2018 4th International Conference on Computing Communication and Automation (ICCCA)

Very Short-Term, Short-Term and Mid-Term Load Forecasting for Residential Academic Institute: A Case Study

Vibhas Kumar Vats
Dept. of Electrical Engineering
NIT Patna
Patna, India
Email: vibhasvats.india@gmail.com

Sneha Rai
Dept. of Electrical Engineering
NIT Patna
Patna, India
Email: sneharai0212@gmail.com

Dibya Bharti
Dept. of Electrical Engineering
NIT Patna
Patna, India
Email: dibya_minu1@rediffmail.com

Mala De
Dept. of Electrical Engineering
NIT Patna
Patna, India
Email: mala.de.in@ieee.org

***Abstract*—In recent years number of smart meters installed in the distribution systems around the world has increased manifold. The large-scale data collected by these smart meters record real-time energy consumption across different nodes of the system. This constitutes a very large set of information. Availability of this very large amount of smart meter data opens up new avenues for multiple operations like load forecasting, demand side management (DSM), error identification etc. For DSM, availability of forecasted load data is indispensable. This paper presents a load forecasting technique that works well for very short term i.e. hour ahead or 15 min ahead load forecasting along with day ahead, month ahead and season ahead case i.e. short term as well as mid-term load forecasting technique. The forecasting is shown for the smart meter data of a practical system available at NIT Patna campus. The uniqueness about NIT Patna campus system is that it is a combination of all possible types of loads like commercial, residential as well as small industry due to the presence of multiple laboratories in the institute. Applicability of the proposed method on NIT Patna campus data shows the method's applicability to any type of distribution system.**

Keywords—Load forecasting, Real time practical system data, Smart metering, Residential Academic Institute, NIT Patna.

I. INTRODUCTION

In recent years, more and more smart meters are being installed in the distribution systems to collect real-time power generation and consumption data. This huge near real-time power consumption data provided new chances for load forecasting and demand side management (DSM) [1]. Today energy market is highly decentralized and includes large number of small and medium capacity renewable generating sources. This structure requires DSM to ensure uninterrupted operation of the critical loads during low generation periods caused by generation intermittency. With the availability of large scale of power consumption data at all nodes of the system every few minutes, these data set can be used to make a more accurate forecasting at every level, starting from a single smart-meter connected to any particular node, up to the complete system. As a result, load forecasting has become an essential part of the planning process of power industry.

This work is supported by the Science & Engineering Research Board, a statutory body of Department of Science and Technology (DST), Government of India, under grants number ECR/2017/001027.

Load forecasting is very important to the utility operators to incorporate DSM and also crucial for consumers so that they can optimize their use of power. Load forecasting is not a new term and the first work on electrical load forecasting based on weather condition was reported way back in 1966 [2]. Since then load forecasting was used for new infrastructure building and short term and long term planning operations [3-5]. More recently, load forecasting methods has changed to accommodate the changing power market scenario, i.e. to incorporate in the proposed techniques the big data available in terms of smart meter data [6-9].

Different classifications of load forecasting was presented over the years. Based on use of time period load forecasting techniques can be classified into four categories [10]:

- very short-term load forecasting (VSTLF) - few minutes to an hour ahead,
- short term load forecasting (STLF) - day or week ahead,
- midterm load forecasting (MTLF) - between 2 weeks to year ahead,
- long term load forecasting (LTLF) - a year to 20 years ahead.

The VSTLF and STLF techniques are important for demand response, hour-ahead scheduling, day-ahead scheduling, unit commitment and energy trading; whereas, LTLF is used for energy system planning [3]. A neural network based load forecasting method is presented in [4] to solve the real world problem of STLF Neural network is integrated with an evolutionary algorithm based on follow the leader concept and its performance is validated by COCO experimental framework. A long short-term memory (LSTM) neural network based LF technique is proposed in [5], which shows improved performance over some other existing load forecasting methods.

The smart meter connected modern grids allow improved monitoring and management electricity use as well as and prediction [6]. In [7] authors have presented a STLF method for residential customer using automatic meter reading. Kalman filtering is used for load prediction from real-time metering data. The results show that the availability of more real-time measurement data improves the accuracy of the load forecast significantly. However, the improved prediction accuracy corresponds to high computational cost. STLF of loads at substations, feeders, and individual users is done in [8] using correlations between demand and the variables influencing it. The uncertainty in demand is also predicted in distributed networks [9].

