# FREQUENCY ANALYSIS OF RAINFALL FOR FLOOD CONTROL IN PATANI, DELTA STATE OF NIGERIA

## O. B. Okeke[1,*] and J. O. Ehiorobo[2]

[1, 2] DEPARTMENT OF CIVIL ENGINEERING, FACULTY OF ENGINEERING, UNIVERSITY OF BENIN, BENIN CITY, NIGERIA.
*E-mail addresses*: [1] *bonavis.benedict@gmail.com,* [2] *jacehi@uniben.edu*

## ABSTRACT

*The Niger Delta Region of Nigeria is within the mangrove forest region and is crisscrossed by series of streams and creeks. As a result of the high rainfall volume within this region there is a tendency for severe flooding to occur. These flood events have severe consequences on lives and properties. It is therefore necessary to carry out a flood frequency analysis in this area in order to predict the potentials for the occurrence of flooding and take necessary measures for their remediation. In this study, peak monthly rainfall for thirty three years from 1981 – 2013 were acquired for Patani area in the Niger Delta region of Nigeria from NIMET and CBN. Five probability distributions, namely Normal, Log normal, Log-Pearson, Gumbel and Foster's Type -1 were selected for frequency analysis of the precipitation data using return periods of 2, 5, 10, 25, 50, 100 and 200 years respectively. The predicted values were subjected to goodness of fit tests using ANFIS tool in the MATLAB 7.0 software. The results of the tests and analyses show that the Gumbel Distribution best describes the precipitation data for the region and can therefore be used for the prediction of flooding within the study area followed by the Foster's Type- I distribution. The Normal Distribution was on the other hand found the least suitable for describing the precipitation data for Patani area.*

*Keywords: Precipitation, Frequency Analysis, Flood Control, Probability Distribution, Return Period*

## 1. INTRODUCTION

Meteorological data have shown that rainfall pattern in Nigeria has changed in the past decades [1].The Intergovernmental Panel on Climate Change has linked the rise in sea level to climate change [2]. Between 1960 and 1970, a mean sea level rise of 0.462m was recorded along the Nigerian coastal water [3, 4]. Flooding of low-lying areas in the Niger-Delta region has been observed and Settlements within the coastal areas have been washed away by coastal erosion. In some places, especially in Forcados, some oil wells have been lost to the ocean due to erosion [5]. The inundation arising from the rise in sea level due to rainfall will increase problems of floods, intrusion of sea-water into fresh water sources and ecosystems, destroying such stabilizing systems as mangroves, and affecting agriculture, fisheries and general livelihoods [6, 7]. Coastal vegetation, especially the mangroves have been lost to coastal erosion in vast majority of the Niger Delta Area [8, 9]. The Niger Delta could lose over 15,000 square kilometres of land by the year 2100 with a one meter rise in sea level [10]. Moreover, it is predicted that Nigeria will lose about $9 billion as a result of the sea level rise [11] while at least 80% of the people of the Niger Delta will be displaced

due to the low level of the region [10]. It is often possible to forecast, with reasonable accuracy, this type of flooding due to the predictability of the tide and track ability of low pressure systems. The duration of this type of flooding is also limited by the cycle of the tides where drainage is available [12]

In statistical rainfall estimation, the rainfall frequency procedures have two purposes: the estimation of design rainfall depths and the assessment of the rarity of observed rainfall events [13]. Design rainfalls are required principally for river flood estimation as they are important component in the design for flood defences, bridges, culverts, and reservoir spillways. Many flood estimates depend on good rainfall frequency information because rainfall records tend to be more plentiful and longer than river flow records especially in Nigeria where river discharge data is scarce [14].

The selection of a specific statistical distribution for describing the population behaviour of a given variable is seldom a simple problem. The specific objectives of the study are to fit different probability distribution (normal, log normal, log pearson, Gumbel and Foster) to the annual peak rainfall data, Predict design for the following return periods (T= 2yrs, 5yrs, 10yrs, 25yrs, 50yrs,

**\* Corresponding author, tel: +234 – 803 – 221 – 7426**

100yrs and 200 years and hence obtain the probability distribution that gives the best fit.

Rainfall prediction in mm is a typical nonlinear regression problem, in which several probability distributions are to be tested and the distribution with the least deviation from the observed maximum rainfall is selected as the best of fit and will be best suitable for rainfall prediction with respect to the study area [15].

## 1.1 Probability Distribution

The rainfall of return periods of 2, 5, 10, 25, 50, 100, and 200 years respectively can be computed using probability distributions. The return period or recurrence interval (T) is the average number of years during which a flood of given magnitude will be equalled or exceeded once and is computed by the Weibull formula given as [16].

$$T = \frac{n+1}{m} \qquad (1)$$

Where n is the number of events, i.e., years of record, m is the order or rank of the event (flood item) When the flood magnitudes (items) are arranged in the descending order (m= 1 for the highest flood, m= n for the lowest flood), T is the recurrence interval (T= n-yr for the highest flood, T= 1 yr for the lowest flood, by California method, 1923).

