Load modeling from smart meter data using neural network methods

Nasrin Kianpoor
Department of Electrical Engineering
UiT – The Arctic University of Norway
Narvik, Norway
Nasrin.Kianpoor@uit.no

Bjarte Hoff
Department of Electrical Engineering
UiT – The Arctic University of Norway
Narvik, Norway
Bjarte.Hoff@uit.no

Trond Østrem
Department of Electrical Engineering
UiT – The Arctic University of Norway
Narvik, Norway
Trond.Ostrem@uit.no

***Abstract*—Electricity load modeling plays a critical role to conduct load forecasting or other applications such as non-intrusive load monitoring. For such a reason, this paper investigates a comparison study of two common artificial neural network methods (Multilayer perceptron (MLP) and radial basis function neural network (RBF-NN) for home load modeling application. The accuracy of load modeling using neural network methods highly depends on chosen variables as the input data set for the networks. For this purpose, data including weather, time, and consumer behavior are considered as the input dataset to train the networks. The results of this study show that the RBF-NN model has higher accuracy in training data. On the other side, the MLP model outperforms in test data. To sum up, the results prove that the load model obtained by MLP has a better performance in terms of mean square and root mean square error indices.**

***Index Terms*—Advanced metering system (AMS), Load modeling, residential load sector, neural network, multilayer perceptron, radial basis function neural network.**

I. Introduction

Smart meters have been installed in many countries such as Australia, Japan, the United States, Canada, Italy, and Norway during the past decade. Only in 2017, 665 million smart meters were installed globally [1]. Buildings equipped with smart meters can generate significant real-time or near real-time data information on occupancy and energy consumption. In Norway, an advanced metering system (AMS) and smart meters have replaced old meters to measure the amount of electricity usage in each Norwegian household. AMS is equipped with an output terminal called a home area network (HAN) port. The HAN port enables the local distribution network and users to monitor several parameters related to the household load characterization such as current, voltage, active and reactive power, etc. Therefore, the meters are not only used for measuring energy bills, but also enable the user to have access to high-resolution AMS data. Historical energy consumption combined with weather forecast data is given an opportunity to improve building load demand modeling because of high interdependency between the weather data and load consumption in a power system.

This work is part of RENEW project (No. 2061340) funded by UiT The Arctic University of Norway.

Residential home load modeling plays a key role in designating real-time building energy management systems and demand response programs. A comprehensive survey of different load modeling methods and their applications is provided by [2]. Home load modeling is divided into two categories, physical-based modeling and measurement-based modeling (data-driven modeling method) in the literature. In the physical-based method, the models are according to the knowledge of physical characteristics of loads and mathematical relations that reflect the real-world situation. However, it is not always possible to quantify all the physical parameters of a system because of limited understanding of a phenomena. Moreover, modeling all the physics of a nature most often leads to a complex model that is not efficient for real-time applications. These drawbacks motivate research on measurement-based modeling methods that use aggregated data to model the load characteristics [3]–[5]. The main capability of this method is that it can model all the physics of the load; however, the obtained model cannot be generalized and most often can only be used for the location in which data are collected. In [6], a measurement-based load modeling using transfer functions for dynamic simulations is proposed. It provides a method that begins with obtaining the currents and voltages from power quality monitoring systems and highlights the issues related to selecting, processing, and resampling of data to estimate the power deviations as a function of the voltage. A measurement-based dynamic load model using the vector fitting technique is presented in [7]. In this research, based on the measurement data, an aggregated load model is presented for dynamic solutions of large power systems. Artificial Neural Network (ANN) methods are some of the strong tools related to the data-driven modeling methods [8]–[10]. They do not represent the physical aspect of the load and require a large amount of measured data to estimate the model. On the other hand, they are highly adaptive and strong tool for modeling complex non-linear systems. An ANN is trained using a collection of input and output datasets. It has been widely used in the context of load modeling, especially for building applications. For example, a dynamic load modeling of an Egyptian primary distribution system using three ANN-based load models is presented in [11], in which the results verify the accurate emulating of load dynamics via ANN

