

MODELLING DROUGHT SERIES IN NORTHERN NIGERIA USING GENERALISED AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY

Chukwu, S. Emeka*, Otache, Y. Martins*, Atemoagbo, O. Precious*

*Department of Agricultural & Bioresources Engineering, Federal University of Technology, Minna, Nigeria.

sechukwu6@gmail.com

ABSTRACT

The existence of time-dependent variance or conditional variance, commonly called heteroscedasticity, in hydrologic time series has not been thoroughly investigated. The various physical mechanisms governing the dynamics of drought series act on a seemingly wide range of temporal and spatial scales; almost all the mechanisms involved present some degree of nonlinearity. Against the backdrop of these issues, this paper deals with modelling the heteroscedasticity in the residuals of the Autoregressive Integrated Moving Average (ARIMA) model using a Generalised Autoregressive Conditional Heteroscedasticity (GARCH) model. Attempt was made to critically evaluate generalised autoregressive conditional heteroscedasticity (GARCH) or volatility of drought series at SPI-3 and SPI-9 timescale resolution. It was also evident that the traditional seasonal Autoregressive Moving Average (ARMA) models are inadequate in describing ARCH effect in SPIs series process. Therefore, the GARCH modelling approach is necessary to capture the heteroscedasticity remaining in the residuals of a ARIMA model. Considering this findings, the potential for a hybrid Autoregressive Moving Average (ARIMA) and Generalised Autoregressive Conditional Heteroscedasticity (GARCH)-type models should be further explored and probably embraced for modelling higher temporal accumulation of SPI-12, SPI-24 and SPI-48 in view of the relevance of statistical modelling in hydrology.

Keywords: Drought, Heteroscedasticity, ARIMA, Stochastics and Autocorrelation

INTRODUCTION

When modelling hydrologic time series, the focus usually is on modelling and predicting the mean behaviour, or the first order moments, and rarely, is it concerned with the conditional variance, or their second order moments; although unconditional season-dependent variances are usually considered (Otache, *et al.* 2012). The increased importance played by risk and uncertainty considerations in water resources management and flood control practice, as well as in modern hydrology theory, however, has necessitated the development of new time series techniques that allow for the modelling of time varying variances. It is not hard to find evidence to argue that a time series with random appearance might be nonlinear dynamic; but it suffices to note that the difficulty is in telling what kind of nonlinear dynamics (Otache, *et al.* 2012). To a large extent, one can think of heteroscedasticity as time-varying variance (i.e., volatility); in this regard, a univariate stochastic process is said to be homoscedastic if variances of the time series are constant. The “conditional”

term as used in the definition of Autoregressive Conditional Heteroscedasticity (ARCH) or Generalised Autoregressive Conditional Heteroscedasticity (GARCH)-type models, implies a dependence on the observations of the immediate past, while autoregressive describes a feedback mechanism that incorporates past observations into the present. Thus, GARCH can be described as a mechanism that includes past variances in the explanation of future variances (Wang, *et al.* 2005). It allows users to model the serial dependence of volatility, as it takes into account excess kurtosis (i.e., fat tail behaviour), which is very common in hydrologic processes or drought series and volatility clustering. Volatility clustering as used here depicts a phenomenon in which large changes tend to follow large changes, and small changes tend to follow small changes. In either of these cases, the changes from one period to the next are typically of unpredictable sign or orientation. Basically, the concept of volatility suggests a time-series model in which successive disturbances, although uncorrelated, are nonetheless serially dependent. A time series

with the ARCH property has two basic components: conditional mean, and a conditional variance function; thus, the nonlinearity in the series may probably come from the nonlinearity of the conditional variance (Otache, *et al.* 2012). As opined by Hauser *et al.* (1998) there is growing interest in nonlinear analysis of hydrologic time series, relatively few efforts have been done in this field. The generalised autoregressive conditional heteroscedasticity (GARCH) approach, which is commonly used in modelling the time variation of the second order moment or the variance of financial time series, can be an appropriate method for nonlinear modelling of hydrologic time series. Although the time variation of the variance of hydrologic variables has been mentioned in the literature, few studies have applied the GARCH approach to model this phenomenon in hydrologic variables. For instance, in a pioneering study, Wang *et al.* (2005) applied the GARCH approach to model the heteroscedasticity of daily and monthly streamflow time series of the upper Yellow River at Tangnaihai in China. Concluded that the conventional linear time series models, the autoregressive model and the deseasonalised autoregressive moving average (ARMA) model, are not sufficient to describe the time-dependent variance of streamflow and that a GARCH model needs to be fitted to the residuals of an ARMA model to capture the time variation behaviour of the streamflow variance. Chen *et al.* (2008) applied linear ARMA and nonlinear ARCH models to model 10-day streamflows of the Wu-Shi River in Taiwan and verified that nonlinear time series models are superior to the traditional linear approaches such as an ARMA model. Reported an increase of the coefficient of efficiency (CE) from 0.28 for an ARMA model to 0.76 for an ARCH model whereas the mean absolute error (MAE) was reduced from 60.45 for an ARMA model to 41.35 for an ARCH model. However, no seasonal or integrated time series model was applied to compare with a GARCH model in their study. In another application of the GARCH model, Romilly (2005) fitted a seasonal

autoregressive integrated moving average (SARIMA) model to the global mean monthly temperature and mentioned the existence of heteroscedasticity in the residuals and the need for applying a GARCH model to remove it from the residuals. However, the model comparison in this study indicated that the SARIMA model performs slightly better than the GARCH model for the global mean temperature time series modelling according to $R^2 = 0.772$, mean absolute error (MAE = 0.084) and the root mean squared error (RMSE = 0.111) for the SARIMA model against $R^2 = 0.771$, MAE = 0.084 and RMSE = 0.112 for the GARCH model.

