JJEE
Volume 4, Number 4, 2018
Pages 186-194
Jordan Journal of Electrical Engineering
ISSN (Print): 2409-9600, ISSN (Online): 2409-9619

Accurate Probability Distribution Model Determination and Forecasting of Peak Load Demand in Nigeria

Ignatius K. Okakwu[1a], Akintunde S. Alayande[2], Patience E. Orukpe[1]

[1]Department of Electrical/Electronics Engineering, University of Benin, Benin, Nigeria
[2]Department of Electrical/Electronics Engineering, University of Lagos, Akoka-Yaba, Nigeria
[a]e-mail: igokakwu@yahoo.com

Received: October 2, 2018 *Accepted: November 8, 2018*

Abstract— This study aims at identifying the best-fit probability distribution and forecasting of the peak load demand in Nigeria. The data used was obtained from the National Control Centre (NCC), Oshogbo, Nigeria for a period of twenty years (1998-2017). Five different probability distribution functions and two forecasting models were used. The probability distribution functions explored include Normal, Log-normal, Gamma, Weibull and Logistic distribution from which the best was determined using two different goodness-of-fit. The two goodness-of-fit used are Akaike Information Criterion (AIC) and Schwartz Bayesian Criterion (SBC) while the two forecasting models include Auto Regression (AR) and Exponential Smoothing (ES). The best model is expected to have the lowest value for AIC, SBC, Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Theil Inequality Coefficient (TIC). The model that satisfies tests adequately was selected as the best fit. Results showed that the Log-normal distribution presents the best fitted distribution with AIC value of 327.5168 and SBC value of 329.5082 followed by Normal distribution with AIC of 327.5987 and SBC of 329.5902; Weibull distribution with AIC of 327.8540 and SBC of 329.8454; Gamma distribution with AIC of 328.0087 and SBC of 330.0002; and Logistic distribution of AIC of 328.3212 and SBC of 330.3127 respectively. The AR gave the best result among the two models with MAPE value of 0.21, MAE value of 12.55, RMSE of 173.40 and TIC value of 0.022. The results from this study will be very useful for decision makers, system operators, load forecasters, scheduling of electricity and potential investors in the power industry in Nigeria.

Keywords— Best of fit, cumulative distribution function, forecasting, peak load, probability distribution function.

I. INTRODUCTION

The electric power infrastructure in Nigeria has become overstretched due to increase in population that has led to increase in energy demands [1]. The knowledge on energy demands of any nation is a necessity before embarking on any major project like building more generating stations, constructing new transmission/distribution lines, installing new transformers, etc. [2]. Unfortunately, in Nigeria, the historical data required to guide one in making decisions about future energy demands are not readily available [3]. There is, therefore, a need to develop an alternative technique, which can predict the load demands in a more efficient manner [4]. Statistical techniques are essential for developing mathematical models to generate synthetic historical records, to forecast energy demand events, to depict intrinsic stochastic characteristics of load demand variables and to fill missing and extending records [5]. Statistical methods have been applied to electricity generation [6], electric load demands and prediction in Akwa Ibom state [7], electricity demands and prediction in Rivers state [8] and distribution system loads [9].

In the past, several probability models have been developed to describe the distribution of events such as peak/average load demand data, wind speed data, solar radiation data, etc. in a country [10]. However, the choice of a suitable model is still a major problem in Engineering and Statistics since there are no general rules as to which distribution(s) to be adopted. Hence, it is necessary to select different probability distribution and forecasting models in order to

