DOI: 10.2478/jwld-2018-0032

JOURNAL OF WATER AND LAND DEVELOPMENT
2018, No. 37 (IV–VI): 127–140
PL ISSN 1429–7426, e-ISSN 2083-4535

Available (PDF): http://www.itp.edu.pl/wydawnictwo/journal; http://www.degruyter.com/view/j/jwld

Received 06.09.2017
Reviewed 11.11.2017
Accepted 30.11.2017

A – study design
B – data collection
C – statistical analysis
D – data interpretation
E – manuscript preparation
F – literature search

Stochastic ARIMA model for annual rainfall and maximum temperature forecasting over Tordzie watershed in Ghana

Mexoese NYATUAME[1, 2] ABCDEF ✉, Sampson K. AGODZO[2] DE

[1] Ho Technical University, Department of Agricultural Engineering, Ho, Volta Region, Box 217, Ghana; e-mail: doga_nyatuame@yahoo.com
[2] Kwame Nkrumah University of Science and Technology, Department of Agricultural and Biosystems Engineering, Kumasi, Ashanti Region, Ghana; e-mail: skagodzo7@usa.net

For citation: Nyatuame M., Agodzo S.K. 2018. Stochastic ARIMA model for annual rainfall and maximum temperature forecasting over Tordzie watershed in Ghana. Journal of Water and Land Development. No. 37 p. 127–140. DOI: 10.2478/jwld-2018-0032.

Abstract

The forecast of rainfall and temperature is a difficult task due to their variability in time and space and also the inability to access all the parameters influencing rainfall of a region or locality. Their forecast is of relevance to agriculture and watershed management, which significantly contribute to the economy. Rainfall prediction requires mathematical modelling and simulation because of its extremely irregular and complex nature. Autoregressive integrated moving average (ARIMA) model was used to analyse annual rainfall and maximum temperature over Tordzie watershed and the forecast. Autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to identify the models by aid of visual inspection. Stationarity tests were conducted using the augmented Dickey–Fuller (ADF), Mann–Kendall (MK) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests respectively. The chosen models were evaluated and validated using the Akaike information criterion corrected (AICC) and also Schwartz Bayesian criteria (SBC). The diagnostic analysis of the models comprised of the independence, normality, homoscedascity, *P–P* and *Q–Q* plots of the residuals respectively. The best ARIMA model for rainfall for Kpetoe and Tordzinu were (3, 0, 3) and (3, 1, 3) with *AICC* values of 190.07 and 178.23. That of maximum temperature for Kpetoe and Tordzinu were (3, 1, 3) and (3, 1, 3) and the corresponding *AICC* values of 23.81 and 36.10. The models efficiency was checked using sum of square error (*SSE*), mean square error (*MSE*), mean absolute percent error (*MAPE*) and root mean square error (*RMSE*) respectively. The results of the various analysis indicated that the models were adequate and can aid future water planning projections.

Key words: *ARIMA, forecasting, rainfall model, temperature, Tordzie watershed*

INTRODUCTION

The most important part of the hydrological cycle is rainfall [RAMANA *et al.* 2013]. It is the result of many complex physical processes that induce particular features and make its observation complex [AKROUR *et al.* 2015]. In the prediction of meteorological information the investigation and analysis of precipitation is so essential [RADHAKRISHNAN, DINESH 2006], and accurate forecast of precipitation is crucial for improved management of water resources, particularly in arid environment [FENG *et al.* 2015].

According to SOMVANSHI *et al.* [2006], rainfall is natural climatic occurrences and its prediction remains a difficult challenge as a result of climatic variability. The forecast of precipitation is particularly relevant to agriculture, growth of plants and development, which profoundly contribute to the economy of Africa. In the statement of the above authors, attempts have been made to predict behavioural pattern of rainfall using autoregressive integrated moving average (ARIMA) technique. ARIMA model is fundamentally a linear statistical technique for modelling the time series and rainfall forecasting with ease to develop future predictions. Though rainfall estimation is an important component of water resources planning, its accurate assessment at locations where rainfall stations are scarce can be very difficult. This makes estimate of rainfall a valid concern using the right method. Thus in the empirical hydro-meteorological modelling of time series data, the emphasis is on modelling and predicting the mean characteristic of the time series using the conventional methods of an autoregressive moving average (ARMA) techniques propounded by BOX *et al.* [2015].

In agricultural planning the understanding of rainfall variability and its prediction has great significance in the agricultural management and helps in decision-making process. Rainfall information is an important input in the hydrological modelling, predicting extreme precipitation events such as droughts and floods, for planning and management of irrigation projects and agricultural production is very important [NIRMALA 2015].

The surface air temperature (SAT) represents an important element of a regional climate. Therefore maximum and minimum values of SAT are usually used as an input in various environmental applications, including agriculture, forestry, fisheries and ecological models to predict likely changes at field and landscape level attributes [KUMARI *et al.* 2012]. During the twentieth century and currently one of the topical issues discussed extensively among researchers and scientists in the field of climate change is changes in SAT. The likely impact of rise in temperature on human beings which is a global phenomenon is being pondered over by a host of scientists.

KAPOOR and BEDI [2013] reported that considering all the climatic variables, forecasting of temperature variability is very essential for diverse applications. Applications of temperature prediction are for climate monitoring, drought detection, agriculture and production, planning in energy industry and many others. In a related study by KENITZER *et al.* [2007], they argued that temperature forecasting is the most essential services delivered by the meteorologist to safeguard life and property of dwellers in a locality and also to improve the efficiency of operations and besides to aid individuals to plan a wide range of activities daily.

In Ghana, several studies have been conducted in recent times to analyse and forecast rainfall and temperatures change using the techniques of ARIMA modelling to assess the changes of rainfall and temperature regime both at the national, regional and watershed levels [ABDUL-AZIZ *et al.* 2013; AFRIFA-YAMOAH 2015; ASAMOAH-BOAHENG 2014]. A search of the literature revealed the gap of no rainfall and temperature analysis and forecast using the ARIMA model exist on the Tordzie watershed. The Tordzie basin is an important basin in the Volta region of Ghana. It has enormous economic benefit to its catchment dwellers. The water from the basin serves as a source of drinking water, washing, watering of animal, irrigation etc. The forecast model will be of tremendous assistance to National Disaster Management Organisation (NADMO). A further analysis of the forecast rainfall and temperature will alert them to prepare for disasters like drought and flood. These hydro-meteorological disasters results in loss of live and properties and millions of dollars are spent in assisting the victims of such natural calamities. A short coming in the agencies responsible for disaster management in Ghana is the weak capacity to forecast this events ahead of time to inform the general populace to prepare or find place of safety. Thus this study is to aid and complement the effort of the decision makers and those responsible for the safeguarding of life and property to plan for the future [AFRIFA-YAMOAH *et al.* 2016]. The result will further engender community resilience in terms of preparedness, mitigation and adaptation strategies. The overall objective of this study is to determine which ARIMA model is the best suited to predict the future rainfall and temperature in the Tordzie watershed.

MATERIALS AND METHODS

STUDY AREA AND DATA USED

The study was carried out in selected network stations in the Tordzie watershed in Ghana. The climatic data from the network of meteorological stations covering two different physiographic areas within the Tordzie watershed was collected for analysis. The network stations in the Tordzie watershed which data were analysed were Kpetoe and Tordzinu with coordinates of (6°54’0” N, 0°69’0”' E) and (5°5’0” N, 0°45’0” E) at elevation 79.0 and 5.4 m respectively. The selection of the physiographic sample area for the study was based on the data availability and the study requirements. Historical rainfall and temperature data were collected from the Ghana Meteorological Services Department. The data for the mentioned two network stations were of good quality as their consistency was checked. The data was continuous, there was no gap in it from the preliminary checks that was conducted to ascertain the data quality and consistency as require for any scientific analysis. The Tordzie watershed is shown in Figure 1.

Tordzie is trans-boundary basin, the area in Ghana is 1865 km^2 which constitute 83.7% and the remaining area in Togo is 363 km^2 which is 16.3% making the total area of the basin 2228 km^2 [WRC 2010]. However, the emphasis of the current study is on the area within Ghana (b).

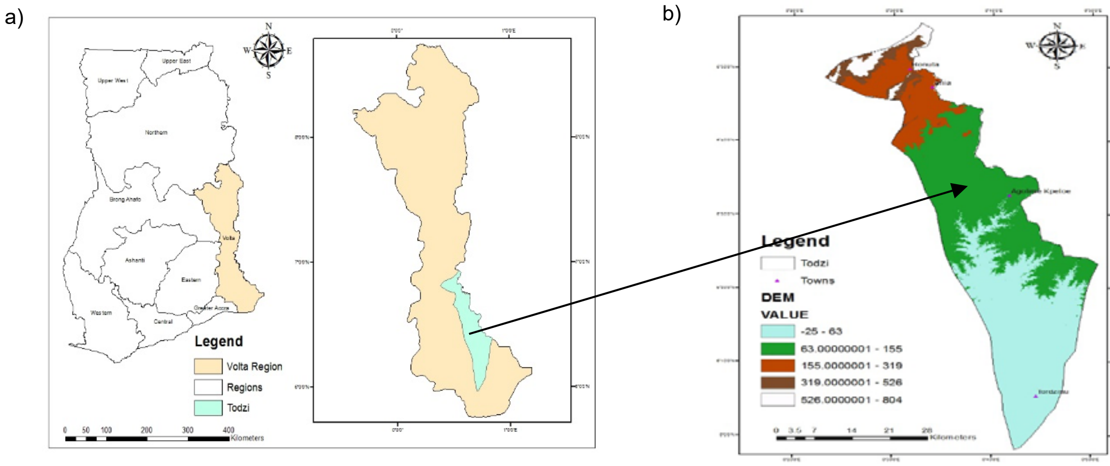

Fig. 1. Study area: a) map of Ghana in relation to Volta region, Tordzie watershed, b) digital elevation model of Tordzie watershed; source: own elaboration

METHODS

Study procedure

The data was organised and tabulated in Excel then XLSTAT, 2015 was used to model the mean annual rainfall and maximum temperature for the period 1984 to 2014 for Kpetoe and Tordzinu portion of the Tordzie watershed. The procedure for the autoregressive integrated moving average (ARIMA) model is described below.

