Scientific Journal Impact Factor (SJIF): 1.711
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745
ISSN (PRINT) : 2393 - 8161

A New Strategy of Load Forecasting Technique for Smart Grids.

Asmaa H. Rabie[1], Ahmed I. Saleh[2], and Khaled M. Abo-Al-Ez[3]

[1] *Department of computer engineering and systems, Mansoura University*
[2] *Department of computer engineering and systems, Mansoura University*
[3] *Department of Electrical Engineering, Mansoura University*

Abstract—Smart grids, or intelligent electricity grids which employ modern IT/communication/control technologies is becoming a global trend nowadays. Load forecasting is a vital part for the power system planning as well as operation to provide intelligence to energy management within smart grids. There are several methods that are used to improve the load forecasting accuracy. Those techniques differ in the mathematical formulation as well as the features used in each formulation. Electric loads may have bad data called outliers' data. As well as, those loads are affected by several features such as weather, season, and events; with some of them having more impact than the others. Therefore, the filtration process is an important stage before forecasting process in order to achieve a faster and more accurate forecasting process than others. This paper intends review different widely used load forecasting techniques and provides a comparative study to show the effectiveness of each of them with simulations in MATLAB. Experimental results have shown that modify/extended logistic model provides the best performance as it gives the minimal error. In split of this advantage, this method doesn't achieve fast prediction as well as optimal minimization of error value. Therefore, a new prediction strategy using data mining techniques will be introduced to give more accurate and fast predicted load value. This strategy is presented as two stages; filtration stage as well as forecasting stage using advanced data mining techniques.

Keywords—Smart grids; load forecasting; time series methods; econometric methods; data mining techniques.

I. INTRODUCTION.

A smart grid is an advanced and intelligent electricity transmission and distribution network that employs modern information, communication , and control technologies to develop economy, efficiency, reliability, and security of the grid. Load forecasting is an important process for the power system planning and operation to provide intelligence to energy management within smart grids. Electric load forecasting is classified in terms of the planning horizon's duration: up to 1 day/week ahead for short-term, 1 day/week to 1 year ahead for medium-term, and more than 1 year ahead for long-term [1] as shown in Figure1. Short-term forecasts are used to schedule the generation of electricity. Medium-term forecasts are used to schedule the fuel purchases. Long-term forecasts are used for planning purposes of the power supply and delivery system. Electric load forecasting methods may be provided in two essential groups: widely used classical methods, and new soft computing techniques based on artificial intelligence [2]. The main problem that demand pattern is almost very complex due to the deregulation of energy markets. Therefore, finding an appropriate forecasting model for a specific electricity network is a difficult task. Thus, it is an important to find an appropriate forecasting model suitable for stable and mature utilities, as well as dynamic and normal or fast growing utilities with very high accuracy and speed. Consequently, the main objective of the development of forecasting models is to improve the forecasting accuracy and reduce it's time. Accurate forecasts lead to substantial savings in operating and maintenance costs, increased reliability of power supply and delivery system, as well as correct decisions for future development [1].

Electric loads may have bad data called outliers' data that is a rare data whose behavior is very exceptional when compared with rest large amount of data [3]. The major objectives of outliers' rejection are to improve model performance as well as accuracy during training phase. In electrical power system, outliers are represented as a bad data which have an unexpected effecting on the loads, ***f For example;*** the existence of a football game in the winter. This event leads the demand of load to be higher than the normal loads in winter. This event was not repeated ,so it must be eliminated from the training data set that is used to learn the system. Thus, it is a main important to eliminate all bad data from electric load data sets before the forecasting process. Electric loads are affected by a variety of features such as time factors, weather conditions, class of customers, special events, population, economic indicators, trends in using new technologies, and electricity price [4]. Several prediction algorithms don't perform well with large amounts of features, where many irrelevant features may be presented in electric load data [5]. Subsequently, removing all irrelevant features by using data mining techniques is an important process before the forecasting process. The major objectives of features selection are to improve model performance as well as to provide faster and more cost-effective models [5]. Filtration process includes outliers' elimination process and features reduction process.

Classical or traditional approaches are based on statistical methods and forecast future value of a variable by using a mathematical combination of the historic information [6]. Classical forecasting techniques such as regression, multiple-regression, exponential smoothing and Iterative reweighted least-squares technique [2]. As well as soft computing techniques such as fuzzy logic (FL), neural networks (NNs), evolutionary algorithms (EAs) like genetic algorithms (GAs), Wavelet Networks, expert system methods and support vector machines [7]. Depending on the degrees of the mathematical analysis used in the forecasting models, these are presented into two basic types, namely: quantitative and qualitative methods. Quantitative techniques rely on forecast future data as a function of past data; they are appropriate when past data are available, as well as qualitative techniques rely on the opinion and judgment of consumers, experts; they are appropriate when past data are not available [2]. Several techniques take considerable time to give a predicted value with a less accuracy value. That is because many irrelevant features may be presented in electric load data. As well as, many bad data may be presented in electric load data. So that, the filtration of electric load data is an important stage before forecasting stage to give an accurate and fast prediction.