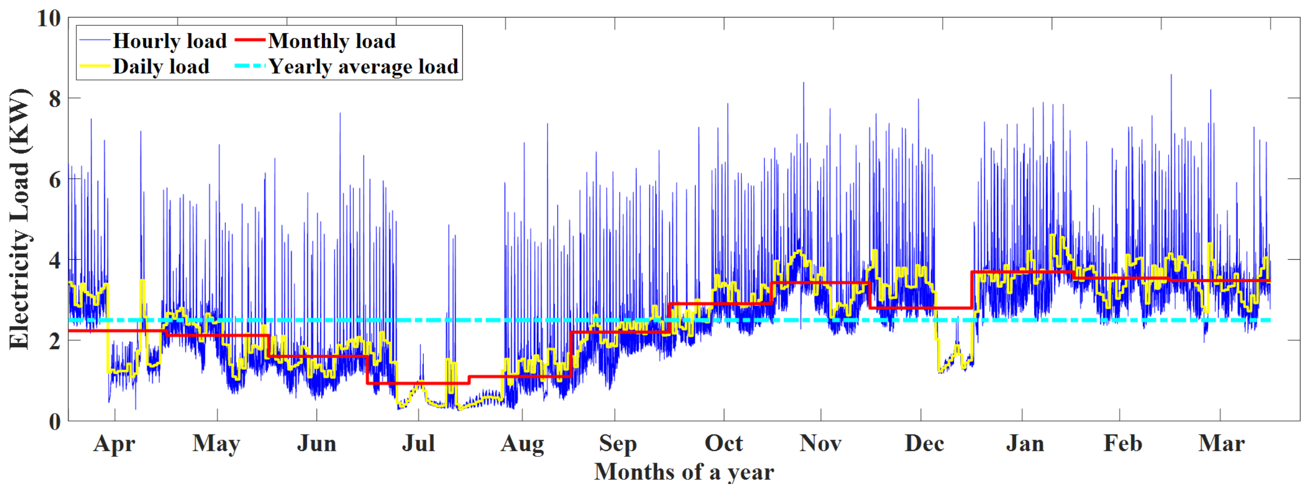

Fig. 1. Electricity load of a single-family home in Narvik, Norway from April 2019 to March 2020.

models. Keyhani et al. [12] propose a method to develop an ANN-based composite model to analyze the power system stability. A mathematical model representing the total charging load at an electric vehicle charging station is presented in [13]; the load model developed using a queuing model followed by a neural network (NN). As mentioned, there are different type of ANN architecture techniques which are employed for modeling the household load demand application. However, the differences between, as well as the strength of these techniques are still unclear in the home load modeling application. Therefore, there is still a need for a comparison study to analyze the strength of each technique for different home load modeling applications such as load forecasting and monitoring.

This main contribution of this article is presenting a comparison study to analyze the advantage and disadvantages of two widely used ANN techniques in the home load modeling application. In this paper, the multilayer perceptron architecture performance is compared with radial basis function neural network architecture in terms of their mean squared error (MSE) and root-mean-square error (RMSE) of the obtained load models. First, electricity usage is analyzed in terms of average electricity consumption with different time intervals basis (hourly, daily, monthly, and annual) for a whole year. Then the effect of occupant behavior on electricity usage is studied by sorting data into weekdays (user is not available at home during working hours) and weekends. Afterward, the data inputs including weather data, time variables, and user behavior are used to train neural networks in order to improve the accuracy of the model. Last but not least, the advantages and disadvantages of each mentioned methods are discussed and shown in the simulation results and also the impact of different inputs on home load modeling is investigated by considering and comparing four different input datasets for NN.

The remaining parts of the paper are organized as follows. In section II, the methodology is explained, including analysis of the historical data and load modeling techniques (MLP and RBF-NN). Simulation results and discussion of the different methods are illustrated in section III. Finally, conclusions are presented in section IV.

