Load Forecasting based on Deep Long Short-term Memory with Consideration of Costing Correlated Factor

Baifu Huang
Department of Electrical Engineering
Guangdong University of Technology
Guangzhou, China
562696940@qq.com

Danqi Wu
Department of Electrical Engineering
Guangdong University of Technology
Guangzhou, China
wudanqi1022@163.com

Chun Sing Lai
School of Civil Engineering
University of Leeds
Leeds, U.K.
c.s.lai@leeds.ac.uk

Xin Cun
Department of Electrical Engineering
Guangdong University of Technology
Guangzhou, China
louiecuncx@163.com

Haoliang Yuan
Department of Electrical Engineering
Guangdong University of Technology
Guangzhou, China
Hunteryuan@126.com

Fangyuan Xu
Department of Electrical Engineering
Guangdong University of Technology
Guangzhou, China
datuan12345@hotmail.com

Loi Lei Lai
Department of Electrical Engineering
Guangdong University of Technology
Guangzhou, China
L.L.Lai@city.ac.uk

Kim-Fung Tsang
Department of Electronic Engineering
City University of Hong Kong,
Hong Kong, China
ee330015@cityu.edu.hk

Abstract—**In Day-ahead Power Market (DAM), Load Serving Entities (LSEs) needs to submit their load schedule to market operator beforehand. For reduction of the total cost, the disparity of the price of DAM and the price of RDM (Real Day Market) should be considered by the LSEs. Therefore, the problem is that a more accurate load-forecasting model sometimes provide a price that has an interspace will lead to a lower cost. Facing this issue, this paper initiates a load forecasting model considering the Costing Correlated Factor (CCF) with deep Long Short-term Memory (LSTM). The target of the forecast model contains both accuracy section and power cost section. At the same time, the construct of LSTM can offset the sacrificed accuracy. Also, this paper uses an Adaptive Moment Estimation algorithm for network training and the type of neuron is Rectified Linear Unit (ReLU). A numerical study based on practical data is presented and the result shows that LSTM with CCF can reduce energy cost with acceptable accuracy level.**

Keywords—**Recurrent Neural Network, Power Market, Load Forecast, Smart Grid, Machine Learning, Demand Response, Market Deregulation**

I. INTRODUCTION

At present, more and more countries have adopted or will adopt power market. In power market, the trade between energy buyers and sellers is driven by price instead of government policy and regulation. The deregulated power system can optimize the asset utilization and improve the social efficiency. Under deregulated environment, power markets develop into complicated systems that contain multiple trading modes. There are three familiar conceptual trading modes namely, power-pool model (day-ahead and real-time markets), bilateral-transaction model (forwards, futures, options and contracts for difference) and mix model [1, 2].

Nord Pool market in Nordic Europe, National Electricity Market in Austria, early British power market [3], PJM, CAISON and ERCOT in America electricity markets use power-pool model [4]. The bilateral-transaction model is adopted in Nordic and California electricity markets. The Nord pool in Scandinavia, the MIBEL in Portugal and Spain, they adopt mixed trade system [5].

The electric power is the concentrated spot in the whole power market, which consists of the day-ahead market and real-time market. In the day-ahead market, the vendors and the customers submit their respective quoted prices and transactional quantities to the day-ahead trade system. The transactional quantities between the disparity of actual power quantity are based on the forecast of future power demand. The calculation of day-ahead prices and real-time prices are based on Locational Marginal Pricing (LMP), which comprises system energy, congestion and loss cost [6].

Electrical load forecasting is integrated to operation and scheme of a utility company that covers the producing plan like electricity generation and purchase in power market, load switches and infrastructure development. According the time duration, the electrical load forecasting will be classified as short-term, medium-term and long-term forecasting [7].

Most of the targets of load forecasting model focus on the prediction accuracy. The tools were developed mainly based on artificial neural network (ANN) [8], fuzzy logic [9], Support Vector Machine (SVM) [10] and other time series forecasting models [11]. For instance, reference [12] designs an ANN-based model to ameliorate load forecasting accuracy in PJM market and ISO New England market [13]. In [14], the authors publish a predictive model with a hybrid adaptive fuzzy neural system.

