TIME SERIES PREDICTION OF WEATHER VARIABLES AND RAINFALL FORECASTING USING THE FUZZY LOGIC BASED RAIN FORECAST MODEL

M. B. Mu'azu[1] and B. G. Bajoga[1]

[1]Department of Electrical and Computer Engineering, Ahmadu Bello University, Zaria

Absract

The Neuro-Fuzzy System developed using the Soft Computing technique of Neuro-Fuzzy is implemented as a pure Fuzzy Logic System as the Rain Forecast Model (RFM). The Rain Forecast Model (RFM) used as its inputs the weather variables: Relative Humidity, Wind Direction, and Wind Speed (from 1993 to 2002) and in order to forecast into the future (2003 to 2010), these weather variables were determined using a hybrid statistical technique comprising of Moving Averages and Exponential Smoothing. Two statistical tests (Root Mean Squared Error (RMSE) and Correlation Factor (R)) were carried out in order to measure the performance of the developed model with the validation data and with the forecasted weather variables. The developed model, using the validation data of 2003 to 2005, produced the following performance results: RMSE=28.02 and R=0.97. This formed the basis for comparing the performance of the developed model when subjected to the forecasted weather variables (2003 to 2010). The performance function results obtained were: RMSE=54.99 and R=0.88**.** Considering the nature of the problem, that is the unpredictability of rainfall, these are reasonable results.

Keywords: Soft Computing, Rain Forecast Model, Rainfall, Wind speed, Wind direction, Relative Humidity

1. Introduction

Weather forecast in general and rainfall forecast in particular, especially on the long term, is a difficult task considering its extreme variability. A number of possible interactions within the atmosphere and between the atmosphere and other influences make it impossible to make long-range forecasts with the same degree of accuracy as short-range forecasts. The root of this problem is referred to as 'Chaos'. This chaotic effect, also called the "butterfly effect" was first described by Edward Lorenz, a meteorologist working at the Massachusetts Institute of Technology, and has become a field of study in its own right (Mu'azu, 2006).

Rainfall is governed by the interaction of the moist tropical maritime air mass and the dry, cool tropical continental air mass. Air mass is defined as a large uniform (with respect to temperature and water vapour) body of air within the atmosphere. Rainfall is characterized by its extreme variability, both of intensity and duration with a temporal and spatial pattern. The characteristic and intensity of the prevailing weather conditions are determined by the surface location of the moisture boundary zone separating the two air masses. The boundary is known as Inter Tropical Divergence (ITD). It is the location of a place in relation to the position of the Inter Tropical divergence (ITD) that determines its weather situation (Mu'azu, 2006).

Rain is one of nature's greatest gifts and is very vital to the agricultural economies and food security of developing countries like Nigeria and also sustenance of the environment. In Nigeria, rainfall determines the zonal pattern of crops and the seasonal activities of farmers. It is then important that rainfall trends be identified as any unforeseen deviations can cause unwanted disruptions. This has assumed an even greater importance due to the threats posed by global warming and greenhouse effect (Mu'azu, 2006).

This work is aimed at the development of a Neuro-Fuzzy based Rain Forecast Model (RFM) and time series forecasting of weather variables using Zaria as a case study. A Neuro-Fuzzy System is an amalgamation of the soft computing techniques of neural nets (with their learning capabilities) and fuzzy logic (with their knowledge base and transparency). It, therefore, combines the advantages of fuzzy logic systems, which deal with explicit knowledge that can be explained and understood, and neural nets, which deal with implicit knowledge, which can be acquired by learning. The objective then is to develop a data driven method using the soft computing technique of Neuro-Fuzzy to forecast the rainfall using the following weather variables: Relative Humidity, Wind Direction and Wind Speed. The Neuro-Fuzzy System developed is implemented (using the fuzzy technology language (FTL)) as a pure fuzzy logic system as the Rain Forecast Model (RFM). In order to forecast into the future, these weather variables are forecasted using a hybrid statistical technique comprising of moving averages and exponential smoothing. The forecast period is between 2003 and 2010.

A time series is a set of observations of a variable at regular intervals over time. A basic requirement for time series is that the data be displayed in the order in which they occurred, since it is possible that successive observations may probably be dependent (Rojas and Pomares, 2004, Monks, 1996). According to Rojas and Pomares (2004), the challenge of predicting future values of a time series spans a variety of disciplines and as such time series techniques find application in such diverse data sets as equity market prices, disease control, meteorological measurements, astronomic observations, etc. Typical statistical time series techniques include simple moving averages and exponential smoothing. According to Easton and McColl (1997) and Roberts (2004), a moving average is a form of average which has been adjusted to allow for seasonal or cyclical components of a time series while exponential smoothing is used to reduce irregularities in the time series data thus providing a means of predicting future values of the time series.

