

TIME SERIES ANALYSIS OF MONTHLY RAINFALL IN NIGERIA WITH EMPHASIS ON SELF-ORGANIZED CRITICALITY

C. C. Nnaji

*Department of Civil Engineering
University of Nigeria, Nsukka*

E-mail: chidozie.nnaji@unn.edu.ng; episcopal2k@yahoo.com

ABSTRACT

Monthly rainfall data of twenty-one years (1980 – 2000) were analyzed for the six regions of Nigeria using the rescaled range (R/S) statistic, the standard fluctuation analysis (FA) and the detrended fluctuation analysis (DFA). The results indicated that the distribution of monthly rainfall has a Hurst exponent of 1 in the short term which is an evidence of self-organized criticality, and a Hurst exponent of about 0.5 in the long run which is an evidence of a random walk. The Hurst exponents decreased from around 1 for small window sizes and reached 0.5 after 36 months and then remained fairly constant. The R/S statistic was found to perform better than the FA and the linear DFA used in the analysis.

Keywords: *Self-Organized Criticality, Hurst exponent, monthly rainfall, crossover, fluctuation analysis.*

INTRODUCTION

It is difficult to represent many natural phenomena which vary randomly with time with an accurate physical model. This is because they exhibit a high degree of randomness, not easily portraying any pattern. However, studies have shown that the so called random phenomena exhibit some correlations which can be unmasked with some recently developed analytical tools. Occurrences such as earthquakes (Bak *et al.*, 2002), landslides (Turcotte *et al.*, 2002), rainfall, DNA (Peng *et al.*, 1994), spot electricity, World Wide Web (Barabasi, 1999), forest fire (Turcotte *et al.*, 2002) financial time series (Czarnecki *et al.*, 2008; Grech and Mazur, 2004; Matos *et al.*, 2008), fragmentation (Turcotte, 1997), fracture (Sornette, 2002),

birth (Sornette, 2002), run off, daily temperature (Vjushin *et al.*, 2002) have been shown not to be subject to pure chance but follow a power law.

Several methods such as the autocorrelation analysis, variance ratios, and spectral analysis have been used for the analysis of time series. However, these methods may fail when trends are present in the system. Trends are systematic deviations from the average data (runoff, rainfall, etc.) that are caused by external processes, e.g. the construction of a water regulation device, the seasonal cycle and climate change.

Monotonous trends may lead to an overestimation of the Hurst exponent or even make uncor-

related data, under the influence of a trend, look like long-term correlated ones. Furthermore, it is difficult to distinguish trends from long-term correlations, because stationary long-term correlated time series exhibit persistent behaviour and a tendency to stay close to the momentary value (Koscielny-Bunde *et al.*, 2003).

The concept of Self-Organized Criticality (SOC) was introduced by Bak *et al.* (1987) to explain the behavior of open and extended driven systems with avalanche-like energy dissipation, as sandpile and earthquake model. Bak *et al.* (1987, 1988) used the agent-based modeling approach to simulate the emerging large-scale dynamics of avalanches when grains of sand were added slowly and randomly to a sandpile. The sandpile system was said to be self-organized because a power-law relationship between the size and the frequency of avalanches appeared without any external guidance (Linkenkaer-Hansen, 2002). SOC refers to the tendency of non-equilibrium systems driven by a low constant energy input to organize themselves to a complex correlated state where all scales are relevant (Jianlin *et al.*, 2005). This concept is deeply related to the idea of scale invariance of the distribution of relaxation events and with the existence of self-regulatory internal mechanisms that drive the system spontaneously to a statistical stationary state.

The extreme complexity of atmospheric processes results from the coupling of several non-linear processes having completely different temporal and spatial characteristic generating correlation that extends throughout the entire system and leading to power law distribution. The aim of the science of self-organization and complexity is to understand the source and character of structures and temporal dynamics in systems that naturally exhibit large variability. A system is said to be self-organized when its structure emerges without external influence. This implies that the constraints on the organization result from the interactions among the components and are internal to the system (Lucas, 2002).

This study aims at ascertaining the presence and range of Self-Organized Criticality in the monthly rainfall pattern of Nigeria using data spanning twenty one year period. The rescaled range (R/S) analysis, the standard fluctuation analysis (FA), and the detrended fluctuation analysis (DFA) will be used to obtain the Hurst exponents, h for the six regions of Nigeria with a view to ascertaining short-term and long-term correlations of rainfall data in these regions.

METHODOLOGY

Monthly rainfall data of twenty one years (1980 – 2000) for six states in each geopolitical region of Nigeria were analyzed using the rescaled range statistic (R/S), the standard fluctuation analysis (FA) and the first order detrended fluctuation analysis (DFA1).

Rescaled Range Analysis (R/S)

This method was first developed by Hurst, a British bureaucrat who spent a lifetime studying the Nile and the problems related to water storage (Hurst, 1951). Mandelbrot (1983) is of the opinion that the rescaled range analysis is a more powerful tool in detecting long range correlation when compared to more conventional methods like autocorrelation analysis, variance ratios, and spectral analysis. In the rescaled range determination of long-term memory, the mean $\langle X \rangle$ of the monthly rainfall data collected is obtained. Then the deviation from the mean value is calculated and the cumulative sum of the deviation is calculated as follows:

$$R(i) = \sum_{t=1}^N (X(i) - \langle X \rangle) \quad (1)$$

Where $R(i)$ = cumulative deviation from mean monthly rainfall; $X(i)$ = monthly rainfall (mm); $\langle X \rangle$ = mean monthly rainfall (mm); N = total number of months since we are dealing with monthly rainfall.

