International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 9 Sep. 2022, pp: 1428-1435 www.ijaem.net ISSN: 2395-5252

Evaluation of Combine Electricity load Demand Models Using Simulation

Victor Okolobah, PhD
Department of Statistics,
Federal Polytechnic, Bida

Date of Submission: 20-09-2022 Date of Acceptance: 30-09-2022

ABSTRACT
The common believed held is that combined models yield better forecast accuracy than stand-alone models. This paper set out to ascertain that claim by adopting the load data for Bida a city in North-Central zone of Nigeria. The Bida load data was reproduced by applying the rand function in MATLAB to generate 366 daily load demands for the year as contained in the original data-set. This was employed to fit a traditional ARIMA model to the simulated data set. To develop the combine EMD-ARIMA model, the simulated data set was first subjected to the EMD sifting algorithm which produced eight IMFs and a residue. An ARIMA model was developed for each of these IMFs which were then prediction and the results of these predictions were pooledd together to obtain the desired EMD-ARIMA. After the usual diagnostic checks were carried out on the two models we then proceed to forecast using both models. The results revealed that the simulated data set were not too different from the original ones and the forecast results by both models were adjudged as very good.The assertion that combine models yield better forecast accuracy than stand-alone models is correct as the ARIMA models produced an average forecast accuracy of 8.259% while that of the combined EMD-ARIMA model is 5.738%. This represents a 30.52% improvement.

Keywords: Forecast, Forecast Accuracy,EMD, IMF, Bida

I. INTRODUCTION

In the last two decades, interest in electricity load forecasting has grown tremendously. This has led to development of several forecasting models. These models could be univariate time series models see (Bracade, Caramia, Falco & Hong, 2019; Zhu, Yang, Mourshed, Guo, Zhou, Chang, Wei & Feng, 2019; Rendon-Sanchez & De Menezes, 2019). Researchers have equally delved into the field of artificial intelligence in developing these load models see (Butt, Hussain, Mahmood & Lone, 2021; Wang, Dou, Yu, Liu, Zhang &Xie, 2020; Hafeez, Alimgeer& Khan, 2020; Lee, Kim & Kim, 2020). Most recently interest has shifted to combine models in Load forecasting (Okolobah, 2022; Pham, Nguyen & Wu, 2021; Zhang, Xu, Cheng, Ding, Liu, Wei & Sun, 2021; Dong, Ma & Fu, 2021). Combine models are models that bring together the advantages of two or more stand-alone models (Okolobah, 2022). It is equally believed that these combine models yield better forecast accuracy than stand-alone models.

Combine models can be very expensive to develop and implement. This is making researchers seek other methods that are more cost efficient in the development of load forecast models. One of these methods that is beginning to gain wide acceptability is the use of simulation methods in load forecasting. Li (2020) and Qader&Qamber (2010) in separate works developed combine models using simulated data sets and the models can forecast load demand well.

Simulation is the imitation of the operation of a real-world process or system over time (Banks, 1999). Simulation is an indispensable problem-solving methodology for the solution of numerous real-world challenges. Whether the problem is an existing one or conceptualized one, both can be modelled using simulation (Banks, 1999). Simulation use is getting more popular dues to advances in software and this has made many businesses to embrace the use of simulation.

Adam reports that simulation is often used in model evaluation while Dielman (1986) and Lin &Granger (1994) used simulation to compare several time series models. In this paper, simulation is used to forecast the accuracy of six time series model already known and existing. This is aimed at ascertaining the claim that combine models yield better forecast accuracy than stand-alone in electricity load forecasting.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 9 Sep. 2022, pp: 1428-1435 www.ijaem.net ISSN: 2395-5252

The remaining part of this paper is structured as follows: Section 2 discusses the methodology of the paper, Section three presents the results while the work in this paper is concluded in Section four.

II. METHODOLOGY

This Section discusses the methodology adopted for the study extensively. Major components of the model such as empirical mode decomposition (EMD), auto-regressive integrated moving average (ARIMA) as well as Simulation were comprehensively discussed.

