**UNIVERSITE JOSEPH KI-ZERBO**

**------------**

**ECOLE DOCTORALE INFORMATIQUE ET CHANGEMENTS CLIMATIQUES**

**BURKINA FASO**

*Unité-Progrès-Justice*

# MASTER RESEARCH PROGRAM

## SPECIALITY: INFORMATICS FOR CLIMATE CHANGE (ICC)

# MASTER THESIS

**Subject:**

# Monthly Rainfall Prediction Using Artificial Neural Network (Case Study: Republic of Benin)

by

**Arsène Nounangnon AIZANSI**

**Major Supervisor:**

Prof. Kehinde O. OGUNJOBI

**Co-Supervisor:**

Dr. Faustin K. OGOU

**Academic year: 2020 - 2021**

# Summary

# Dedication

This thesis is dedicated

To

My beloved parents and my family.

# Acknowledgements

First, I want to thank the Almighty God for His grace and blessings on my life throughout this program.

I would like to express my gratitude to the following individuals and organizations for their contributions to my academic success:

- Prof. Tanga Pierre ZOUNGRANA, Director of the "Ecole Doctorale Informatique et Changement Climatique" (ED-ICC).
- Dr. Ousmane COULIBALY, Deputy Director of ED-ICC.
- Prof. Kehinde O. OGUNJOBI, Director of Research at the WASCAL Competence Center, Burkina Faso, who was my main supervisor.
- Dr. Faustin K. OGOU, co-supervisor.
- West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL).
- Germany's Federal Ministry of Education and Research.
- Agence Nationale de la Météorologie (METEO-BENIN).

Finally, I'd like to thank all of my colleagues in the Informatics for Climate Change for their support, special thanks to Issa KASSOGUE for his useful suggestions and recommendations. I also appreciate Florence KAYODE for facilitating the acquisition of MATLAB software.

# Abstract

Complex physical processes that are inherent to rainfall, lead to the challenging task of its prediction. To contribute to the improvement of rainfall forecasting, artificial neural network (ANN) models were developed to predict monthly rainfall for 6 geographically diverse weather stations across Benin Republic using data from January 1979 to December 2018. Initially, correlation-based feature selection is used to find the most suitable set of predictors for predicting rainfall after identifying the optimal lag times of eight (8) predictors. To compare the forecasting ability of the ANN models, Multiple Linear Regression (MLR) models were developed as a benchmark. The prediction performance was evaluated using root mean squared error, mean absolute error, coefficient of determination ($R^2$) and Nash-Sutcliffe Efficiency. The results revealed that ANN gives better results than MLR methods. The $R^2$ obtained from ANN models during the test period (2015 to 2018) ranges from 0.47 to 0.87. Rainfall predictability was more accurate with 0.4 improvement in $R^2$ at higher latitudes across the country, showing the effect of geographic regions on prediction model results. The input contribution analysis done shows that sea surface temperature, zonal and meridional wind are the most important variables that contribute more to rainfall. In summary, this research has revealed the potential of artificial neural network techniques in predicting monthly rainfall in the Republic of Benin.

**Key words:** Artificial neural network; rainfall prediction; predictors; prediction performance; input contribution analysis.

# Résumé

Les précipitations sont difficiles à prévoir en raison des processus physiques complexes impliqués. Pour contribuer à l'amélioration de la prévision des précipitations, des modèles de réseau de neurones artificiels (RNA) ont été développés pour prédire les précipitations mensuelles de 6 stations météorologiques géographiquement diverses à travers le Bénin en utilisant des données de Janvier 1979 à Décembre 2018. Dans un premier temps, la sélection de prédicteurs basée sur la corrélation est utilisée pour trouver le groupe approprié de prédicteurs après avoir identifié les temps de latence optimaux de huit (8) prédicteurs. Pour comparer la performance de prévision des modèles RNA, des modèles de régression linéaire multiple (RLM) ont été développés comme référence. La performance de prédiction a été évaluée en utilisant l'erreur quadratique moyenne, l'erreur absolue moyenne, le coefficient de détermination ($R^2$) et l'efficacité de Nash-Sutcliffe. Les résultats ont révélé que l'RNA donne de meilleurs résultats par rapport à RLM. $R^2$ obtenu à partir des modèles RNA pendant la période de test (2015 à 2018) varie de 0,47 à 0,87. La prévisibilité des précipitations était plus précise avec une amélioration de 0,4 du $R^2$ à des latitudes plus élevées à travers le pays, montrant l'effet de la topographie sur les résultats des modèles de prédiction. L'analyse de la contribution des prédicteurs a montré que la température de surface de la mer, le vent zonal et méridional sont les variables les plus importantes qui contribuent le plus aux précipitations. En résumé, cette recherche a révélé le potentiel des techniques de réseau de neurones artificiels dans la prévision des précipitations mensuelles en république du Bénin.

**Mots clés:** Réseau de neurones artificiels; prévision des précipitations; prédicteurs; performance des prédictions; analyse de la contribution des prédicteurs.

# List of Tables

# List of Figures

# List of Acronyms

| Acronyms | Definition |
|---|---|
| ACMAD | : African Center of Meteorological Application for Development |
| AGRHYMET | : Agriculture, Hydrology and Meteorology Research Center |
| ANFIS | : Adaptive Network-Based Inference Systems |
| ANN | : Artificial Neural Network |
| ARIMA | : Autoregressive Integrated Moving Average |
| ARMA | : Autoregressive Moving Average |
| ART | : Adaptive Resonance Theory |
| BP | : Back Propagation |
| BPA | : Back Propagation Algorithm |
| CBP | : Cascaded Back Propagation |
| CDS | : Climate Data Store |
| CFS | : Correlation-based Feature Selection |
| $CO_2$ | : Carbon Dioxide |
| CPT | : Climate Predictability Tool |
| DMI | : Dipole Mode Index |
| ENSO | : El Nino Southern Oscillation |
| ERNN | : Elman recurrent neural network |
| ES | : Exponential Smoothing |
| Eva | : Evaporation |
| FFNN | : Feed Forward Neural Network |
| FS | : Feature Selection |
| GARCH | : Generalized Autoregressive Conditional Heteroscedasticity |
| GDP | : Gross Domestic Product |
| GHG | : Green House Gas |

| | |
|---|---|
| GRNN | : Generalized Regression Neural Network |
| HFM | : Hopfield |
| IDNN | : Input delay neural network |
| IOD | : Indian Ocean Dipole |
| IPCC | : Intergovernmental Panel on Climate Change |
| IPO | : Inter-decadal Pacific Oscillation |
| IRI | : International Research Institute for Climate and Society |
| ITCZ | : Inter Tropical Convergence Zone |
| ITD | : Inter Tropical Discontinuity |
| JMA | : Japan Meteorological Agency |
| KNN | : K-Nearest Neighbor |
| LPM | : Linear Perturbation Model |
| LRN | : Layer Recurrent Network |
| MAE | : Mean Absolute Error |
| METEO-BENIN | : Agence Nationale de la Météorologie |
| MLP | : Multi-Layer Perceptron |
| MLR | : Multiple Linear Regression |
| MSE | : Mean squared error |
| NNLPM | : Nearest Neighbor Linear Perturbation Model |
| NOAA | : National Oceanic and Atmospheric Administration |
| NSE | : Nash-Sutcliffe Efficiency |
| NWP | : Numerical Weather Prediction |
| POAMA | : Predictive Ocean Atmosphere Model for Australia |
| PSRSA | : Plan Stratégique de Relance du Secteur Agricole |
| $R^2$ | : Coefficient of determination |
| RBF | : Radial Basis Function |
| RBFN | : Radial Basis Function Network |

| | |
|---|---|
| RBFNN | : Radial Basis Function Neural Network |
| RGPH | : Recensement Général de la Population et de l'Habitat |
| RHmax | : Maximum relative humidity |
| RHmin | : Minimum relative humidity |
| RMSE | : Root Mean Squared Error |
| RR | : Rainfall |
| RT | : Regression Trees |
| SLM | : Simple Linear Model |
| SLR | : Simple Linear Regression |
| SOI | : Southern Oscillation Index |
| SOM | : Self Organizing Map |
| SPEI | : Standardized Precipitation and Evapotranspiration Index |
| SST | : Sea Surface Temperature |
| STTF | : Short-term temperature fluctuations |
| SVM | : Support Vector Machines |
| Tmax | : Maximum air temperature |
| Tmin | : Minimum air temperature |
| Uwind | : Zonal wind |
| VIF | : Variance Inflation Factor |
| Vwind | : Meridional wind |
| WMO | : World Meteorological Organization |

# Chapter 1

# INTRODUCTION

## 1.1 Background

Evidence of events regarding global warming as a result of Green House Gas (GHG) emissions in the atmosphere is very much disputed [1]. Recent studies have also shown that, just in the past century, concentration of atmospheric $CO_2$ had increased significantly by almost 100 ppm, which induced the rising of global average temperature of 0.74 °C compared to the pre-industrialization era [2]. Africa is one of the continents, most vulnerable to climate change and variability, and has worrisome low adaptation capacity [3].

The Republic of Benin is one of Africa's 54 countries, located in West Africa. Its climate is influenced by the Inter Tropical Convergence Zone (ITCZ), which oscillates annually between the northern and southern tropics. South of the ITCZ, the major wind direction is south-westerly, blowing moist air from the Atlantic onto the continent; whereas, north of the ITCZ, the dominating winds are north-easterly, bringing hot and dusty air from the Sahara desert known as the 'Harmattan'. These two completely opposing wind directions dictate the annual West African Monsoon, resulting in a wet season in the north of Benin from May to November, and two wet seasons from March to July and from September to November in the southern regions of Benin [4]. The country has been affected periodically in the past by climatic irregularities resulting in droughts, floods, thunderstorms with strong winds and rains. According to data from the Second National Communication, Benin has experienced more than 15 major disasters since 1984 with significant consequences and numerous losses. For example, drought in 1984; floods in 1985, 1994, 1996, 1997, 2006 and 2010; thunderstorms in 2005 [5]. And because the economic growth of the country is largely dependent on agriculture, fluctuations in rainfall amounts, declining soil fertility, limited access to high quality seeds, the high cost of agricultural inputs, and the low level of mechanization negatively affect the country's agricultural growth and food security [6]. Rainfall quantity is a key factor in agriculture methods of Benin as both less and excessive rainfall might reduce agricultural production. Furthermore, excessive rainfall causes natural disasters that endanger the lives of millions of people. However, Intergovernmental Panel on Climate Change (IPCC) in its fifth report, clearly states that understanding the evolution of precipitation is crucial for the implementation of adaptation strategies in order to manage and reduce the risks of precipitation variability on systems such as agriculture and forestry and to strengthen the level of resilience [7].

Rainfall predictions can be made over a short time, such as predicting an hour or a day into the future, or over a long time such as monthly or a year ahead. Monthly rainfall data offer a more reliable intra-year rainfall distribution than seasonal rainfall data. Monthly rainfall is an important contributing factor to agricultural and hydrological activities [8]. For instance, in Republic of Benin, majority of the population depends on rain fed agriculture. Jaurès et al. in [9] assess the importance of seasonal climate forecasts for maize farmers in the Republic of Benin. Their findings show that, farmers' most important need is accurate seasonal climate forecasts related to the onset, distribution and amount of rainfall between 1 and 2 months before the onset of rains, to improve their production and increase their incomes. Rainfall affect many other sectors including but not limited to water resources management, energy, tourism, health, disaster risk reduction and infrastructure development. Thus, accurate seasonal rainfall forecasting is fundamental for planning and management of many sectors [10].

In Benin Republic, the African Center of Meteorological Application for Development (ACMAD) is the primary source of climate forecasting tailored to the needs of farmers and other socio-economic actors. ACMAD's primary mission, in cooperation with African national meteorology services and AGRHYMET (Agriculture, Hydrology and Meteorology Research Center), is the production of climate information and the provision of short and medium term climate forecasting tools. These forecasts make use of a Climate Predictability Tool (CPT) to link monthly or seasonal rainfall or streamflow data to Sea Surface Temperature (SST). CPT uses multiple linear regressions to link a grid predictor to a predictand. ACMAD makes two types of forecasts available to users:

- Short-term forecast of precipitation (24 hours), updated daily
- Sub-seasonal and seasonal forecasts of precipitation, issued in form of a bulletin

The seasonal precipitation prediction bulletins are determined by consensus of a panel of experts who synthesize various forecasts generated by global climate centers' numerical models [11].

The Sub-seasonal and seasonal forecasts of precipitation delivered actually in Benin are qualitative and probabilistic predictions issued in a categorical format on three scales: the probability of above normal, normal, and below normal precipitation. The forecasting does not provide information about precipitation’s temporal distribution, which is essential for risk management.

However, rainfall prediction using the non-linear multivariate ANN approach has not yet been studied in Republic of Benin. Thus, this work is a first attempt at developing monthly rainfall model by using ANNs to forecast rainfall in Republic of Benin. Moreover, the current model outputs will be deterministic, allowing for quantitative forecasting, as opposed to a probabilistic prediction that just offers above- or below-normal value.

## 1.2 Problem statement

The problem of generating weather predictions is more complicated than that of generating planetary orbit predictions [12]. Nonlinearity is an inherent property of the physical climate system. The series of nonlinear differential equations, which regulate the flow of the atmosphere and oceans, is a well-known example. Rainfall as a meteorological parameter is a very complex nonlinear phenomenon that varies with time and location [8], [13]. The increasing economic cost of adverse weather events provides a strong reason to generate more weather forecasts that are accurate.

Weather forecasting (and rainfall prediction in particular) is one of the most important imperatives and demanding operational responsibilities carried out by meteorological services all over the world. Long-term rainfall forecasting is very important for countries whose economy depends mainly on agriculture, like many of the Third World countries [14], particularly the Republic of Benin. Its accurate prediction can improve the quality of decision making in the energy industry, and for efficient resource planning and management, including agriculture, famine and disease control, rainwater catchment and ground water management.

## 1.3 Research objectives

The aim of the research is to develop a rainfall prediction model based on artificial neural network to predict rainfall one-month ahead using historical data from different synoptic stations in Benin.

The Specific objectives are as follows:

- Determine the most influencing and important meteorological variables that affect rainfall prediction in Benin Republic;
- Use Artificial Neural Networks as an effective approach to construct long-term predictive models for rainfall forecasting from meteorological data in Benin Republic; and validate the results of the model using ground base data;
- Assess the influence of geographical regions on the performance of the model in predicting monthly rainfall in Benin Republic;
- Evaluate the percentage contribution of each meteorological variable to total rainfall.

## 1.4 Research questions

The main research question behind this study is how can we develop a more accurate rainfall prediction model by using Artificial Neural Networks approaches for different synoptic stations in Benin? Other important questions behind this research are:

- What are the meteorological variables that can be used to more effectively predict rainfall?
- What is the most effective neural network architecture for predicting rainfall?
- Does the geographic location affect the model's performance in predicting monthly rainfall?
- What are the contributions in percentage of each meteorological variable to total rainfall?

## 1.5 Significance

Rainfall is one among all weather parameters that mostly affects human life and livelihood around the world. The varying distribution of precipitation over time and space constitutes a serious challenge for practicing economic activities, food insecurity and household poverty reduction. Accurate rainfall prediction is a serious concern in many countries, especially in the Republic of Benin. In the last decade, Benin has been confronted with increasingly unpredictable rainfall and sometimes drastically alters the whole rainy season [15]. One of the key factors behind the economic growth of the country is agriculture. As at 2017, World

Bank [16] reported that agriculture contributes about 25% of the national gross domestic product (GDP) in the country. Agriculture, as it stands, is an important economic sector in Benin, employing over 70% of the population. Since Benin's agriculture directly depends on rainfall, it is crucial to anticipate it earlier and take necessary measures so that it does not spread any adverse effects. Flood or drought events cause significant damage to agriculture, the environment, infrastructure, social life and to GDP. Its prevention can only be done by anticipating it earlier and taking measures accordingly.

The accuracy and exactness of rainfall prediction will help to develop more appropriate strategies for agriculture and water reserves, and would provide information on flood as to implement precautionary measures.

## 1.6 Thesis structure

The thesis is structured into six chapters. The first chapter deals with a general introduction to this thesis, including the background of the study, the problem statement, the objectives of the study, the research questions, the significance of the research and the thesis structure. The literature review follows chapter one. The third chapter briefly presents the study area. The data and methodology are presented in chapter four. Results are presented and discussed in chapter five, while chapter six is the conclusion of the thesis with some perspectives.

# Chapter 2

# LITERATURE REVIEW

## 2.1 Introduction

Forecasting the weather is a process of using the science and technology to make predictions about the state of the atmosphere at a specific location in the future [17]. Soft computational methods and machine learning approaches have seen wide use in weather forecasting. They have been used in predicting rainfall, predicting temperature, and modeling rainfall runoff. The research confirmed that they outperform traditional methods [18]. Numerous algorithms have been developed to simulate and predict rainfall. It includes numerical methods [19], statistical methods [12] and machine learning techniques methods [20].

Rainfall forecasting using numerical weather prediction (NWP) models is far from satisfactory due to inaccurate initial conditions, limited spatial resolution, and parameterization schemes of subscale phenomena [21]. It involves the study of the rainfall processes in order to model the underlined physical law. However, the application of it can be difficult, because rainfall is influenced by a number of meteorological parameters and is limited in both the spatial and temporal dimensions [22]. On the other hand, statistical prediction methods are only suitable for linear applications. These are considered by means of the main disadvantages of using statistical methods when it comes to non-linear time series data [20], [23], since rainfall is known to contain nonlinear as well as stationary data. These issues prompt an examination of the potential of data-driven techniques. Since the emergence of artificial intelligence-based models such as Artificial Neural Networks, a plethora of novel techniques for hydrological modeling have become available.

The Artificial Neural Network (ANN) is a soft computing method that has the capability to identify nonlinear data pattern with learning approach [24]. Not only this method is quite simple and practical to use for prediction, it also has good accuracy. Since the mechanisms of rainfall are still not well understood, ANNs are a good choice worth trying to analyze the relationship between other meteorological parameters and rainfall.

This chapter reviews the existing literature on weather forecasting, focusing on rainfall prediction using artificial neural networks (ANNs) (machine learning methods) and multiple linear regression (statistical methods).

## 2.2 Applications of ANNs in rainfall prediction

Artificial neural network (ANN) models are the most popular and widely used models for rainfall prediction because physical processes affecting rainfall occurrence are highly complex and non-linear [25]. ANNs have the ability to distinguish increasingly dynamic non-linear interactions between input and output variables without requiring an understanding of the underlying physical processes [26]. Some of the most recent ANN applications for rainfall prediction includes:

Ayodele et al. in [27] used artificial neural networks (ANNs) and the Back Propagation (BP) algorithm to forecast seasonal rainfall in Ikeja, Nigeria, using as input variables sea surface temperature (SST), U-wind at (surface, 700, 850, and 1000 hPa), air temperature, specific humidity, ITD, and relative humidity from year 1986 to 2017. Results of the proposed ANN network show an accurate seasonal rainfall forecast for Ikeja with a limited error. Similarly, Abdulkadir et al. in [28] assessed the efficacy of Neural Networks in predicting rainfall patterns in 7 chosen stations in Nigeria using successive rainfall depth data as input. After the model validation, significant correlation coefficients were observed, indicating that ANN may be employed for quantitative rainfall prediction in these locations. In the paper [29], the authors used ANNs to forecast rainfall across 23 sites in Nigeria using 30-year data. The predictability of rainfall was shown to be more accurate at higher latitudes across the country. The paper [30] also developed an ANN based model for monthly rainfall forecasting in Zaria, Nigeria, using temperature, relative humidity, wind speed and sunshine hours as input variables. The authors obtained 81% as correlation value during the testing phase.

Kashiwao et al. in [31] compared multiple linear regression (MLP) to an algorithm composed of random optimization, back propagation, and Radial Basis Function Network (RBFN) for the purpose of forecasting short-term rainfall using data from the Japan Meteorological Agency (JMA). The authors demonstrated that MLP outperformed RBFN. Artificial neural networks (ANNs) have been successfully used in conjunction with the Back Propagation (BP) algorithm to forecast rainfall in Indonesia [32]. Vamsidhar et al. in [33] have also used the BP Neural Network model to predict monthly rainfall in India, with humidity, dew point, and pressure serving as model inputs. The authors achieved 99.79% accuracy during the training phase and 94.28% during the testing phase. They inferred from these findings that future rainfall can be forecast using the same process. The monsoon precipitation has been predicted for the Udupi district of Karnataka using Feed Forward Networks and three related learning algorithms: Back Propagation Algorithm (BPA), Layer Recurrent Network (LRN), and Cascaded Back Propagation (CBP) [34]. In comparison to the other algorithms, the BPA had a smaller mean squared error (MSE). They concluded that their study's findings were generally favorable and that a high degree of precision was achieved. A study investigated the output of the multi-layer perceptron neural network (MLP-NN), the radial base function neural network (RBFNN), and the input delay neural network (IDNN) in predicting weekly and monthly rainfall in Malaysia using temperature and humidity as input variables [35]. Their findings showed that the IDNN model was more accurate when forecasting rainfall.

Using a back propagation neural network model, Geetha et al. [36] predicted rainfall in Chennai, India. The mean monthly rainfall is estimated by their study, and the findings show that the model performed well in both training and independent periods. Aksoy et al. in [37] applied a feed-forward back propagation neural network, a radial basis function neural network, a generalized regression neural network and the MLR at three petrological stations to forecast precipitation in Jordan's arid and semi-arid areas. As input variables, antecedent precipitation and the periodic component were also used. In comparison to the other three models, the feed-forward back propagation neural network was found to be best for monthly rainfall prediction. Lekkas et al. in [38] developed a multilayer back propagation network and discovered that BPN does not always find the right weight and biases for the best solution, while their results confirmed the hypothesis that ANNs would provide qualitative forecasts.

In order to forecast the annual average Indian summer monsoon rainfall, Chattopadhyay et al. in [39] utilized ANN with three separate back propagation learning principles and the asymptotic regression model. Momentum learning, conjugate gradient descent learning, and Levenberg Marquardt learning were the three separate back propagation learning laws. According to the findings, ANN with conjugate gradient descent learning and Levenberg Marquardt learning outperformed the other ones. Patil et al. in [40] investigated the suitability of various ANN models, including radial basis functions and multilayer perceptrons with Levenberg Marquardt and momentum learning laws, for forecasting rainfall using local parameters. Based on feed forward ANNs trained using the Levenberg Back Propagation

Algorithm, Naik et al. in [41] proposed a new approach for weather forecasting. The findings indicated that FFNN is suitable for weather forecasting.

Deo et al. in [42] used a total of 18 predictor variables to apply ANN to the monthly Standardized Precipitation and Evapotranspiration Index (SPEI) prediction in eastern Australia. Monthly rainfall totals, mean temperature, minimum temperature, maximum temperature, evapotranspiration, climate indices (Southern Oscillation Index, Pacific Decadal Oscillation, Southern Annular Mode and Indian Ocean Dipole) and the Sea Surface Temperatures (Nino 3.0, 3.4 and 4.0) were used as predictor variables in the development of ANN. The study discovered that the ANN model is an effective data-driven method for forecasting monthly SPEI.

In Queensland, Australia, Abbot et al. in [43] used ANN and Predictive Ocean Atmosphere Model for Australia (POAMA) to forecast rainfall. Rainfall, maximum and minimum temperatures, the Southern Oscillation Index (SOI), the Inter-decadal Pacific Oscillation (IPO), the Dipole Mode Index (DMI), and El Nino 3.4 were used as input variables. In comparison to the POAMA model, the analysis discovered that the ANN model makes predictions that are more accurate.

Three distinct models for monthly rainfall forecasting were developed by [44]. From 1970 to 2008, the study collected monthly data on precipitation, mean temperature, wind speed, and relative humidity. The findings demonstrated the superiority of ANN models over other models. Maya et al. in [45] forecasted Southwest Indian Monsoon rainfall with ANN, using sea surface temperature, sea level pressure, humidity and zonal (u) and meridional (v) winds. The findings demonstrate a fair degree of precision. Using artificial intelligence methods to predict provincial rainfall, Kumar et al. [46] have discovered that this methodology performs well for monthly and seasonal forecasting. In [21], Ramirez et al. investigated the ANN for daily rainfall prediction in Sao Paulo, Brazil using potential temperature, wind, humidity, air temperature, precipitable water, relative vorticity and moisture divergence flux as input variables. The prediction performance of ANN was compared with the MLR model. The study suggested that ANN is more suitable for rain prediction in Sao Paulo, Brazil.

In [47], the authors investigated rainfall prediction using ANN and ARIMA models with only historical monthly rainfall data as input. The models' accuracy was compared and it was determined that there was no significant difference in either model's ability to forecast monthly rainfall. Karamouz et al. in [48] constructed and applied a time delay recurrent neural network to forecast long-lead seasonal rainfall in three case studies in Iran. The researchers compared the prediction accuracy of the time delay recurrent neural network to that of the autoregressive moving average with exogenous inputs (ARMAX) model and discovered that the time delay recurrent neural network outperformed the ARMAX model in all three scenarios. In [49], Somvanshi et al. demonstrated that the ANN model, which outperforms the ARIMA (Autoregressive Integrated Moving Average) model, could be used as a suitable forecasting method to estimate rainfall.