The profile $R(i)$ is then divided into non-overlapping segments of time length n called window. For each segment (s) of window size n , the rescaled range is obtained as the difference between the maximum value of the profile

$R(i)$ and the minimum value of the profile using equation (2).

$$R(s, n) = \max(X(s_i, n)) - \min(X(s_i, n)) \quad (2)$$

for $1 \leq s_i \leq s$

Where s_i represents different segments and i is an integer.

The rescaled range statistic is obtained as follows:

$$R(s, n) / S(s, n) = \left(\frac{R_{s,n}}{S_{s,n}} \right) \quad (3)$$

Where $R(s, n) / S(s, n)$ = the rescaled range statistic for segment s_i of window size n ; $R(s, n) = \max(X(s_i, n)) - \min(X(s_i, n))$; $S(s, n)$ = standard deviation for segment s_i of window size n ; h = Hurst exponent; n = window size in months.

The average value of $R(n) / S(n)$ is then obtained for the window size n . The window size is changed and the procedure is repeated so that for different window sizes (n), we obtain different values of $R(n) / S(n)$, hence Equation (4).

$$R(n) / S(n) = n^h \quad (4)$$

All parameters have been previously defined. If Equation (4) is log transformed, a linear function is obtained. A linear regression line through a set of points, composed of the log of n (the size of the areas on which the average rescaled range is calculated) and the log of the average rescaled range over a set of regions of size n , is calculated. The slope of the regression line is the estimate of the Hurst exponent, h . The Hurst exponent, h , is equal to 0.5 for Brownian motion also referred to as the random walk phenomenon or Gaussian white noise; $h < 0.5$ is an indication of anti-correlation or anti-persistence, $h > 0.5$ is an indication of positive correlation, while $h = 1$ indicates $1/f$ noise which is a signature of self organized criticality.

Standard Fluctuation Analysis (FA)

In this method, the profile

$$R(i) = \sum_{t=1}^N (X(t) - \langle X \rangle)$$

is divided into $s = \text{int}(N/n)$ non-overlapping segments of window size n starting from the beginning, and s non-overlapping segments starting from the end of the rainfall series. The notation $\text{int}(N/n)$ simply implies that the N/n must be an integer number when creating segments. The fluctuation is obtained by extracting the values of the profile at both ends of each segment. For each segment, the square of the difference of the two end values is obtained so that

$$F^2(n) = (R_{is} - R_{(i-1)n+1})^2 \quad (5)$$

The average of $F^2(i, s)$ over the $2s$ subsequences is obtained for all the segments of a chosen window size (n). Hence, the fluctuation is given by

$$F(n) = \sqrt{\frac{1}{s} \sum_{i=1}^s (R_{is} - R_{(i-1)n+1})^2} \quad (6)$$

This procedure is repeated for different window sizes to obtain a scaling relation between $F(n)$ and n . The fluctuation $F(n)$ can be viewed as the root-mean-square displacement of a random walker on a linear chain after n steps. The standard fluctuation analysis is somewhat similar to the rescaled range analysis described previously except that it focuses on the second moment $F_2(n)$ while the rescaled range method considers the first moment $F_1(n)$. Just like in the rescaled range analysis, the fluctuation follows the power law of Equation (4) so that the slope of $\text{Log } F$ versus $\text{Log } n$ gives the Hurst exponent.

Detrended Fluctuation Analysis

Detrended Fluctuation Analysis (FA) was first proposed by Peng *et al.* (1994) while examining series of DNA nucleotides. The rescaled range analysis and the standard fluctuation

analysis are usually affected by the presence of trends (Vjushin *et al.*, 2002). In order to eliminate trends in a systematic way, several methods have been developed. These methods include the detrended fluctuation analysis (DFA), the multifractal detrended analysis (MF-DFA), the wavelet transform (WT), the detrended moving average (DMA), the centered moving average (CMA), continuous DFA, Fourier DFA, etc. In DFA, unlike the previous two methods described above, trends are eliminated in the process of calculating fluctuations. The deviation from the mean is calculated for each data in the monthly rainfall time series and then the cumulative deviation from the mean is obtained for each data point (see Equation 1). The data generated by Equation (1) is then divided into $K = N/n$ non-overlapping segments of window size (time interval) n . For each segment s , a best linear fit of the profile is determined. The ordinate of the fitted line (local trend) for each segment is denoted by $Y_s(i)$, and the detrended fluctuation is obtained by subtracting the profile $R(i)$ from the corresponding local trend $Y_s(i)$. The fluctuation is given by:

$$F^2(s, n) = \frac{1}{s} \sum_{i=1}^s \left(R_{(s-1)n+i} - Y_s(i) \right)^2 \quad (7)$$

The linear DFA was first proposed by Peng *et al.* (1994) when analyzing correlations in DNA. However, it can be generalized to eliminate higher order trends (Bunde *et al.*, 2000; Kantelhardt *et al.*, 2001). The line of best fit can also be obtained by fitting a polynomial of order k so that the general fluctuation expression for DFA is given by:

$$F^2(s, n) = \frac{1}{s} \sum_{i=1}^s \left(R_{(s-1)n+i} - Y_s^k(i) \right)^2 \quad (8)$$

This approach helps to eliminate trends of the k^{th} order.

RESULTS AND DISCUSSION

The result of the DFA and R/S analyses were found to behave similarly while that of the FA analysis showed a somewhat differing behavior

though they all lead to the same conclusion. The DFA results show that the monthly rainfall for the six regions examined exhibit a positive correlation in the short range (Figures 1 - 3).

