Evaluation of Daily Electric Load Forecasting Algorithms

Saeid Asghari
Electrical Engineering Department, Shahed University, Tehran, IRAN
and S. Seyedtabaii
Electrical Engineering Department, Shahed University, Tehran, IRAN
stabaii@shahed.ac.ir

Abstract
Optimal power system operation requires accurate load demand prediction; so that the generation, transmission, and distribution utilities to operate securely and efficiently. In this paper, several load forecasting algorithms are detailed and their performances are compared. The objective is to use a short length training frame which is more appropriate for short term load forecasting. The problems with each forecasting algorithm is input selection, data preprocessing, feature extraction, prediction network, and training algorithms. In this respect, Empirical Mode Decomposition, neural network, and classic and intelligent training algorithms are arranged together for devising the best set up. Various configurations aiming at best algorithms are employed and simulation results are provided.

Keywords—Load forecasting; Artificial neural network; EMD.

Introduction

Load forecasting plays a vital role in the generation scheduling, system reliability, and power optimization, and economical running of the smart grid. Due to the explosion in the restructuring of power markets within a deregulated economy, competitive power market needs to minimize their required generation reserve gaps. Efficient load forecasting for future demands can reduce the gap, which will help in economic power generation, power operations, power construction planning, and power distribution.

The main reason lying behind load forecasting is economical (Lago, De Ridder et al. 2018). With the increase in load forecasting accuracy, the negative impact on the economy is reduced. In (Gao, Darvishan et al. 2019), it is declared that a 1% Reduction of Mean Absolute Percentage Error (MAPE) can decrease the generation costs from about 0.1% to 0.3%. As the load demand is dependent on some factors like weather conditions, economy, and calendar, it has a highly non-linear and random behavior. The electricity market players need the exact load forecast to optimize utilities and make profits (Singh and Dwivedi 2018).

Based on the prediction time, it can be divided into three types, which are short-term, medium-term, and long-term forecasting. The short- term load forecast (STLF) represents the electric load forecast for a time interval of a few hours to a few days. Short time load forecasting (STLF) provides an accurate load demand for controlling and scheduling the power system. Electricity load forecasting methods can be broadly classified into four categories: (1) Statistical models, (2) knowledge-based expert systems, (3) hybrid models, and (4) artificial intelligence-based models.

To accurately forecast the load demand appropriate load signal features have to be extracted by signal preprocessing. The most commonly used feature extractor in this field is Empirical mode decomposition (EMD). Following feature extraction, a nonlinear prediction model is required to forecast the load demand in a specific time, which due to its non-linear characteristics, neural network models are the best choice.

In (Zhang, Wei et al. 2018), empirical mode decomposition, autoregressive integrated moving average (ARIMA), and wavelet neural network (WNN) optimized by fruit fly optimization algorithm (FOA) are proposed for load forecasting. In (Qiu, Ren et al. 2017), EMD, and deep belief learning NN classifier is applied to one month of data. Application of EMD, the modified generalized regression neural network (GRNN) NSMOCS-GRNN, and Cuckoo search algorithm has been investigated in (Olegario, Coronel et al. 2019). Using denoising and SVR is the subject of study in (Jiang, Zhang et al. 2018) for accurate load forecasting.

Train of NN over data from a very nonlinear process faces trapping in local minima. To avoid such a problem employing intelligent search for nonlinear optimization are common. There are a variety of algorithms such as the Hyper-Spherical Search method (Rahmatian and Seyedtabaii 2019) and the genetic algorithm used in (Ahour, Seyedtabaii et al. 2017).

In this paper, several load forecasting algorithms using a short length training frame are compared. Input choices are the hour of the day, day of the week, holiday/weekend indicator, 168-h (Previous week) lagged load, 24-h Lagged load, previous 24-h average load. Besides, separately predicted temperature and dew point data are also tested. The input preprocessing is conducted and EMD is also tried for feature extraction. The predictor is a neural network that is trained using both classical and intelligent methods. By rearranging the setups, improvements in the results are shown which looks pretty acceptable.;

The remaining of this paper is structured as follows:

First, in Section 2, the empirical mode decomposition and artificial neural network are briefly described. Then, in Section 3, signal preprocessing algorithms are presented. The forecasting methods and the simulation results are discussed in Section 4 and lastly, the conclusion comes in Section 5.