2.1 Empirical Mode Decomposition (EMD)

(Rato et al., 2008)defined empirical mode decomposition as a technique used in decomposing a given signal into basic signals known as intrinsic mode functions (IMFs) while (Dang et al., 2008) described EMD as a tool employed in Mathematics to analyze non-stationary and nonlinear stochastic waves and goes further to say that the EMD process splits a waveform into integral modes of oscillations called intrinsic mode functions (IMFs) which offers good definitions for instantaneous frequency, instantaneous energy, mean trends and oscillations around it. (Huang et al., 1998) proposed two distinct conditions for identifying any function as an IMF and these conditions are:

I. The number of extrema and the number of zero crossings must either be zero or differ by at most one

II. The mean value of the envelope defined by the local maxima and local minima is zero.

The EMD extraction process is referred to as sifting process (Huang et al., 1999); (Okolobah and ismail, 2013a) and (Okolobah and Ismail, 2013b)

2.2.1 The EMD Algorithm

The EMD technique for decomposing a non-linear and non-stationary signal into Intrinsic Mode Functions (IMFs) is summarized in the following five steps:

Step 1: Identify the maxima and minima of the load data, x (t).

Step 2: Construct the lower $e_{min}(t)$ and upper envelopes $e_{max}(t)$ by interpolating with cubic spline.

Step 3: Calculate the mean value by obtaining the average of the upper and lower envelopes

$$m(t) = \frac{[e_{min}(t) + e_{max}(t)]}{2} \quad (2.1)$$

Step 4: Subtract the mean from the original series

$$d(t) = x(t) - m(t) \quad (2.2)$$

Step 5: Check $d(t)$ properties

If $d(t)$ is an IMF, as stated above, then denote it as ith IMF and replace $x(t)$ with the residue $r(t) = x(t) - d(t)$. The ith IMF is often denoted as $C_i(t)$ and **i** is called its index (Zhang et al., 2008)

If it is not, replace $x(t)$ with $d(t)$.

The original data $x(t)$ can be represented as

$$x(t) = \sum_{i=1}^{n} IMF_i + r(t) \quad (2.3)$$

where, **n**is the number of residues and **r(t)**is the final residue.

The EMD algorithm presented above is often referred to as Sifting process. It serves the following two purposes namely to eliminate riding waves and to make the wave profile more symmetric.

If the sifting cycles continue indefinitely the IMF components would lose all its physical components (Huang et al., 1999)To ensure this does not happen and the IMFs retain enough of its components that would make sense of the amplitude and frequency modulations physically, the number of times that the sifting process occurs has to be limited.The stopping criterion that the we iterate up to the pre-defined times after the residue has satisfied the restriction that the zero-crossings and the extrema do not differ by more than one, the sifting process can be stopped by any of the following criteria which must have been determined beforehand: either when the component or residue is so small that it is less than a pre-determined value of a substantial sequence orwhen the residue becomes a monotonic function from which it is no longer possible to extract anymore IMFs. This paper adopted as its stopping criterion that proposed by(Huang et al., 2003). The flowchart for the EMD algorithm is presented in Figure 2.

