Comparison of the performance metrics revealed that gradient boosting model performed better across all the stations. The performance of the gradient boosting Regressor was further improved by hyper-parameter tuning and feature engineering. The grid search hyper-parameter method was applied with n_estimators, learning_rate and max_depth, of 100, 0.2, 8 respectively.

Tables 3-8 present the values of the performance metrics for all the studied stations while Figures 7-12 give the graphical representation. The stations in South Africa are well favoured by the models with gradient boosting model having the best performance. The regular trend pattern observed for all South Africa stations in the section 4.1 is responsible for the optimal fitting of the models. Cape Town has the highest determination co-efficient of 0.84 and minimum RMSE and MAE of 336.51 m and 262 m respectively.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

The co-efficient of determination r^2 and other performance metrics for the stations in the Nigeria indicated that none of these models is good for rain height prediction in the tropical region.

Table 3.
Comparison of models performance metric for Polokwane.

Models	R^2	MSE	RMSE	MAE
Random forest	0.764	55460	235.50	170
Gradient boosting	0.800	56659	238.03	173
XGBoost	0.767	54593	233.65	167
Support vector regressor	0.578	99125	314.84	224
Neural network	0.707	68900	262.49	192
K-Nearest neighbors	0.620	89015	298,36	216

Table 4.
Comparison of models performance metric for Pretoria.

Models	R^2	MSE	RMSE	MAE
Random forest	0.777	56365	237.41	167
Gradient boosting	0.816	45390	213.05	150
XGBoost	0.783	54808	234.11	164
Support vector regressor	0.591	103317	321.43	227
Neural network	0.728	68782	262.27	190
K-Nearest neighbors	0.654	87551	295.89	215

Table 5.
Comparison of models performance metric for Cape Town.

Models	R^2	MSE	RMSE	MAE
Random Forest	0.786	125239	353.89	274
Gradient boosting	0.841	113236	336.51	262
XGBoost	0.790	122919	350.60	272
Support vector regressor	0.613	226837	476.27	354
Neural network	0.778	129957	360.50	278
K-nearest neighbors	0.703	174238	417.42	329

Table 6.
Comparison of models performance metric for Sokoto.

	R^2	MSE	RMSE	MAE
Random forest	0.379	12441	111.54	85
Gradient boosting	0.323	13570	116.49	87
XGBoost	0.332	13396	115.74	87
Support vector regressor	0.185	16339	127.83	98
Neural network	0.094	18178	134.83	104
K-Nearest neighbors	0.162	16796	129.60	98

Table 7.
Comparison of models performance metric for Akure.

Models	R^2	MSE	RMSE	MAE
Random forest	0.456	7652	87.48	68
Gradient boosting	0.447	7772	88.16	68
XGBoost	0.437	7919	88.99	68
Support vector regressor	0.358	9028	95.02	78
Neural network	0.117	12424	111.47	82
K-Nearest neighbors	0.357	9051	95.14	74

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

Table 8.
Comparison of models performance metric for port Harcourt (PH).

Models	R^2	MSE	RMSE	MAE
Random forest	0.458	7496	86.58	65
Gradient boosting	0.455	7503	86.62	65
XGBoost	0.456	7496	86.58	65
Support vector regressor	0.412	8099	90.00	68
Neural network	0.219	10763	103.75	80
K-Nearest neighbors	0.281	9901	99.51	76

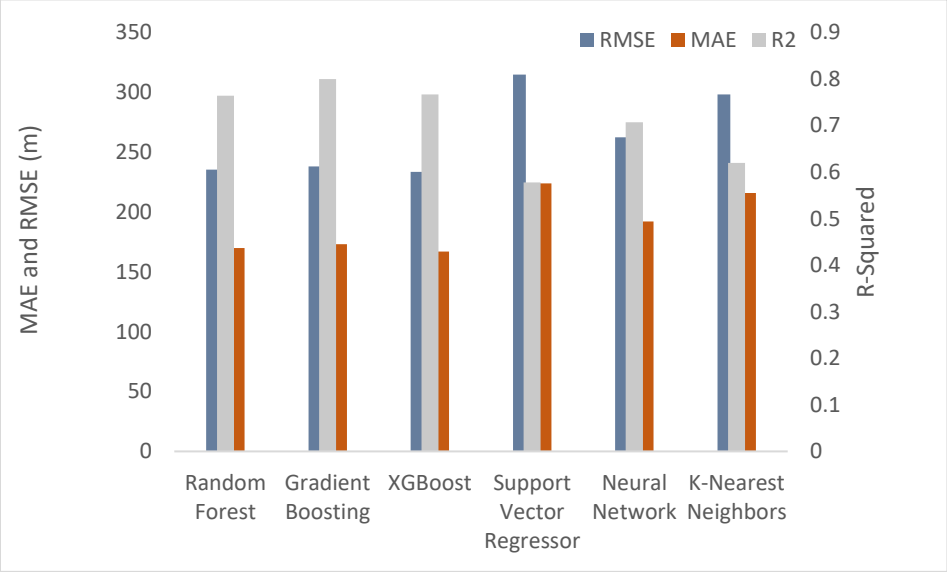


Figure 7.
Performance metrics of the models for Polokwane.

