



Investigating the Non-Stationarity of Historical Data on Rainfall at Ilorin, North Central Nigeria

Sule BF, Ajala OF

Department of Civil Engineering, University of Ilorin, Ilorin

Abstract This study investigated the degree of non-stationarity in the monthly rainfall data of Ilorin meteorological station from 1960 – 2010 (51 years). The methodology adopted included trend test with regression analysis and Mann-Kendell trend analysis, Augmented Dickey Fuller (ADF) test, auto-correlation function (ACF), and spectral analysis. The results showed that the series contained non-significant negative trend. The Autocorrelation function (ACF) plot, and Partial Autocorrelation function (PACF) plot indicated several significant spikes at different lags which confirmed the non-stationarity in the data. The spectrum analysis however indicated that the historical rainfall series of Ilorin contains periodicities of 12 months, 6 months, and 4 months, all which are analysed to be statistically significant. Non-stationarity was therefore generally observed in the form of both monotonic trends and abrupt change over time. It is therefore recommended that non stationarity should be incorporated into subsequent hydrological analysis and design issues based on the historical rainfall data at Ilorin, north central Nigeria.

Keywords Non-Stationarity, Rainfall, Ilorin, Augmented Dickey Fuller (ADF), Partial Autocorrelation function (PACF)

Introduction

Nonstationary behavior of recent climate increases concerns among hydrologists about the currently used design rainfall estimates. Therefore, it is necessary to perform an analysis to confirm stationarity or detect nonstationarity of rainfall data to derive accurate design rainfall estimates for infrastructure projects and flood mitigation works.

Before engineering projects can be undertaken to address water management problems or to take advantage of opportunities for increased economic, ecological, environmental and social benefits, they must first be planned. This involves identifying various alternatives for addressing the problems or opportunities. Next, the various impacts of each proposed alternative need to be estimated and evaluated. A variety of optimization and simulation models and modelling approaches have been developed to assist water planners and managers in identifying and evaluating plans. This work introduces the science and art of testing stationarity of rainfall data in support of water resources planning and management. Its main emphasis is on the practice of developing and using models to address specific water resources planning and management problems. This is done in a way that provide relevant, objective and meaningful information to those who are responsible for making informed decisions about specific water resources management issues.

When a time series is not stationary in some ways, patterns may be hidden in the time series. If a significant trend or cycle can be identified, it can help explain the variance involved in the time series. Reduction of variance is particularly important if the model is to be used as a planning or forecasting tool. Furthermore, the trend or cycle mode (and its associated parameters) often imply physical meaning that can help to identify the cause/effect relationships. One may even attempt to relate a shift in trend and/or cyclic parameters identified from historical data to certain physical changes in the system. Unless a satisfactory physical model is available to test the system response to these changes, stationarity test may be the next best thing to cast some light on the complicated climatological system.

In general, observational and historical hydro-climatologic data are used in planning and designing water resources projects. There is an implicit assumption, so called stationarity implying time invariant statistical



characteristics of the time series under consideration in all water resources engineering works. Such an assumption can no longer be valid because of the changes in global climate as a result of the increase of greenhouse gases in the atmosphere [1]. This has of course, results in major problems (for example, dislocation and inefficiencies) in regional water resources management. However, some investigators [2-7] have reported evidence of trends (possibly due to anthropogenic influences) and long-term variability of climate.

A time series of hydrological data may exhibit jumps and trends owing to what Fatichi, et. al., 2009 [8] call inconsistency and non-homogeneity. Inconsistency is a change in the amount of systematic error associated with the recording of data. It can arise from the use of different instruments and methods of observation. Non-homogeneity is a change in the statistical properties of the time series, its causes can be either natural or man-made. These include alterations to land use, relocation of the observation station, and implementation of flow diversions. The tests for stability of variance and mean verify not only the stationarity of a time series, but also its consistency and homogeneity. In the basic data-screening procedure, these two tests are reinforced by a third one, for absence of trend.

Approximately 0.75 °C global warming has been detected over the last 100 years [3]. This warming cannot be explained by natural variability alone. The main reason for the current global warming is human activities resulting in extensive greenhouse gas emissions to the atmosphere (IPCC 2007). A major question, in the context of global warming, is related to the extreme rainfall events producing floods and droughts. Increases in frequency and magnitude of extreme precipitation events have already been observed in the rainfall records of many regions, irrespective of the mean precipitation trends.

IPCC (2007) [9] reported that the intensity and frequency of the rainfall events are very likely to increase in the future with the exception in the regions that show very significant decreases in rainfall. An increase in frequency and magnitude of extreme precipitation events questions the stationarity in climate, which is one of the main assumptions of frequency analysis of extreme rainfalls. Possible violation of stationarity in climate increases concerns amongst hydrologists and water engineers about the currently used design rainfall estimates in infrastructure projects and flood mitigation works. Therefore, it is essential to perform analysis to confirm stationarity (or detect non-stationarity) of extreme rainfall data.

Khaliq et al. (2006) [10] reported that the classical notions of probability of exceedence and return period are no longer valid under non-stationarity. The possible violation of stationarity in climate increases concerns amongst hydrologists and water resources engineers about the accuracy of design rainfalls, which are derived from frequency analysis of extreme rainfall events under the stationary climate assumption. Erroneous selection of design rainfalls can cause significant problems for water infrastructure projects and flood mitigation works, since the design rainfalls are an important input for design of these projects. Therefore, there is a need to conduct a reliable statist test on the historical rainfall data in order to ascertain its non - stationarity and to give proper information to the concerned water resources professionals.

Olofintoye and Sule (2010) [11] carried out study on the impact of global warming on the rainfall and temperature for some selected cities in the Niger-Delta area of Nigeria. The stations considered were Owerri, Port-Harcourt, Calabar. The meteorological data used for the analysis are maximum and minimum temperature and rainfall. The time series of the data were analysed with the aim of detecting trends in the variables using the nonparametric Man-Kendall test to detect monotonic trends and the Sen's slope estimator was used to develop models for the variable. The study showed that there is evidence of non stationarity in Owerri and rainfall has significantly increased in Calabar over the years. Though the trends in Owerri and Port-Harcourt were not significant, the slope estimates showed a positive trend in the rainfall of the stations. It was concluded that water supply is sustainable under the current climate condition.