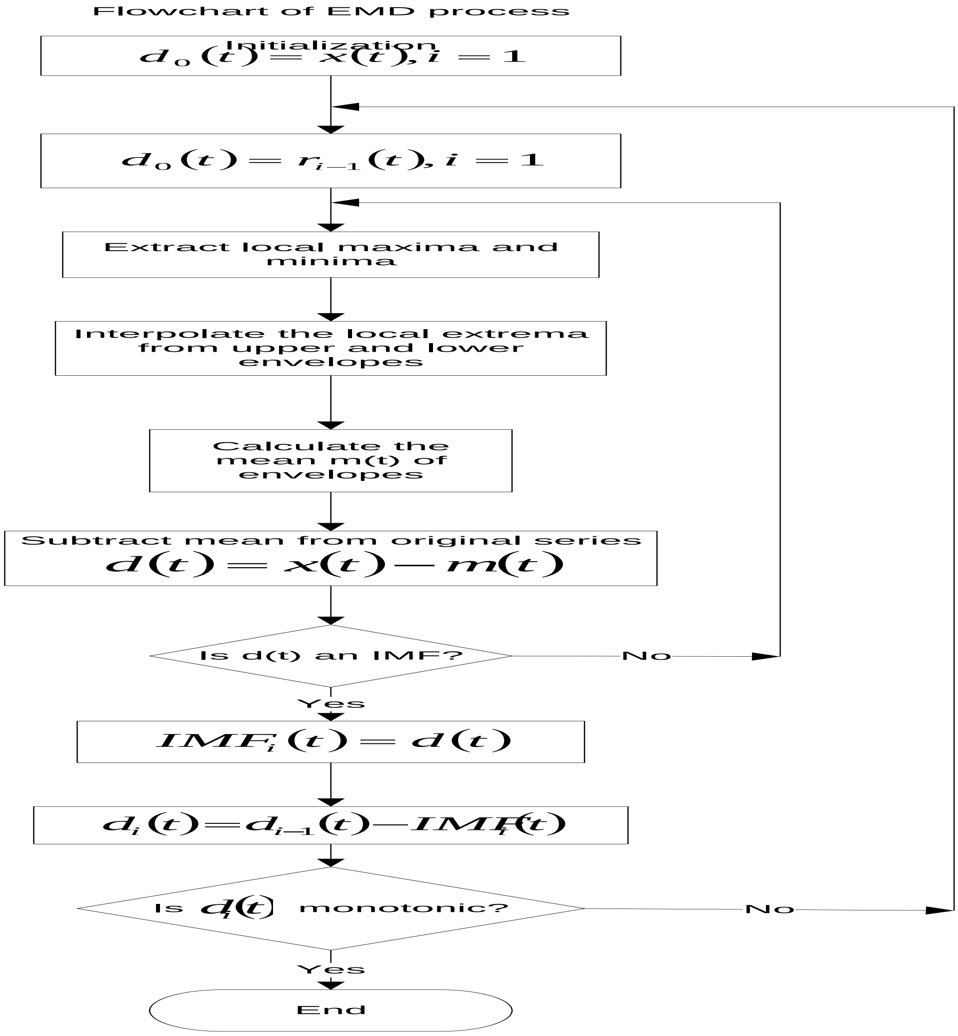

Figure 1: The EMD sifting process

2.2.2 ARIMA MODELS

Time series is a set of observation that is arranged chronologically over time. Many scientific data constitute a time series and this has made time series analysis a popular and ready tool in research and development (Chatfield and Yar, 1991). Time Series models whether univariate, multivariate or structural models are very helpful in model description and forecasting (Prybutok et al., 2000) and (Tsitsika et al., 2007).

One of the most important and widely used time series models is the autoregressive integrated moving average (ARIMA). The popularity of the ARIMA model is due to its statistical properties as well as the popular Box-Jenkins methodology. The ARIMA model can be used when the time series is stationary and there is no missing data within the time series (Box and Jenkins, 1976). A stationary time series is one that has a constant mean, variance and autocorrelation through time that is, seasonal dependencies have been removed by differencing the series.

The parameters of the ARIMA is generally given by: p,d,q which are integer values greater than or equal to zero and stand for the order of the non-seasonal autoregressive (AR) term, the order of the non-seasonal differencing and the order of the non-seasonal moving average (MA) respectively.

The general ARIMA process is of the form

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t + \cdots + \beta_q Z_{t-q} \quad (2.4)$$

By using the backshift operator this can be written as:

$$\varphi(\beta)(1-\beta)^d X_t = \theta(\beta) Z_t \quad (2.5)$$

The model in equation (3.17), describing the d^{th} difference of X_t, is said to be an ARIMA process of order (p,d,q) where p is the order of the non-

seasonal autoregressive term, d is the order of the non-seasonal integration and q is the order of the non-seasonal moving average term.
In addition, the ARIMA modelling is made up of three stages namely model exploration, estimation and diagnostics stages.

