SERBIAN JOURNAL OF ELECTRICAL ENGINEERING
Vol. 18, No. 1, February 2021, 75-94

UDC: 620.9:519.216 DOI: https://doi.org/10.2298/SJEE2101075O

Electricity Consumption Forecast for Tarkwa Using Autoregressive Integrated Moving Average and Adaptive Neuro Fuzzy Inference System

Robert Agyare Ofosu[1], Benjamin Odoi[1], Mercy Asamoah[1]

Abstract: Electricity has become one of the inelastic goods in our world today. The proper functioning of most equipment today relies on electricity. Taking Tarkwa which is a mining community into consideration, the various mines, schools, shops, banks and other companies in the municipality massively rely on electricity for their day to day running. Therefore, knowing the exact amount of electricity to produce and distribute for the smooth running of businesses and basic living is of great necessity. This study compared and formulated a model to forecast and predict the daily electrical energy consumption in Tarkwa for the year 2019. The data used was a monthly dataset for the year 2018 and it comprised of the temperature, wind speed, population and electricity consumption for Tarkwa. The methods used were Artificial Neuro-Fuzzy Inference System (ANFIS) and Autoregressive Integrated Moving Average (ARIMA). The ANFIS was used as a predictor to predict the electricity consumption based on the training and testing of the dependent and independent variables. The ARIMA was used to forecast the dependent and independent variables for 2019. These simulations were done using MATLAB and Minitab. The results of the analysis revealed that the training and testing dataset allowed ANFIS to learn and understand the system but the ANFIS could only forecast the 2019 electricity consumption after the input data to the system was changed to the ARIMA forecasted 2019 independent variables. It was observed that the amount of electricity consumed in 2019 increased by 14%.

Keywords: Electricity consumption, Error metrics, ARIMA, ANFIS, Forecast.

1 Introduction

Ghana generates electrical energy from hydropower, thermal energy and renewable energy sources, with some portions of the generated energy being exported to other countries. Tarkwa is noted as a center of gold and manganese mining city in Ghana. The Municipality has a total land area of 2354 km^2 and about 32% of the entire active population are engaged in agricultural production whilst the remaining 68% find themselves in the area of commerce, private

[1]Faculty of Engineering, University of Mines and Technology, Tarkwa, Ghana.
Emails: raofosu@umat.edu.gh; bodoi@umat.edu.gh; cyasaam96@gmail.com.

informal sector and hospitality industries [1]. GRIDCo and GENSER are the main companies that transmit power to Tarkwa and Electricity Company of Ghana (ECG) is the main distributor with the Bulk Supply Point (BSP) located at Atuabo. 131 kV is transmitted to the substation from GRIDCo. There are two transformers at this substation, a 131 kV transformer and an 11 kV step down transformer. ECG supplies 131 kV to the mining industries for heavy duty operations. For commercial users, the transmitted 131 kV is stepped down to 11 kV which is further reduced to 440/240 V for commercial and residential use. Electricity consumption is considered at this level [2].

The amount of electricity distributed should efficiently meet the public need. There is a problem when the amount of electrical energy supplied is lower than the demand which in turn results in overload. This disequilibrium brings about load shedding, blackouts and other related power issues. These power related problems are dependent on several factors such as lack of adequate power generating equipment, increased number of electricity consuming households, unfavorable weather conditions such as temperature wind speed and demographics. Hence, accurate forecasting of the amount of electricity that will be consumed at a particular time frame should be done to help the utility companies plan ahead to avoid these problems. Electrical load forecasting is the process of predicting the electrical power required to meet the short term, medium term or long-term demand. The electrical load forecasting has both commercial and technical implications and if not done properly, it may lead to bad planning and inefficient operation of the electrical power system. Accuracy of load forecasting is important to both the utility companies as well as the consumers. In addition, the forecast should rely on accurate data and best forecasting practices [3].

In this paper, Adaptive Neuro-Fuzzy Inference System (ANFIS) which combines fuzzy logic and artificial neural network as well as Autoregressive Integrated Moving Average (ARIMA) were implemented in time series forecasting of electrical loads in Tarkwa. These artificial intelligence techniques were implemented to help in the prediction of electricity consumption (demand) for Tarkwa. This was done to bridge the gap between supply and planning so as to help the various power stations know the exact amount of electrical power to send to Tarkwa.

This will curb the problem of power outages and other related issues which occur as a result of an imbalance between the supply and the demand and the lack of adequate machinery to produce the needed power to meet the demand of consumers. This will also help in policy development for the district and help balance the demand and supply to avoid any further future low power problems.

Literature reviewed indicated that many scientists and researchers were able to perform accurate forecasts of electricity consumption using various artificial intelligence and time series techniques such as neural networks optimised by particle swarm optimisation [4] fuzzy logic and Adaptive Neuro-Fuzzy Inference System (ANFIS) [5],[6] comparative analysis of regression analysis, neural networks and least square support vector machines, [7],[8] and use of Adaptive Conditional Density Estimation (ACDE) to perform household electricity demand forecast [9]. However, limited work has been done using ANFIS-ARIMA and no such forecasting has ever been conducted for Tarkwa district in the Western Region of Ghana. This research seeks to develop an ANFIS predictor and an ARIMA forecaster to accurately predict the daily electricity consumption of Tarkwa for 2019.

1.1 Adaptive neuro fuzzy inference system

Adaptive Neuro Fuzzy Inference System (ANFIS) is a multilayer network combination of ANN learning algorithms and Fuzzy Inference Systems (FIS) to map specific input parameters to an output. ANFIS uses conventional FIS interface and the ANN learning capacity to enhance the systems knowledge. The most used FIS model in ANFIS is the Takagi Sugeno model since it is computationally efficient and works well with optimisation and adaptive techniques [10].