Using ANN, Joshi et al. [50] modeled rainfall-runoff. They compared three neural network methods, Feed Forward Back Propagation (FFBP), Radial Basis Function (RBF), and Generalized Regression Neural Network (GRNN), and discovered that the GRNN performed comparably to the FFBP, RBF, and MLR for flow estimation. In a Case Study on Jarahi Watershed, Solaimani [51] investigated Rainfall-runoff Prediction based on Artificial Neural Network in a semiarid area of Iran and discovered that the Artificial Neural Network approach is more suitable and effective to forecast river runoff than the classical regression model. In [52], Chang et al. used radial basis function neural network (RBFNN) to create a rainfall-runoff model for predicting floods three hours in advance. They trained, tested, and validated the network using a dataset of the Lanyoung River obtained during typhoons. Following the analysis, they concluded that the RBFNN methodology is suitable for forecasting flood flow.

In the paper [53], the authors used ANNs to forecast short-term temperature fluctuations (STTF) and discovered that the MLP network has the lowest forecasting error and can be used to model STTF systems. Maqsood et al. in [54] used a soft computing model based on an RBFN to predict weather for southern Saskatchewan, Canada, 24 hours in advance. Hourly weather data for temperature, wind speed, and relative humidity are used to train and test the model. The RBFN's efficiency is compared to that of the multi-layered perceptron (MLP) network, Elman recurrent neural network (ERNN), and Hopfield model (HFM) to assess their suitability for weather prediction. The models' reliabilities are then analyzed using a variety of statistical tests. The findings show that, as opposed to the MLP, ERNN, and HFM, the RBFN provides the most reliable predictions. In a comparative analysis, Maqsood et al. [55] demonstrated that the Hopfield Model (HFM) is comparatively less accurate than the RBFN for weather forecasting problems in southern Saskatchewan, Canada, and that, in addition, neural network ensembles provided the most accurate prediction. Tang et al. in [56] used numerous artificial neural network (ANN) models for prediction and interpretation of meteorology and oceanography data, and they considered the ANN technique to be very useful. In [57], Luk et al. developed and compared three types of artificial neural networks (ANNs) useful for rainfall prediction: multilayer feed forward neural networks, Elman partial recurrent neural networks, and time delay neural networks.

Additionally, many other machine learning algorithms have been used to forecast various weather attributes, including Support Vector Machines (SVMs), K-Nearest Neighbors (KNNs), Adaptive Network-Based Inference Systems (ANFISs), and Regression Trees (RTs) [58], [59].

Few experiments have been carried out on the use of the SVMReg rainfall prediction model. This model was combined with a multi-objective genetic algorithm in [60], [61] to forecast hourly rainfall in Taiwan with a lead time of 1 to 6 hours. As input, the meteorological variables air pressure, air temperature, wind velocity, wind direction and sunshine duration were used. The authors of the paper [62] used the SVMreg to forecast severe rainfall events in India with a lead time of six to forty-eight hours. The SVMReg models were used to forecast warm season rainfall based on thermodynamics in [63]. In [64], the discrete wavelet transform and SVMreg methods were combined to forecast one-day-ahead precipitation in Turkey. The authors of [65] compared the efficiency of the SVMReg model and back propagation neural networks for predicting typhoon rainfall in Taiwan and discovered that the SVMReg model performed significantly better than the back propagation neural networks.

To forecast daily flow in Iran, Eskandarin et al. [66] used k-NN and compared its prediction efficiency to that of an ANN model. The paper [67] evaluated the prediction accuracy of ARMA, ARIMA, artificial neural networks, and nearest neighbor approaches for rainfall and discharge forecasting. In [68], Toth et al. compared the forecast efficiency of k-NN, ANN, and autoregressive moving average models for short-term rainfall prediction in Italy with leading times ranging from 1 to 6 hours. They discovered that using a time series analysis methodology based on ANN significantly improves the accuracy of flood forecasting as compared to using simplistic rainfall prediction approaches. Shamseldin et al. in [69] used the nearest neighbor approach as the nearest neighbor linear perturbation model (NNLPM) to forecast river flow and compared the effects to those from the simple linear model and the linear perturbation model. In [70], the author used the k-NN model to forecast regular mean discharge in a mountain basin in northeastern Italy and compared the accuracy of the estimate to that of the Autoregressive Exogenous model.

The paper [71] proposed a hybrid ANFIS model for monthly rainfall prediction in a Malaysian catchment area. The paper [72] also suggested an ANFIS model for estimating spring precipitation across Australia. In [73], the authors used ANFIS to developed daily rainfall forecasting models for the Indian city of Udaipur. Several machine learning models were compared for forecasting monthly rainfall in Australia, including SVMS and RT [59].

Additional applications using machine learning algorithms in the field of meteorology can be found in [74]–[76].

## 2.3 Applications of MLR in rainfall prediction

Multiple linear regression (MLR) is the most commonly used linear statistical method for examining and modelling the relationship between a dependent variable and a set of predictor variables. Regression applications exist in most fields, including hydrology. MLR models are frequently used for rainfall and flood forecasting. They are also used for comparison with other models.

Bello et al. in [77] made attempt of predicting rainfall over Kano, in Nigeria using ANN and Linear Model for three month lead period. Comparing the results, ANN gave better accuracy. In [78], Sethi et al. predicted early rainfall using the MLR methodology. The model is based on 30 years of climate data from Udaipur City, Rajasthan, India, including rainfall, vapor pressure, average temperature, and cloud cover from 1973 to 2002. The model predicts the amount of rain that will come each month in July (in mm). The experimental findings revealed that the expected and real rainfall amounts are very similar. In the paper [79], the authors investigated the application of ANN and MLR models for long-term seasonal spring rainfall prediction in Victoria, Australia. They select lagged El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) as predictors of rainfall. Both models were assessed and ANN model was found to have lower prediction errors compared to the MLR model. The ANN model also shows higher correlation compared to MLR model. Similarly, in the paper [80], the authors applied the ANN and MLR models for rainfall-runoff prediction in Japan. The results showed that feed forward ANN could describe rainfall-runoff behavior more accurately than the classical regression model. Also in [81], the authors predicted Indian Summer Monsoon rainfall with the ANN and MLR models using Nino indices as predictors. The comparative analysis suggested that ANN model performance is better than the MLR model. Using multiple regression models, Rossel et al. [82] forecasted monthly rainfall in Ecuador, with precipitation, sea surface temperature (SST), and meridional and zonal wind as predictors. In the out-of-sample test set, the created models were utilized to forecast rainfall anomalies. They observed that the wet months of the year had substantial predictive potential, with the strongest prediction occurring from March to May. They discovered that the generated multiple linear models explained 60–82% of the variation in monthly precipitation. Using ANN, Aksoy et al. [37] predicted precipitation in Jordan and compared the prediction performance with the MLR model. Similarly, [39] compared the ANN and the MLR model performance in predicting the annual average Indian summer monsoon rainfall. In the paper [83], the authors forecasted Indian summer monsoon rainfall using MLR and Indian Ocean SST anomalies. They found that the created model was capable of accurately forecasting rainfall. In [84], Ihara et al. investigated the relationship between ENSO and Indian Ocean indices and Indian summer monsoon rainfall using MLR. They observed that when combined with zonal wind anomalies across the equatorial Indian Ocean, Nino3 and zonal wind anomalies are excellent predictors of rainfall in the region. Chattopadhyay in [85] developed a FFNN to estimate average rainfall during summer-monsoon season in India. The proposed network was compared to a MLR model where better accuracy was obtained with proposed FFNN. Using multiple linear regression and nonparametric local polynomial approaches, Singhratina et al. [86] developed a predictive forecasting system for SMR over Thailand. As predictors, SST, sea level pressure (SLP), wind direction, ENSO, and IOD were chosen. According to the experiments, the ratio between measured and forecasted rainfall was 0.6. Lee et al. in [87] have compared rainfall predictions made with neural networks and linear regression model. Results showed that the artificial neural networks produced good predictions while the linear models produced poor predictions. In [88], the authors compared the effectiveness of ANN-based rainfall-runoff modeling to that of the Simple Linear Model (SLM), the seasonally-based Linear Perturbation Model (LPM), and the Nearest Neighbor Linear Perturbation Model (NNLPM), concluding that ANNs could produce more reliable discharge predictions than any of the more established models.

## 2.4 Concluding remarks

These studies demonstrated that ANNs are an effective method for forecasting rainfall amounts on a variety of time scales in advance. It also observed that most researchers have used back propagation technique and found significant results. Although much work has been done on rainfall prediction using Artificial Neural Network (ANN), particularly Multi-Layer Perceptron (MLP) in different countries, so far, no study has established a rainfall-forecasting model for Benin Republic using the non-linear multivariate ANN approach.

The importance of accurate prediction is highlighted by high rainfall variability, which is one of the major challenges that meteorologists face. However, rainfall prediction is a difficult process that requires continual improvement. It is critical to conduct research in this area. Several rainfall prediction methods have been tried; some perform better than others. There is always a need for improvement in the existing models and the development of new methods for reliable rainfall prediction.

# Chapter 3

# STUDY AREA

## 3.1 Study area presentation

The Republic of Benin is located in West Africa between latitudes 6°30’ and 12°30’ north and longitudes 1° and 3°40’ east. It has a land area of 114,763 km$^2$ and a population estimated at 10,008.764 inhabitants (RGPH, 2013). The study covers all the synoptic stations of Benin (Figure **3.1**). These weather stations were chosen as they serve as the primary station for the entire country. The geographical positions of which are presented in Table **3.1**. The climate zones, according to Hounkpèvi et al. [89], were also considered when choosing these weather stations because the Benin Republic experiences a variety of climates due to its size.

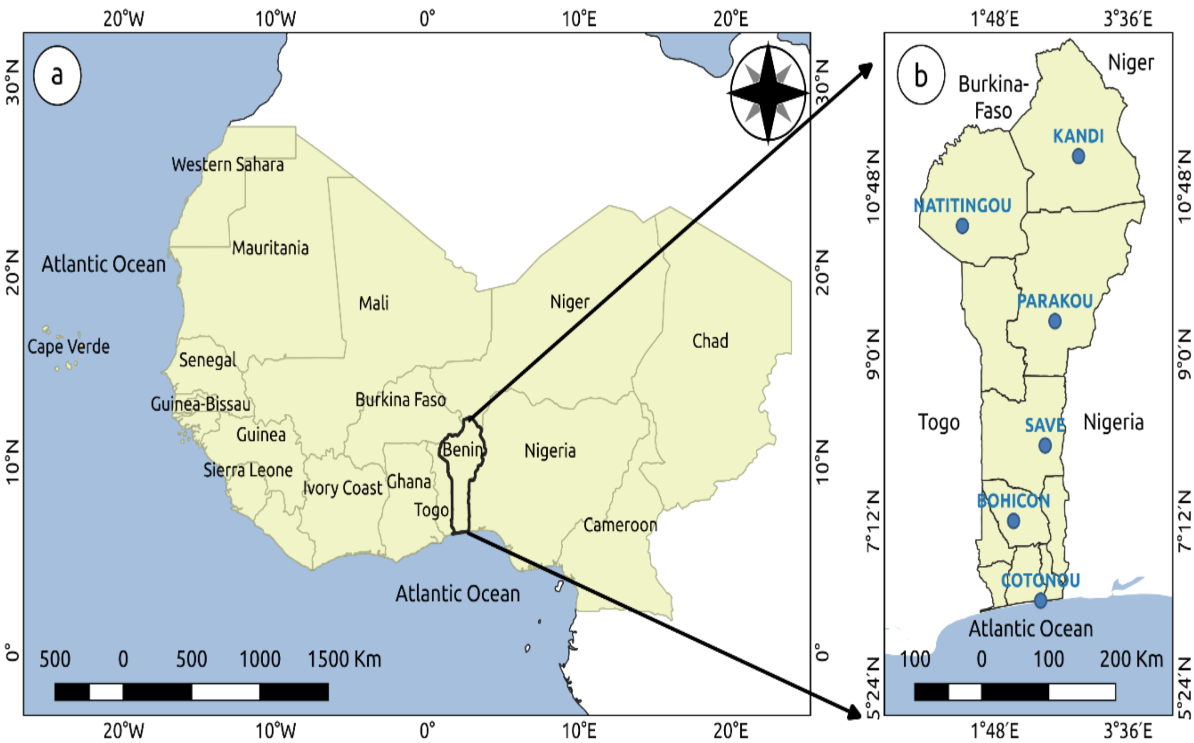



Figure 3.1: Study sites location: (a) Benin’s location in West Africa, (b) Synoptic stations’ location in Benin.

Table 3.1: Geographical Coordinates of Synoptic Stations.

| Climatic zone | Station name | WMO code | Latitude (°N) | Longitude (°E) | Elevation (m) |
|---|---|---|---|---|---|
| Guinean zone | Cotonou | 65344 | 6.35 | 2.38 | 5 |
| | Bohicon | 65338 | 7.17 | 2.07 | 166 |
| Sudano-Guinean zone | Save | 65335 | 8.03 | 2.47 | 199 |
| | Parakou | 65330 | 9.35 | 2.60 | 392 |
| Sudanian zone | Natitingou | 65319 | 10.38 | 1.36 | 460 |
| | Kandi | 65306 | 11.13 | 2.93 | 290 |

### 3.1.1 Climate

Benin’s climatic profile shows two contrasting climatic zones (Guinean vs. Sudanian) and a transitional zone (Sudano-Guinean). The Guinean zone, located between 6°25’ and 7°30’ N, has a subequatorial climate with four seasons (two rainy and two dry). The annual rainfall of approximately 1200 mm is bimodal, occurring mostly between March and July and September and November. The temperature ranges between 25 and 29℃, and the relative humidity ranges between 69% and 97%. The Sudano-Guinean, located between 7°30’ and 9°45’ N, is a transitional zone with two rainy seasons merging into a unimodal regime. Annual rainfall varies between 900 and 1110 mm, the temperature ranges between 25 and 29℃, and the relative humidity ranges from 31% to 98%. The Sudanian region, which lies between 9°45' and 12°25' N, has a tropical dry climate of two seasons of similar duration (rainy and dry). Annual mean rainfall in this region is often less than 1000 mm and occurs mostly between May and September. The temperature varies between 24 and 31℃, and the relative humidity varies between 18% and 99% [89]–[92]. Figure **3.2** illustrates both seasonal patterns using the rainfall data over a period from 1979 to 2018 from the climate zone.

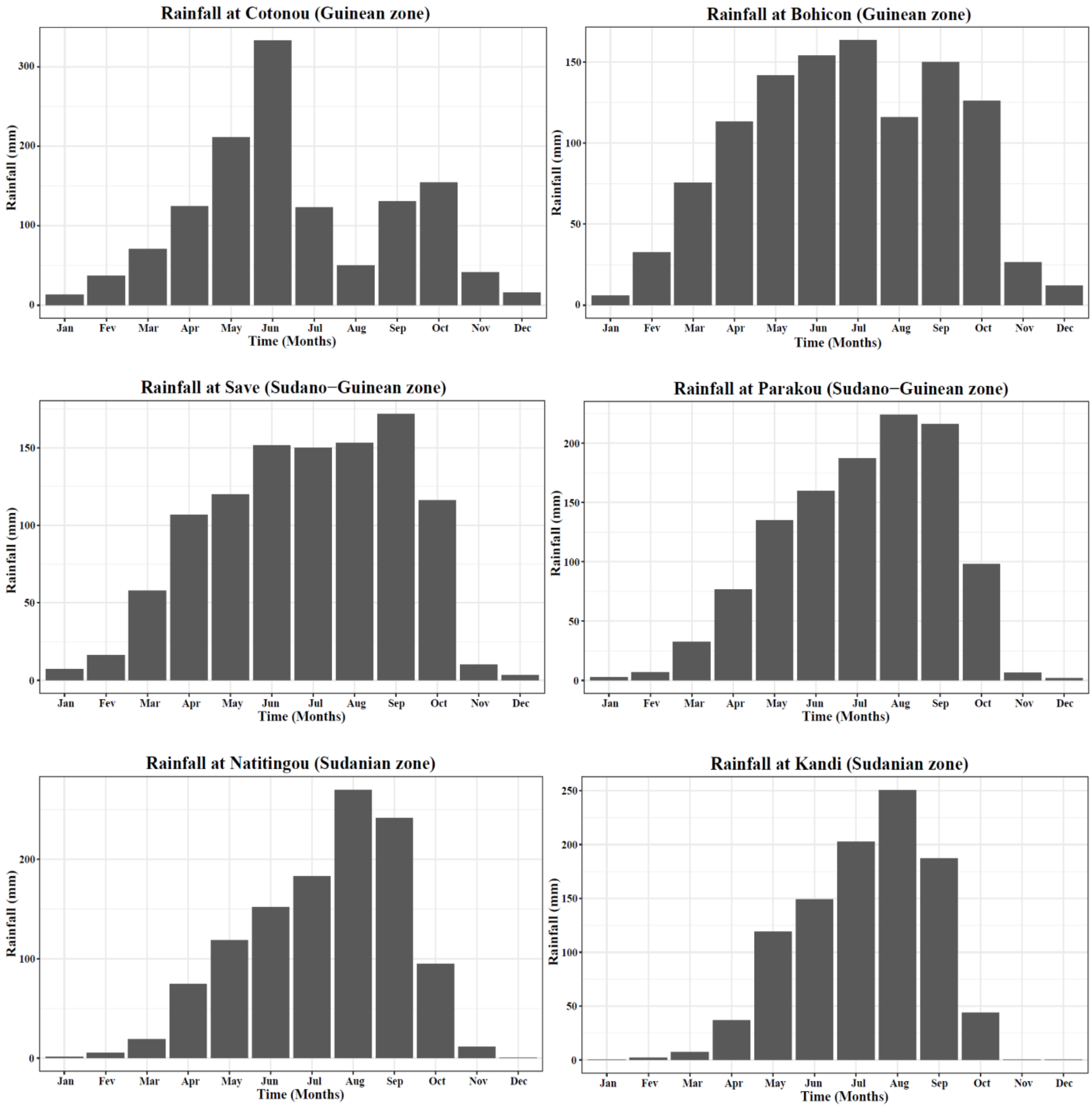



Figure 3.2: Rainfall patterns in different climate zones in Benin Republic.

### 3.1.2 Agriculture

Over the past two decades, the agricultural sector has kept a significant part of the nation's economy steady because of the output's ongoing contribution to GDP. This contribution to GDP, dominated by small farmers, rose from 33.1% in 1995 to 34.9% in 2000 before steadily declining to 32.6% in 2012. Apart from small farmers' domination, new agricultural firms are gaining ground. Despite the agricultural sector's considerable riches, it is confronted with a number of restrictions, the most serious of which are soil depletion and unregulated water usage. Additionally, extreme climate instability is a major limitation on agricultural production. Cotton, cashew nuts, and palm oil are the primary cash crops, while corn, maize, and sorghum are primary food crops. Cotton continues to be the primary cash crop, accounting for 600,000 tons in the 2016-2017 campaign. Along with cotton, cashew nuts and pineapple processed approximately 280,516 and 307,401 tons, respectively, between 2016 and 2017. It's worth noting that oil palm, which was previously one of Benin's top export crops, dropped to 25,971 tons in 2017 [93].

Numerous attempts have been made in the livestock sector during the five years since the Strategic Agricultural Recovery Plan was implemented (PSRSA, 2011-2015). Table egg production rose by approximately 63% from 9,072 tons in 2008 to 14,746 tons in 2015. Between 2008 and 2015, milk and meat production increased at a rate of about 20%, reaching 107,362 and 65,019 tons, respectively, in 2015 [93].

In terms of halieutic demand, the annual average of fish (fishing and aquaculture) is currently estimated at 39,500 tons, while the annual average of halieutic products is 113,000 tons, resulting in a shortfall of more than 73,000 tons offset by frozen fish exports [93].

### 3.1.3 Forest

Benin's woodland cover is currently projected to be 4,625,000 ha, or about 42% of the country's land area. The state forest is split into two sections: protected and classified. Two (2) national parks (Pendjari and W), four (4) fauna reserves (420,000 ha), and 58 listed forests and reforestation perimeters comprise the classified territory (1,436,500 ha). The forestry sector contributes approximately 2.8 percent of GDP. Benin has an extremely rich ecosystem, but is threatened by a variety of factors, including anthropogenic deforestation, bushfires, water degradation, trafficking, exotic species introduction, overgrazing, and overexploitation of terrestrial and marine flora and fauna associated with growth activities. Concerning goods derived from forest resources designated for sale, it is worth noting that the amount of timber exploited between 2001 and 2011 ranges basically between 1,944.8 and 67,272.00 $m^3$, with a peak in 2006 (2,590.7 $m^3$) [93].

### 3.1.4 Water resources

Water is the most commonly utilized substance. Water, a scarce resource, is vital to life. Water is essential for life on Earth. It impacts the environment, economy, and social development. Water is required for domestic consumption, agriculture, industry, irrigation, livestock watering, and electricity. In simply, no water means no life.

Water is a natural resource that is constantly replenished. It is constantly exchanged between the ocean, atmosphere, and land (hydrologic cycle). The hydrologic cycle is initiated when water from the oceans and rivers evaporates. Condensation of the elevated water vapor results in the formation of clouds, which ultimately fall as precipitation to the ground. Following precipitation, some of the water evaporates back into the atmosphere, while others may seep through the surface and form ground water. Ground water either percolates into seas, rivers, and streams or evaporates back into the atmosphere. Surface water and groundwater are two distinct types of water resources. Surface water is found in streams, rivers, lakes, and wetlands, whereas ground water is found underground, in geological formations called aquifers.

Benin's water resources are comprised of both surface and ground water. The surface water resources are dispersed among six (6) watersheds, which are divided into four primary hydrographic units, including the Niger River, Ouémé-Yeoua, Volta, and Mono-Couffo hydrographic groups. Surface water reserves are expected to be 13.106 billion $m^3$ per year. Discontinuous aquifers in the rocky region and continuous aquifers cover around 80% and 20% of Benin's total land area, respectively. Both aquifers are recharged annually with a total of 1.87 billion $m^3$ of water [5], [93].

# Chapter 4

# DATA AND METHODOLOGIES

## 4.1 Rainfall and Atmospheric Data

### 4.1.1 Rainfall data

The predictand in this research is the monthly rainfall calculated using daily rainfall data that was collected from 6 meteorological stations in Benin for 40 years within the period from 1979 to 2018 .The rainfall time series has been taken from the National Meteorological Agency of Benin. The average monthly rainfall varies across these sites from 83.22 mm to 108.82 mm. The statistical summary is given in Table **4.1**.

Table 4.1: Minimum, maximum, average and standard deviation rainfall (mm) of weather stations.

| Station name | Min (mm) | Max (mm) | Mean (mm) | StdDev (mm) |
|---|---|---|---|---|
| Cotonou | 0 | 674.3 | 108.82 | 116.48 |
| Bohicon | 0 | 520 | 93.08 | 79.85 |
| Save | 0 | 402.8 | 88.67 | 80.76 |
| Parakou | 0 | 446.9 | 95.62 | 97.93 |
| Natitingou | 0 | 447.6 | 97.62 | 102.66 |
| Kandi | 0 | 416.9 | 83.22 | 97.80 |

### 4.1.2 Atmospheric data

Monthly atmospheric data were used as predictors of the ANN-based prediction model in this research. The National Meteorological Agency of Benin provided the maximum relative humidity (RHmax), minimum relative humidity (RHmin), maximum air temperature (Tmax), and minimum air temperature (Tmin) data used. ERA5 reanalysis data on Evaporation (Eva), zonal wind (Uwind) and meridional wind (Vwind) were obtained from the Climate Data Store (CDS) (https://cds.climate.copernicus.eu) [94]. The Sea Surface Temperature (SST) data series used is the Extended Reconstructed Sea Surface Temperature (ERSST) version3b from the National Oceanic and Atmospheric Administration (NOAA)/National Climatic Data Center (NCDC) obtained from the International Research Institute for Climate and Society (IRI) data library (https://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCDC/.ERSST/.version3b/) [95]. Table **4.2** presents the description of atmospheric data.

Table 4.2: Description of atmospheric data.

| Parameters | Unit | Level | Data reference | Temporal coverage |
|---|---|---|---|---|
| Maximum relative humidity (RHmax) | % | Surface (1.5 m) | METEO-BENIN | 1979/01 to 2018/12 |
| Minimum relative humidity (RHmin) | % | Surface (1.5 m) | METEO-BENIN | 1979/01 to 2018/12 |
| Maximum air temperature (Tmax) | ℃ | Surface (1.5 m) | METEO-BENIN | 1979/01 to 2018/12 |
| Minimum air temperature (Tmin) | ℃ | Surface (1.5 m) | METEO-BENIN | 1979/01 to 2018/12 |
| Evaporation (Eva) 0.25° x 0.25° grid | mm | surface | ERA5 | 1979/01 to 2018/12 |
| Zonal wind (Uwind) 0.25° x 0.25° grid | m/s | 850 hPa | ERA5 | 1979/01 to 2018/12 |
| Meridional wind (Vwind) 0.25° x 0.25° grid | m/s | 850 hPa | ERA5 | 1979/01 to 2018/12 |
| Sea Surface Temperature (SST) 2° x 2° grid | ℃ | Surface | NOAA NCDC ERSST version3b | 1979/01 to 2018/12 |

## 4.2 Data pre-processing

The downloaded data from CDS and IRI cover a globe and are in a Network common data format (NetCDF) files. The data were extracted using the Python programming language by selecting the data on the research area only. Afterward, the selected data were reprocessed using the software MATLAB R2020b and MS Excel.