By considering the mismatch of the accuracy of traditional load-forecasting model and total power cost, a model is built based on LSTM by considering the Beneficial Correlated Regularization, which improve the accuracy and reduce the energy cost. To begin with, the connections among the individual data within conventional model are made because of the structure of LSTM; in addition to this, the target of the forecast model contains both accuracy section and energy cost section, therefore the model provides a load forecasting solution which ensures the required accuracy and reduce the energy cost simultaneously.

The content of this paper is summarized as below. Section II provides a case about the mismatch of the model accuracy and the total energy cost. A case of LSTM is introduced in Section III. A numerical study to support the performance of proposed model is presented in Section IV. Conclusions are made in Section V.

II. PROBLEM DEFINITION: MISMATCH BETWEEN ACCURACY AND ENERGY COST

A. Power Market

The trading mode consisting of the day-ahead and real-time market is adopted globally, which stimulates participants in managing their schedule to estimate power consumption for day-ahead market. This market matches the trading by considering optimal power system operation (unit commitment and dispatch) with network stability constraints. In real-time market, the disparity between the schedule and real-time consumption will be settled, which is unable to be perceived in the schedule. Locational Marginal Pricing (LMP) is the pricing model for pricing processes in both markets [15].

The Day-ahead Market (DAM) is a forward market in which hourly clearing prices are calculated for each hour of a future operating day. Before daily accounting deadline, the hourly power consumption schedules of LSEs on the target day will be submitted to market operator, which are predicted by load forecasting model. Taking PJM as an example, Table I introduces the daily accounting deadline of PJM market for LSEs in the US [16].

TABLE I   DAILY ACCOUNTING DEADLINE OF PJM MARKET

General	4:00 PM for schedule changes from two business days prior
Power Schedule	Monday-Thursday Operating Days due on two business days later, by 4:00 PM. Friday-Sunday Operating Days due on the following Tuesday, by 4:00 PM.

In Real-time Market (RTM) the clearing prices are calculated at minute's level. LSEs will purchase the power consumption exceeding their day-ahead schedule by Real-time LMPs and will receive revenue for demand deviations below their scheduled quantities [16].

B. Mismatch between Accuracy and Cost

For the hourly consumption schedule in DAM, LSEs will predict their hourly consumption by load forecasting model. The aim of LSEs is to purchase sufficient power for their downstream consumers with least cost. The cost contains two parts, the fee paid for the power in schedule and the fee paid for or received the disparity of the schedule. In fact, using a traditional load forecasting model, a higher-accuracy schedule may not result in a lower cost. The following case study with the traditional ANN load forecasting model is performed to verify this phenomenon.

The load of a large consumer in Pittsburgh US, DAM price and RTM price in the same area are selected for analysis. Data from 2014 to 2015 is adopted for the study. According to the rule of PJM market, the model is used to predict 2-day-ahead load.
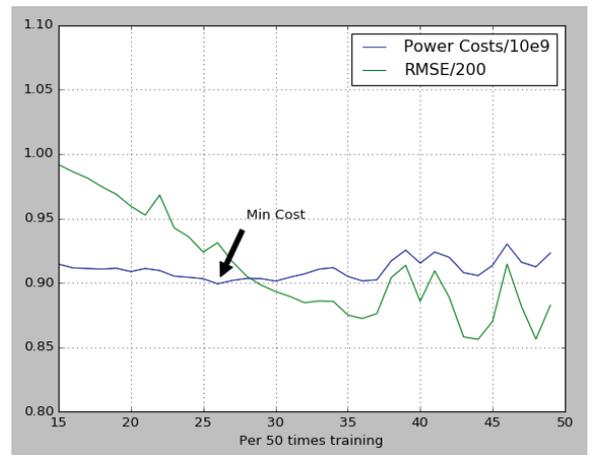

Fig 1 RMSE and Power Cost

RMSE (Root Mean Square Error) of the load forecasting model, the disparity between the forecasting load and real load, reflect the model accuracy. The power cost is calculated by price of DAM and RTM. From Fig 1, the RMSE of the model keeps a downward trend and the accuracy is rising. However the power cost does not fall continuously along with the rising accuracy. At 26*50 time training, the power cost is a minimal and then the power cost begin to increase. This phenomenon reflects the mismatch between accuracy and power cost.