Considering a FIS with two inputs, x and y, one output f, and a feedforward network consisting of five layers. Each layer comprises either square nodes or circular nodes. The adaptive or square nodes consist of parameter sets that can be modified while the parameters of the fixed or circular nodes cannot be modified. The formula for the node functions may vary from layer to layer. A simplified architecture of the ANFIS is illustrated in Fig. 1 [11].

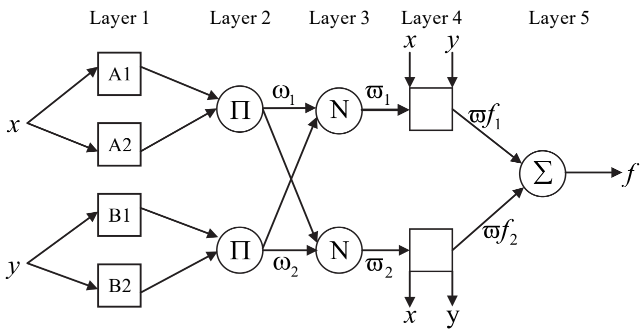

Fig. 1 – *ANFIS architecture.*

The layer 1 (o_i^1) consists of input parameters of membership functions and the formula for its outcome is expressed in (1). The parameters in this layer are known as the premise parameters.

$$o_i^1 = \mu A_i(x), \quad i = 1,2\,, \tag{1}$$

where x is input to node i, A_i is membership function associated with the node.

Layer 2 (o_i^2) consists of the rule nodes where each node calculates firing strength of the rule via multiplication. The formula for the firing angle and the total number of rules used in the structure is given in (2) and (3) respectively.

$$o_i^2 = \omega_i = \mu A_i(x) \times \mu B_i(y), \quad i = 1,2\,, \tag{2}$$

$$R_n = j^i\,, \tag{3}$$

where, x, y are inputs to the structure, A_i, B_i are membership function associated with the node, ω_i is firing strength, R_n is total number of rules, i is number of inputs, j is number of membership functions per input.

In layer 3 (o_i^3), each node computes the ratio of the i-th rule of the firing strength to the total of all the firing strengths. The nodes are referred to as averaging nodes and its outcome is shown in (4).

$$o_i^3 = \varpi_i = \frac{\omega_i}{\omega_1 + \omega_2}, \quad i = 1,2\,. \tag{4}$$

Each node in layer 4 calculates the contribution of the i-th rule towards the overall output and its outcome is expressed in (5).

$$o_i^4 = \varpi_i f_i = \varpi_i (p_i x + q_i y + r_i), \quad i = 1,2\,, \tag{5}$$

where p_i, q_i and r_i are consequent parameters, f_i is algebraic function of consequent parameters.

The final layer comprises the output node which sums the overall contribution of the layers and is expressed in (6) [12].

$$o_i^5 = f = \sum_i \varpi_i f_i = \frac{\sum_i \omega_i f_i}{\sum_i \omega_i}\,. \tag{6}$$

The ANFIS architecture identifies variables by applying two learning algorithms, the backpropagation algorithm and the hybrid learning algorithm. The learning algorithms compare the desired outputs to the measured system outputs and then the systems are tuned to narrow the difference between the two as much as possible. An effectively designed ANFIS is able to solve any nonlinear and complex problems with high precision [13]. Also, a study by the authors of [12] proved that the genfis1 and gaussmf were the best selection for an ANFIS predictor even though the other approaches were equally good [11].

1.2 Autoregressive integrated moving average

Autoregressive Integrated Moving Average (ARIMA) model is a generalisation of an Autoregressive Moving Average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting) [15].

Differencing is a method of transforming a non-stationary time series into a stationary one. The best way to determine whether or not the series is sufficiently differenced is to plot the differenced series and check to see if there is a constant mean and variance [16]. Formula (7) shows a second order differencing equation used to perform differencing on a model sample.

$$y_t^1 1 = y_t - y_{(t-1)} - (y_{(t-1)} - y_{(t-2)}) , \tag{7}$$

where y_t is deterministic time series, y_{t-1} is first differencing and y_{t-2} is second differencing

Adequacy of the model is determined using the Box- Pierce (Ljung- Box) Chi – Square Statistic test. The determining p-values for the test are dependent on the Chi-square values. When the *p*-value ≥ 1, the model is adequate. This means that the model has a confidence level of 95% and the null hypothesis can be rejected assuming a 5% chance of making a mistake. It can also be assumed that the values are showing dependence on each other.

1.3 Error metrics

A comparison of three (3) error metrics were made in order to check for their effect on the models.

Root mean square error

Root Mean Square Error (RMSE) is a measure of spread of the forecast errors about the actual data points. This means that, the RMSE explains how far or near the forecasted values of an estimated model are from the real data points. The formula is given as:

$$RMSE_{forecast} = \sqrt{\sum_{i=1}^{n} \left(\frac{Y_t - \hat{Y}_t}{n} \right)^2} , \tag{8}$$

where $\hat{Y}_t$ is the forecasted value, Y_t the actual data points, and n is the sample size.

Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) is a measure of the size of the error of a forecast in percentage. It is used to measure the accuracy of a forecast and is given as:

$$MAPE_{forecast} = \left(\frac{1}{n} \sum \frac{\left| Y_t - \hat{Y}_t \right|}{\left| Y_t \right|} \right) \times 100\% \,. \tag{9}$$

Mean Absolute Error

Mean Absolute Error (MAE) is the simplest measure of forecast accuracy. Thus, the mean of the absolute errors. The absolute value is the difference between the forecasted values and the actual values. MAE explains how big an error is expected from the forecast on average.

$$MAE = \left(\frac{1}{n} \sum \left| Y_t - \hat{Y}_t \right| \right). \tag{10}$$

It should be noted that statistically, the mean errors are referred to as MAE and could also be referred to as Mean Absolute Deviation (MAD) [17].