Time series prediction is basically a modeling problem, which can better be solved using Soft Computing techniques. This is because it is possible that the underlying relationship between the data may not be known. The first step in the solution is establishing a non-linear mapping between inputs and outputs, after which the model can be used to predict future values based on past and present observations.

2. Study Area

Zaria is located within the Sudan Savannah zone lying on a plateau of about 670.56m above sea level and on latitude 11°8′N and longitude 7°41′E. It lies within a region with distinct dry and wet seasons, which are influenced by two distinct air masses. The wet season occurs in the high sun period and is dominated by the South-West winds coming in around April or May and lasting till around October. About 65% of the rains occur between July and September. The dry season lasts between November and March and is practically rainless. The months of December to February, the harmattan season, are usually cold and dry due to the influence of the continental air mass (dry-dusty wind) from the desert regions of North Africa (Grodsky and Carton, 2002, Mu'azu, 2006)

.

3. Materials and Methods

Using weather variables (Wind Direction, Wind Speed and Relative Humidity) collected from the Nigeria Meteorological Agency (NIMET), Zaria and the Meteorological Unit of the Department of Soil Science, Ahmadu Bello University, Zaria on a monthly average basis over a thirteen- (13) year period (1993 – 2005), a rainfall prediction model is to be developed. Data from 1993 to 2002 is used as the training set whilst the data from 2003 to 2005 is used as the validation data. However, since the exact relationships between these variables are not known, data driven methods are more suitable in developing the prediction model. These methods perform a kind of function fitting by using multiple parameters on the existing information in order to predict the possible relationships in the near future.

The overall methodology adopted in carrying out this investigation involves the following:

i) Data training and development of the Neuro-Fuzzy System using the fuzzyTECH Neuro-Fuzzy module.
ii) Implementation as a pure fuzzy logic system using the Fuzzy Technology Language (FTL) resulting in the development of the Rain Forecast Model (RFM).
iii) Validation of the Rain Forecast Model (RFM) using the validation data.

iv) The weather variables, relative humidity, wind direction and wind speed are forecasted using the time series techniques of simple moving averagesand exponential smoothing. The forecasted values are used to forecast for the rainfall using the Rain Forecast Model (RFM) developed in i) – ii)

The methodology adopted in data training and in developing the Neuro-Fuzzy System using the fuzzyTECH Neuro-Fuzzy module involves the following:

i) Obtain training sample data. The weather data used in this work were obtained from the Nigeria Meteorological Agency (NIMET), Zaria and the Meteorological Unit of the Department of Soil Science, Ahmadu Bello University, Zaria for the period, 1993 – 2005. The data for 1993 – 2002 is used as the test data while that of 2003 – 2005 is used as the validation data.
ii) Cluster the sample data (if necessary). The data obtained may have to be pre-processed to remove redundant data and to resolve conflicts in the data.
iii) Create an empty Fuzzy Logic System i.e. the rule set will have all rules with DoS (Degree of Support) = 0. This means the rule set is COMPLETE but FALSE**;** this is required as Neuro-Fuzzy training can only start with an existing rule set. A DoS value gives the weight for each value to be used in the rule aggregation step of fuzzy inference. The value is between 0 and 1.
iv) Collection of expert knowledge about the process and entering all existing knowledge (if any) in the solution.
v) Selection of the components of the Fuzzy Logic System to be trained. All or specific components can be opened for learning.
vi) Configuring the Neuro-Fuzzy module by specifying the LEARNING METHOD and setting parameters for it. The learning method can be any of **RealMethod, RandomMethod, Batch_Learn** and **Batch_Random.**
vii) Train with the sample data to learn parameters of the Fuzzy System.
viii) Evaluate system performance and validation of results. This is accomplished by testing trained system with sample **test data.** This will help in minimizing the occurrence of "**over training".**
ix) Manual optimization using an interactive approach with the aid of the watch window to eliminate functionally redundant or unnecessary rules**..**

The result of the Neuro-Fuzzy Module training is a 'pure' Fuzzy Logic System, which, can be, implemented on computers, microcontrollers or industrial controllers (Inform GmbH, 2001). The 'pure' fuzzy logic system developed for this work, named the Rain Forecast Model (RFM), is implemented using the Fuzzy Technology Language (FTL).