The probability of occurrence of a flood (having a recurrence interval T-yr) in any year, i.e., the probability of exceedance, is

$$P = \frac{1}{T} \qquad (2)$$

or the percent chance of its occurrence in any one year, i.e., frequency (F) is

$$F = \frac{1}{T} \times 100 \qquad (3)$$

and the probability that it will not occur in a given year, i.e., the probability of non-exceedance,

$$(P')\text{is } P' = 1 - P \qquad (4)$$

### 1.1.1 Normal Distribution

In the application of this distribution, the maximum rainfall series are ranked in a descending order of magnitude. Statistical parameters such as the mean, standard deviation, frequency factor, z and predicted rainfall in mm are determined using the expressions given below.

$$z = w - \frac{2.515517 + 0.80285w + 0.010328w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3} \qquad (5)$$

$$w = \sqrt{\ln\left(\frac{1}{P^2}\right)}, 0 < P \leq 0.\,5\text{,,} \qquad e \text{ is exponential.} \quad (6)$$

In (6), z is standard normal variable. The value $z$ computed using the above equation is given a negative sign; the error in using the equation to estimate the frequency factor is less than 0.00045 [17].

The Statistical Parameters are;

$$\text{Mean} \Rightarrow \bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_{max} \qquad (7)$$

$$\text{Standard Deviation} \Rightarrow \sigma_Q = \sqrt{\frac{\sum_{i=1}^{n}(R_{max} - \bar{R})^2}{n-1}} \qquad (8)$$

Where; $\bar{R}$ is mean of the annual maximum rainfall (mm), $R_{max}$ is annual maximum rainfall (mm), $\sigma_Q$, standard deviation of annual maximum rainfall, n is number of observations. The frequency factor or standard normal variable, z, can be approximated by the empirical relation given in [17]. The predicted rainfalls at various return periods are determined using the mathematical expression:

$$R_T = \bar{R} + z\sigma_Q \qquad (9)$$

### 1.1.2 Lognormal Distribution

In the application of this distribution, the maximum rainfall series are ranked in a descending order of magnitude, the logarithm of each rainfall were computed, statistical parameters such as the mean, standard deviation, frequency factor, z and predicted rainfall in mm are determined using the expressions given below. These predictions are then fitted with actual rainfall for each year.

$$y = logR_{max} \qquad (10)$$

Where, $y$ is in $mm$, as previously defined.

The intermediate variables and standard normal variable corresponding to the ranked annual maximum rainfall were previously determined in equations (5).

The statistical variate and the predicted rainfall are:

$$y_T = \bar{y} + z\sigma_y \qquad (11)$$

$$R_T = 10^{(\bar{y}+z\sigma_y)} \qquad (12)$$

In (11) and (12), $y_T$ is variate of the annual maximum rainfall at return period T (years); $\bar{y}$ is the logarithm mean of annual maximum rainfall (mm); $R_T$, z and $\sigma_y$ have been previously defined. A plot of predicted rainfall against return period, T was made.

### 1.1.3 Log Pearson Type III Distribution:

The exceedance and non-exceedance probability, $(logR - \overline{logR})^2$ and $(logR - \overline{logR})^3$ are calculated on different columns. Statistical parameters such as the mean, standard deviation, frequency factor, z and predicted rainfall in mm are determined using the expressions given in equation (13). These predictions are then fitted with actual rainfall for each year. The Pearson Type III distribution is also called the 'three-parameter Gamma distribution', the frequency factor depend on both the return period, T, and the skewness coefficient, $g_x$. If the skewness coefficient falls between -1 and +1, approximate values of the frequency factor for

the Gamma/Pearson Type III distribution, $K_T$, can be estimated using the relation given in [17] as:

$$K_T = \frac{1}{3k}\{[(\acute{x}_T - k)k + 1]^3 - 1\} \qquad (13)$$

Where $\acute{x}_T = z$ (previously defined) is the standard normal variate corresponding to the return period, T and k is related to the skewness coefficient by:

$$k = \frac{g_x}{6} \qquad (14)$$

When the skewness, $g_x$, is equal to zero, then $K_T = \acute{x}_T$ and the Log Pearson Type III distribution is identical to the Lognormal distribution [17].

The coefficient of skewness $g_x$ [18] is given as:

$$g_x = \frac{n\sum_{i=1}^{n}(y - \bar{y})^3}{(n-1)(n-2)s_y^2} \qquad (15)$$

The value of $y$ with return period $T$, $y_T$ is given by:

$$y_T = \bar{y} + K_T.s_y \qquad (16)$$

Where $\bar{y}$ and $s_y$ are as previously defined.