2 Materials and Methods Used

2.1 Data collection

The 2018 history of electricity consumption patterns in Tarkwa was studied. It was observed that temperature, wind and demographs had a major impact on the amount of electricity consumed in that year and so these factors were considered to predict the electricity consumption of Tarkwa for the subsequent year. Daily data was taken from online websites as well as from ECG Tarkwa. A total of 366 dataset points for the year 2018 was acquired for each variable. The independent variables considered for the simulation were temperature variations, wind speed and Tarkwa demographics and the dependent variable was the maximum daily electricity consumed by Tarkwa in 2018.

2.2 Resource employed

Electricity demand is dynamic, hence to obtain accurate results, ANFIS and ARIMA were selected out of the mass of AI techniques to alleviate this problem because of the ability of ANFIS to learn and understand the system and the statistical accuracy of ARIMA. The resources employed in constructing the forecasting system are of two major categories namely, hardware and software. The statistical software used in this study were MATLAB/ Simulink R2016a and Minitab 18.1.

The independent and dependent variables were divided into training and testing dataset. The training data comprised of independent and dependent variables from January to August of 2018 and the testing data comprised of independent and dependent variables from September to December of 2018 for Tarkwa. The ANFIS controller was not employed in this work but MATLAB codes were rather written to perform the ANFIS prediction. The steps used are shown in the flowchart of Fig. 2.

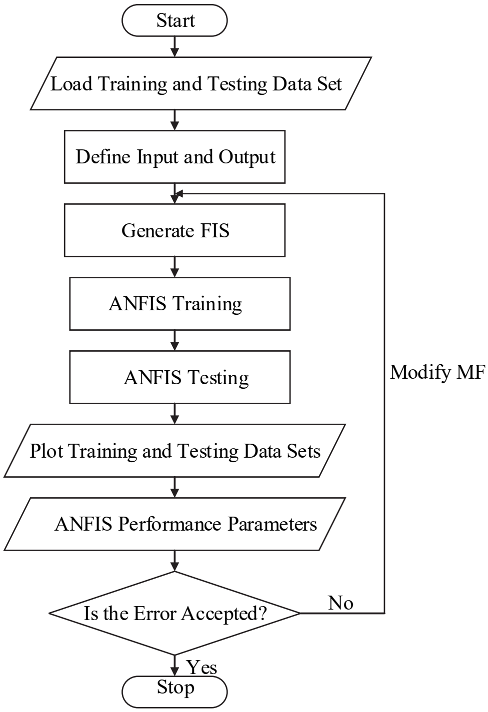

Fig. 2 – *Flowchart of ANFIS model.*

In this research, genfis1 was selected with three (3) MFs of gaussian type with a linear output MF type. Genfis1 and gassmf were selected in running the ANFIS predictor because gaussmf drew smooth curves which allowed for an effective representation of the data points. On the other hand, genfis1 generated all the possible rules and hence produced better results. The predictor was tested by running 100 epochs with an initial step of 0.01, a step size decrease rate of 0.9 and a step size increase rate of 1.1. After the selection process, the predictor displayed the results and the error was calculated.

Due to the inability of the ANFIS to make a forecast on its own, ARIMA was used to generate a forecast of the independent variables which was then applied to the trained ANFIS to predict the corresponding dependent variable or output which in this case is the electricity consumption for 2019. The error term after the program was run was accepted and the trained was set aside. In case the error is not accepted, the program should be run again and the input information for the MFs and number of epochs should be modified. The ANFIS approach is explained in Fig. 2 [18].

Data forecasting using ARIMA

The individual variables were tested for significance and adequacy. The 366 dataset were loaded into the Minitab workspace with a portion of the dataset shown in **Table 1.**

Table 1
Portion of Minitab dataset worksheet.

SN	High Temperature [°C]	Low Temperature [°C]	Wind Speed [km/h]	Population [millions]	Electricity Consumption [MW]
1.	33	23	24	107 389	29 660
2.	32	23	20	107 388	31 080
3.	32	23	21	107 387	30 750
4.	32	22	28	107 386	38 950
5.	30	23	22	107 385	37 920
6.	32	23	21	107 389	32 110
7.	33	24	26	107 384	29 730
8.	32	24	28	107 383	28 730

A try and error analysis was conducted on each variable to determine the ARIMA or ARMA non-seasonal values to use for the test before forecasting. The values with the least p-values were selected as they were found to be the most significant and hence an ARIMA (1, 0, 1) or an ARMA (1, 1) was used to test for all the variables except population which was non-invertible. Fig. 3 shows the ARIMA non-seasonal value selection pane.

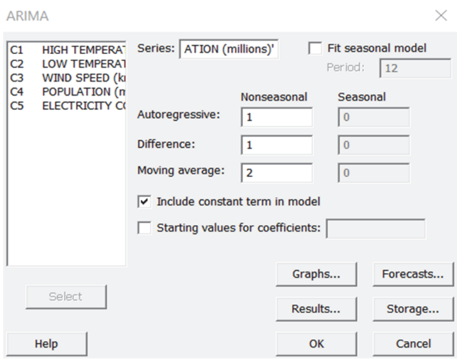

Fig. 3 – *ARIMA non-seasonal value selection pane.*

The general formula used in determining the type of ARIMA to use is as shown in (11)

$$ARIMA(p,q,d) : \nabla^d Y_t = (1-\mathrm{B})^d \varepsilon_t \, , \tag{11}$$

where p is AR of ARIMA, q is I of ARIMA, d is MA of ARIMA, B is back shift notation ($By_t = y_{t-1}$) and ε_t is error term.