Corresponding author's e-mail: igokakwu@yahoo.com

determine which is more suitable and appropriate to provide accurate estimation [11]. In [12], the authors used probability distribution analysis to predict reservoir inflow in hydropower dams in Nigeria. The hydropower dams include Shiroro, Kainji and Jebba dams. The probability distribution models used include Gumbel, Normal, Log-Pearson type III and Log-normal. The choice of an appropriate probability distribution model was based on the goodness of fit test. An investigation to select a proper probability distribution model to describe rainfall distribution in Ibadan metropolis, Nigeria, over a period of 30 years was presented in [13]. The Gamma, Exponential, Normal and Poisson distributions were compared in order to identify the optimal model. The models were evaluated based on Chi-square and Kolmogorov-Smirnov tests. In a similar study, the authors in [14] carried out a study to determine the best-fit probability distributions for peak daily rainfall of selected cities in Nigeria. Different statistical models such as Gumbel, Log-Gumbel, Normal, Log-Normal, Pearson, Log-Pearson distribution were used. The goodness-of-fit test used includes Chi-square, Fisher's test, Correlation coefficient and Coefficient of determination. The model that satisfies the tests correctly was selected as the best fit model; and the results showed that the Log-Pearson type III performed better than the rest. Wind speed data for Ibadan were analyzed statistically using Weibull probability distribution function [15]. The daily, monthly and yearly Weibull probability distribution parameters, mean wind speeds and available power for the location were also determined. The authors in [16] investigated statistically the compressive strength of concrete for a period of 5 years. Among the three distribution models (Shifted Lognormal, Gumbel and Normal distribution), shifted lognormal was found to be the best. The application of Normal, Log-Normal and Log-Pearson type III probability distributions to select the best flood frequency distribution that best fits the annual maximum flood flows of Ona River under Ogun-Oshun river basin development in Nigeria was investigated in [17]. Daily peak electricity demand in South Africa was modeled using regression-SARIMA in [18]. The Normal, log-Normal and Weibull probability distributions were considered, while AIC, Log Likelihood and Estimated Parameters were used for comparison. It turned out that the Normal distribution was the best based on the fact that it had the least AIC. In this study, the best probability distribution function that best models the peak load demand in Nigeria is determined using statistical goodness of fit and predicting the peak load demand using AR and ES models.

II. MATERIALS AND METHODS

A) Data Collection

The Peak load demand data used in this study were obtained from the National Control Centre (NCC), Oshogbo, Nigeria for a period of 1998-2017 as shown in Table 1. The NCC is responsible for the remote control and monitoring of power system operations in Nigeria.

TABLE 1
PEAK LOAD DEMAND IN NIGERIA (MW)

Load Demand			
Year	Peak load	Year	Peak load
1998	2448	2008	3682
1999	2458	2009	3600
2000	2499	2010	3804
2001	2934	2011	4089
2002	3223	2012	4054
2003	3479	2013	4458
2004	3428	2014	4487
2005	3775	2015	4811
2006	3682	2016	5075
2007	3600	2017	5222

B) Methodology

The choice of best-fit is important in selecting a probability distribution function model for any application. It is practically impossible to select all statistical distribution functions because they are too numerous to be accommodated in a paper. In this study, the considered distribution models namely Logistic, Weibull, Normal, Log-Normal and Gamma distributions are commonly used.

B.1. Logistic distribution

Logistic distribution is a continuous probability density function that is symmetric. It is similar in shape to that of normal distribution. The distribution is characterized by two main parameters: location μ and scale σ.
The probability density function (PDF) is given by [19]:

$$f(x) = \frac{e^{-(x-\mu)/\sigma}}{\sigma\left(1+e^{-(x-\mu)/\sigma}\right)^2} \tag{1}$$

The cumulative distribution function (CDF) is given by [19]:

$$F(x) = \frac{1}{1+e^{-\left((x-\mu)/\sigma\right)}} \tag{2}$$

where μ is the mean; σ is the variance; and x is the variable.

B.2. Weibull distribution

The Weibull PDF and its CDF are given by (3) and (4) respectively as [20]:

$$f(x) = \frac{\beta}{\eta}\left[\frac{x}{\eta}\right]^{\beta-1} \times e^{-\left(\frac{x}{\eta}\right)^{\beta}} \tag{3}$$

$$F(x) = 1 - e^{-\left(\frac{x}{\eta}\right)^{\beta}} \tag{4}$$

where β is the Weibull shape; and η is the scale parameter.

B.3. Normal distribution

The normal PDF and CDF are given by (5) and (6) respectively as [19]:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\alpha-\mu}{\sigma}\right)^2} \tag{5}$$

$$F(x) = \frac{1}{2}\left(1 + erf\left(\frac{\alpha-\mu}{\sqrt{2\sigma}}\right)\right) \tag{6}$$

where α is the shape parameter.