Furthermore, extreme rainfall non-stationarity of the storm durations from 6 min to 72 hours of Melbourne City in Australia for the period of 1925-2010 was investigated [3]. Stationary Generalized Extreme Value (GEV) models were constructed to obtain Intensity-Frequency-Duration relationships for the storm durations using data of two time periods: 1925-1966 and 1967-2010 after identifying the year 1967 as the change point year. Design rainfall estimates of the stationary models for the two periods were compared to identify the possible changes. Non-stationary GEV models, which were developed for storm durations that showed statistically significant extreme rainfall trends, did not show advantage over stationary GEV models. There was no evidence of non-stationarity according to stationarity tests, despite the presence of statistically significant extreme rainfall trends.

Study Area

The study area is Ilorin, the capital city of Kwara State, Nigeria and is located on latitude 8° 24'N and 8° 36'N and longitude 4° 10'E and 4° 36'E with an area of about 1188Km². It is situated at a strategic point between the densely populated southwestern and the sparsely populated middle belt of Nigeria. Ilorin is located in traditional zone between the deciduous woodland of the south and dry savanna of North of Nigeria. The climate of Ilorin is characterized by both wet and dry seasons. The temperature of Ilorin ranges from 33°C to 34°C from November



to January while from February to April; the value ranges between 34⁰C to 53⁰C (Ilorin Atlas, 1982). The mean monthly temperatures are very high varying from 25⁰C to 28.9⁰C. The diurnal range of temperature is also high in the area. The rainfall in Ilorin city exhibits greater variability both temporarily and spatially. The average total annual rainfall in the area is about 1234.4mm [12]. Relative humidity at Ilorin in the wet season is between 75 to 80% while it is about 65% in the dry season.

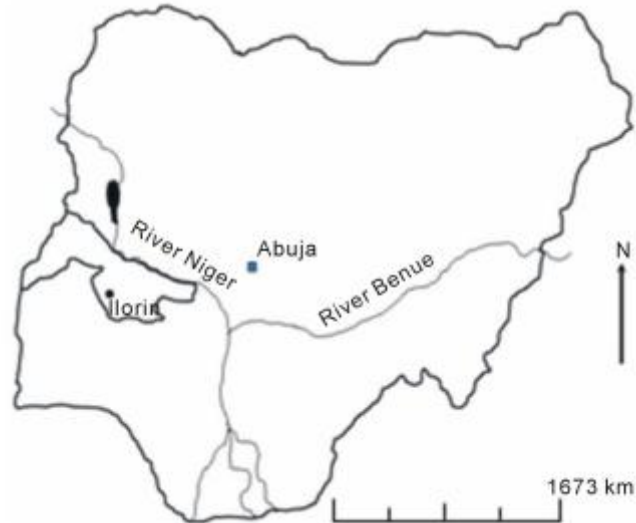


Figure 1: Map of Nigeria Showing the study area (Ilorin)

Theoretical Consideration

Trend Test

Many hydrological time series exhibit trending behavior or nonstationarity. In fact, the trending behavior is a type of nonstationarity. But in this study, they are treated differently. The purpose of trend test is to determine if the values of a series have a general increase or decrease with the time increase, whereas the purpose of stationarity test is to determine if the distribution of a series is dependent on the time.

An important task in hydrological modeling is to determine if there is the existence of any trend in the data is nonstationary [13]. On the other hand, the possible effects of global warming on water resources have been the topic of many recent studies [8,11, 15-16]. Thus, detecting the trend and stationarity in a hydrological time series may help in understanding links between hydrological processes and changes in the global environment. The presence of trends and nonstationarity is undesirable in further analysis. Therefore it is necessary to make sure whether there is presence of trend and nonstationarity or not, and if the presence of trend and nonstationarity is detected, the appropriate pre-processing procedure should be applied. The Least square method, the free hand method and the moving average method are commonly used to estimate trend in a time series. The least square method allows a linear trend to be fitted into a series, which may be used for forecasting. An advantage of the free hand method is that it is simple and does not involve rigorous computation; however the trend line drawn depends on the users' skills and accuracy. One major advantage of the moving average method is that cyclical, seasonal, and irregular variation may be eliminated, thus leaving the trend values, however data are lost at the beginning and end of the series, and moving average is affected by outliers or extreme values.

Mann-Kendall Trend Analysis

The Mann-Kendall test is a non-parametric test for identifying trend in time series data. It compares the relative magnitudes of sample data rather than the data values. The major benefit of this test is that the data need not conform to any particular distribution. Moreover data reported as non-detects can be included by assigning them a common value that is smaller than the smallest measured value in the data set. The procedure for the trend analysis assumes that there exists only one data value per time. When multiple data point exists for a single time period the median value is used. The data values are evaluated as an ordered time series and each data value is compared to all subsequent data values. The initial value of the Mann-Kendall statistic (S) is assumed to be 0 if there is no trend. If a data value from a later value from a later period is higher than a data value from an earlier time period S is incremented by 1. Conversely, if the data value from the data period is lower than a data value



sampled earlier S is incremented by 1. The net result of all the increments and decrements yield the final value of S . The expressions for the Mann-Kendall trend analysis are given in the Equation 1 to 3

Let $i = 1, 2, 3, \dots, n - 1$ and $j = i + 1, i + 2, i + 3, \dots, n$. Each data point T_i is used as a reference point and is compared with all the T_j data points such that:

$$\text{sign}(T) = \begin{cases} 1 & \text{for } T_j > T_i \\ 0 & \text{for } T_j = T_i \\ -1 & \text{for } T_j < T_i \end{cases} \quad (1)$$

The Mann-Kendall statistic (S) is given by

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(T_j - T_i) \quad (2)$$

Calculate the variance of S , σ by the following equation:

$$\sigma^2 = \frac{1}{18} [n(n-1)(2n+5) - \sum_{p=1}^g (t_p - 1)(2t_p + 5)] \quad (3)$$

where n is the number of data points, g is the number of tied groups. Computation of a normalized test statistic Z_s as follows:

$$Z_s = \begin{cases} (S - 1)/\sigma & \text{for } S > 0 \\ 0 & \text{for } S = 0 \\ (S + 1)/\sigma & \text{for } S < 0 \end{cases} \quad (4)$$

The test statistic Z_s is used as a measure of significance of trend. In fact, this test statistic is used to test the null hypothesis, H_0 : There is no monotonic trend in the data. If $|Z_s|$ is greater than $Z_{\alpha/2}$, where α represents the chosen significance level (usually 5%, with $Z_{0.025} = 1.96$), then the null hypothesis is invalid, meaning that the trend is significant.