### 4.2.1 Correlation of rainfall with SSTs

In order to identify basins that have an influence on the rainfall patterns, the correlation between station rainfall and the SSTs for respective stations has been computed (Figure **4.1**). After this, all the best-correlated basin coordinates are noted and recorded (Table **4.3**).

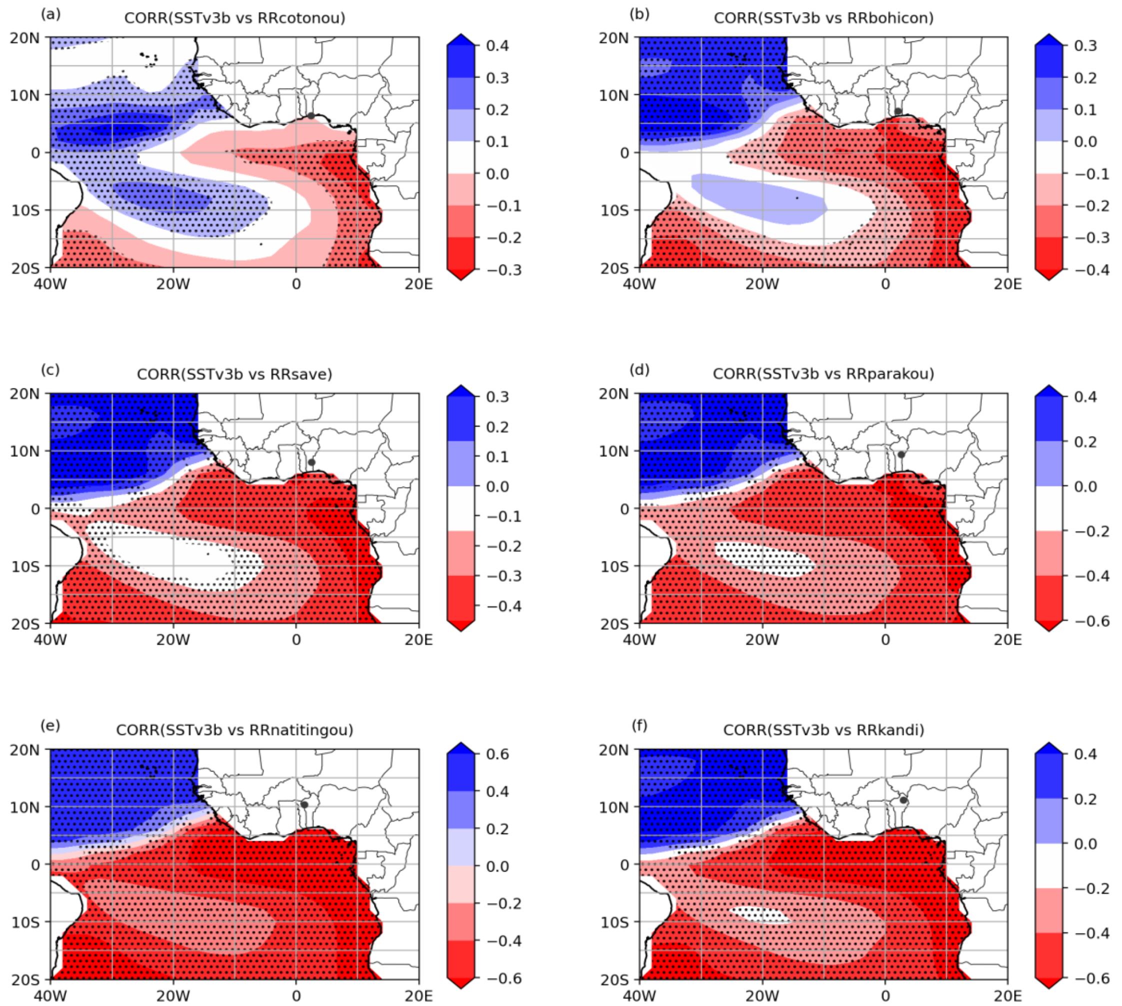



Figure 4.1: Correlation map between station Rainfall and grid point SST (The dotted areas show the areas where the correlations between SST and RF are significant at the 5% level (i.e. $p < 0.05$)).

Table 4.3: Best correlated basins.

| Stations | Basins |
|---|---|
| Cotonou | 4N-6N / 32W-22W |
| Bohicon | 2N-2S / 0E-10E |
| Save | 4N-4S / 10W-10E |
| Parakou | 4N-4S / 10W-10E |
| Natitingou | 4N-4S / 10W-10E |
| Kandi | 4N-4S / 10W-10E |

### 4.2.2 Selection of optimal lag time for each predictor

Meteorological variables are potentially useful predictors of precipitation. To select the predictor variables and identify the months that could be used as input to the ANNs, cross-correlation analyses were carries out for delayed atmospheric data. The correlation coefficients were calculated between each monthly atmospheric data with a lag time from one to eleven months and monthly rainfall. Many previous studies have validated that lagged climate variables are good predictors [96]–[100]. Table **4.4** provides a list of inputs variables, values of the correlation coefficients and times lags (in red color) that were used for the forecasting model. For different climate variables, different months had a significant correlation with rainfall.

Table 4.4: Selected Meteorological variable, lagged months (in red color) and Correlation Coefficients (at 1% significant level) of monthly rainfall.

| Stations | RHmax | RHmin | Tmax | Tmin | Eva | SST | Uwind | Vwind |
|---|---|---|---|---|---|---|---|---|
| Cotonou | -0.443(2) | -0.547(5) | -0.530(10) | -0.431(9) | 0.498(9) | 0.156(11) | -0.519(6) | 0.541(1) |
| Bohicon | 0.396(11) | 0.612(11) | 0.619(5) | 0.458(4) | -0.582(11) | 0.651(5) | 0.504(1) | 0.631(1) |
| Save | 0.509(11) | 0.667(11) | 0.669(5) | -0.520(9) | -0.673(11) | 0.710(5) | 0.646(1) | 0.668(1) |
| Parakou | -0.641(7) | -0.768(6) | 0.751(5) | -0.609(8) | 0.739(6) | 0.783(5) | 0.767(1) | -0.761(7) |
| Natitingou | -0.750(6) | -0.781(6) | 0.795(5) | -0.735(8) | -0.781(11) | 0.820(5) | 0.811(1) | -0.785(6) |
| Kandi | -0.846(6) | -0.793(6) | 0.715(4) | -0.749(7) | -0.840(11) | 0.795(4) | 0.814(1) | 0.727(1) |

### 4.2.3 Feature Selection

The identification of the most significant input variables is important steps in building an optimal ANN model. Feature selection (FS) is a process that selects a subset of original features. Feature selectors are algorithms that are applied to the data prior to it being sent to a machine learning program. FS objective is to reduce the data's dimensionality by eliminating irrelevant and redundant information, allowing the machine learning program to operate more effectively.

In this study to find out the most influencing and important features that affect the monthly rainfall prediction out of the existing ones, the Correlation-based Feature Selection (CFS) [101], [102] algorithm has been implemented with appropriate different search methods such as best first and greedy stepwise [103].

Correlation-based feature selection (CFS) subset evaluator is a filter algorithm that determines the value of a subset of attributes by taking into account each feature's individual predictive potential and the degree of redundancy amongst them. CFS is based on the following hypothesis: "*Good feature subsets contain features highly correlated with the class yet uncorrelated with each other* [101]." A generalized filter algorithm is illustrated below (Figure **4.2**).

**Filter Algorithm**

**input:** $D(F_0, F_1, ..., F_{n-1})$ // a training data set with $N$ features
$S_0$ // a subset from which to start the search
$\delta$ // a stopping criterion

**output:** $S_{best}$ // an optimal subset

**begin**
  **initialize:** $S_{best} = S_0$;
  $\gamma_{best} = eval(S_0, D, M)$; // evaluate $S_0$ by an independent measure $M$
  **do begin**
    $S = generate(D)$; // generate a subset for evaluation
    $\gamma = eval(S, D, M)$; // evaluate the current subset $S$ by $M$
    if ($\gamma$ is better than $\gamma_{best}$)
      $\gamma_{best} = \gamma$;
      $S_{best} = S$;
  **end until** ($\delta$ is reached);
  **return** $S_{best}$;
**end**;

Figure 4.2: Generalized filter algorithm [104].

The algorithm begins its search for a given data set D with a given subset $S_0$ (an empty set, a complete set, or any randomly chosen subset) and proceeds through the feature space using a specific search strategy. Each generated subset S is evaluated and compared to the previous best one using an independent measure M. If it is determined to be superior, it becomes the new best subset. The search iterates until a predefined stopping criterion δ is reached. As a result, the algorithm returns the most recent latest best subset $S_{best}$. Because the filter model uses independent evaluation parameters and is not based on any mining algorithm, it is free of mining algorithm bias and is therefore computationally efficient. CFS is formalized in Eq. (**4.1**) [101].

$$M = \frac{k\bar{r}_{fc}}{\sqrt{k + k(k-1)\bar{r}_{ff}}} \tag{4.1}$$

, where $M$ is the merit of a feature subset S containing $k$ features, $\bar{r}_{fc}$ the average of the correlation between the features and the class (target), and $\bar{r}_{ff}$ the average intercorrelation between features. The numerator indicates how predictive a set of features is, while the denominator indicates how much redundancy exists within them.

## 4.3 Data Normalization

Since each input and output variable used in ANN development have different orders of magnitude, all experimental data must be normalized to be at a comparable range. Therefore, both input and output data were normalized between 0 and 1, by using the Eq. (**4.2**) and (**4.3**) respectively.

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4.2}$$

$$x_n = \frac{log(1+x)}{max(log(1+x))} \tag{4.3}$$

Where $x$, $x_{min}$ and $x_{max}$ are the value to be scaled, the minimum and the maximum value of all records, respectively and $x_n$ is the normalized value.

## 4.4 Implementation of model

In this study, a robust multivariate non-linear Artificial Neural Network (ANN) modelling technique with inputs from several lagged climate variables is developed as new technique to predict one-month ahead rainfall in Benin Republic. To evaluate the forecasting ability of non-linear ANN models, standard Multiple Linear Regression (MLR) models were used as a benchmark. Figure **4.3** shows the steps followed for the implementation of the ANN method as well as MLR prediction method used for comparison.

All models were developed in MATLAB® Version 9.9.0.1495850 (R2020b) for each weather station using a total of 421 records from January 1979 to December 2014 for calibration (training and validation) and evaluated by using test data consisting of 48 records over a 4 year period from January 2015 to December 2018. Numerical experiments were carried out on a personal computer with an Intel ® Core (TM) i5-5200U CPU @ 2.2GHz 2.2GHz, 8.00GB memory and 250GB SSD hard drive, which ran on Microsoft Windows 10. The Python programming language was used to extract the data on each research area and spatial correlation computation. Microsoft ®Excel® 2013 was used for data pre-processing and relative importance computation and some plots. Waikato Environment for Knowledge Analysis (WEKA) Version 3.9.4 © 1999-2019 was utilized to implement a correlation based feature selection subset evaluator in order to determine the most influencing and important predictors.

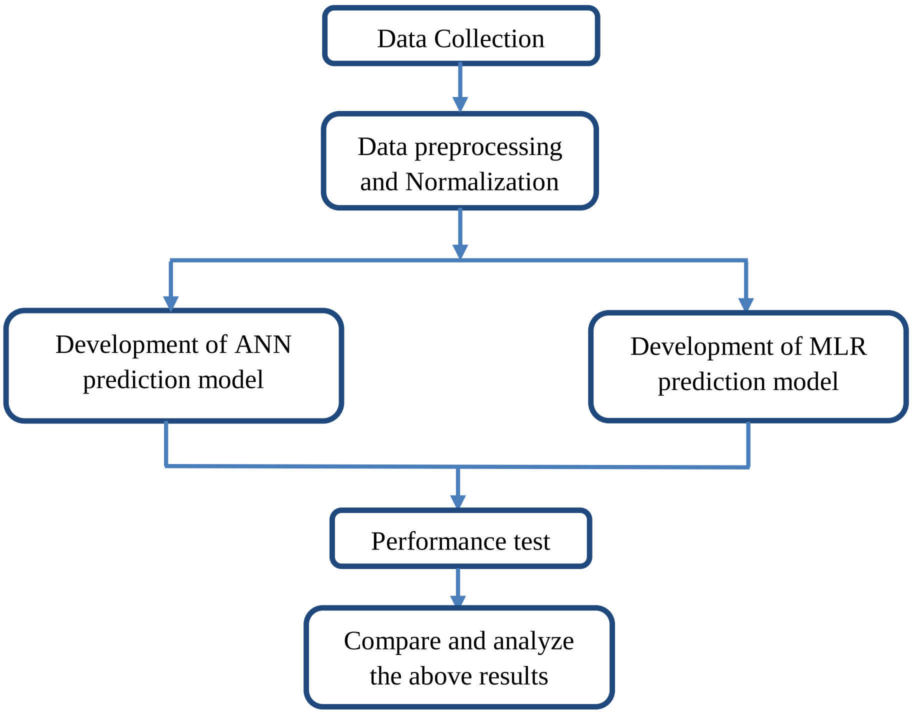



Figure 4.3: Flowchart of the research methodology.

#### 4.4.1 Artificial Neural Network

Artificial Neural Network is a soft computing method that mimics the behavior of biological neural processing [105], [106]. Neural networks were first introduced in 1943 by McCulloch and Pitts [107]. ANN is a nonlinear statistical technique that has become popular among scientists as an alternative technique for predicting and modelling complicated time series, weather phenomena and climate variables [108]. ANN has been widely used in a number of fields, including medicine, engineering, hydrology, etc.

An artificial neural network consists of simple neurons, and links that process information in order to find the relationship between inputs and outputs. The input signal is received through the input neurons. Then, the input units are weighted and added. Finally, the input is transformed into an output unit through the transfer function (Figure **4.4**). The transfer function also known as an activation function is generally divided into two parts, the combination function and the transfer function. The combination function assigns weights to each input and combines the weighted inputs in a single value. The transfer functions produce an output. Transfer functions could be any mathematical function.

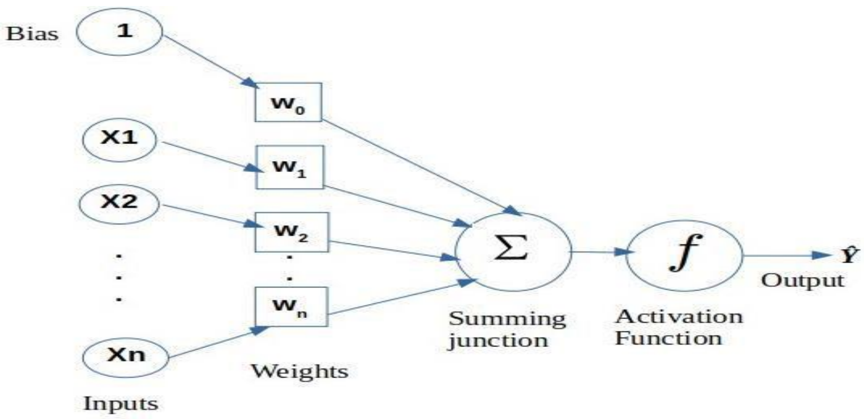



Figure 4.4: Architecture of an Artificial Neural Network.

The most widely used transfer functions are the logarithmic sigmoid (logsig), the tangent sigmoid (tansig), and the linear (purelin) transfer functions (Figure **4.5**).

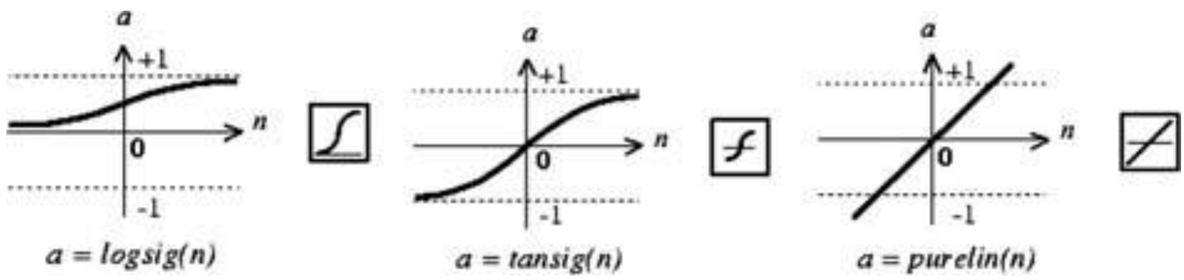



Figure 4.5: Graphical representation for activation (transfer) functions (MATLAB).

**Network Architectures**

Neural Network can be viewed as combination of two or more artificial neurons and the way these individual neurons interconnect is called topology or architecture of the artificial neural network. According to the connection pattern (architecture), two types of neural networks may be identified: feed-forward networks and recurrent/feedback networks. An illustration of a feed-forward network and recurrent/feedback network architectures is given in Figure **4.6**.

a) **Feed-forward networks**: In feed-forward neural network (FFNN) architecture, data moves only in a forward direction from input layer which goes through the hidden layer and then to the output layer. There are three types of networks in this category: Single-layer perceptron, Multilayer perceptron and Radial basis function nets (Figure **4.6**). The single layer neural network is the simplest form of ANN. In a single layer feed-forward neural network the input layer directly links to the output layer. There is no hidden layer in this neural network. On the other hand, a multi-layer feed-forward neural network has one or more hidden layers between the input layer and the output layer. These neural networks are also called multi-layer perceptron (MLP).

b) **Recurrent/feedback networks**: In recurrent/feedback architecture, hidden layers and output layers have recurrent connections. Feedback neural networks have at least one feedback loop, from a hidden layer either to an input layer or from an output layer to a hidden layer. This group includes four distinct network types: Competitive Networks, Kohonen's SOM, Hopfield Networks and ART models.

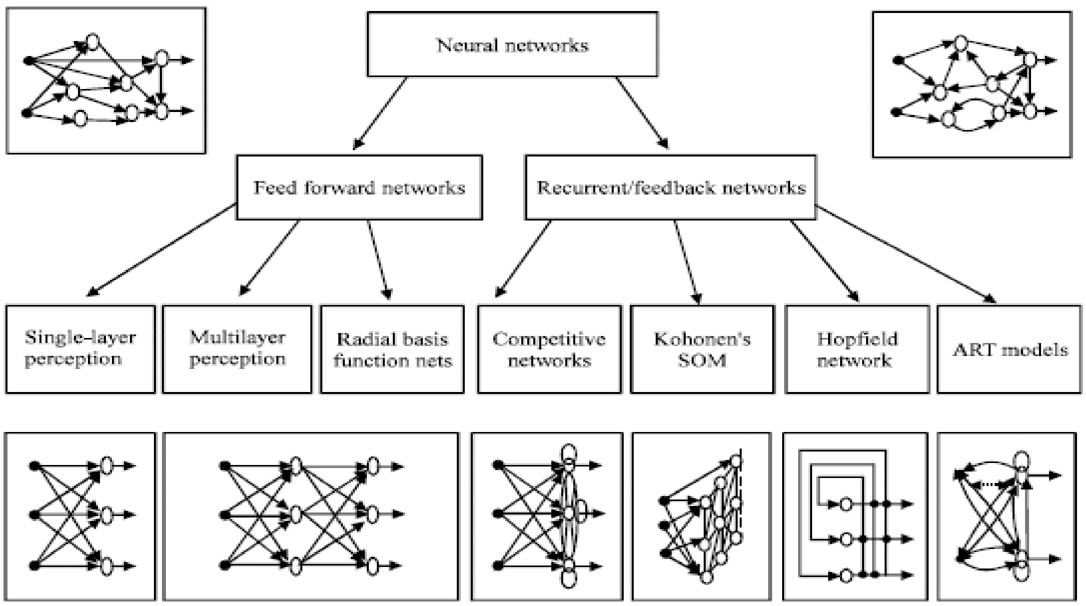



Figure 4.6: Taxonomy of Neural Networks [24].

#### 4.4.1.1 ANN Model Development

In this study multi-layer perceptron (MLP) is implemented as network for rainfall predictions due to its capability of handling complex and nonlinear problems. The multi-layer perceptron (MLP) has been widely used in hydrologic forecasting model [109]–[112].

The number of neurons input and output is dependent on the number of data input and output. Essentially, the input layer exists mainly to receive input data for further processing in the network. The hidden layers are a very important part in a MLP since they provide the nonlinearity between the input and output sets. Three-layered FFNNs is used as ANNs in this study (Figure **4.7**). Any function can be approximated by ANN with single hidden layer as enough freedom is given [113]. One hidden layer has been adopted and the optimal ANN structure was determined by varying the number of hidden neurons from 2 to 10 (see appendix, Table **B.1** to Table **B.6**), since there is no standard style for the determination of the number of hidden layers or the number of neurons. According to [114], this number is determined according to the view of the designer of the model.

Via a linear sum operation, the input data in the input layer is passed to each neuron in the hidden layer, and the product of inputting the linear sum to the activation function is the output of the hidden layer neuron. The same procedure is followed from the hidden layer to the output layer.

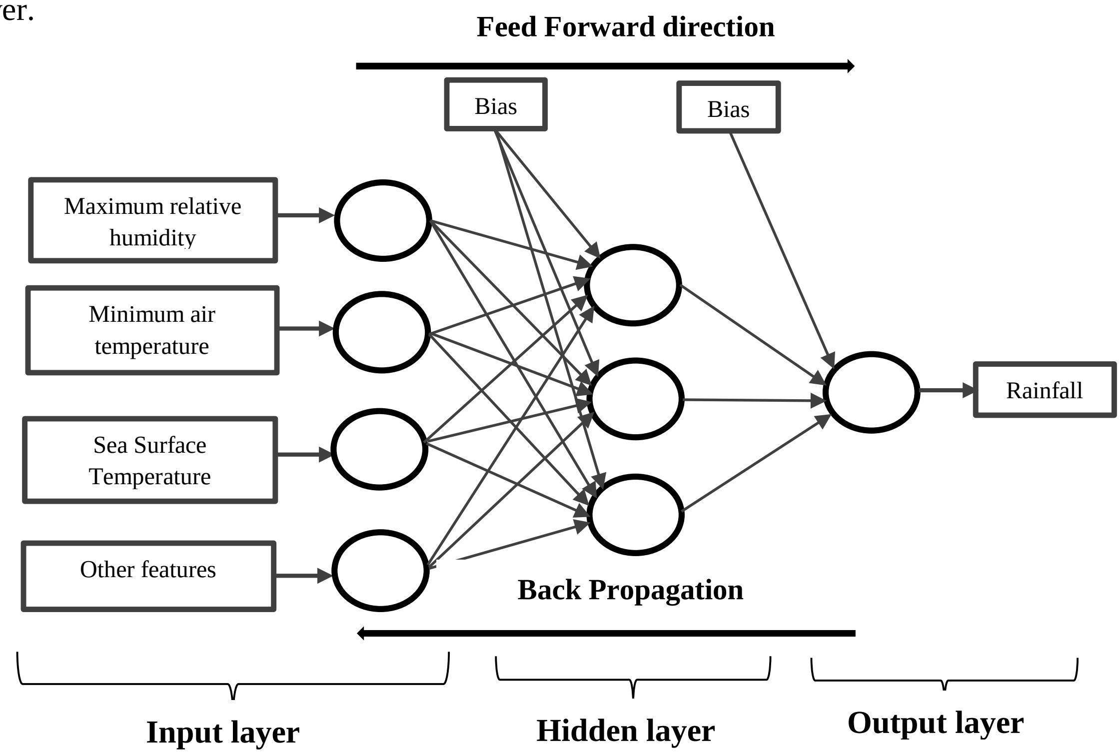



Figure 4.7: Three-layer forward neural networks (FFNNs) with back propagation structure.

The network illustrated in Figure **4.7**, can be expressed mathematically by a linear combination of the transferred input values as in Eq. (**4.4**):

$$\hat{y}_k = f_2\left[\sum_{h=1}^{n} w_{ho} f_1\left(\sum_{i=1}^{m} w_{ih} x_i + w_{bh}\right) + w_{bo}\right] \tag{4.4}$$

,where $\hat{y}_k$ is the forecasted $k^{\text{th}}$ output value, $f_2$ is the activation function for the output neuron, $n$ is the number of hidden neurons, $w_{ho}$ is the weight connecting the $h^{\text{th}}$ neuron in the hidden layer and neuron in the output layer, $f_1$ is the activation function for the hidden neuron, $m$ is the number of input neurons, $w_{ih}$ is the weight connecting the $i^{th}$ neuron in the input layer and $h^{th}$ neuron in the hidden layer, $x_i$ is the $i^{th}$ input variable, $w_{bh}$ is the bias for the $h^{th}$ hidden neuron, and $w_{bo}$ is the bias for output neuron [115], [116]. In this study $f_1$ is considered as logsig function, which is a nonlinear function and $f_2$ is considered as tansig function, which is also nonlinear function.

The data is divided into two sets: the calibration (training and validation) and the testing sets. The training set is used to update the network weights and biases; the validation set, used to guarantee the generalization capability of the model; and the test set, used to check the generalization. The multi-layer feed-forward neural networks have the ability to learn through training. Training a multilayer perceptron is a procedure to find the combination of weights, which results in the smallest error. To obtain the optimal weights (parameters) of the neural network structure, Levenberg-Marquardt Back Propagation algorithm was used to train the network. This algorithm is a second-order optimization method that is usually faster and more efficient than any other technique based on Back Propagation [117].