Theoretical Background

Empirical mode decomposition

Empirical Mode Decomposition is an adaptive and efficient non-linear and non-stationary time-frequency data analysis method. It is based on the simple assumption that any signal consists of different subtasks. Non-linear and non-stationary time series can be decomposed into a group of zero mean and quasi-periodic signals, where each component is called IMF and a residue component. The detailed process of EMD is explained in the following steps (Yaslan and Bican 2017).

1- With a given time series signal x(t), create its upper and lower envelopes by a cubic-spline interpolation of local maxima and minima.
2- Calculate the mean of the upper and lower envelopes m1.
3- Subtract the mean from the original time series to obtain the first component $h(t) = x(t) - m(t)$.
4- Repeat steps 1 to 3 by considering h(t) as new x(t) until one of the following stopping criteria is satisfied: i) m(t) approaches zero, ii) the numbers of zero-crossings and extrema of h(t) differs at most by one, or iii) the predefined maximum iteration is reached.
5- Treat h(t) as an IMF and compute residue signal: r(t) = x(t) - h(t).
6- Use the residual signal r(t) as new x(t) to find the next IMF. Repeat steps 1 to 5 until all IMFs are obtained. Finally, the original TS signal is decomposed as:

$$x(t) = \sum_{i=1}^{n} c(i) + r_n \tag{1}$$

Where the number of functions n in the set depends on the original TS signal.

Artificial neural network

Neural networks as a nonlinear black-box model are employed for very complex classification (Qaedi and Seyedtabaii 2012) (Seyedtabaii 2012) and highly nonlinear detection tasks (Seyedtabaii 2012). Thus, it has a strong capacity to be used in load forecasting cases.

The network used has three fundamental layers: an input layer with the same number of neurons as the dimension of input features; a hidden layer comprised of neurons with nonlinear activation function; and an output layer which aggregates the outputs from the hidden layer neurons. The final output, y, and the hidden layer output are expressed by the following equations:

$$y = g\left(\sum_{j=1}^{h} w_{jo} v_j + b_j\right),\ \ v_j = f\left(\sum_{i=1}^{n} w_{ij} x_i + b_i\right) \tag{2}$$

where x_i is the input to the neuron; f() and g() are nonlinear activation functions; vj is the output of hidden layer neuron j; y is the output of this short time load forecasting network (SLFN); n and h are the number of input features and the number of the hidden layer neurons, respectively; w_{ij} is the weight of the connection between the input variable i and the neuron j of the hidden layer; w_{jo} is the weight of the connection between the hidden layer neuron j and the output; b_i and b_j are the biases.

The signal preparation

In this section the test data set and the signal preprocessing algorithms used are discussed.

Data set

Various datasets are available for electricity load forecasting among them is the New England version. Fig. 1 shows a randomly selected nine weeks of it.

Outlier existence check

The outliers are data points located far outside the range of the majority of the data such as glitches, data-entry errors, and inaccurate measurements can produce outliers in real data samples. The outliers can significantly affect the analysis of data samples. If we suspect that the data that we want to analyze contains outliers, we can discard the outliers or replace them with the values typical for that data sample. Before we discard or replace the outliers, we try to verify that they are actual errors. The outliers can be a part of the correct data sample, and discarding them can lead to incorrect conclusions. In this paper, an outlier is a value that is more than three scaled median absolute deviations (MAD) away from the median. For a random variable vector A made up of N scalar observations, the median absolute deviation (MAD) is defined as:

$$MAD = median\left(\left|A_i - median(A)\right|\right),\ i = 1,2,\ldots,N \tag{3}$$

The median value is the mean of the middle two numbers in sorted order.

The center value is the median of the data, and the upper and lower thresholds are three scaled MAD above and below the median. As shown in Fig. 2, the dataset has no outliers and all single points are within the confidence lanes.

Normalization

To eliminate the excessively deviating instances, a mapping algorithm is used to map the load profile L to a small L' range. The algorithm is given by

$$K_{pre} = \left(L'_{\max} - L'_{\min}\right)/\left(L_{\max} - L_{\min}\right)$$
$$L'_i = K_{pre}(L_i - L_{\min}) \tag{4}$$

Where K_{pre} is a mapping ratio from the original load L to the small range L'. L_{max} and L_{min} indicate the maximum value and minimum value of L, respectively. L'_{max} and L'_{min} indicate the maximum value and minimum value of L', respectively. In this paper, L' is normalized to the range (0,1), which means L'_{max} and L'_{min} are set to 1 and 0, respectively. With the inverse function of the mapping algorithm (1) and (2), the forecasting result $\hat{L}$ can be computed.