Model Exploration Stage:
This stage is the process of determining the appropriate model and the orders of the model. This is normally obtained by inspecting the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF). Another approach to it is by using goodness-of-fit statistic to select the best model from a class of fitted model structures and orders.
Model Estimation Stage:
This is the estimation of the coefficients of the model. Several estimation techniques are employed to achieve this such as the Least Squares Estimation (LSE) method and the method of maximum likelihood Estimation (MLE).

Model Diagnostic Checking Stage:
The interest here is to investigate if the residuals of the selected model are normally distributed and also to check if the model parameters are significant. (Box and Jenkins, 1976) said the best models are those with the fewest number of parameters in the class of models that fit the data.
Differencing
Data transformation was done by differencing this is to make the data stationary. Differencing is a technique for processing data that is used in the removal of trends and seasonal components of time series data. Consider the first difference between pairs of observation that is separated by appropriate denoted by:

$$\nabla X_t = X_t - X_{t-1} = (1-B)X_t \quad (2.6)$$

where:
B is the backward shift operator.
Taking the order p difference gives:

$$\nabla^d X_t = (1-B)^d X_t \quad (2.7)$$

Model Diagnostics
The Autoregressive (AR) and Moving Average (MA) terms are determined after any autocorrelation left in the differenced series. If the sample size is large enough, the autocorrelation of residuals sequence $Y_1, \dots, Y_n$ having a finite variance is approximately independently and identically distributed (i.i.d.) and have a distribution$N(0, \frac{1}{n})$. By inspecting the sample correlations of residuals, the consistency of the observation residuals can be tested with the iid noise. The null hypothesis of iid noise is rejected if at least two or three out of sample of forty falls outside the bounds $\pm 1.96/\sqrt{n}$ or if one sample falls far from the bounds (Brockwell and Davis, 2002).

2.2.3. EMD-ARIMA model
The steps involved in building the EMD-ARIMA model for load demand data forecasting is presented as follows:
Step 1: Extract the IMFs and residue component from the load data using the sifting processing described in 2.2.1.
Step 2: for each IMF and residue component obtained in Step 1, an appropriate ARIMA model is developed. The model developed at this step is referred to IMF-ARIMA models.
Step 3: This is the final step for obtaining the EMD-ARIMA model which is the desired model. Here, the predictions obtained from the models in Step 2 together into one.

III. DATA PRESENTATION AND RESULTS
The results obtained from the experimentation carried out in this paper are presented in this Section beginning with the data which are presented as Time Series time plot.

3.1. Time Plot of the research data
As state in the introductory Section, the time plot of the data collected for this study is presented in Figure 2

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 9 Sep. 2022, pp: 1428-1435 www.ijaem.net ISSN: 2395-5252

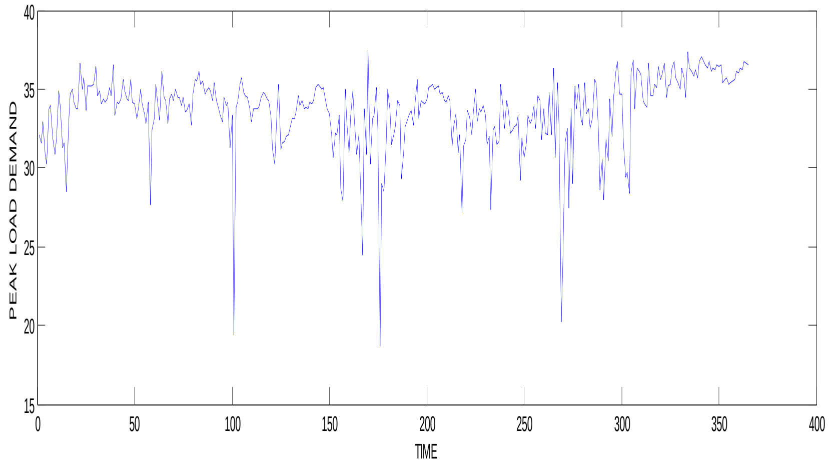

Figure 2: Time plot of the Electric Peak load demand for Bida

It can be observed from Figure 2 that the Load data for Bida exhibits a unique pattern which is depicted by the downward spikes in the time plot. This downward spikes indicate the times when supply drops or there is no supply at all in the area. There are several of these which points to the fact that there is need to plan for the load demand of the area in order to reduce these downward spikes. In addition these downward spikes are indications of system failures which invariably translate to loss of power for the utility concern whose aim at all times is to reduce system failure so as to improve service delivery to its customers and gain their confidence.

