Towards the Establishment of Robust Load Forecasting Mechanism in Tanzania Grid: Effect of Air Temperature and Daytime on Electricity Consumption in Residential Buildings

Hussein Bakiri*, Hellen Maziku*, Nerey Mvungi*, Ndyetabura Hamisi*, Massawe Libe**

Hello professor. Professor Nerrey Mvungi: 0000-0002-8254-3994. Dr Libe MASSAWE: 0000-0003-3367-4813.Dr Ndyetabura Hamisi: 0000-0002-7366-5877.

*Department of Computer Systems Engineering, College of Information and Communication Technologies, researcher at the department, P.O.Box 33335, Dar es Salaam, Tanzania

**Department of Electronics and Telecommunication Engineering, College of Information and Communication Technologies, researcher at the department, P.O.Box 33335, Dar es Salaam, Tanzania

(hussein.bakiri@ifm.ac.tz, maziku.hellen@udsm.ac.tz, nhmvungi@udsm.ac.tz, yhamisi@udsm.ac.tz, massawe.libe@udsm.ac.tz)

‡ Corresponding Author; Hussein Bakiri, Postal address, Tel.: +255 718 572 299; E-mail adress: hussein.bakiri@ifm.ac.tz

Received: 10.02.2021 Accepted:15.03.2021

Abstract- The current phenomenon of electric power management has been towards the adoption of smart grid technologies to achieve efficient utility management processes such as transmission and distribution. Electric load forecasting has become an important aspect of smart grid technologies due to its capability of anticipating the power demand of a particular domain. The effective design of any load forecasting mechanism requires a critical investigation of electricity consumption determinants, following the diversification of social, economic, meteorological, and demographic grounds. Many research works have attempted to investigate the effect of temperature and daytime on electricity usage, targeting a specific country. These works have reported the existence of various degrees of causality based on a particular country under investigation. The variation in the findings from different research works has been a motivation to establish this study which aims at examining the impact of air temperature and daytime on load consumption in Tanzania. Four-years (2015-2018) load and weather data have been collected from Tanzania Electric Supply Company (TANESCO) and Tanzania Meteorological Agency (TMA) respectively. The k-mean algorithm is used to detect outliers and missing values in the load dataset before further processing. Furthermore, the Shapiro-Wilk normality test method is applied to identify data distribution patterns which in turn leads us to the correct choice for Spearman's rank correlation method. Results indicate that there is no linear relationship between electricity consumption and air temperature in residential buildings. Furthermore, the findings indicate the existence of a strong causality degree between electricity consumption and daytime in Tanzania.

Keywords Smart grid, load forecasting, electricity consumption determinants, temperature, daytime

1. Introduction

The smart grid can be considered as a modern electric power grid infrastructure for enhanced efficiency and reliability through automated control, high-power converters, modern communications infrastructure, sensing, and metering technologies to optimize generation, transmission, distribution, and service availability [1]. In Tanzania, the primary distribution is automated leaving the secondary one not properly cyber-managed. In the effort to establish smart distribution technologies, the University of Dar es Salaam, in conjunction with Swedish International Development Agency (Sida), has facilitated research focusing on Automatic Fault Clearance (AFC) in the secondary electric distribution network. Electric load forecasting is an important agency within the AFC mechanism which helps to anticipate power consumption demand, and thus be useful for providing efficient means of load management to distributed energy as well as service restoration. The effective design of load demand forecasting mechanism requires critical analysis on the factors affecting electricity consumption,

The effect of daytime and temperature on electricity consumption from various countries has been investigated in several research works including [2]–[6] and [2], [7]–[10] respectively. The existence of several research studies on the aforementioned research works follows the fact that countries experience different social, economic, technological, and meteorological grounds. The variation in the findings from several research contexts concerning how electricity consumption associates with temperature and daytime, is the main motivation of this paper. However, the effective design of load forecasting model requires a critical investigation of load consumption drivers [11]. This paper aims at investigating the effect of temperature and daytime on electricity consumption in Tanzania. To our knowledge, this is the first study in Tanzania attempting to analyze the effect of air temperature and time of a day on short-run electricity consumption variation.

Having known the impact of daytime and temperature on electricity demand in Tanzania, may pave a way for future research works that would attempt to propose load management models such as forecasting, electricity pricing, load shifting, load balancing, and maintenance schedule.

2. Literature Survey

2.1. The Trend of Load Forecasting Mechanisms in Developing Countries

Research works that investigated factors influencing energy consumption in developing countries have reported that; developing countries experience different economic, geographical, and cultural grounds from the developed countries. Therefore, the existing load forecasting approaches may not be appropriate when subjected to different context [12]. Following this fact, it can be observed that the load forecasting mechanism in developing countries needs to be modeled in such a way that it accommodates country-specific electricity consumption determinants.

It is reported that the recent soft computing techniques in forecasting, such as neural networks, depends on the quality of the labeled data, that is; the more meaningful the data is, the more accurate the developed forecasting model will be [13]. [2], [14] report the presence of missing values and outliers in the load dataset due to frequent power outages and faults. Therefore, it can be inferred from the analytical results presented in Fig.1 that the trend of load forecasting mechanisms in developing countries needs to be towards the inclusion of data cleansing pre-processer.

Further implication of the application of outlier preprocessing mechanism can be inferred from the observations through 19 forecasting models applied in the developing countries as seen in Table 4 in the appendix.

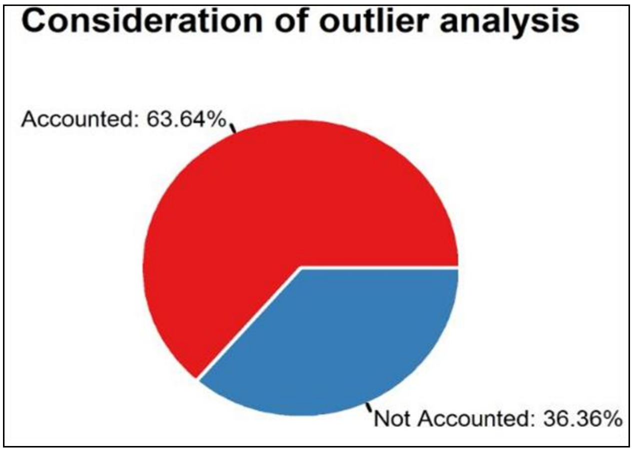

Fig. 1. The analytical report of the inclusion of outlier preprocessing as observed from 19 forecasting models in developing countries

2.2. Factors Affecting Electricity Consumption in Developing Countries

As presented in Table 3 in the appendix, electricity consumption in developing countries is driven by several factors depending on a particular country. Among the 20 sampled developing countries, 16 investigated the electricity consumption determinants on long-term basis, while 4 studies focused on short-run factors. Among 16 studies that focused on long-run determinants, GDP has been identified to cause significant effect in 9 publications (similar to 50%) followed by income and price of electricity by 16.67% (observed in 3 studies) each. However, population, urbanization, number of customers, and foreign exports exert less impact in the long-run (5.56% each) as presented in Fig. 2.