B.4. Log-normal distribution

The Log-Normal PDF and CDF are given by (7) and (8) respectively as [19]:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{ln\,x-\mu}{\sigma}\right)^2} \tag{7}$$

$$F(x) = \int_0^x f(x)\,dx = \frac{1}{2} + \frac{1}{2} erf\left(\frac{ln\,x-\mu}{\sigma\sqrt{2}}\right) \tag{8}$$

where *erf* is the error function.

B.5. Gamma distribution

Gamma distribution is an extension of the exponential distribution whose PDF and CDF are expressed by (9) and (10) respectively as [21]:

$$f(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} e^{-\frac{x}{\beta}} \tag{9}$$

$$F(x) = \frac{1}{b^{a}\Gamma(a)} \int_0^x t^{a-1} e^{-\frac{t}{b}} dt \tag{10}$$

where Γ is the gamma function.

B.6. Exponential smoothing model

A simple ES can be expressed as [22]:

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t \tag{11}$$

where $\hat{Y}_{t+1}$ is the forecast value for period *t*+1 made at time *t*; Y_t is the actual value in period *t*; and α is the smoothing constant (0< α <1). In general, ES can be written as in (12):

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1-\alpha)\hat{Y}_t + \alpha(1-\alpha)^2\hat{Y}_{t-1} + \cdots \alpha(1-\alpha)^{t-1}Y_1 + (1-\alpha)^t Y_0 \tag{12}$$

In a compact form, (12) is reduced to:

$$\hat{Y}_{t+1} = \alpha \sum_{k=0}^{t-1}(1-\alpha)^k Y_{t-k} + (1-\alpha)^t Y_0 \tag{13}$$

B.7. Autoregressive model

The AR model for order 1 is expressed as [23]:

$$Y_t = \beta_o + \beta_1 Y_{t-1} + e_i \tag{14}$$

The estimated model is given in (15) and subsequently in (16). Substituting (15) into (16) gives (17):

$$\hat{Y}_t = \hat{\beta}_o + \hat{\beta}_1 Y_{t-1} \tag{15}$$

where

$$\hat{\beta}_o = \bar{Y}_t - \hat{\beta}_1 \bar{Y}_{t-1} \tag{16}$$

$$\hat{\beta}_1 = \frac{n\sum_{t=2}^{n} Y_t Y_{t-1} - (\sum_{t=2}^{n} Y_t)(\sum_{t=2}^{n} Y_{t-1})}{n\sum_{t=2}^{n} Y_{t-1}^2 - (\sum_{t=2}^{n} Y_{t-1})^2} \tag{17}$$

C) Goodness-of-Fit Criteria

The most common techniques for goodness-of-fit criterion are Akaike Information Criterion (AIC) and Schwartz Bayesian Criterion (SBC). Akaike information criteria are given below [24]:

$$AIC = -2\,ln(L) + 2K \tag{18}$$

$$SBC = -2\,ln(L) + K\,ln(n) \tag{19}$$

where L is the likelihood function; K is the number of parameters to be estimated; and n is the number of observations. The best model is expected to have the lowest value for AIC and SBC, respectively.

D) Performance Evaluation of Forecasting Models

Four measures were used to evaluate the performance of the models presented in this study. The smaller the error, the better the models are for forecasting. These models are [25]:

$$\text{Root Mean Square Error (RMSE)}= \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_t - y_t)^2}{n}} \tag{20}$$

$$\text{Mean Absolute Error (MAE)}= \sum_{t=1}^{n}\frac{|\hat{y}_t - y_t|}{n} \tag{21}$$