Stationary Tests

One major purpose of any stationarity test is to determine whether the mean and/or variances of observations vary with time significantly. Almost in all of the analyzing linear and nonlinear time series analysis practitioners assumed stationarity for time series. Also in most of the applications in hydrological modeling, an assumption of stationarity thus testing stationarity for justification of using those models is necessary [16]. On the other hand, sometimes the investigation of nonstationarity may give us some insights into the underlying physical mechanism of a process. Therefore, testing for stationarity is an important topic of time series in the field of hydrology. Although, there are numerous methods of stationarity tests in hydrology, it is worthy to note that these methods originated from the uncertainties in economic studies.

Generally, a time series is said to be stationary if the expected value $E(Y_t) = \mu$, and the variance, $\text{var}(Y_t) = \sigma^2$ for all t . In other words, if $y_1, y_2, y_3, \dots, y_n$ values of the time series fluctuate around a constant mean with constant variation, the time series is stationary. If the n values do not seem to fluctuate around a constant mean, then it is non-stationary.

If the time series is not stationary, it can be transformed to stationary with the following technique called data differencing or standardization among others. Hence, from the given (observed) series Z_t , we create the new series,

$$Y_t = Z_{(t)} - Z_{(t-1)} \quad (5)$$

where $t = 2, 3, 4, \dots, n$

The differenced data will therefore contain one less point than the original data for the first difference, although it is worthy to note that the data difference can be done more than once before the series becomes stationary.

Stationarity is an important condition for any hydrological models [16]. In practice, the mean and variance should be constant as a function of time before performing the analysis. Otherwise, past effects would accumulate and the values of successive y_t 's would approach infinity making the process non-stationary. For a first order non-stationarity, the observations with hydrological models should be sieved first by differencing the observations d times, of y_t as the time series to obtain stationary data. This is usually done with the transformation as shown in equation (5) above.

Methods of Analysis

Augmented Dickey Fuller (ADF) Test

One of the most important stationarity test based on statistics is the augmented Dickey-Fuller (ADF) unit root test first proposed by Dickey and Fuller (1979) and modified by Said and Dickey (1984) [17-18], which tests for the null hypothesis of the presence of unit roots in a series (difference stationarity).



In general, Augmented Dickey Fuller (ADF) test is used to examine the stationarity of dataset in order to overcome the problem of spurious regression that is common in time series analysis of nonstationary variables. The model of the ADF test with the constant term and trend is as follows:

$$\Delta Y_t = \Psi Y_{t-1} + e_t \quad (6)$$

Where $t = 1, 2, 3, \dots, n$

Where Y_t is the value of rainfall at time t , n is the number of observations, and e_t are assumed to be white noise error term

$\Psi = \rho - 1$: If $\rho \geq 1$, Y_t is a nonstationary series and the variance of Y_t increases with time and approaches infinity. If $\rho < 1$ then Y_t is a (trend) stationary series thus, the hypothesis of (trend) stationarity can be evaluated by testing whether the absolute value of ρ is strictly less than one.

In ADF test the null and alternative hypotheses can be written as $H_0 : \Psi = 0$ and $H_1 : \Psi < 0$.

By estimating the above model (equation 6) and test for the significance of the Ψ coefficient, there are two basic conditions;

- i. If the null hypothesis is accepted, $\Psi = 0$, then Y_t is not stationary.
- ii. If the null hypothesis is rejected, $\Psi \neq 0$, then Y_t is stationary

Auto Covariance (Ψ_k)

From the work carried out by Nason (2006). This is defined mathematically as,

$$\Psi_k = cov(Y_t, Y_{t+k}); \quad (7)$$

where $k =$ considered time lag

$$= E[(Y_t - \mu)(Y_{t+k} - \mu)] \quad (8)$$

Therefore, for $k = 0$,

$$\begin{aligned} \Psi_0 &= cov(Y_t, Y_t) \\ \Psi_0 &= \sigma_y^2 \end{aligned} \quad (9)$$

Auto Correlation (ρ_k)

The Auto Correlation between Y_t and Y_{t+k} is

$$\rho_k = \frac{cov(Y_t, Y_{t+k})}{\sigma_{y_t} \sigma_{y_{t+k}}} \quad (10)$$

Hence, for series that is stationary in mean and covariance, the denomination of equation (10) becomes σ_x^2 (since $\sigma_{y_t} = \sigma_{y_{t+k}}$)

Therefore,

$$\rho_k = \frac{cov(Y_t, Y_{t+k})}{\sigma_y^2} = \frac{Y_k}{Y_0} \quad (11)$$

Similarly, for $k = 0$,

$$\rho_0 = \frac{Y_0}{Y_0} = 1 \quad (12)$$

This implies that, lag 0 auto-correlation is 1

Auto-Correlation Function (Correlogram)

Autocorrelations are numerical values that indicate how a data series is related to itself over time. More precisely, it measures how strongly data values at a specified number of periods apart are correlated to each other over time. The number of periods apart is usually called the "lag". For example, an autocorrelation at lag 1 measures how values 1 period apart are correlated to one another throughout the series. An autocorrelation at lag-2 measures how the values that are 2-periods apart are correlated throughout the series. Autocorrelations may range from +1 to -1 [19]. A value close to +1 indicates a high positive correlation while a value close to -1 implies a high negative correlation. These measures are most often evaluated through graphical plots called "correlograms". A correlogram plots the auto- correlation values for a given series at different lags.

Auto Correlation Function (ACF) indicates the memory of a process [13], that is, how far into the time is it able to remember what has happened before now. For example, it answers the question of how far into the time does Y_{t+k} depends on Y_t in the series. Essentially, correlogram is a plot of ρ_k versus k , and it gives very important information about the process such as stationarity and periodicity of the hydrological process.