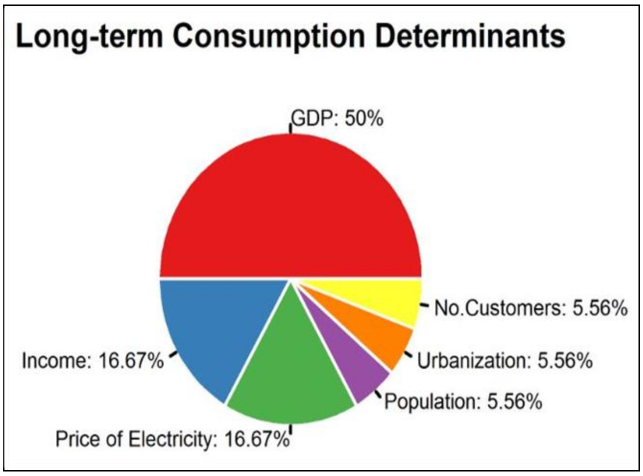

Fig. 2. The Analytical results of common long-term electricity consumption determinants in 20 developing countries

Conversely, temperature as a factor, was found to produce high degree of correlation in short-term basis as observed in four (4) among the 20 publications. Daytime and calendar events follow after the temperature as seen in Fig. 3. Therefore, it can be concluded that GDP, price of electricity, and income are the main electricity consumption drivers in developing countries for long-term basis. Furthermore, weather (temperature), daytime, and calendar events are the common determinants for short-term forecasting in developing countries.

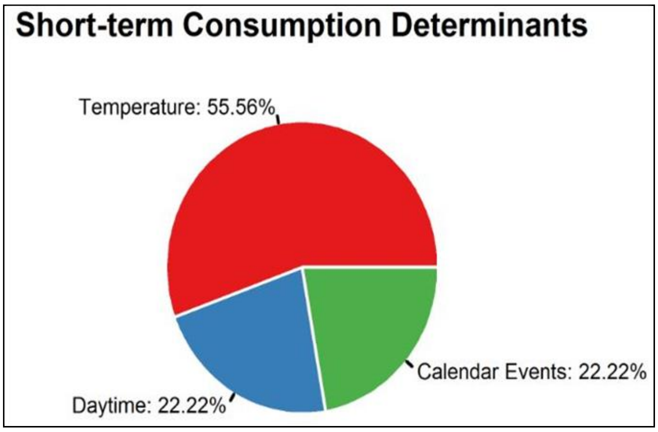

Fig. 3. The Analytical results of common short-term electricity consumption determinants in 20 developing countries

2.3. Related Works

2.3.1. The Relationship between Air Temperature and Electricity

The relationship between electricity consumption and the daily air temperature in Spain is investigated by [15] using the population-weighted temperature index method. The findings show that the electricity consumption is influenced by temperature change more significantly in winter than in summer (correlation coefficient values for summer and winter being 0.79 and 0.87 respectively). The authors use aggregated data from all economic sectors (industrial, residential, and commercial).

A study to investigate the relationship between electricity demand and temperature in the European Union is analyzed by [7] and researchers found that temperature has a large influence on electricity consumption. A short-term regression model that determines the influence of temperature on daily peak electricity demand for South Africa is developed by [8] using piecewise linear regression and report that the electricity consumption increases significantly for temperature values below 18ºC and slightly for values above 22ºC. [9] investigate the impact of daily temperature on power consumption for Kragujevac (Serbia) using visual inspection and regression techniques and observes a strong correlation between changes in external temperature and electricity consumption in residential and commercial sectors.

Research to investigate factors affecting short-run load forecasting is conducted in Pakistan by [2] using regression analysis and visual inspection in which the authors found that the consumption is affected by temperature, relative humidity, precipitations, wind speed, cloud cover, and light intensity. The study concludes that positive correlation is observed for temperature higher than 25°C and negative for temperature less than 20°C. The authors further report that, the relationship is least significant for temperature between 20°C and 25°C. Furthermore, they point out that the relationship between temperature and electricity consumption is very complex as such they that cannot be modeled linearly.

A study to investigate the impact of weather variation on energy use in the short-run for two residential houses (house 1 with advanced efficiency features and house 2 with more advanced efficiency features) in the USA using regression analysis has been conducted by [10]. The authors find out that there is a one-unit increase in degree heating and cooling day increases energy use by 9% and 5% for houses 1 and 2 respectively.

2.3.2. The Relationship between Electricity Consumption and Daytime

The electricity consumption trend concerning the time of a day has been investigated by [2]–[6]. The studies found that the daily consumption pattern varies according to the building type, characteristics of the studied area, and economy of a particular country under consideration. Studies conducted by [2], [3] in residential buildings found that the peak load is observed between 16:00 and 23:00.

From the analytical review of literature, it can be pointed out that research findings differ from study to study. Furthermore, the direction of consumption with temperature depends on whether it is the residential or commercial sector. It has further been noticed that in some countries (such as in the USA) the impact of air temperature on electricity consumption depends upon weather season (winter and summer). The difference in the extent of the driver’s effects on load consumption has motivated us to establish the study to examine the impact of load consumption determinants in Tanzania.

2.4. Data Analysis Techniques

2.4.1. The k-mean Algorithm

The k-mean is the simplest unsupervised machine learning algorithm that solves some clustering problems [16]. Given the number of centroids, clusters, and centroid values, data points close to each centroid can be identified. The k-mean clustering algorithm is trying to find out, for a given number of clusters (say k clusters), members of a vector (say v) comprised of vector values (x_i where i=1, 2, …., n) that are closer to centroids (say, S_j where j =1,2,….k). The clustering in the k-mean algorithm is achieved by determining the Euclidian distance between x_i and S_i shown in equation (1). The x_i belongs to S_j if at all the distance is minimum [17].

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\left(d(x_i,S_j)\right)^2 \quad (1)$$

Where $d(x_i,S_j)$ denotes Euclidean distance between the points x_i and the centroid S_j.

2.4.2.Shapiro-Wilk Normality Test

The normality test is an undeniable phenomenon since many statistical methods are built under the assumption that data follow the normal distribution. The three major categories of normality test methods are graphical methods, numerical methods, and formal normality tests [18].