3.1 Rain Forecast Model

The Fuzzy Technology Language (FTL) is a hardware and vendor independent description language for Fuzzy Logic Systems with definition formats and no loops and branches. Such chip manufacturers as Intel, SGS-Thomson, Siemens, Texas Instruments, Microchip, etc support the Fuzzy Technology Language (FTL). The Fuzzy Technology Language (FTL) consists basically of two entities:

i) **Objects:** Each object consists of an object name and an object body in **"{ }".** Within an object body, other objects and slots can be defined.

ii) **Slots:** A slot consists of a slot name to the left of an **"="** and a value for the slot to the right.

Comments are put into **"/* … */"** marks (Von Altrock, 1995).

The general flow of the Rainfall Forecast Model using the Fuzzy Technology Language (FTL) is described as follows:

PROJECT{

/*Main section for project data*/

SHELLOPTIONS{

/*Section containing terminal configurations for the project*/

}/* SHELLOPTIONS**/*

MODEL{

/*Section containing the entire model definitions (Variable and Object Section)*/

VARIABLE_SECTION{

/*Section containing all definitions of linguistic variables*/

LVAR{

TERM{

}/*TERM*/

}/*LVAR*/

}/*VARIABLE SECTION*/

OBJECT_SECTION{

/*Section containing all structural elements of a Fuzzy Logic System*/

RULEBLOCK{

RULES{

}/*RULES*/

}/*RULEBLOCK*/

}/*OBJECT SECTION*/

}/*MODEL*/

}/*PROJECT*/

The Rain Forecast Model is shown in Figure 1

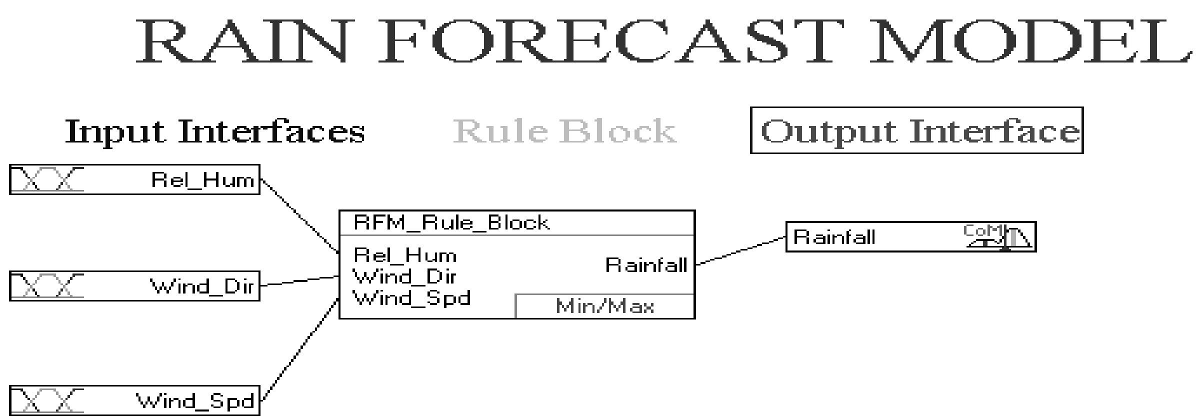

Figure 1: Structure of the Rain Forecast Model Fuzzy Logic System

Time Series Analysis

Time series analysis relies, at least in part, on understanding or exploiting the dependence among the observations. Because of this reliance, the goal of time series prediction can, therefore, be stated thus: Given a sequence up to time t, $x(1), x(2), \ldots, x(t)$, find the continuation $x(t+1), x(t+2), \ldots$ (Rojas and Pomares, 2004). The method to be employed in finding this continuation is a hybrid of the simple moving averages method and that of exponential smoothing.

3.2 Simple Moving Averages

The basic components of a time series are:

i) TREND, which is the underlying direction (upward or downward tendency) and rate of change in a time series, when allowance has been made for the other components.
ii) SEASONAL EFFECTS, which describe similar variations occurring during corresponding periods.
iii) CYCLICAL FACTORS, which describe any regular fluctuations.

The simple moving averages method is a smoothing technique that uses a type of average that is adjusted to allow for seasonal or cyclical components of a time series and which helps make the long term trends clearer (Monks, 1996, Roberts, 2004).