ARCH-type models can generate accurate forecasts of future volatility, especially over short horizons, therefore providing a better estimate of the forecast uncertainty which is available for water resource management and flood control Wang *et al.* (2005); as such, ARCH-type models could be very useful for hydrologic time series modelling and fairly capable of reproduce the asymmetric periodic behaviour with large fluctuations around large streamflow and small fluctuations around small streamflow are available (Otache, *et al.* 2012). These fluctuations though, basically can be handled with the conventional time series models that take season-dependent variance into account, such as Periodic Autoregressive Moving Average (PARMA) models and deseasonalised Autoregressive Moving Average (ARMA) models. However, little attention has been paid so far by the hydrologic community to test and model the possible presence of the ARCH effect with which large fluctuations tend to follow large fluctuations, and small fluctuations tend to follow small fluctuations in either standardised precipitation index (SPI) time series vis-à-vis drought series and dearth of evidence of related studies in the proposed regional. In this regard therefore, it is imperative to realised that both conditional heteroscedasticity and asymmetric seasonality in the mean and variance could also be probable sources of nonlinearity within the

overall context of nonlinear determinism itself (Otache, 2008).

Against the backdrop of all this, the objective of this study is to test for the existence of ARCH effect in both the SPI-3 and SPI-9 drought series in northern, Nigeria. In addition, if there is the existence of ARCH effect, propose an ARMA-GARCH error model for the drought series since the ARCH theory admits non-normality of the unconditional distribution of the data. Thus, with the assumption of normality of the conditional distribution, an ARCH-type structure could be built to capture the time-dependent variances.

2.0 MATERIALS AND METHODS

2.1 Study Location and Data Assembly

The study location is located between 10°N - 14°N and 4°E - 7°5'E. It belongs to arid and semi-

arid region of the country, precisely of a predominantly Sudan vegetation. It is characterised by a distinct bi-seasonal weather pattern; i.e., wet and dry. The wet season starts in April and ends in October, while the dry season starts in November and ends in March (Sombroek and Zonneveld, 1971). Figure 1 shows the map of Hydrological Areas (HAs) in Northern Nigeria. For this study, historical rainfall time series of over 48years (1971-2019) from the region was used. To do this, mean monthly rain gauge rainfall values (i.e., point rainfall) for substantial decadal time period were collected from NIMET and River Basin Development Authority Zonal offices across the catchment States of Katsina, Kano, Minna and Sokoto respectively

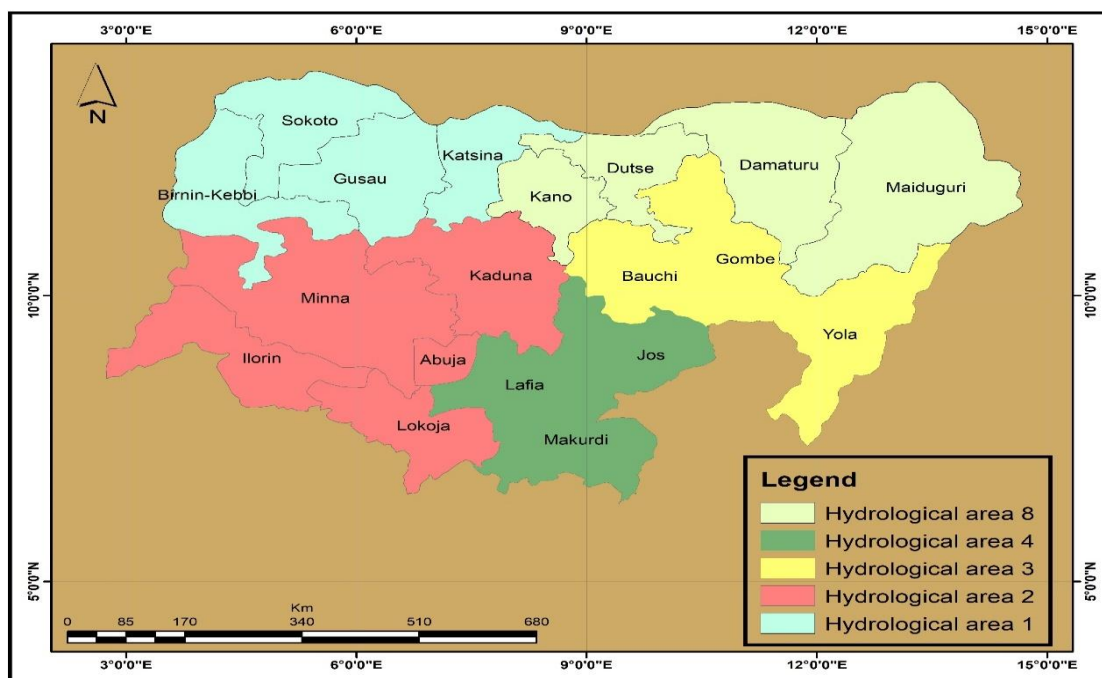


Figure 1: Hydrological areas in Northern Nigeria

2.2 SPI Computation

In the computation of the Standardised Precipitation Index (SPI), a gamma probability density was fitted to the given frequency of precipitation (rainfall) totals. The SPI computed for several accumulation time (SPI-3 and SPI-9). The parameters of the gamma probability distribution were first applied to find the

cumulative probability of the observed series in line with Mckee et al. (1995).