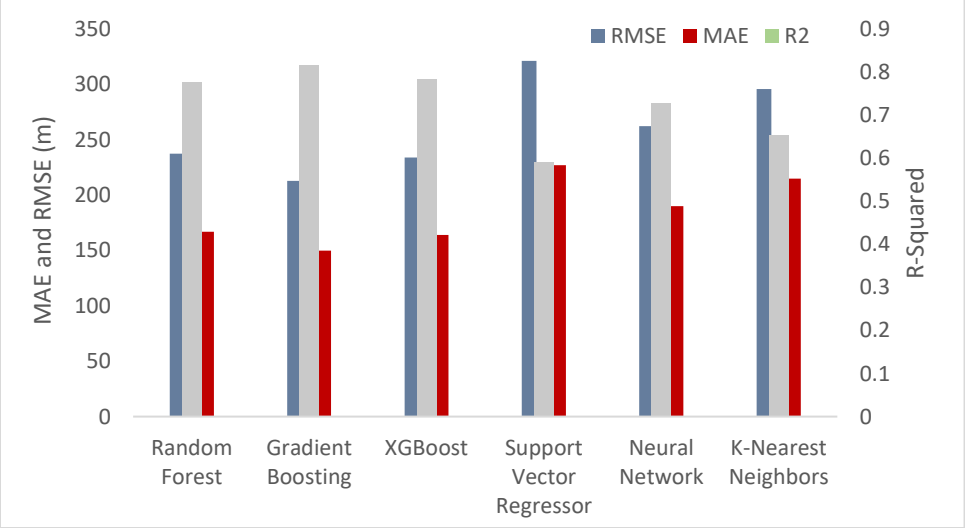


Figure 8.
Performance metrics of the models for Pretoria.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

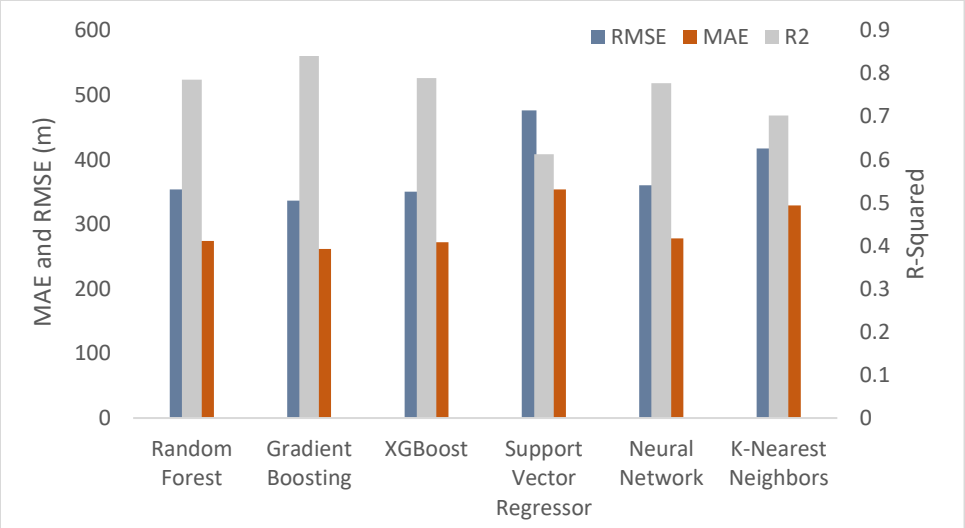


Figure 9.
Performance metrics of the models for cape Town.