As can be easily verified from these figures, the slopes of $\text{Log } F$ versus $\text{Log } n$ suddenly transit from a region of high values in the short-term to lower values in the long-term. Jos, Kano and Bauchi which represent the three regions in the north, the Hurst exponent for $0 \leq n \leq 24 \text{ months}$ tends to unity but decreases as n increases. For these regions, Hurst exponent tends to 0.5 for $24 \leq n \leq 252 \text{ months}$ which is an indication of a random walk. This implies that the monthly rainfall exhibits self-organized criticality for fine temporal resolutions (short range) and Gaussian white noise for coarser resolution (long range). For the R/S analysis, the Hurst exponent tends to unity for $0 \leq n \leq 12 \text{ months}$ and 0.5 for $12 \leq n \leq 252 \text{ months}$ for the same regions. For Lagos (South-West), Enugu (South-East) and Calabar (South-South) which represent the three regions of the south, the Hurst exponent for $0 \leq n \leq 14 \text{ months}$ tends to unity, and 0.5 for $14 \leq n \leq 252 \text{ months}$. For the R/S analysis, the Hurst exponent tends to unity for $0 \leq n \leq 12 \text{ months}$ and 0.5 for $12 \leq n \leq 252 \text{ months}$ in these regions. The analysis above becomes clearer by successively estimating h at each window size and comparing the values.

Chen *et al.* (2002) and Kantelhardt, *et al.*, (2002) observed that many records do not exhibit a simple monofractal scaling behavior, which can be accounted for by a single scaling exponent. In some cases, there exist crossover (time) scales separating regimes with different scaling exponents (Hu *et al.*, 2001). Systems exist which exhibit long-range correlations on small scales and another type of correlations or uncorrelated behavior on larger scales. Figures 1 and 2 distinctly reveal the crossovers for the regions under study. A crossover is usually indicated by a clear change in the slope of the plot of $\text{Log } F$ versus $\text{Log } n$. If a crossover does not exist, larger window sizes will yield larger

Hurst exponents, h , but the slope of the plot will remain constant. A crossover usually indicates a change in correlation pattern of a data set as time scale changes. Some data may have positive correlation in the short-term but anti-correlation or even pure randomness in the long-term. So crossovers can be used to monitor the

variations in correlation patterns of a data set in relation to time scale. This study has revealed that rainfall in Nigeria follows this pattern, exhibiting self-organized criticality on small scales, but as time scale increases the correlation decays giving rise to pure randomness.