The Box–Jenkins method of ARIMA modelling was used in this study. The procedure adopted by Box and Jenkins in the ARIMA modelling was an iterative process the best models were determined through trial and error. But, the iterative process has been made easy by the aid of statistical software packages. The said model has three parts and they are: autoregressive (AR), integrated (I) and moving-average (MA). The AR part denotes the autocorrelation between current and past observations while the MA part describes the autocorrelation structure of error (residuals). In the report of HASMIDA [2009], the integrated part denotes the level of differencing needed to transform a non-stationary series into a stationary series. The non-seasonal ARIMA model is normally denoted (p, d, q) where the AR part of the model is represented by p, while d denotes the level of differencing required and q is the MA part. For non-seasonal components of a seasonal ARIMA model, the MA operator is written as:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \quad (1)$$

Where: q = the order of non-seasonal MA operator; θ_j, j = 1, 2, …, q = the MA parameters, B = the backward shift operator such that

$$BZ_t = Z_{t-1} \quad (2)$$

The AR operator is written as

$$\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots \emptyset_p B^p \quad (3)$$

Where: p = the order of non-seasonal AR operator, $\emptyset_i$, I = 1, 2… p = the non-seasonal AR parameters.

The non-seasonal ARIMA model for a set of equidistant measurements $Z = [Z_1, Z_2 \ldots Z_n]$ can be written as

$$\emptyset(B)(1-B)^d(Z_t) = \emptyset(B)a_t \quad (4)$$

Where: d = the number of differences; t = discrete time; a_t = the white noise.

The general ARIMA (p, d, q) model is:

$$U_t = \emptyset_1 U_{t-1} + \emptyset_2 U_{t-2} + \cdots + \emptyset_p U_{t-p} + \\ + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots \theta_q \varepsilon_{t-q} \quad (5)$$

$$U_t = X_t - X_{t-d} \quad (6)$$

Where: $\emptyset_p$ = autoregressive parameter; ε_t = residual or white noise; θ_q = moving average parameter; X_t = dependable variable; U_t = d^{th} difference of the dependable variable.

The first step in building the model was to establish whether there is any stationarity in the observed data. Visual inspection of the ACF (autocorrelation function) and PACF (partial autocorrelation function) indicated whether the series was stationary or not.

Stationarity tests

According to BOX *et al.* [2015] if a series is non stationary, then it necessitates a differencing to be carried out to transform it to a stationary series in order to continue with the ARIMA modelling. The tests carried out to identify that consist of:

The unit root's presence in the time series data was carried out. The standard tests for unit root is the augmented Dickey–Fuller (ADF) test. This test is based on estimates from an augmented autoregression. The selection of lag length k is one of the key issues in the ADF test [HUANG *et al.* 2016]. The presence of a unit root is an indication of non-stationarity of the series. The test was carried at 5% significant level. If the series is non-stationary differencing is required to convert it to a stationary data.

Another test for stationarity is the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. It tests for the null hypothesis of stationarity as opposed to the ADF test which tests for the null hypothesis of non-stationarity. It also tests for a unit root of the series [HUANG *et al.* 2016].

The Mann–Kendall (MK) test is principally conducted to test for a trend in the time series data. The presence of a trend in the series indicates a non-stationary series and must be transformed into a stationary series for an ARIMA process to continue. On the other hand if the series is stationary then the series is modelled as an ARMA process which do not require a differencing to be conducted [HUANG *et al.* 2016].

Differencing. It is required in fitting ARIMA model to attend stationarity in both the mean and variance. To attend stationarity in the data, it could be done by log transformation and differencing of the original data [HUANG *et al.* 2016]. In this study, stationary first difference (d = 1) of the original data was carried out to achieve stationarity. The ACF and PACF of the differenced data were observed and tested for stationarity

Independence. The fundamental assumption of the residuals of an ARIMA model are that they white noise. A white noise is a serially uncorrelated variable. If a series has a white noise it indicates uncorrelated random shock with a mean of zero and a variance which is constant [HUANG *et al.* 2016]. An independent residual indicates that no information can be extracted from the data. The independence of the series is arrived at by the visual inspection of the correlogram of the residuals. A correlogram with values close to zero is an indication that the residuals are uncorrelated and independent.

Homoscedasticity. If the variance of a disturbance term of each observation in a series is constant it is described as homoscedastic. A homoscedastic residuals have their variances stable. The probability of the disturbance terms attaining a given positive or negative value is same in all observations, meaning they have the same dispersion [HUANG *et al.* 2016].

Forecasting. Statistical methods and for that reason ARIMA models are good for short-term forecasting because the historical data normally exhibit inertia and do not show drastic changes [MONTGOMERY *et al.* 2008]. Short-term forecasting is based on identifying, modelling and extrapolating the patterns found in the data.

Analysis of ACF and PACF

One of the techniques used to identify the models was the visual inspection of the series, which included the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The mean annual rainfall and temperature data were used as the input variable. The auto covariance function, the autocorrelation coefficients and the partial correlation coefficients were computed from the said input variables and the series with its ACF and PACF were plotted using XLSTAT software.

Model diagnostic

To evaluate the model in order to select the best model for each category of data, the criteria used by KHADR [2011], for selecting the best model was used. The criteria used were:

1. Akaike information criteria (AIC)
2. Schwarz-Bayesian information criteria (SBC)

The Akaike information criterion corrected (*AICC*) was established by AKAIKE [1974] to choose the best model among the class of plausible models. The models with the lowest value of *AICC* and SBC were selected as the most suitable model and used for the forecasting. The equations governing the above mentioned criteria [SCHWARZ 1978] are:

$$AICC(p,q,P,Q) = N\ln(\sigma^2) + \frac{2(p+q+P+Q+1)N}{(N-p-q-P-Q-2)} \quad (7)$$

$$SBC = -2\log(L) + (p+q+P+Q-2)\ln(N) \quad (8)$$

Where: N = the number of observations; L = the likelihood function of the ARIMA models; δ = the mean square error.

When the number of samples is low, SBC is usually a better criterion than AIC. However, the study of MISHRA and DESAI [2005] and ALAM *et al.* [2014] indicated that *AICC* which is the revised version of AIC acts well even by low number of samples.

Model estimation

It is required to determine the model that best fits the data being analysed. This was achieved by observing the ACF and the PACF of the differenced data. The models were considered with both p and q starting from one to three. The models tested were: (1, 1, 1), (1, 1, 2), (1, 1, 3), (2, 1, 1), (2, 1, 2), (2, 1, 3), (3, 1, 1), (3, 1, 2) and (3, 1, 3). The model having the lowest *AICC* value was chosen as the best for each station.

Model diagnostic analysis

In order to be sure that the models were representative for the observed rainfall and temperature data and can be used to forecast the future data, the models were subjected to diagnostic tests. The best chosen models with their corresponding model parameter values are presented. Diagnostic checks were carried out to determine whether the models fit the data very well. If the model fits well the residuals should be uncorrelated (white noise) with a constant variance and also the residuals should be normally distributed [HUANG *et al.* 2016].