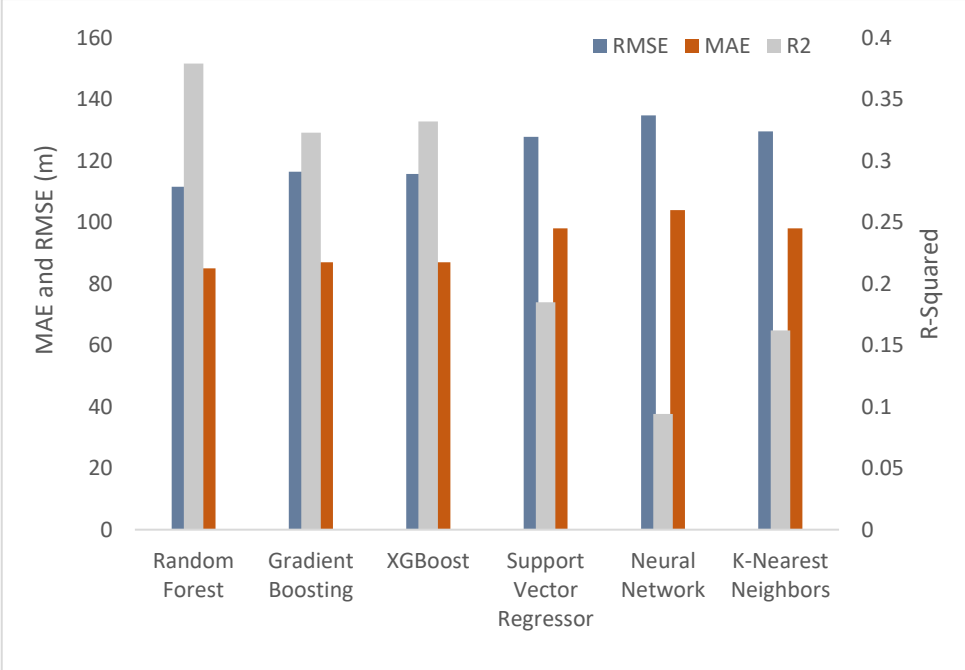


Figure 10.
Performance metrics of the models for Sokoto.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

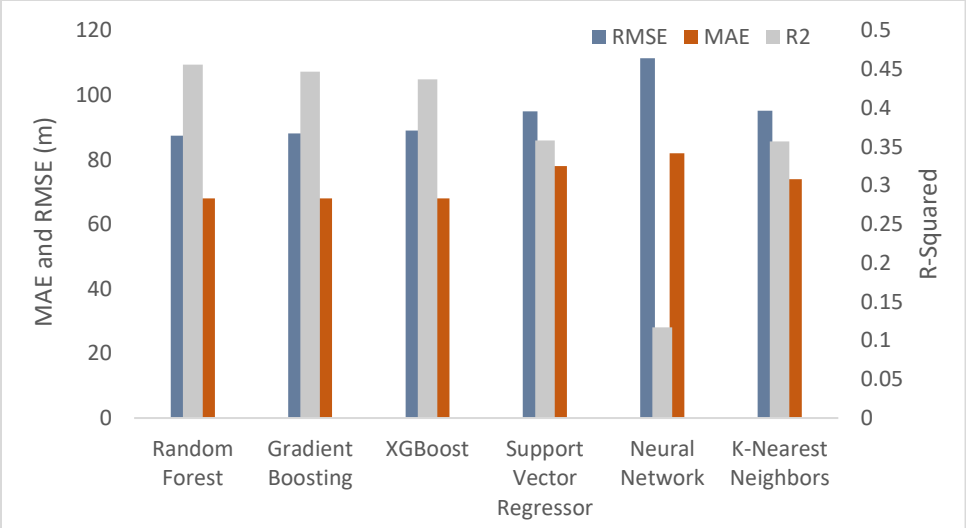


Figure 11.
Performance metrics of the models for Akure.

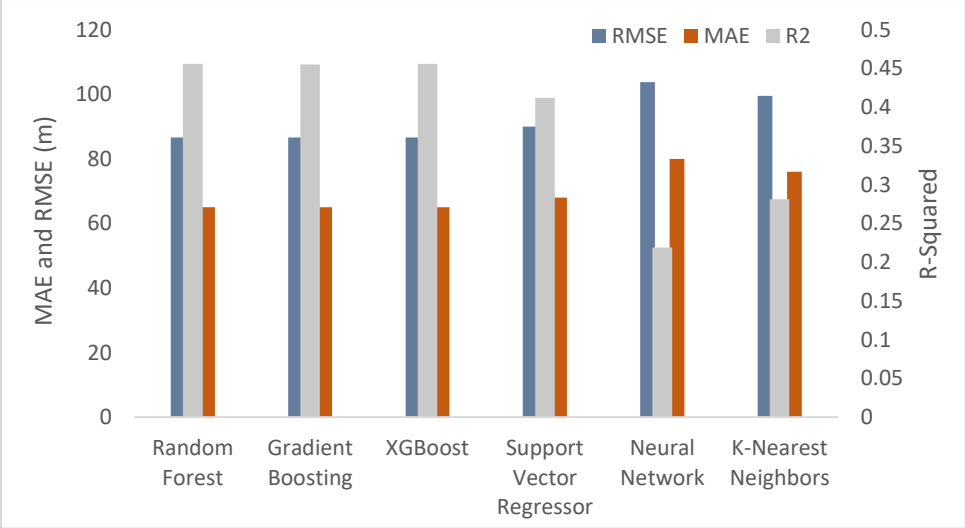


Figure12.
Performance metrics of the models for PH.