In the simple moving average method, only the **n** most recent periods of data points need to be maintained. At the end of each period, the oldest period's data is discarded and the newest period's data is added to the database. The database is then divided by **n** and used as a forecast for the next period (Monks, 1996, Roberts, 2004).

In order to determine the moving average $x(t+1)$ for a period $(t+1)$, the following expression is used:

$$x(t+1) = \frac{[D_t + D_{t-1} + \ldots + D_{t-n-1}]}{n}, \quad n \le t \qquad 1$$

Where, n = number of observations used in the calculation
D = time series data

A critical issue in using the simple moving average method is in the choice of **n**. It can be determined by selecting a value that minimizes the Mean Squared Error (MSE) of the forecasting, which is given by (Roberts, 2004):

$$MSE = \frac{\sum_{t=1}^{n} (D_t - x_t)^2}{n} \qquad 2$$

The simple moving average method was used for each of the weather variables on a monthly basis, that is, say for relative humidity, April 1993 to April 2010 and so on. A moving 'window' of three-year, five-year, seven-year and nine-year width (n=3, n=5, n=7 and n=9) respectively was applied on the validation data of 2003 to 2005 in order to determine the choice of n to minimize the Mean Squared Error (MSE). The result is as shown in Table 1.

Table 1: MSE Values For n=3, n=5, n=7 and n=9

	2003				**2004**				**2005**			
	n=3	n=5	n=7	n=9	n=3	n=5	n=7	n=9	n=3	n=5	n=7	n=9
REL_HUM	111.1	73.9	59.96	**53.08**	285.4	87.68	94.4	**60.09**	150.1	43.94	30.97	**15.65**
WIND_DIR	685.4	116.33	63.34	**38.76**	928.6	222.4	179.58	**103.2**	1342	401.78	292.44	**160.6**
WIND_SPD	568.6	185.1	251.33	**152.5**	285.4	191.6	222.89	**125.5**	150.1	81.97	40.395	**79.66**

On the basis of Table 1, the choice of **n=9** was decided upon since, on the overall, it exhibits the least Mean Squared Error (MSE) values. The Mean Squared Error (MSE) values obtained, as in Table 1, also indicate that the higher the value of **n** the lower the value of the Mean Squared Error (MSE). It should also be noted that in this case the highest possible value of n is 9. The selected value is then used to determine preliminary forecast values for the weather variables from 2003 to 2010. The exponential smoothing method is then applied to obtain the forecast values to be used with the developed Rain Forecast Model (RFM).

3.3 Simple Exponential Smoothing Method

Exponential smoothing is a forecasting technique that weighs past data in an exponential manner so that the most recent data carries more weight. This means that an exponentially smoothed moving average is based not on a sequential average of individual time periods but on the most recent data and the average prior to it adjusted by a smoothing constant, α. The smoothing constant, α, is a number in the range [0, 1] with smaller values resulting in less weight given to the current time period. It can be determined on a trial-and-error basis but Roberts (2004) has determined an empirical expression for the smoothing constant as:

$$\alpha = \frac{1}{\frac{n}{2} - 1} \qquad 3$$

Where n is the moving window width.

The smoothed series is recursively updated as new observations are recorded using the following expression:

$$S(t) = \alpha D(t) + (1 - \alpha) S(t - 1) \qquad 4$$

Where S(t) is the value of the smoothed series at period t.
The current smoothed value is an interpolation between the previous smoothed value (S(t-1)) and the current observation, where α controls the closeness of the interpolated value to the most recent observation.

Based on the choice of n=9 from Table 1, and from equation 3, a value of α=0.286 is used to determine the forecast values of the weather variables using Microsoft Excel®. A sample Microsoft Excel® spreadsheet showing the procedure is as shown in Figure 2.