The auto-correlation function may indicate some fluctuations. For a purely (stochastic) random series, ρ_k must equal to 0, for all k (since Y_t and Y_{t+k} are not related at all). The cursory look at the graph of the data and structure of autocorrelation function may provide clues for the presence of stationarity [18]. If the series is stationary, the spikes decay fairly rapidly (within 4 or 5 lags, in most hydrologic applications), while the decay is very slow if the time series is not stationary (see Figure 2 and Figure 3)



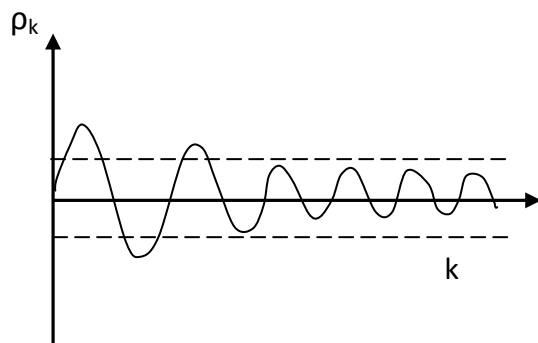


Figure 2: Stationary time series

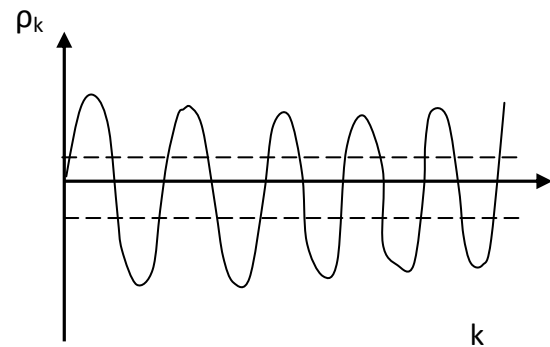


Figure 3: Non-stationary series

Spectral Analysis / Frequency Domain Analysis (FDA)

The observed series Y_t can be transformed from time domain into a frequency domain and then do the analysis in the frequency domain. Often, this is very advantageous especially in determining the periodicity inherent in the data. In general, frequency domain analysis (otherwise known as spectral analysis) is mainly carried out in order to identify the periodicities in the data.

Identification of these periodicities is an important problem in hydrology [20]. Although the correlogram (Auto-correlation function) decay slowly for a nonstationary data which could also give an important clue about the presence of periodicity in the observed series, the identification of the exact and the significant periodicity is better done by spectral analysis.

The basic premise of which the frequency domain analysis is based is on the fact that the observed time series is a random sample of a process over time and is made up of oscillations of all possible frequencies.

According to Mujumdar (2011), the time series Y_t as consisting of different frequencies is expressed as;

$$Y_t = \alpha_0 + \sum_{k=1}^{\frac{N-1}{2}} [\alpha_k \cos(2\pi f_k t) + \beta_k \sin(2\pi f_k t)] + \epsilon_t \quad (13)$$

$$\text{Where } f_k = k/N \quad (14)$$

N = number of observations

$$\text{Periodicity (P); } P = 1/f_k \quad (15)$$

Therefore, in the above expression, we expressed α_0 , α_k , and β_k as shown below,

$$\alpha_0 = \bar{y} \quad (16)$$

$$\alpha_k = \frac{2}{N} \sum_{t=1}^N Y_t \cos(2\pi f_k t) \quad k = 1, 2, 3, \dots, N/4 \quad (17)$$

$$\beta_k = \frac{2}{N} \sum_{t=1}^N Y_t \sin(2\pi f_k t) \quad k = 1, 2, 3, \dots, N/4 \quad (18)$$

A variance spectrum divides the variances into number of intervals or bands of frequency. Spectral density (I_k) is the amount of variance per interval of frequency.

$$I(k) = \frac{N}{2} [\alpha_k^2 + \beta_k^2] \quad (19)$$

Angular Frequency;

$$\omega_k = \frac{2\pi k}{N} \quad (20)$$

A plot of $I(k)$ by ω_k is called spectrum, and a peak in the spectrum indicates an important contribution to variance at frequencies close to the peak. Prominent spikes therefore indicate periodicities. Essentially, what is done in spectral analysis is to convert Y_t which is the observed time series to a sort of Fourier transformation where sinusoidal representing of time can be used.

Statistical Significance of Periodicities

A wrong conclusion may be made that all the prominent spikes in the transformed time series are significant, hence the prominent spikes need to be analysed for their statistical significance. The periodicities which are eventually tested to be significant need to be removed from the original series to get a new series (Z_t).

The periodicities are tested for significance by defining a statistic ' \cap ' as follows;

$$\cap = \frac{y^2(N-2)}{4\hat{\rho}_1} \quad (21)$$

$$\text{where, } y^2 = \alpha^2 + \beta^2, \text{ and} \quad (22)$$

$$\hat{\rho}_1 = \frac{1}{N} \sum_{t=1}^N \{Y_t - \hat{\alpha} \cos(\omega_k t) - \hat{\beta} \sin(\omega_k t)\} \quad (23)$$



The periodicity corresponding to ω_k is significant at level α only if,

$$\hat{\rho} \geq F(2, N - 2) \quad (24)$$

where F denotes F -distribution at 2 degrees of freedom.

Also, it is important to know that this test examines one periodicity at a time and should be carried out on a series from which all periodicities previously found to be significant are removed.