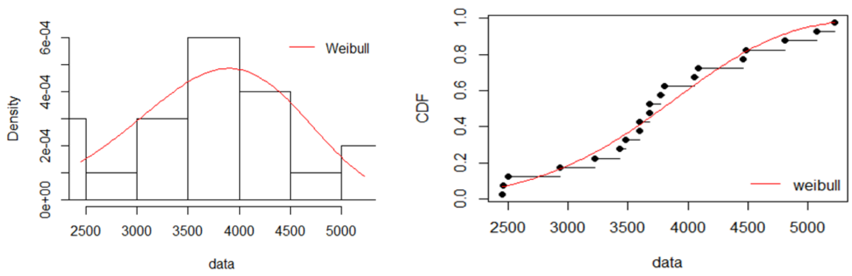
$$\text{Mean Absolute Percentage Error (MAPE)}= \sum_{t=1}^{n}\frac{\left|\frac{\hat{y}_t - y_t}{y_t}\right|}{n} \times 100\% \tag{22}$$

$$\text{Theil Inequality Coefficient (TIC)}= \frac{\sum_{t=1}^{n}\sqrt{\frac{(\hat{y}_t - y_t)^2}{n}}}{\sqrt{\frac{\sum_{t=1}^{n}\hat{y}_t^2}{n}}+\sqrt{\frac{\sum_{t=1}^{n}y_t^2}{n}}} \tag{23}$$

where $\hat{y}_t$ is the actual value; y_t is the forecasted value; n is the number of observations; and t is the period.

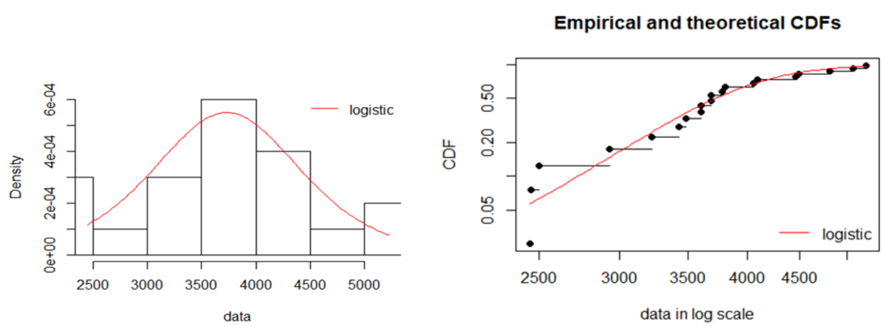
III. Results and Discussion

Figs. 1-5 show the statistical PDF and its corresponding CDF for the different models used in this study. Fig. 6 shows the actual and predicted peak load in Nigeria for the period of 1998-2017. Fig. 7 depicts the forecasted peak load for the period of 2018-2030.



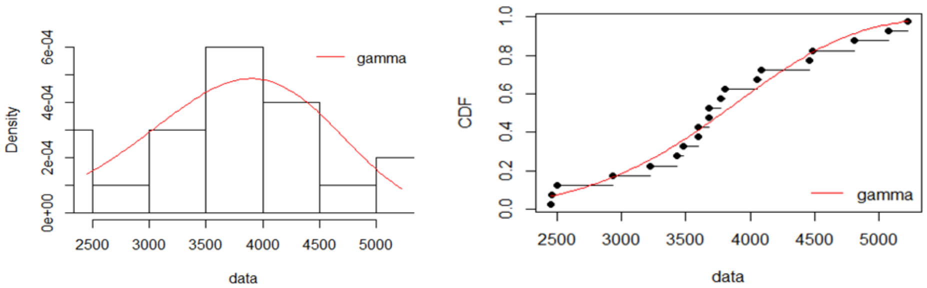
(a) (b)

Fig. 1. a) Weibull probability distribution function, b) Weibull cumulative distribution function



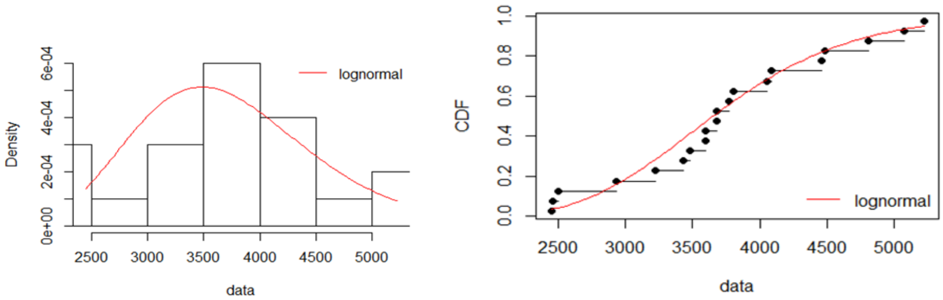
(a) (b)