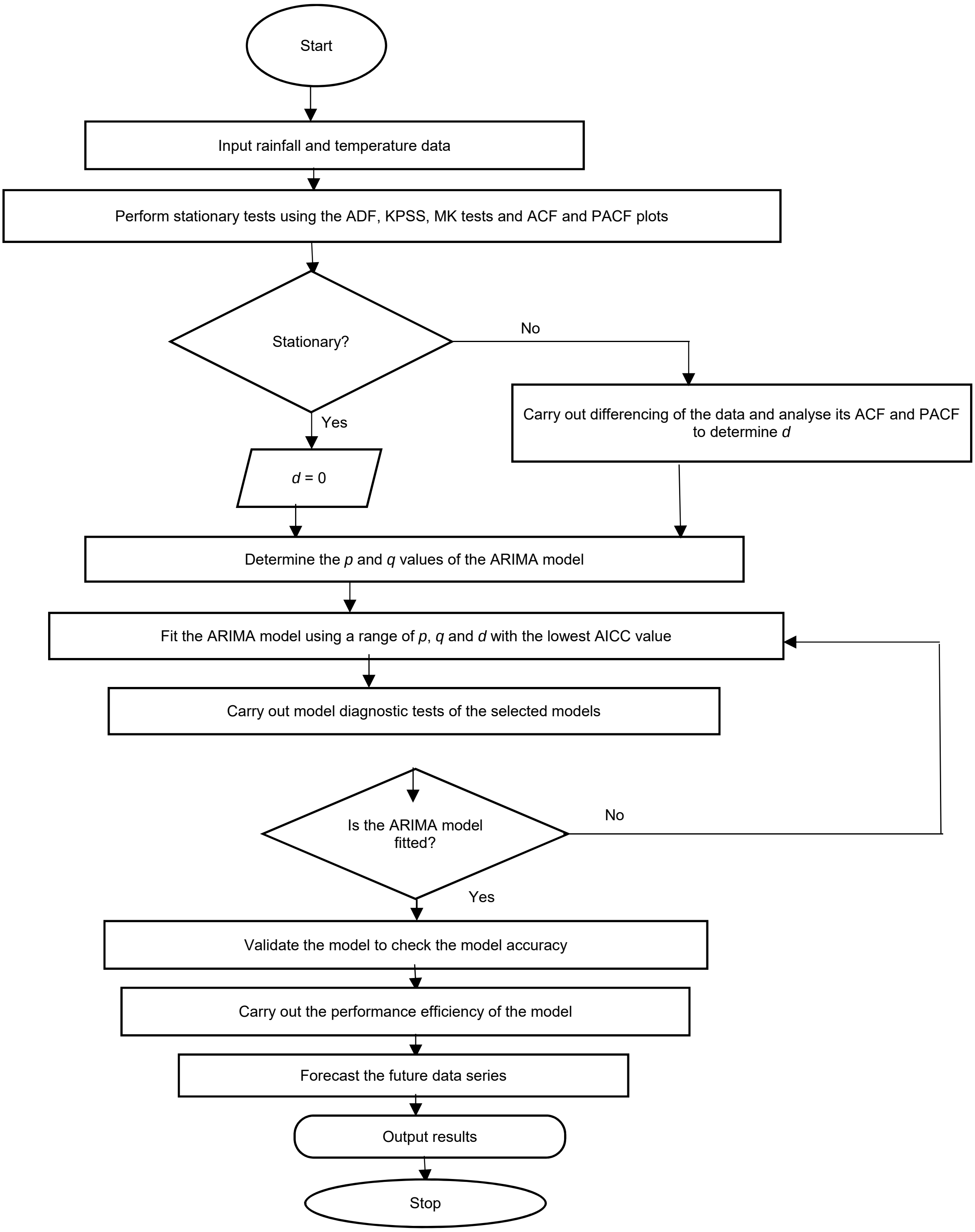

Fig. 2. Flow chart of ARIMA modelling; source: own elaboration

Plot of residuals of ACF and PACF. After establishing the appropriate model with the above mentioned criteria, goodness of fit of the model was determined by plotting ACF and PACF of the residuals. A white noise residuals was an indication of the model fit being adequate. A residual with a zero mean and uncorrelated is termed white noise.

Calibration and validation. Calibration was done to assess the models for quality and accuracy of prediction. Calibration procedures were carried out to strengthen the models performance. Part of the observed field data, that is data from the meteorological station (from 1984 to 2004) was used for calibration of the model, while the remaining data were used to validate the model (2004–2014). Validation test was conducted after finalising calibration data which is important for testing the simulation capacity of the model. The model validation is in fact the extension of the calibration process. Thus validation was carried out without any further adjustments to the calibrated model parameters. The model was validated for the period 2004 to 2014. The selected suitable models were validated using the available data.

Model validation and forecasting. The validated time series data was then used to forecast the future time series data of rainfall and temperature for ten years. The significance of the validation procedure was to determine the reliability of the model for forecasting of rainfall and temperature time series data. The procedure entails the comparison of the actual/observed data series with the forecasted data series to see how the model simulate the actual observation.

Performance evaluation criteria. The models were evaluated and validated using the following performance criteria: root mean square error (*RMSE*), sum of squares error (*SSE*), mean square error (*MSE*) and mean absolute percent error (*MAPE*). The equations of the performance criteria are stated below:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_p - Y_o)^2 \tag{9}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_o - Y_p}{Y_o}\right| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_o - Y_i)^2} \tag{11}$$

$$SSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_o - Y_p\right)^2 \tag{12}$$

Where: Y_p = the predicted values; Y_o = the observed values; n = the number of observation.

At the forecasting stage, the estimated parameters were tested for their validity using the above error statistics. The procedure is summarised in Fig. 2.

RESULTS AND DISCUSSION

The time series plot of Figures 3 and 4 of mean annual rainfall and maximum temperature on two network stations (Kpetoe and Tordzinu) encounters problem of non-stationarity. In order to use the ARIMA procedure for the modelling, the non-stationarity was removed by differencing to make the data set stationary. The time series plot portray an increasing but waving maximum temperature whiles the converse is the case for mean annual rainfall. The difference in the magnitude of time series plot of the said network stations is due to the differences in the ecological zones.

The stationarity tests were conducted on the mean annual rainfall and the maximum temperature data series to check and confirm the autocorrelation function (ACF) and partial autocorrelation function (PACF) analysis on the non-stationarity of the data series. The results of augmented Dickey–Fuller (ADF), Mann–Kendall (MK) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests respectively are presented in Table 1. The tests buttressed that all the mean annual rainfall and maximum temperature data series were not stationary for Kpetoe and Tordzinu respectively. The ADF and KPSS tests indicated the data series had a unit root. Additionally, the MK test also revealed a trend in the data series which further buttressed the non-stationarity. From the pattern of the ACF and PACF plots as shown in Figures 5 and 6, the mean annual rainfall and the mean maximum temperature series respectively were non-stationary. The patterns exhibited a very slow decay which was an indication of possible non-stationarity. Besides some of the spikes were outside the confidence interval and some very close to it. This indicated the values were significant and were not white noise [HUANG *et al.* 2016]. Thus differencing was carried

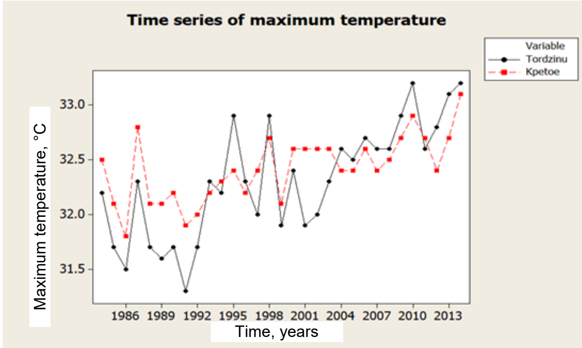

Fig. 3. Time series plot of maximum temperature (1984–2014); source: own study

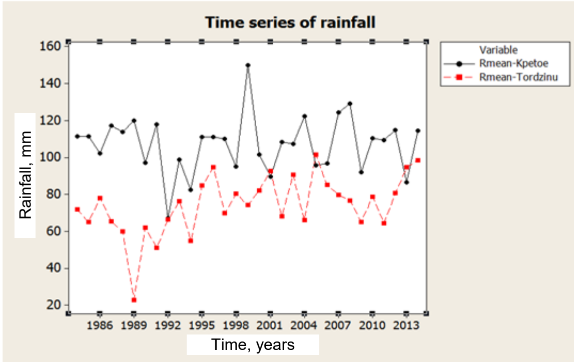

Fig. 4. Time series plot of mean annual rainfall (1984–2014); source: own study

Table 1. Stationary tests

Station	Augmented Dickey–Fuller test		Kwiatkowski–Phillips–Schmidt–Shin test		Mann–Kendall test		Remarks
	p-value (rainfall)	*p*-value (max temp)	*P*-value (rainfall)	*P*-value (max temp)	*P*-value (rainfall)	*P*-value (max temp)	
Kpetoe	0.3727	0.2628	0.8293	0.0005	0.9156	0.0018	non-stationary
Tordzinu	0.1854	0.0264	0.0068	0.0006	0.0060	0.0003	non-stationary

Source: own study.

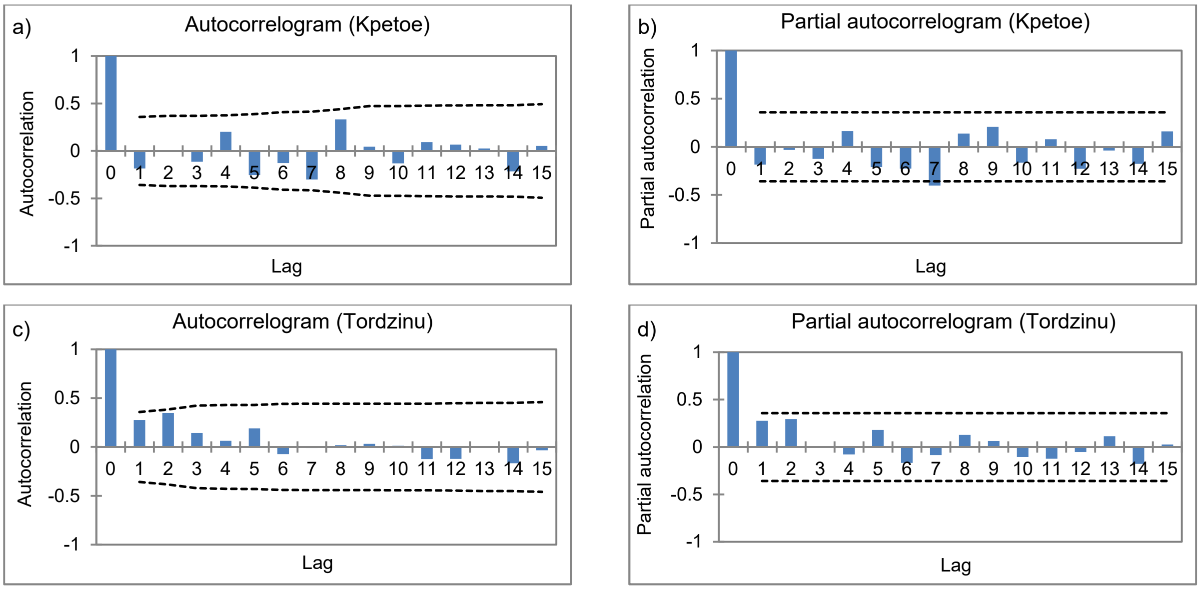

Fig. 5. Autocorrelation function (a, c) and partial autocorrelation function (b, d) plot of rainfall series; source: own study