2.3 Tests for ARCH Effect

2.3.1 Identification of Appropriate Model

The detection for the ARCH effect of a drought series is actually a test of serial independence applied to the serially uncorrelated fitting error of

some model, usually a linear stochastic autoregressive (AR) model. Suppose a stochastic process Y_t is generated by an $AR(p)$ process:

$$Y_t = \sigma_0 + \sum_{j=1}^p \sigma_0 Y_{t-j} + \varepsilon_t \quad (1)$$

There exists an information set $\Psi_{t-1} = \{y_{t-1}, y_{t-2}\}$

$$\varepsilon_t / \Psi_{t-1} \sim N(0, h) \quad (2)$$

Where

$$h_t = \theta_0 + \sum_{i=1}^k \theta_i \varepsilon_{t-i}^2 \quad (3)$$

$$\theta_0 > 0, \theta_i \geq 0, i = 1, \dots, k$$

to ensure, that the conditional variance is positive. The process Y_t is called $AR(p)$ with $ARCH(k)$ errors. It is important to note that the conditional variance function might be complicated for a long series, which means that k in Equation (3), the lag of the conditional variance (ε^2), is large. If this situation exists, the computation becomes increasingly burdensome and interpretation, difficult (Otache, *et al.* 2012).

2.3.2 Model Adequacy Testing

Thus to test for the presence or otherwise of ARCH effect, the Engle's Lagrange Multiplier (LM) test and Ljung-Box-Pierce Q-test for a departure from randomness based on autocorrelation function of the data series were adopted and implemented using R programming language. The test statistic for LM test is given by TR^2 , where R is the sample multiple correlation coefficients computed from the regression of ε_t^2 on a constant, and $\varepsilon_{t-1}^2, \dots, \varepsilon_{t-q}^2$ T is the sample size. Under the null hypothesis that there is no ARCH effect, the test statistic is asymptotically distributed as Chi-square distribution with q degrees of freedom. On the other hand, Ljung-Box- Pirece test is a modification of the Box-Pierce test. The Ljung-Box-Pirece test for ARCH effect evaluated the autocorrelation function of the squares of the data, and tests whether the first L autocorrelations for the squared residuals are collectively small in magnitude. For fixed, sufficiently large L , the

Ljung-Box- Pierce Q-statistic is defined according as

$$Q = N(N+2) \sum_{k=1}^L \frac{\hat{r}^2 k(\varepsilon^2)}{N-k} \quad (4)$$

where, N is the sample size, and $\hat{r}^2 k(\varepsilon^2)$ is the squared sample autocorrelation of squared residual series at lag k . Like the Engle's LM test, under the null hypothesis of a linear generating mechanism for the data, namely, no ARCH effect, the test statistic is asymptotically $\chi^2(l)$ distributed.

2.4 Modelling Framework and Development

Attempts have been made by several researchers to model both economic series and hydroclimatic processes by combining ARMA models with ARCH errors typical of (Wang, *et al.*, 2005, Weiss, 1984 and Hauser, *et al.*, 1998). The same framework was adopted in this study; here, the SPI-3 and SPI-9 series was modelled by using ARIMA- GARCH. In this regard, the ARIMA- GARCH model employed can be interpreted as a direct combination of an ARIMA model which was used to model the mean behaviour and the ARCH, for ARCH effect in the residual series from the model. The general GARCH (p, q) model for the conditional variance of innovations is:

$$\lambda_t = k + \sum_{i=1}^p G_i \lambda_{t-i} + \sum_{j=1}^q A_j \varepsilon_{t-j}^2 \quad (5)$$

With constraints:

$$\begin{cases} \sum_{i=1}^p G_i + \sum_{j=1}^q A_j < 1 \\ k > 0 \\ G_i \geq 0 \quad i = 1, 2, \dots, p \\ A_j \geq 0 \quad j = 1, 2, \dots, q \\ G_i = \text{is the ARCH compt} \end{cases} \quad (6)$$

where,

ε_t denotes a real-valued discrete time stochastic process, p and q represent the order of the GARCH(p, q) conditional variance model; k is a constant, GARCH: represents the p -element coefficient vector G_i and ARCH represents the q -element of the coefficient vector A_j .

3. 0 RESULTS AND DISCUSSION

3.1 Heteroscedasticity and modelling of SPI series

3.1.1 Determination of Appropriate Arima Model for Heteroscedasticity Analysis

Figure 2 show that the linear stochastic models fit the respective SPIs series appropriately, as there are no significant serial correlations. Despite this though, while the residuals seemed statistically

uncorrelated according to the autocorrelation function (ACF) values (*i.e.*, Figure 3 shows that the autocorrelation values are not identically distributed; that is, the residuals are not independent and identically distributed over time. This situation, basically, connotes volatility, a common behaviour of GARCH process.