Fig. 2. a) Logistic probability distribution function, b) Logistic cumulative distribution function



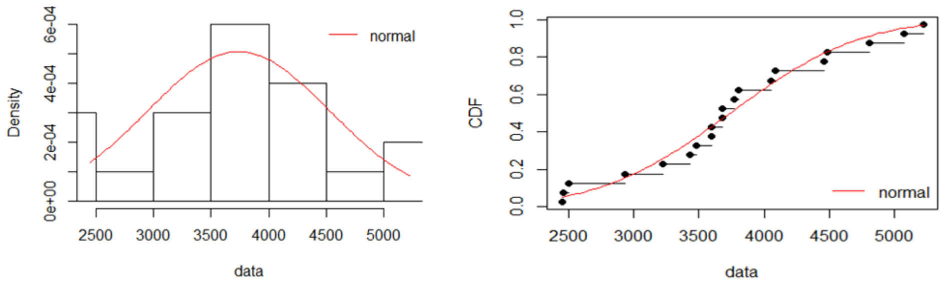
(a) (b)

Fig. 3. a) Gamma probability distribution function, b) Gamma cumulative distribution function



(a) (b)

Fig. 4. a) Lognormal probability distribution function, b) Lognormal cumulative distribution function



(a) (b)

Fig. 5. a) Normal probability distribution function, b) Normal cumulative distribution function

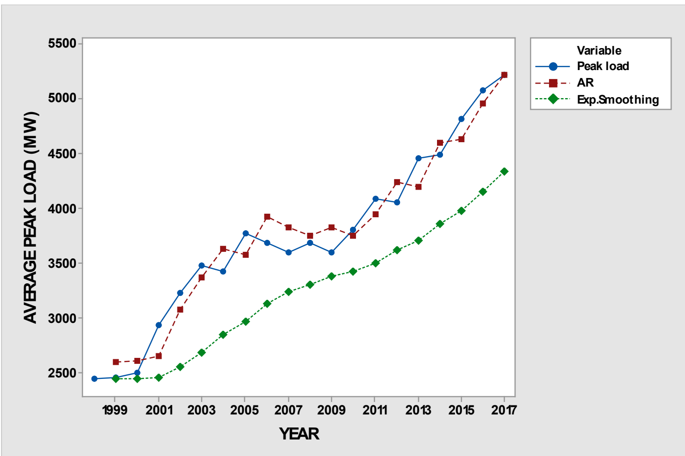


Fig. 6. Actual and predicted peak load in Nigeria between 1998 and 2017

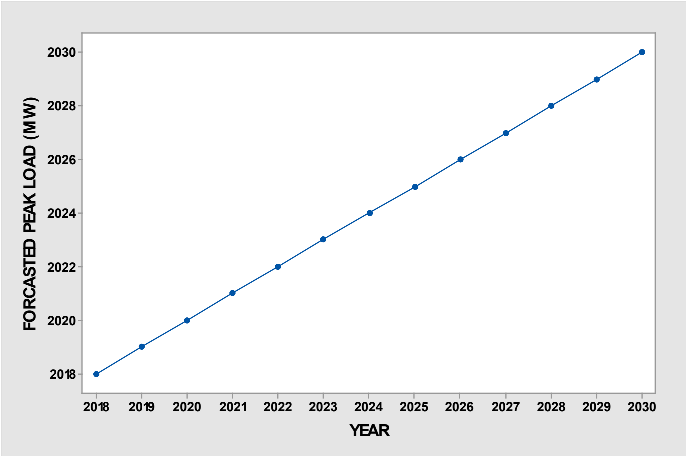


Fig. 7. Forecasted peak load in Nigeria

Table 2 shows the error values of the two models used to predict the peak load demand. It is apparent that both AR and ES models demonstrated promising results with AR having MAPE value of 0.21, MAE value of 12.55, RMSE value of 173.40 and TIC value of 0.022 as against ES with MAPE value 0.38, MAE value of 23.30, RMSE value of 602.38 and TIC value of 0.078. That AR has the least errors implies that AR model performed better than the ES. Hence it is adopted for load forecasting as shown in Fig. 7.