$K_T$ is the frequency factor (z) of the Pearson Type III distribution with return period T given by[18]. The value of the original variable, x, with the return period T, $x_T$, is then given by:

$$x_T = lny_T^{-1} = e^{y_T} \qquad (17)$$

### 1.1.4 Gumbel Distribution

In the application of this distribution, the exceedance and non-exceedance probability are determined on different columns. Statistical parameters such as the mean, standard deviation, frequency factor, k and predicted rainfall in mm are determined using the expressions given below. These predictions are then fitted with actual rainfall for each year. The probable rainfall for a given return period is given as [19];

$$R_T = \bar{R} + K\sigma \qquad (18)$$

$$y_T = -lnln\left(\frac{T}{T-1}\right) \qquad (19)$$

$$k = \frac{y_T - y_n}{\sigma_n} \qquad (20)$$

$$T = \frac{1}{P} = \frac{1}{1 - e^{-e^{-y}}} \qquad (21)$$

$$y = \frac{(R - \bar{R}) + 0.45\sigma}{0.7797\sigma} \qquad (22)$$

$$P' = 1 - P = 1 - e^{-e^{-y}} \qquad (23)$$

Here, $R_T$ is the probable rainfall with a return period of T years, $P'$ is probability of non-exceedance, $k$ is frequency factor, and $y$ is a reduced variate. $y_n, \sigma_n$ = expected mean and standard deviations of reduced extremes to be found from Gumbel's table of reduced mean ($y_n$) and reduced standard deviation ($\sigma_n$) as functions of sample size n as given in [20]. For desired $T$ and the number of years of record n, the value of $K$ can be directly taken from table of Frequency factor (K) for Gumbel's method (extremal value: Type-I distribution) as given in [21, 22].

### 1.1.5 Foster's Type-I distribution

In the application of this distribution, the maximum rainfall series are ranked in a descending order of magnitude. Statistical parameters such as the mean (eqn. 7), standard deviation (eqn. 8), frequency factor, k and predicted rainfall in mm were determined using the expressions given below.

$$C_v = \frac{\sigma}{\bar{R}} \qquad (24)$$

$$C_s = \frac{\sum_{i=1}^{n}(R - \bar{R})^3}{(n-1)\sigma^3} \qquad (25)$$

Note that $C_s = 2C_v$

$C_s$ adjusted to allow for the period of record [23]

$$\bar{C}_s = C_s\left(1 + \frac{k}{n}\right) \qquad (26)$$

Where $\sigma$ is the standard deviation of data, $C_v$ and $C_s$ are skewness values, $\bar{R}$ is mean rainfall in mm.

Foster type I, in the table for Skew Curve Factors (K values) for Foster's Type I curve as given in [20] are used to determine the value of k using $\bar{C}_s$, $k$ is modified by Foster type III, in table for Skew Curve Factors (K values) for Foster's Type-III curve. $R_T$ is computed as;

$$R_T = \bar{R} + K\sigma \qquad (27)$$

Here $R_T$ is the probable rainfall with a return period of T years, $\bar{R}$ is the mean rainfall

From the value of skewness, $C_s$, the values of the factor K giving the variations of the peak rainfall from the mean rainfall at various percentages of time (i.e., probability $P\%$) can be obtained by referring to Foster tables. From this, the flood frequency curve can be drawn; and the probability and recurrence interval ($T = 1/P \times 100$) of a desired flood magnitude can be read off from the curve.

### 2. DESCRIPTION OF STUDY AREA

Patani is the headquarters of the Local Government Council in Delta State within the Niger Delta region of Nigeria. It lies between Latitude 05°6'N to 05°21'N and Longitude 06°00'E to 06°23'E. Patani is a small community in delta state having an area of 217km² and a population of 67,707 based on the 2006 population census. The Local Government is bounded in the North by Ndokwa East LGA, Bomadi LGA in the South, Ughelli North LGA in the West, and then Sagbama Local Government Area of Bayelsa State to the East. The major economic activities are fishing, farming and petty trading.

The rainfall pattern for Patani area from historical hydrological data shows that peak rainfall occurs around the month of June, July, August and September. The months of July and August are the peak of the wet season. The months of October to February make up the dry season in this area. The location map of the study area is shown in Figure 1.
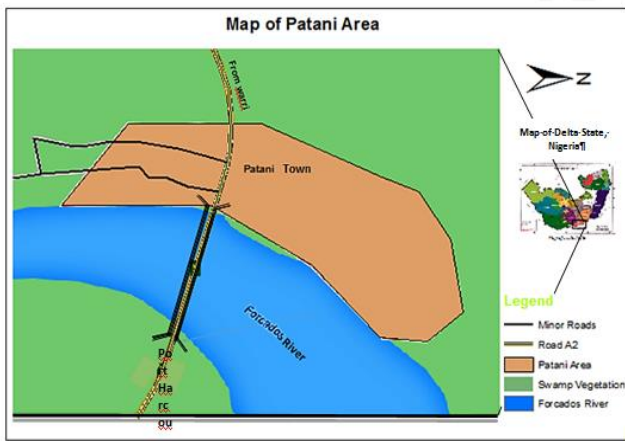
*Figure 1: Location map of Patani (google earth, satellite imagery)*