II. Data Analysis and Modeling

In this section, the methodology of the paper is presented in two parts. In the first part, the data analysis of the electricity load for one year is performed. In the second part, the MLP and RBF-NN structures are explained.

A. Data Analysis

In this paper, a case study is presented on a residential building located in the arctic climate of Narvik, Northern Norway. The data are from a single-family home (detached house) to investigate correlations between load consumption and some factors, including time of day, days of a week, months of a year, outside temperature, wind speed and holiday. The following main devices are included in the building model:

- Plug-in hybrid electric vehicle (Mitsubishi Outlander PHEV 2014) with a 12 kWh T-shaped lithium-ion battery pack and 3.6 kW one-directional charger station.
- Air-to-air inverter controlled heat pump.
- Electric hot water tank (200-liter tank with a 2 kW resistive heating element).
- One portable electric radiator (Thermostat controlled 1.5 kW).
- Four electric floor heating cables.
- Kitchen and laundry appliances (Dishwasher, washing machine, oven, stove, el-kettle, etc.).

The collected data are hourly load consumption from 01-04-2019 to 31-03-2020. The hourly, daily, monthly and annual averages of energy consumption are shown in Fig. 1. The daily average of the load is plotted in yellow to have a better view of the load consumption of each day. In this plot, the increase and decrease in electricity consumption are more distinct than the hourly load plot (Blue).

The Norwegian energy system is largely based on electricity, including heating and hot water. Hence, the energy consumption during cold months is higher than other months. According to Fig. 1, the average of the monthly load varies between 0.93 KW and 3.69 KW. As can be seen, the average monthly load starts to increase from September as the weather is getting cold and it peaks in January, February, and March that have the lowest temperatures in the year. The average annual amount of electricity usage is 2.5 KW, which is plotted in dashed cyan. From April to September the annual load is higher than the monthly load and from October to March it is lower than the monthly load.

In Fig. 2, the difference between electricity consumption of each day at the time t ($E_i(t)$) and the average electricity consumption of that particular day ($E_{avg,i}$) is plotted. The reasons behind it are to cancel the baseload and distinguishing the peak load periods. It is calculated according to (1) and denoted by $Ed_i(t)$.

$$\begin{cases} Ed_i(t) = E_i(t) - E_{avg,i} \ , & \begin{array}{l} i = 1:365 \\ t = 1:24 \end{array} \end{cases} \tag{1}$$

In Fig. 2 the load data is split into two parts, weekdays and weekends. The daily consumer pattern from Monday to Friday is quite similar; hence all the weekdays are put in one category. Saturday and Sunday are put in one category because of a similar consumer pattern. As can be seen from Fig. 2 (a), there are two peak load periods during weekdays. The first one occurs in the morning from 7:00 to 9:00, as the occupant is getting ready to leave home, it is normally around 8:00. The other peak load period occurs from 17:00 to 2:00. At this period, the occupant comes back home and the reason for high electricity consumption during midnight is the charging of an electric vehicle.

On weekends, the electricity consumption pattern changes. The pattern of weekends is clearly different from the pattern of weekdays. There is one peak load period which is from 16:00 to 2:00 which is broader than the weekdays load period.

B. Modeling

The first step to designing a proper energy system is finding reliable models for each component of the energy system. This section is aimed towards obtaining reliable models for the electricity load demand in a residential building. The developed model can be used for further studies related to different applications such as monitoring, fault detection and energy management systems.

In this paper, two different Neural Network architectures (MLP and RBF-NN) are used to model the electricity load of a building. In the following, the structures of these methods are explained.