4.3. Correlation between Atmospheric Variables and Rain Height

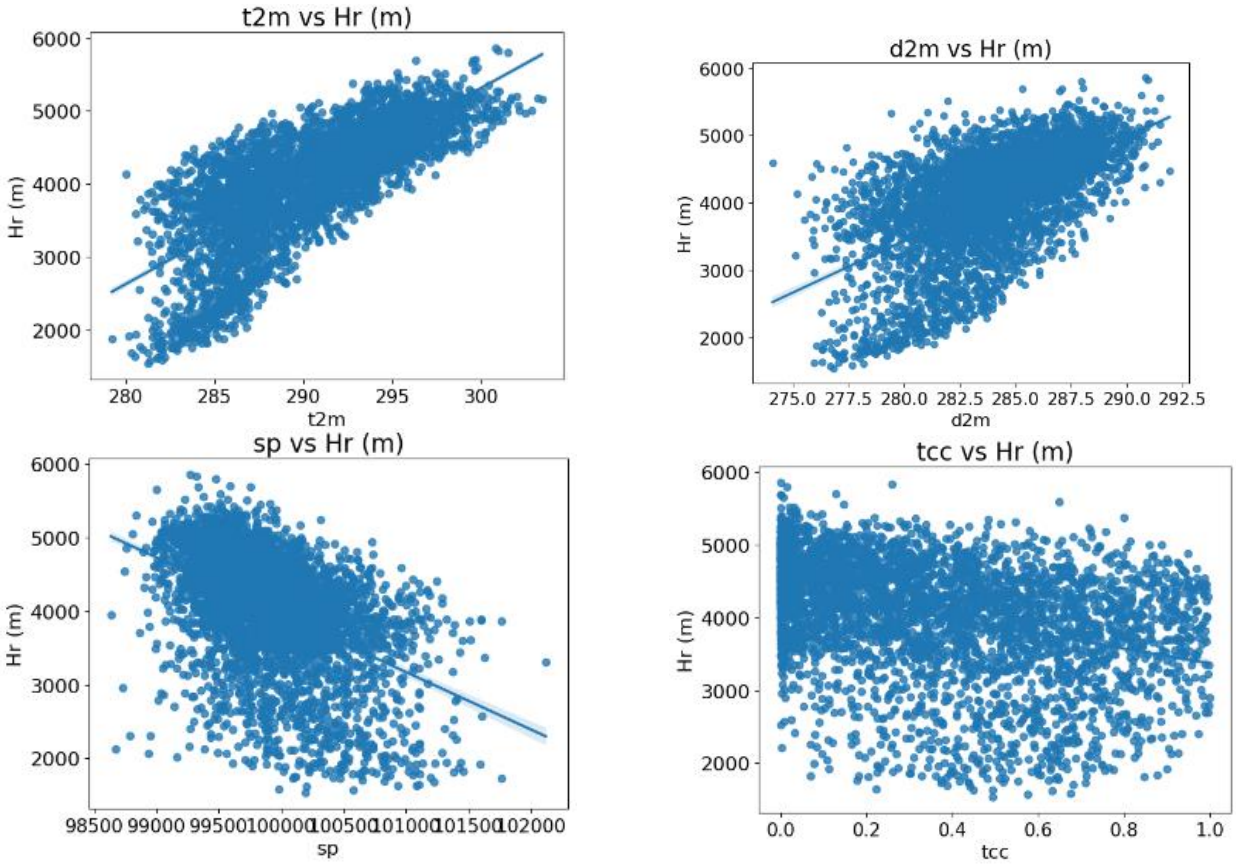
The relationship between the input variables and rain height was investigated through correlation. Table 9 presents the correlation coefficient between each of the atmospheric earth surface variables and rain height for a typical station (Cape Town), since similar results were obtained for all stations. Figures 13a–h indicate the scatter plots of the variables. Surface temperature and due point temperature have very strong positive correlations of 0.79 and 0.59 on rain height, as presented in Table 9. This is due to the effect of solar radiation on the melting layer. An increase in temperature reduces the bright band and raises the height of the water region in the atmosphere. The path travelled by water droplets before the earth surface becomes higher, i.e., the rain eight, However, this does not imply that the total volume of water reaching the earth's surface is increased proportionately, thus a weak correlation of about 0.41. Surface pressure and Earth's surface solar radiation also have similar effects, as they both enhance the melting of the ice, thus increasing the rain height. This observation was corroborated by [38] and [39],

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

who reported strong positive correlations between the pair of meteorological observables. Oher variables which exhibit positive but very poor correlation are total cloud cover, total precipitation.

Table 9.
Correlation co-efficient between earth surface variables and rain height on a typical day for cape town station.