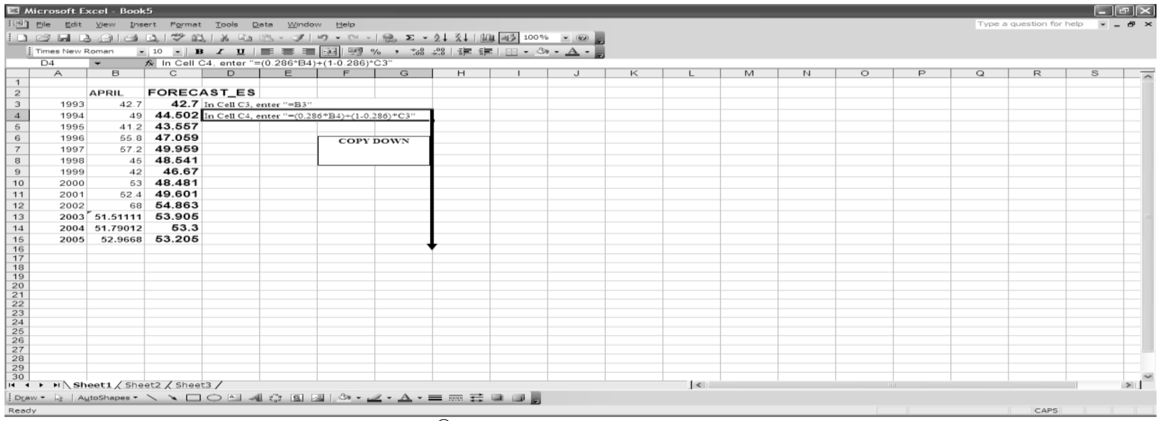

Figure 2: Sample Microsoft Excel® Spreadsheet Showing Exponential Smoothing Procedure

4. Testing with rain forecast model (RFM)

The forecasted values of the weather variables are then used to forecast the monthly rainfall amounts for 2003 to 2010 using the developed Rain Forecast Model (RFM). Sample watch window outputs for October 2005, April 2006, August 2008 and July 2010 are as shown in Figures 3 – 6.