Shapiro-Wilk test is one of the formal normality test methods. Other formal normality test methods include Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling tests. In the study to compare the three formal methods conducted by [18], the Shapiro-Wilk test was found to be superior over the counterparts. More background concepts about the Shapiro-Wilk normality test can be found in [19]. The original mathematical representation of the Shapiro-Wilk test is given in equation (2).

$$w = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad (2)$$

Where $x_{(i)}$ are the ordered sample values, a_i are constants generated from mean, variance and covariance, w is a W statistics, and n is the sample size.

The interpretation of the Shapiro-Wilk test is based on the p-value inferred from the w value that depicts the degree of the distribution. The distribution is normal if the p-value is greater than 0.05 and otherwise if the data deviates from normal distribution.

2.5. The Spearman-Rank Correlation

Correlation analysis between quantitative variables can be conducted using scatter diagrams, product-moment (Karl Pearson), regression, least squares, and rank-based (Spearman's rank and Kendall's tau) methods [20]. All the techniques are based on statistical methods and the underlying difference is whether the relationship is linear or non-linear, the nature of variable measurements (continuous vs ordinal), and the number of ties.

In recent years, the product-moment and rank-based methods have been widely used in analyzing the correlation between quantitative variables. However, the Karl Pearson method is ideal for describing linear relationships and working only with a continuous scale. In real-world data, non-linear relationships and different data measurement scales are expected. The rank-based method is ideal in the situation for non-normally distributed data as well as working better for both continuous and ordinal scaled measurements [21]. Spearman's and Kendall's tau methods are compared extensively in the study by [21] and the latter found to work better if the data contains a large number of ties.

The original formulation of spearman's method is presented in equation (3) in which the number of ties is not taken into account [21].

$$r_s = 1-((6*\sum_{i=1}^{N} D_i^2)/(N*(N^2-1)) \quad (3)$$

Where r_s is the spearman's coefficient, N is the total number of elements in a sample, and D is the rank's difference in each iteration.

Furthermore, equation (3) has been modified in [21] to accommodate the number of ties in the dataset and the new equation is presented in equation (4).

$$r_s = \frac{\sum_{i=1}^{n}\{(x_i-\bar{x})(y_i-\bar{y})\}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (4)$$

Where; x_i is a rank of an individual item y_i of y values. $\bar{x}$ and $\bar{y}$ are average ranks and n is the total number of samples.

Correlation coefficients show the degree of strength between the variables under investigation. Correlation coefficient values generally range from -1 to +1 representing strong negative to strong positive correlation [22].

3. Methodology

3.1. Research Design and Data Description

This study is based on a correlational research design in which the effect of air temperature and day time to power consumption is investigated. The archival data is used as a method of data collection. Furthermore, data is processed quantitatively using statistical techniques. Fig. 4 shows the research design phases undergone in this study.

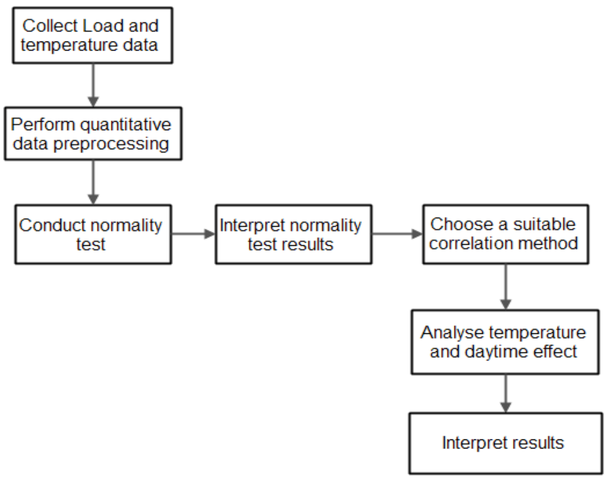

Fig. 4. Research design processes undertaken in this study

Electricity consumption demand in Tanzania has been slightly increasing from 2015 to 2018 as seen from Fig. 5, especially in urban areas, such as in the city of Dar es salaam. The increase in consumption might be caused by intermittent fluctuation in population, new connections, industrialization, and economic growth in general. The subnetwork is connected by three transformers, named BBQ Village SS1, Abiudi Street, and Kimweri (with the power of 315KVA, 200KVA, and 100KVA respectively) serving an average of 143 customers each, with a mean consumption of 30KWh per every 20 minutes

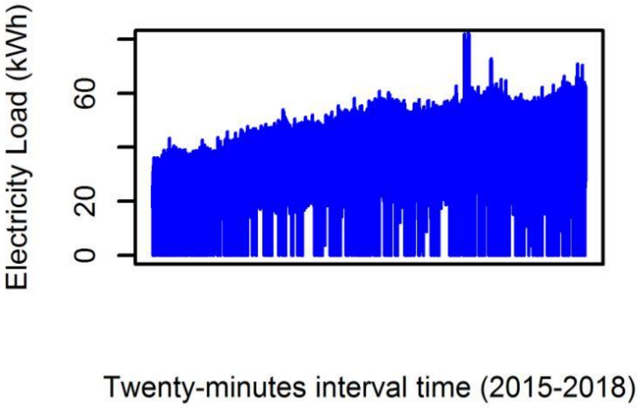

Fig. 5. Load consumption growth for 2015-2018, represented by Mikocheni area in Dar es salaam

3.2. Data Collection

The twenty-minute interval load data from 2015 to 2018 has been collected from the Mikocheni area in Dar es Salaam through Automatic Meter Reading (AMR) device situated at one of the three distribution transformers. The study area comprises of 143 houses, characterized as residential buildings. Similarly, the three-hours interval temperature data from 2015 to 2018 was collected from TMA.

3.3. Data Analysis and Pre-Processing

Shapiro-Wilk test outwits the counterparts and is found to be superior in identifying distribution trends. The Shapiro-Wilk test method from equation (2) is used in this study to examine load distribution trend before deciding which correlation technique fits the data pattern.

The k-mean algorithm shown in equation (1) is used to detect the presence of outliers and missing values in load data. The determination of outliers is aided by the knowledge from the TANESCO expert of which all values below 15 are considered to be outliers. Three clusters are formed (cluster 1, cluster 2, and cluster 3) containing centroid values of 8, 25, and 40 of which all data that cluster 1 contains are termed as outliers or missing values. Depending on the number of outliers detected, the “pruning” method is applied as a means of data cleansing.