From the above discussion it can be seen that the load forecasting methodologies has evolved from its early introduction in '60s to present day, from a simple load forecasting based on weather data to smart meter connected distribution system load forecasting. In present day, large data availability within a small time frame makes the analysis process different from earlier days. Much work is going on smart meter based load predication; but most of these works are done on a particular load type, residential [7] or commercial and also for a particular time horizon, STLF or MTLF. The present work targets to proposed a load forecasting technique using smart metering data that will be fast, time horizon independent, i.e., applicable to both VSTLF, STLF and MTLF and applicable to a combination of residential, commercial and small industrial load.

The important contributions of the paper are given below:

- Firstly, a method is developed for VSTLF, STLF as well as MTLF using near real-time smart meter data.
- Secondly, the computational time is almost independent of the size of the data set used. Hence, with increasing size of data the accuracy of prediction increases without much increase in computational time.
- Finally, the forecasting technique uses all types of loads, a combination of residential, commercial and industrial load which is a part of the residential academic institute load. Real-world data set from NIT Patna is used for all analysis which consists of smart meter readings installed at different important locations of the campus.

The proposed method is implemented using Python.

The remainder of this paper is organized as follow: The statistical load forecasting method is presented in Section II. The real-life meter reading data set from NIT Patna campus meters are presented in Section III. The validation of the effectiveness of the developed approaches are shown in Section IV. Finally, the conclusion is drawn in Section V.

II. The Load Forecasting Method

Load forecasting crucial for planning and management of any electric power system. Precise load forecasting is very important for economic and reliable operation of power system. There are many forecasting techniques based on regression analysis [2], neural network [4], evolutionary algorithm [5], etc. A large amount of data collected by smart meters makes it very difficult to use for electric load forecasting directly. Therefore, it is necessary to establish some kind of interrelation between this smart meter big amount of data collected. In present paper, multiple linear regression and polynomial regression methods of statistical analysis is used to forecast using available smart meter data.

A. The Estimation Technique

A large set of data containing meter readings is available, we classify them in to training data and test data. Training data is used to train the model i.e. to predict function (f^) and test data is used to test it. There exists two types of statistical learning methods, parametric method and non-parametric method. In parametric model there are two main steps as follows:

- First, a function form is assumed, for example, assuming a linear relation.
- Then the model is fitted or trained using the data available to determine the coefficients of the linear(assumed) model.

The advantage is that we need to estimate less complex function by assuming the function in first place. And a potential disadvantage is that since we have assumed the function without actually knowing the function, it does not guarantee to be in the true form of 'f'. The problem can be eliminated by choosing more flexible function but that might lead to overfitting the data. Overfitting is an undesirable situation where the model follows the noises hidden in data In this case, the estimated data will not be accurate for any new observations which were not present in the training set.

Non-parametric model does not assume the functional form of 'f'. But it looks for an estimate of 'f' as close to the data points as possible. It provide better fit of the model. But the disadvantage is, a very large number of observations, much more than parametric method, is needed to get a proper estimate of 'f'.

B. Prediction Accuracy

With increase in flexibility for properly fitting the model, the interpretability is compromised. More flexible model is hard to interpret and vice-versa. In linear model is easy to understand the function 'f'.

C. Linear Regression

A simple equation in slope and intercept form is used to establish the proposed model. The relationship is given be (1)

$$Y \approx \beta_0 + \beta_1 X \tag{1}$$

where, β0 (Intercept) and β1 (slope) are known are coefficients or parameters. The training data is used to find out the approximate value of ^β0 and ^β1 using least square estimation.

Equation (1) becomes $y\hat{} = \beta\hat{}_0 + \beta\hat{}_1 x$.

D. Estimating the Coefficients

Training data is used to estimate the value of coefficients. The measure of closeness of the estimated coefficients values are checked by Least Square method. For the i^{th} value of X we have $y_i\hat{} = \beta\hat{}_0 + \beta\hat{}_1 x_i$ then the i^{th} residual is given by difference of i^{th} observed value and its corresponding predicted value by the linear model, by equation

$$\epsilon_i = y_i - y\hat{}_I \tag{2}$$

The residual sum of squares (RSS) is given by:

$$RSS = \epsilon_{21} + \epsilon_{22} + \cdot\cdot\cdot + \epsilon_{2n} \tag{3}$$

In least squares method the values of β_0 (Intercept) and β_1 (slope) are chosen such that RSS is minimized.