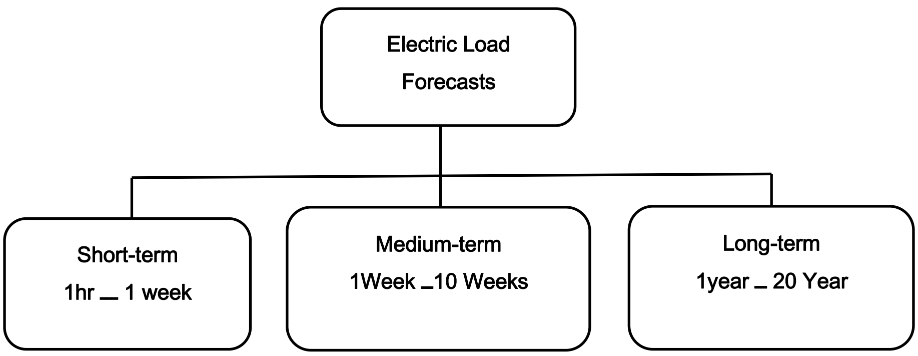

Figure1. Classification of electric load forecasting

In this paper, a review of different widely used classical electricity load forecasting techniques such as time-series models and econometric model, are presented. Thereafter, a comparative analysis of those methods is introduced to show the effectiveness of each method. Also, a new prediction strategy to give an accurate and fast predicted load value will be introduced to solve the problems of those techniques. The rest of the paper is organized as follows: *Section2* shows an overview of the widely used classical load forecasting methods. *Section3* shows the mathematical formulation of each method .*Section4* introduces a new load forecasting strategy. *Section5* shows a case study to compare the reviewed methods as well as new proposed model results in MATLAB simulation. *Section6* discusses of the results. Conclusions are discussed in *Section7*.

II. OVERVIEW OF THE WIDELY USED CLASSICAL LOAD FORECASTING METHODS.

Widely used classical models are such as time-series models and econometric model. Time-series models are quantitative techniques, which are depended only on the extrapolation of past observations of the load. While the econometric approach combine both economic theory and statistical techniques [8].

A. Time Series Models.

Time series models analyze the load evolution to detect the dynamic characteristics of the load and to extrapolate them in to the future by a mathematical tool. So, it is often useful to decompose the load in to components and to treat these components separately [8]. Also, it is a sequence of observations made through time, in the form of vector or scalar [9]. Time series prediction can be represented as a problem of a model formation that generates a mapping between the input and the output values. After such model is formed, it can be used to forecast the future values depended on the previous and current values [9]. Thus, the advantage of time series models is their structural simplicity. Where, they do not require collection of data on multiple variables [10]. In addition, observations on the variable under study are completely sufficient. Despite of these advantages, they do not describe a cause-and-effect relationship [10]. So, a time series does not provide insights into why changes occurred in the variable.

B. Econometric Methods.

The econometric approach combines economic theory and statistical techniques in order for forecasting electricity demand. Therefore, this approach estimates the relationship between energy consumption (dependent variables) and factors influencing consumption. In addition to that, the relationships are estimated by the least-square method or time series methods. Thus, one of the options in this framework is to aggregate the econometric approach, when consumption in different sectors (residential, commercial, industrial, etc.) is calculated as a function of weather, economic, price and other variables, and then estimates are assembled using recent historical data [7],[11], and [12]. The econometric advantages are that it provides detailed information on future levels of electricity demand, why future electricity demand increases, and how electricity demand is affected by all the various factors [7], [11]. Despite of these advantages, for an econometric forecast to be accurate, the changes in electricity remain the same in the forecast period as in the past [7], [11]. Thus, its failure is to recognize the interdependence between prices and quantity. So, this assumption that is called constant elasticity may be hard to justify especially where very large electricity prices changes, make customers more sensitive to electricity prices [7].

III. MATHEMATICAL FORMULATION OF CLASSICAL LOAD FORECASTING METHODS.

Widely applied mathematical methods for load forecasting are based on [8], [4]: A) *Curve Fitting such as:* Linear Curve Fitting (LCF) or Straight line, Time Polynomial (second order curve fitting (SCF) and third order curve fitting (TCF) …..Etc.), and Exponential Curve Fitting (ECF). B) *Gompertz Model*. C) *Logistic Model (*extended or modified Logistic Model). Also, there are basic long term loads forecasting techniques (LTLF) such as *Econometric Model.* In this section the mathematical formulation of each of the forecasting methods are presented.