3.4. Correlation Analysis

The spearman’s correlation test method expressed in equation (4) is used to examine the association between temperature and electricity usage. Spearman’s method has a great ability to handle ties as well as nonlinearly distributed data. The values of correlation coefficients are then used to determine the direction of correlation between temperature and electricity usage. Furthermore, visual analysis techniques, such as graph plots are also used to investigate the direction of change between load and temperature. Correlation analysis is conducted in the R programming language.

3.5. Analyzing the Effect of Daytime

Visual analysis techniques are used to investigate the trend of variation between electricity consumption and time of the day. The graph plots show how electricity is consumed at midnight, dawn, morning, afternoon, evening, and night. The graphs are plotted in the R programming language.

4. Results and Discussion

4.1. Load Distribution in Tanzania

Following the outlier detection process, empirical results in this study indicate presence of 5207 outliers and missing values in a sample of 105192 load data (5.17852% of the data are corrupted). The detected outliers and missing values are found to be unevenly distributed throughout the dataset, such that the pruning process becomes sounding. The occurrence of corrupted entries in the Tanzania load data conforms to recent observations in various research studies such as [2], that the dataset in developing countries is usually characterized by the presence of missing values and outliers.

The results after running a Shapiro-Wilk test in R, indicates that the load consumption trend is not normally distributed, because in all four cases (2015-2018), the p-value (the probability value of *W* statistics) is found to be less than 0.05 as shown in Table 1. Furthermore, the p-values in Table 1 which are less than 2.2e-16 negates the null hypothesis (that the distribution is normal) for all values of p in all four years under investigation.

Table 1. The Normality test results for 2015-2018 data using the Shapiro-Wilk method

Year	Confidence Intervals	Interpretation
2015 (Jan– Dec)	W = 0.86403, p-value < 2.2e-16	The Null Hypothesis is rejected; Not normally distributed.
2016(Jan – Dec)	W = 0.85805, p-value < 2.2e-16	The Null Hypothesis is rejected; Not normally distributed.
201 (Jan – Dec)	W = 0.88961, p-value < 2.2e-16	The Null Hypothesis is rejected; Not normally distributed.
2018(Jan – Dec)	W = 0.89076, p-value < 2.2e-16	The Null Hypothesis is rejected; Not normally distributed.

Having identified the actual distribution trend of load data, helped us to determine the appropriate method to apply in further processing. This is because various statistical methods are based on the distribution of the data. Furthermore, it is the normality test result that has led this study to choose the spearman-rank correlation method since the approach is well suited for data that does not fall under normal distribution.

4.2. Load versus Temperature

After running the spearman's correlation test for load vs temperature data, the following results were observed and presented in Table 2. For all four cases (2015, 2016, 2017, and 2018) a negative correlation is observed. The correlation values of -0.2967123, -0.2334719, -0.2350168, and -0.2412897 indicate the existence of no relationship between load consumption and temperature in residential buildings. Concerning the p-values (2.2e-16) observed from the experimentation, it provides strong confidence for rejecting the null hypothesis (that is, there is existence of a correlation between the variables).

Fig. 6 presents graphical plots of electricity consumption versus air temperature in the four years. In all the four plots the regression line (shown in red) seems to be non-linear and data points are randomly distributed. The graph plots provide further insight of abnormal relationship between electricity and temperature.

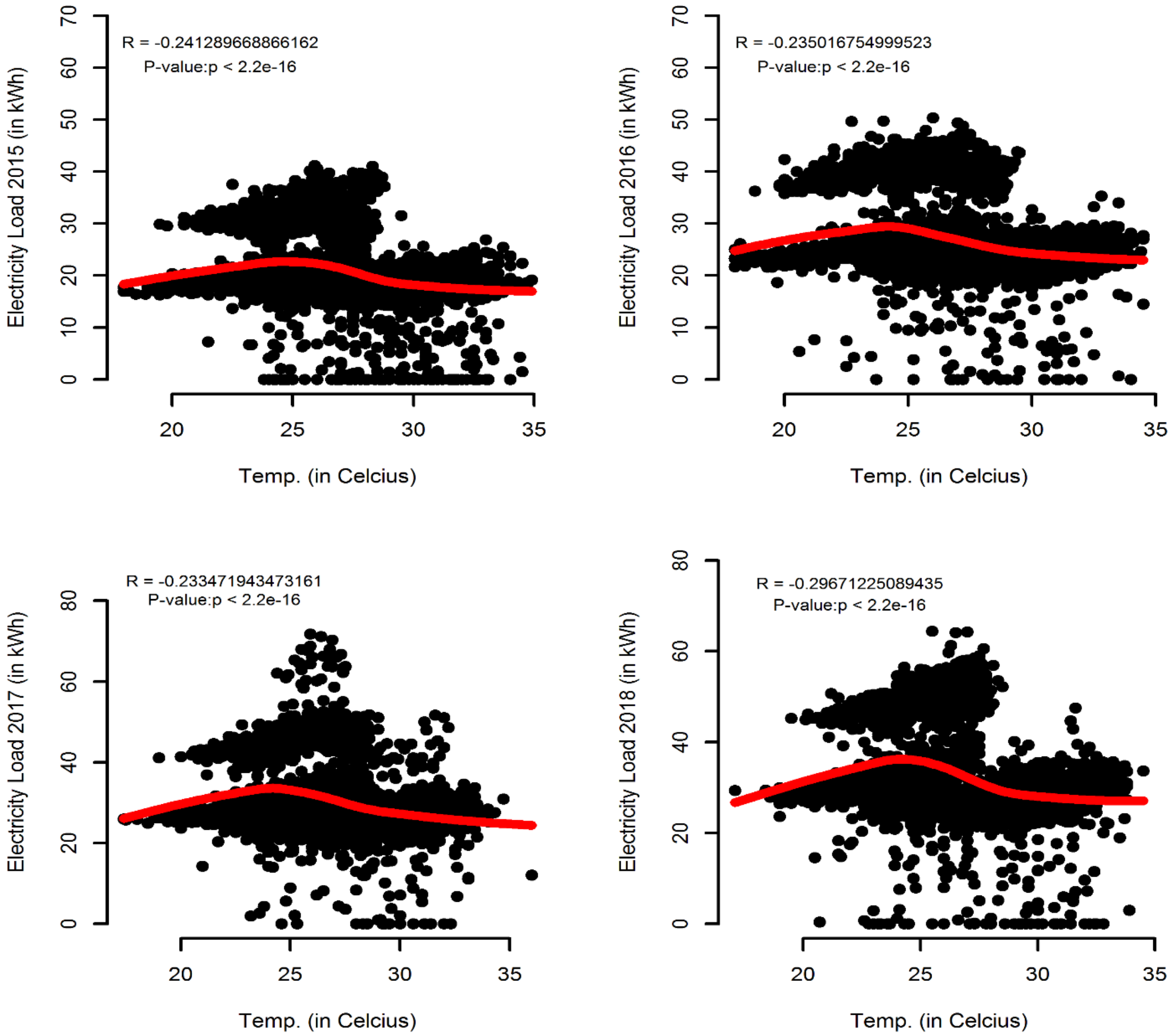

Fig. 1. Correlation analysis results for load versus temperature (2015-2018)

The relationship between electricity variation and temperature can also be deduced from the average load and average temperature graphs for all four years. The data trend for electricity consumption from 2015 to 2018 seems to be nonstationary (as seen in Fig. 5) while that of air temperature observed to be stationary as seen in Fig. 7.