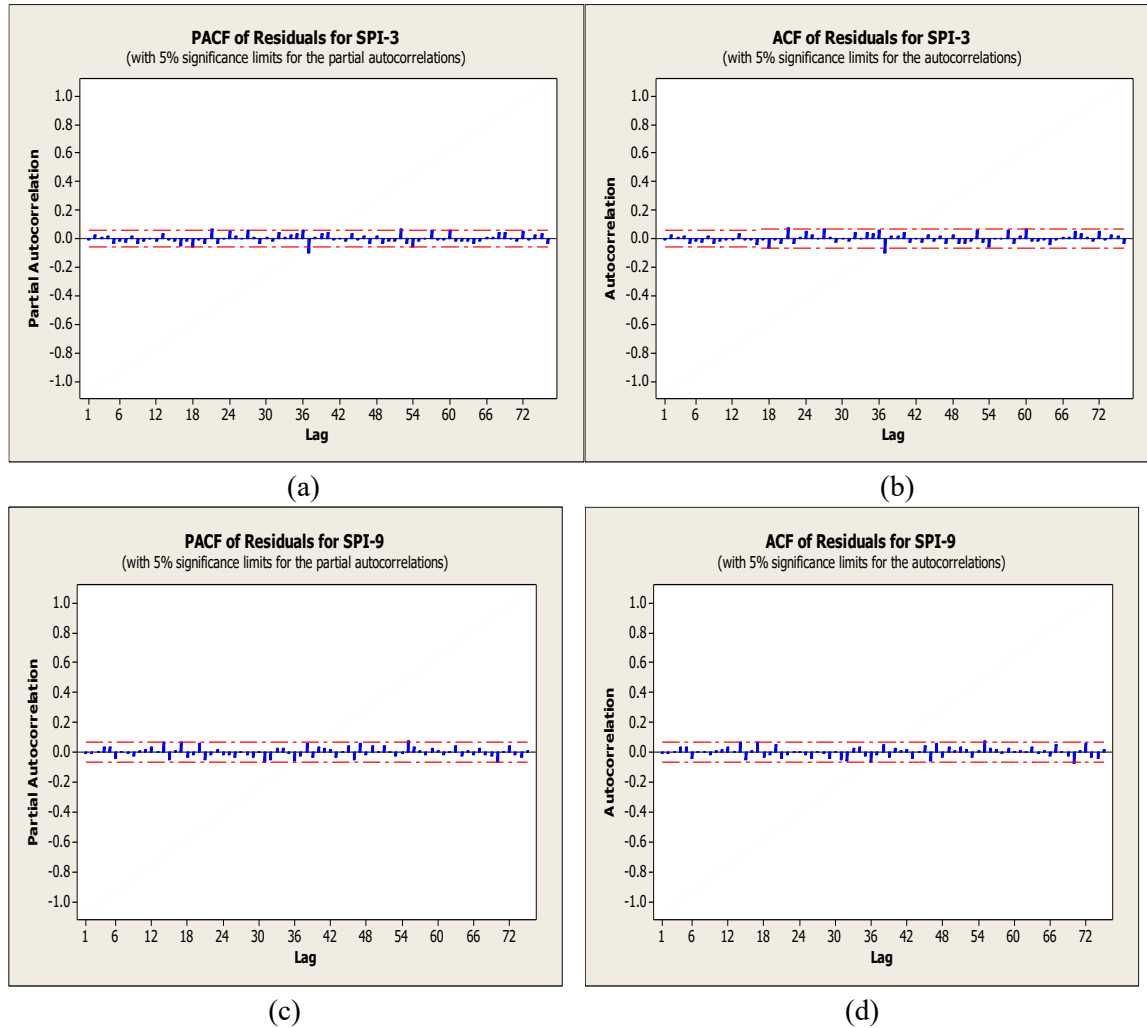
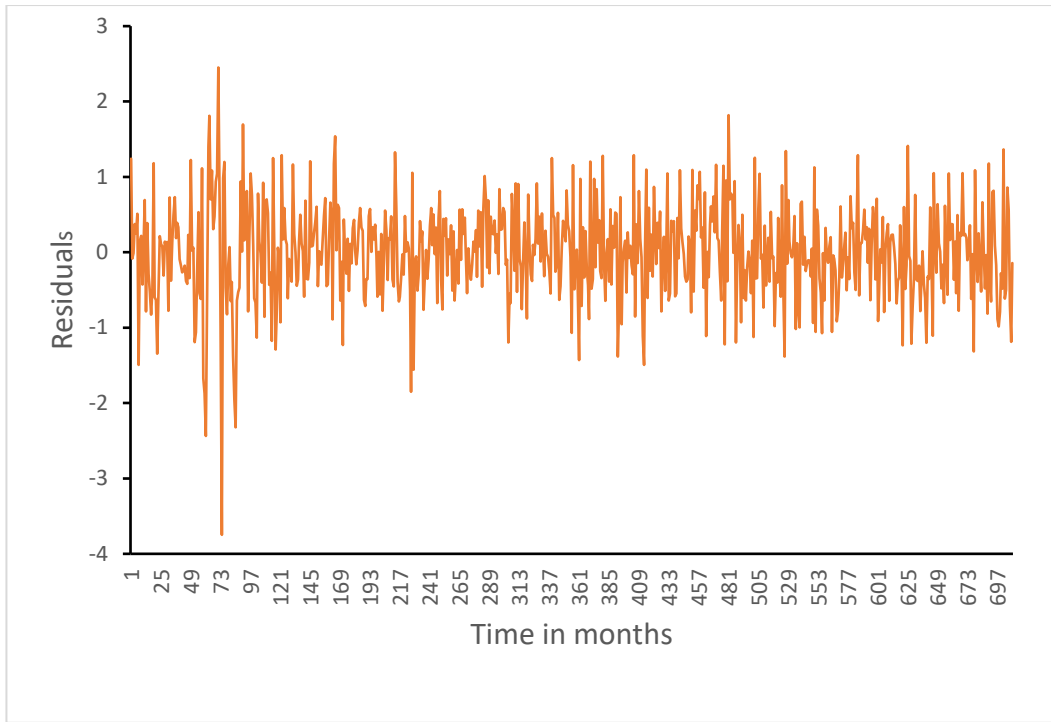
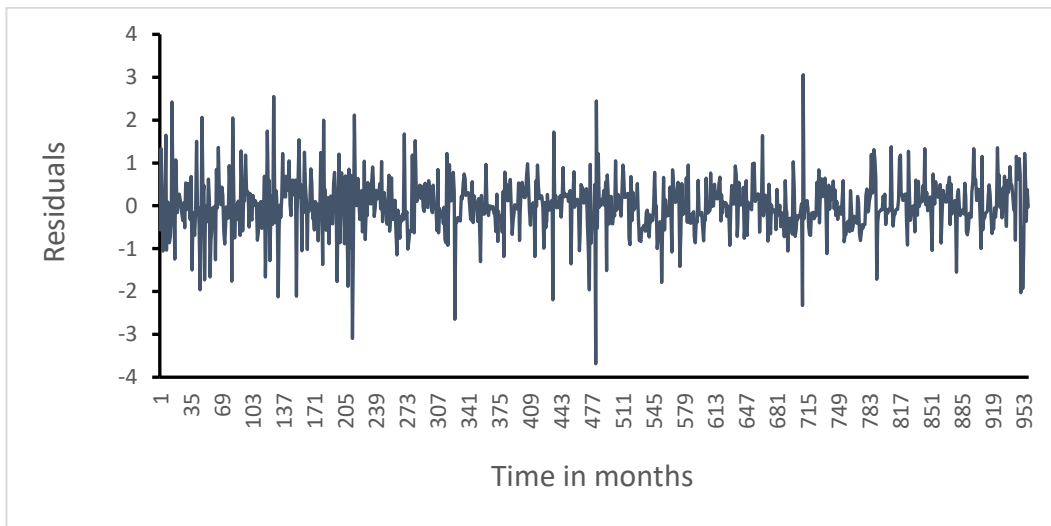