A. Curve Fitting Techniques.

Approximating an unknown function with sample data is an important practical problem. Curve fitting methods was applied to forecast an unknown function using a finite set of sample data, where a function is constructed to fit sample data points [13]. So, curve fitting approach may be used to find the load of the target year [4]. Advantages of this approach are that it is simple to understand and it is inexpensive to implement. Despite of these advantages, it implicitly assumes that the trends in various load driving parameters remain unchanged during the study period [4]. There are several methods of curve fitting. In addition, interpolation is a special case of curve fitting where an exact fit of the existing data points is expected. In addition, the term "regression" is used to include many different models of curve fitting [13].Once a model is established, acceptability of the model must be tested. There are several measures to test the goodness of a model such as mean absolute percent error (MAPE) ….etc. [13]. Most popular curve fitting models are [8], [4]:

- *Linear Curve Fitting (LCF) or Straight line or linear regression:*

$$PD(t)=a+b*t \qquad (1)$$

Where, ***a*** and ***b*** are fitting parameters. And, ***PD(t)*** represent load at time ***t***.

- *Polynomial Curve Fitting:*

$$PD(t)=\sum_{i=0}^{n} bi * t^{i} \qquad (2)$$

If (n=2), then that is second order curve fitting (SCF). If (n=3), then that is third order curve fitting (TCF).
Where, ***bi*** represents fitting parameters according to each order. And, ***PD(t)*** represent load at time ***t***.

- *Other nonlinear fits (Exponential Curve Fitting (ECF)):*

$$PD(t)=a*(1-e^{bt}) \qquad (3)$$

Where, ***a*** and ***b*** are fitting parameters. And, ***PD(t)*** represent load at time ***t***.

B. Gompertz Model.

A Gompertz curve or Gompertz function is a sigmoid function. It is a type of mathematical method for a time series, in which growth is slowest at the start and end of a time period. So, the right-hand or future value asymptote of the function is approached much more gradually by the curve than the left-hand or lower valued asymptote, unlike the simple logistic function where both asymptotes are approached by the curve symmetrically. Thus, it is a special case of the generalized logistic function [14]. Gompertz model formulation as follow [8]:

$$PD(t)= e^{b(c^{\wedge}t)} \quad (4)$$

Where, ***a*** and ***b*** are fitting parameters. And, ***PD(t)*** represent load at time ***t***.

C. Logistic Model.

A logistic function or logistic curve is a common "S" shape (sigmoid curve) [15], with equation as follow [8]:

$$PD(t)=K/(1+m*e^{at}) \quad (5)$$

Where, ***a***, ***m*** and ***K*** are fitting parameters. And, ***PD(t)*** represent load at time ***t***.

In this technique, time is the major effected factor on predicted load. In *Extended or modified Logistic Model*, more weights such as temperature weight and event weight (Ramadan event) may be attached to the loads towards the end of the past period. So, the prediction may be improved, with equation as follow [16]:

$$PD(W) = \frac{K+AW}{[1+B\exp(CW+DW^{n})]}\ ,\ W=t+W_r+W_x \quad (6)$$

Where ***PD(W)*** is the annual peak demand, ***W*** comprising the effect of time, maximum temperature and Ramadan effects on peak demand, ***t*** is the time (in years), ***Wr and Wx*** are weighting factors representing the effect of Ramadan and maximum temperature on the peak demand, ***A,B,C,D and n*** are constant coefficients.

D. Econometric Model.

Econometric approach estimates the relationship between energy consumption (dependent variables) and factors influencing consumption [7]. A typical nonlinear estimation as follow [4] is:

$$D_i=a\ (\text{per capita income})_i^{\,b}\ (\text{population})_i^{\,c} \quad (7)$$

Where ***i*** represent the year and ***a, b, and c*** are the parameters to be determined from the historical data.

Once this relationship is generated, the future values of the driving variables (i.e. per capita income, population, etc.) should be projected. Also, D_i for a future year can then be determined [4].