TABLE 2
FORECASTING MODEL PERFORMANCE FOR PEAK LOAD

Model	MAPE	MAE	RMSE	TIC
AR	0.20906049	12.54971	173.4002	0.022349
Exp. Smoothening	0.379986597	23.30267	602.3789	0.077639

It can be observed from Figs. 1-5 (a-b) that all the probability distribution models fit the data used as they all follow the same shape. Though several statistical distributions were tested, only the five that best approximate the peak load data are presented in Table 3 due to space limitation.

TABLE 3
GOODNESS-OF-FIT CRITERIA

Statistical Distribution	Goodness-of-Fit Criteria		Order of Goodness of Fit
	AIC	SBC	
Gamma	328.0087	330.0002	4th
Log-Normal	327.5168	329.5082	1st
Logistic	328.3212	330.3127	5th
Normal	327.5987	329.5902	2nd
Weibull	327.8540	329.8454	3rd

It can be observed from Table 3 that the log-normal distribution presents the best statistical goodness-of-fit in the modelling of the peak load data with AIC of 327.5168 and SBC of 329.5082. Log-normal distribution was closely followed by Normal distribution with AIC of 327.5987 and SBC of 329.5902; Weibull distribution with AIC of 327.8540 and SBC of 329.8454; Gamma distribution with AIC of 328.0087 and SBC of 330.0002; and Logistic distribution of AIC of 328.3212 and SBC of 330.3127, respectively. That the above values of AIC and SBC for all the five models are very close to each other further shows that all the models fit the pattern of the peak load.

IV. CONCLUSIONS

The most suitable PDF function and load forecasting models for modelling peak load demands in Nigeria have been studied in this paper. The results obtained show that the best distribution function was Log-normal, followed by Normal, Weibull, Gamma and Logistic distribution respectively. The outcome of the choice of model was relied upon by two goodness-of-fit criteria, namely AIC and SBC. These goodness-of-fit test criteria have been identified as the best probability distribution that could provide accurate peak load demand estimations in Nigeria. Furthermore, the result for the forecasting models identifies AR model as the best for predicting peak load demands compared to ES model due to its relatively small error. The results of this study will be useful for policy makers, system operators, load forecasters, scheduling of electricity and investors who are interested in the power industry in Nigeria.

REFERENCES

[1] B. Onakoya and O. Onakoya, "Energy consumption and Nigerian economic growth: an empirical analysis," *European Scientific Journal*, vol. 9, no. 4, pp. 25-40, 2013.

[2] A. Momoh, S. Meliopoulos, and R. Saint, *Centralized and Distributed Generated Power Systems– A Comparison Approach*, Power Systems Engineering Research Center, 2012.

[3] S. Sambo, "Matching electricity supply with demand in Nigeria," *International Association for Energy Economics, Fourth Quarter*, pp. 32-36, 2008.

[4] K. Srivastava, S. Pandey, and D. Singh, "Short-term load forecasting methods: a review," *Proceedings of International Conference on Energy Trends in Electrical, Electronics and Sustainable Energy Systems*, pp. 130-138, 2016.

[5] A. Alam, K. Emura, C. Farnham, and J. Yuan, "Best-fit probability distributions and return periods for maximum monthly rainfall in Bangladesh," *Journal of Climate Science*, vol. 6, no. 9, pp. 1-16, 2018.

[6] E. Imo, C. Chukwu, and I. Abode, "Statistical analysis of electricity generation in Nigeria using multiple linear regression model and box-Jenkins autoregression model of order 1," *International Journal of Energy and Power Engineering*, vol. 6, no. 3, pp. 28-33, 2017.

[7] C. Idoniboyeobu and B. Ekanem, "Assessment of electric load demand and prediction of future load demand: a case study Akwa Ibom state of Nigeria," *Asian Journal of Scientific Research*, vol. 7, no. 4, pp. 525-536, 2014.