In this model, an ARIMA (1, 0, 1) was seen to be the best model to use. Formula (12) gives the equation used to carry out the ARIMA (1, 0, 1) modelling.

$$\begin{aligned} & ARIMA(1,0,1) = ARMA(1,1) \\ & = y_t = c + \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \ , \\ & = c + \phi_1 B y_t + \varepsilon_t - \theta_1 B \varepsilon_{t-1} \end{aligned} \tag{12}$$

where c is constant term, $\phi_1 By_{t-1}$ is the AR part of ARIMA with a coefficient of ϕ, ε_t is error term and $\theta_1 B\varepsilon_{t-1}$ is the MA part of ARIMA with a coefficient of θ_1.

After all the variables were found to be significant and adequate using the p-values and the Ljung- Box Test, the forecast was carried out. The ARIMA approach is explained in Fig. 4 [16].

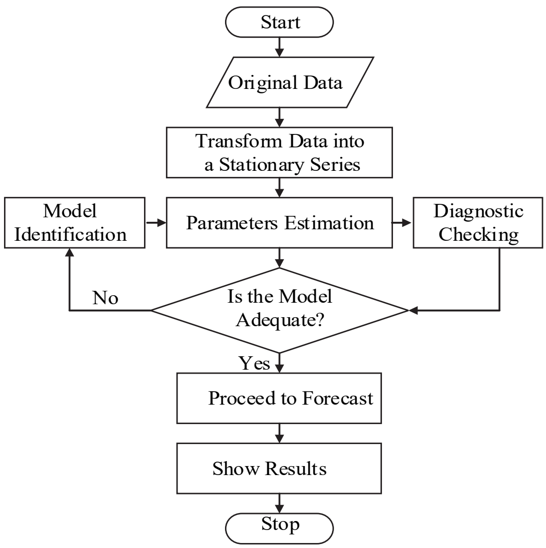

Fig. 4 – *ARIMA Model Flowchart.*

2.3 Simulations

The ANFIS and ARIMA were simulated with the dependent and independent variables and they were tested for significance, adequacy and errors. Due to the inability of the ANFIS to make the forecast with past data, the ARIMA was used to forecast the independent and dependent variables to obtain their corresponding 2019 values. The 2019 dataset for the independent variables was then inserted back into the ANFIS code to work with the trained data and forecast the 2019 values for the dependent variable. Both software gave a forecast of the dependent variable hence, an analysis was carried out to determine which technique was best to forecast electricity consumption for Tarkwa.

3 Results and Discussion

3.1 Trend analysis

Before the dataset was forecasted, a trend analysis was conducted to know the nature of the dataset. In measuring the accuracy of the model based on the variability of the error term, the three diagnostic checkings were considered that is the Mean Absolute Deviation (MAD), Mean Square Deviation (MSD) and the Mean Absolute Percentage Error (MAPE). Based on the Thumb Rule, the MAPE was the best measurement among them. The smaller the MAPE value, the more accurate is the model hence, the exact model of the variable was determined by choosing the trend model with the lowest MAPE value.

High temperature

Figs. 5 – 7 show the analysis carried out on the High Temperature dataset using linear, quadratic and exponential differencing. It was observed that the quadratic trend model had the smallest MAPE value of 4.92862. This indicated that the high temperature dataset was in a quadratic form.

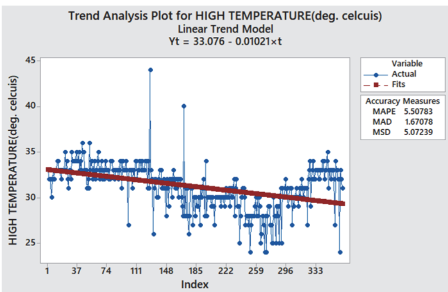

Fig. 5 – *Linear trend model for high temperature.*

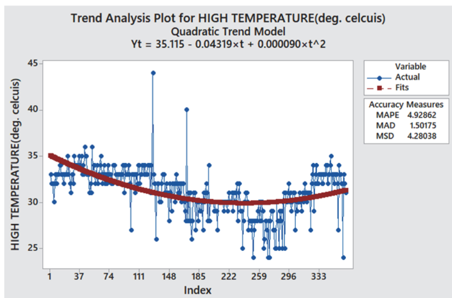

Fig. 6 – *Quadratic trend model for high temperature.*

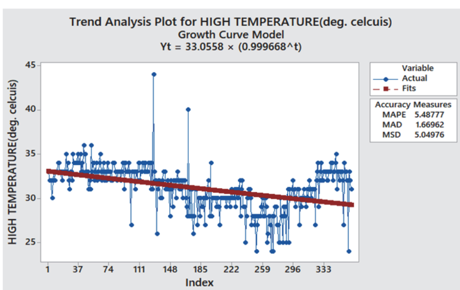

Fig. 7 – *Exponential growth curve trend model for high temperature.*

Low temperature

Figs. 8 – 10 show the analysis carried out on the Low Temperature dataset using linear, quadratic and exponential differencing. It was observed that the quadratic trend model had the smallest MAPE value of 11.2410. This indicated that the low temperature dataset was in a quadratic form.