IV. THE PROPOSED LOAD FORECASTING STRATEGY.

In this section, a new strategy based on an accurate load forecasting for power system in smart grids will be described. Load forecasting is an essential process for the power system planning and operation in smart grid. In addition, accurate forecasting will enable the utility provider to plan how to use or assign the resources like fuel in advance as well as to take control actions such as switching on/off demand response appliances. Also, this allow him to give the correct decisions for future development. Initially, it is an essential process to study the electrical power system and the electrical load as well. Moreover, the influential conditions or features on the load must be known and well understood. The best forecasting model will be established after understanding of overall system as well. After that, collecting data of this system is an important process. This data is considered as loads from smart meters under the affection of available collected conditions or features. Then, the representation of this collected data in proper form for forecasting model is an essential process. Representation process means the formulation data in the fashionable form for forecasting model to understand and deal with it. In this paper, formulation data is considered as tables, which represent training and testing data sets. Thus, the proposed model become already to deal with this data.

Dealing with this data may not be enabled model to perform its task well. So that, after the representation data stage, it is an important to filtrate the data before dealing with the model. In filtration stage, there are two types of filtration data that is had a bad effect on forecasting model performance; outliers data and irrelevant features. Outliers' represent a rare data whose behavior is very exceptional when compared with rest large amount of data. This data had a bad effect during the training phase to forecasting model. Therefore, it's an important process to filtrate training data from this bad data before the learning phase. So, it is essential to apply an advanced outliers' rejection methodology before forecasting model to filtrate data sets from all outliers' data. On the other hand, presenting irrelevant features badly affect the prediction algorithms. After that, forecasting model can perform its task well using data sets after passing it through filtration stage. Finally, new proposed model is considered as several stages; (i) understanding well the studied system, (ii) collecting and representing data, (iii) filtrating data from outliers, (iv) filtrating data from irrelevant features, and (v) applying forecasting model using filtrate data from previous stages. Those stages and proposed models are illustrated in Figure2.