Results and Discussion

Statistical Analysis

Statistical analysis was conducted on the monthly rainfall data for Ilorin meteorological station. The parameters obtained include; mean standard deviation (S.D), median, and skewness, maximum and minimum values. The statistical summary for the meteorological variable is as presented in Table 4.1

Table 1: Monthly statistical summary for precipitation (mm) at Ilorin

Parameter	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	6.74	9.51	43.53	104.4	159.5	192.0	158.6	144.1	222.6	144.3	15.89	7.82
Median	0.00	1.20	33.30	98.90	156.00	184.9	154.2	133.7	238.6	154.6	1.70	0.00
STD	16.38	16.83	28.61	51.72	68.25	67.37	72.02	84.12	70.47	76.94	33.17	24.99
CV	2.43	1.77	0.66	0.50	0.43	0.35	0.45	0.58	0.32	0.53	2.09	3.20
Min	0.00	0.00	0.00	19.60	42.30	72.50	62.20	26.90	62.30	9.20	0.00	0.00
Max	73.30	90.70	117.7	250.3	355.7	382.5	325.9	350.0	366.4	335.5	156.1	148.3
Skew	2.90	2.91	0.78	0.88	0.65	0.84	0.85	0.67	-0.33	0.17	3.20	4.34

The maximum and minimum values of rainfall values for the data series are 382.5mm and 0.00mm respectively, which indicate that Ilorin meteorological station exhibits a very high range in rainfall distribution. Range of values for standard deviation and skewness coefficient were 67.74 and 4.17 respectively, while the mean and median values are as shown in Table 1.

Statistical analysis of rainfall time series data involves the computation of the rainfall mean rainfall, standard deviation and coefficient of variation. This measures rainfall variability around its mean value. Table 1 shows the statistical analysis results.

Time Plot

In order to understand and to extract more information contained in the historical rainfall data of the study area; the data was divided into two before the time plots, thereby having two different realizations. The first realization is gotten from 1960 to 1984 rainfall data, while the second realization is from 1985 to 2010 rainfall data series.

The time plots of the rainfall data in Ilorin meteorological station for the two realizations are as shown in Figure 4 and Figure 5 and shows that there is a systematic changes in the data series known as trend. The trend in the data is decreasing globally but not always the case because it was realized from Figure 5 that there is a shift in the hydrological record that produced sudden increase in trend from 1985 to 2010. This shows that the preliminary test on the data indicates non-stationarity due to trend components.

However, the peak monthly rainfall obtained from Figures 4 and 5 for the considered meteorological station indicated that the major outlier value in the series is 382.5mm and occurred in June, 1966. From the two realizations, major outliers occurred in 1960s and it is the decade with the highest number of major outliers.

The time series plot forms the basis for a comprehensive analysis of time series data. This plot shows temporal distribution of accumulated monthly rainfall with time and also indicates peak rainfall values (major outliers). Outliers are data exhibiting outright deviation from the mean rainfall value and are indicative of months with extreme wet condition; they are series with a very high monthly value implying peak monthly rainfall in the series.

Mann Kendall Trend Analyses Result

The non-parametric Mann-Kendall trend analysis was used to analyze the trend pattern of the rainfall data in the study area. The Mann-Kendall statistic (S) is obtained using Equation 9 and it was used to determine the structure of the trend exhibited; however, variance (σ^2) of S was obtained from the Equation 10.

Null hypothesis (H_0) was used to check if there is monotonic trend in the data or not, while the significance level was chosen as 5% and the $Z_{\alpha/2}$ from the normal distribution table is $Z_{0.025} = 1.96$, hence, the Z_s obtained from the calculation was compared with the $Z_{0.025}$ (1.96) to know whether to accept or reject the null hypothesis.



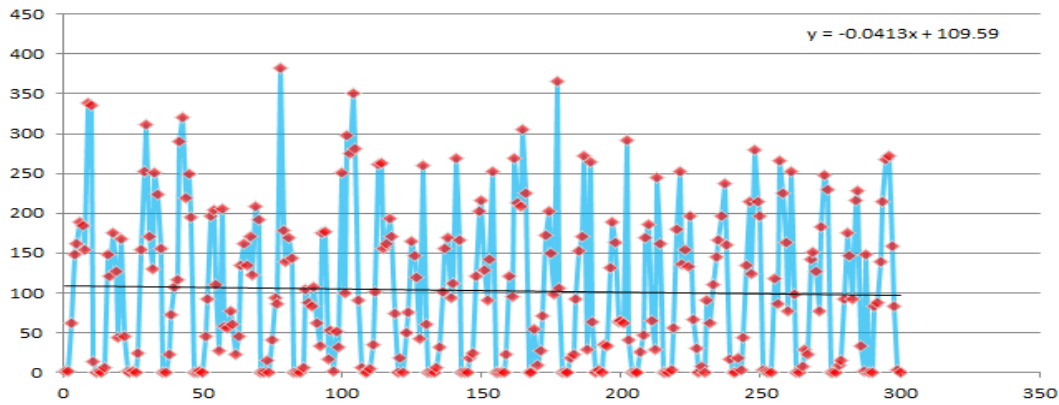


Figure 4: Time Plot of Monthly Rainfall Data (First Realization) from 1960 – 1984

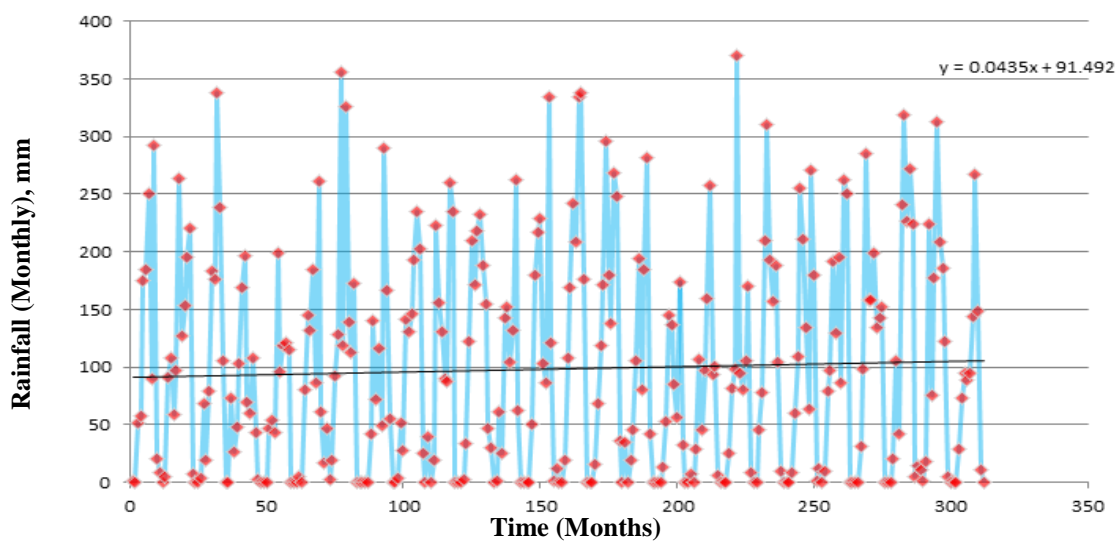


Figure 5: Time Plot of Monthly Rainfall Data (Second Realization) from 1985 – 2010