*Table 1: Monthly Rainfall Data for Patani from 1981 – 2013 arranged in descending order of magnitude (NIMET, CBN Bulletin, 2014)*

| RANK | YEAR | AVERAGE |
|---|---|---|
| 1 | 1999 | 283.0500 |
| 2 | 1995 | 278.1500 |
| 3 | 2002 | 276.0167 |
| 4 | 1997 | 268.9417 |
| 5 | 2008 | 259.4000 |
| 6 | 1990 | 258.2333 |
| 7 | 2004 | 255.3333 |
| 8 | 1992 | 255.0250 |
| 9 | 1982 | 254.7417 |
| 10 | 2006 | 252.5917 |
| 11 | 1993 | 250.8833 |
| 12 | 1985 | 248.0667 |
| 13 | 2007 | 245.2141 |
| 14 | 1991 | 243.3917 |
| 15 | 1986 | 242.9833 |
| 16 | 1994 | 233.9167 |
| 17 | 1987 | 231.3000 |
| 18 | 2009 | 231.2583 |
| 19 | 2012 | 230.8333 |
| 20 | 2013 | 230.1750 |
| 21 | 2000 | 226.8500 |
| 22 | 1988 | 225.7333 |
| 23 | 1996 | 225.4750 |
| 24 | 2011 | 225.3750 |
| 25 | 1984 | 224.7917 |
| 26 | 2010 | 215.1000 |
| 27 | 1989 | 214.2917 |
| 28 | 1983 | 210.1333 |
| 29 | 1998 | 204.0417 |
| 30 | 2001 | 199.1833 |
| 31 | 2005 | 197.3833 |
| 32 | 1981 | 195.8250 |
| 33 | 2003 | 194.1167 |

## 2.1 Methodology

Rainfall data used in this study were obtained from the Nigerian Meteorological Agency, NIMET and annual bulletins published by the Central Bank of Nigeria (CBN). In this study, peak monthly rainfall for thirty three years from 1981 – 2013 were acquired for Patani area in the Niger Delta region of Nigeria from NIMET and CBN. Five

probability distributions, namely Normal, Log normal, Log-Pearson, Gumbel and Foster's Type -1 were selected for frequency analysis of the rainfall data using return periods of 2, 5, 10, 25, 50, 100 and 200 years respectively. Table 1 below shows the average monthly rainfall for the area from 1981 – 2013 ranked in a descending order of magnitude.

There are several probability distributions but the five probability distributions were employed due to their wide use and acceptability for hydrological data analysis. These distributions were further used in predicting rainfall for all 33 years and comparison was made between the observed maximum and the model predicted values for each year. These predicted values were also plotted against return periods. Goodness of fit test was carried out using the Adaptive Neuro-Fuzzy Inference System (ANFIS).

The Adaptive Neuro-Fuzzy Inference System, ANFIS, a tool in MATLAB was utilised in the goodness of fit test to determine and rank from the various Probability Distribution the best distribution for the rainfall analysis. The ANFIS tool utilises the Root Mean Square Error method in error analysis. The best rainfall frequency distribution was then selected for the rainfall prediction for various return periods of 2, 5, 10, 25, 50, 100 and 200 years respectively.

## 2.2 Rainfall Data modelling using ANFIS tool
### 2.2.1 Partitioning Data

The data set is obtained from Table 2. The dataset is then partitioned into a training set (row 1-16 indexed distributions) and a checking set (row 17-33 indexed distributions).

trn_fit=goodness_fit_data(1:16,:);

chk_fit=goodness_fit_data(17:end,:);

### 2.2.2 Input Selection

The function exhsrch performs an exhaustive search within the available inputs to select the set of inputs that most predicts rainfall in Patani. The first parameter to the function specifies the number of input combinations to be tried during the search. Essentially, exhsrch builds an ANFIS model for each combination and trains it for one epoch and reports the performance achieved. In this analysis, exhsrch is used to determine the one most influential input attribute in predicting the output.

exhsrch(1, trn_fit, chk_fit, input_names);

For better generalization, we always prefer a model with a simple structure. Therefore we will stick to the two-input ANFIS for further exploration.

new_trn_fit = trn_fit(:, [input_index, size(trn_fit,2)]);

new_chk_fit = chk_fit(:, [input_index, size(chk_fit,2)]);

We then extract the selected input attributes from the original training and checking datasets as can be seen in Figure 4.

### 2.2.3 Training the ANFIS Model

The function exhsrch only trains each ANFIS for a single epoch in order to be able to quickly find the right inputs. Now that the inputs are fixed, we can spend more time on ANFIS training (100 epochs).The genfis1 function generates an initial FIS from the training data, which is then fine-tuned by ANFIS to generate the final model.

in_fismat = genfis1(new_trn_fit, 2, 'gbellmf');

[trn_out_fismat trn_error step_size chk_out_fismat chk_error] = ...

anfis(new_trn_fit, in_fismat, [100 nan 0.01 0.5 1.5], [0,0,0,0], new_chk_fit, 1);

ANFIS returns the error with respect to training data and checking data in the list of its output parameters. The plot of the errors provides useful information about the training process. This plot is presented in Figure 6.

```
>> [x, y] = min(chk_error);
>> plot(1:100, trn_error, 'g-', 1:100, chk_error, 'r-', y, x, 'ko');
>> title('Training (green) and checking (red) error curve');
>> xlabel('Epoch numbers');
>> ylabel('RMS errors');
```