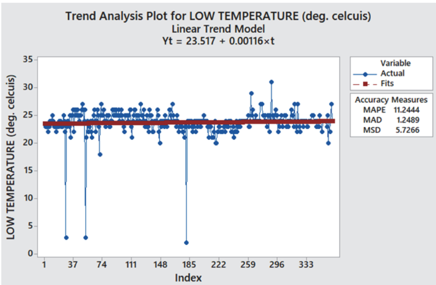

Fig. 8 – *Linear trend model for low temperature.*

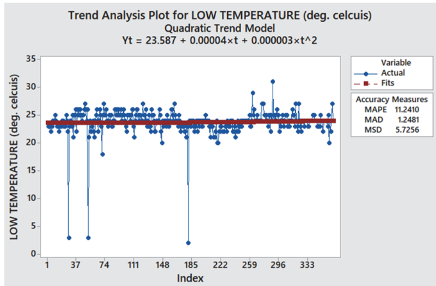

Fig. 9 – *Quadratic trend model for low temperature.*

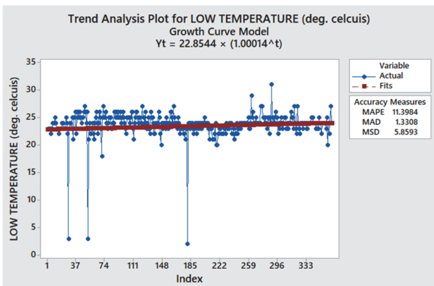

Fig. 10 – *Exponential growth curve trend model for low temperature.*

Wind speed

Figs. 11 and 12 show the analysis carried out on the Wind Speed dataset using linear and quadratic differencing.

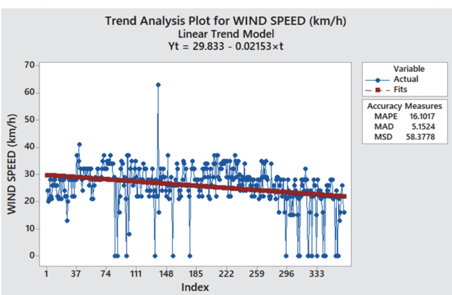

Fig. 11 – *Linear trend model for wind speed.*

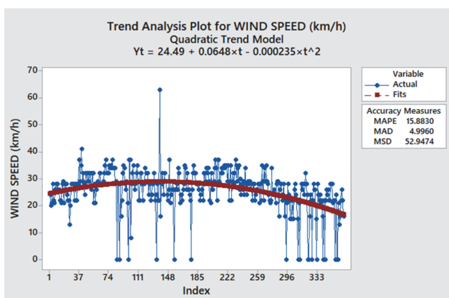

Fig. 12 – *Quadratic trend model for wind speed.*

Exponential Growth was not compatible with the model. It was observed that the quadratic trend model had the smallest MAPE value of 15.8830. This indicated that the wind speed dataset was in a quadratic form.

Electricity consumption

Figs. 13 – 15 show the analysis carried out on the Electricity Consumption dataset using linear, quadratic and exponential differencing. It was observed that the exponential growth curve trend model had the smallest MAPE value of 11.208. This indicated that the electricity consumption dataset was in an exponential form.

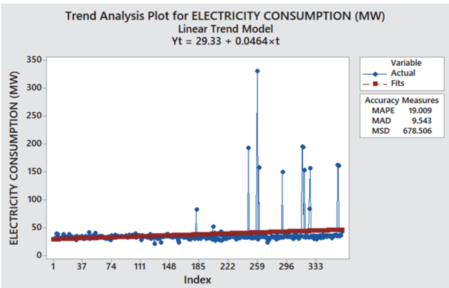

Fig. 13 – *Linear trend model for electricity consumption.*

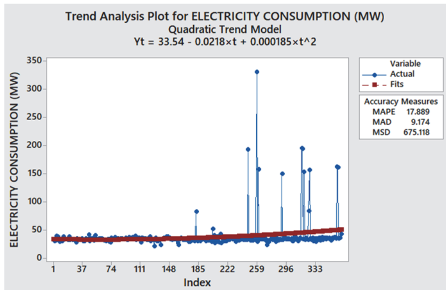

Fig. 14 – *Quadratic trend model for electricity consumption.*

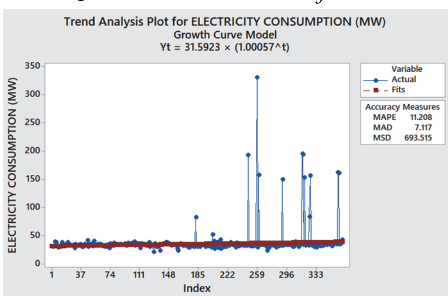

Fig. 15 – *Exponential growth curve trend model for electricity consumption.*

ANFIS training

The ANFIS was first used to train and test the independent and dependent variables. Fig. 16 shows the results obtained from the program after training the data. It shows a graph of the error in the system indicating the Mean Square Error (MSE), Standard Deviation Error (SDE), Root Mean Square Error (RMSE) and the Mean. Fig. 17 shows the results obtained from the program after testing the data and just as stated above for the train data, the same process is considered for the test data.

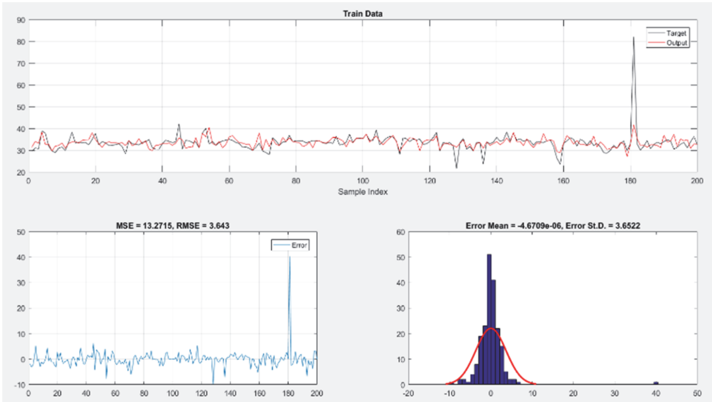

Fig. 16 – *Train data output results.*

The first graph shows the target output of the trained data to the actual output obtained. A spiked graph was obtained which indicated that the dataset was volatile. This means that the pattern of the dataset and its variance were constantly changing. For the error, a MSE, RMSE, error Mean and SDE of 13.2715, 3.643, -4.6709e-06 and 3.6522 were obtained respectively. These errors fell within an acceptable range between 0 and 100 hence the error was accepted and the training results were feasible. The mean and standard deviation gave a uniformly distributed graph.