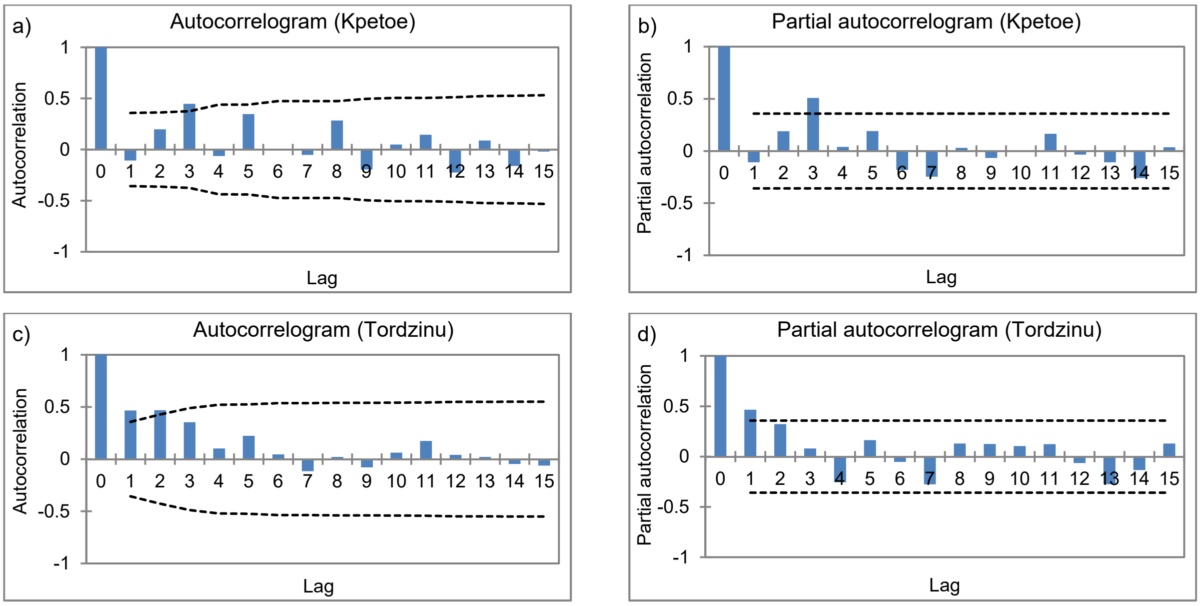

Fig. 6. Autocorrelation function (a, c) and partial autocorrelation function (b, d) plot of maximum temperature series; source: own study

out to convert the data to a stationary one. The implication of the above results is that ignoring non-stationarity of the data series will have led to the selection of sub-optimal model whose estimate may be misleading. This means the accuracy of the model depended on the stationarity information. The assertion is in agreement with the report of JUNG and SHAH [2015] who study the implication of non-stationarity on predictive models and HENDRY and PRETIS [2016] who considered the implication of non-stationarity on empirical modelling and forecasting.

The first diagnostic check was the plot of histogram of the residuals (Figs. 7, 8) which confirmed the normality of the residuals for rainfall and maximum temperature respectively. The other requirement of a good model is that the residuals should be homoscedastic. The homoscedasticity plots were indicated in Figures 9 and 10 respectively of rainfall and maximum temperature models. The results of Breusch–Pagan test as shown in Table 2 confirmed the homoscedascity plots. A residual with constant variance is said to be homoscedastic. Homoscedasticity is a determinant of the model's ability to predict variables consistently. A heteroscedastic residuals cannot provide predictions that are reliable [HUANG *et al.* 2016].

Table 2. Results of Breusch–Pagan test

Stations	*p*-value (rainfall)	*p*-value (temperature)	Interpretation
Kpetoe	0.9171	0.8565	homoscedastic
Tordzinu	0.3766	0.1998	homoscedastic

Source: own study.

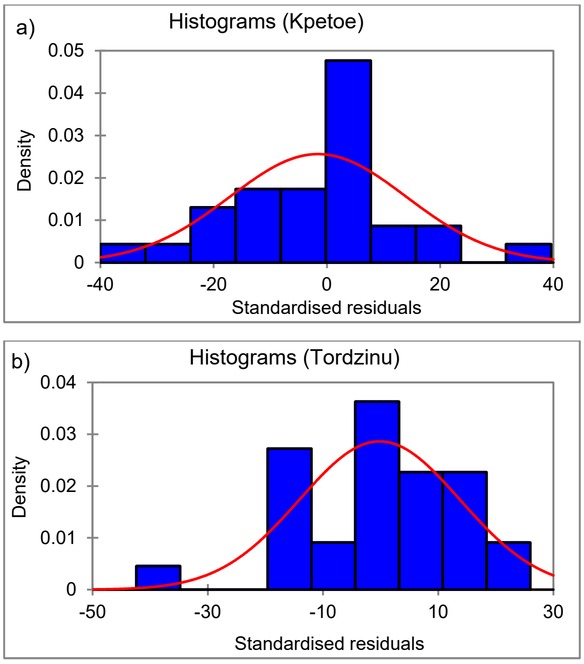

Fig. 7. Histogram of residuals (rainfall) for:
a) Kpetoe, b) Tordzinu; source: own study

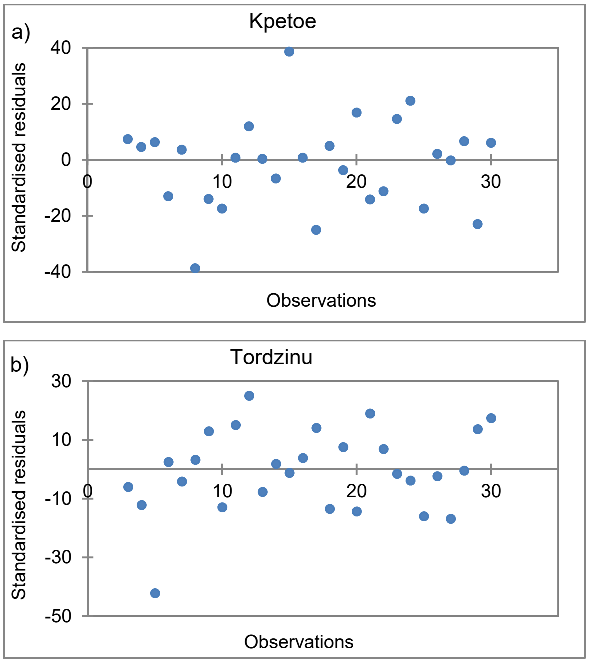

Fig. 9. Distribution of standard residuals (rainfall) for:
a) Kpetoe, b) Tordzinu source: own study

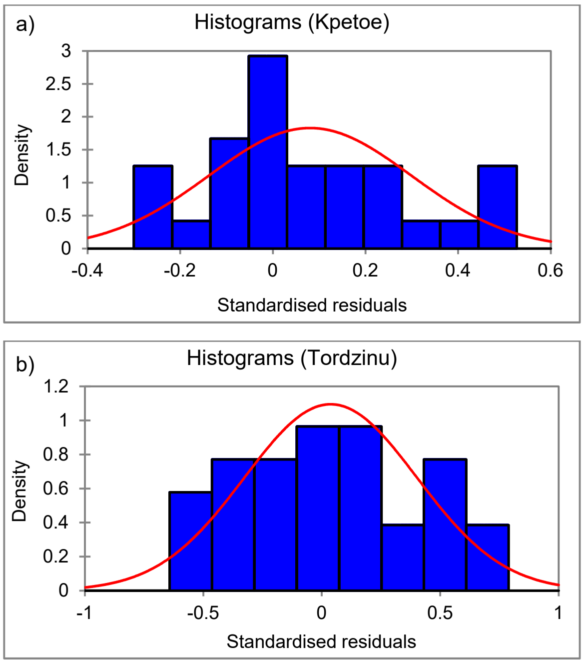

Fig. 8. Histogram of residuals (temperature) for:
a) Kpetoe, b) Tordzinu; source: own study