E. Assessment of Accuracy of the Model

Two parameters are used to evaluate the quality of linear regression: residual standard error (RSE), R^2 static and Mean squared error (MSE).

RSE is the estimate of standard deviation of ϵ, given by -

$$RSE = \sqrt{RSS/(n-2)} \tag{4}$$

RSE is the measure of absence of fit (absolute model of lack of fit) of the model ($Y = \beta_0 + \beta_1 X + \epsilon$) to the data. Purportedly, smaller the RSE, prediction is better fit with the data. From (5) we see that RSE is in same units of Y, so it is easy to define a value range of RSE that can be considered ad a good RSE. Hence, R^2 Statistic is used to measure fit. The value of R^2 is represented by a proportion as shown in (6) and so its value remains within the range 0 and 1. To calculate R^2, we use the formula

$$R_2 = 1 - RSS/TSS \tag{5}$$

where TSS = total sum of squares = $\sum (y_i - y_{avg})^2$.

III. The NIT Patna Campus Smart Grid

Smart meters are connected at the central substation and different strategic locations of NIT Patna campus which collects the data for every 15 seconds interval. This data consists of active and reactive load connected at any node in the network along with voltages, power factor, load current etc. for three phases. A sample load pattern is shown in Fig. 1. The figure shows the active load distribution in different phases connected to the girls' hostel. The hostel being a residential load and 15.09.2018 being Saturday (weekend), the load is almost steady. The load at a different node on the same day shows a different pattern (Fig. 2). This figure represents the load of the student activity center (SAC) which shows very less load at night (lighting loads in building and grounds) and considerable increase in load during morning and evening hours. Similar load patterns are available for other node throughout the day. A sample data is shown in Table 1.

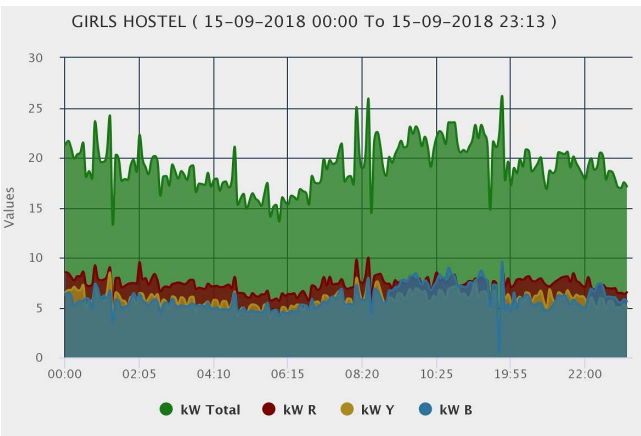

Fig 1. Load data for the Girls' Hostel for a day (15.09.2018)

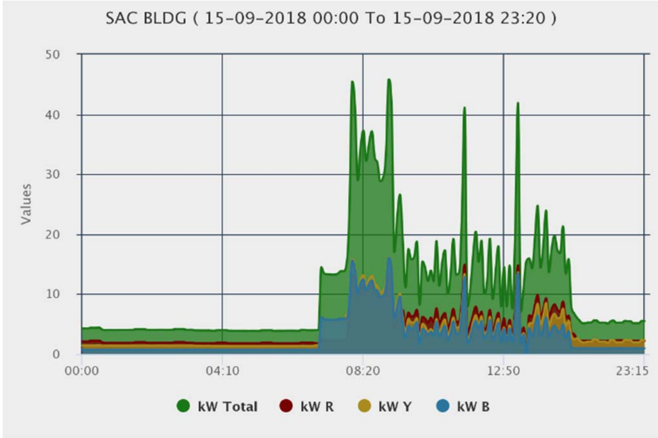

Fig. 2. Load data for the SAC building for a day (15.09.2018)