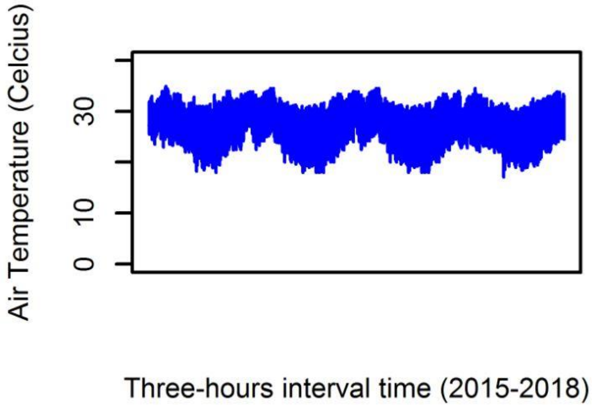

Fig. 7. Temperature variation (2015-2018)

Existing research works have identified the existence of a positive correlation between temperature and electricity consumption especially when it comes to commercial or office buildings. Research studies indicate that the direction of causality between electricity usage and the temperature is influenced by whether the study area is concerned with commercial or residential buildings. Furthermore, the literature report that the degree of causality is also influenced by the climatic condition of a particular area. The results of this study show that there is no correlation between electricity consumption and temperature in residential buildings as far as Tanzania's context is concerned. Results in this study may, therefore, pave a way to load analyst and utility companies to design robust load forecasting model.

4.3. Electricity Consumption versus Daytime

From the plotted graph in Fig. 8, the electricity consumption trend forms a common pattern for all four years. The graph plot indicates that there is a definite association between daytime and power usage in residential buildings based on the 2015 to 2018 data. The notable trend is observed in the mid-night, dawn, morning, afternoon, evening, and night hours as seen in the graph plot.

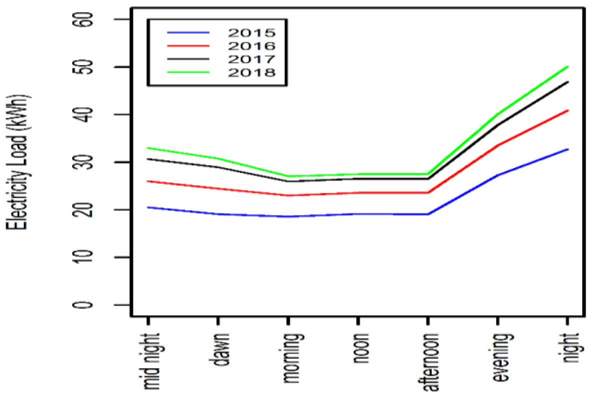

Fig. 8. The daily load consumption at Mikocheni, Dar es salaam from 2015-2018

As presented in Fig. 9, it can also be seen that there is an abrupt power demand increase at night for all four years. This may be due to dwellers being back to their homes and therefore followed by sudden switch-on of the electric appliances such as air conditioners, cookers, washing machines, and fans. Minimum consumption is observed at the morning and daytime hours wherein most of the dwellers are out of their homes for work and load shifts to the workplaces.

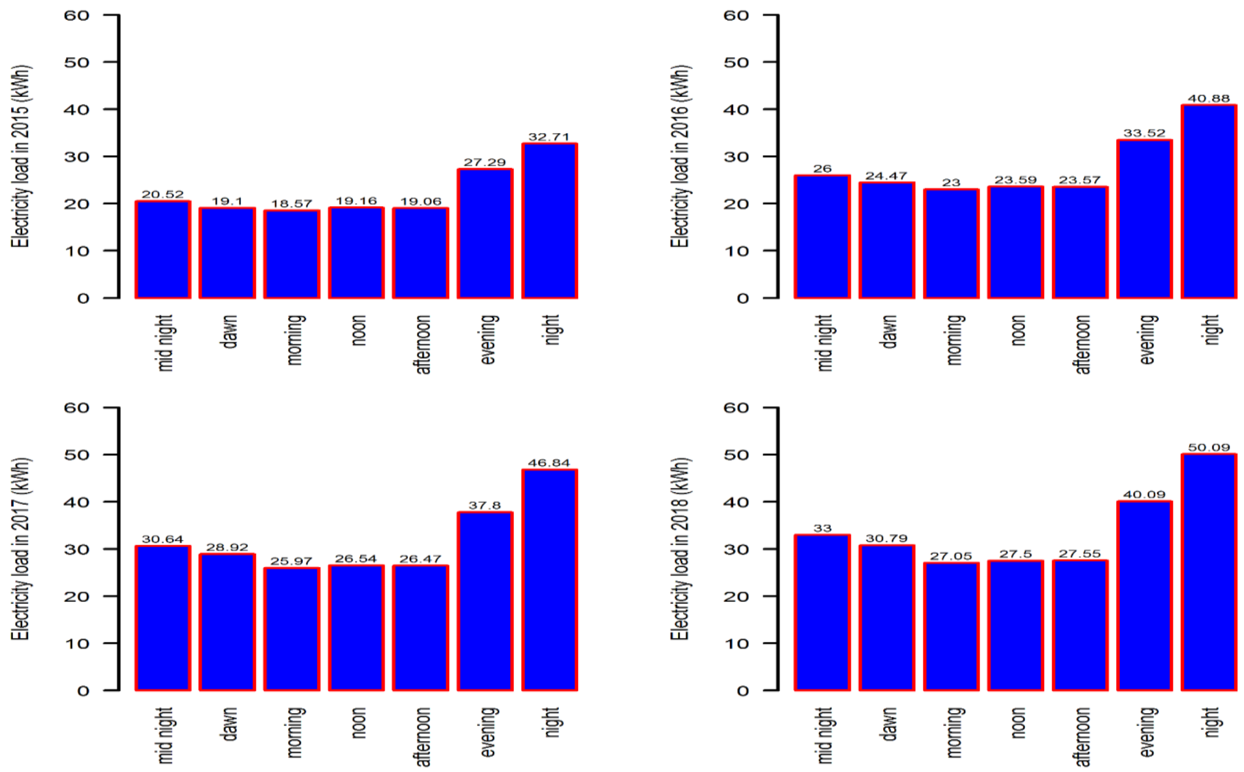

Fig. 9. Bar charts presenting peak load with regard to daytime from 2015-2018

Having found the existing association between daytime and electricity consumption in Tanzania paves a way to achieving the following; firstly, the utility company may effectively plan for load shifting, electricity pricing, and maintenance schedule. Secondly, the designers of electricity usage monitoring systems, particularly in Tanzania, may use findings from this study to model an efficient forecasting mechanism with consideration of the daytime.