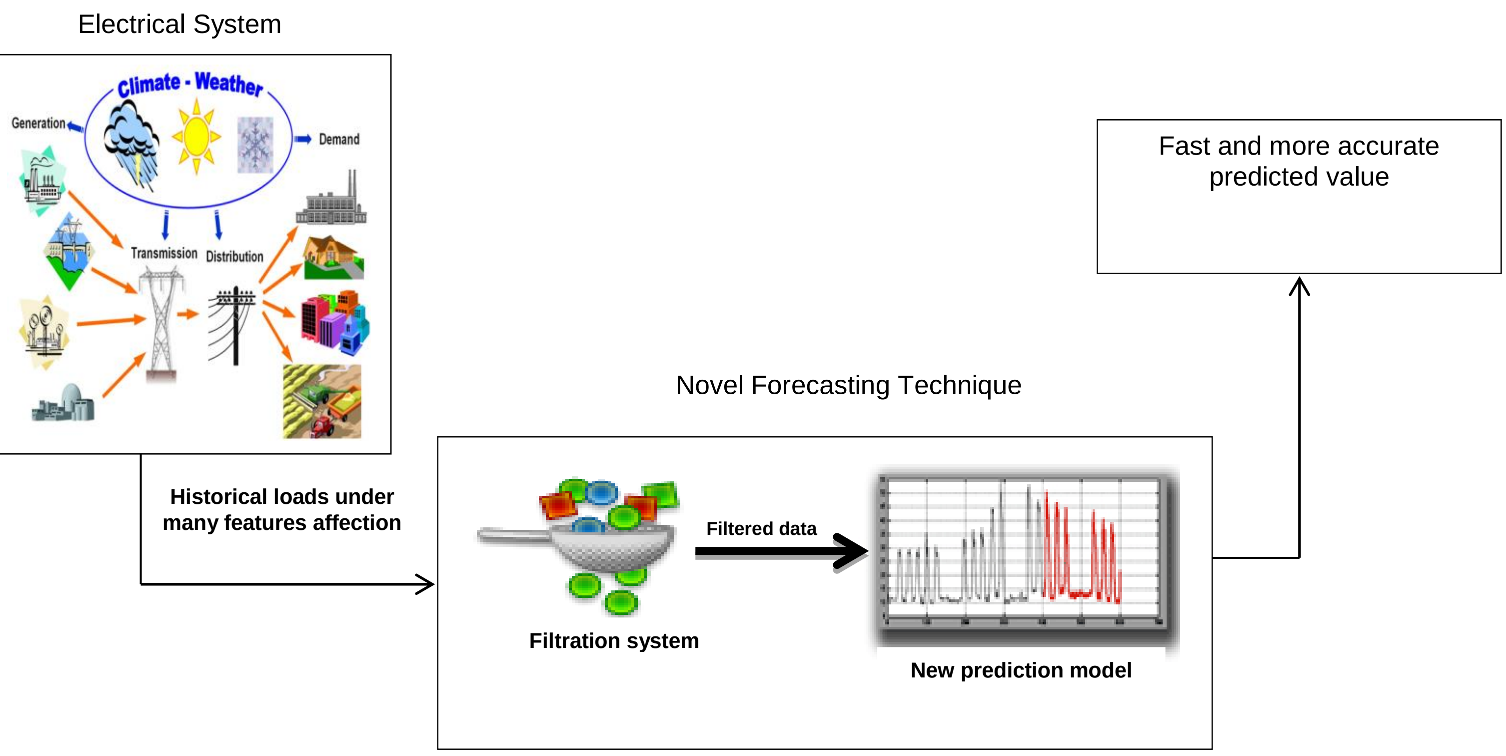

Figure2. Load forecasting strategy in the smart grid

V. SIMULATION RESULTS.

In this section a case study is used for comparing our proposed model with different classical forecasting methods. The case study is a region composed of agricultural and residential utilities to forecast the load of this large scale utility. With the data presented in Table1 [4], that's consists of actual load and the parameters affecting the forecasted load of future such as Gross Domestic

Product (GDP) and population. The calculations for any model are down in MATLAB as shown in a flowchart in Figure3. As well as, we was depending on DELL INSPIRON N5010 - CI3 device with 320GB Storage Capacity, 4 GB DDR3 RAM Storage, Intel CPU Brand, Core i3 CPU Family, 370M CPU Type, and 3MB CPU Cache .