III. DEEP LSTM CONSTRUCTION

A. Model Construction

This paper proposes a Long Short-term Memory based Recurrent Neural Network (LSTM) model for load forecasting. The model is implemented on real-time electricity data provided by the PJM. The predictions have hourly frequency and hence are highly beneficial for use by LSEs. The hourly predictions provide a more accurate description of load demand in the forthcoming weeks and maximum load demand can also be computed on a daily basis using the data obtained. For performing time series predictions, Recurrent Neural Networks (RNNs) are one of the most widely used models [16]. However, RNNs suffer from an inherent problem of vanishing gradient

descent. To overcome this problem and additionally formulate long-term dependencies between training samples, LSTM is used which significantly increases the precision of the proposed model [18].

Contrary to RNN model, the hidden layers of LSTM have a complex structure. The LSTM structure consists of three gates, namely, input gate, forget gate and output gate . In each gate, there is a sigmoid function to transform the data. In Equation (1), the input data contains h_{t-1} (the last output state), x_t (the No.t input data) and f(Activation function) and the input function is:

$$z = f(W_z[h_{t-1}, x_t] + b_z) \qquad (1)$$

Equation (2) is the input gate function which filters the useless information.

$$i = \text{sigmoid}([W_i[h_{t-1}, x_t] + b_i]) \qquad (2)$$

Equation (3) and Equation (4) represent the forget gate and output gate.

$$f = \text{sigmoid}([W_f[h_{t-1}, x_t] + b_f]) \qquad (3)$$

$$o = \text{sigmoid}([W_o[h_{t-1}, x_t] + b_o]) \qquad (4)$$

The model state is calculated by Equation (5) which contains the C_{t-1}(the last state).

$$c_t = f \cdot c_{t-1} + t \cdot z \qquad (5)$$

And Equation (6) is the output function within which the output is calculated by c_t and the forget gate. The structure of LSTM is shown in Fig. 2 below:

$$h_t = o \cdot f(c_t) \qquad (6)$$

For the RNN to remember the long-team meaningful information effectively, the performance of the input gate and forget gate is very important. The effect of the forget gate is to forget the former useless information, which filters the information on the basis of the x_t and the h_t. After forgetting, the information need to be supplied by the input gate in which x_t and h_{t-1} decide and provide the useful information into c_{t-1} to build the c_t.
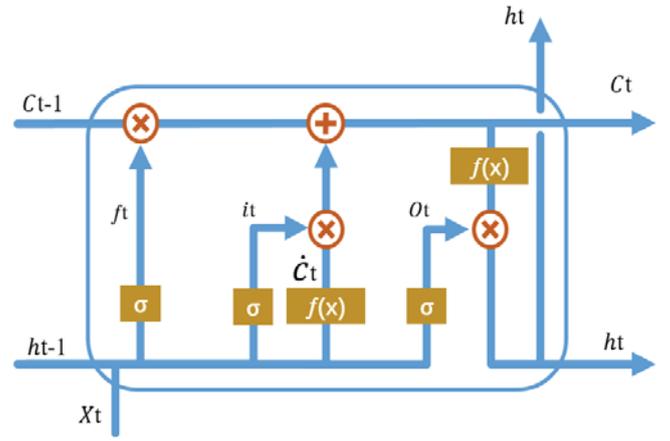

Fig. 2 Structure of LSTM

Deep LSTM based on the RNN is an advanced LSTM model. To improve the learning ability, the deep LSTM model has several layers of LSTM. Between the input and output at every moment, there are several LSTM which can extract more hidden information. Every layer of LSTM, every LSTM parameters are the same and the LSTM parameters in a layer are different from the parameters of other layers.