TABLE I. Hourly load pattern at the administrative building on a working day

				Feeder Name :		MAIN BLDG[]
National Institute of Technology - Patna				From Date:-		14-09-2018 00:00
Patna				To Date:-		14-09-2018 23:59
Log Date / Time	**kW Total**	**P.F**	**Volts LL**	**Volts LN**	**Amps Ave.**	**Frequency**
14-09-2018 00:00	0.00	1.000	425.60	245.70	0.00	0.05
14-09-2018 01:00	0.00	1.000	427.00	246.50	0.00	0.05
14-09-2018 02:00	0.00	1.000	423.90	244.70	0.00	0.05
14-09-2018 03:00	0.00	1.000	425.90	245.80	0.00	0.05
14-09-2018 04:00	0.00	1.000	422.50	243.90	0.00	0.05
14-09-2018 05:00	0.00	1.000	422.20	243.70	0.00	0.05
14-09-2018 06:00	0.00	1.000	425.10	245.40	0.00	0.05
14-09-2018 07:00	0.00	1.000	418.70	241.70	0.00	0.05
14-09-2018 08:00	0.00	1.000	420.30	242.60	0.00	0.05
14-09-2018 09:00	12.34	0.990	416.00	240.10	17.30	0.05
14-09-2018 10:00	20.86	-0.990	410.30	236.80	29.46	0.05
14-09-2018 11:00	27.99	0.990	413.20	238.50	39.24	0.05
14-09-2018 12:00	28.83	0.990	410.30	236.80	40.77	0.05
14-09-2018 13:00	23.86	0.990	417.10	240.80	33.17	0.05
14-09-2018 14:00	28.15	0.990	417.10	240.80	39.10	0.05
14-09-2018 15:00	26.26	0.990	416.30	240.30	36.56	0.05
14-09-2018 16:00	28.18	0.990	413.40	238.60	39.55	0.05
14-09-2018 17:00	28.19	0.990	414.20	239.10	39.52	0.05
14-09-2018 18:00	4.06	1.000	411.80	237.70	5.80	0.05
14-09-2018 19:00	1.80	0.990	411.70	237.70	2.55	0.05
14-09-2018 20:00	0.00	1.000	423.00	244.20	0.00	0.05
14-09-2018 21:00	0.00	1.000	422.10	243.60	0.00	0.05
14-09-2018 22:00	0.00	1.000	426.50	246.20	0.00	0.05
14-09-2018 23:00	0.00	1.000	416.10	240.20	0.00	0.05

Table 1 shown hourly load for a working day 14.09.2018. The hourly load is shown instead of more smaller spans to keep the size of the table within limit. The loads have non-zero value only during working hours (between 9 am to 7 pm) as it is a purely official building. The corresponding power factor, volatge, current and voltage alongwith frequency is also shown in the table. There are more parameters which get stored with this data, this includes quality index like THD also.

IV. Application and Validation

For the NIT Patna system described in previous section, the proposed load forecasting method is tested and validated and the results are presented below.

A. Comparison of Predicted data and Actual Data

The actual load on a day at main building node for few hours are shown in Table II. Data for the whole day couldn't be shown due to limitation of space. From the table we see that the actual load and the predicted load are close. For any hour of the day, there are 6 data sets which means the prediction is done for 10 min interval. This means for a day ahead prediction number of data point used training the prediction model is 144. The model is also tested for a larger set of data points, where every 3 min interval data was used to train the model. It was found that the accuracy of the model has improved when we use a 10 min interval model, i.e. 144 data points. Accuracy analysis of the present model is detailed later in this section.

The available load data are also tested to predict load for 15 days ahead, month ahead as well as for seasonal prediction. The model gives acceptable result for all cases.

TABLE II. ACTUAL AND PREDICTED RESULTS FOR MAIN BUILDING ON 4TH SEPTEMBER 2018

Date	Hour of Day	Actual Load	Prediction
4	8	7.524	7.351
4	8	11.205	10.950
4	8	12.522	12.227
4	9	13.795	13.614
4	9	18.357	18.132
4	9	18.762	18.512
4	9	18.813	18.535
4	9	21.910	21.648
4	9	25.195	24.959
4	10	24.577	24.277
4	10	24.083	23.789
4	10	24.790	24.574
4	10	24.623	24.412
4	10	24.690	24.487
4	10	23.554	23.386
4	11	24.470	24.212
4	11	23.588	23.362
4	11	24.513	24.290
4	11	24.495	24.316
4	11	26.570	26.434
4	11	24.565	24.424
4	12	24.750	24.683
4	12	24.370	24.280
4	12	24.995	24.888
4	12	24.073	23.898
4	12	21.343	21.256
4	12	19.848	19.751
4	13	18.220	18.105
4	13	19.542	19.426
4	13	20.430	20.326
4	14	23.353	23.037
4	14	23.627	23.349
4	14	23.173	22.951
4	14	23.683	23.376
4	15	24.752	24.365
4	15	25.067	24.816
4	15	24.182	23.919
4	15	22.685	22.459
4	15	21.043	20.878
4	15	21.593	21.415

The actual and predicted load for 04.09.2018 is also presented in Fig. 3 and Fig. 4 below. Fig. 3 represents the prediction based on linear regression analysis whereas Fig, 4 shows the prediction results by polynomial regression analysis. The comparison of accuracy of these two regression-based models is also presented below (Table III).