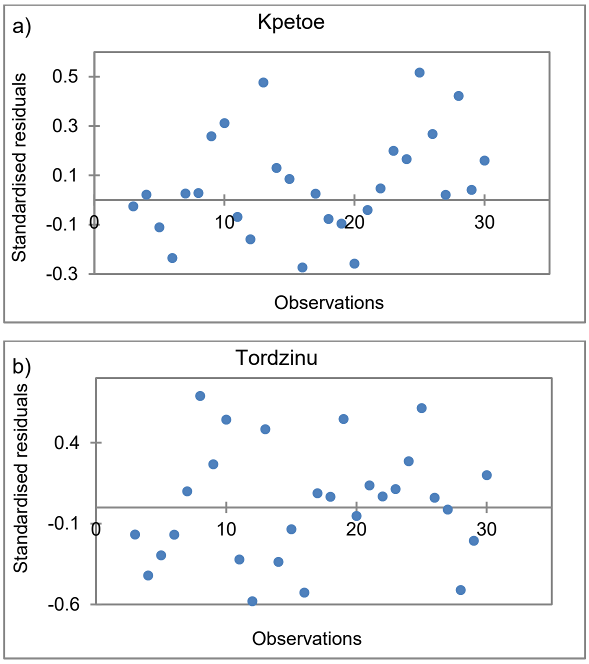

Fig. 10. Distribution of standard residuals (temperature) for:
a) Kpetoe, b) Tordzinu; source: own study

The normality of residuals distribution was essential to produce a satisfactory confidence interval for the forecast. The results of Shapiro–Wilk, Anderson–Darling, Jarque–Bera and Lilliefors tests respectively further confirmed the normality of the residuals (Tab. 3). The *p*-values of the aforementioned tests were more than 0.05 at 95% confidence interval. The *P–P* and *Q–Q* plots of the residuals also attested the normality (Figs. 11–14). The *Q–Q* plot of standardised residuals was based on gamma distribution assumption for a data set. The *Q–Q* plot compared the observed data with the fore-casted data by plotting their quantile against each other. The data set almost lying in straight line is an indication that the two distributions were similar. The implication is that the fitted models are correct as their standardised residuals were from the same gamma distribution. The few data points far from the straight line may be due to deviation from the mean stemming from variability of the date set (heavy or low rainfall). The attribution made is corroborated by GEORGE *et al.* [2016]. The computed Hessian standard errors and all the estimated model

Table 3. Results of normality tests

Stations	Shapiro–Wilk test		Anderson–Darling test		Jarque–Bera test		Lilliefors test	
	p-value (rainfall)	*p*-value (temperature)	*p*-value (rainfall)	*p*-value (temperature)	*p*-value (rainfall)	*p*-value (temperature)	*p*-value (rainfall)	*p*-value (temperature)
Kpetoe	0.8089	0.2260	0.5186	0.2512	0.8023	0.5378	0.3625	0.1076
Tordzinu	0.2728	0.3623	0.6158	0.5234	0.1473	0.5934	0.9017	0.7085

Source: own study.

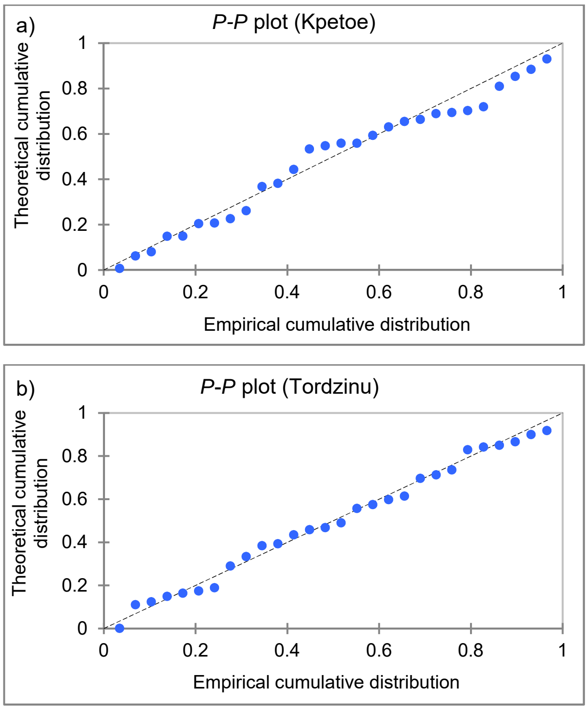

Fig. 11. The *P–P* plot of residuals (rainfall) for: a) Kpetoe, b) Tordzinu; source: own study

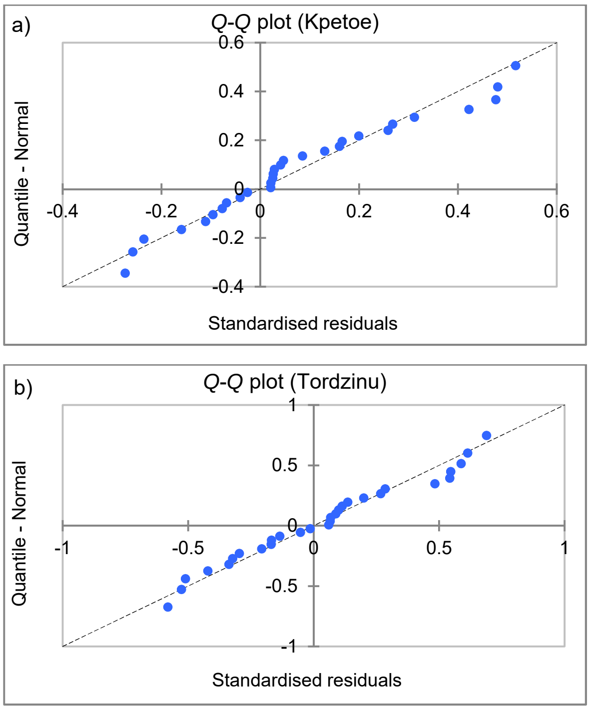

Fig. 13. The *Q–Q* plot of standardised residuals (max temperature) for: a) Kpetoe, b) Tordzinu; source: own study

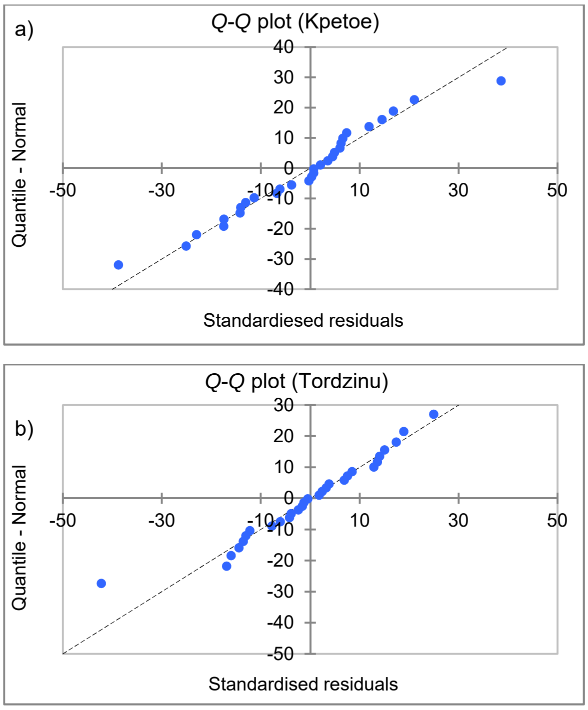

Fig. 12. The *Q–Q* plot of the residuals (rainfall) for: a) Kpetoe, b) Tordzinu; source: own study

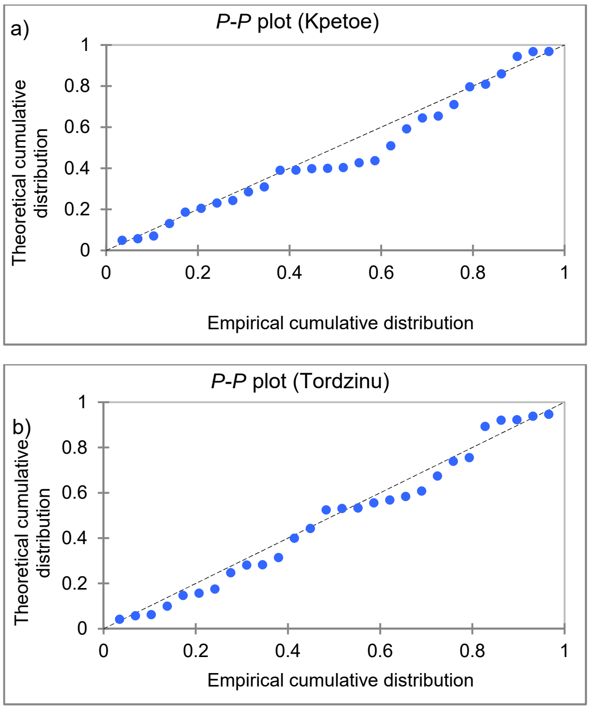

Fig. 14. The *P–P* plot of standardised residuals (max temperature) for: a) Kpetoe, b) Tordzinu; source: own study

parameters were within the confidence band. The implication of the non-normality of the data series is that the inference or prediction made with the model may be unreliable and misleading. The evidence of normality is corroborated by the histogram of the residuals which is almost bell shape (Fig. 7–8).

OBSERVED AND FORECAST SERIES

The models accuracy was checked by comparing the observed series with the forecast series generated by the ARIMA model. A forecast series of a lead time of ten years were generated with a confidence interval of 95%. The forecast series for both rainfall and temperature data for the observed and forecast series for the physiographic stations on Tordzie watershed selected for this study were shown in Figures 15 and 16 respectively. The generated forecast data series followed the observed data set, with variability as is generally reported by other studies. The variability in the rainfall pattern as generated by the forecast model can be attributed to the global warming and the land use and land cover changes. The forecasted rainfall is on slow declining trend while the temperature is on a rise. This assertion agrees with [LOGAH *et al.* 2013; NKRUMAH *et al.* 2014; OWUSU, WAYLEN 2009; 2013]. The implication of the declining rainfall is an increasing drought leading to food insecurity [ŁABĘDZKI, BĄK 2017; NYATUAME, AGODZO 2017], also the variability in the forecast rainfall value could also be responsible for hydrological drought as has been reported by a similar study by [BĄK, KUBIAK-WÓJCICKA 2017].