Figure 2: Correlogram for (a) ACF of SPI-3 ARIMA (1,1,3) x (1,1,1)₁₂ (b) PACF of SPI-3 ARIMA (1,1,3) x (1,1,1) (c) ACF of SPI-9 ARIMA (2,1,2) and (d) ACF of SPI-9 ARIMA (2,1,2)



(a)



(b)

Figure 3: Segments of the Residual Series from (a) SPI-3 ARIMA (1,1,1,3) X (1, 1, 1)₁₂ model, and (b) SPI-9 ARIMA (2,1,2) model.

Though, the residuals seemed almost uncorrelated over time, the squared residuals (SR) and Standardised Squared Residual (SSR) are clearly correlated and appeared to be identically same, there is no distinctive different between SR and SSR in both SPI-3 and SPI-9 as shown in (Figure 4). The autocorrelation structures of both squared residual series still exhibit traces of strong seasonality as shown in Figure 5 with a lot of Spikes exceeding the confident bounds at 5%

significant limits. In line with this, Figure 6 shows P-values of the Engle's test, as all the values are less than 0.05 significant level. The physical implication of this is that the variance of residual series is conditional on its past history; that is, the residual series may probably exhibit ARCH effect. In other words, the heteroscedasticity of the residuals of the ARIMA (1,1,3) x (1,1,1)₁₂ and ARIMA (2,1,2) of SPI-3 and SPI-9 respectively, is significant, and there is a needed to apply a

GARCH approach to remove it from the residuals
of the ARIMA (1,1,3) x (1,1,1)₁₂ and ARIMA
(2,1,2) in the models.