[8] A. Briggs and K. Ugorji, "Assessment of electricity demand and prediction model for the future: Rivers state," *European Journal of Mechanical Engineering Research*, vol. 4, no. 1, pp. 1-23, 2017.

[9] R. Singh, C. Pal, and A. Jabr, "Statistical representation of distribution system loads using Gaussian mixture model," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 29-37, 2010.

[10] I. Graabak and M. Korpas, "Variability characteristics of European wind and solar power resources—a review," Energies, vol. 9, no. 6, pp. 1-31, 2016.

[11] S. Singh and B. Masuku, "Sampling techniques and determination of sample size in applied statistics research: an overview," *International Journal of Economics, Commerce and Management*, vol. 2, no. 11, pp. 1-22, 2014.

[12] O. Olukanni and W. Salami, "Fitting probability density functions to reservoir inflow at hydropower dams in Nigeria," *Journal of Environment Hydrology*, vol. 16, no. 35, pp. 1-4, 2008.

[13] A. Oseni and F. Ayoola, "Fitting the statistical distribution for daily for daily rainfall in Ibadan, based on Chi-square and Kolmogorov-Smirnov goodness-of-fit tests," *European Journal of Business and Management*, vol. 4, no. 16, pp. 62-70, 2012.

[14] O. Olofintoye, A. Sule, and A. Salami, "Best-fit probability distribution model for peak daily rainfall of selected cities in Nigeria," *New York Science Journal*, vol. 2, no. 3, pp. 1-12, 2009.

[15] O. Rauff and F. Nymphas, "A statistical approach to estimate wind speed distribution in Ibadan, Nigeria," *Physical Science International Journal*, vol. 11, no. 2, pp. 1-14, 2016.

[16] S. Silvestri, G. Gasparini, T. Trombetti, and C. Ceccoli, "Statistical analysis towards the identification of accurate probability distribution models for the compressive strength of concrete," *Proceedings of the World Conference on Earthquake Engineering, Beijing*, pp. 1-8, 2008.

[17] A. Ewemoje and S. Ewemoje, "Best distribution and plotting positions of daily maximum flood estimation at Ona river in Ogun-Oshun river basin, Nigeria," *Agricultural Engineering International*, vol. 13, no. 3, pp. 1-13, 2011.

[18] D. Chikobvu and C. Sigauke, "Regression-SARIMA modeling of daily peak electricity demand in South Africa," *Journal of Energy in Sothern Africa*, vol. 23, no. 3, pp. 23-30, 2012.

[19] T. Ayodele, "Determination of probability distribution for modelling global solar radiation: case study of Ibadan, Nigeria," *International Journal of Applied Science and Engineering*, vol. 13, no. 3, pp. 233-245, 2015.

[20] D. Fadare, "A Statistical analysis of wind energy potential in Ibadan, Nigeria, based on Weibull distribution function," *The Pacific Journal of Science and Technology*, vol. 9, no. 1, pp. 110-119, 2008.

[21] M. Lawal, A. Adelakun, and K. Obisesan, "The use of gamma distribution to evaluate water pollutants in Asejire reservoir, Ibadan," *Foundation Journal of Natural and Applied Sciences*, vol. 2, no. 2, pp. 38-46, 2013.

[22] M. Jumba, C. Joel, and J. Mungatu, "Use of exponential smoothing technique in estimation of returns in a financial portfolio (a case of the Matatu public transport business in Kenya)," *American Journal of Theoretical and Applied Statistics*, vol. 4, no. 6, pp. 484-495, 2015.

[23] Z. Baharudin and N. Kamel, "Autoregressive method in short term load forecast," *Proceedings of IEEE International Conference on Power and Rnergy*, pp. 1603-1608, 2008.

[24] H. Acquali, "Comparison of akaike information criterion (AIC) and bayesian information criterion (BIC) in selection of an asymmetric price relationship," *Journal of Development of Agricultural Economics*, vol. 2, no. 1, pp. 1-6, 2010.

[25] X. Liu and J. Fang, "Long term load forecasting based on a time-variant ratio multi-objective optimization fuzzy time series model," *Mathematical Problems in Engineering*, vol. 2013, pp. 1-7.