## 3. RESULTS ANALYSIS AND DISCUSSION

### 3.1 Results Analysis

The plot in Figure 2 shows the summary of the five probability distribution results against various return periods. This revealed that there was very little difference between the various distributions for each year with Foster, Gumbel and Log-Pearson showing relatively close and higher predictions when compared to the Log-normal and normal distributions.
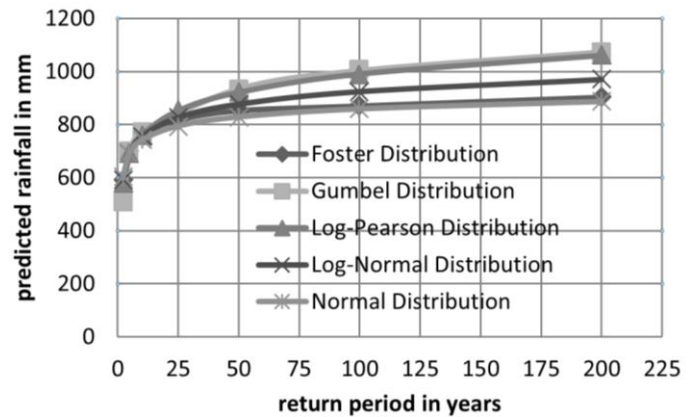


Figure 2: Estimated rainfall against various return periods

### 3.1 Goodness of Fit Tests

#### 3.1.1 Analysis of Probability Distributions Using ANFIS

The training data are presented in Table 2. It contains as inputs, rainfall predictions of five different probability distribution methods ranked in a descending order of magnitude and as output the observed maximum rainfall recorded from 1981 to 2013. The dataset were then partitioned into a training set (row 1-16 indexed distributions) and a checking set (row 17-33 indexed distributions).

Table 2: ANFIS training data of five probability distribution methods for Patani

| YEAR | INPUTS | | | | | OUTPUT |
|------|--------|--------|--------|--------|--------|--------|
|      | NORMAL DISTRIBUTION | LOG NORMAL DISTRIBUTION | LOG PEARSON DISTRIBUTION | GUMBEL DISTRIBUTION | FOSTER DISTRIBUTION | OBSERVED MAX. RAINFALL |
| 1993 | 810.9198 | 849.6425 | 877.0318 | 910.7934 | 847.5280 | 868.4000 |
| 2004 | 774.7338 | 798.0157 | 802.9396 | 827.6864 | 794.6653 | 795.7000 |
| 1992 | 750.9944 | 765.8628 | 769.2876 | 784.9327 | 774.4440 | 758.3000 |
| 1990 | 732.6185 | 741.8665 | 739.2307 | 756.6969 | 755.9988 | 733.6000 |
| 1986 | 717.2693 | 722.3999 | 714.6717 | 748.4662 | 749.7138 | 726.4000 |
| 2012 | 703.8664 | 705.8199 | 698.8003 | 692.5662 | 695.0348 | 677.5000 |
| 2011 | 691.8149 | 691.2368 | 684.0464 | 686.9648 | 691.2923 | 672.6000 |
| 1983 | 680.7487 | 678.1115 | 660.5466 | 684.9071 | 689.8666 | 670.8000 |
| 1997 | 670.4248 | 666.0914 | 642.9923 | 679.6486 | 686.1241 | 666.2000 |
| 1999 | 660.6715 | 654.9315 | 629.1412 | 663.4159 | 673.5779 | 652.0000 |
| 1985 | 651.3619 | 644.4538 | 618.0628 | 660.6724 | 671.3324 | 649.6000 |
| 2008 | 642.3981 | 634.5237 | 608.9218 | 644.4397 | 657.0752 | 635.4000 |
| 1988 | 633.7018 | 625.0361 | 601.6375 | 643.5251 | 656.2198 | 634.6000 |
| 2002 | 625.2079 | 615.9063 | 595.1221 | 635.9804 | 649.0200 | 628.0000 |
| 1987 | 616.8607 | 607.0641 | 589.6902 | 631.1791 | 644.2795 | 623.8000 |
| 1994 | 608.6107 | 598.4497 | 584.8103 | 623.2914 | 636.1886 | 616.9000 |