In most of the studies, the applications related to rainfall forecasting utilized feed forward neural network (FFNN) that incorporates the standard static multilayer perceptron (MLP) trained with the back-propagation algorithm [96]. Many algorithms could be used to train a multi-layer perceptron, but the back propagation algorithm and the algorithms derived from it are the most computationally straightforward algorithms for training the multi-layer perceptron. The back propagation algorithm involves a forward propagating step followed by a back-propagating step. Both the forward and back propagation steps are done for each pattern presentation during training. The forward propagation step begins with the presentation of an input pattern to the input layer of the network, and continues as activation level calculation propagates forward through the hidden layers in each successive layer. Every processing unit sums its inputs and then applies a transfer function to compute its output. The output layer of units then produces the output of the network. The back-propagation step begins with the comparison of the network's output pattern to the target vector, when the difference of error is calculated. The back propagation step then calculates error values for hidden units and changes for their incoming weights , starting with the output layer and moving backward through the successive hidden layers. In this back propagation step, the network corrects its weight in such a way as to decrease the observed error. Figure **4.8** shows the FFNN with back propagation step.

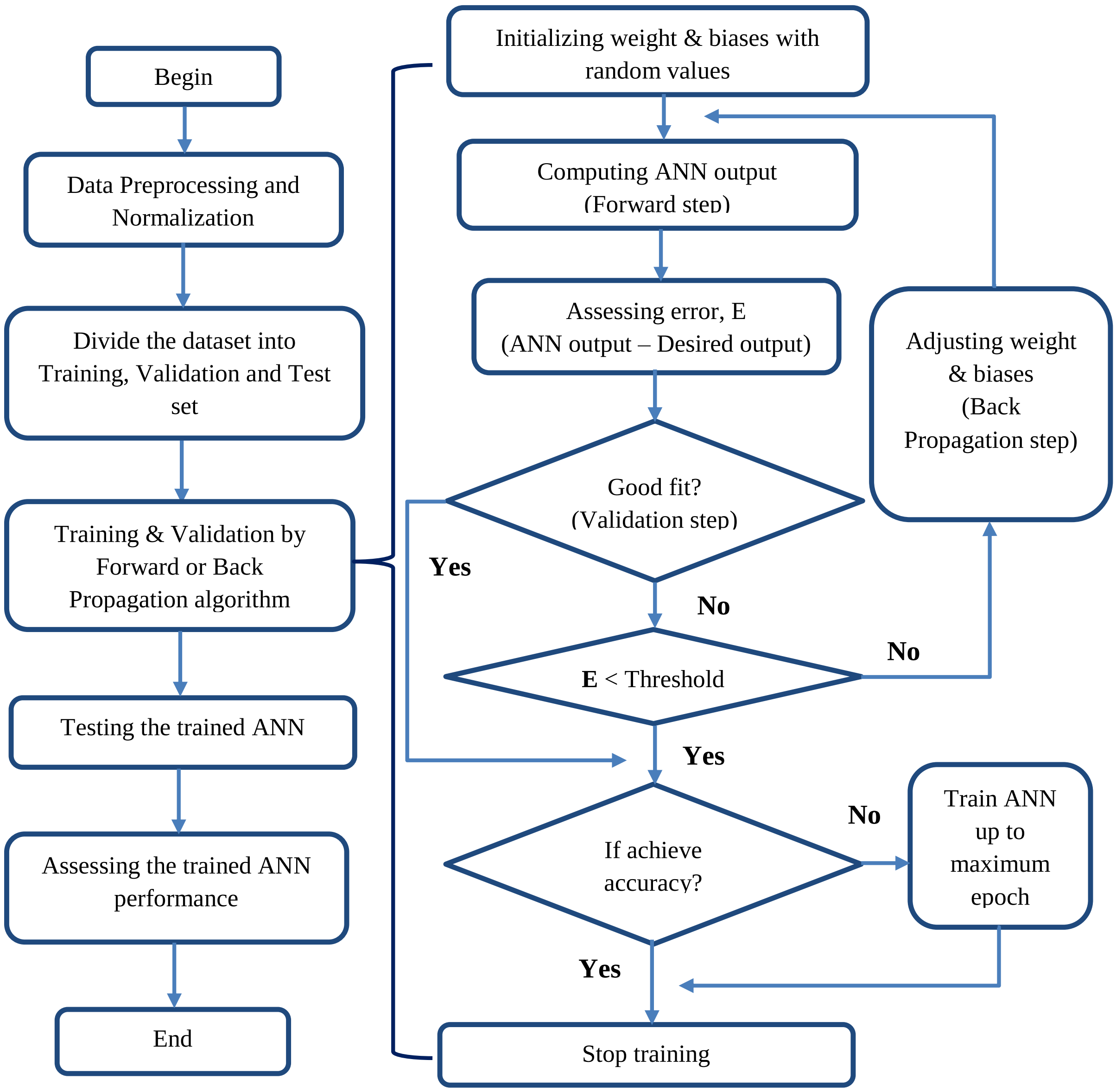



Figure 4.8: Flowchart of the rainfall prediction using FFNN with back propagation.

The amplitude of the gradient **E** during the training phase and the number of training epochs are used to end the training. If the gradient magnitude is less than the threshold, then the training process will stop. When the number of training epochs reaches the maximum, the training will also stop. There is often the possibility of having an over-fitted model in ANN modeling. The early stopping method was applied to avoid the over-fitting problem, in which, the output error of training and validation in each epoch is monitored and training is stopped when the validation error is minimum [118], [119]. Figure **4.9** illustrates the early-stop process in MATLAB when creating the models.

The MATLAB code was developed to take advantage of the programming capabilities of MATLAB in order to read in the input files, use the neural network, implement the early-stop technique, and calibrate and test the models. The first page of this code is shown in Figure **4.10**.

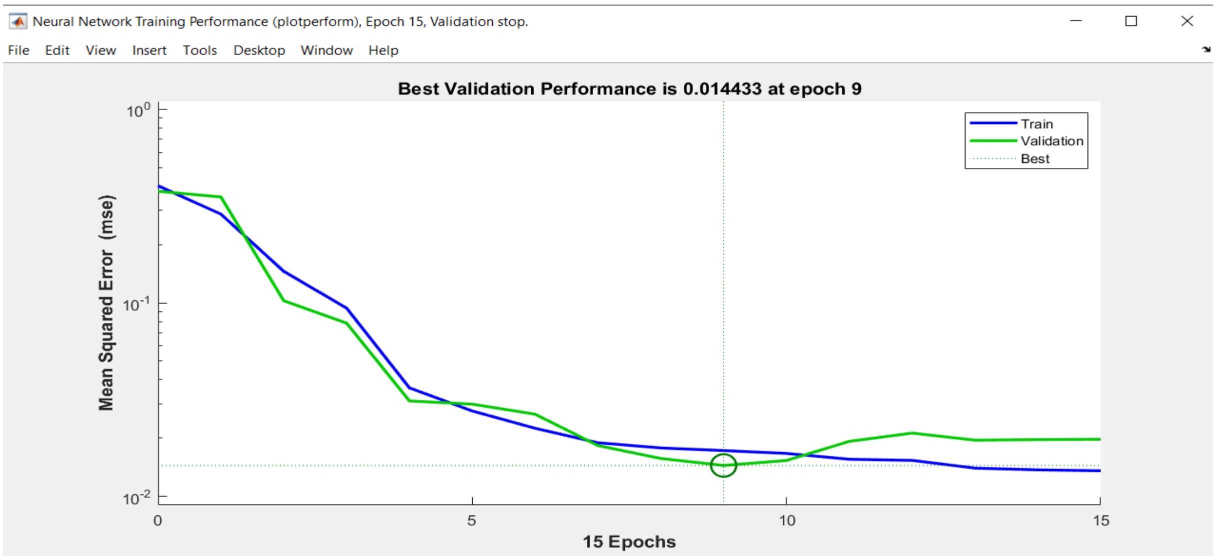



Figure 4.9: MSE curve sample for ANN training in MATLAB using the early-stop strategy.

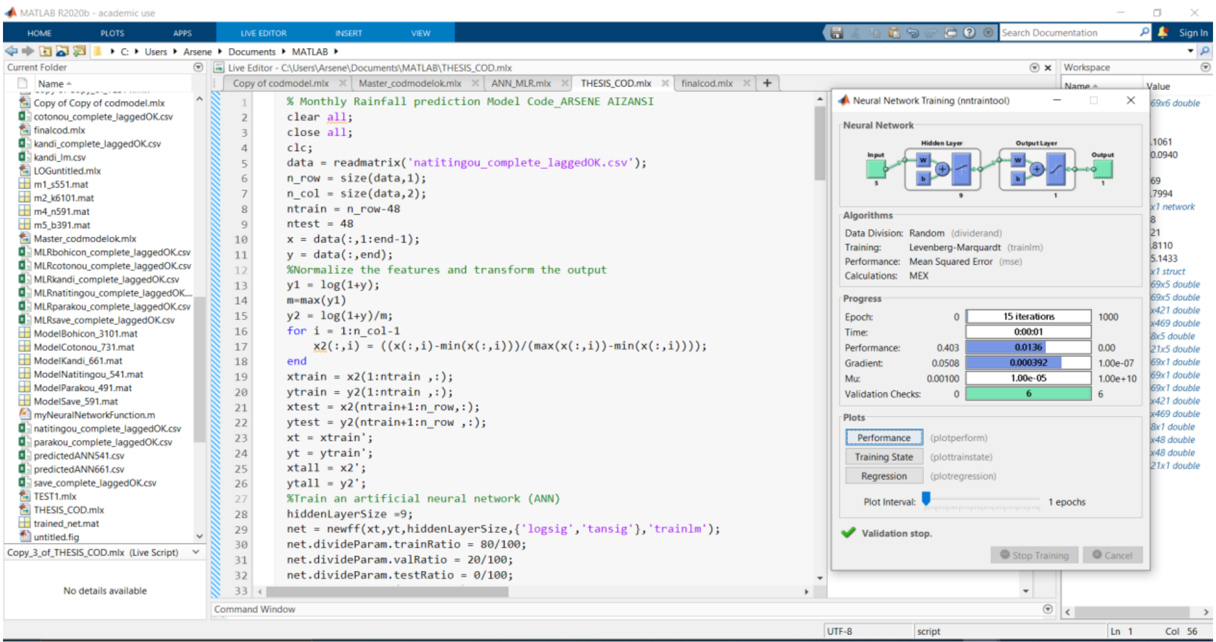



Figure 4.10: First page of the MATLAB code used for the models.

### 4.4.2 Multiple Linear Regression

Multiple Linear Regression (MLR) is one of the simplest and widely used statistical methods for rainfall forecasting. It is used to predict a dependent or predictand variable based on several other independent or predictor variables through the least square method. Linearity is assumed between the response and predictor variables. The MLR model can be presented mathematically by Eq. (**4.5**):

$$\hat{y} = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + C \tag{4.5}$$

, where $\hat{y}$ is the predicted value of the dependent variable (Monthly rainfall in our case) and $X_1$, $X_2$,…, $X_n$ are different independent or predictor variables (like Relative humidity, Air temperature, SST, Uwind etc.) and $a_1$, $a_2$, …, $a_n$ are the regression coefficients and $C$ is constant (intercept on the $\hat{y}$ axis).

Verification of multicollinearity is another critical step in multiple regression modeling. When two or more predictor variables are strongly correlated, multicollinearity occurs, resulting in significant differences in parameter estimated in response to minor changes in the data or the model. De Veaux et al. in [120] noted that "neural networks are generally indifferent to multicollinearity issues." As a result, the neural network model is considered to be independent of the statistical regression model's limitations [79], and collinearity between predictor variables is typically ignored when a machine learning model focuses on predictive capacity [121].

The indicators used to identify multicollinearity among predictors are tolerance (T) and variance inflation factor (VIF):

$$Tolerance = 1 - R^2 \qquad VIF = \frac{1}{Tolerance} \tag{4.6}$$

, where $R^2$ is the coefficient of determination.

A tolerance of less than 0.20-0.10 or a VIF greater than 5-10 indicates a multicollinearity problem. In this study, it was decided that the variables have multicollinearity if VIF > 10 [122].

In this study, MLR is used as yardstick to compare the forecasting ability by the ANN models.

## 4.5 Statistical evaluation of model performance

All models were tested for prediction efficiency by comparing observed and predicted rainfall using four different forecast accuracy measures calculated from the test datasets: the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), the Coefficient of determination (R2) and Nash-Sutcliffe Efficiency (NSE).

The RMSE represents the error between model predictions and target values. It can be computed with Eq. (**4.7**) and ranges from 0 to ∞. Since there is no criterion for a "good" value, lower values of RMSE are preferable.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(PR_i - OB_i)^2}{n}} \tag{4.7}$$

Where *PR* and *OB* are model predictions and observed values respectively, and *n* is the number of target values.

The MAE is simply the average of the absolute errors between model predictions and target values. It can be computed with Eq. (**4.8**) and ranges from 0 to ∞. Similar to RMSE, lower values of MAE indicate a good correlation between model predictions and actual data. The small values of RMSE and the MAE indicate small deviations of the predictions from actual observations. The RMSE and MAE metrics are commonly used in statistical modelling due to their theoretical importance. It is well known that the MAE is less sensitive to outliers than the RMSE [123].

$$MAE = \frac{\sum_{i=1}^{n}|PR_i - OB_i|}{n} \tag{4.8}$$

The Coefficient of determination ($R^2$), also known as the Goodness of fit, indicates the proportion of the variance in the dependent variable that is predictable from the independent variable. It gives information about the length of the relationship between the models' predicted data and the observed data, and ranges from 0 to 1, with 1 being the best fit between predictions and actual data. It computed with Eq. (**4.9**)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(OB_i - PR_i)^2}{\sum_{i=1}^{n}(OB_i - \overline{PR})^2} \tag{4.9}$$

$\overline{PR}$ is the mean of predicted values.

The Nash-Sutcliffe Efficiency (NSE) was proposed by Nash and Sutcliffe in 1970. It can be defined as a measure of how the observed variance is simulated. It can be computed with Eq.(**4.10**). The NSE can range from -∞ to 1. An efficiency NSE = 1 means a perfect prediction. An efficiency of 0 means that the model predictions are as reliable as the observed mean, while an efficiency -∞ < NSE < 0 indicates that the observed data mean is a better predictor than the model predictions.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(OB_i - PR_i)^2}{\sum_{i=1}^{n}(OB_i - \overline{OB})^2} \tag{4.10}$$

$\overline{OB}$ is the mean of observed values.

Table **4.5** presents the Nash-Sutcliffe Efficiency coefficient rating on monthly time step given by [124].

Table 4.5: NSE ratings for recommended statistics for a monthly time step

| **Performance rating** | **NSE** |
|---|---|
| Very Good | 0.75<NSE≤1.0 |
| Good | 0.65<NSE≤0.75 |
| Satisfactory | 0.50<NSE≤0.65 |
| Unsatisfactory | NSE≤0.50 |

Source: [124]

## 4.6 Relative Importance of independent variables

The relative importance of input variables is the contribution of each input variables to the prediction of the dependent variable. It is a good index to establish the relations between input variables and the output value provided by a Multilayer Perceptron [125]. Several methods have been developed by researchers for calculating the relative importance of input variables. It includes Garson’s algorithm [126], [127], connection weight approach [128], partial derivatives [129] and others [130], [131].

In this study, Olden’s connection weight method is applied. The choice is predicated on several reasons: Olden and Jackson [128] demonstrated it is more accurate than the Garson's algorithm; [132] compared various methods for determining the contribution of input variables in ANNs.

Their work demonstrated that the connection weights approach was the least biased among others, owing to the fact that it is obtained directly from the weights of the neural network's connections. This position was corroborated by [125], [133].

Olden's method calculates the product of the raw input-hidden and hidden-output connection weights between each input node and output node, and sums the products across all hidden nodes. The details of the method are given in Eq. (**4.11**)

$$RI_i = \sum_{h=1}^{n} w_{ih} w_{ho} \tag{4.11}$$

Where $\boldsymbol{RI_i}$ is the relative importance of input neuron $\boldsymbol{i}$, $\boldsymbol{w_{ih}}$ and $\boldsymbol{w_{ho}}$ are final weights of the connection from input neuron to hidden neurons and from hidden neurons to output neuron respectively, $\boldsymbol{h}$ is the total number of hidden neurons, and $\boldsymbol{o}$ is the output neuron. The connection weight method has the advantage to consider the direction of the input–output interaction. A positively or negatively high value of $\boldsymbol{RI}$ indicates a significant predictor variable for the output of the neural network model.

## 4.7 Summary

In this chapter, a description of rainfall data and meteorological variables was provided. Data preprocessing including cross-correlation between rainfall and other meteorological variables for optimal lag time selection and the predictor's selection techniques were also provided. Detailed information on the implementation of models and the performance measures used to evaluate each model were provided. Finally, the relative importance of input variables on rainfall prediction is discussed.

# Chapter 5

# RESULTS AND DISCUSION

This chapter presents the results of the most influencing input variables of each station. Secondly, the monthly rainfall prediction results of the ANN and MLR models within each station are presented. The chapter further assesses the ANN model performance compared to MLR model. All models were developed using appropriate input variables. The models were trained using data from January 1979 to December 2014 and tested using data from January 2015 to December 2018 in all 6 locations. The prediction performance of models was evaluated by comparing observed and predicted rainfall. Then compared the predictions zone-wise to assess the influence of geographic regions on the performance of models. The relative importance of each input variable on rainfall was finally assessed in this work.

The structure of the chapter is as follows. Feature selection results are presented in section 5.1, followed by the monthly rainfall prediction results for each location in section 5.2 and comparison of both ANN and MLR models in section 5.3. Relative importance of input variables are also discussed in section 5.4. The chapter is concluded in section 5.5.

## 5.1 Features Selection Results

Table **5.1** shows the Correlation-based Feature Selection subset evaluator with the search methods beside the number and order of the selected attributes for all 6 weather stations.

Table 5.1: The results of attributes selection.

| Stations | Attributes evaluator | Search method | No of Selected attributes | Selected attributes |
|---|---|---|---|---|
| Cotonou | Correlation based Feature Selection subset evaluator | Best First & Greedy Stepwise | 7 | RHmax,RHmin,Tmax,Tmin,Eva,Uwind,Vwind |
| Bohicon | | | 3 | Eva,SST,Vwind |
| Save | | | 5 | Tmin,Eva,SST,Uwind,Vwind |
| Parakou | | | 4 | Eva,SST,Uwind,Vwind |
| Natitingou | | | 5 | Tmin,Eva,SST,Uwind,Vwind |
| Kandi | | | 6 | RHmax,Tmax,Tmin,Eva,SST,Uwind |

As shown in Table **5.1**, the same choice of dataset attributes is obtained for both Best First and Greedy Stepwise search methods (see appendix, Figure **A.1** & Figure **A.2**). Only input variables that affect significantly the long-term rainfall prediction out of the 8 variables have been used in this study.

## 5.2 Monthly rainfall prediction

In this section, the monthly rainfall prediction results for each model in predicting monthly rainfall is presented with appropriate input variables in each location. Then, it summarizes the performance of all models. The ANN model performance with MLR model were finally compared using computational results and time series plots.

### 5.2.1 Monthly rainfall prediction by ANN models

The optimal ANN model structure for each location (see appendix, Figure **A.3**) was determined by setting the hidden layer number to one and evaluating a different number of hidden neurons ranging from 2 to 10 and selecting the best based on the lowest root mean squared error value obtained from the test datasets (see Table **B.1** to Table **B.6** in appendix).

Table **5.2** summarizes the prediction performance of the ANN model with the best input variables in each location. The performance measure RMSE for the ANN model in predicting monthly rainfall ranges from 38.86 mm to 68.00 mm, MAE from 26.90 mm to 48.30 mm, $R^2$ from 0.47 to 0.87 and NSE from 0.45 to 0.85. These results show that the model provides unsatisfactory prediction in Bohicon, satisfactory prediction in Cotonou and Save, good prediction in Parakou and very good prediction in Natitingou and Kandi based on NSE performance rating according to [124] (see Table **4.5**).

Table 5.2: Optimal ANN model prediction performance on test set (Period 2015-2018).

| Stations | Model | Input variables | RMSE | MAE | $R^2$ | NSE |
|---|---|---|---|---|---|---|
| Cotonou | ANN (7-3-1) | RHmax,RHmin,Tmax,Tmin,Eva,Uwind,Vwind | 68.00 | 48.30 | 0.64 | 0.62 |
| Bohicon | ANN (3-10-1) | Eva,SST,Vwind | 51.00 | 37.47 | 0.47 | 0.45 |
| Save | ANN (5-5-1) | Tmin,Eva,SST,Uwind,Vwind | 51.00 | 34.71 | 0.56 | 0.55 |
| Parakou | ANN (4-9-1) | Eva,SST,Uwind,Vwind | 56.45 | 37.48 | 0.68 | 0.67 |
| Natitingou | ANN (5-9-1) | Tmin,Eva,SST,Uwind,Vwind | 38.86 | 26.90 | 0.87 | 0.85 |
| Kandi | ANN (6-10-1) | RHmax,Tmax,Tmin,Eva,SST,Uwind | 45.52 | 27.06 | 0.83 | 0.83 |

Figure **5.1** illustrates the regression plot of the ANN model to predict monthly rainfall during the testing phase (2015-2018). The dashed line in each plot represents the perfect result predicted = observed rainfall data. The blue line represents the best fit linear regression line between predicted and observed data. The regression plot was used to determine the correlation between the predicted rainfall and the observed rainfall values using Pearson's correlation R. If R = 1, predicted and the observed rainfall have a perfect linear relationship. If R is close to zero, then there is no linear relationship between them. The correlation coefficient between the ANN models predicted values and observed data for stations Cotonou, Bohicon, Save, Parakou, Natitingou and Kandi are 0.80, 0.69, 0.75, 0.82, 0.93 and 0.91, respectively and significant at 1% level (i.e. $p < 0.01$).