After fitting the models the residual plots of ACF and PACF were examined and it was observed that the residuals were within the confidence intervals which is an indication of a good fit and the adequacy of the model. That is correlogram analysis of a plot of ACF and PACF. For a good forecasting model, the residuals left after fitting the model, must satisfy the requirements of a white noise process [HUANG *et al.* 2016]. From Figures 17 and 18 it was clear that the correlogram of the ACF and PACF of the residuals of the stations for both rainfall and temperature data fell within the confidence interval. This was an indication that they were not significant and that the residuals were independent and thus satisfying the residual criterion. Besides no patterns were observed in the residuals which further buttressed the point that the models could be used to represent the observed data. It means the residuals were very small in magnitude and have no pattern or trend. The residual is the difference between the observed and the forecast data. The implication is that the forecast value is as close as the observe data further indicating the performance efficiency of the model. Residuals are therefore employ to validate models. The study of NOBRE and SINGER [2007] is consistent with the above assertion. Also, the observed ACF and PACF plots indicated that one order differencing is adequate. Further differencing of higher orders revealed higher standard deviations, an indication of over differencing. Thus the minimum standard deviations were achieved with differencing order one (d = 1). Therefore the preliminary ARIMA (p, d, q) was selected.

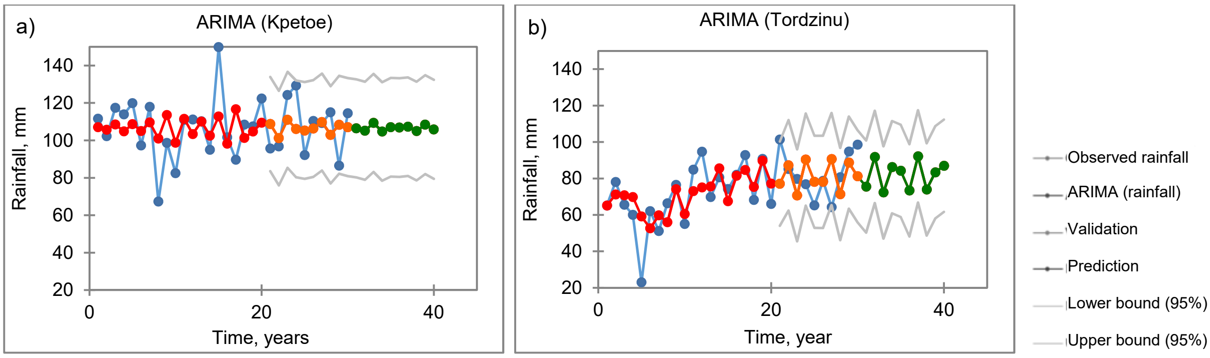

Fig. 15. Observed, synthetic and forecast series of mean rainfall for: a) Kpetoe, b) Tordzinu; source: own study

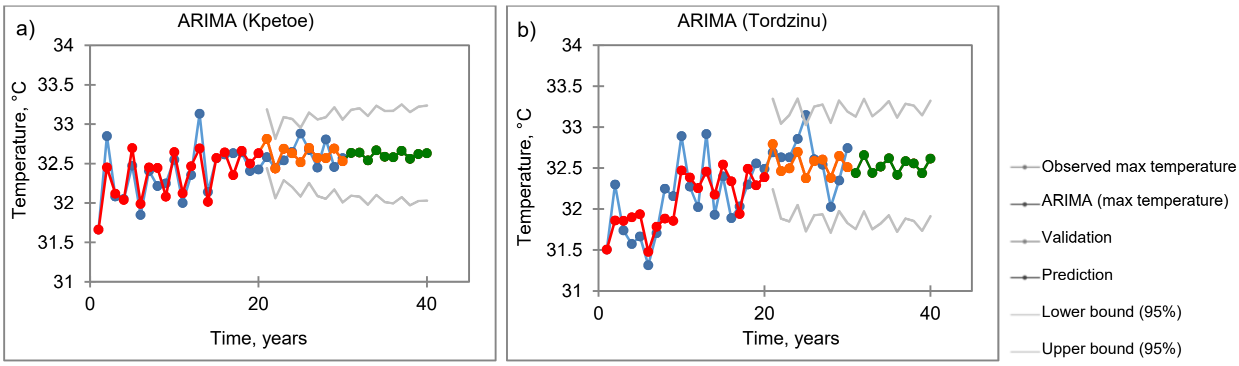

Fig. 16. Observed, synthetic and forecast series of maximum temperature for: a) Kpetoe, b) Tordzinu; source: own study

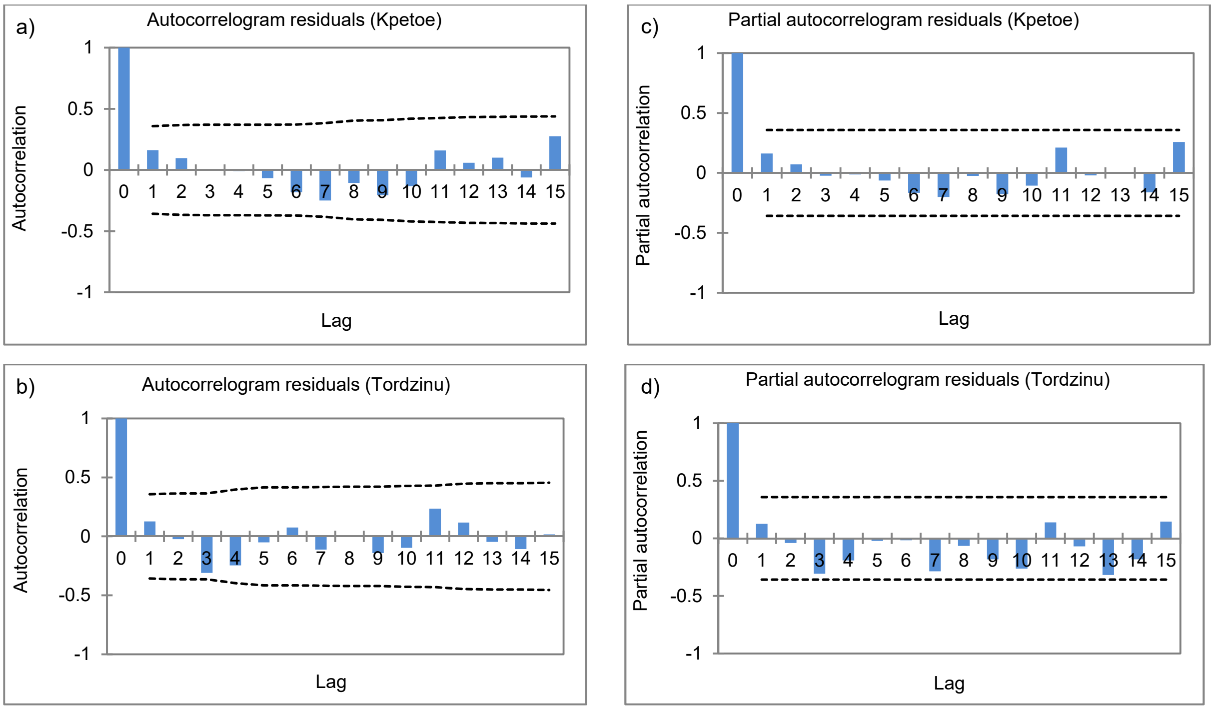

Fig. 17. Residual plots of autocorrelation function (a, b) and partial autocorrelation function (c, d) of maximum temperature; source: own study

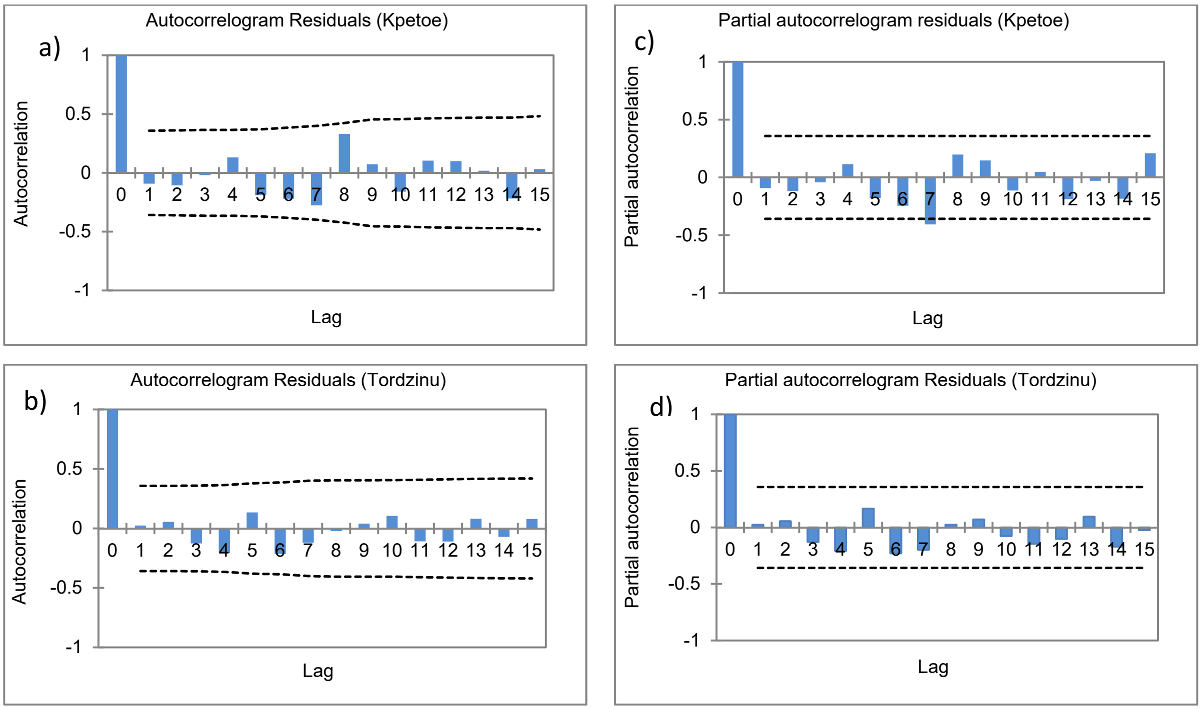

Fig. 18. Residual plots of autocorrelation function (a, b) and partial autocorrelation function (c, d) of annual rainfall; source: own study