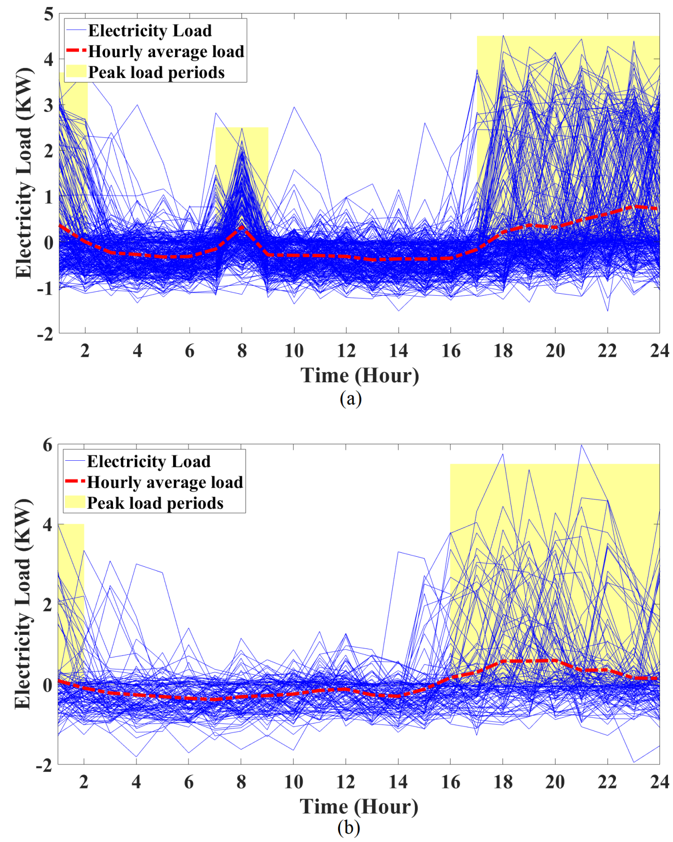

Fig. 2. Electricity load consumption of the sampled data; a) From Monday to Friday; b) Saturday and Sunday

1) MLP: The Multi-Layer Perceptron model is the most common neural network architecture. It is a strong tool for modeling and function approximation applications like load modeling [14], [15]. The MLP structure, including input, hidden and output layers is shown in Fig. 3. Evidently, it has a feedforward architecture so that the output of one layer is the input of the next layer. In the MLP structure, the output of the neural network is calculated as follows [16]

$$y = \sum_{j=1}^{n} z_j f_j (\sum_{i=1}^{m} w_{ji} x_i + b_j) \tag{2}$$

where z and w are weight matrices, n and m are the number of neurons and inputs, respectively; b is a bias term and x is the input of NN. In (2), f is an activation function which is an S-shaped curved sigmoid function as expressed below

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

There is not a straightforward method to calculate the number of neurons in the hidden layer of MLP. Therefore, it can be obtained empirically by trial and error. The MLP structure that is used in this paper has six inputs, one hidden layer with 500 neurons, and one output. Inputs are variables that have an impact on the electric consumption

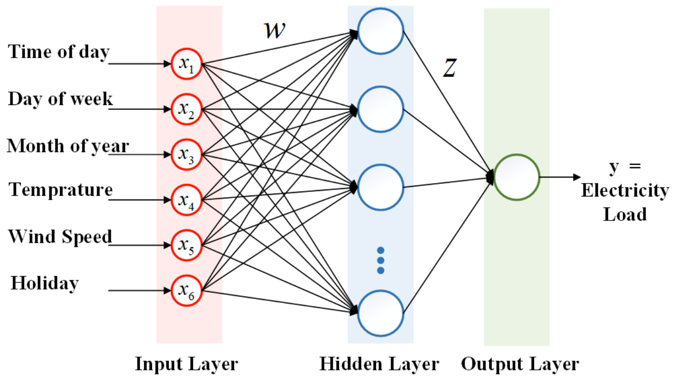

Fig. 3. Multilayer perceptron neural network structure.