| YEAR | INPUTS | | | | | OUTPUT |
|---|---|---|---|---|---|---|
| | NORMAL DISTRIBUTION | LOG NORMAL DISTRIBUTION | LOG PEARSON DISTRIBUTION | GUMBEL DISTRIBUTION | FOSTER DISTRIBUTION | OBSERVED MAX. RAINFALL |
| 1995 | 600.4122 | 590.0101 | 580.6360 | 612.6601 | 624.7472 | 607.6000 |
| 1991 | 608.6107 | 598.4497 | 557.5674 | 603.6292 | 614.5178 | 599.7000 |
| 2005 | 616.8607 | 607.0641 | 537.7061 | 595.2843 | 604.7160 | 592.4000 |
| 1998 | 625.2079 | 615.9063 | 520.4290 | 576.1937 | 582.3100 | 575.7000 |
| 2009 | 633.7018 | 625.0361 | 505.4562 | 576.0794 | 582.1853 | 575.6000 |
| 1982 | 642.3981 | 634.5237 | 491.9004 | 567.1628 | 572.1741 | 567.8000 |
| 2007 | 651.3619 | 644.4538 | 479.9850 | 547.1577 | 549.1890 | 550.3000 |
| 2010 | 660.6715 | 654.9315 | 469.3163 | 542.9280 | 544.3237 | 546.6000 |
| 2006 | 670.4248 | 666.0914 | 459.7059 | 521.8941 | 520.3406 | 528.2000 |
| 1996 | 680.7487 | 678.1115 | 451.3946 | 513.3205 | 510.8908 | 520.7000 |
| 1984 | 691.8149 | 691.2368 | 443.1203 | 508.0620 | 505.2771 | 516.1000 |
| 2013 | 703.8664 | 705.8199 | 435.9035 | 486.7994 | 480.5454 | 497.5000 |
| 2000 | 717.2693 | 722.3999 | 429.2895 | 457.3062 | 450.4718 | 471.7000 |
| 1989 | 732.6185 | 741.8665 | 423.2080 | 435.8150 | 433.9492 | 452.9000 |
| 2003 | 750.9944 | 765.8628 | 417.6087 | 416.6101 | 416.4842 | 436.1000 |
| 1981 | 774.7338 | 798.0157 | 412.4144 | 387.5741 | 405.3681 | 410.7000 |
| 2001 | 810.9198 | 849.6425 | 407.5970 | 322.9862 | 389.6608 | 354.2000 |
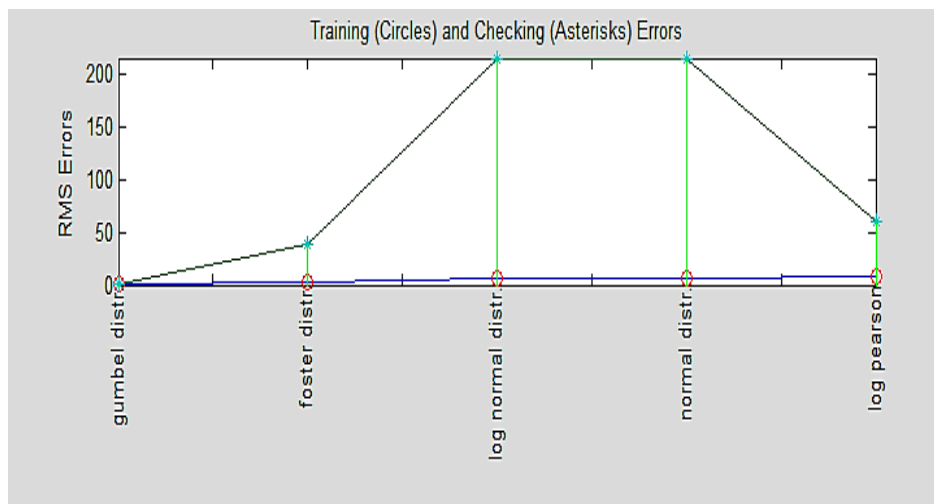


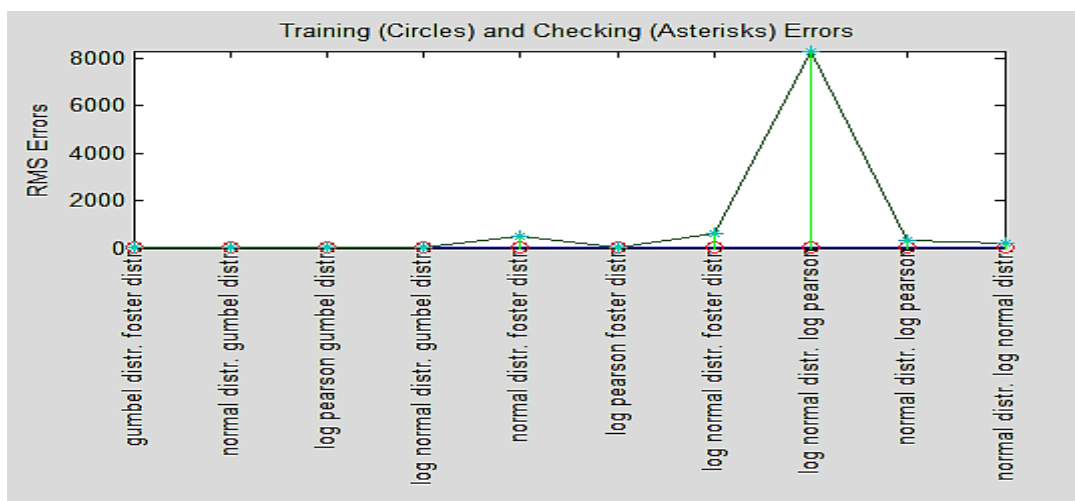Figure 3: Every input variable's influence on observed maximum rainfall



Figure 4: All two input variable combinations and their influence on observed maximum rainfall.

The input variable to the extreme left of Figure 3 has the least error or in other words the most relevant with respect to the output.

Figure 3 and results from the function clearly indicate that the input attribute 'Gumbel's distribution' is the most prominent. The training and checking errors are comparable, which implies that there is no over fitting. This means we can push a little further and explore if we can select more than one input attribute to build the ANFIS model.