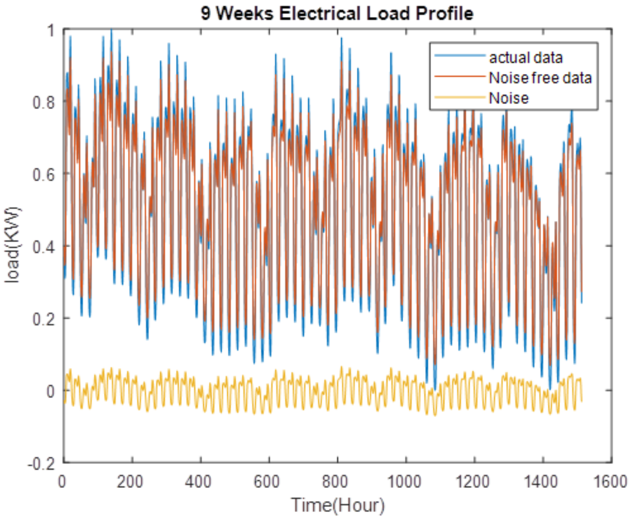

Fig. 1. **A sample load signal**

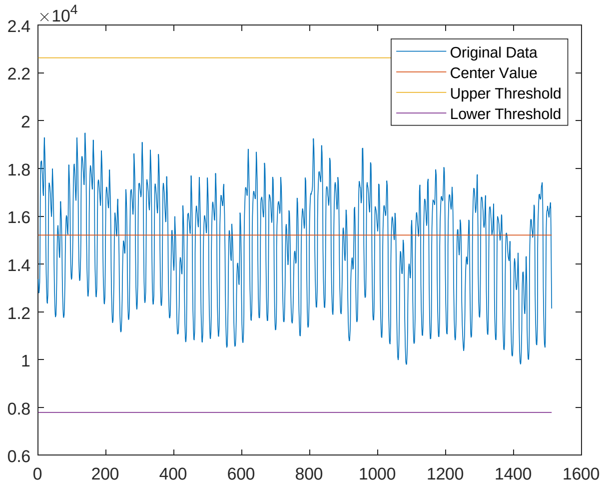

Fig. 2. **The original signal and the threshold values.**

Signal dependency analysis

The correlation coefficient between load signal and independent variables to be assigned as input to the forecasting neural network is presented in Table 1 where there is a strong correlation between 24-h lagged load and the load. The other important signal in this respect is 168-h (Previous week) lagged load following by the previous 24-h average load. The hour of the day and holiday/weekend are the other effective signals. Among the important signals is the temperature which is predicted by the weather forecasting system. Other factors show weak links with the load pattern.

Considering the autocorrelation and partial autocorrelation functions, it seems that the maximum values are correlated to the day before, the last week, and two weeks ago as it is shown in Fig. 3.

Algorithms efficiency

The algorithm's efficiency is measured through mean absolute percentage error (MAPE) given below,

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y'_t - y_t}{y_t}\right| \tag{6}$$

Where y_t' is the predicted value of the corresponding y_t, and n is the number of data points in the testing time series.

TABLE I. The correlation coefficients between load signal and other variables

Input parameter	Correlation coefficient (r)
Dry bulb temperature	0.19
Dew point temperature	0.07
Hour of the day	0.51
Day of the week	0.03
Holiday/weekend indicator (0 or 1)	0.26
168-h (Previous week) lagged load	0.85
24-h Lagged load	0.90
Previous 24-h average load	0.56

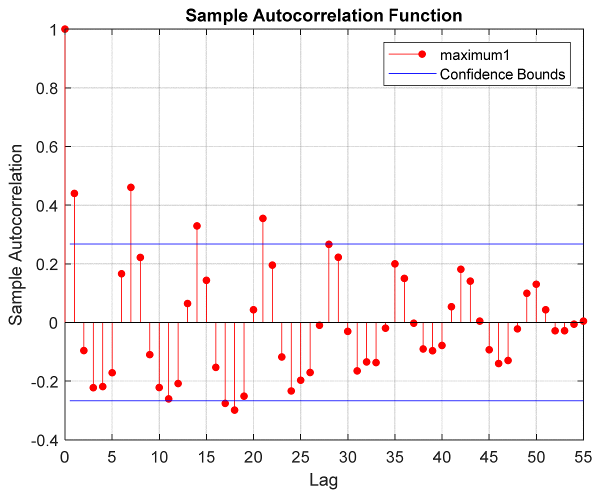

Fig. 3. **The daily peak values autocorrelation plot**

Methods and results

In this section, several algorithms are detailed and their performances are compared. The following input data are common to all tests while some may additional inputs be injected to.