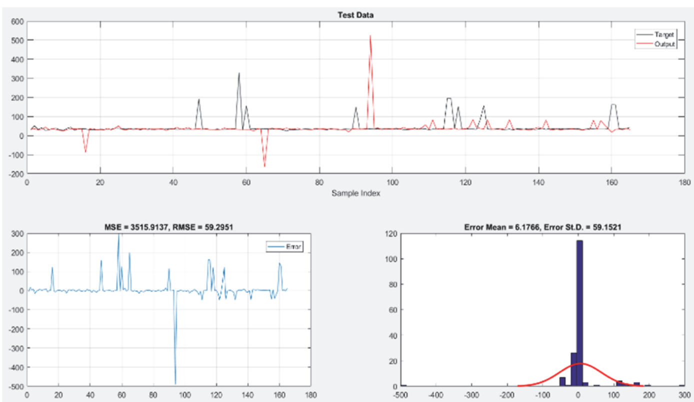

Fig. 17 – *Test data output results.*

From Fig. 17 the first graph shows the target output of the test data to the actual output obtained. A spiked graph was obtained which indicated that the dataset was volatile. This means that the pattern of the dataset and its variance were not constant and was constantly changing. For the error, a MSE, RMSE, error Mean and SDE of 3515.9137, 59.2951, 6.1766 and 59.1521 were obtained respectively. These errors fell within an acceptable range between 0 and 100 hence, the error was accepted and the training results were feasible. The mean and standard deviation gave a uniformly distributed graph. Fig. 18 shows the regression analysis of the training and test data showing the line of best fit against the plotted data.

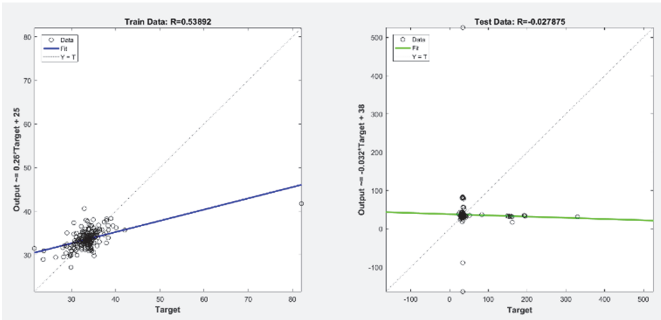

Fig. 18 – *Plots of regression results.*

A regression analysis was obtained for both the training data and the test data and a line of best fit was obtained to see the relationship between the dataset. A coefficient of regression value of 0.53892 was obtained for the training data and 0.027875 was obtained for the test data. Having a value between 0.05 ≤ R ≤1 showed that the training and test data were well correlated.

3.2 ARIMA analysis

ARIMA (1, 0, 1) was selected as the best model based on its p-value. A diagnostic checking was conducted using the Ljung- Box test. The Thumb Rule for checking the adequacy of the model is such that the p-value should be greater than the significance level chosen (5% level of significance). All the variables had p-values ≤ 0.05. This implied that all the variables were highly significant and a small change will affect the model significantly. In considering the adequacy of the ARIMA model for high temperature, it was observed that the p-values were 0.371, 0.411, 0.565 and 0.583 at their respective lags of 12, 24, 36 and 48 which are all greater than the significance level chosen. This showed the model was adequate. Formula (13) hence shows the ARIMA equation obtained for the high temperature variable.

$$y_t = 0.2323 + 0.9904 y_{t-1} + \varepsilon_t + 0.09114 \varepsilon_{t-1} . \tag{13}$$

Table 2 and **Table 3** show the final p-value estimates, Mean Square, and the Ljung- Box statistic test results for the independent and dependent variables. The analysis showed that all the variables were adequate and significant.

In considering the adequacy of the ARIMA model for low temperature, it was observed that the p- values were 0.911, 0.917, 0.039 and 0.092 at their respective lags of 12, 24, 36 and 48 which are all greater than the significance level chosen. This showed the model was adequate. Formula (14) hence shows the ARIMA equation obtained for the low temperature variable.

$$y_t = 2.7052 + 0.8860 y_{t-1} + \varepsilon_t + 0.78060 \varepsilon_{t-1} . \tag{14}$$

Table 2

ARIMA analysis results for high temperature.

Type	Coefficient	SE Coefficient	T-Value	P-Value
AR 1	0.99068	0.00852	116.32	0.00
MA 1	0.84580	0.03140	26.92	0.00
Constant	0.29240	0.01430	20.40	0.00
Mean	31.38	1.54		
Number of observations: 365				
Residual Sums of Squares				
DF	SS	MS		
362	1132.43	3.12827		
Back forecasts excluded				

Table 3

Diagnostic checking for high temperature.

Lag	Chi- Square	DF	P-Value
12	9.75	9	48.000
24	21.81	21	0.411
36	31.04	33	0.565
48	42.40	45	0.583

Table 4 and **Table 5** show the final p-value estimates, Mean Square, and the Ljung- Box statistic test results for the independent and dependent variables. The analysis showed that all the variables were adequate and significant.

Table 4

ARIMA analysis results for low temperature.