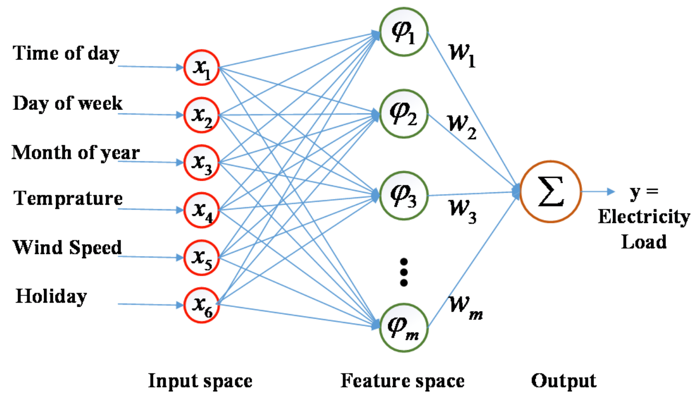

Fig. 4. Radial-basis neural network structure.

in a residential home including, time of day, weekday, month, outdoor temperature, wind speed and holiday time. The network parameters including bias and weight values are estimated according to minimizing a loss function (usually a quadratic function) during the training process. The Levenberg-Marquardt optimization is a steepest-descent backpropagation algorithm, which is employed based on the gradient of the loss function with respect to the network parameters.

2) RBF-NN: Another widely used ANN architecture for load modeling is the radial basis function neural network technique, because it can capture the nonlinearities and uncertainties of a system. It consists of three parts, input space, feature space, and output. In RBF-NN, data are transferred to a feature space using nonlinear functions. This mapping is done by $\varphi(x)$, which is a nonlinear mapping of x. The output of RBF-NN is a combination of weighted kernels which is given as follows [17]

$$y = \sum_{i=1}^{m} w_i \varphi_i(x) \tag{4}$$

where w_i is the weight value of i-th neuron in the RBF-NN. Moreover, the following Gaussian function is defined for the activation function

$$\varphi_i(x) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp^{\{-(x-c_i)^T(x-c_i)/2\sigma_i^2\}} \tag{5}$$

where c_i and σ_i are vectors of the center and radius of the activation function of the RBF-NN, respectively.

Likewise, the MLP, the RBF-NN model used in this study has six inputs, and one output. The trained RBF-NN is obtained with maximum 1200 neurons of the feature space.

III. Result

In this section, the simulation results of obtained home load models using MLP and RBF-NN are presented. The case study is a single-family detached house located in the arctic climate in Northern Norway. The main devices that exist in the building are listed in section II, part A.

All the simulations are done in MATLAB R2020a, Neural Network Toolbox.

A. Comparison results of different methods

In this part, the same input dataset including outdoor temperature, wind speed, user pattern, time of day, weekday, and month is inputted into the networks to compare their performance on the modeling of the electrical load of the home. In both methods, the arrangement of data is changed to allocate data to each category (training, validation and test data) randomly. In MLP algorithm, data is divided into three parts; 70% for training, 15% for validation and 15% for test. In RBF-NN technique, 70% is allocated for training and 30% for test.

The home load modeling performance of the different proposed methods is assessed by three widely used metrics including MSE, RMSE and regression (R) that is correlation between ANN outputs and actual values.

$$MSE = \frac{\sum_{k=1}^{N} (T_k - Y_k)^2}{N} \tag{6}$$

$$RMSE = \sqrt{MSE} \tag{7}$$

where T_k is the actual value, Y_k is the output of the model, and N is size of data.

Fig. 5 shows one week of the test data. As shown in Fig. 5 (a), the black line is the output of the MLP network, and the red line is the actual value of electricity consumption.The blue line in Fig. 5 (b) is the error between output and target. As it is evident in Fig. 5 (a) and (b), the model obtained by MLP is successfully tracking the actual values.

Furthermore, Fig. 6 demonstrates the real load and modeled load by the RBF-NN method on a random range equivalent to a one-week measured of test data. Likewise, the MLP, the developed model by RBF-NN is able to follow targets trajectory.