- d1: Hour of the day;
- d2: Day of the week;
- d3: Holiday/weekend indicator (0 or 1);
- d4: 168-h (Previous week) lagged load;
- d5: 24-h Lagged load;
- d6: Previous 24-h average load;

Algorithm 1.

In the first test algorithm, the inputs are d1 to d6 plus the prediction of the

- dry-bulb temperature;
- Dew point temperature;

which is provided by the weather forecasting systems as it is used in (Singh and Dwivedi 2018). Four years of data are considered for training and from two other years, randomly, the 15th week is set aside for the test. Prediction is conducted using a network with 6 nodes. The result of the days of the week load forecasting has been depicted in Fig. 4. The best MAPE in 10 rounds of the algorithm execution is 1.79%, with the average MAPE value of 1.84%.

Algorithm 2.

In the second algorithm, a shorter length of data which is more appropriate for a short term load forecasting is utilized. Eight weeks of data before the 15th week is assigned for training and the 15th week for the test. The inputs are the same as Algorithm 1, including d1 to d6 plus the prediction of the

- dry-bulb temperature;
- Dew point temperature;

A network of 6 nodes is used for prediction. The outcomes have been shown in Fig. 5. The best MAPE equal to 3.1598% and mean MAPE of 3.71% after 10 runs are the algorithm performance; which is worse than the result of Algorithm 1. The weakness is due to the shorter length of the training algorithm. By incorporating appropriate steps the results despite shorter training frame will be improved.

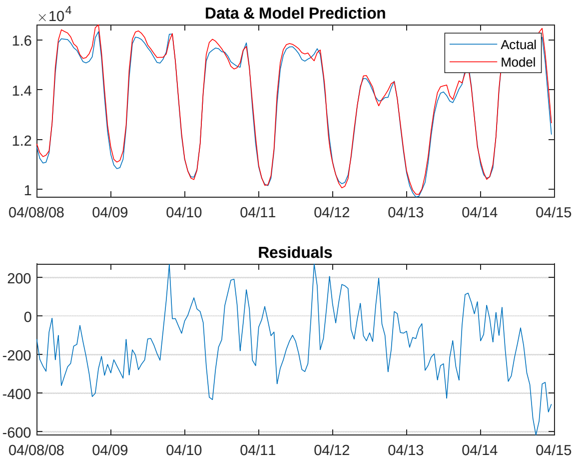

Fig. 4. **Days of a week load prediction by Algorithm 1.**

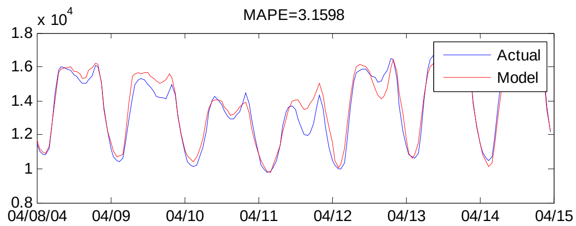

Fig. 5. **Days of a week load prediction by Algorithm 2.**

Algorithm 3.
In this algorithm, the prediction of the temperature and the dew point are discarded and their roles are allocated to the network to handle it. So, just the d1 to d6 inputs feed the prediction network, and again a network of 6 nodes is used.
The number of the training set is 8*7*24 and the number of the test set is 1*7*24. The results have been exhibited in Fig. 6. The best MAPE of 3.78% and an average of 4.75% after 10 runs are the algorithm consequences.
Applying this algorithm to the 18th week of the dataset renders different results as it has been expected. The best result is 2.3415 while the average of 10 runs is 2.6143
By normalizing the data using

$$ndata=(x-\bar{x})/max(abs(x-\bar{x}))$$

and executing the algorithm improvement in the results is accomplished. The best MAPE now reaches 3.27 and the average of MAPE after 10 runs reduces to 4.38. The result has been shown in Fig. 7.