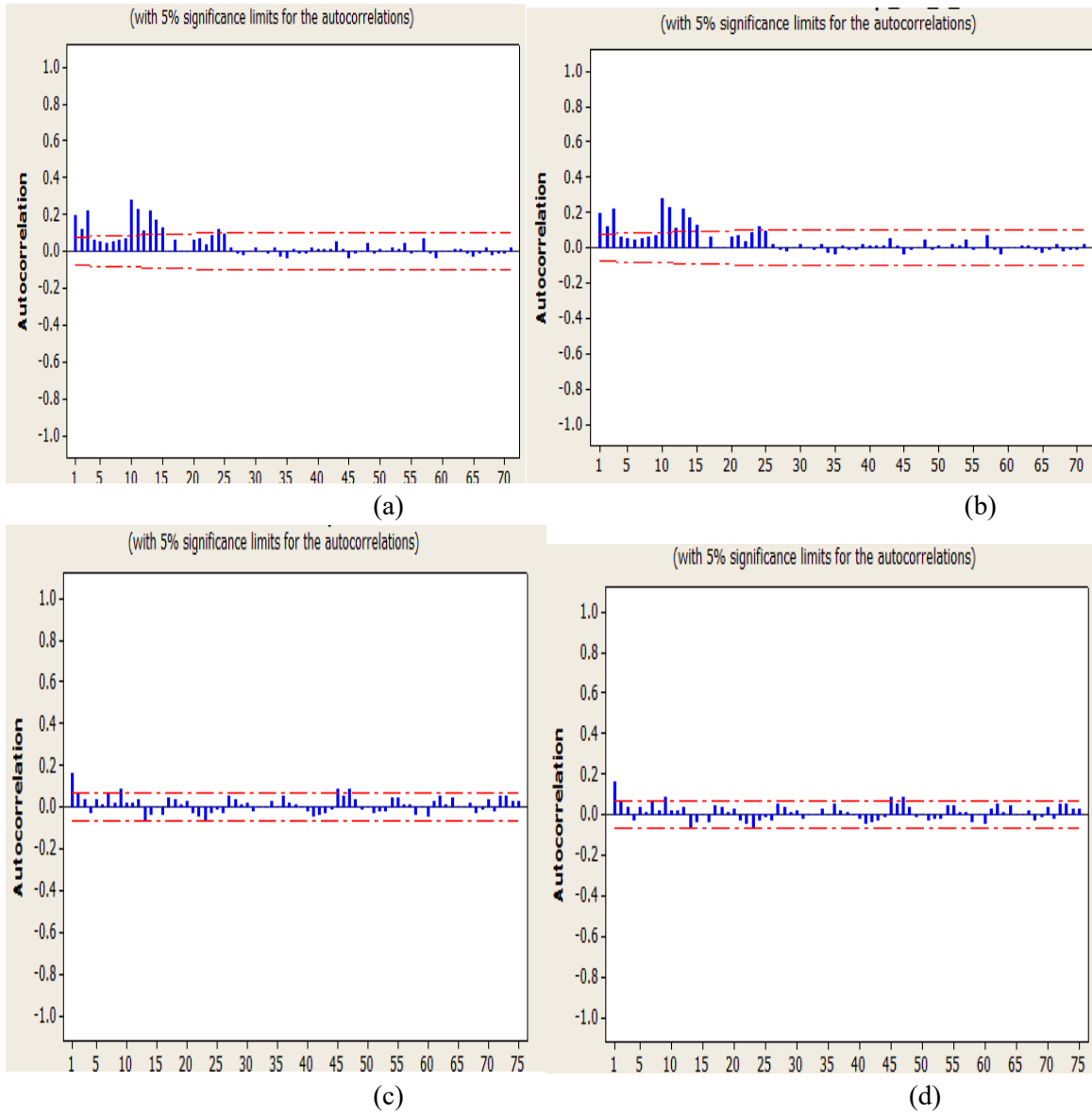


Figure 5: Autocorrelation Function of (a) SPI-3 SR (b) SPI-3 SSR (c) SPI-9 SR and (d) SPI-9 SSR

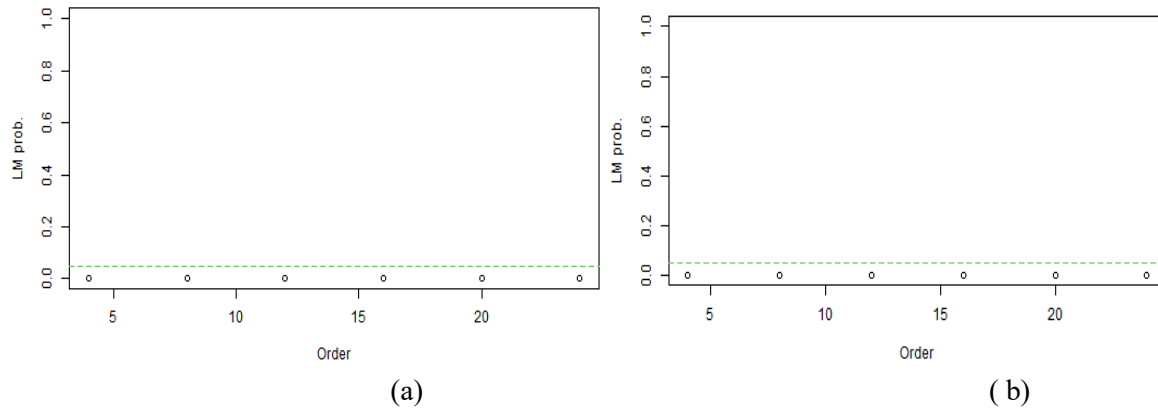


Figure 6: P-values of the Engle's test for SR (a) SPI-3 ARIMA (1,1,3) x (1,1,1)₁₂ and (b) SPI-9 ARIMA (2,1,2)

3.1.2 Application of GARCH Model

According to the AIC, ACF and PACF structures, two GARCH models were found suitable; GARCH (1,2), *i.e.*, GARCH(1) ARCH(2) and GARCH(5,3) models were suitable for ARIMA (1,1,3) x (1,1,1)₁₂ of SPI-3. Whereas, the preliminary GARCH models fitted to SPI-9 ARIMA composed of ARIMA(3,1,1)-ARCH(4), and ARIMA (2,1,2)-GARCH(1,1).The adopted

models in both instances were selected based on the smallest AIC value for each of the cases considered as well as on the basis of model parameter parsimony and significance, the latter(s) were preferred in both cases. Table 1 and Table 2 shows the parameters for ARIMA (1,1,3)x(1,1,1)₁₂-GARCH(5,3) and ARIMA (2,1,2)-GARCH(1,1) for the SPI-3 and SPI-9 respectively.

Table 1: ARIMA (1,1,3) x (1,1,1)₁₂-GARCH (5,3) Model Fitting for SPI-3

Parameters	Parameter values	Standard error	t-value
C	0.041653	0.033009	1.2619
K	0.583496	0.124335	4.6929
GARCH(1)	0.298277	0.114177	2.6124
ARCH(1)	0.220329	0.061153	3.6029

Table 2: ARIMA (2,1,2)-GARCH (1,1) Model Fitting for SPI-9

Parameters	Parameter values	Standard error	t-value
C	0.332857	0.006262	3.06825
K	0.121267	0.040312	5.91624
GARCH(1)	0.441494	0.164781	2.67927
ARCH(1)	0.156477	0.072255	2.16563