In Fig 3, the structure of the deep LSTM is given. The input x passes into l layer and then across the fully connected layer to form the output. The h is the initial output of the LSTM which is given as zero.
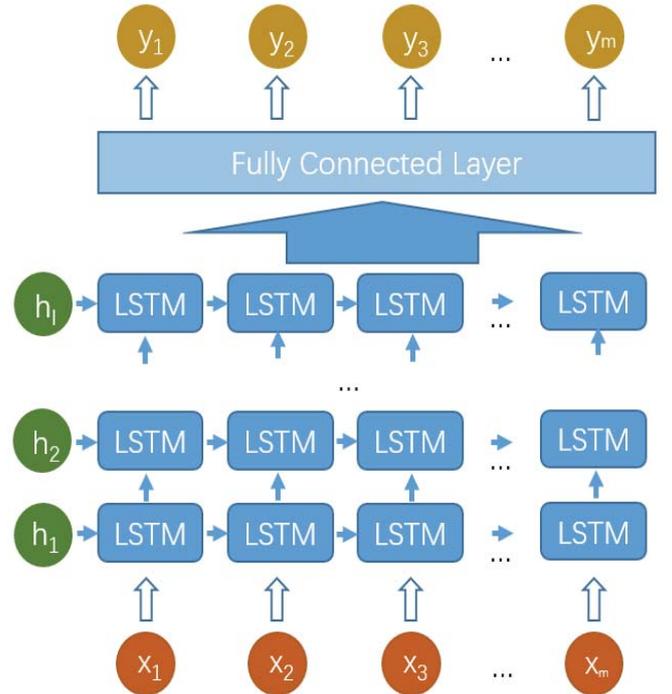

Fig. 3 Structure of Deep LSTM

B. Case study

Feature space of model input consists of historical load section and target day weather section. In historical load section, the latest weekday's 24-hour load data are selected as model input. In target day weather section, hourly temperature and humidity are selected as model input under the condition that the weather forecast accuracy is sufficiently high. The weather data is obtained from 'Local Climatological Data' from National Oceanic and Atmospheric Administration (NOAA) [19]. According to the rules of PJM submissions schedule of DAM in Table I, the historical load data of different target weekday are different, which is a mix step load forecasting data process.

For testing the effect of LSTM and several layers, several experiments have been carried out. The first one is to compare influences of diverse activation functions and the second one is to test several layers. This study uses 2014-2015 year data as training set and the 2015-2016 year data as the testing set. The results for testing set are given in Tables III. Test 1 parameters and values are given in Table II below:

TABLE II   MODEL PARAMETERS AND VALUES

HIDDEN_SIZE	20
LAYERS_NUM	1
Learning rate	0.001
Optimizer	Adam
TRAINING_STEPS	2000

The model selects three typical activation functions to take a test using the common parameters in Table I and the result is in the Table III. Observing Table III, it can be noticed that the function ReLU and function tanh perform better than function sigmoid on the condition of Table II.

TABLE III   RESULT OF DIFFERENT ACTIVATION FUNCTION

Activation Function	Mean Absolute Percentage Error (MAPE)
sigmoid	0.0812
ReLU	0.0802
Tanh	0.0806

In Test 2, tests on the influence in different layers of deep LSTM were carried out and the result is given in Fig. 4. Further tests were set up by including changes in the number of layers from 2 to 4,select function ReLU and function tanh.

From Fig. 4, we notice that the performance in different layers are changing. Since the information are abstracted further after every layer, the information produced by a surface LSTM may contain useless part and noise and an excessively deep model may filter too much information. A proper deep model can obtain abundant and useful information. Function ReLU and function tanh both got the lowest MAPE with 2 layers and the result of the function ReLU with 2 layers has the best performance in the Test 2.
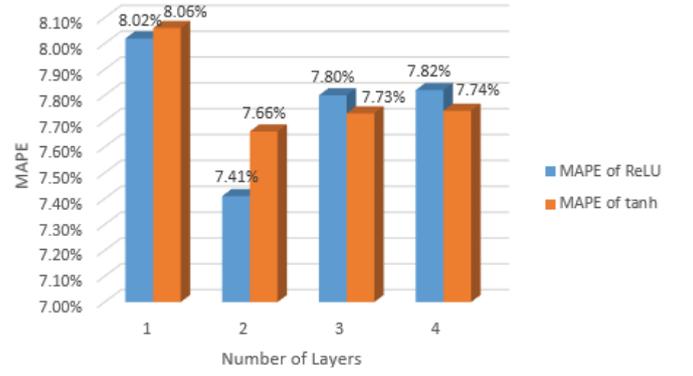