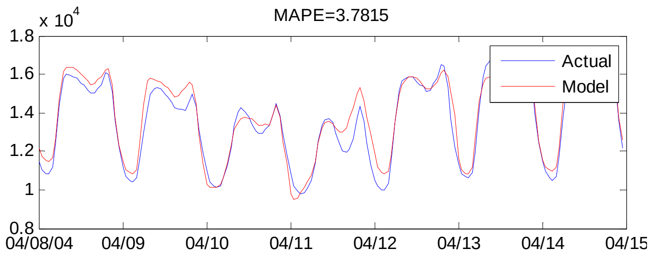

Fig. 6. **Days of a week load prediction by Algorithm 3.**

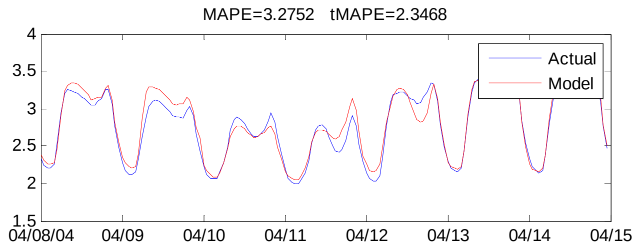

Fig. 7. **Days of a week load prediction by normalization and Algorithm 3.**

Algorithm 4.

In this algorithm, d1 to d6 are used for input and they are normalized.
Eight weeks before the 15th week (8*7*24) is used for training and the 15th week is used for a test similar to the other algorithms.
The mfile forecasting program is linked to the MATLAB Optimtool where its genetic algorithm program trains the 6-node network. The training is supervised and during its course, as it is possible, different mutation and crossover methods are assigned. Following such provisions, the load forecasting merit of MAPE=1.5646 is obtained. The result of forecasting and the prediction error has been depicted in Fig. 8.

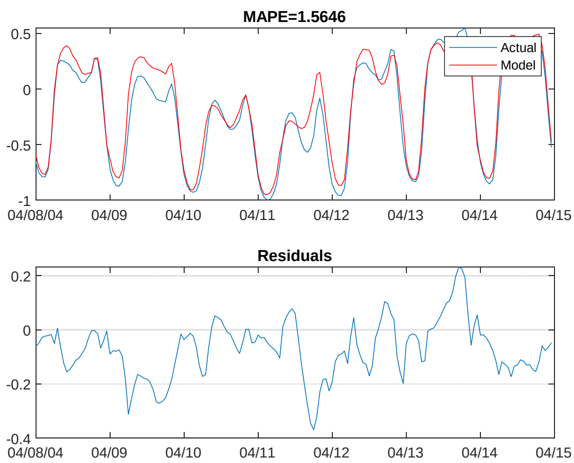

Fig. 8. **Days of a week load prediction and prediction error by Algorithm 4.**

Algorithm 5.

In this algorithm, the denoised inputs of d1 and d6 are used and the temperature and the dew point predictions are waived.
The denoising is conducted using EMD (Yaslan and Bican 2017). Generally, in EMD denoising one of the IMF components is discarded, but there is no consensus on the subject. Recently, in (Premanode, Vongprasert et al. 2013), a new digital filter is presented and showed that the Average

IMF (aIMF) of a dataset is normally distributed. Based on the Pearson chi-square, Kolmogorov-Smirnov and Anderson-Darling normality tests, aIMF contains the white Gaussian noise characteristics more than all IMFs and corresponds to the highest noise in the data. So, for noise removal, aIMF is subtracted from the original dataset as defined below,

$$aIMF = \frac{1}{N}\sum_{i=1}^{N} IMF_i(t)$$
$$X_N(t) = X(t) - aIMF(t) \qquad (5)$$

Where $X_N(t)$ is the filtered data sample; aIMF is the Average IMF that corresponds to noise and $X(t)$ is the original signal sample at time t.
The filtered signal feeds a NN predictor of 6 nodes. Again, eight weeks are for training and one week for the test is allocated. Now, the average result of 10 runs stands at MAPE=1.8422. Fig. 9 shows the result and prediction error.