B. Accuracy of the model

Two different regression-based load forecasting models are considered. One multiple linear regression model (Fig. 4.) and the other one is polynomial regression model (Fig. 5). Comparing the results presented in Table III we see that both the models produce similar accuracy level.

TABLE III. COMPARISON OF MODEL ACCURACY

	R2 score	***Mean Absolute Error***	***Median Absolute Error***
Linear model	0.8202	3.5122	1.9712
Polynomial model	0.8218	3.4896	1.9426
Best value of errors	1	0	0

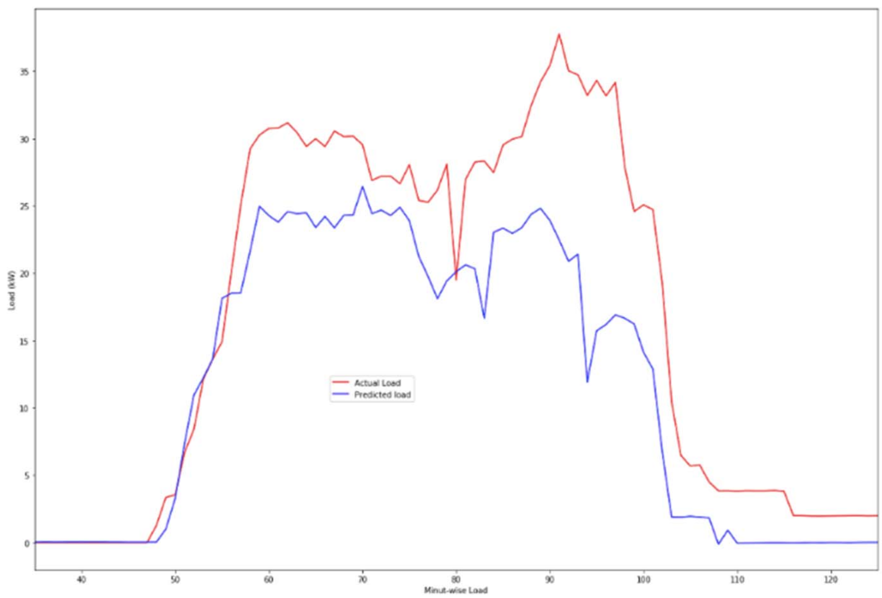

Fig. 3. The actual and predicted load for a day with linear regression model

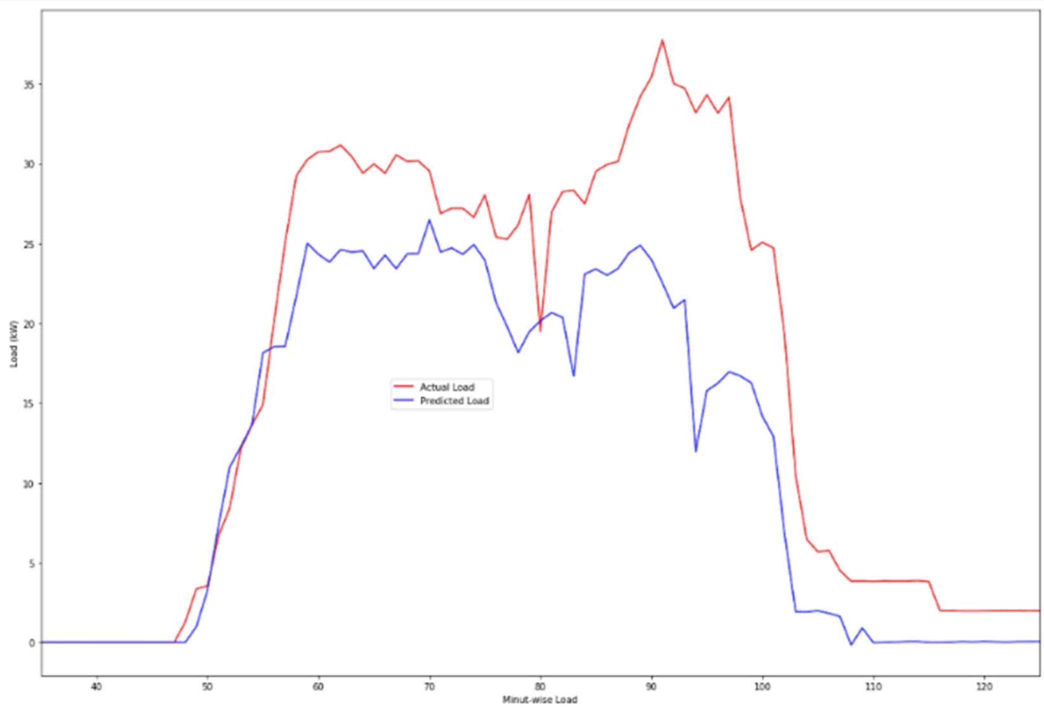

Fig. 4. Actual and predicted load with polynomial regression model

V. Conclusion

The paper presents a regression based very simple and fast method of load forecasting which is applicable to any time horizon. The method is verified for real-world smart meter data to prove the effectiveness of the proposed method. The main contribution of the work is – the proposed method is used for a residential academic institute distribution system which is a combination of all possible loads connected in any typical distribution system. This forecasted data will lead to use for management of load day ahead and planning for month ahead. In future, we plan to manage load more efficiently by applying DSM for the NIT Patna campus. The forecasting model also be extended to predict intermittent generation level based on PV and wind along with the predicted demand level and these load and generation data will be used for DSM and demand response for the system. This means, the individual big loads present in an academic institute, which is largely composed of laboratory loads, can be managed in a way that work mutually exclusive to each other.

References

[1] L. Wen, K. Zhou, S. Yang, and L. Li, “Compression of smart meter big data: A survey”, Renewable and Sustainable Energy Reviews, vol. 91, pp. 59-69, August 2018.