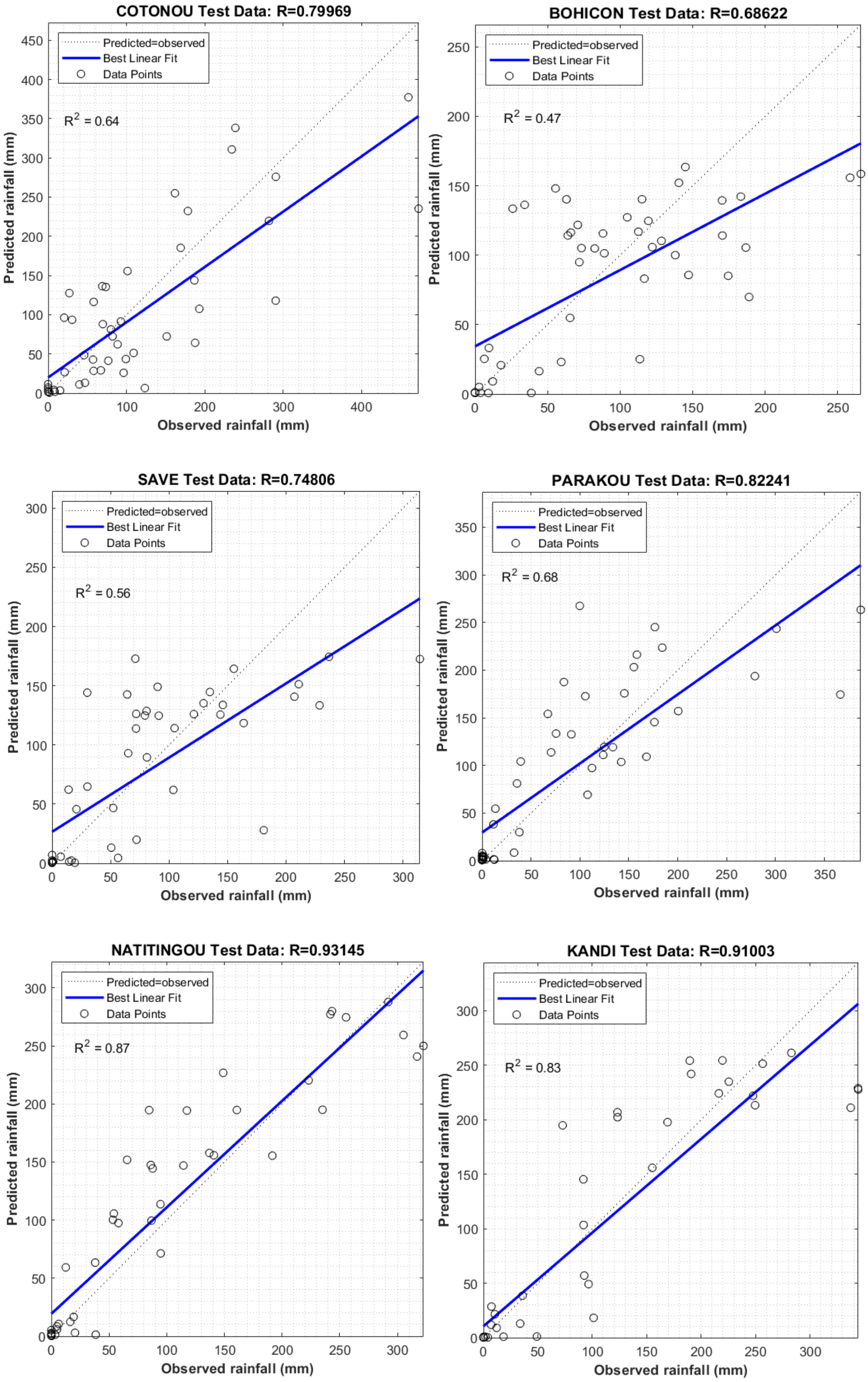
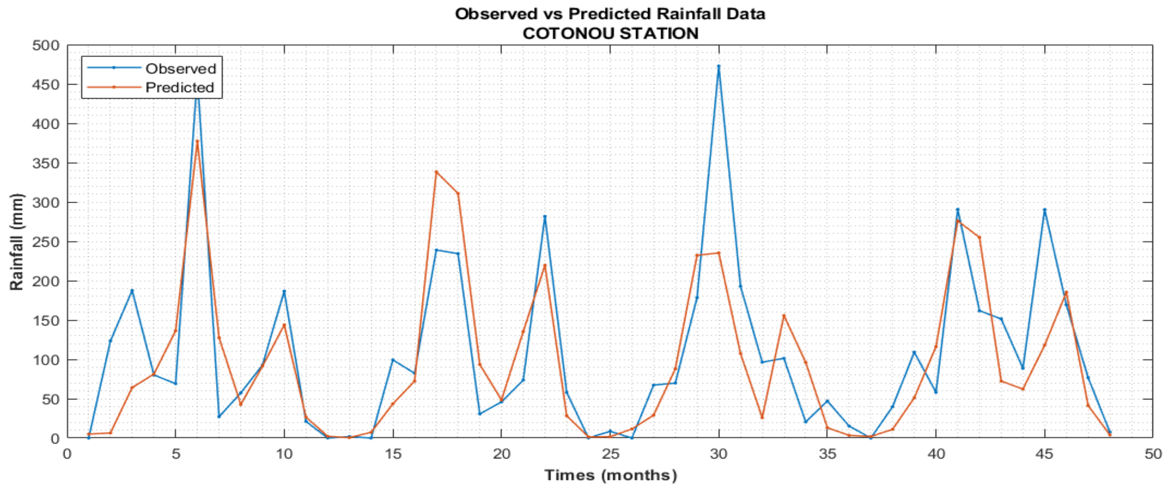
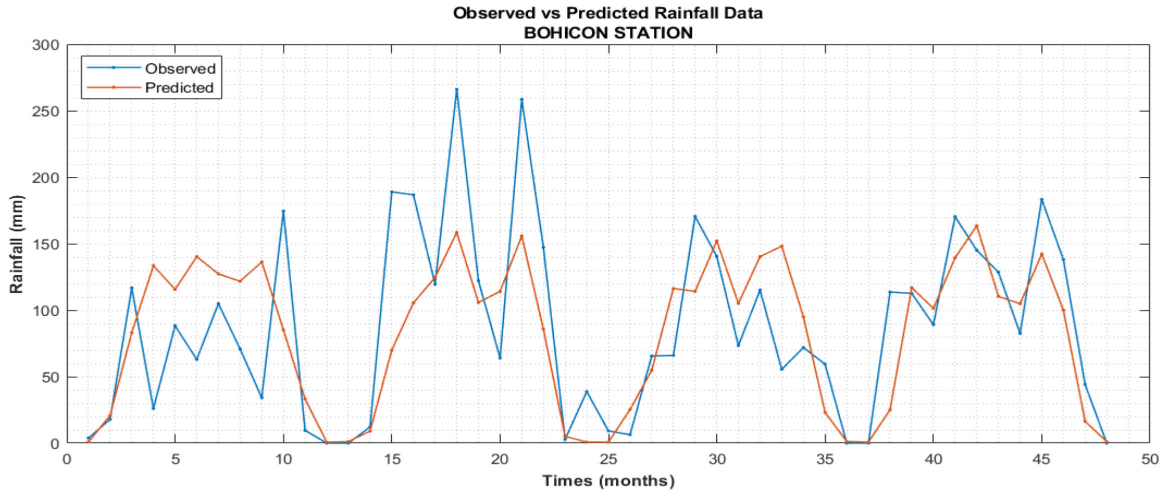
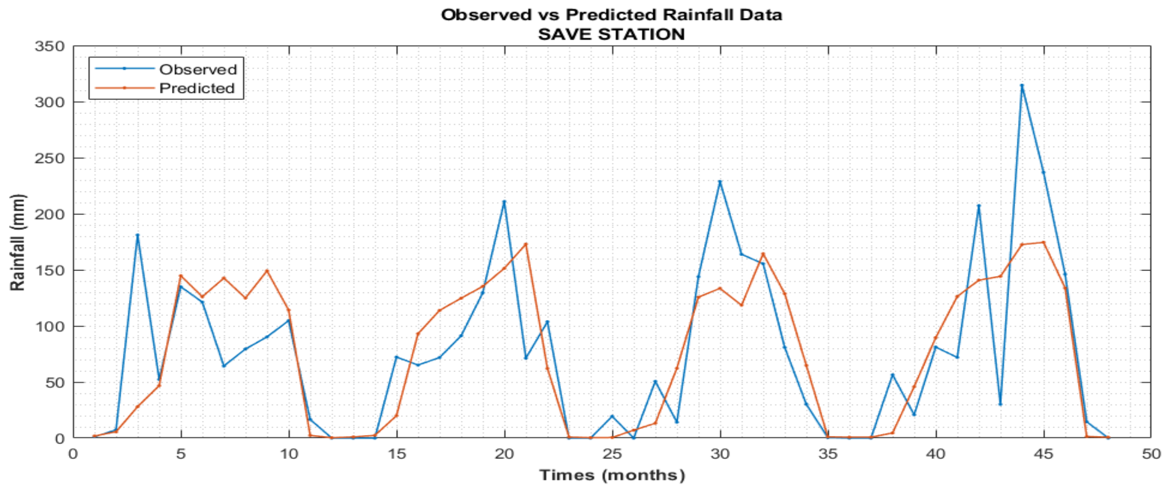



Figure 5.1: Scatter plot of actual versus ANN predicted monthly rainfall in all stations in the study area.

The graphical display of the model predictions and the observed rainfall in the test period (2015-2018) is given in Figure **5.2**. The figure shows that the model predictions reasonably follow the series patterns at all locations. However, the model appeared to underestimate some monthly rainfall values when compared with actual monthly rainfall and some of the highest monthly rainfall values failed to be replicated.



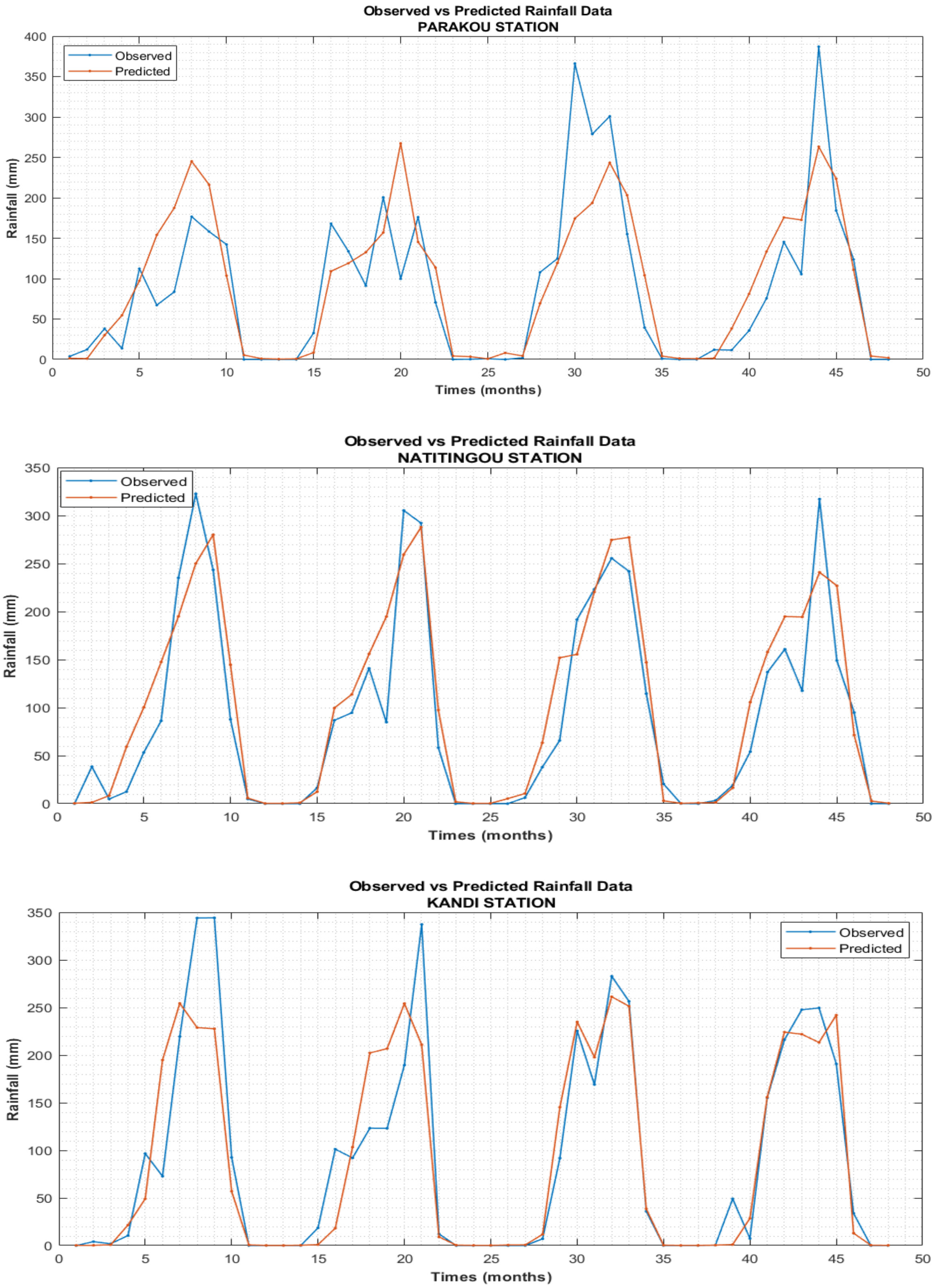



Figure 5.2: Observed rainfall vs. ANN model predictions in all stations in the study area (Period 2015-2018).

### 5.2.2 Monthly rainfall prediction by MLR models

The present study investigated the multicollinearity by calculating the values of the VIF between the predictor variables, as shown in Table **5.3**.

Table 5.3: Initial input variables and their VIF values.

| Features | Cotonou (7 variables) | Bohicon (3 variables) | Save (5 variables) | Parakou (4 variables) | Natitingou (5 variables) | Kandi (6 variables) |
|---|---|---|---|---|---|---|
| | VIF_Values | | | | | |
| RHmax | 1095.37 | - | - | - | - | 67.84 |
| RHmin | 218.96 | - | - | - | - | - |
| Tmax | 1134.92 | - | - | - | - | 561.44 |
| Tmin | 728 | - | 622.94 | - | 209.87 | 107.59 |
| Eva | 51.87 | 43.67 | 62.56 | 33.71 | 39.81 | 20.2 |
| SST | - | 46.76 | 831.74 | 33.67 | 307.84 | 987.09 |
| Uwind | 3.65 | - | 2.77 | 5.15 | 5.71 | 6.9 |
| Vwind | 3.98 | 1.69 | 3.52 | 5.48 | 5.75 | - |

The final result is shown in Table **5.4**, which is obtained by removing one by one each feature with VIF > 10, beginning with the highest.

Table 5.4: Final input variables and their VIF for the MLR models.

| Features | Cotonou (3 variables) | Bohicon (2 variables) | Save (3 variables) | Parakou (3 variables) | Natitingou (3 variables) | Kandi (2 variables) |
|---|---|---|---|---|---|---|
| | VIF_Values | | | | | |
| RHmax | - | - | - | - | - | - |
| RHmin | - | - | - | - | - | - |
| Tmax | - | - | - | - | - | - |
| Tmin | - | - | - | - | - | - |
| Eva | 2.66 | 1.06 | 1.07 | - | 4.37 | 1.03 |
| SST | - | - | - | 4.37 | - | - |
| Uwind | 2.87 | - | 2.01 | 4.7 | 3.49 | 1.03 |
| Vwind | 3.49 | 1.06 | 1.92 | 4.41 | 5.37 | - |

Table **5.5** summarizes the prediction performance of the MLR model with the best final input variables in each location. The performance measure RMSE for the MLR model in predicting monthly rainfall ranges from 55.42 mm to 111.79 mm, MAE from 37.36 mm to 66.13 mm, $R^2$ from 0.34 to 0.77 and NSE from -0.29 to 0.66. These results show that the MLR model provides unsatisfactory prediction in Cotonou, Bohicon, Save, Parakou, and Natitingou and good prediction in Kandi based on NSE performance rating Table **4.5**.

Table 5.5: MLR model prediction performance on test set (Period 2015-2018).

| **Stations** | **MLR Model Equation** | **RMSE** | **MAE** | **$R^2$** | **NSE** |
|---|---|---|---|---|---|
| Cotonou | 0.37127+0.20441Eva-0.25474Uwind+0.37748Vwind | 83.1 | 55.09 | 0.49 | 0.43 |
| Bohicon | 0.56007-0.61095Eva+0.43135Vwind | 60.23 | 42.00 | 0.34 | 0.24 |
| Save | 0.54972-0.6317Eva+0.24479Uwind+0.32832Vwind | 55.42 | 40.87 | 0.51 | 0.47 |
| Parakou | 0.13969+0.63089SST+0.41773Uwind-0.21288Vwind | 111.79 | 66.13 | 0.44 | -0.29 |
| Natitingou | 0.97397-0.62129Eva+0.25192Uwind-0.49201Vwind | 91.01 | 54.01 | 0.67 | 0.18 |
| Kandi | 0.68026-0.83092Eva+0.48502Uwind | 63.80 | 37.36 | 0.77 | 0.66 |

A scatter plot of actual versus MLR predicted monthly rainfall values in the test period (2015-2018) is shown in Figure **5.3**. The regression analysis shows that how close are the output value to the actual target values. The correlation coefficient between the MLR models predicted values and observed data for stations at Cotonou, Bohicon, Save, Parakou, Natitingou and Kandi are 0.70, 0.59, 0.72, 0.66, 0.82 and 0.88, respectively and significant at 1% level (i.e. $p < 0.01$).

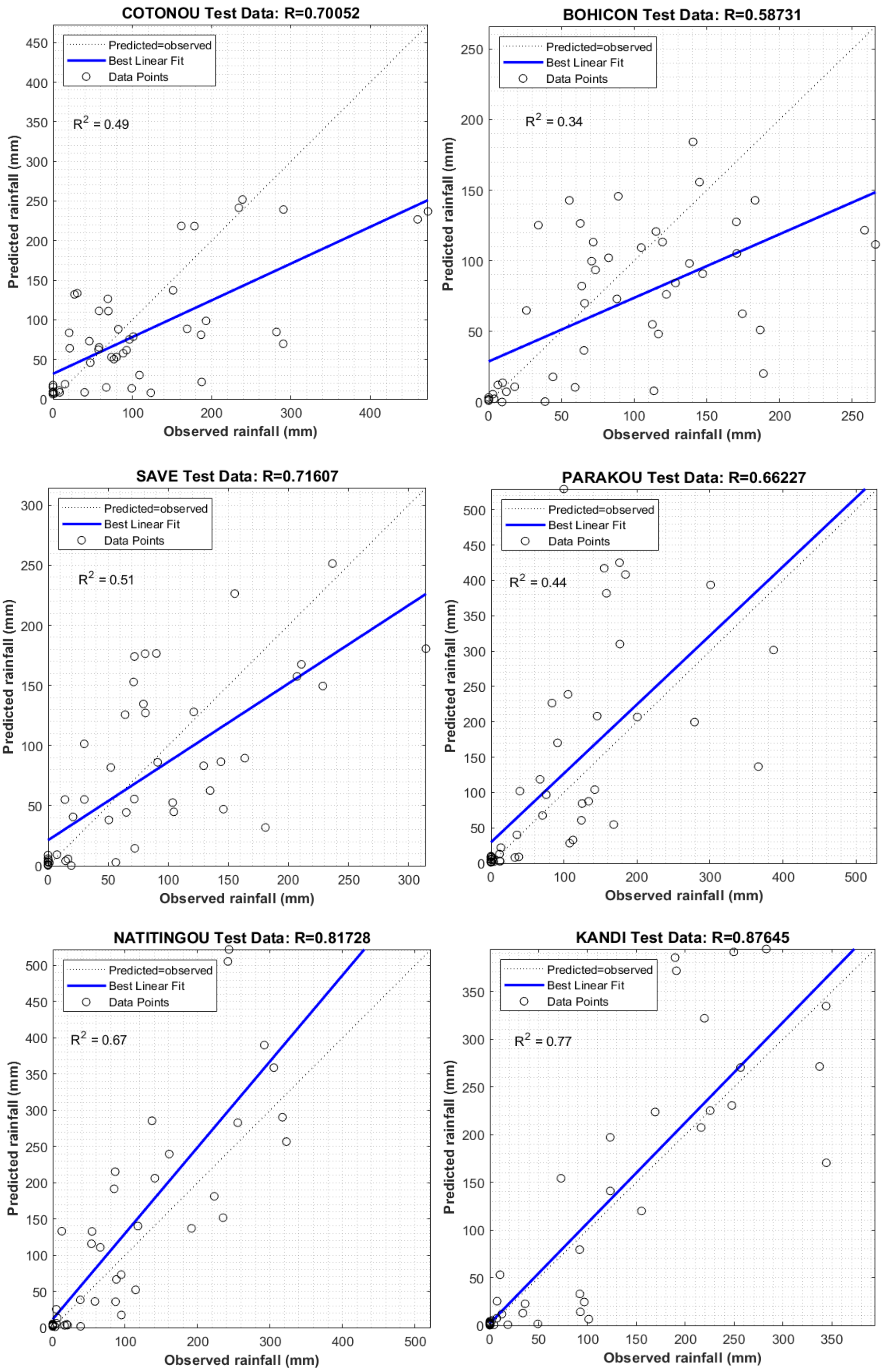
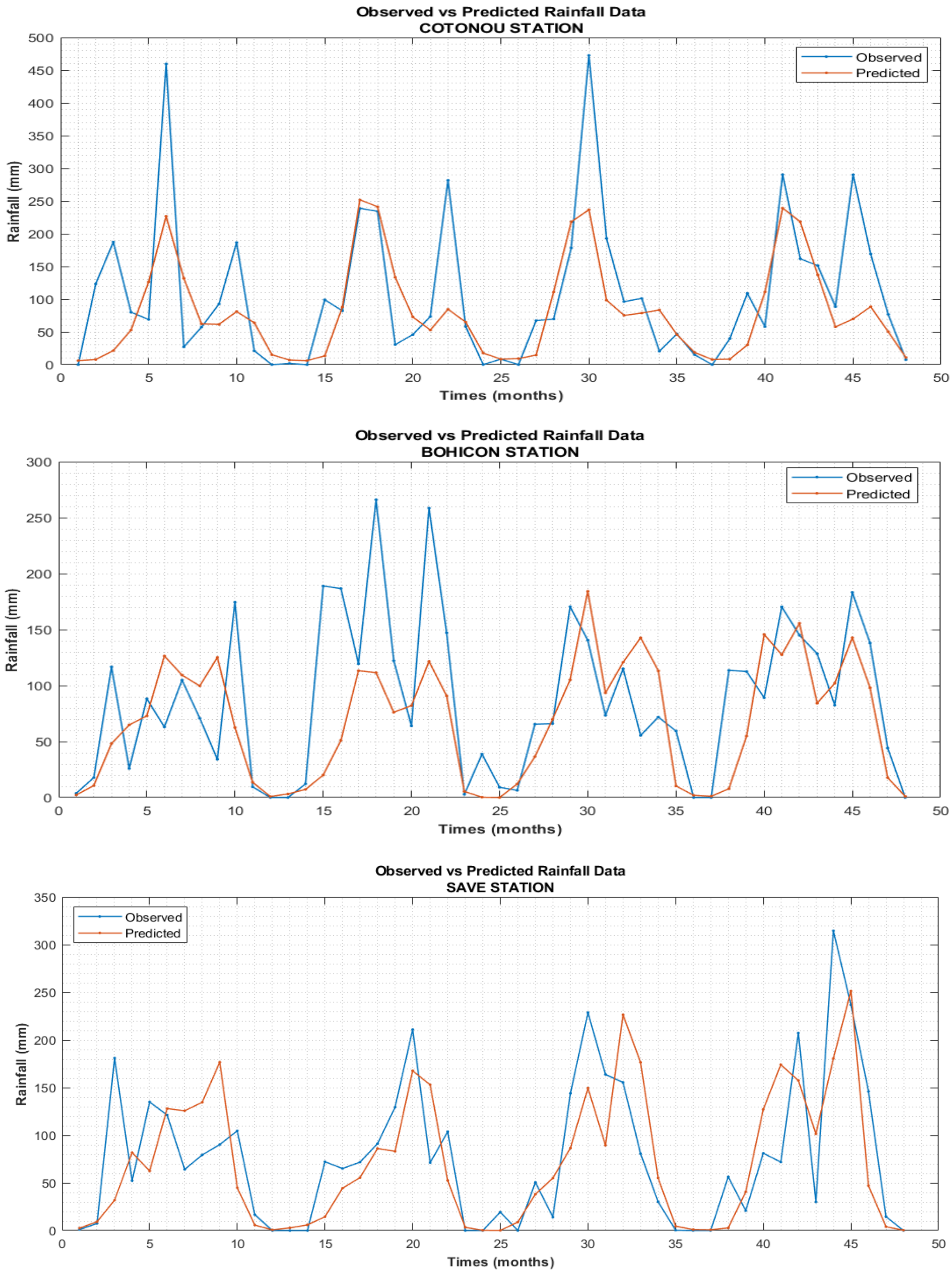



Figure 5.3: Scatter plot of actual versus MLR predicted monthly rainfall in all stations in the study area.

The graphical display of the model predictions and the observed rainfall is given Figure **5.4**.



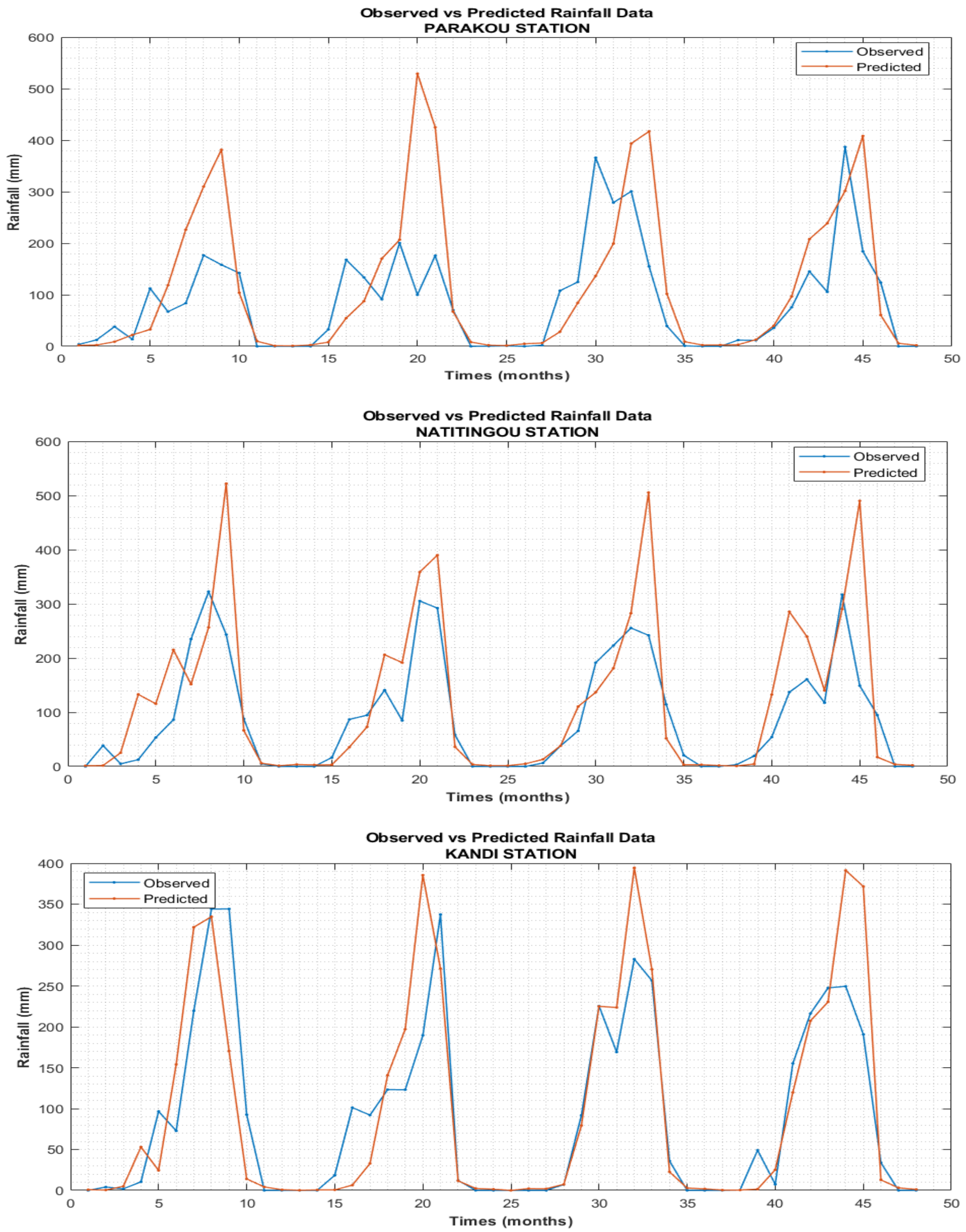



Figure 5.4: Observed rainfall vs. MLR model predictions in all stations in the study area (Period 2015-2018).