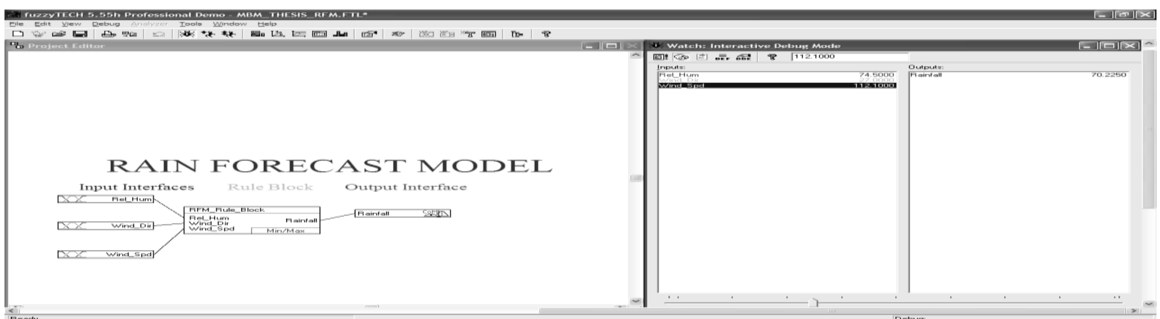

Figure 3: Watch Window Output For October 2005

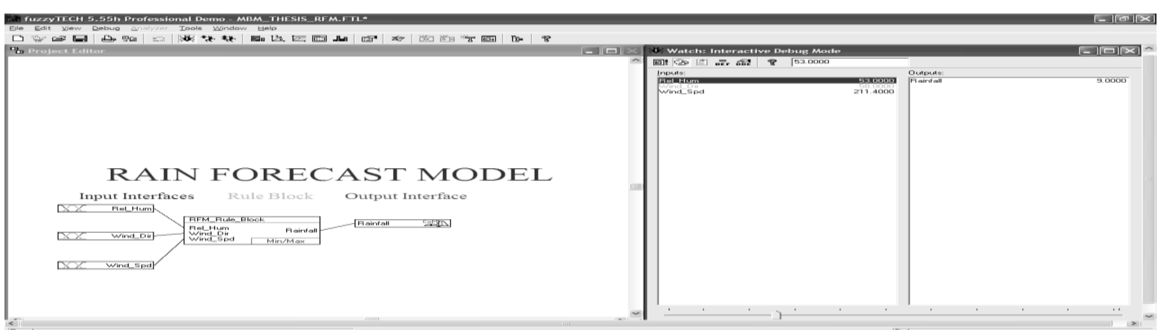

Figure 4: Watch Window Output For April 2006

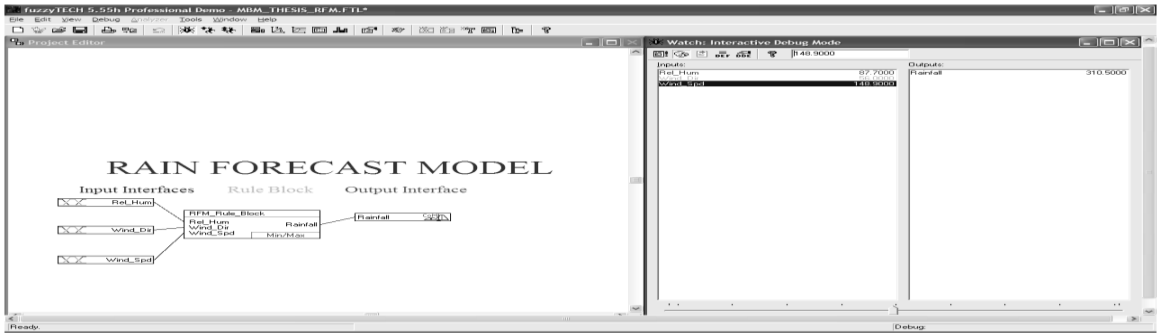

Figure 5: Watch Window Output For August 2008

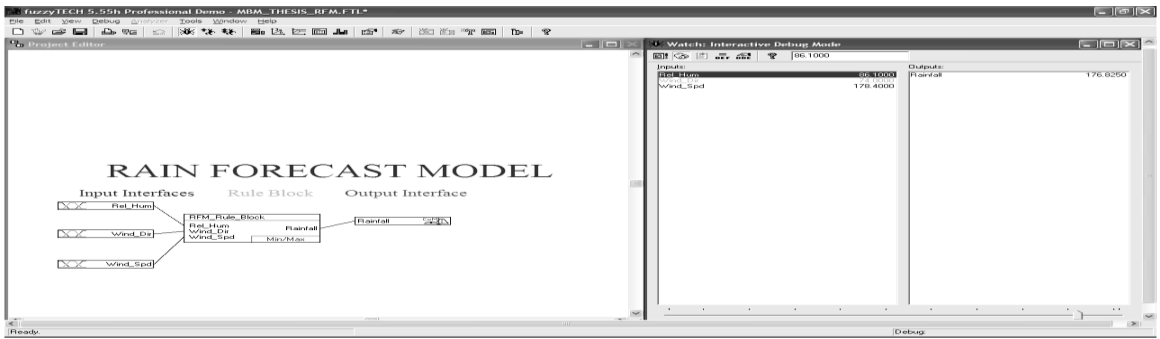

Figure 6: Watch Window Output For July 2010

A plot of the rainfall data, forecasted (2003 to 2010) and actual (2003 to 2005) is shown in Figure 7. Figure 8 shows a plot of the rainfall data, actual (2003 to 2005), forecasted (2003 to 2010) and that obtained from subjecting the actual weather variables of 2003 to 2005 to the Rain Forecast Model (RFM).