PERFORMANCE EFFICIENCY OF THE MODELS

The models performance efficiency were evaluated using sum of squares error (*SSE*), mean absolute percentage error (*MAPE*), mean square error (*MSE*) and root mean square error (*RMSE*) for both rainfall and temperature models respectively. The values for the respective models for each of the stations from each physiographic area of the studied watershed are presented in Table 4. In the table mentioned, the (*MAPE*) which is an unbiased statistic was employed to evaluate the ability of the model to predict correctly. Its low value is an evidence of the models adequacy. It is reported in literature that the smaller the value the better the model's performance [GALAVI *et al.* 2013; VALIZADEH *et al.* 2014].

Table 4. Best autoregressive integrated moving average models and goodness of fit statistics

Best model	Goodness of fit statistics			
	Kpetoe		Tordzinu	
	(3, 0, 3) (rainfall)	(3, 1, 3) (temperature)	(3, 1, 3) (rainfall)	(3, 1, 3) (temperature)
AICC	190.07	23.81	178.23	36.10
SBC	187.71	20.24	174.66	32.53
MAPE	11.27	0.43	17.34	0.74
SSE	4525.44	0.69	2651.55	1.50
MSE	226.27	0.04	139.56	0.08
RMSE	15.04	0.19	11.81	0.28

Source: own study.

The forecast model for Kpetoe rainfall is:

$$Y_t = -1.69Y_{t-1} - 1.55Y_{t-2} - 0.81Y_{t-3} - 1.64\varepsilon_{t-1} + \\ - 1.67\varepsilon_{t-2} - 0.89\varepsilon_{t-3} + 3.46 \qquad (13)$$

Where: Y_t = the forecast of rainfall for Kpetoe for time t years; Y_{t-1} = the forecast of rainfall for Kpetoe for previous year; Y_{t-2} = the forecast of rainfall for Kpetoe for previous two years; Y_{t-3} = the forecast of rainfall for Kpetoe for previous three years; ε_{t-1} = the previous one year residuals of rainfall for Kpetoe; ε_{t-2} is the previous two years residuals of rainfall for Kpetoe; ε_{t-3} is the previous three years residuals of rainfall for Kpetoe.

The model for Tordzinu rainfall is:

$$Y_t = -1.14Y_{t-1} - 0.36Y_{t-2} + 0.41Y_{t-3} - 0.53\varepsilon_{t-1} + \\ + 0.53\varepsilon_{t-2} + 0.99\varepsilon_{t-3} + 1.31 \qquad (14)$$

The forecast model Kpetoe mean maximum temperature is:

$$Y_t = -1.92Y_{t-1} - 1.68Y_{t-2} - 0.53Y_{t-3} - 0.75\varepsilon_{t-1} + \\ + 0.75\varepsilon_{t-2} + \varepsilon_{t-3} + 0.03 \qquad (15)$$

The forecast model for Tordzinu temperature is:

$$Y_t = -0.75Y_{t-1} - 0.23Y_{t-2} + 0.50Y_{t-3} - 0.07\varepsilon_{t-1} + \\ + 0.07\varepsilon_{t-2} + \varepsilon_{t-3} + 0.04 \qquad (16)$$

The Table 5 provides the model input parameters that was input into Equation (5) to obtain Equations (13–16). The models were used for the forecasting. The difference between the observed values and the forecasted values follows the normal distribution and other performance efficiency values which is an indication of the reliability of the models.

APPLICATION OF THE FORECAST MODELS

The developed models could be used for water resources planning and management. Sound water management planning and cropping system design can be achieved with an understanding of the statistical properties of long-term records of major climatic parameters like rainfall and temperature.

The probable climate change impacts on rainfall and temperature assessment and its forecast is crucial for disaster alertness, planning of irrigation infrastructural and development. The studied watershed has not received adequate research attention in climate change impact assessment and for that reason this study and the model developed will serve as a foundation for decision making on the watershed.

Table 5. Autoregressive integrated moving average (ARIMA) model parameters

Model	Parameter	Estimate (coefficient)	Hessian standard error
ARIMA (3, 0, 3) Kpetoe rainfall	Φ_1	–1.69	0.88
	Φ_2	–1.55	0.95
	Φ_3	–0.81	0.31
	Θ_1	1.64	0.82
	Θ_2	1.67	0.76
	Θ_3	0.89	0.50
ARIMA (3, 1, 3) Tordzinu rainfall	Φ_1	–1.14	0.73
	Φ_2	–0.36	2.55
	Φ_3	0.41	1.66
	Θ_1	0.53	0.36
	Θ_2	–0.53	0.36
	Θ_3	–0.99	0.30
ARIMA (3, 1, 3) Kpetoe temperature	Φ_1	–1.92	0.29
	Φ_2	–1.68	0.39
	Φ_3	–0.53	0.30
	Θ_1	0.75	0.29
	Θ_2	–0.75	0.32
	Θ_3	–1.00	0.29
ARIMA (3, 1, 3) Tordzinu temperature	Φ_1	–0.75	0.25
	Φ_2	–0.23	0.30
	Φ_3	0.50	0.23
	Θ_1	0.07	0.29
	Θ_2	–0.07	0.29
	Θ_3	–1.00	0.29

Explanations: Φ_1, Φ_2 Φ_3 = autoregressive parameters; Θ_1, Θ_2, Θ_3 = moving average parameters.
Source: own study.

The analysis carried out provides vital information in addressing projected climatic changes and their impacts on the fresh water resources of the watershed.

An exact knowledge of the future water resources of a watershed is a strategic information which is required for long-term planning of a watershed water users and food security issues of its users among other needs. Modelling tools permit this quantification feasible. The implication of the changing climate that is generally reported may be impacting at the local level and rendering the indigenous knowledge of predicting the climate pattern a Herculean task. Thus integrating the scientific procedure with that of the indigenous know-how is ideal in surmounting the challenge. The variability in the forecasted rainfall is in agreement with the earlier findings of NYATUAME *et al.* [2014] who reported that the climate change might have been responsible for the variability in the rainfall pattern in the Volta region of Ghana.

CONCLUSIONS

In this study, the annual rainfall and maximum temperature of Tordzie watershed was modelled using ARIMA models for the two stations on the basin. The results showed a slight decrease and in oscillatory manner in the rainfall for the future up to 2024 (2014–2024) in most part of the watershed. Interestingly, the maximum temperature forecast revealed a gentle upward increase but oscillating in nature. The upward fluctuating trend in the maximum temperature might be the reason for the decreasing rainfall in the said watershed. However, further comprehensive multivariate analysis employing a digitized data is required for a definite pronouncement on the rainfall and temperature situation. The models have been evaluated and validated after tentative identification and diagnostic tests were performed on them and the selected models proved adequate and suitable for the forecast of future annual rainfall and temperature values in the Tordzie watershed which can aid decision makers establish priorities in terms of water demand management.

REFERENCES

ABDUL-AZIZ A., ANOKYE M., KWAME A., MUNYAKAZI L., NSOWAH-NUAMAH N. 2013. Modeling and forecasting rainfall pattern in Ghana as a seasonal ARIMA process: The case of Ashanti region. International Journal of Humanities and Social Science. Vol. 3. No. 3 p. 224–233.

AFRIFA-YAMOAH E. 2015. Application of ARIMA models in forecasting monthly average surface temperature of Brong Ahafo Region of Ghana. International Journal of Statistics and Applications. Vol. 5. No. 5 p. 237–246.

AFRIFA-YAMOAH E., SAEED B.I., KARIM A. 2016. Sarima modelling and forecasting of monthly rainfall in the Brong Ahafo Region of Ghana. World Environment. Vol. 6(1) p. 1–9.

AKAIKE H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control. Vol. 19. Iss. 6 p. 716-723.

AKROUR N., CHAZOTTES A., VERRIER S., MALLET C., BARTHES L. 2015. Simulation of yearly rainfall time series at microscale resolution with actual properties: Intermittency, scale invariance, and rainfall distribution. Water Resources Research. Vol. 51. Iss. 9 p. 7417–7435. DOI 10.1002/2014WR016357.