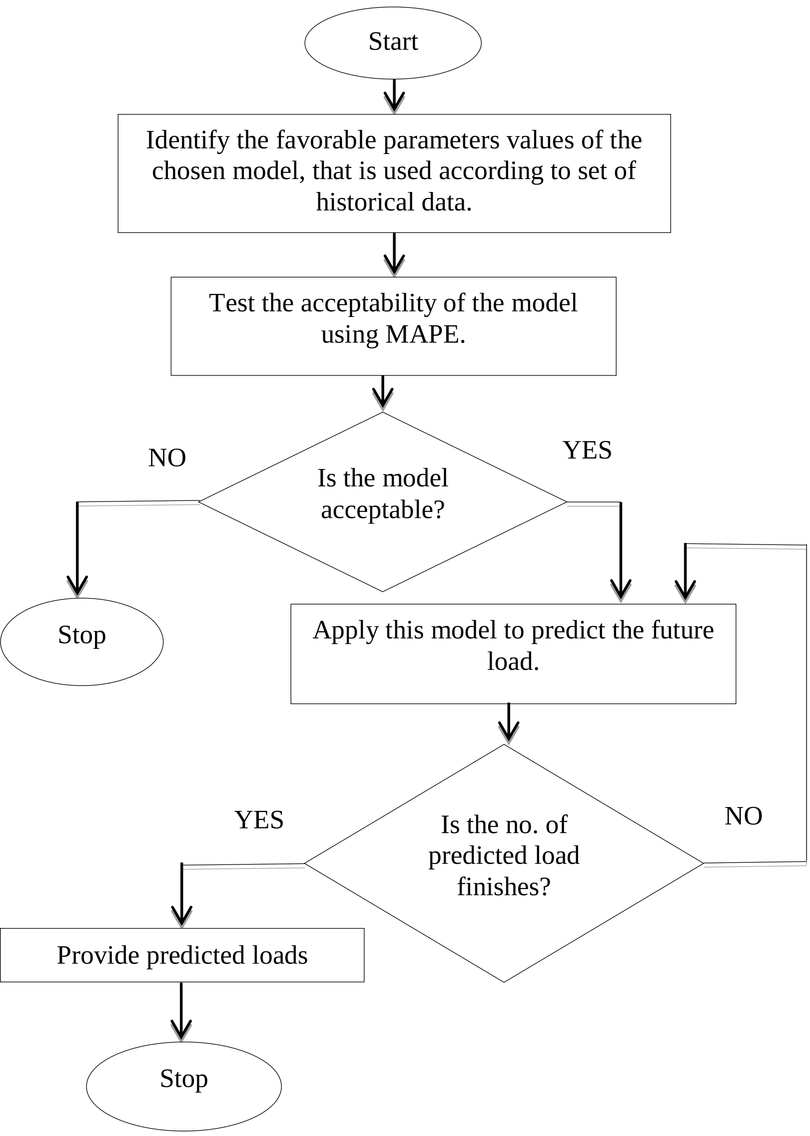

Figure3. Steps used for the used forecasting method

Table1. Data of the case study

No.	Year	Actual load (MW)	GDP (106R_)	Population/1000
1	1980	2934	219,191	36,393
2	1981	3242	209,919	37,814
3	1982	3773	178,149	39,291
4	1983	3741	170,281	40,826
5	1984	4171	191,667	42,420
6	1985	4884	212,877	44,077
7	1986	5625	208,516	45,798
8	1987	6672	212,686	47,587
9	1988	7487	193,235	49,445
10	1989	7999	191,312	50,662
11	1990	8738	180,823	51,909
12	1991	9184	191,503	53,187
13	1992	10,276	218,539	54,496
14	1993	11,205	245,036	55,837
15	1994	12,064	254,822	56,656
16	1995	13,383	258,601	57,478
17	1996	14,369	259,876	58,331
18	1997	15,251	267,534	59,187
19	1998	16,109	283,807	60,055
20	1999	17,465	291,769	61,070
21	2000	18,821	300,140	62,103
22	2001	19,805	304,941	63,152
23	2002	21,347	320,069	64,219
24	2003	23,062	330,565	65,301
25	2004	24,750	355,554	66,300
26	2005	27,107	379,838	67,315
27	2006	29,267	398,234	68,345
28	2007	32,217	413,765	69,254
29	2008	34,107	437,344	70,313
30	2009	34,894	464,308	71,410
31	2010	37,639	496,313	72,483

The prediction behavior is expressed in terms of the error using mean absolute percent error (MAPE) as follows [16], [8], and [4]:

$$\text{Error} = \left|\frac{\text{Forecasted} - \text{Actual}}{\text{Actual}}\right| * 100\ \%, \text{ and average of error } = \frac{\sum_{i=1}^{n} Errori}{n} \qquad (8)$$

To measure the accuracy of each method:

$$\text{Accuracy=100-Error} \qquad (9)$$

The more accurate forecasting model would be the one with low error value. The MATLAB simulation results are shown in Table2, and Table3. The results are also presented in Figure4.

Table2. Simulation results for the load forecasting methods

No.	Year	Actual load (MW)	GDP (106R_)	Population/1000	Forecast (MW)							
					LCF	Time polynomial		ECF	Gompertz	econometric	Modify/extended logistic	Proposed Model
						SCF	TCF					
28	2007	32,217	413,765	69,254	26,074	29,325	31,130	25,639	41,104	8754	32,188	32,200
29	2008	34,107	437,344	70,313	27,025	30,919	33,474	26,536	44,994	8828	34,663	34,107
30	2009	34,894	464,308	71,410	27,975	30,919	35,982	27,431	49,290	8892	34,888	34,970
31	2010	37,639	496,313	72,483	28,926	34,231	38,661	28,326	54,037	8919	38,263	37,700
32	2011	—	507,728	73,840	29,877	35,948	41,518	29,219	5929	9187	40,963	38,250
33	2012	—	527,529	74,822	30,828	37,706	44,562	30,111	6510	9295	34,438	40,950
34	2013	—	548,103	75,717	31,779	39,505	47,800	31,002	7154	9377	36,913	36,880
35	2014	—	569,479	76,826	32,730	41,345	51,239	31,891	7868	9514	39,163	39,200
36	2015	—	591,689	77,848	33,680	43,226	54,888	32,779	8660	9625	39,613	39,650
37	2016	—	614,764	78,883	34,631	45,148	58,754	33,666	9540	9738	40,963	40,990
38	2017	—	638,740	79,932	35,582	47,111	62,845	34,552	10,518	9852	40,063	40,060

Table3. Prediction behavior

No.	Year	LCF	SCF	TCF	ECF	Gompertz	Econometric	Modify/extended logistic	Proposed Model
28	2007	19.07	8.98	3.37	20.42	27.5861	72.8290	0.0900	0.05
29	2008	20.77	9.35	1.86	22.2	31.9203	74.1169	1.6302	0
30	2009	19.83	6.70	3.12	21.39	41.2555	74.5178	0.0172	0.2
31	2010	23.15	9.06	2.71	24.74	43.5674	76.3035	1.6579	0.16
Average error (%)		20.7	8.52	2.77	22.19	36.0823	74.4418	0.8488	0.1025