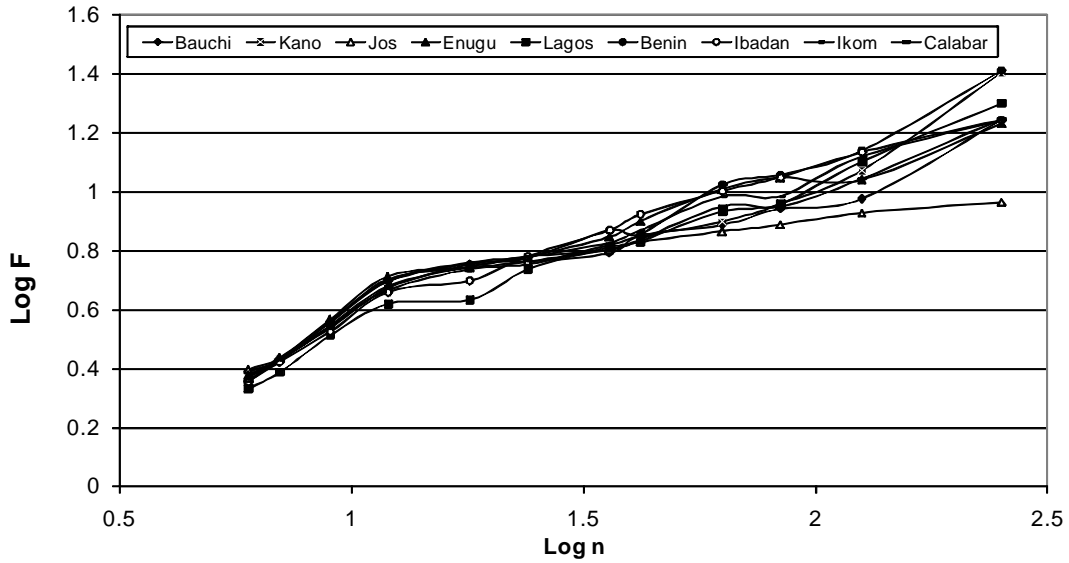


Fig. 1: Plot of R/S Analysis for the six regions

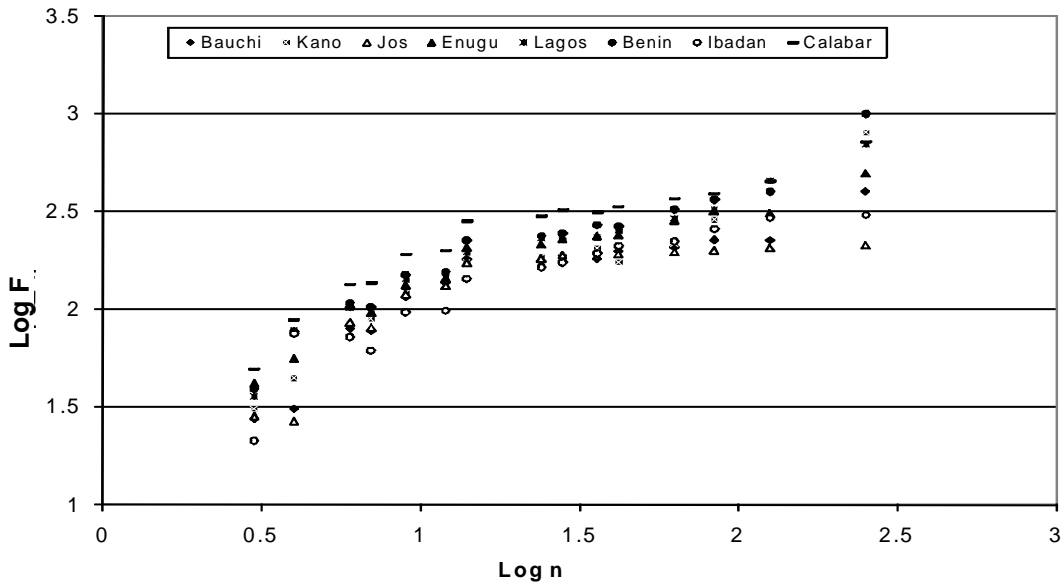


Fig. 2: Plot of DFA Analysis for the six regions

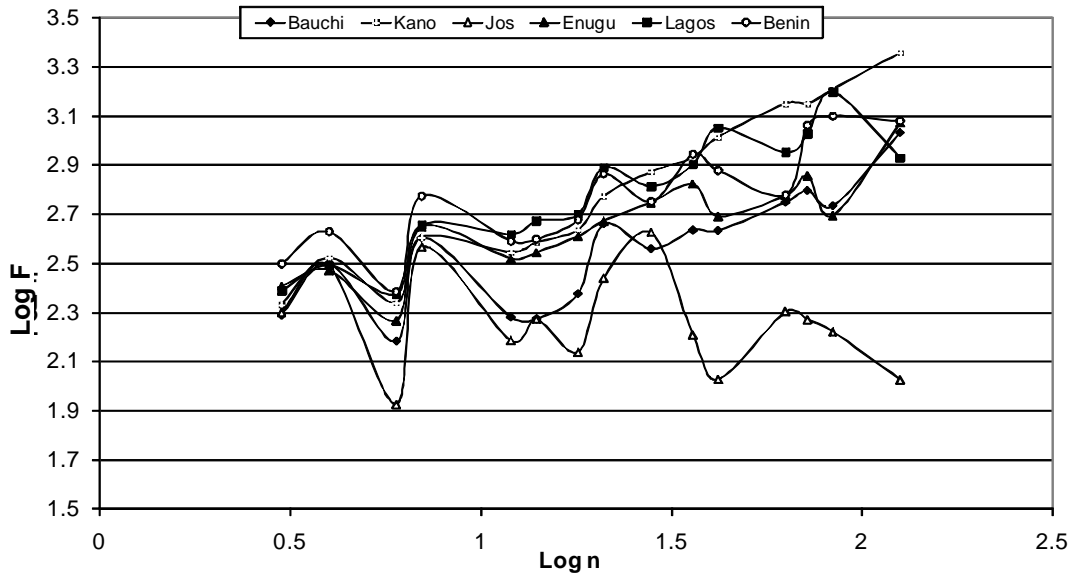


Fig. 3: Plot of FA Analysis for the six regions

Even though researchers have claimed that the DFA gives better results than the R/S analysis and the FA (Grech and Mazur, 2004), the results obtained (Figures 4 – 9) and 1 - 2) seem to indicate otherwise for this study.

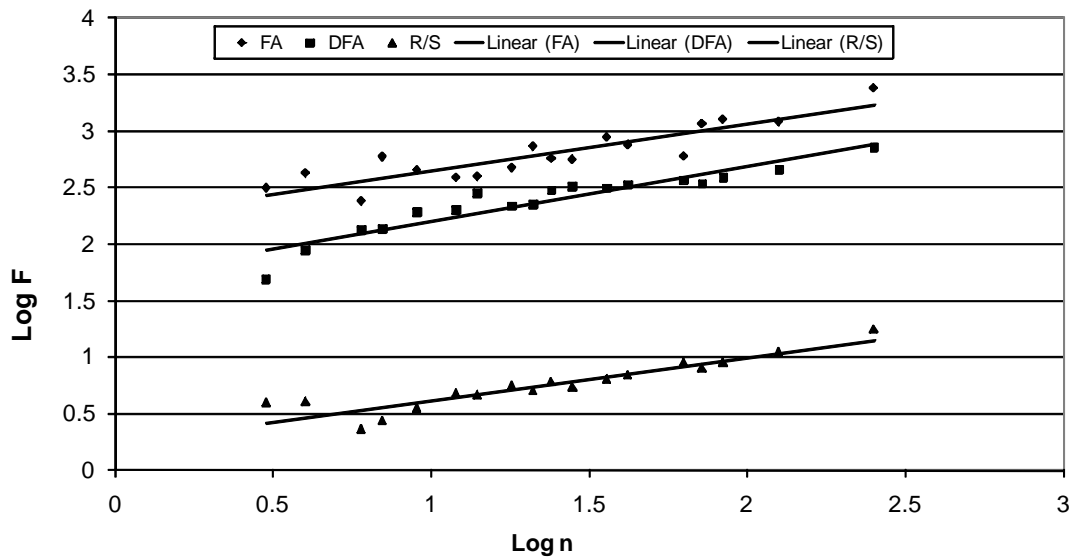


Fig. 4: Plots of R/S, FA and DFA for Calabar (South-South)