[2] G Heinemann, D Nordman, and E. Plant, “The relationship between summer weather and summer loads: A regression analysis,” IEEE Trans. Power App. Syst. 1966; PAS-85(11): 1144-1154.

[3] J Többena, and T Schröder, “A maximum entropy approach to the estimation of spatially and sectorally disaggregated electricity load curves,” Applied energy, volume 225, pp. 797-813, September 2018.

[4] P. Singh, and P. Dwivedi, “Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem”, Applied Energy, vol. 217, pp. 537-549, May 2018.

[5] Y. Cheng, C. Xu, D. Mashima, V. L. L. Thing, and Y. Wu, “PowerLSTM: Power Demand Forecasting Using Long Short-Term Memory Neural Network”, Advanced Data Mining and Applications, ADMA 2017, Lecture Notes in Computer Science vol 10604, Springer.

[6] K. S. Cetin, and Z. O’Neill, “Smart Meters and Smart Devices in Buildings: a Review of Recent Progress and Influence on Electricity Use and Peak Demand”, Current Sustainable/Renewable Energy Reports, vol. 4, Issue 1, pp 1–7, March 2017.

[7] M. Ghofrani, M. Hassanzadeh, M. Etezadi-Amoli and M. S. Fadali, "Smart meter based short-term load forecasting for residential customers," 2011 North American Power Symposium, Boston, MA, 2011, pp. 1-5.

[8] B. Hayes, J. Gruber and M. Prodanovic, "Short-Term Load Forecasting at the local level using smart meter data," 2015 IEEE Eindhoven PowerTech, Eindhoven, 2015, pp. 1-6.

[9] M. Majidi and K. Zare, "Integration of Smart Energy Hubs in Distribution Networks under Uncertainties and Demand Response Concept," in IEEE Transactions on Power Systems. doi: 10.1109/TPWRS.2018.2867648.

[10] S. R. Khuntia, J. L. Rueda and M. A. M. M. van der Meijden, "Forecasting the load of electrical power systems in mid- and long-term horizons: a review," IET Generation, Transmission & Distribution, vol. 10, no. 16, pp. 3971-3977, December 2016.

[11] Wang Y, Chen Q, Kang C, Xia Q. “Clustering of electricity consumption behavior dynamics toward big data applications”, IEEE Trans Smart Grid, vol. 7, pp. 2437–2447, 2016.