5. Conclusion and Recommendations

In this work, the effect of daytime and air temperature on electricity consumption in Tanzania's residential buildings has been investigated using both correlation and visual analysis tools, based on the four-years data (2015-2018). Our empirical results indicate the existence of a non-linear relationship between temperature and electricity usage. Observations from graph plots indicate that daytime has a great significant influence on power consumption in Tanzania. Further interpretation from graph plots indicates high electricity usage at night and falling along midnight and dawn. The empirical results lead us to the conclusion that electricity consumption has no linear relationship with air temperature when it comes to residential areas in Tanzania. Unlike the findings in other research works that claim presence of weak relationship, this paper reports the absence of correlation between temperature and electricity consumption. Furthermore, the outstanding observation in this study is on the size of correlation coefficient which is -0.25 which indicating that there is no correlation between temperature and electricity usage.

The findings in this research work can help a utility company to guarantee efficient operations such as service disbursement, load management, and power usage monitoring. Furthermore, having known the impact of daytime and temperature on power demand, paves a way for future research works that would attempt to propose load management models such as forecasting, electricity pricing, load shifting, load balancing, and maintenance schedule. This work is limited to a sample load and air temperature data from one small residential area in Dares Salaam that would constrain the generalization for the entire country, since regions may experience varying social-economic and meteorological grounds. Future research should consider various study areas of the country as well as touch both residential and commercial buildings.

Acknowledgments

The authors would firstly, like to extend their sincere gratitude to Tanzania Electricity Supply Company (TANESCO) for immense hand especially concerning workshops, and field visits. Secondly, intimate appreciation to Sida for masterminding smart grid researches in Tanzania. lastly, it will be amess not to mention an entire iGrid research group at the University of Dar es Salaam for their keen support.

References

[1] C. J. Mtokambali and Z. Jun, "Smart Grid Technologies and Future Motivators Influencing Change in the Electricity Sector in Tanzania," *Int. J. Sci. Res.*, vol. 3, no. 2, pp. 39–43, 2014.

[2] M. U. Fahad and N. Arbab, “Factor Affecting Short Term Load Forecasting,” *J. Clean Energy Technol.*, vol. 2, no. 4, pp. 305–309, 2014.

[3] M. Li, D. Allinson, and M. He, “Energy & Buildings Seasonal variation in household electricity demand : A comparison of monitored and synthetic daily load profiles,” vol. 179, pp. 292–300, 2018.

[4] M. S. Gul and S. Patidar, “Understanding the energy consumption and occupancy of a multi-purpose academic building,” *Energy Build.*, vol. 87, pp. 155–165, 2015.

[5] A. Almuhtady, A. Alshwawra, M. Alfaouri, W. Al-kouz, and I. Al-hinti, “Investigation of the trends of electricity demands in Jordan and its susceptibility to the ambient air temperature towards sustainable electricity generation,” *Energy. Sustain. Soc.*, vol. 1, pp. 1–18, 2019.

[6] S. Firth, K. Lomas, A. Wright, and R. Wall, “Identifying trends in the use of domestic appliances from household electricity consumption measurements,” *Energy Build.*, vol. 40, pp. 926–936, 2008.

[7] M. Bessec and J. Fouquau, “The non-linear link between electricity consumption and temperature in Europe : A threshold panel approach ☆,” *Energy Econ.*, vol. 30, no. January 2007, pp. 2705–2721, 2008.

[8] D. Chikobvu, “Modelling influence of temperature on daily peak electricity demand in South Africa,” *J. Energy South. Africa*, vol. 24, no. 4, pp. 63–70, 2013.

[9] S. Savi, M. Boji, Z. Djordjevi, and D. Nikoli, “The impact of the mean daily air temperature change on electricity consumption,” pp. 8–13, 2015.

[10] M. G. Fikru and L. Gautier, “The impact of weather variation on energy consumption in residential houses,” *Appl. Energy*, vol. 144, pp. 19–30, 2015.

[11] M. Lorena and M. Lochinvar, “A review of the development of Smart Grid technologies,” *Renew. Sustain. Energy Rev.*, vol. 59, pp. 710–725, 2016.

[12] S. C. Bhattacharyya and G. R. Timilsina, “Modelling energy demand of developing countries: Are the specific features adequately captured?,” *Energy Policy*, vol. 38, no. 4, pp. 1979–1990, 2010.

[13] L. K. Hotta, “Effect of Outliers on Forecasting Temporally Aggregated Flow Variables,” *Soc. Estad~stica e I~vestigacidn Oper.*, vol. 13, no. 2, pp. 371–402, 2004.

[14] S. Saab, E. Badr, and G. Nasr, “Univariate modeling and forecasting of energy consumption: The case of electricity in Lebanon,” *Energy*, vol. 26, no. 1, pp. 1–14, 2001.

[15] D. De Termodina, U. De Vale, U. De Vale, D. De Termodina, and U. De Vale, “Daily Air Temperature and Electricity Load in Spain,” *Am. Meteorol. Soc.*, vol. 40, pp. 1413–1421, 2001.

[16] U. Habib, G. Zucker, M. Blochle, F. Judex, and J. Haase, “Outliers detection method using clustering in buildings data,” *IECON 2015 - 41st Annu. Conf. IEEE Ind. Electron. Soc.*, no. 838688, pp. 694–700, 2015.

[17] R. Kumari, “Anomaly Detection in Network Traffic using K- mean clustering,” in *International Conference on Recent Advances in Information Technology*, 2016, pp. 4799–8579.

[18] N. M. Razali and B. W. Yap, “Power Comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Anderson-Darling tests,” no. November, 2014.

[19] A. S. S. Shapiro and M. B. Wilk, “An Analysis of Variance Test for Normality (Complete Samples) Published by : Biometrika Trust Stable URL : http://www.jstor.org/stable/2333709,” *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[20] V. Bewick, L. Cheek, and J. Ball, “Statistics review 7: Correlation and regression,” *Crit. Care*, vol. 7, no. 6, pp. 451–459, 2003.