Fig. 4. Results for different number of layers

IV.   CONSIDERING COSTING CORRELATED FACTOR

A. Model Construction

From the test in Section III, It can be observed that an energy cost from a better load forecasting model may have a space to compress the cost. According to the prices in PJM which contains DAM and RTM, the loss function not only contains the traditional loss but also CCF. The loss based on the accuracy is given in Equation (7) below. Within that, $y_{forecasting}$, a vector of 1*24, represents the forecasting value and y_{real}, the vector of 1*24 represents the real value.

$$loss_{accuracy} = \sqrt{\sum_i (y_{forecasting} - y_{real})^2} \qquad (7)$$

To investigate the influence of power cost over the prediction, take the total power cost as the loss in Equation (8). Within that, P_{DAM} is the hourly price in DAM and P_{RTM} is the hourly price in RTM.

$$loss_{cost} = y_{forecastiong} \times P_{DAM} + (y_{real} - y_{forecastiong}) \times P_{RTM} \qquad (8)$$

By preventing the excessive influence of total power cost over the forecasting accuracy, this gives the $loss_{cost}$ a factor as the CCF. So the total loss function is given by Equation (9) as below:

$$Loss = loss_{accuracy} + \alpha \times loss_{cost} \qquad (9)$$

The significance of α is a control parameter to balance $loss_{accuracy}$ and $loss_{cost}$. If α is too small, the primary loss is $loss_{accuracy}$ and the effect of $loss_{cost}$ in the processing is weak. On the contrary, an enormous α will lost the accuracy and the forecast of the total price will be negative continuously. The model will limit buying power when the DAM price is lower than the RTM price and sell back when the DAM price is higher than the RTM price.

The optimization method we use is the Adam, Adaptive Moment Estimation, a method for stochastic optimization which can calculate adaptive learning rate of every parameter. This method contains g_t (gradients stochastic objective at time step t), m_t (the biased first moment estimate and initialized value is 0), v_t (the biased second raw moment estimate and initialized value is 0), $\hat{m}_t$ (the bias-corrected first moment estimate) and $\hat{v}_t$ (the bias-corrected second raw moment estimate). The Equations (10) to (15) are the parameter-updated formula and more details and information can be found in [20].

$$g_t = \nabla_\theta f_t(\theta_{t-1}) \tag{10}$$

$$m_t = \beta_1 \cdot m_{t-1} + (1-\beta_1) \cdot g_t \tag{11}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1-\beta_2) \cdot g_t^2 \tag{12}$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{13}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{14}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \varepsilon} \hat{m}_t \tag{15}$$

From [20], good default settings for the tested machine learning problems are $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$. The Adam algorithm is different from the traditional stochastic gradient descent. Random gradient descent keeps a single learning rate to update all weights and the learning rate does not change during training. Theadaptive learning rate for different parameters is calculated from the biased first moment estimation and biased second raw moment estimate.

B. Case Study

First, a research in the traditional ANN and deep LSTM model without the CCF is implemented. The energy costs in different MAPE by traditional ANN and deep LSTM model are recorded and compared in Fig. 5.