This figure shows that the MLR model predictions follow the series patterns at all locations. However, the model appeared to under and overestimate some monthly rainfall values when compared with actual monthly rainfall and some of the highest monthly rainfall values failed to be replicated.

## 5.3 Comparison of ANN vs. MLR models in predicting monthly rainfall

Table **5.6**, summarizes the performance of ANN and MLR models in predicting monthly rainfall in each location. Results presented in this table clearly demonstrate that the ANN model is superior to the MLR models for monthly rainfall predictions in all locations, regarding its higher accuracy and lower RMSE. It shows also that predictions of rainfall are affected by variations of geographical characteristics of each region with better accuracy at higher latitudes. Ewona et al. in [29] reported similar results by applying ANN to predict rainfall over 23 stations in Nigeria with more accurate rainfall predictability at higher latitudes over Nigeria. The correlation coefficients were in ascending order from south to north. Similar results were also reported by applying ANN to predict rainfall at 7 locations in Nigeria [28].

Table 5.6: The performance results of ANN and MLR models for monthly rainfall predictions.

| **Stations** | **Measures** | **ANN** | **MLR** |
|---|---|---|---|
| Cotonou | RMSE (mm) | 68.00 | 83.1 |
| | MAE (mm) | 48.30 | 55.09 |
| | $R^2$ | 0.64 | 0.49 |
| | NSE | 0.62 | 0.43 |
| Bohicon | RMSE (mm) | 51.00 | 60.23 |
| | MAE (mm) | 37.47 | 42.0 |
| | $R^2$ | 0.47 | 0.34 |
| | NSE | 0.45 | 0.24 |
| Save | RMSE (mm) | 51.00 | 55.42 |
| | MAE (mm) | 34.71 | 40.87 |
| | $R^2$ | 0.56 | 0.51 |
| | NSE | 0.55 | 0.47 |
| Parakou | RMSE (mm) | 56.45 | 111.79 |
| | MAE (mm) | 37.48 | 66.13 |
| | $R^2$ | 0.68 | 0.44 |
| | NSE | 0.67 | -0.29 |
| Natitingou | RMSE (mm) | 38.86 | 91.01 |
| | MAE (mm) | 26.90 | 54.01 |
| | $R^2$ | 0.87 | 0.67 |
| | NSE | 0.85 | 0.18 |
| Kandi | RMSE (mm) | 45.52 | 63.8 |
| | MAE (mm) | 27.06 | 37.36 |
| | $R^2$ | 0.83 | 0.77 |
| | NSE | 0.83 | 0.66 |

In summary, based on the performance measures, the ANN model is the most suitable models in finding the pattern and trends of the observations compared to MLR models at all sites.

## 5.4 Quantification of Relative Importance of Input variables

After determining the optimal architecture of ANN model, relative contribution of input variables to the output value was quantified based on Connection Weight Approach. Values of connection weights from input neurons to hidden neurons and from hidden to output neurons are presented in Table **5.7**, Table **5.8**, Table **5.9**, Table **5.10**, Table **5.11** and Table **5.12** respectively for Cotonou, Bohicon, Save, Parakou, Natitingou and Kandi station. Each weight represents the intensity of signal transfer between one neuron and another.

Table 5.7: Final connection weights ANN (7-3-1) for Cotonou.

| | **Input-hidden** | | |
|---|---|---|---|
| | Hidden1 | Hidden2 | Hidden3 |
| RHmax | 0.6029 | -0.5006 | 0.7512 |
| RHmin | 1.57 | -0.0777 | 0.7757 |
| Tmax | 0.4688 | 0.3247 | -1.0913 |
| Tmin | 1.3478 | -2.8975 | -1.8713 |
| Eva | 1.5187 | 1.7495 | 0.2605 |
| Uwind | -0.8671 | -1.0781 | -0.4233 |
| Vwind | -2.6421 | -1.6665 | -2.9855 |
| | **Hidden-output** | | |
| | Hidden1 | Hidden2 | Hidden3 |
| RR | -2.0657 | 1.2763 | -1.8153 |

Table 5.8: Final connection weights ANN (3-10-1) for Bohicon.

| | **Input-hidden** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 | Hidden10 |
| Eva | 0.4345 | 2.8197 | 2.6776 | -8.7032 | 2.8265 | 12.0448 | 3.2495 | -9.3089 | 5.5719 | -7.8413 |
| SST | 7.7395 | -7.2502 | -6.4583 | 0.4942 | -0.9139 | -22.2234 | -3.9207 | 13.0747 | -7.5138 | 0.4249 |
| Vwind | -12.3011 | -6.0863 | 8.4828 | -2.9385 | 12.4151 | 8.5419 | -2.4212 | -7.193 | -2.931 | -10.9765 |
| | **Hidden-output** | | | | | | | | | |
| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 | Hidden10 |
| RR | 1.0435 | -1.0631 | 1.4897 | 0.1421 | -0.8031 | -0.4166 | -2.5721 | 0.5119 | 1.8042 | -0.7406 |

Table 5.9: Final connection weights ANN (5-5-1) for Save.

| | | | | | |
|---|---|---|---|---|---|
| | | | **Input-hidden** | | |
| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 |
| Tmin | 1.0952 | 8.7495 | 0.3290 | -6.6732 | 3.9595 |
| Eva | 1.7204 | -0.2800 | -0.5883 | 0.6607 | -1.0629 |
| SST | 3.6210 | 1.4111 | -1.5362 | -1.4084 | 4.7639 |
| Uwind | -1.6831 | 1.3315 | -0.7823 | 0.1693 | 7.3492 |
| Vwind | -2.2873 | -0.6366 | -0.1617 | 1.1592 | -3.5536 |
| | | | **Hidden-output** | | |
| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 |
| RR | -1.6814 | -1.9431 | -3.4594 | -2.2914 | 3.8588 |

Table 5.10: Final connection weights ANN (4-9-1) for Parakou.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | **Input-hidden** | | | | |
| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 |
| Eva | -4.3537 | 4.5578 | -4.0405 | 4.4761 | 4.4087 | 5.3952 | 3.452 | -7.1901 | 0.2158 |
| SST | 2.3357 | 3.4412 | -1.6277 | -2.6704 | 7.7587 | -5.1444 | 9.0741 | -18.87 | -0.4033 |
| Uwind | -2.1363 | -23.9713 | -3.437 | 0.4607 | 1.1607 | -3.1521 | 6.6396 | 6.8375 | 4.8907 |
| Vwind | 6.7335 | 10.509 | -6.9976 | -5.7764 | -9.9541 | -5.3753 | -6.5482 | 1.8425 | 1.2095 |
| | | | | | **Hidden-output** | | | | |
| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 |
| RR | 2.8739 | -0.6933 | -1.8936 | 4.3055 | 0.1991 | -1.466 | 0.2383 | -0.6793 | 1.618 |

Table 5.11: Final connection weights ANN (5-9-1) for Natitingou.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | **Input-hidden** | | | | |
| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 |
| Tmin | -4.0191 | 3.2623 | 0.4344 | -1.3599 | -2.9425 | 1.0696 | -2.5716 | -2.2951 | -1.1065 |
| Eva | -1.8379 | -0.7948 | 0.0694 | 3.5189 | 0.7586 | 1.4816 | 2.5101 | -1.9319 | 3.4029 |
| SST | -1.9440 | -0.0803 | -1.8732 | 1.9846 | 0.3310 | -0.9419 | 0.5837 | 1.5773 | 1.2603 |
| Uwind | 2.1484 | 2.6534 | -1.2771 | -0.8562 | 1.6690 | -2.1529 | -1.4958 | 2.2028 | 1.8594 |
| Vwind | 1.8363 | 1.2262 | 1.2691 | 0.4743 | 2.8208 | -2.1023 | -2.3742 | 1.5901 | -1.7816 |
| | | | | | **Hidden-output** | | | | |
| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 |
| RR | 2.1152 | 1.5373 | -2.9561 | 0.6790 | -1.4991 | -0.8257 | 0.5222 | 0.9326 | -1.3204 |

Table 5.12: Final connection weights ANN (6-10-1) for Kandi.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Input-hidden** | | | | | | | | | | |
| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 | Hidden10 |
| RHmax | -2.8415 | 1.0823 | 0.5966 | -0.0213 | -1.6892 | 0.7677 | -1.0271 | -1.8128 | 2.7704 | -0.7132 |
| Tmax | -2.0593 | 3.4120 | -1.9469 | 2.5175 | -2.2657 | -2.0056 | -2.3609 | 1.3877 | 0.5173 | 2.1065 |
| Tmin | -2.2405 | -0.6725 | -0.2927 | 1.3556 | 1.9483 | -1.7805 | 1.4399 | 1.6513 | 1.3982 | 1.4970 |
| Eva | -1.0633 | 0.8473 | -0.1967 | -0.0805 | 2.1254 | 1.5459 | 2.4012 | 2.0798 | 0.7111 | 2.0401 |
| SST | 1.8362 | 0.4844 | 1.2893 | -2.5614 | 1.1194 | 1.4778 | -0.7186 | 0.3433 | 1.3493 | -2.2924 |
| Uwind | 0.3387 | -1.6186 | 1.9317 | -1.7639 | -0.1713 | -1.0470 | 0.5155 | -1.7319 | 1.0738 | 0.2938 |
| **Hidden-output** | | | | | | | | | | |
| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 | Hidden10 |
| RR | 1.5236 | 0.0337 | -0.2226 | -1.0623 | 0.2863 | -1.1307 | -1.1118 | 0.1370 | -0.5728 | 1.3393 |

Following the methodology described, the sum of the products of the connection weights and rank of the input variables are presented in Table **5.13**, Table **5.14**, Table **5.15**, Table **5.16**, Table **5.17** and Table **5.18** respectively for Cotonou, Bohicon, Save, Parakou, Natitingou and Kandi station.

Table 5.13: Connection weights products, relative importance and rank of inputs for Cotonou.

| | **Hidden1** | **Hidden2** | **Hidden3** | **Sum** | **Rank** |
|---|---|---|---|---|---|
| RHmax | -1.2454 | -0.6389 | -1.3637 | -3.2480 | 6 |
| RHmin | -3.2431 | -0.0992 | -1.4081 | -4.7504 | 7 |
| Tmax | -0.9684 | 0.4144 | 1.9810 | 1.4271 | 2 |
| Tmin | -2.7842 | -3.6981 | 3.3970 | -3.0853 | 5 |
| Eva | -3.1372 | 2.2329 | -0.4729 | -1.3772 | 4 |
| Uwind | 1.7912 | -1.3760 | 0.7684 | 1.1836 | 3 |
| **Vwind** | **5.4578** | **-2.1270** | **5.4196** | **8.7504** | **1** |

Table 5.14: Connection weights products, relative importance and rank of inputs for Bohicon.

| | **Hidden1** | **Hidden2** | **Hidden3** | **Hidden4** | **Hidden5** | **Hidden6** | **Hidden7** | **Hidden8** | **Hidden9** | **Hidden10** | **Sum** | **Rank** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Eva | 0.4534 | -2.9976 | 3.9888 | -1.2367 | -2.2700 | -5.0179 | -8.3580 | -4.7652 | 10.0528 | 5.8073 | -4.3431 | 3 |
| **SST** | **8.0762** | **7.7077** | **-9.6209** | **0.0702** | **0.7340** | **9.2583** | **10.0844** | **6.6929** | **-13.5564** | **-0.3147** | **19.1317** | **1** |
| Vwind | -12.8362 | 6.4703 | 12.6368 | -0.4176 | -9.9706 | -3.5586 | 6.2276 | -3.6821 | -5.2881 | 8.1292 | -2.2892 | 2 |

Table 5.15: Connection weights products, relative importance and rank of inputs for Save.

| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Sum | Rank |
|---|---|---|---|---|---|---|---|
| Tmin | -1.8414 | -17.0016 | -1.1382 | 15.2913 | 15.2790 | 10.5891 | 3 |
| Eva | -2.8925 | 0.5440 | 2.0352 | -1.5139 | -4.1015 | -5.9288 | 4 |
| SST | -6.0882 | -2.7419 | 5.3145 | 3.2273 | 18.3830 | 18.0947 | 2 |
| **Uwind** | **2.8299** | **-2.5872** | **2.7062** | **-0.3880** | **28.3590** | **30.9199** | **1** |
| Vwind | 3.8457 | 1.2370 | 0.5592 | -2.6562 | -13.7126 | -10.7269 | 5 |

Table 5.16: Connection weights products, relative importance and rank of inputs for Parakou.

| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 | Sum | Rank |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Eva | -12.5121 | -3.1599 | 7.6511 | 19.2718 | 0.8778 | -7.9094 | 0.8226 | 4.8842 | 0.3492 | 10.2753 | 3 |
| SST | 6.7126 | -2.3858 | 3.0822 | -11.4974 | 1.5448 | 7.5417 | 2.1624 | 12.8184 | -0.6525 | 19.3262 | 2 |
| **Uwind** | **-6.1395** | **16.6193** | **6.5083** | **1.9835** | **0.2311** | **4.6210** | **1.5822** | **-4.6447** | **7.9132** | **28.6744** | **1** |
| Vwind | 19.3514 | -7.2859 | 13.2507 | -24.8703 | -1.9819 | 7.8802 | -1.5604 | -1.2516 | 1.9570 | 5.4891 | 4 |

Table 5.17: Connection weights products, relative importance and rank of inputs for Natitingou.

| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 | Sum | Rank |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Tmin | -8.5014 | 5.0151 | -1.2840 | -0.9233 | 4.4110 | -0.8832 | -1.3428 | -2.1405 | 1.4610 | -4.1880 | 4 |
| Eva | -3.8875 | -1.2218 | -0.2050 | 2.3893 | -1.1372 | -1.2234 | 1.3107 | -1.8018 | -4.4930 | -10.2698 | 5 |
| SST | -4.1120 | -0.1235 | 5.5375 | 1.3475 | -0.4961 | 0.7777 | 0.3048 | 1.4710 | -1.6641 | 3.0428 | 2 |
| **Uwind** | **4.5444** | **4.0791** | **3.7753** | **-0.5813** | **-2.5019** | **1.7777** | **-0.7811** | **2.0544** | **-2.4551** | **9.9115** | **1** |
| Vwind | 3.8842 | 1.8850 | -3.7517 | 0.3220 | -4.2286 | 1.7359 | -1.2398 | 1.4830 | 2.3524 | 2.4426 | 3 |

Table 5.18: Connection weights products, relative importance and rank of inputs for Kandi.

| | Hidden1 | Hidden2 | Hidden3 | Hidden4 | Hidden5 | Hidden6 | Hidden7 | Hidden8 | Hidden9 | Hidden10 | Sum | Rank |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RHmax | -4.3291 | 0.0365 | -0.1328 | 0.0226 | -0.4837 | -0.8681 | 1.1419 | -0.2484 | -1.5868 | -0.9552 | -7.4030 | 6 |
| Tmax | -3.1374 | 0.1151 | 0.4335 | -2.6742 | -0.6488 | 2.2678 | 2.6248 | 0.1901 | -0.2963 | 2.8213 | 1.6960 | 2 |
| Tmin | -3.4135 | -0.0227 | 0.0652 | -1.4400 | 0.5579 | 2.0132 | -1.6008 | 0.2263 | -0.8008 | 2.0049 | -2.4104 | 4 |
| Eva | -1.6201 | 0.0286 | 0.0438 | 0.0855 | 0.6086 | -1.7480 | -2.6696 | 0.2850 | -0.4073 | 2.7324 | -2.6611 | 5 |
| SST | 2.7976 | 0.0163 | -0.2870 | 2.7209 | 0.3205 | -1.6710 | 0.7990 | 0.0470 | -0.7728 | -3.0703 | 0.9002 | 3 |
| **Uwind** | **0.5161** | **-0.0546** | **-0.4301** | **1.8737** | **-0.0491** | **1.1839** | **-0.5732** | **-0.2373** | **-0.6150** | **0.3935** | **2.0080** | **1** |

The results obtained indicating the relative importance of various input variables on the output of the network (rainfall), are presented in Figure **5.5**. Based on Connection Weight Approach, the most influential climate variable, among other variables, on rainfall is Vwind for Cotonou SST for Bohicon and Uwind for Save, Parakou, Natitingou and Kandi. This goes along with the recommendation of Ogou et al. [134] which is to consider SST, Uwind and Vwind at 850 hPa when analyzing rainfall over Republic of Benin. Small variation in those parameters can drastically change the amount of the rainfall and vice versa. The sign (positive or negative values) helps indicate the direction in which each input affects the output. The positive sign indicates the likelihood that increasing this input variable will increase the output parameter, while the negative sign indicates the possibility that increasing this input variable will decrease the output variable.

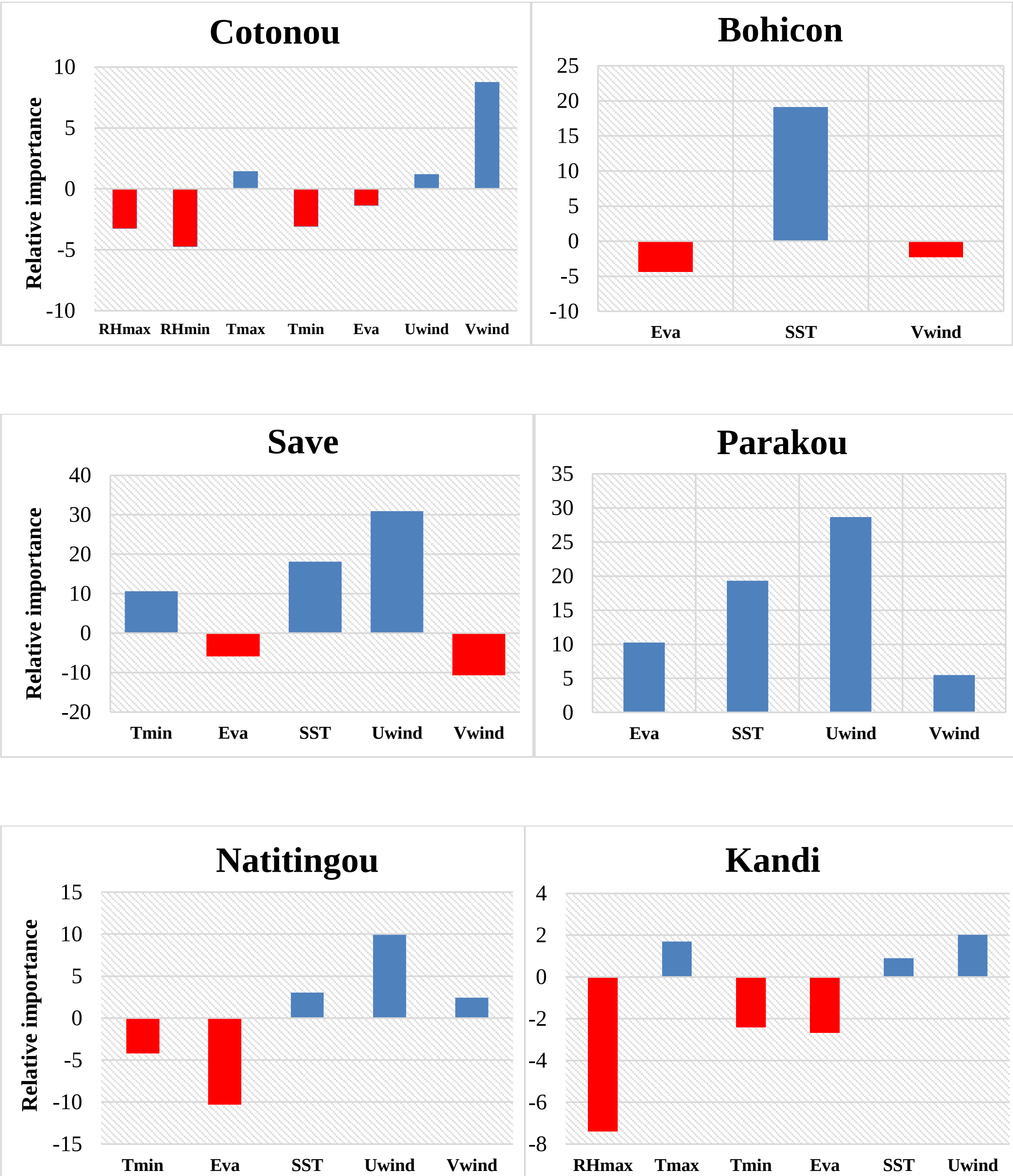



Figure 5.5: Relative importance of each input variables in the ANN models.

## 5.5 Summary

This chapter reports on the results of the proposed ANN and MLR for monthly rainfall predictions over four years in the Republic of Benin. Considering the different climate zones, six geographically diverse weather stations were used: two each from Guinean, Sudano-Guinean and Sudanian zones. The purpose of choosing these locations was to investigate the performance of the prediction models by zones, which experience different climate and hydrological regimes. All selected prediction models were developed for each weather station using training sets and evaluated using the test sets. The prediction performance of the models was evaluated by comparing observed and predicted rainfall using performance measures such as RMSE, MAE, $R^2$ and NSE. The computational results of the proposed model were compared with that obtained from MLR model.

The results reported show that in all climate zones the ANN model is most suitable for monthly rainfall predictions when compared to MLR model. Results demonstrate that the ANN model is an efficient method for monthly rainfall predictions and is superior to MLR model in all locations used in this study.

The prediction performance of all models varied with changes in geographic regions. Predictions have the lowest deviation from the actual observations in Sudanian climate zone while having the highest in Guinean zone. This may be due to the increased instability and extremes of rainfall in the Guinean zone of the study site. The results presented suggest that the 8 meteorological parameters considered are suitable predictors for monthly rainfall predictions in Benin Republic.

In order to determine the amount of influence of each input meteorological parameters, the relative importance of these variables have been determined based on weights between produced artificial neurons in the network. This process is done using the Connection Weight Approach (Eq. (**4.11**))

It was found that SST, Uwind and Vwind are the most important variables that contribute more on rainfall.

# Chapter 6

# CONCLUSION AND PERSPECTIVES

## 6.1 Conclusion

Rainfall plays a vital role in deciding the weather conditions and is a deciding factor for natural disasters such as floods, drought etc. Agricultural sectors can avail the benefit of knowing weather conditions in advance and take precautionary steps accordingly. This directly helps in the improvement of the national economy as well. The aim of this thesis is to develop a non-linear ANN model to improve the accuracy of rainfall prediction in Benin. This study also investigated the influence of geographic regions on the performance of models in predicting monthly rainfall in Benin and the contribution of each input variables on total rainfall. Accurate rainfall prediction is a serious concern in many countries, especially in Benin, where the climate is variable. Any change in the probability of rainfall (heavy rainfall or drought) has important implications for future resource planning, management and investment. Predicting rainfall is a complex process, needing continual improvement. In this thesis, non-linear multivariate ANN models have been developed for monthly rainfall predictions. For the development of the model, the best predictors with their optimal lag times were determined through the cross-correlation analysis between the input and output variables and the correlation-based feature selection. The prediction performance of the proposed ANN model with multivariate MLR was assessed using root mean squared error, mean absolute error, the coefficient of determination and Nash-Sutcliffe efficiency as performance metrics. The findings of this thesis were as follows:

- These eight meteorological variables are suitable predictors for monthly rainfall predictions in Benin Republic.
- The non-linear ANN is superior to the MLR model in all locations used in this study. MLR is not accurate model for rainfall prediction.
- The prediction performance of all models varies considerably both within and across climate zones. Predictions have the lowest deviation from the actual observations in sudanian climate zone and the highest in guinea zone.

- SST, Uwind and Vwind are the most important variables that contribute more on rainfall.

Findings confirm that data-driven modeling is an effective technique for rainfall prediction. Nonlinear models are the best appropriate for such predictions. The findings from the analysis of the ANN models corroborate this conclusion. Additionally, the results demonstrate that the ANN model is an effective tool for rainfall prediction, particularly for severe rainfall prediction. The results obtained in this research are useful to understand the relationship between rainfall and meteorological variables in different climate zones in the Benin Republic and may lead to further development in the modeling and predictions of monthly rainfall. In conclusion, this study demonstrated the feasibility of predicting monthly rainfall in advance for the study locations using ANNs and lagged meteorological variables.

## 6.2 Research Contributions

This thesis contributed to the understanding of monthly rainfall forecasting in the Benin Republic. This was done through the development of a new monthly forecasting model. A large contribution of this work is the quantitative prediction of monthly rainfall. This kind of forecasting allows policy-makers and local populations to better plan their activities and to act appropriately ahead of time in order to mitigate impacts. Another contribution of this research is that it helps in understanding that nonlinear models can also be used instead of the usual linear methods. The findings of this study will be used to:

- predict rainfall pattern for other areas in Benin that is not covered in this study.
- provide information that would be useful to decision-makers in formulating policies to mitigate the challenges of managing rainwater supplies, soil degradation, drought and flood.
- provide information for the early warning system in each study area.