Intuitively, we can simply select 'Gumbel's distribution' and 'Foster's distribution' directly since they have the least errors as shown in the plot. However, this may not necessarily be the optimal combination of two inputs that result in the minimal training error. To verify this, we can use exhsrch to search for the optimal combination of two input attributes as shown in Figure 4. The results from exhsrch indicate that 'Gumbel's distribution' and 'Foster's distribution' form the optimal distributions of two input attributes as shown in Figures 4 and 5 respectively.

### 3.1.2 Analysis using the ANFIS Model

The variable chk_out_fismat represents the plot of the ANFIS model at the minimal checking error during the training process. The input-output surface of the model is shown in the Figure 5 below.
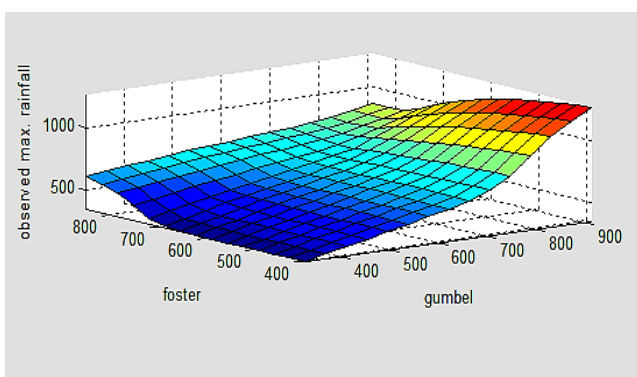


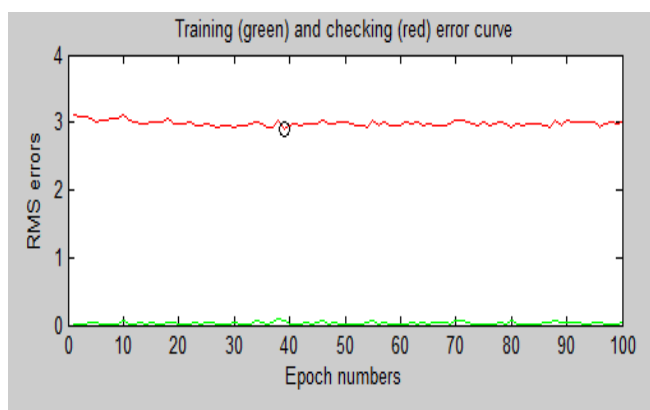*Figure 5: Input-Output surface for trained FIS*



*Figure 6: ANFIS training and checking errors*

chk_out_fismat = setfis(chk_out_fismat, 'input', 1, 'name', 'gumbel');
chk_out_fismat = setfis(chk_out_fismat, 'input', 2, 'name', 'foster');
chk_out_fismat = setfis(chk_out_fismat, 'output', 1, 'name', 'observed max. rainfall');
% Generating the FIS output surface plot
gensurf(chk_out_fismat)

The surface plot generated in Figure 5 shows the graphical level of performance of the predicted values of the best two probability distributions in contrast with the maximum observed values presented in Table 2. The undulated surface describes the deviation of values predicted from the actual after training was done.

Figure 6 shows the error curves for 100 epochs of ANFIS training. The green curve gives the training errors and the red curve gives the checking errors. The minimal checking error occurs at about epoch 38, which is indicated by a circle. Notice that the checking error curve goes up after 39 epochs, indicating that further training over fits the data and produces worse generalization

### 3.2 Discussion

Examination of the summary of the five probability distribution results against the annual maximum observation reveals that in many cases there was very little difference between the various distributions for each year as the goodness of fit assessment for all 33 years indicated that no one distribution ranked consistently best.

The checking set partitioned was used to validate the outcome of the training set for better accuracy as the Root Mean Square Error revealed that the Gumbel distribution performed best in predicting rainfall in Patani as this was clearly illustrated in Figure 3. Further combination analysis showed that the Gumbel and Foster distribution can be selected in the order outlined.

In Figure 3, the legends clearly describe the training and checking errors as could be seen. The training and checking errors are comparable, which implies that there is no overfitting which simply means that these two errors can be easily distinguished in most of the distributions apart from Gumbel where there is a slight overlap. The reverse can be seen in Figure 4 which is characterised by overfitting of these errors for most of the distribution combinations making them not easily distinguishable; ANFIS then ranks them from left to right. As a result, a further push will be unnecessary.

### 4. CONCLUSION AND RECOMMENDATION

It is of high probability that the predicted maximum rainfall in Patani occurs in the months of July and August.

Examination of the five probability distribution results against the annual maximum observation reveals that in many cases there was very little difference between the various distributions for each year. The goodness of fit assessment for all 33 years also indicated that no one distribution ranked consistently best. However, the overall ranks for the combined years show that Gumbel's distribution was best to describe the peak rainfall.

The Gumbel's distribution, from the analysis is the best probability distribution for the Patani area and should be employed in future rainfall prediction. The Foster's distribution also proved to show slight deviation from the observed maximum rainfall after the Gumbel's distribution.

Information derived from this study can be used by government agencies in developing flood protection infrastructures in Patani to improve the economic condition of the people. It can also be used for flood risk analysis and flood hazard management.