To prove the superiority of GARCH (5,3) and GARCH (1,1) over conventional ARIMA, for both SPI-3 and SPI-9 respectively in terms of preservation of seasonality in the variance of the residuals. The supposed GARCH (5,3) and (1,1) model were fitted to the residuals of ARIMA (1,1,3) x (1,1,1)₁₂ and ARIMA (2,1,2) of the models respectively. Figure 7 shows the results for test for ARCH effect, the ACF of the SRs and SSRs ARIMA (1,1,3) x (1,1,1)₁₂-GARCH (5,3) and ARIMA (2,1,2)-GARCH (1,1). indicated no significant autocorrelation structure, in both SRs as all the coefficient at different lags are within the confidence interval, especially that of the SPI-3 series. Figure 7 shows no significant autocorrelation structure, in both standardised residual (SR) and squared standardised residual (SSR) as all the coefficient at different lags are within the confidence interval, especially for that of the SPI-3 series, even though there is no autocorrelation in the SRs (Figure 7(a and c)), which implies that the ARCH effect has been removed, but in the SSR (Figure 7(b and d)) there is still presence of weak autocorrelation which is

attributable to random process common in hydrologic time series. In addition, The P-values of the Engle's test and Ljung-Box test in Tables 3 - 6 attested to the adequacy of the choosen GARCH models in capturing constant varying variance. The Tables indicated very clearly that the ARCH effect has been removed as p-value in all the considered instances exceeded significant value of 0.05 at a substantial margin. This connote, that the specified GARCH model as well as the choosen linear stochastic model was appropriate and able to remove the seasonal heteroscedasticity from the seasonal hydrologic time series. However, the findings vividly establish the fact that data pre-whitening strategy is critical in linear stochastic modelling of hydro-climatic variables. Hence, because of the seasonality of the time series employed, there is need for the adoption of deseasonalisation pre-processing approach; both of the time sequence should be cleaned via classical harmonic analysis for the deseasonalisation process so as to remove the implication of periodicity.

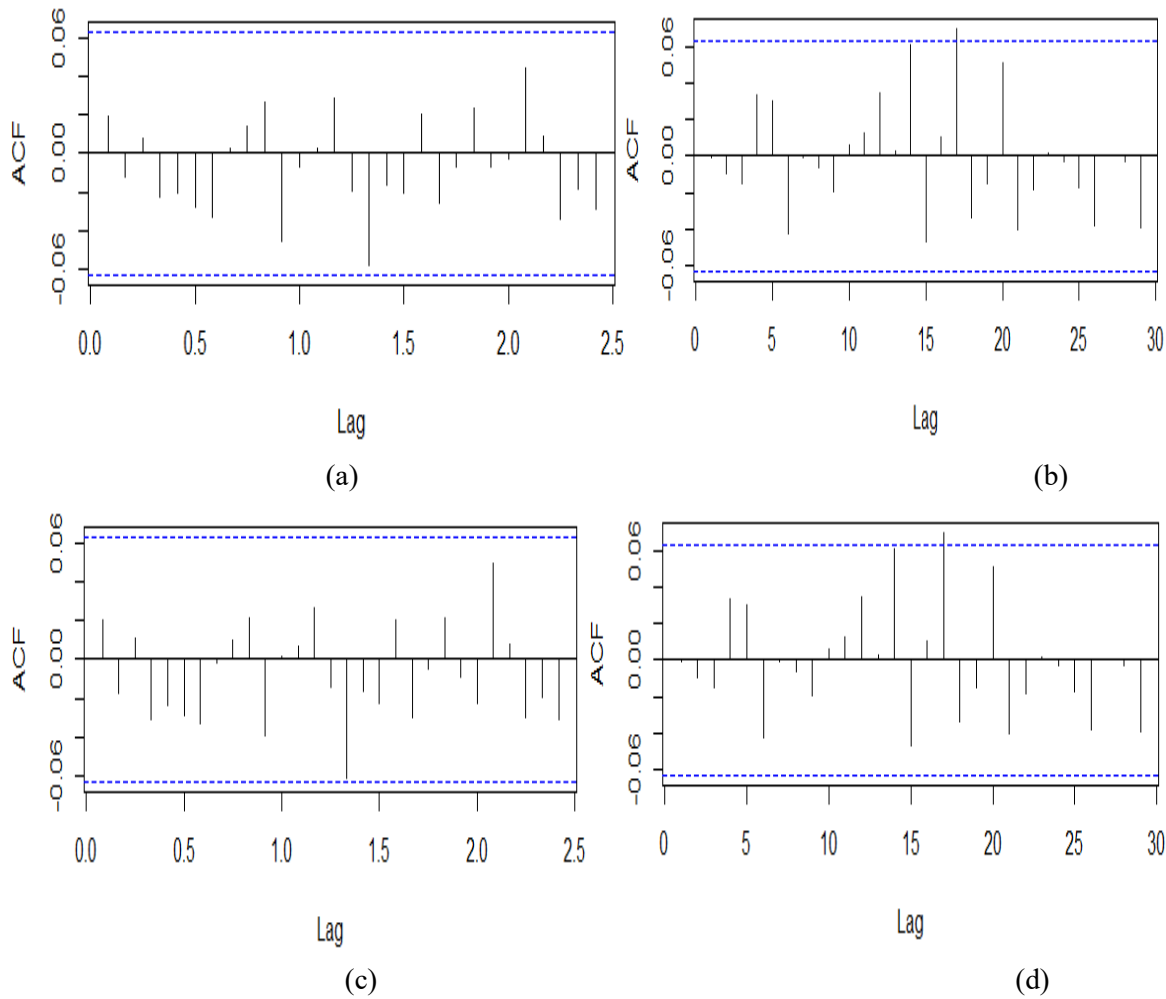


Figure 7 ACF of (a) SR of ARIMA (1,1,3) x (1,1,1)₁₂ -GARCH (5,3) (b) SR of ARIMA (2,1,2)-GARCH (1,1) (c) SSR of ARIMA (1,1,3) x (1,1,1)₁₂ -GARCH (5,3) and (d)) SSR of ARIMA (2,1,2)-GARCH (1,1)