3.2 Implementation the Simulation

The simulation was conducted using the rand function in M ATLAB to generate 366 random numbers between 18.7 and 37.5 which happens to be the minimum and maximum values of the real-life Bida load demand data which we intend to imitate and this was replicated 30 times.The justification for using 30 replicates is found in (Zhang, Patuwo& Hu, 2001; Mao, Harrison & Dixon, 1999) who generated 30 and 20 replications respectively, which were employed in two separate Time Series studies. The idea behind replicating 30 times is to generated 30 different sets for the 30 simulations carried out in the study.After which these simulated figures were employed in building the appropriate ARIMA and EMD-ARIMAmodels by using the parameters obtained for configuration as stated in Section 2.

To obtain the EMD-ARIMA models the simulated datasets were passed through the EMD sifting process described in Sub-Section 2.2.1, this yielded 9 IMFs which were combined together and used to develop one single ARIMA model called the EMD-ARIMA model. The procedure is summarized in Steps 1 to 4.

Step 1: we employed the rand function in MATLAB to generate the data to be simulated.
Step 2: we employed the MATLAB software and the parameters which are explicitly stated in Table 1 to develop appropriate ARIMA model.
Step 3: we again pass the simulated data through the EMD sifting algorithm to obtain the IMFs. Thereafter these IMFs are combined to obtain the EMD-ARIMA model
Step 4: having developed the models as stated in Step 2 and Step 3, we then proceed to obtain the MAPE for each of the models.
Step 5: results obtained in Step 4 were compared with one another using the criteria put forward by Lewis (1982)using MAPE as given in equation 3.1

$$MAPE = \frac{100}{n} \sum_{i}^{n} \left| \frac{x_i - f_i}{x_i} \right| \quad (3.1)$$

Table 1: Parameters used for developing ARIMA and EMD-ARIMA models

INPUTS	BEST FIT MODELS	AIC VALUES
IMF1	ARIMA(1,1,1)	1073.2
IMF2	ARIMA(0,1,1)	483.86

IMF3	ARIMA(2,1,4)	285.36
IMF4	ARIMA(4,1,3)	-1101.22
IMF5	ARIMA(4,1,3)	-2077.83
IMF6	ARIMA(3,1,2)	3092.13
IMF7	ARIMA(3,1,2)	-4321.08
IMF8	ARIMA(1,2,1)	-6079.57
IMF9	ARIMA(0,1,3)	-9181.32

3.2.1 Simulation results for ARIMA and EMD-ARIMA models

A comparison was carried out using simulated data for the stand-alone ARIMA model and the combined EMD-ARIMA model. This is aimed at validating the claim that combine models performs more than stand-alone models and in general that the combined models improve the forecast accuracy of stand-alone models in load forecasting. The results from 30 replicates of the simulated datasets of both models are presented in Table 2.

Table 2: MAPE obtained from simulated datasets of ARIMA and EMD-ARIMA

S/N	TRADITIONAL ARIMA	EMD-ARIMA
1	8.100	5.450
2	8.390	7.131
3	8.031	5.260
4	10.120	6.124
5	8.143	5.072
6	10.790	8.063
7	8.121	6.153
8	6.252	5.160
9	6.130	6.010
10	8.110	6.21
11	11.210	5.150
12	7.562	6.850
13	8.000	5.671
14	9.144	6.610
15	8.568	5.322
16	7.420	4.110
17	7.823	5.250
18	8.950	6.116
19	8.213	6.001
20	10.004	6.784
21	6.997	4.327
22	7.200	5.110
23	8.700	6.270
24	8.190	5.002
25	7.431	4.977
26	7.700	5.011
27	8.100	5.500
28	8.342	6.001
29	7.110	6.235
30	8.900	5.219
AVERAGE	**8.259**	**5.738**

It can be observed from Table 2 that EMD-ARIMA model performs better than traditional ARIMA model judging by the MAPE posted by both models in the table. EMD-ARIMA overall MAPE is 5.738% while that of traditional ARIMA is 8.259%. This represents a 30.52% improvement.