## 5. REFERENCES

[1]  Oladipo, E. O. "An Indication of Abrupt Change of Rainfall and Its Potential Impact on Energy Development in Nigeria". In: Umolu, J. C. (ed). *Global Climate Change: Impact on Energy Development*, DAMTECH Nigeria Limited, Nigeria. 1995.

[2]  Camille Parmesan and Gary Yohe "A globally coherent fingerprint of climate change impacts across natural systems" Nature, *International Weekly Journal of Science* 421, 37-42 2003.

[3]  Udofa, I. M. and Fajemirokun, F. A., "On a Height Datum for Nigeria." *Proceedings International Symposium on Geodetic Measurements and Computations*, Ahmadu Bello University, Zaria, Nigeria, 1978.

[4]  Ehiorobo, J. O.,Izinyon, O.C. and Ilaboya F. R. "Effect of Climate Change on the River Flow regimes in the Mangrove and Tropical Rainforest of the Niger Delta region of Nigeria". *Research Journal in Engineering and Applied Sciences* (RJEAS), USA, Vol. 2(4), , pp 256-261. 2013.

[5]  Awosika, L. F., "Impacts of global climate change and sea level rise on coastal resources and energy development in Nigeria". In: Umolu, J. C. (ed). *Global Climate Change: Impact on Energy Development,*. DAMTECH Nigeria Limited, Nigeria. 1995.

[6]  Okali, D. and Eleri, E. O., "Climate Change and Nigeria: A guide for Policy Makers", *The publication of the Nigerian Environmental Study Action Team* (NEST) 2004.

[7]  Ehiorobo, J. O. and Izinyon O. C. "Monitoring of Soil Loss from Erosion using Geoinformatics and Geotechnical Engineering Methods". *Journal of Civil Engineering and Architecture*, David Publishing, USA, Vol. 7 (1), pp 78 – 84. 2013.

[8]  T. K.S. Abam "Regional hydrological research perspectives in the Niger Delta" *Hydrological Sciences Journal*, Volume 46, 2001 – Issue 1, Page 13-25, 2009.

[9]  C. Kuenzer, S. van Beijma, U. Gessner, S. Dech, "Land surface dynamics and environmental challenges of the Niger Delta, Africa: Remote sensing-based analyses spanning three decades (1986–2013)" *Applied Geography*, Volume 53, September 2014, Pages 354-368, 2014.

[10]  Awosika, L. F., French, G. T., Nicholls, R. J. and Ibe, C. E., "The impact of Sea level Rise on the Coastline of Nigeria", In: *Proceedings of IPCC Symposium on the Rising Challenges of the Sea.* Magaritta, Venezuela, pg 14-19. 1992.

[11]  Enete, Ifeanyi C, Ezenwanji, E. E. "Implications of climate variability on water resources of Nigeria: A review" *Journal of Geography and Regional Planning*, Vol. 4(13), pp. 678-682, 4 November, 2011.

[12]  Dance, K. W. and H.R.N. Hynes, "Some Effects of Agricultural Land Use on Stream Insect Communities". *Environ. Pollution.* (Series A), Vol.22, 14-28. 1988.

[13]  Dawei Han, "Concise Hydrology". Dawei Han & Ventus Publishing Aps, ISBN 978-87-7681-536-3. 2010.

[14]  T. A. Ewemoje, O. S. Ewemooje "Best distribution and plotting positions of daily maximum flood estimation at Ona River in Ogun-Oshun river basin, Nigeria" Agricultural Engineering International: *CIGR Journal*, Vol 13, No 3 2011.

[15]  Izinyon, O. C., Ihimekpen, N. and Igbinoba, G.E., "Flood Frequency Analysis of Ikpoba River Catchment at Benin City Using Log-Pearson Type III Distribution", *Journal of Emerging trends in Engineering and Applied sciences* (JETEAS), 2 (1): pp. 50-55. 2011.

[16]  Weibull W. "A Statistical Theory of the Strength of Materials" *The Royal Swedish Institute for Engineering Research,* part I and II 1939.

[17]  Chin, D. A., *"Water Resources Engineering"*. 2nd Edition, Pearson Prentice Hall, Upper Saddle River, New Jersey, 2006.

[18]  Haan, C. T., *"Statistical Methods in Hydrology"*, Iowa State University Press. Ames, 1994.

[19]  Pearson, K., "Contributions to the Mathematical Theory of Evolution, II: Skew Variation in Homogeneous Material", *Philos. Trans. Royal Soc. London,* ARRAY Vol.186, , pg 343-414. 1895.

[20]  Raghunath, H. M. *"Hydrology: Principles, Analysis, Design"*. Revised Second Edition, New Age International (P) Ltd. Publishers, 2006.

[21]  Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J., Mykytha, E. F., "A probability distribution and its uses in fitting data", *Technometrics* Vol. 21, , pg 201-214. 1979.

[22]  Ehiorobo J. O. and Nosa Uso "Flood Frequency Analysis in the River Niger Basin at Onitsha Bridge Head". *Journal of Engineering Research,* UNILAG, Vol. 19 (1) pp 33 - 44. 2014.

[23]  Foster H. A., "Theoretical Frequency Curves and their Application to Engineering Problems". *Trans. ASCE*, Vol. 87, , pp. 142-173. 1924.