Earth surface variables (Unit)	Symbol	Correlation co-efficient (r^2)
Surface temperature (K)	t2m	0.76
Due point temperature (K)	d2m	0.59
Surface pressure (Pa)	sp	-0.47
Surface solar radiation (J/m^2)	ssrd	0.62
Total cloud cover (0-1)	tcc	0.36
Total precipitation (m)	tp	0.41
Wind speed (m/s)	wind_speed	-0.04
Relative humidity (%)	relative_humidity	-0.51



Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

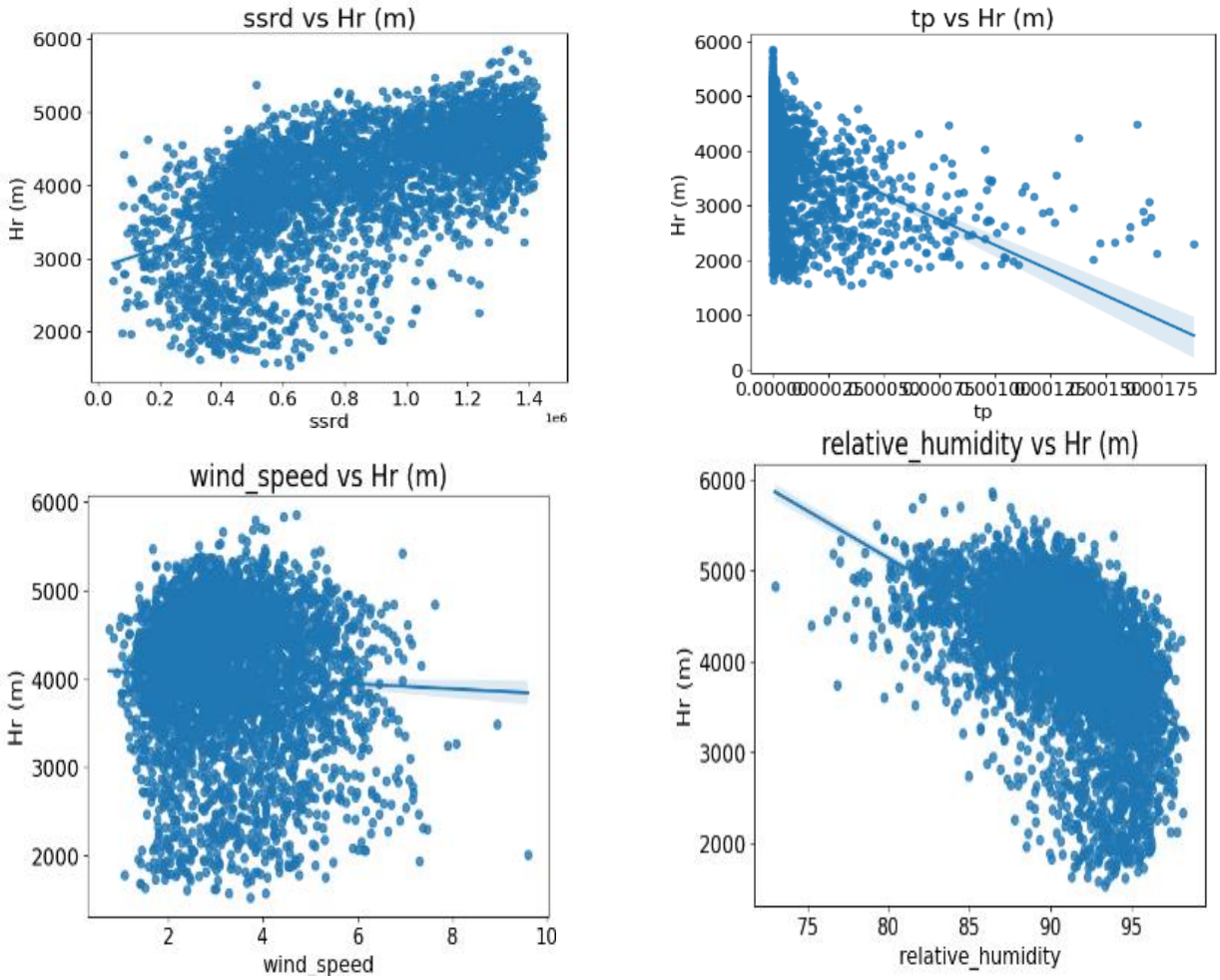


Figure13a-h.
The scatter plots of the atmospheric variables on rain height for cape town station.

4.4. Feature Importance of the Input Variables

The contributions of each input variable used for the modeling were analyzed and weighted through feature importance calibrations. The results obtained are presented in Figures 14 (e.g). Among all the surface atmospheric variables, the earth surface temperature makes over 60% contribution to the regression fit across all the stations, as depicted in Figures 14 (e-g). This implies that temperature has the most significant influence on rain height variability. This is a manifestation of the strong positive correlation reported in Section 4.3. All other variables contribute less than 20% to the rain height model. For instance, in Polokwane, the total precipitation, surface pressure, earth surface due point temperature, wind speed, relative humidity, and total clover contribute about 15.4%, 15.1%, 13.8%, 8.5%, 6.1%, 2.8%, and 2.2%, respectively.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