[21] M. Puth and M. Neuh, “Effective use of Spearman ’ s and Kendall ’ s correlation coef fi cients for association between two measured traits,” vol. 102, pp. 3–10, 2015.

[22] L. Hernández, C. Baladrón, J. M. Aguiar, B. Carro, A. Sánchez-Esguevillas, and J. Lloret, “Artificial neural networks for short-term load forecasting in microgrids environment,” *Energy*, vol. 75, pp. 252–264, 2014.

[23] F. Egelioglu, A. A. Mohamad, and H. Guven, “Economic variables and electricity consumption in Northern Cyprus,” *Energy*, vol. 26, no. 4, pp. 355–362, 2001.

[24] P. Holtedahl and F. L. Joutz, “Residential electricity demand in Taiwan,” *Energy Econ.*, vol. 26, no. 2, pp. 201–224, 2004.

[25] A. Z. Al-Garni, Y. N. Al-Nassar, S. M. Zubair, and A. Al-Shehri, “Model for electric energy consumption in Eastern Saudi Arabia,” *Energy Sources*, vol. 19, no. 4, pp. 325–334, 1997.

[26] S. Abosedra, A. Dah, and S. Ghosh, “Electricity consumption and economic growth, the case of Lebanon,” *Appl. Energy*, vol. 86, no. 4, pp. 429–432, 2009.

[27] H. T. Pao, “Forecast of electricity consumption and economic growth in Taiwan by state space modeling,” *Energy*, vol. 34, no. 11, pp. 1779–1791, 2009.

[28] P. K. Narayan and R. Smyth, “Multivariate granger causality between electricity consumption, exports

and GDP: Evidence from a panel of Middle Eastern countries," *Energy Policy*, vol. 37, no. 1, pp. 229–236, 2009.

[29] S. S. Adebola, "Electricity Consumption and Economic Growth: Trivariate investigation in Botswana with Capital Formation," *Int. J. Energy Econ. Policy*, vol. 1, no. 2, pp. 32–46, 2011.

[30] A. A. Aziz, N. H. Nik Mustapha, and R. Ismail, "Factors affecting energy demand in developing countries: A dynamic panel analysis," *Int. J. Energy Econ. Policy*, vol. 3, no. SPECIAL ISSUE, pp. 1–6, 2013.

[31] S. A. Solarin, "Multivariate causality test of electricity consumption, capital formation, export, urbanisation and economic growth for Togo," *Energy Stud. Rev.*, vol. 21, no. 1, pp. 109–132, 2014.

[32] P. K. Adom and W. Bekoe, "Modelling electricity demand in Ghana revisited: The role of policy regime changes," *Energy Policy*, vol. 61, pp. 42–50, 2013.

[33] M. A. Momani, "Factors Affecting Electricity Demand in Jordan," *Energy Power Eng.*, vol. 05, no. 01, pp. 50–58, 2013.

[34] F. Karanfil and Y. Li, "Electricity consumption and economic growth: Exploring panel-specific differences," *Energy Policy*, vol. 82, no. 1, pp. 264–277, 2015.

[35] G. A. Mabea, "Modelling Residential Electricity Demand for Kenya," *J. Econ. Sustain. Dev.*, vol. 5, no. 4, pp. 145–153, 2014.

[36] R. I. Thamae, L. Z. Thamae, and T. M. Thamae, "Dynamics of Electricity Demand in Lesotho: A Kalman Filter Approach," *Stud. Bus. Econ.*, vol. 10, no. 1, pp. 130–139, 2015.

[37] Y. Keho, "What drives energy consumption in developing countries? The experience of selected African countries," *Energy Policy*, vol. 91, no. December 2015, pp. 233–246, 2016.

[38] J. Mawejje and D. N. Mawejje, "Electricity consumption and sectoral output in Uganda : an empirical investigation," *J. Econ. Struct.*, vol. 2040, 2016.

[39] H. B. M. Albiman, S. Najat, "The relationship between energy consumption, CO2 emissions and economic growth in Tanzania," *Int. J. Energy Sect. Manag.*, vol. Vol.9, no. Issue: 3, p. pp.361-375, 2015.

[40] N. M. Odhiambo, "Energy consumption and economic growth nexus in Tanzania: An ARDL bounds testing approach," *Energy Policy*, vol. 37, no. 2, pp. 617–622, 2009.

[41] A. Z. Al-Garni, Y. N. Al-Nassar, S. M. Zubair, and A. Al-Shehri, "Model for electric energy consumption in Eastern Saudi Arabia," *Energy Sources*, vol. 19, no. 4, pp. 325–334, 1997.

[42] Y. Chakhchoukh, P. Panciatici, P. Bondon, and L. Mili, "Robust short-term load forecasting using projection statistics," *CAMSAP 2009 - 2009 3rd IEEE Int. Work. Comput. Adv. Multi-Sensor Adapt. Process.*, pp. 45–48, 2009.

[43] D. K. Chaturvedi, S. A. Premdayal, and A. Chandiok, "Short-Term Load Forecasting Using Soft Computing Techniques," *Int. J. Commun. Netw. Syst. Sci.*, vol. 03, no. 03, pp. 273–279, 2010.

[44] K. Kandananond, "Forecasting electricity demand in Thailand with an artificial neural network approach," *Energies*, vol. 4, no. 8, pp. 1246–1257, 2011.

[45] C. Sotomane, L. Asker, and V. Massingue, "ICT for automated forecasting of electrical power consumption: A case study in maputo," *2011 IST-Africa Conf. Proceedings, IST 2011*, no. January, 2011.

[46] C. Sigauke and D. Chikobvu, "Prediction of daily peak electricity demand in South Africa using volatility forecasting models," *Energy Econ.*, vol. 33, no. 5, pp. 882–888, 2011.

[47] C. A. Amlabu, J. U. Agber, C. O. Onah, and S. Y. Mohammed, "Electric Load Forecasting : A Case Study of the Nigerian Power Sector," *Int. J. Electr. Power Energy Syst.*, vol. 2, no. 10, pp. 23–27, 2013.

[48] C. A.Moturi and F. K. Kioko, "Use of Artificial Neural Networks for Short-Term Electricity Load Forecasting of Kenya National Grid Power System," *Int. J. Comput. Appl.*, vol. 63, no. 2, pp. 25–30, 2013.

[49] S. M. Awan, M. Aslam, Z. A. Khan, and H. Saeed, "Machine Learningbased Generic Load Forecastingmodel for Noisy Data: Lesco Case Study With Weather Influence.," *Pak. J. Sci.*, vol. 66, no. 2, pp. 121–129, 2014.