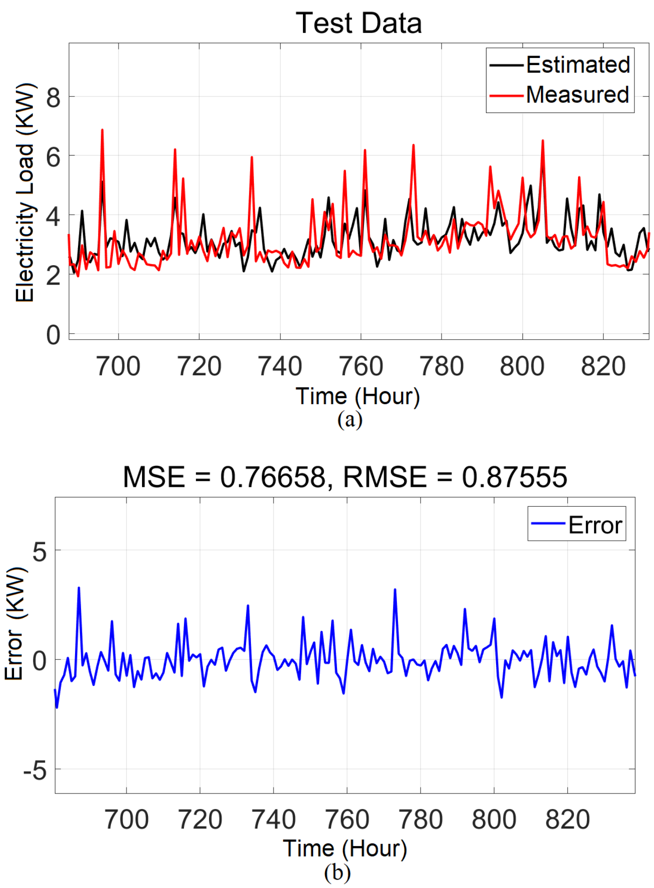

Fig. 5. One week of the test data; a) Target and output of the MLP with one hidden layer; b) Error

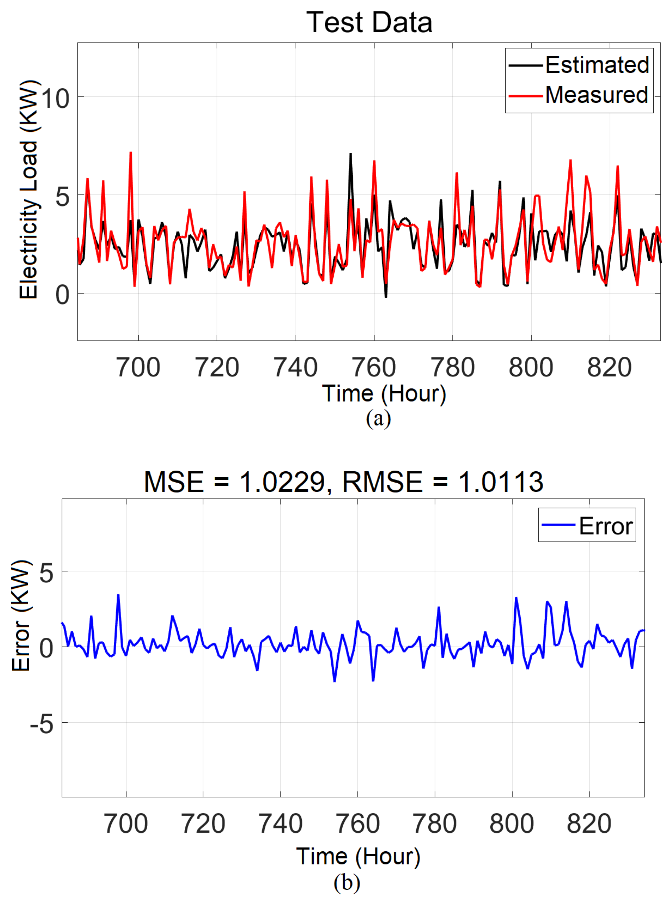

Fig. 6. One week of the test data; a) Target and output of the RBF-NN; b) Error

TABLE I
ERROR AND REGRESSION FOR DIFFERENT METHODS

	Train data			**Test data**		
	MSE	**RMSE**	**R**	**MSE**	**RMSE**	**R**
MLP	0.52	0.72	0.85	0.76	0.87	0.79
RBF-NN	0.32	0.56	0.91	1.02	1.01	0.70

For better understanding, the results related to MLP and RBF-NN performance are summarized in Table I. As shown in Table I, RBF-NN has higher accuracy in training data.