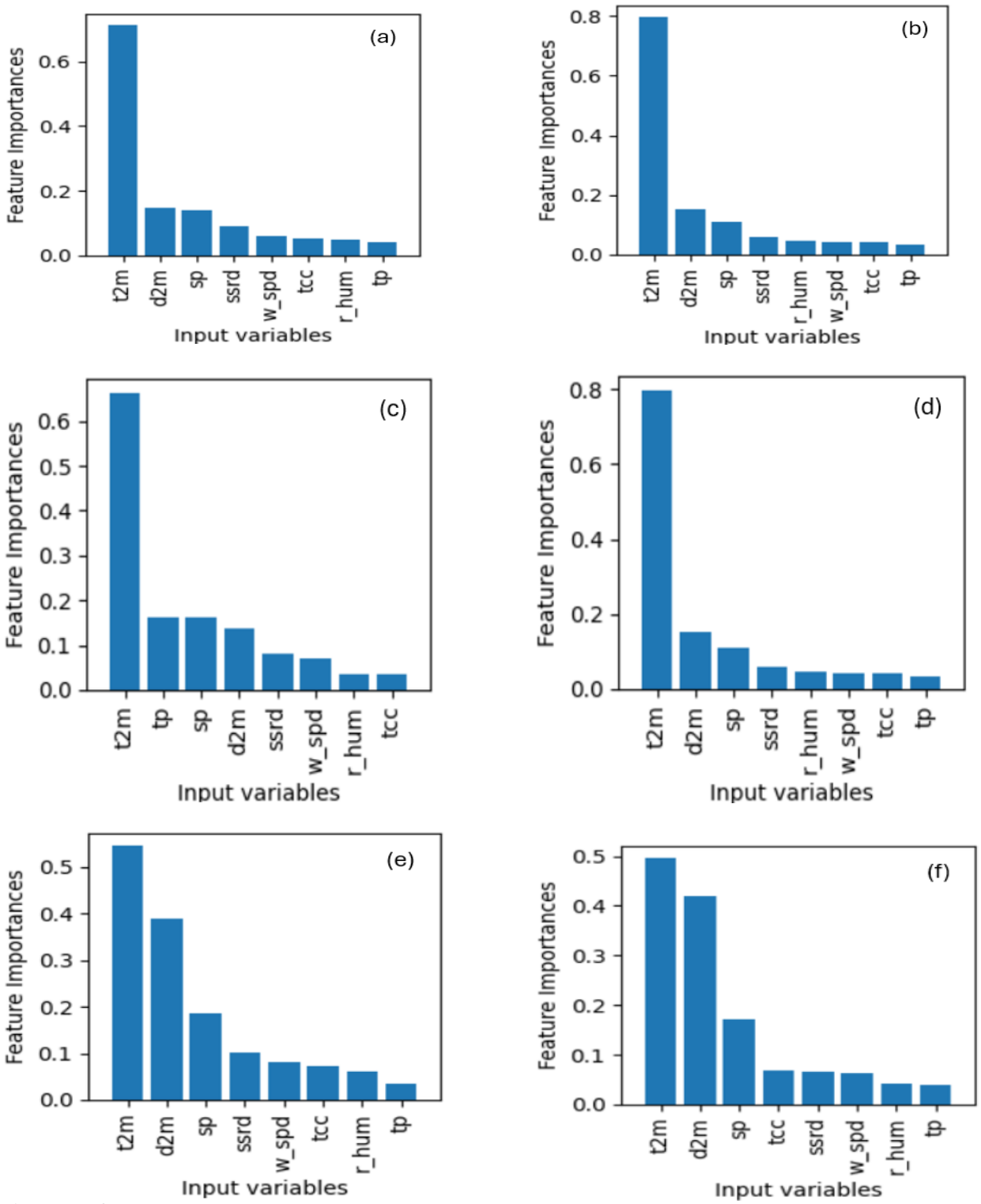


Figure.14a-f
The feature importance of each input variable in the model for; (a) Polokwane, (b) Pretoria, (c) Cape Town, (d) Sokoto, (e) Akure, and (f) Port Harcourt (in progress)

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 520-537, 2024
DOI: 10.55214/25768484.v8i6.2117

5. Conclusion

The study examines the annual trend of rain height over the studied years. Findings show that South African stations possess a well-defined sinusoidal wave-like pattern, with peak and trough values occurring in the summer (December–January) and winter (May–August), respectively. In Nigeria, rain height is generally low during the rainy season and high in the dry season. Generally, temperature has the greatest influence on rain height due to the high rate of ice melting in the bright band during solar active periods. The prediction accuracy of the machine learning models increases as we move away from the equator. All the South African stations are situated farther away from the equator compared to the Nigerian stations; hence, better performance metrics were observed. The gradient boosting model out performed other five models considered in all the stations. Cape Town, which is the farthest station from the equator, possesses the best performance metrics with r^2, MSE, RMSE, and MAE of 0.84, 113,236 m^2, 336.5 m, and 262 m, respectively. Overall, the research provides an alternative means of computing rain height for better prediction of rain-induced attenuation, which is a major threat to earth-space on radio links.