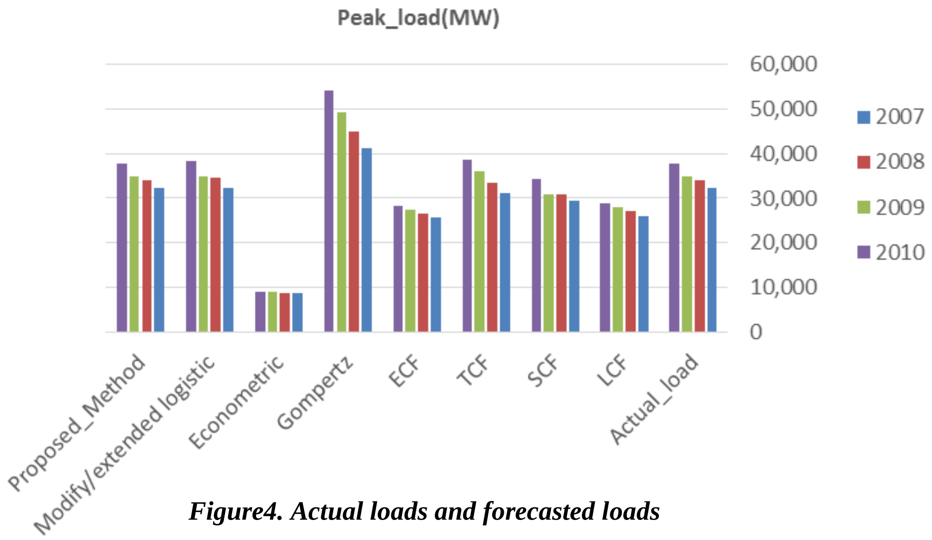

Figure4. Actual loads and forecasted loads

Volume 02, Issue 12, [December – 2015] ISSN (Online):2349–9745 ; ISSN (Print):2393-8161

In previous case study, several forecasting techniques are applied. Initially, the favorable parameters values for each technique are identified using a historical dataset in Table1. Then, each technique was used with another historical dataset in Table1 in order to test the acceptability of the model using mean absolute percent error (MAPE).Thus, the predicted loads from each technique shown in Table2 and the effectiveness of each of them shown in Table3.Therefore, The errors observed for various approaches are shown in Table3 and Figure5. The accuracy values of each technique are presented in Figure6. As shown, Extended logistic and TCF are ranked as the best choices, in terms of, the prediction behavior. From table3: the worst model, which represents the highest error value such as an econometric model with average error equal to 74.4418 %. On the other hand, the best classical model, which represents the least error value such as Modify/extended logistic with average error equal to 0.8488. Applying these traditional techniques consumes large time and effort, so it is complex in computation. Also, each one of those techniques doesn't provide optimal accuracy. Therefore, we introduce a new load forecasting strategy in order to achieve a faster and more accurate forecasting process. This proposed model gives a highest accuracy or a less error value with average error equal to 0.1025.