## 6.3 Limitation and Future Work

Among the existing parameters that influence the accuracy of rainfall prediction, only 8 are considered for the present study, namely minimum relative humidity, maximum relative humidity, minimum air temperature, maximum air temperature, evaporation, Sea Surface

Temperature, zonal and meridional wind at 850 hPa. Also many areas within the study region aren't considered due to the limitations of the synoptic stations available.

From the monthly rainfall forecasting results, an ANN approach to forecasting is an interesting solution to climate forecasting in the Benin Republic. Regarding the obtained results from ANN models, there is a need to continue their development by adopting specific training techniques in order to achieve better forecasts. Future work will embrace many more variables to test the ANNs model to forecast the rainfall and extend the study to a larger scale (e.g., Guinea coast countries, West Africa, African continent). Such models should be able, in particular, to predict extreme rainfall events, which are a challenge for all existing models.

# Bibliography


[1] A. Asfaw, B. Simane, A. Hassen, and A. Bantider, "Variability and time series trend analysis of rainfall and temperature in northcentral Ethiopia : A case study in Woleka sub-basin," *Weather and Climate Extremes*, vol. 19, pp. 29–41, 2018, doi: 10.1016/j.wace.2017.12.002.

[2] UNFCCC, "Climate Change: Impacts, Vulnerabilities and Adaptation in Developing Countries.," 2007.

[3] IPCC, *Bilan 2007 des changements climatiques. Contribution des Groupes de travail I, II et III au quatrième Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du clima*. Genève, Suisse, 2007.

[4] C. Mcsweeney, M. New, and G. Lizcano, "UNDP Climate Change Country Profiles: Benin," *UNDP Climate Change Country Profiles: Benin, text, January 2006; 304 East 45th Street, Fl. 9 New York, NY 10017.*

[5] MEHU, "Deuxieme communication nationale de la republique du benin sur les changements climatiques," 2011.

[6] *FAO; ICRISAT; CIAT. 2018. Climate-Smart Agriculture in Benin. CSA Country Profiles for Africa Series. International Center for Tropical Agriculture (CIAT); International Crops Research Institute for the Semi-Arid Tropics (ICRISAT); Food and Agriculture Or*. Rome, Italy. 22p.

[7] IPCC, "Summary for policymakers. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B.," pp. 1–32, 2014.

[8] J. B. Omotosho, A. A. Balogun, and K. Ogunjobi, "Predicting monthly and seasonal rainfall, onset and cessation of the rainy season in West Africa using only surface data.," *Int J Climatol*, vol. 20, pp. 865–880, 2000.

[9] C. Jaurès, K. Asomanin, and A. Mensah-bonsu, "Contingent valuation study of the benefits of seasonal climate forecasts for maize farmers in the Republic of Benin , West Africa," vol. 6, pp. 1–11, 2017, doi: 10.1016/j.cliser.2017.06.007.

[10] D. G. Savakar and R. C., "Rainfall Prediction Based On Rainfall Statistical Data," *Int. J.*

*Recent Innov. Trends Comput. Commun*, vol. 4, no. 3, pp. 270–275, 2016.

[11] M. Hamatan, “Synthèse et Evaluation des Prévisions Saisonnières en Afrique de l’Ouest DEA Sciences de l‘Eau dans l‘Environnement Continental. Université Montpellier 2, 2002, 115pp.”

[12] J. R. Holton and G. J. Hakim, “An introduction to dynamic meteorology,” *Academic press*, vol. 88, 2012.

[13] N. Khalili, S. R. Khodashenas, K. Davary, and F. Karimaldini, “Daily Rainfall Forecasting for Mashhad Synoptic Station using Artificial Neural Networks,” *2011 International Conference on Environmental and Computer Science*, vol. 19, pp. 118–123, 2011.

[14] A. Ajith, S. N. Philip, and J. K. Babu, “Will we have a wet summer? Soft computing models for long-term rainfall forecasting,” *Modeling and Simulation*, pp. 1044–1048, 2001.

[15] M. Boko, F. Kosmowski, and E. W. Vissin, “Les Enjeux du Changement Climatique au Bénin,” *Konrad-Adenauer-Stiftung, Programme pour le Dialogue Politique en Afrique de l’Ouest*, p. 71, 2012.

[16] “World Bank. 2017. World Development Indicators 2017. Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/26447 License: CC BY 3.0 IGO.” http://hdl.handle.net/10986/26447 (accessed May 14, 2021).

[17] P. Lynch, “The origins of computer weather prediction and climate modeling,” *Journal of Computational Physics*, vol. 227, pp. 3431–3444, 2008, doi: 10.1016/j.jcp.2007.02.034.

[18] D. R. Nayak, A. Mahapatra, and P. Mishra, “A Survey on Rainfall Prediction using Artificial Neural Network,” *Int. J. Comput. Appl*, vol. 72, no. 16, pp. 32–40, 2013.

[19] R. Kimura, “Numerical weather prediction,” vol. 90, pp. 1403–1414, 2002.

[20] G. Shrivastava, “Application of Artificial Neural Networks in Weather Forecasting : A Comprehensive Literature Review,” vol. 51, no. 18, pp. 17–29, 2012.

[21] M. C. V. Ramirez, H. Fraga de Campos Velho, and N. Jesus Ferreira, “Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region,” *Journal of hydrology*, vol. 301, no. 1, pp. 146–162, 2005, doi: 10.1016/j.jhydrol.2004.06.028.

[22] M. Richard and K. G. Rao, “Artificial neural networks in temporal and spatial variability studies and prediction of rainfall,” *ISH J. Hydraul. Eng.*, vol. 20, no. 1, pp. 1–6, 2014, doi: 10.1080/09715010.2013.806400.

[23] U. Yolcu, E. Egrioglu, and C. H. Aladag, “A new linear & nonlinear artificial neural network model for time series forecasting,” *Decision Support Systems*, vol. 54, no. 3, pp. 1340–1347, 2013, doi: 10.1016/j.dss.2012.12.006.

[24] A. Jain K., J. Mao, and K. M. Mohiuddin, “Artificial Neural Network: A Tutorial,” *Communications*, vol. 29, pp. 31–44, 1996.

[25] J. Adamowski and K. Sun, “Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds,” *Journal of Hydrology*, vol. 390, no. 1–2, pp. 85–91, 2010, doi: 10.1016/j.jhydrol.2010.06.033.

[26] M. H. Gholizadeh and M. Darand, “Forecasting precipitation with artificial neural networks (case study: Tehran),” *Journal of Applied Sciences*, vol. 9, no. 9, pp. 1786–1790, 2009, doi: 10.3923/jas.2009.1786.1790.

[27] A. P. Ayodele and E. E. Precious, “Seasonal Rainfall Prediction in Lagos, Nigeria Using Artificial Neural Network,” *Asian Journal of Research in Computer Science*, vol. 3, no. 4, pp. 1–10, 2019, doi: 10.9734/ajrcos/2019/v3i430100.

[28] T. S. Abdulkadir, A. . Salami, A. . Aremu, A. . Ayanshola, and D. . Oyejobi, “Assessment of Neural Networks Performance in Modeling Rainfall,” *Journal of Research in Forestry, Wildlife & Environment*, vol. 9, no. 1, pp. 12–22, 2017.

[29] I. O. Ewona, J. E. Osang, U. I. Uquetan, E. O. Inah, and S. O. Udo, “Rainfall prediction in Nigeria using Artificial Neural Networks,” *International Journal of Scientific & Engineering Research*, vol. 7, no. 1, pp. 1157–1169, 2016.

[30] J. O. Folorrunsho, “Application of artificial neural network model in forecasting Zaria rainfall,” 2014.

[31] T. Kashiwao, K. Nakayama, S. Ando, K. Ikeda, and M. Lee, “A neural network-based local rainfall prediction system using meteorological data on the Internet : A case study using data from the Japan Meteorological Agency,” *Applied Soft Computing Journal*, vol. 56, pp. 317–330, 2017, doi: 10.1016/j.asoc.2017.03.015.

[32] Mislan, Haviluddin, S. Hardwinarto, Sumaryono, and M. Aipassa, “Rainfall Monthly

Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan - Indonesia," *Procedia Computer Science*, vol. 59, no. Iccsci, pp. 142–151, 2015, doi: 10.1016/j.procs.2015.07.528.

[33] E. Vamsidhar, K. Varma, P. S. Rao, and R. Satapati, "Prediction of Rainfall Using Backpropagation Neural Network Model," *International Journal on Computer Science and Engineering*, vol. 02, no. 04, pp. 1119–1121, 2010.

[34] K. Abhishek, A. Kumar, R. Ranjan, and S. Kumar, "A Rainfall Prediction Model using Artificial Neural Network," *2012 IEEE Control and System Graduate Research Colloquium, ICSGRC 2012*, pp. 82–87, 2012, doi: 10.1109/ICSGRC.2012.6287140.

[35] A. El-Shafi, A. Noureldin, M. Taha, A. Hussain, and M. Mukhlisin, "Dynamic versus static neural network model for rainfall forecasting at Klang River Basin , Malaysia," *Hydrology and Earth System Sciences*, vol. 16, pp. 1151–1169, 2012, doi: 10.5194/hess-16-1151-2012.

[36] G. Geetha and R. S. Selvaraj, "Prediction Of Monthly Rainfall In Chennai Using Back Propagation Neural Network Model," *International Journal of Engineering Science and Technology*, vol. 3, no. 1, pp. 211–213, 2011.

[37] H. Aksoy and A. Dahamsheh, "Artificial neural network models for forecasting monthly precipitation in Jordan," *Stochastic Environmental Research and Risk Assessment*, vol. 23, no. 7, pp. 917–931, 2009, doi: 10.1007/s00477-008-0267-x.

[38] D. F. Lekkas, C. Onof, M. J. Lee, and E. A. Baltas, "Application of artificial neural networks for flood," vol. 6, no. 3, pp. 205–211, 2004.

[39] S. Chattopadhyay and G. Chattopadhyay, "Comparative study among different neural net learning," vol. 280, no. April, pp. 273–280, 2008, doi: 10.1002/met.

[40] Y. C. Patil and A. A. Ghatol, "Rainfall Forecasting Using Local Parameters Over A Meteorological Station: An Artificial Neural Network Approach," *International Journal of Engineering Research & Industrial Applications*, vol. 3, no. II, pp. 341–356, 2010.

[41] A. R. Naik and P. S. K. Pathan, "Weather Classification and Forecasting using Back Propagation Feed-forward Neural Network," *International Journal of Scientific and Research Publications*, vol. 2, no. 12, pp. 1–3, 2012.

[42] R. C. Deo and M. Sahin, "Application of the artificial neural network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using

hydrometeorological parameters and climate indices in Eastern Australia," *Atmospheric Research*, 2015, doi: 10.1016/j.atmosres.2015.03.018.

[43] J. Abbot and J. Marohasy, "Input selection and optimisation for monthly rainfall forecasting in Queensland , Australia , using artificial neural networks," *Atmospheric Research*, vol. 138, pp. 166–178, 2014, doi: 10.1016/j.atmosres.2013.11.002.

[44] S. A. Muttaleb Alhashimi, "Prediction Of Monthly Rainfall In Kirkuk Using Artificial Neural Network And Time Series Models," *Journal of Engineering and Sustainable Development*, vol. 18, no. 1, pp. 129–143, 2014.

[45] L. P. Maya, V. P. Kalavampara, and N. B. Alungal, "Long Range Forecast on South West Monsoon Rainfall using Artificial Neural Networks based on Clustering Approach," *International Journal of Information Technology and Computer Science(IJITCS)*, vol. 6, no. 7, pp. 1–8, 2014, doi: 10.5815/ijitcs.2014.07.01.

[46] D. N. Kumar, M. J. Reddy, and R. Maity, "Regional Rainfall Forecasting using Large Scale Climate Teleconnections and Artificial Intelligence Techniques," *Journal of Intelligent Systems*, vol. 16, pp. 307–322, 2007.

[47] J. Farajzadeh, A. Fakheri, and S. Lot, "Modeling of monthly rainfall and runoff of Urmia lake basin using feed-forward neural network and time series analysis model," *Water Resources and Industry*, vol. 8, pp. 38–48, 2014, doi: 10.1016/j.wri.2014.10.003.

[48] M. Karamouz, S. Razavi, and S. Araghinejad, "Long-lead seasonal rainfall forecasting using time-delay recurrent neural networks : a case study," *Hydrological Processes*, vol. 22, pp. 229–241, 2008, doi: 10.1002/hyp.

[49] K. V.Somvanshi, P. O.Pandey, K. P. Agrawal, N.V.Kalanker1, M. R. Prakash, and C. Ramesh, "Modelling and prediction of rainfall using artificial neural network and ARIMA techniques," *J. Ind. Geophys. Union*, vol. 10, no. 2, pp. 141–151, 2006.

[50] J. Joshi and V. M. Patel, "Rainfall-Runoff Modeling Using Artificial Neural Network ( A Literature Review )," *National Conference on Recent Trends in Engineering & Technology*, 2011.

[51] K. Solaimani, "Rainfall-runoff Prediction Based on Artificial Neural Network ( A Case Study : Jarahi Watershed )," vol. 5, no. 6, pp. 856–865, 2009.

[52] F. Chang, J. Liang, and Y. Chen, "Flood forecasting using radial basis function neural networks," *Systems, Man, and Cybernetics, Part C: Applications and Reviews,IEEE*

*Transactions on*, vol. 31, no. 4, pp. 530–535, 2001.

[53] M. Hayati and Z. Mohebi, "Application of Artificial Neural Networks for Temperature Forecasting," *World Academy of Science, Engineering and Technology*, vol. 28, pp. 275–279, 2007.

[54] I. Maqsood, M. Riaz, G. H. Huang, and R. Abdalla, "Application of soft computing models to hourly weather analysis in," *Engineering Applications of Artificial Intelligence*, vol. 18, pp. 115–125, 2005, doi: 10.1016/j.engappai.2004.08.019.

[55] I. Maqsood, M. R. Khan, and A. Abraham, "Weather Forecasting Models Using Ensembles of Neural Networks," *Springer*, pp. 33–42, 2003.

[56] B. Tang, W. W. Hsieh, and K. Haines, "A neural network atmospheric model for hybrid coupled modelling," *Climate dynamics*, vol. 17, pp. 445–455, 2001.

[57] K. C. Luk, J. E. Ball, and A. Sharma, "An Application of Artificial Neural Networks for Rainfall Forecasting," *Mathematical and Computer modelling*, vol. 33, pp. 683–693, 2001.

[58] J. Lee and J.-H. Lee, "Constructing Efficient Regional Hazardous Weather Prediction Models through Big Data Analysis," *The International Journal of Fuzzy Logic and Intelligent Systems*, vol. 16, no. 1, pp. 1–12, 2016, doi: 10.5391/ijfis.2016.16.1.1.

[59] A. M. Bagirov and A. Mahmood, "A Comparative Assessment of Models to Predict Monthly Rainfall in Australia," *Water Resources Management*, vol. 32, no. 5, pp. 1777–1794, 2018, doi: 10.1007/s11269-018-1903-y.

[60] L. Gwo Fong and J. Bing Chen, "A real-time forecasting model for the spatial distribution of typhoon rainfall," *Journal of Hydrology*, vol. 521, pp. 302–313, 2015, doi: 10.1016/j.jhydrol.2014.12.009.

[61] L. Gwo Fong, J. Bing Chen, and C. Chia Chuan, "Development of an effective data-driven model for hourly typhoon rainfall forecasting," *Journal of Hydrology*, vol. 495, pp. 52–63, 2013, doi: 10.1016/j.jhydrol.2013.04.050.

[62] N. Munir Ahmad and G. Subimal, "Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier," *Theoretical and Applied Climatology*, vol. 114, no. 3–4, pp. 583–603, 2013, doi: 10.1007/s00704-013-0867-3.

[63] M. Andrew, D. Jamie, and Z. Song, "Warm-season thermodynamically-driven rainfall

prediction with support vector machines," *Procedia Computer Science*, vol. 20, pp. 128–133, 2013, doi: 10.1016/j.procs.2013.09.250.

[64] K. Ozgur and C. Mesut, "Precipitation forecasting by using wavelet-support vector machine conjunction model," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 4, pp. 783–792, 2012, doi: 10.1016/j.engappai.2011.11.003.

[65] L. Gwo Fong, C. Guo Rong, W. Ming Chang, and C. Yang Ching, "Effective forecasting of hourly typhoon rainfall using support vector machines," *Water Resources Research*, vol. 45, no. 8, pp. 1–11, 2009, doi: 10.1029/2009WR007911.

[66] A. Eskandarin, H. Nazarpour, M. Teimouri, and M. Z. Ahmadi, "Comparison of Neural Network and K-Nearest Neighbor Methods in Daily Flow Forecasting," *Journal of Applied Sciences*, vol. 10, no. 11. pp. 1006–1010, 2010, doi: 10.3923/jas.2010.1006.1010.

[67] A. Brath, A. Montanari, and E. Toth, "Neural networks and non-parametric methods for improving realtime flood forecasting through conceptual hydrological models," *Hydrology and Earth System Sciences*, vol. 6, no. 4, pp. 627–640, 2002, doi: 10.5194/hess-6-627-2002.

[68] E. Toth, A. Brath, and A. Montanari, "Comparison of short-term rainfall prediction models for real-time flood forecasting," *Journal of Hydrology*, vol. 239, no. 1–4, pp. 132–147, 2000, doi: 10.1016/S0022-1694(00)00344-9.

[69] A. Y. Shamseldin and K. M. O'Connor, "A nearest neighbour linear perturbation model for river flow forecasting," *Journal of Hydrology*, vol. 179, no. 1–4, pp. 353–375, 1996, doi: 10.1016/0022-1694(95)02833-1.

[70] G. Giorgio, "A comparison of parametric and non-parametric methods for runoff forecasting," *Hydrological Sciences Journal*, vol. 35, no. 1, pp. 79–94, 1990, doi: 10.1080/02626669009492406.

[71] Z. M. Yaseen *et al.*, "Rainfall Pattern Forecasting Using Novel Hybrid Intelligent Model Based ANFIS-FFA," *Water Resources Management*, vol. 32, no. 1, pp. 105–122, 2018, doi: 10.1007/s11269-017-1797-0.

[72] F. Mekanik, M. A. Imteaz, and A. Talei, "Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals," *Climate Dynamics*, vol. 46, no. 9–10, pp. 3097–3111, 2016, doi: 10.1007/s00382-015-

2755-2.

[73] M. A. Sojitra and P. A. Pandya, “Comparative Study of Daily Rainfall Forecasting Models Using Adaptive-Neuro Fuzzy Inference System (ANFIS),” *Current World Environment*, vol. 10, no. 2, pp. 529–536, 2015, doi: 10.12944/cwe.10.2.19.

[74] Z. M. Yaseen, A. El-shafie, O. Jaafar, H. A. Afan, and K. N. Sayl, “Artificial intelligence based models for stream-flow forecasting: 2000-2015,” *Journal of Hydrology*, vol. 530, pp. 829–844, 2015, doi: 10.1016/j.jhydrol.2015.10.038.

[75] F. Fahimi, Z. M. Yaseen, and A. El-shafie, “Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review,” *Theoretical and Applied Climatology*, vol. 128, no. 3–4, pp. 875–903, 2017, doi: 10.1007/s00704-016-1735-8.

[76] V. Nourani, A. Hosseini Baghanam, J. Adamowski, and O. Kisi, “Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review,” *Journal of Hydrology*, vol. 514, pp. 358–377, 2014, doi: 10.1016/j.jhydrol.2014.03.057.

[77] A. Bello and M. Mamman, “Monthly rainfall prediction using artificial neural network: A case study of Kano, Nigeria,” *Environmental and Earth Sciences Research Journal*, vol. 5, no. 2, pp. 37–41, 2018, doi: 10.18280/eesrj.050201.

[78] N. Sethi and K. Garg, “Exploiting Data Mining Technique for Rainfall Prediction,” *International Journal of Computer Science and Information Technologies*, vol. 5, no. 3, pp. 3982–3984, 2014.

[79] F. Mekanik, M. A. Imteaz, S. Gato-trinidad, and A. Elmahdi, “Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes,” *Journal Of Hydrology*, vol. 503, pp. 11–21, 2013, doi: 10.1016/j.jhydrol.2013.08.035.

[80] A. El-Shafie, M. Mukhlisin, A. A. Najah, and M. R. Taha, “Performance of artificial neural network and regression techniques for rainfall-runoff prediction,” *International Journal of the Physical Sciences*, vol. 6, pp. 1997–2003, 2011, doi: 10.5897/IJPS11.314.

[81] R. P. Shukla, K. C. Tripathi, A. C. Pandey, and I. M. L. Das, “Prediction of Indian summer monsoon rainfall using Niño indices : A neural network approach,” *Atmospheric Research*, vol. 102, no. 1–2, pp. 99–109, 2011, doi: 10.1016/j.atmosres.2011.06.013.

[82] F. Rossel and E. Cadier, “El Niño and prediction of anomalous monthly rainfalls in

Ecuador," *Hydrological Processes*, vol. 23, pp. 3253–3260, 2009, doi: 10.1002/hyp.7401.

[83] Y. Sadhuram and T. V. Ramana Murthy, "Simple multiple regression model for long range forecasting of Indian summer monsoon rainfall," *Meteorology and Atmospheric Physics*, vol. 99, no. 1–2, pp. 17–24, 2008, doi: 10.1007/s00703-007-0277-0.

[84] C. Ihara, Y. Kushnir, M. A. Cane, and V. H. De La Pena, "Indian summer monsoon rainfall and its link with ENSO and Indian Ocean climate indices," *International Journal of Climatology*, vol. 27, pp. 179–187, 2007, doi: 10.1002/joc.1394.

[85] S. Chattopadhyay, "Feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India," *Acta Geophys*, vol. 55, no. 3, pp. 369–382, 2007, doi: 10.2478/s11600-007-0020-8.

[86] N. Singhratina, B. Rajagop, M. Clark, and K. Kumar, "Seasonal forecasting of Thailand summer monsoon," *International Journal of Climatology*, vol. 25, pp. 649–664, 2005, doi: 10.1002/joc.1144.

[87] S. Lee and P. M. Wong, "Rainfall Prediction Using Artificial Neural Networks," *Journal of Geographic Information and Decision Analysis*, vol. 2, no. 2, pp. 233–242, 1998.

[88] A. Y. Shamseldin, "Application of a neural network technique to rainfall-runoff modelling," *Journal of Hydrology*, vol. 199, pp. 272–294, 1997.

[89] A. Hounkpèvi *et al.*, "Climate and potential habitat suitability for cultivation and in situ conservation of the black plum (Vitex doniana Sweet) in Benin, West Africa," *Int. J. Agri. Agri. R.*, vol. 8, no. 4, pp. 67–80, 2016.

[90] B. Fandohan, A. E. Assogbadjo, R. L. Gle, and B. Sinsin, "Effectiveness of a protected areas network in the conservation of Tamarindus indica ( Leguminosea – Caesalpinioideae ) in Benin," *African Journal of Ecology*, vol. 49, pp. 40–50, 2010.

[91] C. P. Gnanglè, R. G. Kakaï, A. E. Assogbadjo, S. Vodounnon, J. Afouda, and N. Sokpon, "Tendances Climatiques Passées, Modélisation, Perceptions Et Adaptations Locales Au Benin," *Climatologie*, vol. 8, pp. 27–40, 2011.

[92] A. E. Assogbadjo, R. G. Kakaï, F. G. Vodouhê, C. A. M. S. Djagoun, J. T. C. Codjia, and B. Sinsin, "Forest Policy and Economics Biodiversity and socioeconomic factors supporting farmers ' choice of wild edible trees in the agroforestry systems of Benin ( West Africa )," *Forest Policy and Economics*, vol. 14, no. 1, pp. 41–49, 2012, doi:

10.1016/j.forpol.2011.07.013.

[93] MCVDD, “Troisième communication nationale du Bénin à la Convention-cadre des Nations Unies sur les Changements Climatiques,” Cotonou, 2019.

[94] H. Hersbach *et al.*, “ERA5 monthly averaged data on pressure levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS).,” 2019. .

[95] T. M. Smith, R. W. Reynolds, P. Thomas C., and L. Jay, “Improvements to NOAA’s Historical Merged Land-Ocean Surface Temperature Analysis (1880-2006),” *Journal of Climate*, vol. 21, pp. 2283–2296, 2008.

[96] J. Abbot and J. Marohasy, “Application of Artificial Neural Networks to Rainfall Forecasting in Queensland , Australia,” vol. 29, no. 4, pp. 717–730, 2012, doi: 10.1007/s00376-012-1259-9.1.1.1.

[97] J. Abbot and J. Marohasy, “Application of artificial neural networks to forecasting monthly rainfall one year in advance for locations within the Murray Darling basin,” *Int. J. Sustain. Dev. Plan*, vol. 12, no. 8, pp. 1282–1298, 2017, doi: 10.2495/SDP-V12-N8-1282-1298.