In contrast to training data, when comparing the MSE, RMSE and R values of the test data, the MLP method has a better modeling performance with lower error values than the other method according to Table I.

B. Impact of different input datasets on modeling

In this section, four different input datasets are considered to investigate the impact of each of them on the modeling of electric load consumption.

In the first dataset, all the inputs including temperature, wind and holiday are applied to the system. In the second dataset, wind is not considered as an input in order to check its impact on the modeling of load consumption. In the third dataset, holiday is not one of the inputs to check how much impact it has if the occupant behavior is removed and the last dataset is without considering temperature as an input.

All these datasets are applied to both MLP and RBF-NN networks, and the performance of systems with different input datasets are shown in Table II and Table III.

When comparing different results in Table II and Table III, it is obvious that wind speed has the lowest impact on the load modeling, because there is no significant difference between metrics of Case 1 and Case 2. Consequently wind speed as one of the inputs of ANN does not have a significant impact on the accuracy of load modeling. This result might differ from building to building, depending on how airtight the building is. On the other side, temperature and holiday inputs have the highest impact on the load modeling, since when they are not considered as inputs, there are considerable difference between the metrics of them for Case 1. That means for better performance in home load modeling using ANN methods, temperature and user behavior are two important factors that should be considered as NN inputs. The

TABLE II
ERROR AND REGRESSION FOR DIFFERENT DATASETS INPUT IN MLP METHOD (T: TEMPERATURE; W: WIND; H: HOLIDAY)

MLP		**Train data**			**Test data**		
		MSE	**RMSE**	**R**	**MSE**	**RMSE**	**R**
Case 1	T, W, H	0.52	0.72	0.85	0.76	0.87	0.79
Case 2	H, T	0.53	0.72	0.85	0.79	0.89	0.77
Case 3	T, W	0.59	0.77	0.83	0.99	0.99	0.71
Case 4	H, W	0.62	0.78	0.82	0.95	0.97	0.75

TABLE III
ERROR AND REGRESSION FOR DIFFERENT DATASETS INPUT IN RBF-NN METHOD (T: TEMPERATURE; W: WIND; H: HOLIDAY)

RBF-NN		**Train data**			**Test data**		
		MSE	**RMSE**	**R**	**MSE**	**RMSE**	**R**
Case 1	T, W, H	0.32	0.56	0.91	1.02	1.01	0.70
Case 2	H, T	0.38	0.62	0.89	1.10	1.04	0.70
Case 3	T, W	0.39	0.62	0.89	1.26	1.12	0.63
Case 4	H, W	0.57	0.75	0.84	1.20	1.10	0.65

effects of temperature and user behavior on the electricity consumption are also seen in Fig. 1. During the months when the temperature is low, the monthly average load is higher than annual average load, and when the occupant is on a vacation, the power consumption is much lower than normal situations.

IV. CONCLUSION

In this paper, a comparison between MLP and RBF-NN networks for home load modeling has been presented. Detailed analysis of the simulation results illustrated that both ANN methods have acceptable performance in the case of home load modeling application, but in comparison to each other MLP excels because of its higher accuracy in test data. The comparative analysis results for different input datasets showed that wind speed has the lowest impact on the accuracy of load modeling and temperature and occupant behavior have the highest effect on the efficiency of the model. The results were presented for a single-family home in Arctic climate condition, and more measurements are needed to verify if the same result can be expected on the majority of buildings.