Type	Coefficient	SE Coefficient	T-Value	P-Value
AR 1	0.8860	0.0781	11.35	0.00
MA 1	0.7860	0.1040	7059.00	0.00
Constant	2.7052	0.0264	102.48	0.00
Mean	23.7280	0.2320		
Number of observations: 365				
Residual Sums of Squares				
DF	SS	MS		
362	2002.07	5.53057		
Back forecasts excluded				

Table 5

Diagnostic checking for low temperature.

Modified Box- Pierce (Ljung- Box) Chi- Square Statistic			
Lag	Chi-Square	DF	P-Value
12	4.00	9	0.911
24	12.75	21	0.917
36	48.66	33	0.039
48	58.06	45	0.092

In considering the adequacy of the ARIMA model for wind speed, it was observed that the *p*-values were 0.001, 0.051, 0.033 and 0.042 at their respective lags of 12, 24, 36 and 48 which are all greater than the significance level chosen. This showed the model was adequate. Formula (15) hence shows the ARIMA equation obtained for the low temperature variable.

$$y_t = 0.2323 + 0.9904 y_{t-1} + \varepsilon_t + 0.09114 \varepsilon_{t-1} . \quad (15)$$

Table 6 and **Table 7** show the final p-value estimates, Mean Square, and the Ljung- Box statistic test results for the independent and dependent variables. The analysis showed that all the variables were adequate and significant.

Table 6

ARIMA analysis results for wind speed.

Type	**Coefficient**	**SE Coefficient**	**T-Value**	**P-Value**
AR 1	0.9904	0.0156	63.54	0.00
MA 1	0.9114	0.0335	27.23	0.00
Constant	0.2323	0.0423	5.49	0.00
Mean	24.3100	4.4300		
Number of observations: 365				
Residual Sums of Squares				
DF		SS	MS	
362		19347.1	53.4451	
Back forecasts excluded				

Table 7

Diagnostic checking for wind speed.

Modified Box- Pierce (Ljung- Box) Chi- Square Statistic			
Lag	Chi-Square	DF	P-Value
12	27.29	9	0.001
24	32.55	21	0.051
36	49.42	33	0.033
48	62.69	45	0.042

In considering the adequacy of the ARIMA model for electricity consumption, it was observed that the p- values were 0.016, 0.329, 0.131 and 0.150 at their respective lags of 12, 24, 36 and 48 which are all greater than the significance level chosen. This showed the model was adequate. Formula (16) hence shows the ARIMA equation obtained for the low temperature variable.

$$y_t = 013.66 + 0.639 y_{t-1} + \varepsilon_t + 0.477 \varepsilon_{t-1}. \quad (16)$$

Table 8 and **Table 9** show the final *p*-value estimates, Mean Square, and the Ljung- Box statistic test results for the independent and dependent variables. The analysis showed that all the variables were adequate and significant.

Table 8

Arima analysis results for electricity consumption.

Type	**Coefficient**	**SE Coefficient**	**T-Value**	**P-Value**
AR 1	0.639	0.175	3.65	0.000
MA 1	0.477	0.200	2.39	0.018
Constant	13.660	0.713	19.16	0.000
Mean	37.860	1.980		
Number of observations: 365				
Residual Sums of Squares				
DF	SS	MS		
362	245448	678.032		
Back forecasts excluded				

Table 9

Diagnostic checking for electricity consumption.

Modified Box- Pierce (Ljung- Box) Chi- Square Statistic			
Lag	Chi-Square	DF	P-Value
12	20.39	9	0.016
24	23.28	21	0.329
36	42.20	33	0.131
48	54.83	45	0.150

3.3 Forecasting

The 2019 Electricity Consumption for Tarkwa was forecasted using both ARIMA and ANFIS on the 366 datasets for 2019 of both the independent and dependent variables. A graph of the forecasted electricity consumption for Tarkwa was obtained as shown in Fig. 19. It illustrates a graph of the days in the year 2019 against the consumed Megawatts (MW) with the ANFIS and ARIMA prediction.

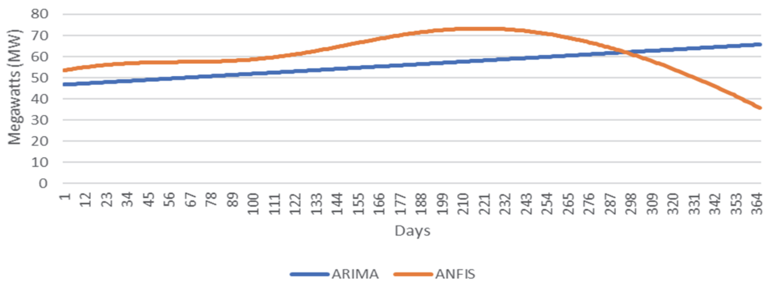

Fig. 19 – *Graph of days against demanded MW for ARIMA and ANFIS.*

3.4 Simulation results and discussion

The results from the ARIMA proved that all the datasets were significant having p–values ≤ 0.05 and adequate having p–values ≥ 0.05. This proved that any change in the dataset will affect the results. Also, when the forecasted dataset was incorporated in the MATLAB code to get the ANFIS prediction, the trained dataset was used alongside the new input from ARIMA to predict the 2019 electricity consumption for Tarkwa.

A graph of ANFIS and ARIMA was obtained for the predicted dataset. The graph was varying continuously until day 298. At day 298, the two graphs met and the ANFIS graph started decreasing while the ARIMA graph remained constant with time. Considering the quadratic and exponential graphs obtained from the trend analysis, it implied that the nature of the output graph should also be increasing and decreasing with time. Also, the MSE obtained for the ANFIS was smaller than that obtained for ARIMA. This proved that the ANFIS gave a better prediction than the ARIMA. Finally, it was observed that the consumption for 2019 increased. The percentage increase was calculated as follows.

$$\%\,\text{increase} = \frac{\text{increase}}{\text{originalnumber}} \times 100 = \frac{33.8115}{34.67} \times 100 = 14.469\%\,. \qquad (17)$$

It can be seen that the electricity consumption increased by 14% after the prediction was done.