Table 3. Ljung-Box test for SR of ARIMA (1,1,3) x (1,1,1)₁₂ -GARCH (5,3) of SPI-3

H	p-value	Statistic
0	0.9597	0.002554
0	1.0000	0.007535
0	1.0000	0.012112

Table 4. Engle's LM test for SR of ARIMA (1,1,3) x (1,1,1)₁₂ -GARCH (5,3) of SPI-3

H	p-value	Statistics
0	0.9612	0.002372
0	0.9998	0.005882
0	1.0000	0.008436

Table 5: Ljung-Box test for SSR of ARIMA (2,1,2)-GARCH (1,1) of SPI-9

H	p-value	Statistic
0	0.8191	0.05229
0	0.9826	0.31038
0	0.9993	0.39830

Table 6: Engle's LM test for of SSR of ARIMA (2,1,2)-GARCH (1,1) of SPI-9

H	p-value	Statistic
0	0.8560	0.03294
0	0.9905	0.07922
0	0.9991	0.11509

However, it can be assumed that the seasonal heteroscedasticity may be the result of the seasonality of the atmospheric and climatic factors and components that influence the seasonal variance of hydrological variables. Based on the results, it can be opined that GARCH that has the potential to capture nonlinear dynamics, and to a large extent reduce serial dependence and leptokurtosis, this is in accord with the findings of Otache *et al.*(2012); the third moment for a regular ARCH process is zero; therefore, unconditional distributions of the series must be symmetric.

CONCLUSION

It is obvious from available literature that discussions on the nonlinear mechanism, conditional heteroscedasticity in hydrological processes have been given adequate attention. Modelling data with time varying conditional variance could be attempted in various ways, non-parametric and parametric. Towards this end, the parametric approach was adopted here. The existence of the ARCH effect was verified in the residual series from linear models fitted to SPI-3 and SPI-9 series. Following from the tests, it is shown that the ARCH effect is caused by seasonal variation. It is also evident that the traditional seasonal ARIMA models are inadequate in describing ARCH effect in the SPI-3 and SPI-9. In the light of this, and taking the seasonal variances inspected in the residuals from linear stochastic models fitted to the SPI-3 and SPI-9 series into consideration, the idea of an ARIMA-GARCH model with seasonal components is worth exploring. The ARIMA-GARCH model is basically a combination of an ARIMA model, which is used to model the mean behaviour, and GARCH model, the ARCH effect in the residuals from the ARIMA model. Thus, to preserve the seasonal variation in variance in the residuals, the ARCH model should not be fitted directly to the residual series, but rather, to seasonally standardized residuals. In addition, findings here shows the adequacy of the chosen GARCH models in capturing constant varying variance.

There is clear indication that the ARCH effect has been removed as p-value in all the considered instances exceeded significant value of 0.05 at a substantial margin. This connote, that the specified GARCH model as well as the chosen linear stochastic model was appropriate and able to remove the seasonal heteroscedasticity from the seasonal hydrologic time series. However, the findings vividly establish the fact that data pre-whitening strategy is critical in linear stochastic modelling of hydro-climatic variables. Hence, because of the seasonality of the time series employed, there is need for the adoption of deseasonalisation pre-processing approach; both of the time sequence should be cleaned via classical harmonic analysis for the deseasonalisation process so as to remove the implication of periodicity.

REFERENCES

- Liu, C.H., Su, H.C. (2008). A nonlinear Time Series Analysis Using Two-stage genetic algorithms for streamflow forecasting. *Hydrological Processes*. 22, 3697–3711.
- Hauser, M. A. and Kunst, R. M. (1998). Fractionally Integrated Models with ARCH Errors: With an Application to the Swiss 1-Month Euromarket Interest Rate. *Review of Quantitative Finance and Accounting*, 10(1) 95-113. doi:10.1023/A:1008252331292
- McKee, T. B., Doesken, N. J. and Kleist, J. (1995). Drought monitoring with multiple time scales. Preprints, 9th Conference on Applied Climatology, 15-20 January, Dallas, Texas, American Meteorological Society, 233-236.
- Otache, M. Y. (2008). “Contemporary Analysis of Benue River Flow Dynamics and Modelling,” Ph.D. Dissertation, *Hohai University, Nanjing*.

Otache, M. Y., Ahaneku, I. E., Mohammed, A. S.
and Musa, J. J. (2012). Conditional
Heteroscedasticity in Streamflow Process:
Paradox or Reality? *Open Journal of
Modern Hydrology*, 2, 79-90.
doi.org/10.4236/ojmh.2012.24010.

Romilly P. (2005). Time series modeling of
global mean temperature for managerial
decision-making. *Journal of Environmental
Management*. 76, 61–70.

Wang, W., Van Gelder, P. H. A. J. M., Verjling,
J. K. and Ma, J. (2005). Testing and
Modelling. Autoregressive Conditional
Heteroscedasticity of Streamflow
Processes. *Nonlinear Processes in
Geophysic.*, 12, 55-66. doi:10.5194/npg-12-
55-2005.

Weiss, A. A. (1984). ARMA Models with ARCH
Errors, *Journal of Time Series Analysis*,
Vol. 5, No. 2, 1984, 129- 143.
doi:10.1111/j.14679892.1984.tb00382.x.