From Fig. 5, it can be observed that not only the MAPE but also the energy cost of deep LSTM are better than traditional ANN result. Similarly, in the performance of deep LSTM, the energy cost will not follow the performance of MAPE. But in the area of MAPE which the traditional ANN cannot achieve, the energy cost of deep LSTM is lower than the traditional ANN energy cost. The lowest energy cost of deep LSTM is 3.24E+08, a saving about 2.6% of the lowest power cost of ANN (3.3354E+8). So the deep LSTM indeed increases the accuracy and decrease the energy cost.
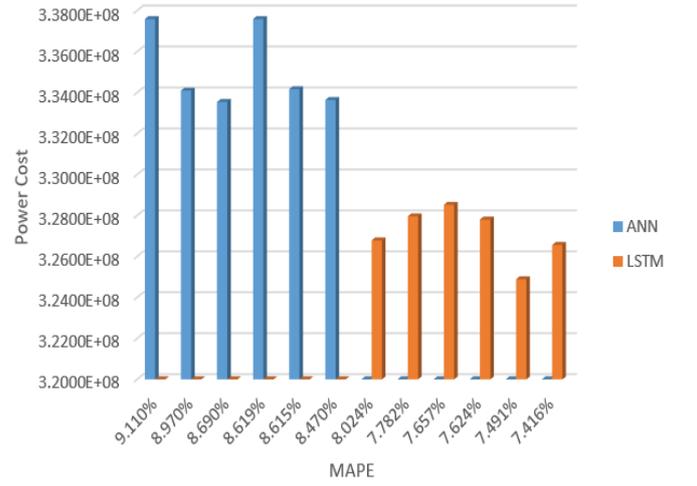

Fig. 5. Result of ANN and LSTM

Secondly, the model combines CCF with the deep LSTM and make some load forecasting tests. The parameter α of CCF is set from 0 to 0.05 and the interval is 0.001. The results of load forecasting model in diverse α are given in Fig. 6.
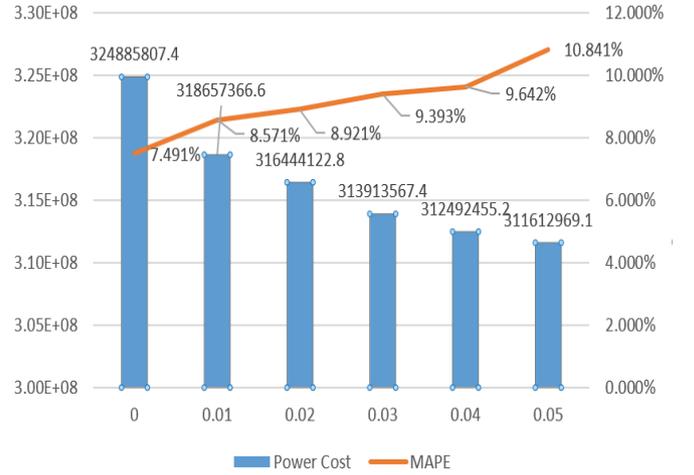

Fig. 6 Results of diverse α

From Fig. 6, as the values for parameter α increase from 0 to 0.05, the loss$_{cost}$ proportionally rises, the MAPE increases but the power cost falls. When the parameter α is 0.05, the power cost saves 4.1% of the lowest cost in deep LSTM and saves 6.6% of the lowest cost in ANN. But when α becomes larger, the accuracy will be degraded further, which may threaten the safety of the grid. So the model only reduces the energy cost in the condition of accuracy. If the condition of accuracy is 90%, the parameter α will be 0.04 approximately and this saves 6.3% of the lowest cost in ANN.

V. CONCLUSIONS

This paper introduces a new model for load forecasting and the model consists of deep LSTM and CCF. The LSTM structure can utilize the hidden connection of independent data and deep layers can dig out the hidden information to improve

the accuracy. When CCF is introduced in the deep LSTM model, the accuracy is considered while the power cost is emphasized too.

In this paper, it can be observed that mismatch between the accuracy and energy cost exists from the Test 1 and Fig. 5. Using the hidden connection of independent data, it is proved that the accuracy can be improved by the deep LSTM and the energy cost falls to some degrees relative to the traditional ANN model. Utilizing the mismatch between the accuracy and energy cost, an appropriate CCF model sacrifice accuracy to a certain extent or in a certain condition which can be complemented by deep LSTM to reduce the energy cost further.

In the future, more in-depth research will be performed, including mining deeper information structure with more diverse data combination to improve the accuracy, and a model with consideration of dynamic load forecasting instead of fixed 24 hours.