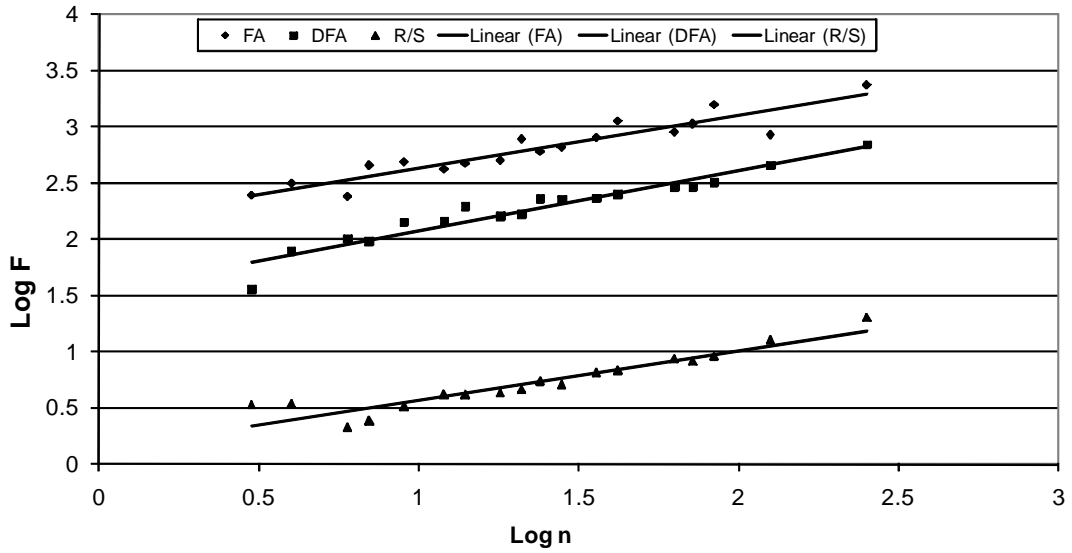


Fig. 5: Plots of R/S, FA and DFA for Lagos (South-West)

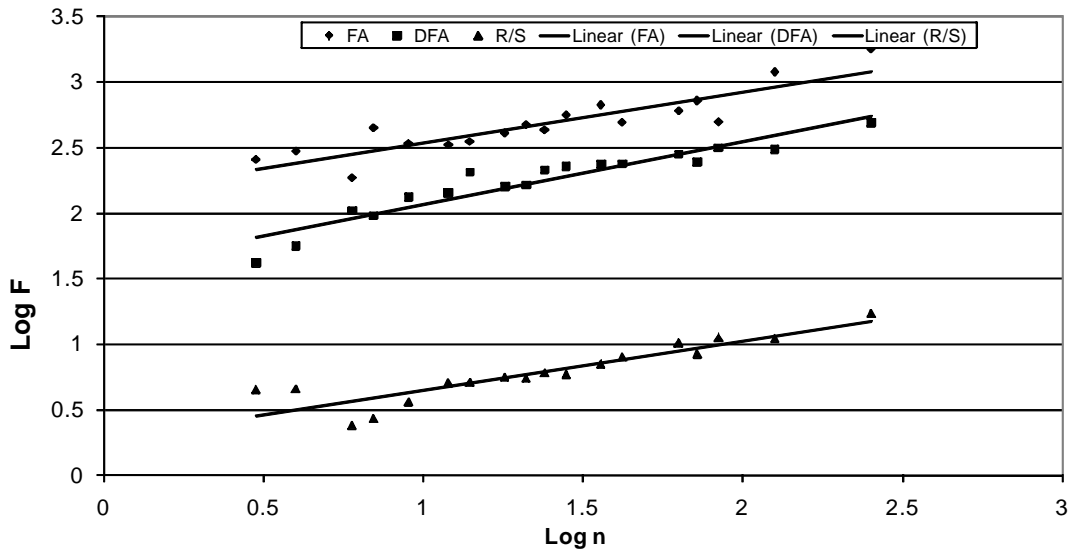


Fig. 6: Plots of R/S, FA and DFA for Enugu (South-East)

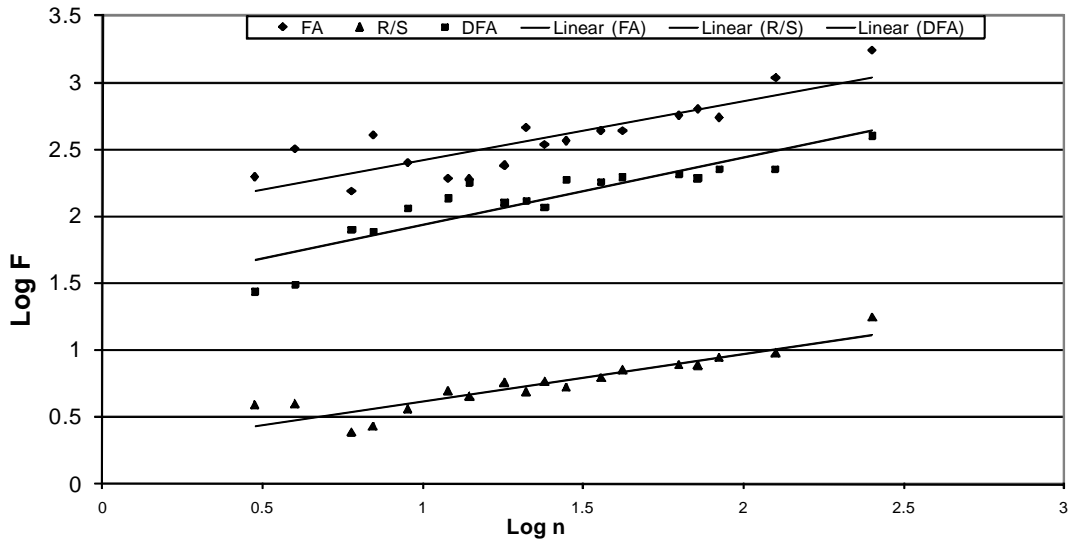


Fig. 7: Plots of R/S, FA and DFA for Bauchi (North-East)