4 Limitations

The following are some of the problems incurred during this research:

1. The system did not work with the population dataset. This was because the data collected for Tarkwa's population were significantly the same throughout the selected time frame, hence, all the analysis carried out on this dataset gave 100% accuracy which is logically not correct in mathematical terms as there is no perfect system.

2. The problem of insufficient data received for Tarkwa and also missing data for some of the days in the year, posed some difficulties in this research.

5 Conclusion and Recommendations

Conclusively, that ANFIS was seen to be a better technique for prediction because of its ability to learn and understand the system. ARIMA was seen to be a better tool for giving a future projection of the data due to the inability of ANFIS to perform such function. Based on the outcomes, electricity consumption for Tarkwa increased by a minimum of 14% for 2019 and beyond.

It is therefore recommended that GRIDCo and ECG should plan and find ways to supply more power to consumers in Tarkwa to cater for the increase in demand and curb power outages.

For further research, more independent variables should be considered in performing the forecast. This research can also be modified to perform forecasts for a larger time frame and larger dataset should be considered. Finally, other artificial intelligence techniques can be integrated to forecast the electrical consumption of Tarkwa.

6 Acknowledgments

The authors appreciate the University of Mines and Technology for making available the necessary resources for the completion of this research work.

7 References

[1] Tarkwa Nsuaem Municipal Assembly,
Available at: http://www.ghanadistricts.com/Home/District/215, Accessed: April 1, 2019.

[2] About the Power Sector in Ghana, Available at: http://www.ecgonline.info, Accessed: March 20, 2019.

[3] Electric Load Forecasting: Advantages and Challenges, Available at: http://engineering.electrical-equipment.org/electrical-distribution/electric-load-forecasting-advantages-challenges.html, Accessed: January 23, 2019.

[4] M. A. Farahat, A. F. Abd Elgawed, A. Ibrahim, H. M. M. Mustafa: Short Term Load Forecasting Using Neural Networks and Particle Swarm Optimization, Journal of Electrical Engineering, Vol. 15, No. 3, 2015, pp. 68 – 77.

[5] H. H. Çevik, M. Çunkaş: Short-Term Load Forecasting Using Fuzzy Logic and ANFIS, Neural Computing and Applications, Vol. 26, No. 6, August 2015, pp. 1355 – 1367.

[6] H.- W. Peng, S.- F. Wu, C.- C. Wei, S.- J. Lee: Time Series Forecasting with a Neuro-Fuzzy Modelling Scheme, Applied Soft Computing, Vol. 32, July 2015, pp. 481 – 493.

[7] F. Kaytez, M. C. Taplamacioglu, E. Cam, F. Hardalac: Forecasting Electricity Consumption: A Comparison of Regression Analysis, Neural Networks and Least Squares Support Vector

Machines, International Journal of Electrical Power and Energy Systems, Vol. 67, May 2015, pp. 431 – 438.

[8] S. Saravanna, S. Kannan, C. Thangaraj: Prediction of India's Electricity Demand Using ANFIS, ICTACT Journal on Soft Computing, Vol. 5, No. 3, April 2015, pp. 985 – 990.

[9] F. Amara, K. Agbossou, Y. Dube, S. Kelouwani, A. Cardenas, J. Bouchard: Household Electricity Demand Forecasting Using Adaptive Conditional Density Estimation, Energy and Buildings, Vol. 156, December 2017, pp. 271 – 280.

[10] J. Tziu Dzib, E. J. Alejos Moo, A. Bassam, M. Flota-Bañuelos, M. A. Escalante Soberanis, L. J. Ricalde, M. J. López-Sánchez: Photovoltaic Module Temperature Estimation: A Comparison between Artificial Neural Networks and Adaptive Neuro Fuzzy Inference Systems Models, Proceedings of the International Symposium on Intelligent Computing Systems, Mérida, Mexico, March 2016, pp. 46 – 60.

[11] M. Sahin, R. Erol: A Comparative Study of Neural Networks and ANFIS for Forecasting Attendance Rate of Soccer Games, Mathematical and Computational Applications, Vol. 22, No. 4, November 2017, pp. 1 – 12.

[12] N. Talpur, M. N. M. Salle, K. Hussain: An Investigation of Membership Functions on Performance of ANFIS for Solving Classification Problems, Proceedings of the International Research and Innovation Summit, Melaka, Malaysia, May 2017, pp. 1 – 8.

[13] R. A. Ofosu, E. Normanyo, L. Obeng: Temperature Control of Heaters in Cable Extrusion Machine Using PSO-ANFIS Controller, Proceedings of the IEEE AFRICON 2019, Accra, Ghana, September 2019, pp. 1 – 9.

[14] How to Create an ARIMA Model for Time Series Forecasting with Python,
Available at: https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/, 2017, Accessed: February 28, 2021.

[15] Time Series Forecasting - ARIMA Models, Available at:
https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06
Accessed: January 23, 2019.

[16] S. Twumasi-Ankrah, B. Odoi, W. Adoma Pels, E. H. Gyamfi: Efficiency of Imputation Techniques in Univariate Time Series, International Journal of Science, Environment and Technology, Vol. 8, No. 3, June 2019, pp. 430 – 453.

[17] N. O. Bushara, A, Abraham: Using Adaptive Neuro-Fuzzy Inference System (ANFIS) to Improve the Long- Term Rainfall Forecasting, Journal of Network and Innovative Computing, Vol. 3, No. 1, 2015, pp. 146 – 158.