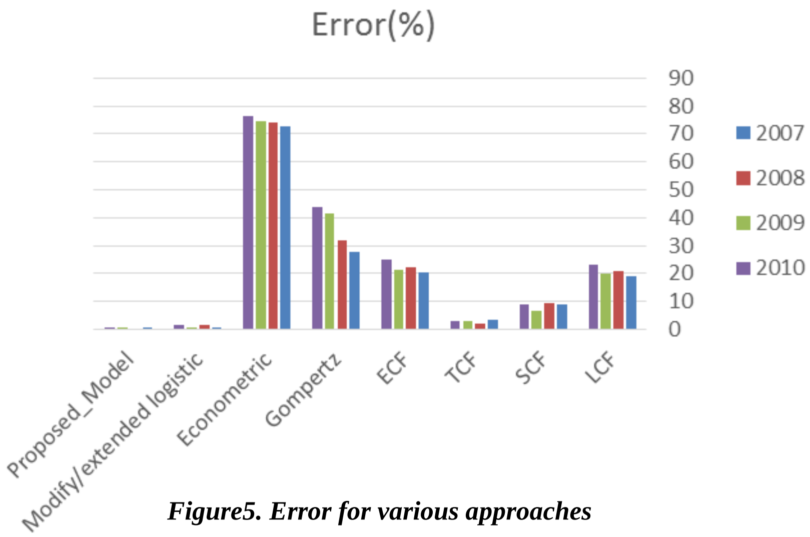

Figure5. Error for various approaches

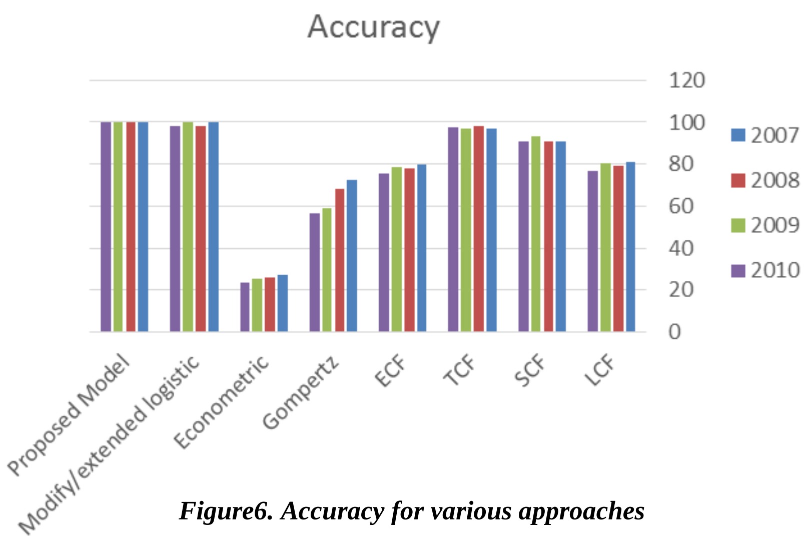

Figure6. Accuracy for various approaches

VI. CONCLUSION.

In this paper, new load forecasting strategy is introduced to provide a faster and more accurate forecasting than classical load forecasting methods. The mathematical presentation of each classic method is thoroughly presented. The methods are divided to be based on time series or econometric models. The models are tested using a case study of a typical electric grid. The comparison was performed based on the least error using mean absolute percent error. Simulations were carried on using MATLAB software. The extended logistic and TCF are ranked as the best choices, in terms of, the prediction behavior. On the other hand, the worst technique is an econometric approach as shown in previous results. Despite of this, the applying of these techniques consumes large time and effort, so it is complex in computation. Therefore, in a future work selecting the most influential features on the electric loads; eliminating outliers' data from data set for less forecasting time and high accuracy should be introduced. This can be accomplished by applying new soft computing based load forecasting strategies using data mining techniques with the aims of achieving minimum time penalty as well as better grid performance. So, the advanced model which is contents of modern filtration model and new prediction model achieving a faster and more accurate forecasting process than classical methods.

REFERENCES

[1] E. Almeshaiei and H. Soltan, "A Methodology for Electric Power Load Forecasting", Alexandria Engineering Journal, Vol. 50, 2011.

[2]A. Kumar, S. Khatoon, and M. Muazzam" An Overview Electricity Demand Forecasting Techniques", National Conference on Emerging Trends in Electrical, Instrumentation & Communication Engineering, Vol.3, No.3, 2013.

[3] F. Angiulli, and C. Pizzuti," Fast Outlier Detection in High Dimensional Spaces", Proc. Int',l Conf. Principles of Data Mining and Knowledge Discovery (PKDD ',02), pp. 15-26, 2002.

[4] H. Seifi, and M. Sadegh , "Electric Power System Planning Issues, Algorithms and Solutions", Power Systems, ISBN: 978-3-642-17988-4 (Print) 978-3-642-17989-1 (Online), 2011.

[5] S. Beniwal, and J. Arora," Classification and Feature Selection Techniques in Data Mining," International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 6, ISSN: 2278-0181, 2012.

[6] N. Amjady, "Short-Term Hourly Load Forecasting Using Time Series Modeling with Peak Load Estimation Capability", IEEE Transactions on Power Systems, Vol. 16, No. 3, 2001.

[7] L.Ghods and M.Kalantar, " Different Methods of Long-Term Electric Load Demand Forecasting; A Comprehensive Review", Iranian Journal of Electrical & Electronic Engineering, Vol. 7, No. 4, 2011.

[8] M. Kandil, S. El-Debeiky, and N. Hasanien, "Overview and Comparison of Long-Term Forecasting Techniques for a Fast Developing Utility: Part I", Electric Power Systems Research, Vol. 58, No. 1, 2001.

[9] B. Stojanovi´c, M. Boˇzi´c, and M. Stankovi´c," Mid-Term Load Forecasting Using Recursive Time Series Prediction Strategy With Support Vector Machines", Facta Universitatis, Series:ELEC. ENERG. Vol. 23, No. 3, 2010.

[10] Deloitte Consulting, "Long-Term Load Forecasting Options In Georgia", Usaid Hydropower Investment Promotion Project (HIPP), 2013.

[11]L.Ghods and M.Kalantar ," Methods for Long-Term Electric Load Demand Forecasting; A Comprehensive Investigation" , Industrial Technology, 2008. ICIT 2008. IEEE International Conference on.

[12] E. Badar," Comparison of Conventional and Modern Load Forecasting Techniques Based on Artificial Intelligence and Expert Systems " International Journal of Computer Science Issues, Vol. 8, No. 3, 2011.

[13] K. Karabulut, A. Alkanb, A. Yilmaz," Long Term Energy Consumption Forecasting Using Genetic Programming", Mathematical and Computational Applications, Vol. 13, No. 2, 2008.

[14] http://en.wikipedia.org/wiki/Gompertz_function.

[15] http://en.wikipedia.org/wiki/Logistic_function.

[16] M. Kandil, S. El-Debeiky, and N. Hasanien," Long-Term Load Forecasting for Fast Developing Utility Using a Knowledge-Based Expert System ", IEEE Transactions on Power Systems, Vol. 17, No. 2, 2002.