[98] H. S. Badr, B. F. Zaitchik, and S. . Guikema, “Application of Statistical Models to the Prediction of Seasonal Rainfall Anomalies over the Sahel,” *J. Appl. Meteorol. Climatol*, vol. 53, pp. 614–637, 2013, doi: 10.1175/JAMC-D-13-0181.1.

[99] H. M. Rasel, M. A. Imteaz, I. Hossain, and F. Mekanik, “Comparative study between linear and non-linear modelling techniques in Rainfall Forecasting for South Australia,” *21st International Congress on Modelling and Simulation*, pp. 2012–2018, 2015.

[100] D. J. Peres, C. Iuppa, L. Cavallaro, A. Cancelliere, and E. Foti, “Significant wave height record extension by neural networks and reanalysis wind data,” *Ocean Modelling*, vol. 94, pp. 128–140, 2015, doi: 10.1016/j.ocemod.2015.08.002.

[101] M. A. Hall, “Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning,” *Proc. 17th Int’l Conf. Machine Learning*, pp. 359–366, 2000.

[102] M. A. Hall and L. A. Smith, “Feature Subset Selection : A Correlation Based Filter Approach,” pp. 1–4, 1998.

[103] R. Kohavi and G. H. J. B, “Wrappers for Feature Subset Selection,” *Artificial Intelligence*, vol. 97, pp. 273–324, 1997, doi: 10.1007/978-3-642-39038-8_27.

[104] L. Huan and Y. Lei, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, 2005, doi: 10.1117/12.942023.

[105] Paras, S. Mathur, A. Kumar, and M. Chandra, "A Feature Based Neural Network Model for Weather Forecasting," *International Journal of Computational Intelligence*, vol. 4, no. 3, pp. 209–216, 2001.

[106] D. Graupe, *Principle of Artificial Neural Networks*, 3rd ed. World Scientific, UK, 2013.

[107] W. S. McCulloch and W. Pitts, "A Logical Calcul of the Ideas Immanent In Nervous Activity," *The bulletin of Mathemetical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[108] F. Mekanik and M. A. Imteaz, "Capability of Artificial Neural Networks for predicting long-term seasonal rainfalls in east Australia," *20th International Congress on Modelling and Simulation*, pp. 1–6, 2013.

[109] C. W. Dawson and R. L. Wilby, "Hydrological modelling using artificial neural networks," *Prog. Phys. Geogr*, vol. 25, pp. 80–108, 2001, doi: 10.1177/030913330102500104.

[110] N. J. De Vos and T. H. M. Rientjes, "Constraints of artificial neural networks for rainfall-runoff modelling : trade-offs in hydrological state representation and model evaluation Input units Hidden layer Output layer," *Hydrology and Earth System Sciences*, vol. 9, pp. 111–126, 2005.

[111] S. M. Sumi, M. F. Zaman, and H. Hirose, "A rainfall forecasting method using machine learning models and its application to the fukuoka city case," *International Journal of Applied Mathematics and Computer Science*, vol. 22, no. 4, pp. 841–854, 2012, doi: 10.2478/v10006-012-0062-1.

[112] S. K. Singh, S. K. Jain, and A. Bárdossy, "Training of Artificial Neural Networks Using Information-Rich Data," *Hydrology*, vol. 1, pp. 40–62, 2014, doi: 10.3390/hydrology1010040.

[113] H. R. Maier and G. C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications," *Environmental Modelling and Software*, vol. 15, no. 1, pp. 101–124, 2000, doi: 10.1016/S1364-8152(99)00007-9.

[114] G. López, M. A. Rubio, M. Mart, and F. J. Batlles, "Estimation of hourly global

photosynthetically active radiation using artificial neural network models," *Agricultural and Forest Meteorology*, vol. 107, pp. 279–291, 2001.

[115] T. Kim and J. B. Valde, "Nonlinear Model for Drought Forecasting Based on a Conjunction of Wavelet Transforms and Neural Networks," *J. Hydrol. Eng,* vol. 8, no. 6, pp. 319–328, 2004.

[116] J. Y. Sung, J. Lee, I. Chung, and J. Heo, "Hourly Water Level Forecasting at Tributary Affected by Main River Condition," *Water*, vol. 9, pp. 1–17, 2017, doi: 10.3390/w9090644.

[117] M. T. Hagan and M. B. Menhaj, "Training Feedforward Networks with the Marquardt Algorithm," *IEEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 5, no. 6, pp. 989–993, 1994.

[118] K. C. Luk, J. E. Ball, and A. Sharma, "A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting," *Journal of Hydrology,* vol. 227, no. 1–4, pp. 56–65, 2000, doi: 10.1016/S0022-1694(99)00165-1.

[119] W. S. Sarle, "Stopped Training and Other Remedies for Overfitting," *Proceedings of the 27th Symposium on the Interface*, pp. 1–10, 1995.

[120] R. D. De Veaux and L. H. Ungar, "Multicollinearity: A tale of two nonparametric regressions," *Springer*, vol. 89, pp. 393–402, 1994, doi: 10.1007/978-1-4612-2660-4_40.

[121] L. Xu, N. Chen, X. Zhang, and Z. Chen, "A data-driven multi-model ensemble for deterministic and probabilistic precipitation forecasting at seasonal scale," *Climate Dynamics*, vol. 54, no. 7–8, pp. 3355–3374, 2020, doi: 10.1007/s00382-020-05173-x.

[122] F. J. Lin, "Solving multicollinearity in the process of fitting regression model using the nested estimate procedure," *Quality and Quantity,* vol. 42, no. 3, pp. 417–426, 2008, doi: 10.1007/s11135-006-9055-1.

[123] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting,* vol. 22, pp. 679–688, 2006, doi: 10.1016/j.ijforecast.2006.03.001.

[124] D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, "Model Evaluation Guidelines For Systematic Quantification Of Accuracy In Watershed Simulations," *American Society of Agricultural and Biological Engineers,*

vol. 50, no. 3, pp. 885–900, 2007.

[125] J. De Ona, "Extracting the contribution of independent variables in neural network models : a new approach to handle instability," *Neural Comput. Appl*, vol. 25, pp. 859–869, 2014, doi: 10.1007/s00521-014-1573-5.

[126] D. . Garson, "Interpreting neural network connection weights," *Artif. Intell. Expert*, vol. 6, pp. 47–51, 1991.

[127] A. T. C. Goh, "Back-propagation neural networks for modeling complex systems," *Artificial Intelligence in Engineering*, vol. 9, pp. 143–151, 1995.

[128] J. D. Olden and D. A. Jackson, "Illuminating the '" black box "': a randomization approach for understanding variable contributions in artificial neural networks," *Ecological Modelling*, vol. 154, pp. 135–150, 2002.

[129] I. Dimopoulos, J. Chronopoulos, and S. Lek, "Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city ( Greece )," *Ecological Modelling*, vol. 120, pp. 157–165, 1999.

[130] S. Lek, M. Delacoste, I. Dimopoulos, J. Lauga, and S. Aulagnier, "Application of neural networks to modelling nonlinear relationships in ecology," *Ecological Modelling*, vol. 90, pp. 39–52, 1996.

[131] M. Gevrey and I. Dimopoulos, "Review and comparison of methods to study the contribution of v ariables in artificial neural network models," *Ecological Modelling*, vol. 160, pp. 249–264, 2003.

[132] J. D. Olden, M. K. Joy, and R. G. Death, "An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data," *Ecological Modelling*, vol. 178, pp. 389–397, 2004, doi: 10.1016/j.ecolmodel.2004.03.013.

[133] M. J. Watts and S. P. Worner, "Using artificial neural networks to determine the relative contribution of abiotic factors influencing the establishment of insect pest species," *Ecological Informatics*, vol. 3, pp. 64–74, 2008, doi: 10.1016/j.ecoinf.2007.06.004.

[134] F. K. Ogou, K. Batebana, B. A. Ogwang, Z. M. M. Sein, V. Ongoma, and J. P. Ngarukiyimana, "Investigation of the influence of Atlantic Ocean on rainfall variability over Benin republic, West Africa," *Ethiopian Journal of Environmental Studies & Management*, vol. 9, no. 1, pp. 70–79, 2016, doi: http://dx.doi.org/10.4314/ejesm.v9i1.7.

# Appendix

## Additional Figures

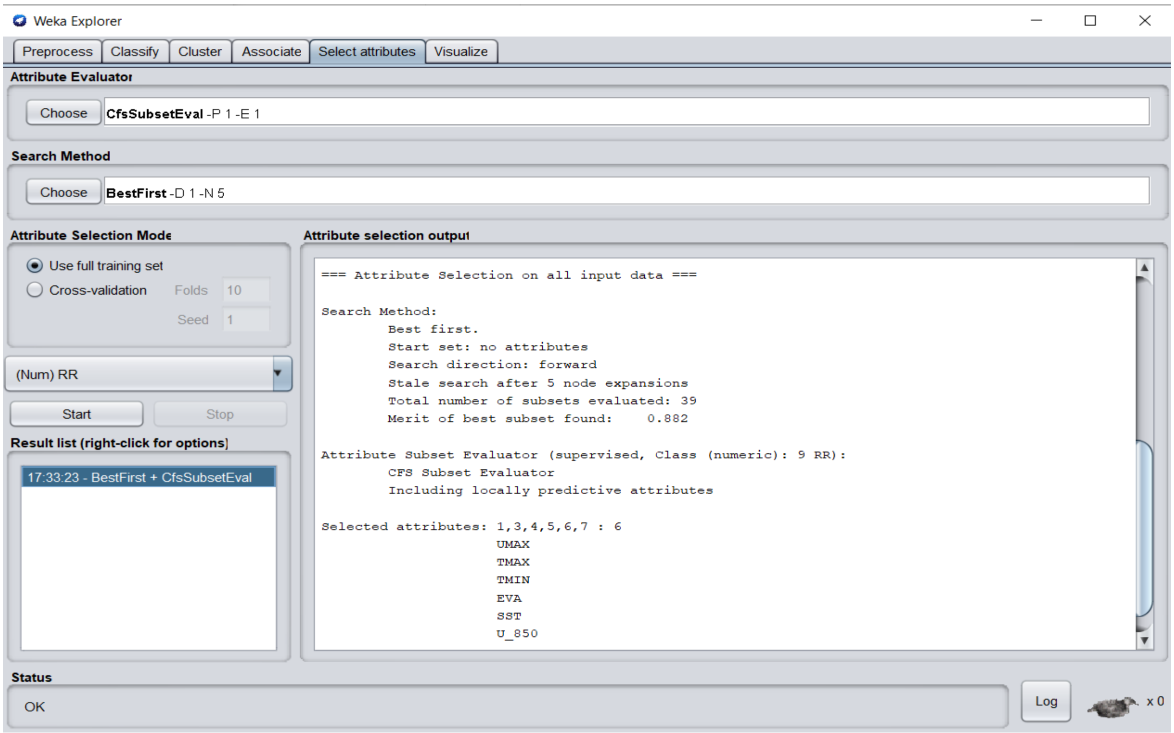



Figure A.1: CFS result for Kandi using Best First search method.

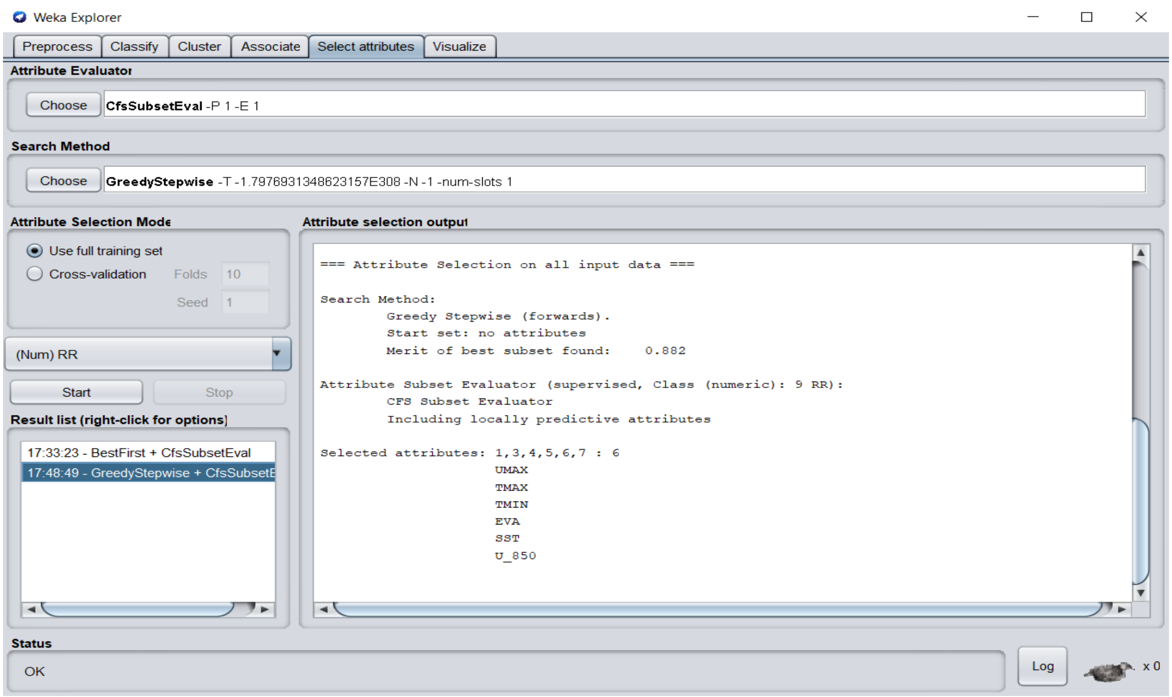
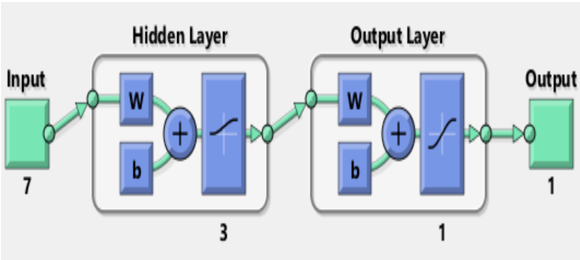
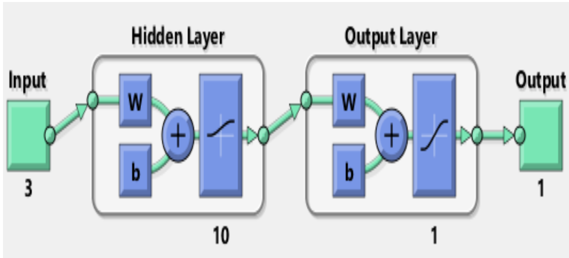



Figure A.2: CFS result for Kandi using Greedy Stepwise search method.

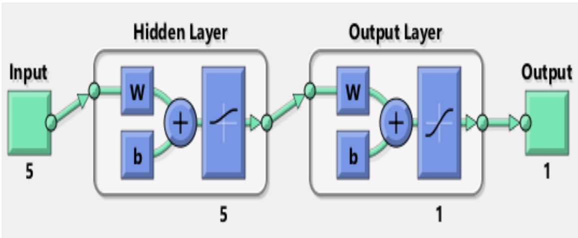
**Cotonou ANN (7-3-1)**



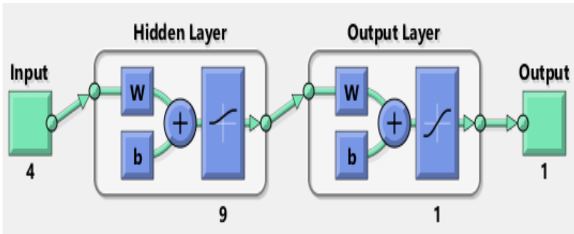

**Bohicon ANN (3-10-1)**



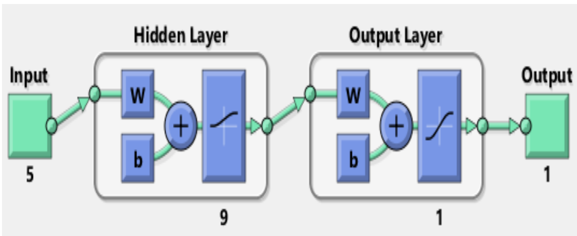

**Save ANN (5-5-1)**




**Parakou ANN (4-9-1)**




**Natitingou ANN (5-9-1)**




**Kandi ANN (6-10-1)**

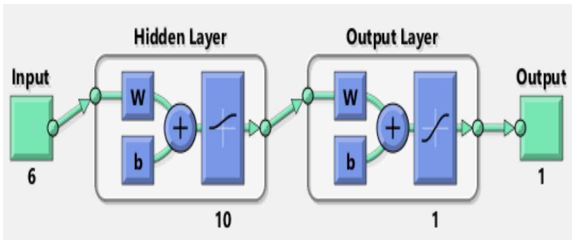



Figure A.3: Optimal proposed ANN architectures for each station.

# Additional Tables

Table B.1: The performance for training-validation and testing parts of the ANN model for Cotonou.

| Number of Hidden Nodes | RMSE (mm) | | MAE (mm) | | $R^2$ | | NSE | |
|---|---|---|---|---|---|---|---|---|
| | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing |
| 2 | 90.09 | 70.25 | 57.63 | 47.37 | 0.44 | 0.62 | 0.40 | 0.60 |
| **3** | **91.69** | **68.00** | **57.22** | **48.30** | **0.45** | **0.64** | **0.38** | **0.62** |
| 4 | 92.20 | 69.40 | 57.28 | 47.91 | 0.44 | 0.63 | 0.37 | 0.61 |
| 5 | 93.27 | 69.70 | 57.99 | 46.13 | 0.43 | 0.64 | 0.36 | 0.60 |
| 6 | 90.73 | 69.07 | 56.50 | 42.70 | 0.46 | 0.65 | 0.39 | 0.61 |
| 7 | 91.31 | 68.09 | 56.84 | 48.18 | 0.42 | 0.63 | 0.38 | 0.62 |
| 8 | 91.94 | 69.61 | 57.36 | 46.71 | 0.43 | 0.62 | 0.37 | 0.60 |
| 9 | 91.24 | 70.65 | 58.25 | 48.94 | 0.42 | 0.62 | 0.38 | 0.59 |
| 10 | 90.43 | 70.35 | 57.05 | 46.68 | 0.41 | 0.61 | 0.39 | 0.59 |

Table B.2: The performance for training-validation and testing parts of the ANN model for Bohicon.

| Number of Hidden Nodes | RMSE (mm) | | MAE (mm) | | $R^2$ | | NSE | |
|---|---|---|---|---|---|---|---|---|
| | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing |
| 2 | 60.28 | 52.75 | 40.55 | 37.83 | 0.48 | 0.44 | 0.43 | 0.42 |
| 3 | 60.21 | 52.98 | 40.55 | 37.83 | 0.47 | 0.44 | 0.43 | 0.41 |
| 4 | 60.16 | 51.94 | 40.47 | 36.54 | 0.48 | 0.46 | 0.44 | 0.43 |
| 5 | 59.02 | 52.85 | 39.97 | 38.15 | 0.49 | 0.43 | 0.46 | 0.41 |
| 6 | 59.20 | 51.16 | 39.78 | 36.12 | 0.49 | 0.47 | 0.45 | 0.45 |
| 7 | 62.12 | 53.78 | 41.42 | 36.42 | 0.47 | 0.43 | 0.40 | 0.39 |
| 8 | 60.27 | 53.83 | 40.47 | 38.10 | 0.48 | 0.42 | 0.43 | 0.39 |
| 9 | 59.67 | 52.00 | 40.24 | 35.13 | 0.50 | 0.46 | 0.44 | 0.43 |
| **10** | **58.75** | **51.00** | **39.51** | **37.47** | **0.50** | **0.47** | **0.46** | **0.45** |

Table B.3: The performance for training-validation and testing parts of the ANN model for Save.

| Number of Hidden Nodes | RMSE (mm) | | MAE (mm) | | $R^2$ | | NSE | |
|---|---|---|---|---|---|---|---|---|
| | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing |
| 2 | 52.97 | 52.42 | 37.07 | 37.54 | 0.58 | 0.54 | 0.56 | 0.52 |
| 3 | 53.09 | 52.19 | 37.46 | 36.95 | 0.59 | 0.53 | 0.56 | 0.53 |
| 4 | 52.62 | 51.59 | 36.71 | 35.84 | 0.59 | 0.54 | 0.57 | 0.54 |
| **5** | **53.47** | **51.00** | **37.26** | **34.71** | **0.57** | **0.56** | **0.55** | **0.55** |
| 6 | 53.13 | 52.53 | 37.35 | 37.15 | 0.58 | 0.54 | 0.56 | 0.52 |
| 7 | 53.32 | 52.24 | 37.00 | 36.11 | 0.58 | 0.56 | 0.56 | 0.53 |
| 8 | 53.36 | 51.65 | 36.92 | 36.57 | 0.59 | 0.56 | 0.56 | 0.54 |
| 9 | 53.36 | 51.87 | 36.19 | 35.95 | 0.58 | 0.55 | 0.56 | 0.54 |
| 10 | 53.17 | 51.79 | 36.51 | 36.99 | 0.58 | 0.56 | 0.56 | 0.54 |

Table B.4: The performance for training-validation and testing parts of the ANN model for Parakou.

| Number of Hidden Nodes | RMSE (mm) | | MAE (mm) | | $R^2$ | | NSE | |
|---|---|---|---|---|---|---|---|---|
| | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing |
| 2 | 54.15 | 60.19 | 35.02 | 38.14 | 0.71 | 0.63 | 0.69 | 0.62 |
| 3 | 52.74 | 57.19 | 33.87 | 38.62 | 0.72 | 0.66 | 0.71 | 0.66 |
| 4 | 52.12 | 57.97 | 33.79 | 38.91 | 0.73 | 0.66 | 0.71 | 0.65 |
| 5 | 56.03 | 57.45 | 36.23 | 38.85 | 0.69 | 0.66 | 0.67 | 0.66 |
| 6 | 53.23 | 58.85 | 33.81 | 40.83 | 0.72 | 0.65 | 0.70 | 0.64 |
| 7 | 54.26 | 58.56 | 35.05 | 38.91 | 0.71 | 0.65 | 0.69 | 0.64 |
| 8 | 56.50 | 58.85 | 35.76 | 36.66 | 0.70 | 0.64 | 0.67 | 0.64 |
| **9** | **52.28** | **56.45** | **33.78** | **37.48** | **0.72** | **0.68** | **0.71** | **0.67** |
| 10 | 52.93 | 58.39 | 34.44 | 37.38 | 0.72 | 0.65 | 0.71 | 0.65 |

Table B.5: The performance for training-validation and testing parts of the ANN model for Natitingou.

| Number of Hidden Nodes | RMSE (mm) | | MAE (mm) | | $R^2$ | | NSE | |
|---|---|---|---|---|---|---|---|---|
| | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing |
| 2 | 45.65 | 41.08 | 28.70 | 27.15 | 0.81 | 0.84 | 0.80 | 0.83 |
| 3 | 45.97 | 40.65 | 29.00 | 26.95 | 0.81 | 0.86 | 0.80 | 0.84 |
| 4 | 45.81 | 40.25 | 29.00 | 25.72 | 0.81 | 0.86 | 0.80 | 0.84 |
| 5 | 46.79 | 39.28 | 29.38 | 26.11 | 0.81 | 0.86 | 0.79 | 0.85 |
| 6 | 47.34 | 39.64 | 29.75 | 26.27 | 0.80 | 0.86 | 0.79 | 0.85 |
| 7 | 46.31 | 40.77 | 29.52 | 27.41 | 0.80 | 0.87 | 0.80 | 0.84 |
| 8 | 47.67 | 41.15 | 30.17 | 25.65 | 0.80 | 0.85 | 0.79 | 0.83 |
| **9** | **47.18** | **38.86** | **29.63** | **26.9** | **0.79** | **0.87** | **0.79** | **0.85** |
| 10 | 46.78 | 41.44 | 29.46 | 27.59 | 0.80 | 0.86 | 0.79 | 0.83 |

Table B.6: The performance for training-validation and testing parts of the ANN model for Kandi.

| Number of Hidden Nodes | RMSE (mm) | | MAE (mm) | | $R^2$ | | NSE | |
|---|---|---|---|---|---|---|---|---|
| | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing | Training + Validation | Testing |
| 2 | 41.63 | 47.11 | 25.18 | 28.12 | 0.81 | 0.82 | 0.81 | 0.81 |
| 3 | 42.66 | 47.41 | 25.72 | 28.51 | 0.81 | 0.82 | 0.80 | 0.81 |
| 4 | 41.85 | 49.79 | 25.08 | 28.97 | 0.81 | 0.81 | 0.81 | 0.79 |
| 5 | 43.27 | 48.64 | 26.58 | 28.68 | 0.81 | 0.80 | 0.80 | 0.80 |
| 6 | 42.49 | 49.16 | 26.67 | 27.82 | 0.80 | 0.80 | 0.80 | 0.80 |
| 7 | 49.89 | 48.35 | 30.59 | 28.55 | 0.79 | 0.82 | 0.73 | 0.80 |
| 8 | 42.47 | 48.76 | 25.35 | 29.66 | 0.81 | 0.81 | 0.80 | 0.80 |
| 9 | 44.29 | 49.24 | 26.59 | 29.01 | 0.79 | 0.81 | 0.79 | 0.80 |
| **10** | **47.34** | **45.52** | **28.85** | **27.06** | **0.78** | **0.83** | **0.76** | **0.83** |

# Table of Contents