Table 2: Mann-Kendall Result for Rainfall Data in Ilorin

Data	Autocorrelation factor	Kendall's S	Z _s	Trend	Significance status
Rainfall	0.090	-54	0.4550	Negative	Non-significant

The results of the Mann Kendall analysis for all the variables are presented in Table 2 above. The value of Mann Kendall's S of rainfall was -54, and a Z_s of 0.4550 which is less than the normalized test statistic (Z_{0.025}) of 1.96, which made it statistically non-significant confirming the existence of a negative (decreasing) trend in the data on a global scale as earlier observed in the time plots.

Regression Analysis

The detection of the trend's significance was also analyzed using the regression approach which involve simple Linear Regression model. The linear regression analysis was carried out with the aid of R-statistical software on the rainfall data series (Y_t) of Ilorin meteorological station using rainfall as the dependent variable and time (month) as the independent variable. Hence, the following was obtained,

Table 3: Regression Analysis Summary for Ilorin Meteorological Station

Data	Trend	Slope	Regression Model
Rainfall	Decreasing	-0.0115	Y _t = 104.32 - 0.0115t

The regression analysis also conformed to the Mann-Kendall Trend test which shows that the data contained a decreasing trend and not stationary.

Augmented Dickey-Fuller Test

ADF test was used to confirm the stationarity status of the rainfall data in the study area by testing whether a unit root is present in an autoregressive model; hence, after the analysis, the following results were obtained using significant p-value level of 0.05 (that is, 95% confident limit).

Dickey-Fuller = -12.3864, Lag order = 8, p-value = 0.01

Alternative hypothesis: stationary

Recall from section 4.1,

The null and alternative hypotheses can be written as $H_0 : \Psi = 0$ and $H_1 : \Psi < 0$ respectively.

- i. If the null hypothesis is accepted, $\Psi = 0$, then y_t is not stationary.
- ii. If the null hypothesis is rejected, $\Psi \neq 0$, then y_t is stationary

DECISION: after estimating the model (equation 13) as shown above, test for the significance of the Ψ coefficient, the p-value of 0.01 (approximately, $\Psi = 0$) was obtained. It showed that the test is significant at 0.05; therefore the null hypothesis is accepted and it can be concluded that the rainfall data of Ilorin meteorological station is non-stationary.

Autocorrelation Function (Correlogram)

Auto-correlation Function (ACF) indicates the memory of a process [13], that is, how far into the time is it able to remember what has happened before now. It simply answers the question of how far into the time does Y_{t+k} depends on Y_t in the series. Essentially, correlogram is a plot of ρ_k versus k , and it gives very important information about the process such as stationarity and periodicity of the hydrological process. Here, Equation 14 – Equation 18 were used in order to get an appropriate autocorrelation plot for the considered meteorological data up to lag k of $(N/2)$.

The table shown below (Table 4) only displayed the first 12 months, and the procedure followed.

Table 4: Autocorrelation Analysis of Rainfall Data for $k = 1$

Y_t	$Y_t - \bar{Y}$	Y_{t+1}	$Y_{t+1} - \bar{Y}$	$(Y_t - \bar{Y})(Y_{t+1} - \bar{Y})$	$(Y_t - \bar{Y})^2$
1	-99.7819	1	-99.7819	-99.7819	9956.427568
1	-99.7819	62.5	-38.2819	-2392.61875	9956.427568
62.5	-38.2819	148.3	47.5181	7046.93423	1465.503868
148.3	47.5181	161.9	61.1181	9895.02039	2257.969828
161.9	61.1181	189.5	88.7181	16812.07995	3735.422148
189.5	88.7181	184.4	83.6181	15419.17764	7870.901268
184.4	83.6181	154.9	54.1181	8382.89369	6991.986648
154.9	54.1181	338.4	237.6181	80409.96504	2928.768748
338.4	237.6181	335.5	234.7181	78747.92255	56462.36145
335.5	234.7181	13.7	-87.0819	-1193.02203	55092.58647
13.7	-87.0819	0.5	-100.2819	-50.14095	7583.257308
0.5	-100.282	1	-99.7819	-99.7819	10056.45947

One of the most reliable ways to examine stationarity is by plotting the graph of sample auto-correlation at lag k (ρ_k) against k ; this was as shown in Figure 6.

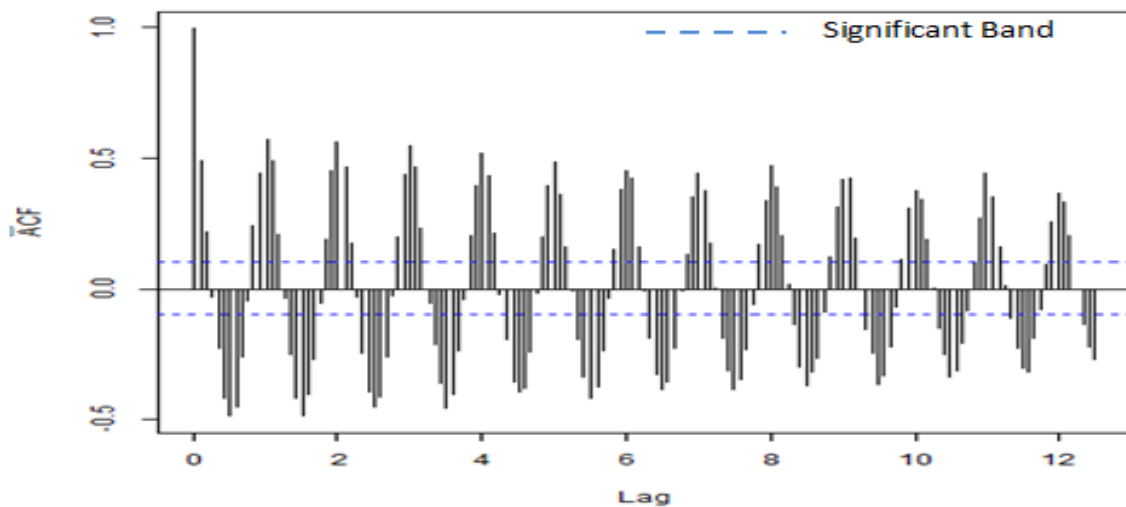


Figure 6: Autocorrelation Function Plot

Figure 6 represents the Auto-Correlated Function (ACF) plot of rainfall data of Ilorin metrological station. Hence, it was observed that there are several significant lags on the ACF plots far above the significant band of $\pm 1.96/\sqrt{N}$; therefore, all the data is not stationary when considering the original series. Since the non stationarity of the data has been suspected, there is need to stationarize the data by standardization or trend difference before an appropriate model of the data can be built.