REFERENCES

[1] Y. Wang, Q. Chen, T. Hong, and C. Kang, “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges,” IEEE Trans. Smart Grid, vol. 10, no. 3, pp. 3125–3148, 2019.

[2] D. Arif, A., Wang, Z., Wang, J., Mather, B., Bashualdo, H. and Zhao, “Load modeling—A review,” IEEE Trans. Smart Grid, vol. 9, no. 6, pp. 5986–5999, 2018.

[3] J. W. Lee, S.H., Son, S.E., Lee, S.M., Cho, J.M., Song, K.B. and Park, “Kalman-Filter Based Static Load Modeling of Real Power System Using K-EMS Data,” J. Electr. Eng. Technol., vol. 7, no. 3, pp. 304–311, 2012.

[4] D. J. Renmu, H., Jin, M. and Hill, “Composite load modeling via measurement approach,” IEEE Trans. power Syst., vol. 21, no. 2, pp. 663–672, 2006.

[5] B. N. Hu F, Sun K, Del Rosso A, Farantatos E, “Measurement-based real-time voltage stability monitoring for load areas,” IEEE Trans. Power Syst., vol. 31, no. 4, pp. 2787–2798, 2015.

[6] I. F. Visconti, D. A. Lima, J. M. C. de Sousa Costa, and N. R. de BC Sobrinho, “Measurement-based load modeling using transfer functions for dynamic simulations,” IEEE Trans. Power Syst., vol. 29, no. 1, pp. 111–120, 2013.

[7] G. K. Kontis, E. O., Papadopoulos, T. A., Chrysochos and A. I., Papagiannis, “Measurement-based dynamic load modeling using the vector fitting technique,” IEEE Trans. Power Syst., vol. 33, no. 1, pp. 338–351, 2017

[8] M. Yousefi, N. Kianpoor, A. Hajizadeh, and M. Soltani, “Smart Energy Management System for Residential Homes Regarding Uncertainties of Photovoltaic Array and Plug-in Electric Vehicle,” in IEEE International Symposium on Industrial Electronics, 2019.

[9] Y. J. Chang, G.W., Chen, C.I. and Liu, “A neural-network-based method of modeling electric arc furnace load for power engineering study,” IEEE Trans. Power Syst., vol. 25, no. 1, pp. 138–146, 2009.

[10] M. Yousefi, A. Hajizadeh, and M. N. Soltani, “A comparison study on stochastic modeling methods for home energy management systems,” IEEE Trans. Ind. Informatics, vol. 15, no. 8, pp. 4799–4808, 2019.

[11] A. H. Abdelaziz, A. Y., Badr, M. A. L., and Younes, “Dynamic load modeling of an Egyptian primary distribution system using neural networks,” Int. J. Electr. Power Energy Syst., vol. 29, no. 9, pp. 637–649, 2007.

[12] A. Keyhani, W. Lu, and G. T. Heydt, “Composite neural network load models for power system stability analysis,” in In IEEE PES Power Systems Conference and Exposition, 2004, pp. 1159–1163.

[13] K. Hafez, O., and Bhattacharya, “Integrating EV charging stations as smart loads for demand response provisions in distribution systems,” IEEE Trans. Smart Grid, vol. 9, no. 2, pp. 1096–1106, 2016.

[14] A. Engelbrecht, Computational intelligence: an introduction, 2007.

[15] M. N. Du, K. L., and Swamy, Neural Networks in a Softcomputing Framework, 2006.

[16] R. C. Hippert, H. S., Pedreira, C. E., and Souza, “Neural networks for short-term load forecasting: A review and evaluation,” IEEE Trans. power Syst., vol. 16, no. 1, pp. 44–55, 2001.

[17] B. M. Cecati, C., Kolbusz, J., Różycki, P., Siano, P., and Wilamowski, “A novel RBF training algorithm for short-term electric load forecasting and comparative studies,” IEEE Trans. Ind. Electron., vol. 62, no. 10, pp. 6519–6529, 2015.