Funding:
This research is supported by Tshwane University of Technology South Africa.

References

[1] Y. Wen, L. Liao, D. Wolff, R. Meneghini, and T. Schuur. "Joint Collaboration on Comparing NOAA's Ground-Based Weather Radar and NASA–JAXA's Spaceborne Radar," *Bulletin of the American Meteorological Society, vol. 104, no. 8. 2023.* E1435–E1451. https://doi.org/10.1175/BAMS-D-22-0127.1

[2] L. Liao, D. Wolff, R. Meneghini, and T. Schuur, T. "Systematic Differences in Reflectivity between NOAA WSR-88D and NASA–JAXA DPR," *Journal of Applied Meteorology and Climatology.*

[3] Z. Yu, X. Zheng, and T. Sun. "Three-Dimensional Slope Imaging Method for Ground-Based Real-Aperture Radar. *Sensors, vol. 21, no.* 10. 2021 https://doi.org/10.3390/s21103511

[4] Y. Hong, K. Hsu, and S. Sorooshian. "Remote Sensing Precipitation: Sensors, Retrievals, Validations, and Applications," *SpringerLink.* 2022. https://doi.org/10.1007/978-3-319-07184-0

[5] Y. Zhao, X. Liu, M. Xian, and T. Gao. "Statistical Study of Rainfall Inversion Using the Earth-space Link at the Ku Band: Optimization and Validation for 1 Year of Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 9486-9494, 2021

[6] D. A. Short, and K. Nakamura.2000). "TRMM Radar Observations of Shallow Precipitation over the Tropical Oceans," *Journal of Climate*, vol. 13, no. 23, pp. 4107-4124, 2000. http://dx.doi.org/10.1175/1520-0442(2000)013<4107:TROOSP>2.0.CO;2

[7] T. Iguchi, N. Kawamoto, and R. Oki. "Detection of Intense Ice Precipitation with GPM/DPR," *America Meteorological Society.* Vol. 35, no. 3, pp. 491-502, 2018. DOI: https://doi.org/10.1175/JTECH-D-17-0120.1

[8] M. Kojima, T. Miura, K. Furukawa, Y. Hyakusoku. T. Ishikiri, H. Kai, T. Iguchi, and H. Hanado. "Dual-frequency precipitation radar (DPR) development on the global precipitation measurement (GPM) core observatory," *Proc. SPIE 8528, Earth Observing Missions and Sensors: Development, Implementation, and Characterization* II, 85281A, 2012 https://doi.org/10.1117/12.976823

[9] D. S. Zrnic, and A. V. Ryzhkov. "Polarimetry for weather surveillance radars," *Bulletin of the American Meteorological Society*, vol. 80, no. 3, pp. 389-406, 1999 https://doi.org/10.1175/1520-0477(1999)080<0389>2.0.CO;2

[10] R. Meneghini, H. Kim, L. Liao, J. A. Jones, J. M. Kwiatkowski. "An initial assessment of the surface reference technique applied to data from the Dual-Frequency Precipitation Radar (DPR) on the GPM satellite," *Journal of Atmospheric and Oceanic Technology, vol. 32*, no. 11, pp. 2281-2296, 2015 https://doi.org/10.1175/JTECH-D-15-0044.1

[11] J. S. Mandeep, "Atmospheric effects on Ku-band satellite communication links during the Malaysian rainy season," *International Journal of Satellite Communications and Networking, vol. 26, no. 5*, pp.447-462, 2008. https://doi.org/10.1002/sat.893

[12] Y. B. Lawal, J. S. Ojo, S. E. Falodun, and E. O. Olurotimi. "Geoclimatic characterization and latitudinal dependence of rain heights over Nigeria," *Journal of Physics: Conference Series. 2034* 012010, 2021. IOP Publishing https://doi.org/10.1088/1742-6596/2034/1/012010

[13] K. Nalinggam, H. Ahmad, and S. Abdullah, "The role of satellite-derived data in rainfall estimation over the tropics: case study over Malaysia," *Hydrological Processes, vol. 27, no.* 24, pp. 3453-3465, 2013. https://doi.org/10.1002/hyp.9433

[14] J. Goldhirsh, and I. Katz, "Useful experimental results for earth-satellite rain attenuation modelling," *IEEE Trans Antennas & Propagation*, AP-27 pp. 413–415, 1979

[15] G. O. Ajayi, and F. Barbaliscia. "Prediction of Attenuation Due to Rain: Characteristics of the 0 °C Isotherm in Temperate and Tropical Climates," *Int. J. of Sat. Comm*, vol. 8, pp. 187–196, 1990.