Spectral Analysis / Frequency Domain Analysis (FDA)

Spectral analysis was used in this section to estimate spectral density function (spectrum) of the rainfall time series. It is essentially a modification of Fourier analysis so as to make it suitable for stochastic rather than deterministic functions of time. The spectrum (that is, a plot of $I_{(k)}$ versus ω_k) was obtained by using Equation 26 and Equation 27, and is as shown in Figure 7.

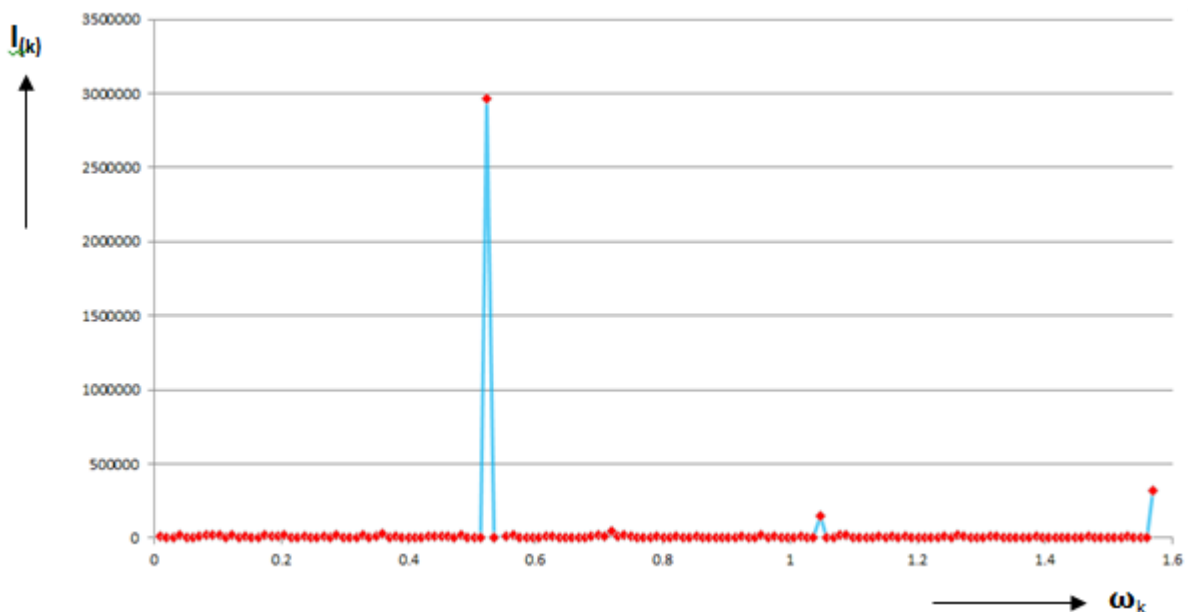


Figure 7: Line spectrum (plot of $I_{(k)}$ versus ω_k)

From the spectrum, prominent peaks (spikes) represent the periodicities inherent in the data.

Recall, Equation 21, 22, and 27; one may come to the conclusion that,

$$P = 2\pi/\omega_k$$

Using equation 32 and from Figure 7, it is confirmed that the peak correspond to $\omega_k= 0.5236$ indicating a periodicity of 12months, the peak correspond to $\omega_k= 1.0473$ indicates a periodicity of 6 months, while the spike correspond to $\omega_k= 1.571$ indicates the periodicity of 4 months.

Therefore, it can be seen clearly that the historical rainfall data of Ilorin meteorological station contained 12 months, 6 month, and 4 months periodicity.

Significant Test for the Identified Periodicities

In hydrology, a prominent spike does not necessarily mean that the periodicity is significant in the data series, it is therefore necessary to have a statistic to test if the spikes are statistically significant or not. As earlier discussed, Equations 28, 29, and 30 were used to do the test. Because of the rigorous computations involved, Microsoft Excel and MATLAB software were used to perform the analysis.

The parameters required for the estimations as seen in Equation 28 to Equation 30 were computed as shown in Table 5 while the outcome status of the analysis was displayed in Table 6.

Table 5: Parameter estimates for significant periodicity test

ω_k	α	β	γ^2	\bar{p}
0.5236	-80.0428	-57.3901	9700.47	181.3405
1.0473	-15.6183	-15.8919	496.48	116.6881
1.571	-16.6401	27.6597	1041.95	116.6574

Table 6: Significance Status of Inherent Periodicities

S/No.	ω_k	Statistic	F(2, N-2)	Status
1	0.5236	8157.7032	3.01	Significant
2	1.0473	648.8513	3.01	Significant
3	1.571	1362.086	3.01	Significant



From the analysis carried out on the transformed data series, it is confirmed that the historical rainfall data in Ilorin meteorological station contained a significant periodicities of 12 months, 6 months, as well as 4 months. This however indicates that the data is not stationary and the degree of non-stationarity (that is, the deterministic components) in the data needs to be removed when using it for any hydrologic design, planning, or modeling.