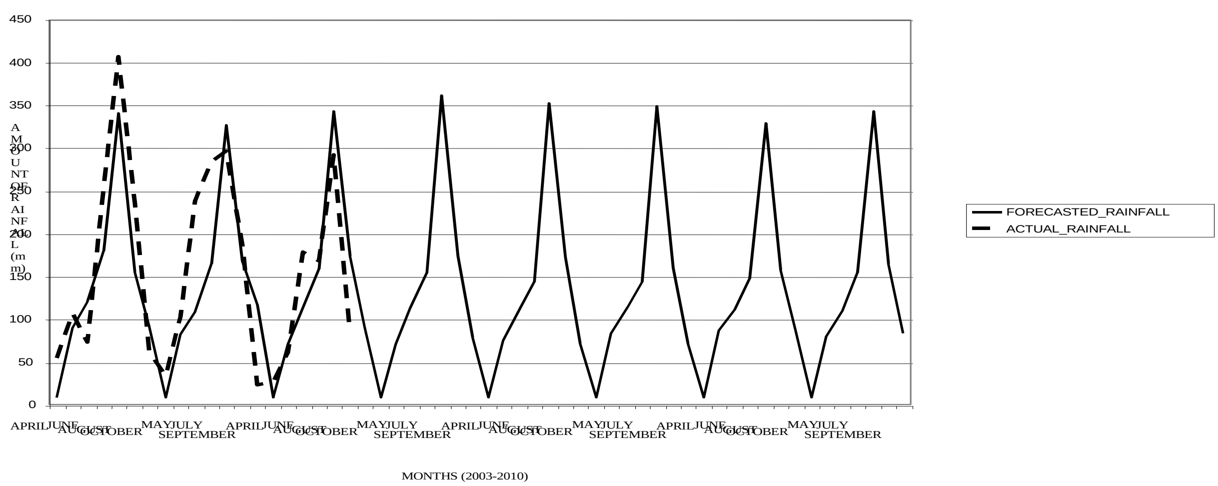

Figure 7: Plot of Actual (2003-2005) And Forecasted (2003-2010) Rainfall Data

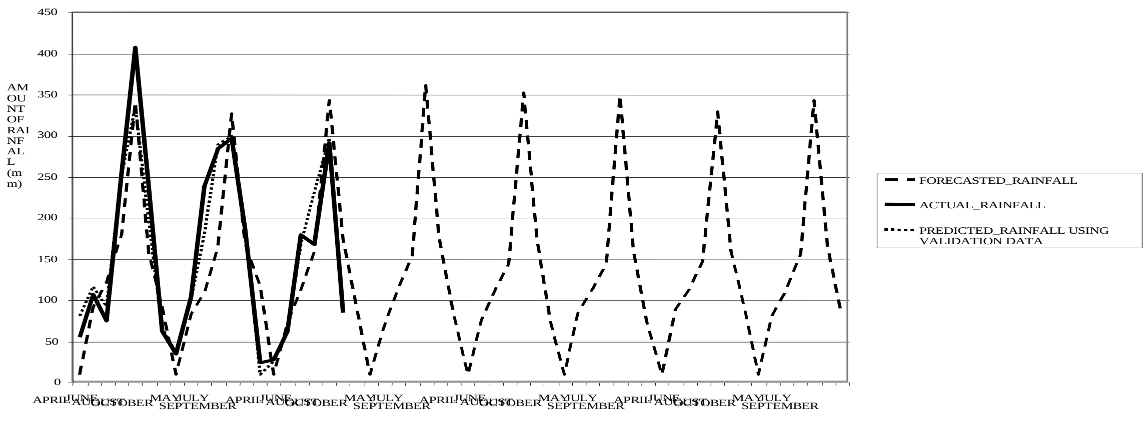

Figure 8 Plot of Actual (2003-2005), Validation (2003-2005) and Forecasted (2003-2010) Rainfall Data

5. Analysis and discussion of findings

The plot of the actual rainfall data for the validation period and that forecasted using the Rain Forecast Model (RFM) is shown in Figure 12. The Root Mean Square Error (RMSE) between the actual rainfall and forecasted rainfall generated by the Rain Forecast Model (RFM) is used as the performance function. The Root Mean Square Error (RMSE) is determined from:

$$RMSE = \sqrt{MSE} \qquad 5$$

Where, $$MSE = \frac{\sum_{t=1}^{n}\left(Actual\_Rainfall_t - Predicted\_Rainfall_t\right)^2}{n} \qquad 6$$

In addition to this, the Pearson's correlation factor, R, between the actual rainfall and forecasted rainfall generated by the Rain Forecast Model (RFM) has also been determined. The correlation coefficient, a number between 0 and 1, is a measure of how well trends in the validation values using the developed Rain Forecast Model (RFM) follow trends in the actual data. What constitutes a good value of R is dependent on the problem domain. In this case, R > 0.8 is an indication of a good model considering the volatility of weather variables in general and rainfall in particular.

The Root Mean Square Error (RMSE) is 28.02, while the correlation factor, R is 0.97.

The Root Mean Square Error (RMSE) between the actual rainfall and forecasted rainfall generated by the Rain Forecast Model (RFM) based on the forecasted weather variables is used as the performance function. In this case also, the model is only be tested for the validation period (2003 to 2005). In addition to this, the Pearson's correlation factor, R, between the actual rainfall and forecasted rainfall generated by the Rain Forecast Model (RFM) has also been determined.