The 30 replicates of the ARIMA model were found to be very consistent except for S/N 4 with MAPE 10.126, S/N 6 with MAPE of 10.790, S/N 11 with MAPE of 11.210 and S/N 20 with MAPE of 10.004. The percentage of these four out of 30 is negligible considering that these constitutes 13.3% of the 30 replicates in this case. The results by the EMD-ARIMA were much better as only one of the simulation outcomes fell out of range and that is S/N 2 with MAPE 7.131. This single outcome represents 3.3% of the 30 replicates. This points to the fact that the simulated data sets are not too far from the original data set they are meant to model or imitate.

IV. CONCLUSION

This study would conclude thus that simulation which reduces cost in terms of energy, time and money is as good technique for developing electricity load demand forecast models and that indeed combine models yield better forecast accuracy when compared with stand-alone models.

Acknowledgement

We would like to thank the Tertiary Education Trust Fund (TETFUND) for the institutional Based Research (IBR) grant that aided this research work. We equally wish to thank the management of the Federal Polytechnic, Bida for allowing us use her facilities to carry out the study and lastly, we thank the management of Abuja Electricity Distribution Company for providing the data used for the study.

REFERENCES

[1]. Banks, J. (1999), "Introduction to Simulation", Proceedings of the Winter Simulation Conference, PP. 7-13.

[2]. Bracale, A., Caramia, P., Defalco, P., and hong (2019) "multivariate quantile regression for short – term probabilistic load forecasting," IEEE transactions on power systems, 35(I), PP. 628-638.

[3]. Box, G. & G.M. Jenkins. Time Series Analysis: Forecasting and Control. 1976.

[4]. Brockwell, P.J. & R. A. Davis, Introduction to Time Series and Forecasting: Taylor & Francis US. 2002.

[5]. Butt, F.M, Hussain, L, Mahmood, A, & lone, K.J(2021), "artificial intelligence based accurately load forecasting system to forecast short an d medium-term load demands", Mathematical Biosciences and Engineering, 18(1), PP.400-425.

[6]. Chatfield, C. & M. Yar, (1991), "prediction intervals for Multiplicative Holt Winters, International Journal of Forecasting, 7(1), PP. 31-37.

[7]. Chen, Y, & Zhang, D. (2021) "Theory-guided deep learning for electrical load forecasting (TgDLF) via ensemble Long Short-Term Memory," Advances in Applied Energy. 1, 1004.

[8]. Dang, C., Z. Nenadic& G.S. Kasssab (2008), "A Comparative analysis of coronary nd aortic flow waveforms", Annals of biomedical engineering, 36(6), PP. 933-946.

[9]. Dong, Y, Ma, X, & Fu, T (2021) "Electrical load forecasting: A deep learning approach based on K – nearest neighbors", Applied soft computing 99, 106900.

[10]. Hafeez, G, Alimgeer, K.S & Khan, I (2020) "Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid," Applied Energy 269, PP. 114915.

[11]. Haung, N.E; Shan, S.R., Long, M.C. Wu, H.H., Shih, Q., Zheng, N.C, Yen, C.C., Tung, H. & Liu, H. (1998), "The Empirical Mode Decomposition and the Hilbert Spetrum for nonlinear and non-stationary time series analysis", Royal Society, London, Vol. 454, PP. 903-995.

[12]. Huang, N.E., Z. Shen & S.R. Long (1999), "A New view of nonlinear water waves: The Hilbert Spectrum 1", Annual review of fluid mechanics, 31(1), PP.417-457.