[16] E. Matricciani, "Rain attenuation predicted with a two layer rain model," *Electron Lett (UK)*, vol. 29, pp. 72–73, 1993.

[17] A. I. O. Yussuff, and N. H. H. Khamis, "Comparative analysis of bright band data from TRMM and ground radar data in Malaysia," *International Journal of Networks and Communications*, vol. 3, Issue 4, pp. 99-109, 2013.

[18] O. J. Olaniran, and G. N. Sumner, "A Study Of Climatic Variability In Nigeria Based On The Onset, Retreat, And Length Of The Rainy Season," *International Journal Of Climatology*, vol. 9, pp. 253-269, 1988

[19] N. P. Iloeje. "A New Geography of Nigeria," *New Revised Edition, Longman Nig. Ltd., Lagos*, p. 200, 2001.

[20] A. Taiwo, E. U. Utah, and T. E. Sombo. "Monthly Variation and Annual Trends of Rainfall across Major Climatic Zones in Nigeria," *IOSR Journal of Applied Physics, e-ISSN: 2278-4861*, vol. 10, no. 4, pp. 15-28, 2018.

[21] J. H. Friedman. "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189-1232, 2001. https://doi.org/10.1214/aos/1013203451

[22] J. H. Friedman. "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367-378, 2002. https://doi.org/10.1016/S0167-9473(01)00065-2

[23] T. Chen, and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016

[24] A. J. Smola, and B. Schölkopf. "A tutorial on support vector regression," *Statistics and Computing*, vol.14, pp. 199-222, 2004. https://doi.org/10.1023/B.0000035301.49549.88

[25] V. Vapnik. "The Nature of Statistical Learning Theory," *Springer-Verlag New York*, Inc. 1st ed., 1995. https://doi.org/10.1007/978-1-4757-2440-0

[26] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik. "Support Vector Regression Machines," *Advances in Neural Information Processing Systems*, vol. 9, pp. 155-161, 1997. https://doi.org/10.5555/2998981.2999003

[27] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533-536, 1986. https://doi.org/10.1038/323533a0

[28] I. Goodfellow, Y. Bengio, and A. Courville. "Deep Learning" *MIT Press*, Available online: http://www.deeplearningbook.org/

[29] N. S. Altman. "An Introduction to Kernel and Nearest Neighbor Nonparametric Regression," *The American Statistician*, vol. 46, pp. 175-185, 1992.

[30] T. Hastie, R. Tibshirani, and J. Friedman. "The Elements of Statistical Learning: Data Mining," *Inference, and Prediction, 2nd ed.* Stanford, CA: Stanford University, 2009. https://web.stanford.edu/~hastie/Papers/ESLII.pdf https://doi.org/10.1007/978-0-387-84858-7

[31] ITU-R P.839-3. "Rain height model for prediction methods," *P Series Radiowave Propagation. International Telecommunication Union, Radiocommunication Section (ITU-R)*, Geneva, Switzerland, P.839-3, 2001.

[32] P. D. Tyson, and R. A. Preston-Whyte. "The Weather and Climate of Southern Africa" *Oxford University Press*. 2000.

[33] A. C. Kruger, and S. Shongwe. "Temperature trends in South Africa 1960–2003," *International Journal of Climatology*, vol. 24, no. 15, pp. 1929-1945, 2004. https://doi.org/10.1002/joc.1096

[34] M. New, B. Hewitson, D. B. Stephenson, A. Tsiga, A. Kruger, A. Manhique, and R. Lajoie, "Evidence of trends in daily climate extremes over southern and west Africa," *Journal of Geophysical Research: Atmospheres*, vol.111, no. D14, 2006. https://doi.org/10.1029/2005JD006289

[35] R. K. Udo. "Geographical Regions of Nigeria," *Heinemann Educational Books*, London. 1970.

[36] A. A. Adebayo. "The Environment and Climate Change," *The Costs of Inaction. Environment Institute report*, 2011

[37] E. A. Adefisan, and A. A. Abatan. "Rainfall distribution and change detection across climatic zones in Nigeria," *Journal of Geography and Regional Planning*, vol. 8, no. 4, pp. 123-135. https://doi.org/10.5897/JGRP2014.0432

[38] K. Paulson, and A. Al-Mreri. "Trends in the Incidence of Rain Height and the Effect on Global Satellite Telecommunication," *IET Microwaves Antennas & Propagation*, vol. 5, no. 14, pp. 1710-1713, 2011. https://doi.org/10.1049/iet-map.2010.0507

[39] Y. B. Lawal, S. E. Falodun and J. S. Ojo. "Temporal Evolution of Atmospheric Parameter-Profiling on Rain Height Over Two Geoclimatic regions in Nigeria," *Journal of Atmospheric and Solar Terrestrial Physics (JASTP)*. Vol. 211, 105482. 1364-6826/©2020 Elsevier Ltd, 2020. https://doi.org/10.1016/j.jastp.2020.105482