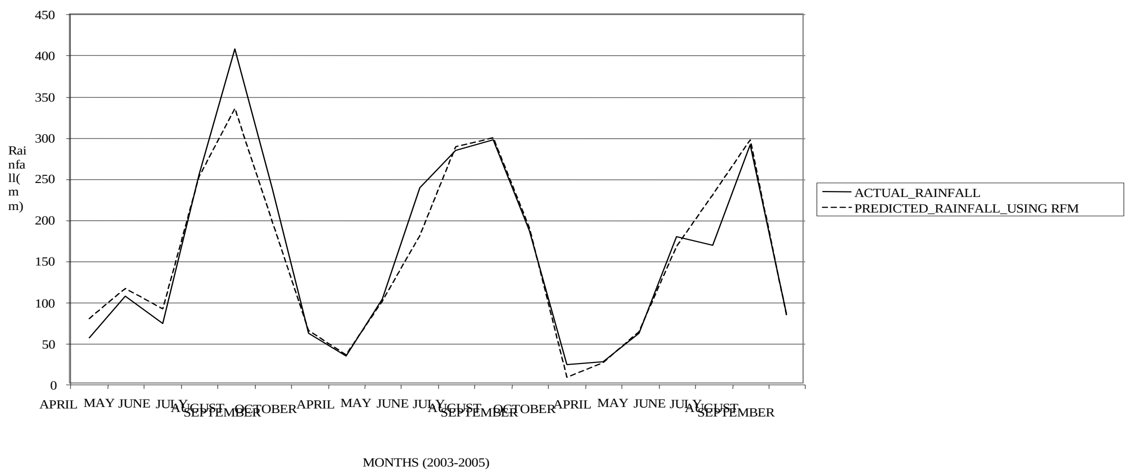

Figure 8: Plot of Actual Rainfall And Forecasted Rainfall Using Validation Data of 2003 – 2005

The correlation factor, here, is a measure of how well trends in the forecasted values using the developed Rain Forecast Model (RFM) follow trends in the actual data. The Root Mean Square Error (RMSE) is 54.99, while the correlation factor, R is 0.88. This is a good indication of the efficiency of the model

6. Conclusion

Weather is very difficult to predict despite the data and information that may be available. This is because weather can change very quickly (e.g. if the wind changes direction slightly, the rain may pass over and fall elsewhere) which means the forecast can quickly be out of date. In general, climate and rainfall in particular, are highly non-linear phenomena in nature. Soft Computing techniques are well suited for developing the prediction model as a result of the non-linear nature of rainfall. The idea of using a Neuro-Fuzzy approach came both, from Neural Nets' abilities to learn and generalize from sets of historical patterns, and the complexities of the problem (unpredictability of rainfall), which is more suitable to be modeled using fuzzy techniques.

The Rain Forecast Model (RFM) used as its inputs the weather variables and in order to forecast into the future, these weather variables were forecasted using a hybrid statistical technique comprising of Moving Averages and Exponential Averages. The forecast period is between 2003 and 2010. A fuzzy time series technique was also applied to forecast for the rainfall so as to have a basis for comparison with the developed Rain Forecast Model (RFM).

Two statistical tests were carried out in order to measure the performance of the developed model and that of the fuzzy time series method namely Root Mean Squared Error (RMSE) and Correlation Factor (R). The developed model was tested with the validation data of 2003 to 2005 and the following performance function results were obtained: RMSE=28.02 and R=0.97. This formed the basis for comparing the performance of the developed model when subjected to the forecasted weather variables for the same period. The performance function results obtained were: RMSE=54.99 and R=0.88. Considering the nature of the problem, that is the unpredictability of rainfall, these are indicators of a good model.

References

Easton, V. J. and Mccoll, J. H. (1997), "Time Series Data", STEPS' Statistics Glossary v1.1.Accessed from **www.stats.gla.ac.uk/steps/glossary/time_series.html**

Grodsky, S. A., and Carton, J. A., (2002), "Effect of Intra-seasonal Wind Fluctuations in The West Africa Monsoon on Air-Sea Fluxes" Accessed from **www.soc.soton.ac.uk/JRD/MET/WGASF/workshop/WABSTRACTS/Grodsky.php3**

Inform GmbH (2001), fuzzyTECH® 5.5 User's Manual, Inform Software Corporation, pp1 - 147.

Monks, J. G. (1996), Schaum's Outline of Theory and Problems of Operations Management, 2.ed, McGraw-Hill Inc, pp39 – 43.

Mu'azu, M. B. (2006), "Forecasting and Modeling Statistical Phenomena using NeuroFuzzy Logic: A Case Study of Rainfall Forecasting for Zaria", Unpublished Ph.D Dissertation, Department of Electrical Engineering, Ahmadu Bello University, Zaria.

Roberts, P. (2004), "Moving Averages: The Heart of Trend Analysis", Alchemist, Journal of The London Bullion Market Association, Issue 33, pp12-14.www.lmba.org.uk/publications/alchemist/alch33_averages.pdf

Rojas, I. and Pomares, H. (2004), "Soft Computing Techniques for Time Series Forecasting", European Symposium on Artificial Neural Networks, Bruges, Belgium, ESANN'2004 Proceedings, pp93 – 102.

Von-Altrock, C.(1995), Fuzzy Logic and NeuroFuzzy Applications Explained, Prentice Hall PTR, pp29 – 319.