[13]. Huang, N.E., M-L. C. Wu, S.R Long, S.S. Shen et.al (2003), "A Confidence Limit for the empirical mode decomposition and Hilbert Spectral analysis", Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences, 459(2037), PP.2317-2345.

[14]. Lee, J.W, Kim, H.J, & Kim, M.K (2020) "design of shirt term load forecasting based on ANW using Big data," Transactions of Korean institute of Electrical Engineers, 69, PP. 792-799.

[15]. Lewis, C.D. (1982), Industrial and Business forecasting methods: A practical guide to exponential smoothing and curve fitting: Butterworth Scientific London.

[16]. Li, C. (2020), "Designing a short-term load forecasting model in the Urban Smart Grid System", Applied Energy, Vol. 266, 114850.

[17]. Mao, A., C.G. Harrison & Dixon, T.H. (1999) "Noise in GPS coordinate time series", Journal of Geophysical Research: Solid Earth (1978-2012), 104(B2), PP. 2797-2816.

[18]. **Okolobah, V.A**. (2022), "A Combined EMD and Dynamic Regression Model for forecasting Electricity Load Demand",

Nigeria Journal of Applied Arts and Sciences (NIJAAS), Volume 15 Issue 4(1)
[19]. Okolobah, V.& Z. Ismail (2013a), "An Empirical Mode Decomposition Approach to Peak Load Demand Forecasting", Indian Journal of Science and Technology, Vol. 6, Issue 9, PP.5201-5207.
[20]. Okolobah, V.& Z. Ismail (2013b), "A New Approach to Peak Load Forecasting based on EMD and ANFIS", Indian Journal of Science and Technology, Vol. 6, Issue 12.
[21]. Quader, M.R. &Qamber, I.S.(2010), "Long-term Load forecasting for the Kingdom of Bahrain Using Monte-Carlo Method", Journal of Association of Arab Universities for Basic and Applied Sciences, Vol. 9(1), PP. 12-17.
[22]. Pham, M.H, Ngugen, N & Wu, Y.K (2021) "A novel short-term load forecasting method by combining deep learning with singular spectrum analysis," IEEE Access, PP. 73736-73746.
[23]. Rato, R.T., M.D. Ortigueira& A. G. Batista. The EMD and its use to identify system modes. 2008
[24]. Rendon – Sanchez J.F & De Menezer, L.M (2019) "Structural combination of seasonal exponential smoothing forecasts applied to load forecasting," European Journal of Operational Research, 275(3), PP. 916-924.
[25]. Tsitsika, E.V., C.D. Maravelias& J. Harlbous (2007), "Modeling and forecasting pelagic fish production using univariate and multivariate ARIMA models", Fisheries Science, 73(5), PP.979-988.
[26]. Wang, W, Dou, F, Yu, X, Liu, G, Zhnag, Q &Xie, D (2020) " load forecasting based on SVR under electricity market reform,"lop conference series: Earth and Environmental science, 467(I).
[27]. Zhang, G.P., B.E. Patuwo& M.Y.Hu (2001), "A Simulation study of Artificial Neural Networks for nonlinear time-series forecasting", Computers & Operations Research, 28(4), PP. 381-396.
[28]. Zhang, H. Xu, R, Cheng, L, Ding, T, Liu, L, Wei, Z. Sun, G. (2021) "Residential load forecasting based on LSTM fusing Self-Attention Mechanism with Pooling," Energy 229, 120222.
[29]. Zhu, Ji, Yang, Z, Mourshed, M, Guo, Y, Zhou, Y, Chang, Y, Wei, Y & Feng, S. "Electric vehicle charging load forecasting: A comparative study of dep learning approaches," Energies,12(14), PP. 2692
[30]. Zhang, G.P., B.E. Patuwo& M.Y.Hu (2001), "A Simulation study of Artificial Neural Networks for nonlinear time-series forecasting", Computers & Operations Research, 28(4), PP. 381-396.