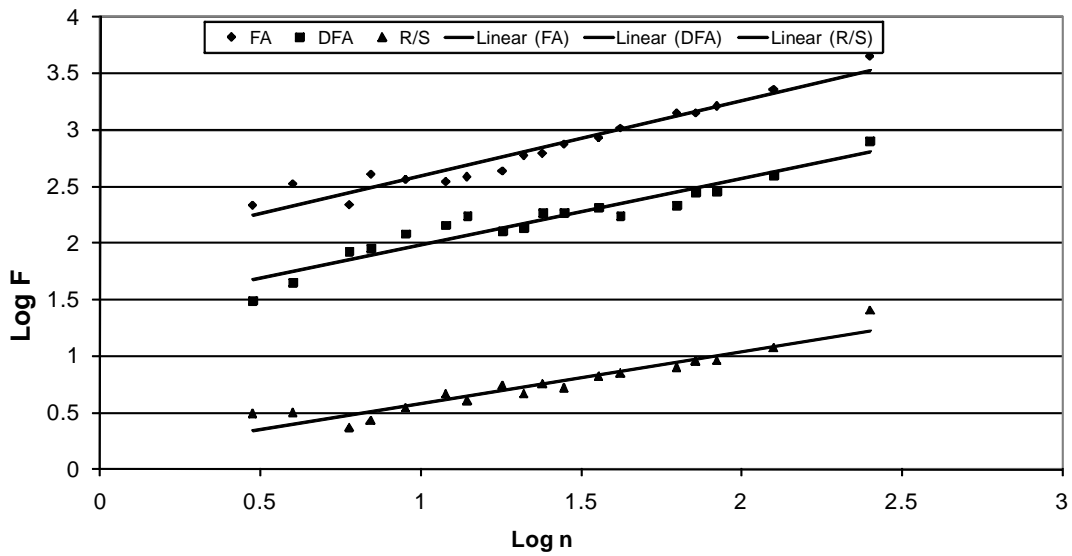


Fig. 8: Plots of R/S, FA and DFA for Kano (North-West)

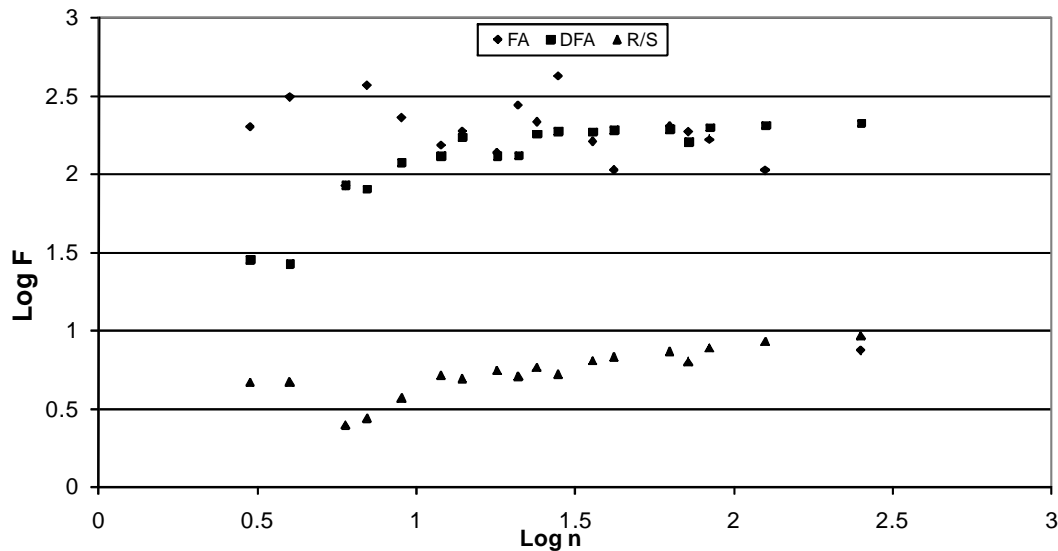


Fig. 9: Plots of R/S, FA and DFA for Jos (North-Central)

Figures 4 – 9 indicate that apart from variation in fluctuations, there are no marked variations in the values of h for the three methods. Besides, the R/S method seemed to predict the presence of SOC better. The DFA indicates that monthly rainfall exhibits SOC biannually for the North and during a fourteen month period for the South, the R/S analysis consistently indicates an annual SOC for both regions which is a more accurate representation of the yearly seasonal rainfall in the country. Kristoufek (2009) showed that rescaled range analysis can still stand the test against new methods. The R/S is known to usually overestimate Hurst exponent.

Despite this overestimation, studies have shown that even if R/S shows higher values of Hurst exponent than DFA, the standard deviations are lower for R/S so that the confidence intervals are narrower. Nevertheless, both methods show very similar estimates (Figures 4 - 9), when the bias is taken into consideration, whereas they are more correlated with growing time series length (Kristoufek, 2009). However, the result of the DFA could become more accurate and consistent if higher order DFAs (DFA_n) are

used instead of the linear DFA used in this study. It was also observed that the Hurst exponents calculated using DFA have a stronger tendency to converge to a single line at 0.5 (Figure 10) for all the regions and window sizes, while those from the R/S and FA analyses maintained clearly distinct lines near 0.5 (Figs 11 and 12) for different regions. This implies that even though the R/S analysis was more accurate in predicting the crossovers, the DFA gave better and finer results after the crossovers.

The result of the standard fluctuation analysis (FA) did not exhibit any distinct demarcation between the region of SOC and the region of random walk as did the other two methods (Figure 3). But it tends to show that monthly rainfall is purely random. Though the three methods used in this study give results of varying degrees to an extent, they all seem to point to randomness in monthly rainfall distribution over the 21 year period of study. Figures 10 to 12 are plots of Hurst exponent (h) versus window size (n) which show that for all the regions, Hurst exponent decreases from around 1 to 0.5, remaining at 0.5 at $n = 42$ months for the