ALAM N., MISHRA P., JANA C., ADHIKARY P.P. 2014. Stochastic model for drought forecasting for Bundelkhand region in Central India. Indian Journal of Agricultural Sciences. Vol. 84(1) p. 79–84.

ASAMOAH-BOAHENG M. 2014. Using SARIMA to forecast monthly mean surface air temperature in the Ashanti Region of Ghana. International Journal of Statistics and Applications. Vol. 4(6) p. 292–298.

BĄK B., KUBIAK-WÓJCICKA K. 2017. Impact of meteorological drought on hydrological drought in Toruń (central Poland) in the period of 1971–2015. Journal of Water and Land Development. No. 32 p. 3–12. DOI 10.1515/jwld-2017-0001.

BOX G. E., JENKINS G.M., REINSEL G.C., LJUNG G.M. 2015. Time series analysis: forecasting and control. 5th ed. John Wiley & Sons. ISBN 1118675029 pp. 712.

FENG Q., WEN X., LI J. 2015. Wavelet analysis-support vector machine coupled models for monthly rainfall forecasting in arid regions. Water Resources Management. Vol. 29. Iss. 4 p. 1049–1065.

GALAVI H., MIRZAEI M., SHUI L.T., VALIZADEH N. 2013. Klang river-level forecasting using ARIMA and ANFIS models [online]. Journal of America Water Works Association. Vol. 105. Iss. 9 p. E496–E506. DOI 10.5942/jawwa.2013.105.0106.

GEORGE J., LETHA J., JAIRAJ P. 2016. Daily rainfall prediction using generalized linear bivariate model – A case study. Procedia Technology. Vol. 24 p. 31–38.

HASMIDA H. 2009. Water quality trend at the upper part of Johor River in relation to rainfall and runoff pattern. MSc Thesis. Universiti Teknologi Malaysia pp. 155.

HENDRY D.F., PRETIS F. 2016. All change! The implications of non-stationarity for empirical modelling, forecasting and policy. Unviversity of Oxford, Oxford Martin School Policy Paper Series pp. 28.

HUANG Y.F., MIRZAEI M., YAP W.K. 2016. Flood analysis in Langat River Basin using stochatic model. International Journal of GEOMATE. Vol. 11 p. 2796–2803.

JUNG K., SHAH N.H. 2015. Implications of non-stationarity on predictive modeling using EHRs. Journal of Biomedical Informatics. Vol. 58 p. 168-174.

KAPOOR P., BEDI S.S. 2013. Weather forecasting using sliding window algorithm. ISRN Signal Processing. Vol. 2013 (2013). Article ID 156540 pp. 5.

KENITZER S., ROSENFELD J., HEIDEMAN K. 2007. Bulletin of the America Meteorological Society: Annual Report 2006. Boston 88.4 (Apr 2007) p. 1–27.

KHADR M. 2011. Water resources management in the context of drought (An application to the Ruhr River Basin in Germany). PhD Dissertation. Civil Engineering Department.University of Wuppertal, Germany (edited).

KUMARI K.A., BOIROJU N.K., GANESH T., REDDY P.R. 2012. Forecasting surface air temperature using neural networks. International Journal of Mathematics and Computer Applications Research. Vol. 3 p. 65–78.

LOGAH F., OBUOBIE E., OFORI D., KANKAM-YEBOAH K. 2013. Analysis of rainfall variability in Ghana. International Journal of Latest Research in Engineering and Computing. Vol. 1. Iss. 1 p. 1–8.

ŁABĘDZKI L., BĄK B. 2017. Impact of meteorological drought on crop water deficit and crop yield reduction in Polish agriculture. Journal of Water and Land Development. No. 34 p. 181–190. DOI 10.1515/jwld-2017-0052.

MISHRA A.K., DESAI V.R. 2005. Drought forecasting using stochastic models. Stochastic Environmental Research and Risk Assessment. Vol. 19. Iss. 5 p. 326–339.

MONTGOMERY D.C; JENNINGS C.L; KULAHCI M. 2008. Introduction to time series analysis and forecasting. John Wiley and Sons, Inc. ISBN 0471653977 pp. 472.

NIRMALA M. 2015. Computational models for forecasting annual rainfall in Tamilnadu. Applied Mathematical Sciences. Vol. 9. Iss. 13 p. 617–621.

NKRUMAH F., KLUTSE N.A.B., ADUKPO D.C., OWUSU K., QUAGRAINE K.A., OWUSU A., GUTOWSKI W. 2014. Rainfall variability over Ghana: Model versus rain gauge observation. International Journal of Geosciences. Vol. 5. Iss. 7 p. 673–683.

NOBRE S.J., SINGER M.J. 2007. Residual analysis for linear mixed models. Biometrical Journal. Vol. 49. Iss. 6 p. 863–875.

NYATUAME M., AGODZO S. 2017. Analysis of extreme rainfall events (drought and flood) over Tordzie Watershed in the Volta Region of Ghana. Journal of Geoscience and Environment Protection. Vol. 05. No. 09 p. 275–295.

NYATUAME M., OWUSU-GYIMAH V., AMPIAW F. 2014. Statistical analysis of rainfall trend for Volta Region in Ghana. International Journal of Atmospheric Sciences. Vol. 2014. Article ID 203245 pp. 11.

OWUSU K., WAYLEN P.R. 2009. Trends in spatio-temporal variability in annual rainfall in Ghana (1951–2000). Weather. Vol. 64. Iss. 5 p. 115–120.
OWUSU K., WAYLEN P.R. 2013. The changing rainy season climatology of mid-Ghana. Theoretical and Applied Climatology. Vol. 112. Iss. 3–4 p. 419–430.
RADHAKRISHNAN P., DINESH S. 2006. An alternative approach to characterize time series data: Case study on Malaysian rainfall data. Chaos, Solitons and Fractals. Vol. 27. Iss. 2 p. 511–518.
RAMANA R.V., KRISHNA B., KUMAR S., PANDEY N. 2013. Monthly rainfall prediction using wavelet neural network analysis. Water Resources Management. Vol. 27. Iss. 10 p. 3697–3711.
SCHWARZ G. 1978. Estimating the dimension of a model. The Annals of Statistics. Vol. 6(2) p. 461–464.
SOMVANSHI V., PANDEY O., AGRAWAL P., KALANKER N., PRAKASH M.R., CHAND R. 2006. Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. The Journal of Indian Geophysical Union. Vol. 10. No. 2 p. 141–151.
VALIZADEH N., EL-SHAFIE A., MIRZAEI M., GALAVI H., MUKHLISIN M., JAAFAR O. 2014. Accuracy enhancement for forecasting water levels of reservoirs and river streams using a multiple-input-pattern fuzzification approach. The Scientific World Journal. Vol. 2014. Article ID 432976 pp. 9.
WRC 2010. Annual report. Water Resources Commission. ISBN 978-0-621-40211-7 pp. 156.

Mexoese NYATUAME, Sampson K. AGODZO

Stochastyczny model ARIMA do prognozowania rocznego opadu i maksymalnej temperatury w zlewni Tordzie w Ghanie

STRESZCZENIE

Prognozowanie opadu i temperatury jest trudnym zadaniem z powodu zmienności tych parametrów w czasie i przestrzeni, a także nieznajomości wszystkich czynników wpływających na opady w regionie czy w danej miejscowości. Prognozowanie opadów jest ważne dla rolnictwa i gospodarki zlewniowej, mających znaczący wkład w gospodarkę regionu. Przewidywanie opadu wymaga modelowania matematycznego i symulacji z powodu jego skrajnie nieregularnego i złożonego charakteru. Do analizy i prognozowania rocznych opadów i maksymalnej temperatury w zlewni Tordzie wykorzystano autoregresyjny zintegrowany model średniej ruchomej (ARIMA). Do zidentyfikowania modeli metodą oglądu wizualnego użyto funkcji autokorelacji (ACF) i cząstkowej autokorelacji (PACF). Testy stacjonarności przeprowadzono za pomocą testów Dickeya–Fullera (ADF), Manna–Kendalla (MK) i Kwiatkowskiego–Phillipsa–Schmidta–Shina (KPSS). Wybrane modele poddano ocenie i walidacji z użyciem skorygowanego kryterium Akaike (*AICC*) i Bayesowskiego kryterium Schwartza (SBC). Diagnostyczna analiza modeli obejmowała niezależność, normalność, homoscedastyczność, wykresy *P–P* i *Q–Q* dla reszt. Najlepsze modele ARIMA dla opadu w Kpetoe i Tordzinu miały postać (3, 0, 3) i (3, 1, 3), gdy wartości *AICC* równe odpowiednio 190,07 i 178,23. Modele dla maksymalnej temperatury w Kpetoe i Tordzinu miały postać (3, 1, 3) i (3, 1, 3), a ich odpowiednie wartości *AICC* wynosiły 23,81 i 36,10. Wydajność modelu sprawdzano, wykorzystując sumę błędu kwadratowego (*SSE*), średni błąd kwadratowy (*MSE*), średni bezwzględny błąd procentowy (*MAPE*) i pierwiastek ze średniego błędu kwadratowego (*RMSE*). Wyniki różnych analiz wykazały, że modele są odpowiednie i mogą stanowić pomoc w przyszłej gospodarce wodnej.

Słowa kluczowe: *ARIMA, modele opadu, prognozowanie, temperatura, zlewnia Tordzie*