REFERENCES

[1] Algarvio, H., Lopes, F., Sousa, J., & Lagarto, J. (2017). "Multi-agent electricity markets: retailer portfolio optimization using markowitz theory," Electric Power Systems Research, vol.148, pp.282-294, 2017.

[2] M. Y. Hassan, M. P. Abdullah, A. S. Arifin, F. Hussin and M. S. Majid, "Electricity market models in restructured electricity supply industry," Power and Energy Conference, 2008, pp.1038-1042, pp. 1-3 Dec. 2008.

[3] B. Cory, "Power system restructuring and deregulation: trading, performance and information technology," Power Engineering Journal, vol. 16, page 22, Feb. 2002.

[4] P. Zou, Q. Chen, Q. Xia, C. He and C. Kang, "Incentive compatible pool-based electricity market design and implementation: A Bayesian mechanism design approach," Applied Energy, vol. 158, pp.508-518, Nov. 2015.

[5] L. Mari, N.Nabona and A.Pages-Bernaus, "Medium-term power planning in electricity markets with pool and bilateral contracts," European Journal of Operational Research, vol. 260, pp.432-443, July 2017.

[6] M. Kohansal and H. M. Rad, "Price-maker economic bidding in two-settlement pool-based markets: The case of time-shiftable loads," IEEE Trans. Power Systems, vol.31, no. 1, pp. 695-705, Jan. 2016

[7] G. Mitchell, S. Bahadoorshngh, N. Ramsamooj and C. Sharma, "A comparison of artificial neural networks and support vector machines for short-term load forecasting using various load types," IEEE Manchester PowerTech, pp. 1-4, June 2017.

[8] T. Pinto, H. Morais, T. M. Sousa, T. Sousa, Z. Vale, I. Praca, R. Faia and E. J. S. Pires, "Adaptive portfolio optimization for multiple electricity markets participation," IEEE Trans. Neural Network and Learning Systems., vol. 27, no. 8, pp. 1720-1733, Aug. 2016.

[9] D. M. Minhas, R. R. Khalid and G. Frey, "Short term load forecasting using hybrid adaptive fuzzy neural system The performance evaluation," IEEE Int. Conf. PES, pp. 468-473, June 2017.

[10] Cruz E. Borges, Yoseba K. Penya, Ivan Fernandez, "Evaluating combined load forecasting in large power systems and smart grids," IEEE Trans. Indus Inform. , vol. 9, no. 3, pp. 1570-1577, Aug. 2017.

[11] D. H. Vu, K.M. Muttaqi, A.P. Agalgaonkar, A. Bouzerdoum, "Short-term electricity demand forecasting using autoregressive based time varying model incorporating representative data adjustment," Applied Energy, vol. 205, pp. 790-801, 2017

[12] Serkan Buhan and Isik Cadirci, "Multistage wind-electric power forecast by using a combination of advanced statistical methods," IEEE Trans. Ind. Informat., vol. 11, no.5, Oct. 2015

[13] X. Xia, X. Rui, X. Bai, H. Wang, F. Jin, W. Yin, J. Dong and H. Lee, "One-day-ahead load forecast using an adaptive approach," IEEE Int. Conf. SOLI, pp. 382-387, Nov. 2014.

[14] Agrawal, Rahul Kumar, F. Muchahary, and M. M. Tripathi. "Long term load forecasting with hourly predictions based on long short-term-memory networks." IEEE Texas Power and Energy Conference IEEE, pp. 1-6, 2018.

[15] PJM, "PJM in Schedule User Guide", http://www.pjm.com/library/manuals.aspx, visited on 20th Aug. 2017.

[16] PJM, "Historical load data", http://www.pjm.com/markets-and-perations/ops-analysis/historical-load-data.aspx, visited on 20th Aug. 2017.

[17] L. Medsker and L. Jain, "Recurrent neural networks," Design and Applications, 2001.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[19] Michael A. Nielsen, "Neural Networks and Deep Learning," Determination Press, ch. 3, sec. 3, pp.83-87, 2015.

[20] Kingma, Diederik P, and J. Ba. Adam: "A method for stochastic optimization," Computer Science, 2014.