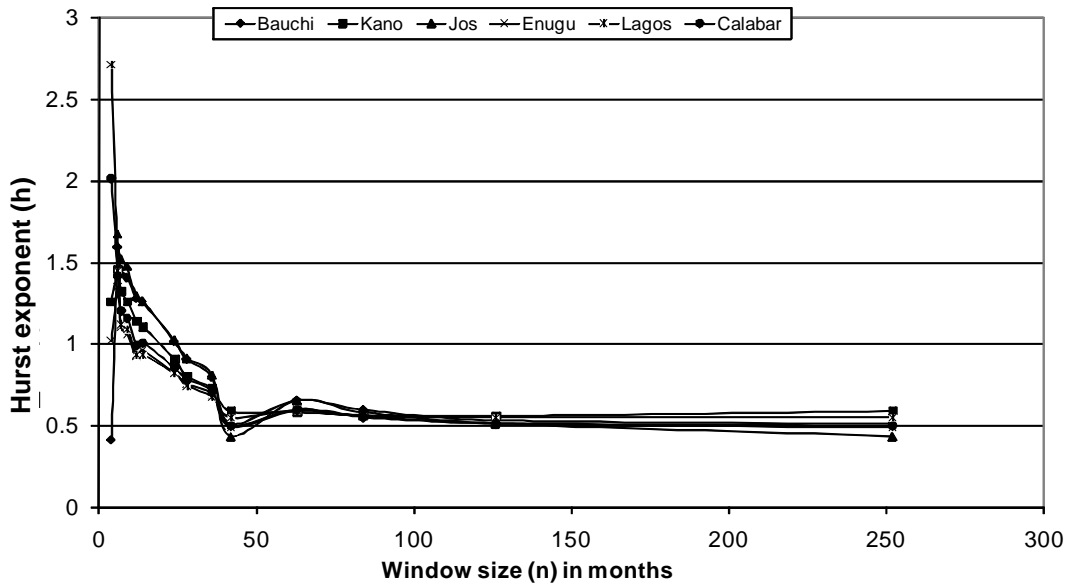


Fig. 10: Plot of Hurst exponent (h) versus windows size (n) for DFA analysis

DFA and 36 months for the R/S analysis, respectively. This seems to show that monthly rainfall has a memory span of about three years. This implies that the rainfall of a particular year will be influenced by that of the previous year to an extent, and will, in turn, influence that of the following year. However, as

time passes, this influence decays so that, when monthly rainfall data for so many years are considered, they appear to follow a random pattern. This can be interpreted to mean that monthly rainfall exhibits positive correlations in the short range and anti-persistence in the long run.

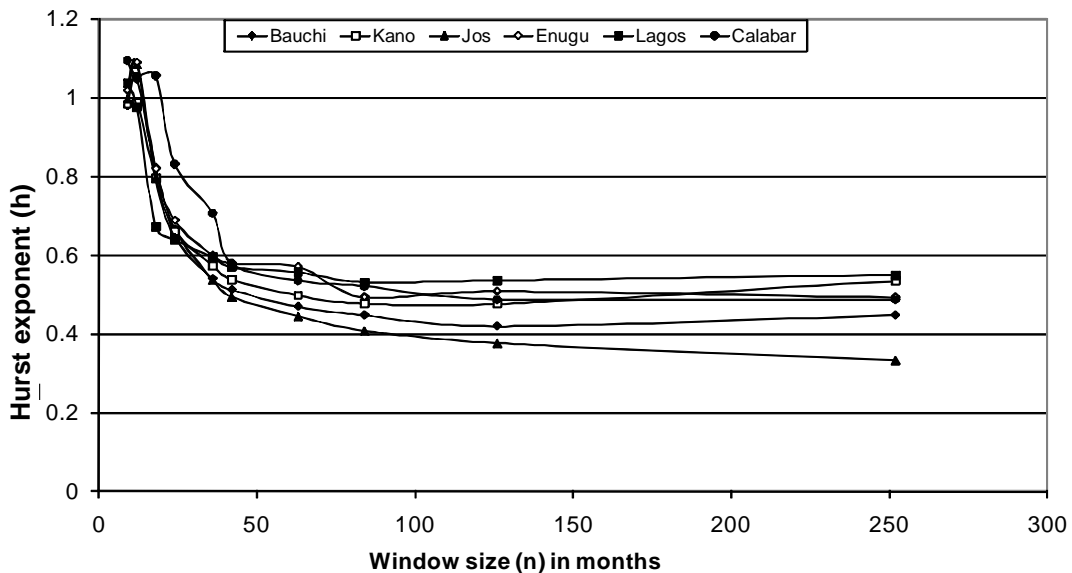


Fig. 11: Plot of Hurst exponent (h) versus windows size (n) for R/S analysis

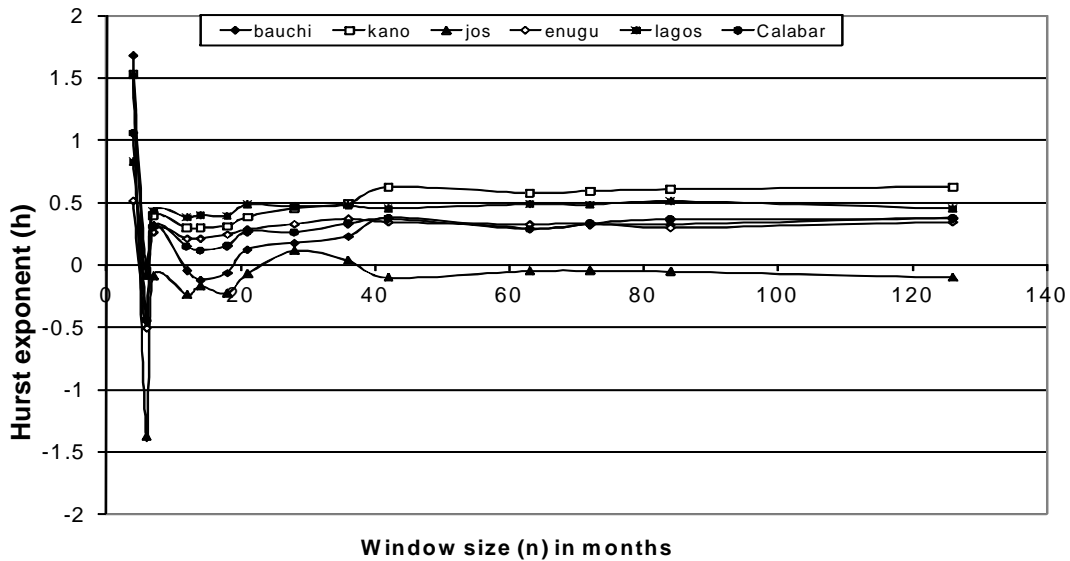


Fig. 12: Plot of Hurst exponent (h) versus windows size (n) for FA analysis

In most developing countries (Nigeria inclusive) there is a dearth of hydro-meteorological data as a result of absence of functional and elaborate hydro-meteorological stations. This has made Engineers and Hydrologists to resort to unreasonable interpolations in attempts to fill in missing data. However, results from this study show that rainfall has a memory span which must be taken into consideration during interpolation. The implication is that there is no substitute for accurate data gathering especially when dealing with random processes.