Conclusion

Systems for management of water resources throughout the developed and developing world are mostly designed and operated under the assumption of stationarity. Stationarity—the idea that natural systems fluctuate within an unchanging envelope of variability—is a foundational concept that permeates training and practice in water-resource engineering. It implies that any variable (hydro-meteorological data) has a time-invariant (or 1-year-periodic) probability density function (pdf), whose properties can be estimated from the instrument record. The meteorological data used for the purpose of this study is the historical rainfall data of Ilorin, Nigeria. The data was subjected to linear regression analysis, Mann-Kendall analysis, Augmented Dickey-Fuller (ADF), Autocorrelation analysis, and Spectral analysis (Frequency Domain Analysis); all of which aimed at investigating stationarity / non-stationarity inherent in the data. The data was divided into two realizations in order to have deeper understanding of the years that contribute to the trend of the data series.

The initial trend test on the two realizations of the time plots indicated that the rainfall data contained different pattern and fluctuations. The trend of the monthly series was investigated using Mann-Kendall test. The result showed that the series contains non-significant negative trend. Similarly, the Autocorrelation function (ACF) plot, and Partial Autocorrelation function (PACF) plot indicated that there are several significant spikes at different lags confirming that the data is not stationary.

The data was then transformed from time domain into a frequency domain using Fourier transform after which spectral analysis was performed on the data in the frequency domain to detect the particular periodicities embedded in the data series. The results of the spectrum analysis indicated that the historical rainfall series of Ilorin contains periodicities of 12 months, 6 months, and 4 months, all of which were statistically significant.

Based on the outcome of the analysis performed, it could be recommended that non-stationarity and significant periodicities inherent in the data should be removed by standardization before it could be used for any hydrological modeling.

References

- [1]. Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J. (2008), “Stationarity Is Dead: Whither Water Management?” Policy Forum, Science Vol. 319, pg. 573-574.
- [2]. Uba, T. S., and Bakari, H. R. (2015), An Application of Time Series Analysis in Modeling Monthly Rainfall Data for Maiduguri, North Eastern Nigeria; *Mathematical Theory and Modeling*, ISSN 2225 – 0522, Vol.5, No.11, pg. 24-33.
- [3]. Yilmaz A. G., Hossain, I., and Perera, B. J. C. (2014), Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne, *Hydrol. Earth Syst. Sci.*, 18, 4065–4076, www.hydrol-earth-syst-sci.net/18/4065/2014/ doi:10.5194/hess-18-4065-2014.
- [4]. Daniela S., Magda, K., Silvia, K., Ján S., and Kamila H. (2010), Analysis of Discharge and Rainfall Time Series in the Region of the KláštorskéLúky Wetland in Slovakia, Department of Land and Water Resources Management, Slovak University of Technology Bratislava, Slovak Republic.
- [5]. Gurudeo A. T. and Mahbub, I. (2010), Time Series Analysis of Rainfall and Temperature Interactions in Coastal Catchments, *Journal of Mathematics and Statistics* 6 (3): ISSN 1549-3644, pg. 372-380.
- [6]. Oduro-Afriyie, K. and Adukpo, D. C. (2006), Spectral Characteristics of the Annual Mean Rainfall Series in Ghana, *West Africa Journal of Applied Ecology (WAJAE)* –ISSN: 0855-4307 Volume 9.
- [7]. Ideiã, S. M. and Santos, C. A. (2005), Analysis of Precipitation Time Series Using the Wavelet Transform, *Sociedade&Natureza, Uberlândia, Special Issue*, 736-745.
- [8]. Ekpoh, I. J. and Nsa, E. (2011), “Extreme Climatic Variability in North-western Nigeria: An Analysis of Rainfall Trends and Patterns”, *Journal of Geography and Geology*, Vol. 3, No. 1.
- [9]. IPCC, 2007: Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 104 pp.
- [10]. Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J.-C., Gachon, P., and Bobee, B. (2006): Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review, *J. Hydrology*, 329, 534–552.



- [11]. Olofintoye O. O., and Sule B. F. (2010), "Impact of Global Warming on the Rainfall and Temperature in the Niger Delta of Nigeria", USEP: Journal of Research Information in Civil Engineering, Vol.7, No.2, pg. 33-48.
- [12]. Olofintoye, O., and Adeyemo, J. (2012), Development and Assessment of a Fourier Approximation Model for the Prediction of Annual Rainfall in Ilorin, Nigeria, Department of Civil Engineering and Surveying, Durban University of Technology, Durban, South Africa.
- [13]. Wang, W., Van-Gelder, P.H.A.J.M., and Vrijling, J.K. (2005), Trend and Stationarity Analysis for Streamflow Processes of Rivers in Western Europe in the 20th Century, IWA International Conference on Water Economics, Statistics, and Finance, Rethymno, Greece, pg. 451-461.
- [14]. Adewuyi, S. A., Okojie, L.O., Folorunso, B., and Bada, B.S. (2014), "Effect of climate change on yam and cassava production in Oyo state, Nigeria: A co-integration model approach", International Journal of Agricultural Economics & Rural Development - 6 (1): pg. 54-66.
- [15]. Ajadi, B.S., Adeniyi, A., and Afolabi, M. T. (2011), "Impact of Climate on Urban Agriculture: Case Study of Ilorin City, Nigeria", Global Journal of Human Social Science, Volume 11, Issue 1, Version 1.0, ISSN:0975-587X.
- [16]. Chisimkwuo, J., Uchekukwu, G., and Okezie, S. C. (2014), Time Series Analysis and Forecasting of Monthly Maximum Temperatures in South Eastern Nigeria, International Journal of Innovative Research & Development, Vol 3, Issue 1, pg. 165-171.
- [17]. Dickey, D. A. and Fuller, W. A. (1979), "Distributions of the Estimators for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association, Volume 74, Issue 366, pg. 427-431.
- [18]. Said, E. S., and Dickey, D. A. (1984), "Testing for unit roots in autoregressive-moving average models of unknown order", Biometrika Vol. 71, Issue 3, pg. 599-607.
- [19]. Ajala, O. M. (2014), Statistical Analysis of Recorded Road Traffic Accidents and Casualties in Lagos State from 2006-2012; B.Sc. Project, Statistics Department, University of Ilorin, Nigeria
- [20]. Edwin, A. I., Martins, O. Y. (2014), "Stochastic Characteristics and Modelling of Monthly Rainfall Time Series of Ilorin, Nigeria", Open Journal of Modern Hydrology, 2014, 4, pg. 67-79, <http://dx.doi.org/10.4236/ojmh.2014.43006>