[50] S. Katara, A. Faisal, and G. M. Engmann, "A Time Series Analysis of Electricity Demand in Tamale , Ghana," vol. 4, no. 6, pp. 269–275, 2014.

[51] B. Kichonge, G. R. John, T. Tesha, and I. S.N. Mkilaha, "Prediction of Tanzanian Energy Demand using Support Vector Machine for Regression (SVR)," *Int. J. Comput. Appl.*, vol. 109, no. 3, pp. 34–39, 2015.

[52] A. Hussain, M. Rahman, and J. A. Memon, "Forecasting electricity consumption in Pakistan: The way forward," *Energy Policy*, vol. 90, pp. 73–80, 2016.

[53] G. Okoboi and J. Mawejje, "Electricity peak demand in Uganda: insights and foresight," *Energy. Sustain. Soc.*, vol. 6, no. 1, 2016.

[54] A. T. Ali, E. B. M. Tayeb, and Z. M. Shamseldin, "Short Term Electrical Load Forecasting Using Fuzzy Logic," *Int. J. Adv. Eng. Technol.*, vol. 03, no. 11, pp. 131–138, 2016.

[55] S. Karthika, V. Margaret, and K. Balaraman, “Hybrid short term load forecasting using ARIMA-SVM,” *2017 Innov. Power Adv. Comput. Technol. i-PACT 2017*, vol. 2017-Janua, pp. 1–7, 2017.

[56] A. K. Kür, “Forecasting Uganda ’ s Net Electricity Consumption Using a Hybrid PSO-ABC Algorithm,” 2018.

[57] A. A. Salami, A. S. Akoda Ajavon, K. A. Dotche, and K.-S. Bedja, “Electrical Load Forecasting Using Artificial Neural Network: The Case Study of the Grid Inter-Connected Network of Benin Electricity Community (CEB),” *Am. J. Eng. Appl. Sci.*, vol. 11, no. 2, pp. 471–481, 2018.

Appendix

Table 2. Correlation analysis results from programming in R

Spearman's rank correlation rho (2015)

```
data:  x and y
S = 5369677157, p-value < 2.2e-16
alternative hypothesis: true rho is not
equal to 0
sample estimates:
       rho
-0.241267123
```

Spearman's rank correlation rho (2016)

```
data:  x and y
S = 5118309320, p-value < 2.2e-16
alternative hypothesis: true rho is
not equal to 0
sample estimates:
       rho
-0.235034719
```

Spearman's rank correlation rho (2017)

```
data:  x and y
S = 5.167e+09, p-value < 2.2e-16
alternative hypothesis: true rho is not
equal to 0
sample estimates:
       rho
-0.23340168
```

Spearman's rank correlation rho (2018)

```
data:  x and y
S = 4311884526, p-value < 2.2e-16
alternative hypothesis: true rho is
not equal to 0
sample estimates:
       rho
-0.296712897
```

Table 3. Determinants of short and long-run load consumption from 20 developing countries

SN.	Author & Year	Country	Factors Investigated	Significant Drivers
1	[23]	N. Cyprus	Economic variables	Number of customers, price of electricity
2	[24]	Taiwan	Household disposable income, population growth, the price of electricity and the degree of urbanization	income elasticity in long-run; Cooling degree-day in short-run
3	[25]	Saudi Arabia	Weather, global solar radiation, population, and gross domestic product per capita	Temperature
4	[26]	Lebanon	Economic growth	NONE
5	[27]	Taiwan	Economic growth	Real GDP
6	[28]	Middle Eastern countries	Exports and gross domestic product(GDP)	Exports and gross domestic product
7	[29]	Botswana	GDP	GDP
8	[30]	16 countries	Income, price, economic structure, and CO2 emission	Income and price
9	[31]	Angola	Economic growth, urbanization	Urbanization
10	[32]	Ghana	Not clearly justified	Industry efficiency, industry value added, and real per capita GDP
11	[33]	Jordan	Not clearly justified	Demographic, technological, environmental and national energy pricing factors
12	[8]	South Africa	Not clearly justified	Temperature
13	[34]	13 countries	Economic activities	GDP
14	[2]	Pakistan	Behavior of the consumer load, total losses in transmission lines transmission lines.	Time factor, weather, economy and random disturbances
15	[35]	Kenya	Real disposable income and residential electricity prices	Income elasticity
16	[36]	Lesotho	Price and income elasticities	Economic growth
17	[37]	12 Sub-Saharan African countries	Real GDP per capita, industrial output, imports, foreign direct investment, credit to private sector, urbanization population	Economic growth, industrial output, population
18	[38]	Uganda	Sectoral output growth	GDP
19	[39]	Tanzania	CO2 emissions and economic growth	Significant impact
20	[40]	Tanzania	Economic growth	Economic growth

Table 4. Summary of analytical results from 19 applied load forecasting models in developing countries

SN.	Article	Country	Method Used	Handling Outliers	Data Cleansing Method
1	[41]	Saudi Arabia	Regression Analysis	X	N/A
2	[14]	Lebanon	Autoregressive (AR1)	√	High-pass filter model
3	[23]	Northern Cyprus	multiple regression analysis	X	N/A
4	[42]	India	AutoRegressive Model	√	Mahalanobis Distance, AutoRegressive, Gaussian distribution
5	[43]	India	GNN model	√	Wavelet Transform method
6	[44]	Thailand	ARIMA, ANN and Multiple Linear Regression (MLR)	X	N/A
7	[45]	Mozambique	similar-day method	X	N/A
8	[46]	South Africa	regression-SARIMA	X	N/A
9	[47]	Nigeria	least squares technique	X	N/A
10	[48]	Kenya	Artificial Neural Network		N/A
11	[49]	Pakistan	hybrid ANN - SVM	√	Box-plot and Weighted Moving Average
12	[50]	Ghana	ARIMA	X	N/A
13	[51]	Tanzania	Support Vector Machine for Regression (SVR)	X	N/A
14	[52]	Pakistan	Holt- Winter and Autoregressive Integrated Moving Average (ARIMA)	X	N/A
15	[53]	Uganda	double exponential forecasting	X	N/A
16	[54]	SUDAN	fuzzy logic approach	X	N/A
17	[55]	India	hybrid ARIMA-SVM	√	Percentage Error (PE) and Deviation method
18	[56]	Uganda	Particle Swarm Optimization (PSO) and Artificial Bee Colony (ABC) algorithm	X	N/A
19	[57]	Benin	ANN-MLP, ANN-RBF	X	N/A