CONCLUSION

The monthly rainfall in the different regions of Nigeria exhibit a tendency to SOC in the short range of between 12 and 14 months, and a positive correlation between one and three years, but a tendency of randomness in the long run. The DFA and the R/S analysis proved to be better tools for isolating a region of SOC from a seeming random time series. The R/S analysis was more consistent in bringing out the seasonal rainfall pattern in Nigeria than the other two methods. The standard fluctuation analysis did not give very good results compared to the other two methods.

REFERENCES

- Bak, P., Christensen K., Danon L. and Scanlon T. (2002). Unified scaling law for earth quakes. *Phys Rev Lett* 88:178501-1-178501-4
- Bak P., Tang C. and Wiesenfeld K. (1988). Self organized criticality. *Phys Rev A* 38:364-374.
- Bak P., Tang C. and Wiesenfeld, K. (1987). Self-organized criticality: An explanation of 1/f noise. *Phys Rev Lett*. 59:381-384.
- Barabasi, L. R. (1999). Emergence of scaling in random networks. *Science* 286:509-512.
- Braun, P., Bunde, A. and Havlin, S. (2003). Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies. Preprint submitted to Elsevier Science.
- Bunde, A., Havlin, S. and Kantelhardt. J. W., Penzel, T., Peter, J., and Voigt, K. (2000). Correlated and uncorrelated regions in heart-rate fluctuations during sleep. *Phys.*

- Rev. Lett.*, 85(17):3736-3739.
- Chen, Z., Ivanov, P., Hu, K. and Stanley, H. E. (2002). Phys. Rev. E 65, preprint physics/0111103.
- Czarnecki, L., Grech, D. and Pamula, G. (2008). Comparison study of global and local approaches describing critical phenomena on the Polish stock exchange market, *Acta Physica Polonica A*.
- Grech, D. and Mazur, Z. (2004). Can one make any crash prediction in finance using the local Hurst exponent idea?, *Physica A*, Vol. 336.
- Hu, K., Ivanov, P., Chen, Z., Carpena, P. and Stanley, H. (2001) Effect of trends on detrended fluctuation analysis, *Physical Review E*, Vol. 64, 2001 Phys. Rev. E 64, 011114.
- Hurst, H.E. (1951). Long term storage capacity of reservoirs, *Transactions of the American Society of Engineers*, Vol. 116: 770 - 779
- Jianlin, Z., Guangming Z., Xing, Z., Guohe, H. and Yimin, J. (2005). Self organized criticality behaviour of acid deposition. *Air Water and Soil Pollution*, 162: 295 - 313.
- Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H., Havlin, S. and Bunde, A. (2001). Detecting long-range correlations with detrended fluctuation analysis. *Physica A*, 295:441-454.
- Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Bunde, A., Havlin, S. and Stanley, H. E (2002). Multifractal Detrended Fluctuation Analysis of Non-stationary Time Series. *Physica A* 316: 87 - 114.
- Koscielny-Bunde, E., Bunde, A., Havlin, S., Roman, H.E., Goldreich, Y. and Schellnhuber, H.-J., (1998). Indication of a universal persistence law governing atmospheric variability. *Phys. Rev. Lett.*, 81(3):729-732.
- Kristoufek, L. (2009). R/S analysis and DFA: Finite sample properties and confidence intervals. Munich Personal RePEc Archive, Paper No. 16446.
- Linkenkaer-hansen, K. (2002). Self-organized criticality and stochastic resonance in the human brain. Ph.D Thesis submitted to the Department of Engineering Physics and Mathematics, Helsinki University of Technology, Finland.
- Lucas, C. Self-organizing systems (SOS) FAQ (Version 2.7). <http://www.calresco.org/sos/-sosfaq.htm>. May 24, 2002.
- Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*. W. H. Freeman, New York.
- Matos, J., Gama, S., Ruskin, H., Sharkasi, A. and Crane, M. (2008). Time and scale Hurst exponent analysis for financial markets, *Physica A*, Vol. 387(15).
- Peng, C., Buldyrev, S., Havlin, S., Simons, M., Stanley, H. and Goldberger, A. (1994). Mosaic organization of DNA nucleotides, *Physical Review E*, 49(2):1685.
- Sornette, D. (2002) Proc. Natl. Acad. Sci. USA 99, Suppl. 1:2522–2529.
- Turcotte, D. L. (1997). *Fractals and Chaos in Geology and Geophysics* Cambridge Univ. Press, Cambridge, U.K. (2nd Ed.)
- Turcotte, D. L., Malamud, B. D., Guzzetti, F. and Reichenbach, P. (2002) Proc. Natl. Acad. Sci. USA 99, Suppl. 2522-2529.
- Vjushin, D., Govindan, R. D., Brenner, S., Bunde, A., Havlin, S. and Schellnhuber, H. J. (2002). Lack of scaling in global climate

models. *J. Phys. Condens. Matter*

Time Series Analysis of Monthly Rainfall... 151

14:2275 – 2282.