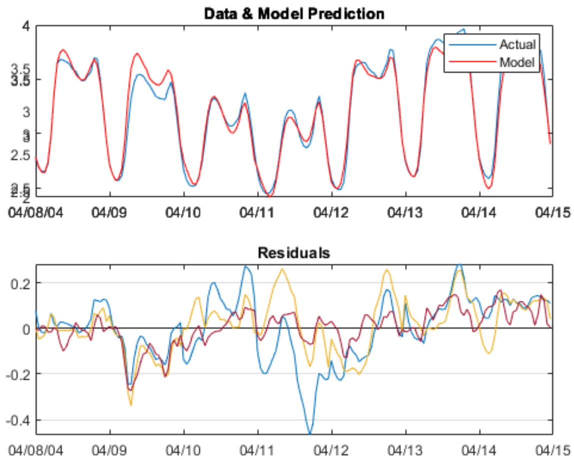

Fig. 9. **Days of a week load prediction and prediction error by Algorithm 5.**

TABLE I. **Comparison of the proposed method with other methods**

Method	Error(MAPE)
Algorithm 1: 2 years training+temp.+Dew point	1.84%
Algorithm 2:8 weaks training+temp.+Dew point	3.71%
Algorithm 3:8 weaks training	4.38%
Algorithm 4: 8 weaks training+ GA trainer	1.56%
Algorithm 5: 8 weaks training+EMD denoising	1.84%

Conclusion

In this paper, the various algorithms for load forecasting using a short frame of data for training are devised and tested. The result indicates that the temperature and the dewpoint improve the results, but these are themselves are the result of a weather forecasting system. In this respect, the application of EMD for denoising and feature extraction are assessed. Application of the intelligent neural network training evaluated. In the end, algorithms are introduced that despite using a short training frame provide acceptable results.

Acknowledgment

This work has been partially supported by the research department of Shahed University, Tehran, IRAN.

References

Ahour, J. N., S. Seyedtabaii and G. B. Gharehpetian (2017). "Determination and localisation of turn-to-turn fault in transformer winding using frequency response analysis." IET Science, Measurement & Technology **12**(3): 291-300.

Gao, W., A. Darvishan, M. Toghani, M. Mohammadi, O. Abedinia and N. Ghadimi (2019). "Different states of multi-block based forecast engine for price and load prediction." International Journal of Electrical Power & Energy Systems **104**: 423-435.

Jiang, H., Y. Zhang, E. Muljadi, J. J. Zhang and D. W. Gao (2018). "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization." IEEE Transactions on Smart Grid **9**(4): 3341-3350.

Lago, J., F. De Ridder, P. Vrancx and B. De Schutter (2018). "Forecasting day-ahead electricity prices in Europe: the importance of considering market integration." Applied Energy **211**: 890-903.

Olegario, C. C., A. D. Coronel, B. D. Gerardo and R. P. Medina (2019). "Towards An Enhanced Backpropagation Network for Short-Term Load Demand Forecasting." International Journal of Electrical Engineering and Applied Sciences (IJEEAS) **2**(1): 65-70.

Premanode, B., J. Vongprasert and C. Toumazou (2013). "Prediction of exchange rates using averaging intrinsic mode function and multiclass support vector regression." Artif. Intell. Research **2**(2): 47-61.

Qaedi, S. and S. Seyedtabaii (2012). "Improvement in power transformer intelligent dissolved gas analysis method." World Academy of Science, Engineering and Technology **61**: 1144-1147.

Qiu, X., Y. Ren, P. N. Suganthan and G. A. Amaratunga (2017). "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting." Applied Soft Computing **54**: 246-255.

Rahmatian, M. and S. Seyedtabaii (2019). "Multi-machine optimal power system stabilizers design based on system stability and nonlinearity indices using Hyper-Spherical Search method." International Journal of Electrical Power & Energy Systems **105**: 729-740.

Seyedtabaii, S. (2012). "Improvement in the performance of neural network-based power transmission line fault classifiers." IET generation, transmission & distribution **6**(8): 731-737.

Seyedtabaii, S. (2012). "Performance evaluation of neural network based pulse-echo weld defect classifiers." Measurement Science Review **12**(5): 168-174.

Singh, P. and P. Dwivedi (2018). "Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem." Applied energy **217**: 537-549.

Yaslan, Y. and B. Bican (2017). "Empirical mode decomposition based denoising method with support vector regression for time series prediction: a case study for electricity load forecasting." Measurement **103**: 52-61.

Zhang, J., Y.-M. Wei, D. Li, Z. Tan and J. Zhou (2018). "Short term electricity load forecasting using a hybrid model." Energy **158**: 774-781.