

Probabilistic Load Forecasting using an Improved Wavelet Neural Network Trained by Generalized Extreme Learning Machine

Mehdi Rafiei, Taher Niknam, *Member, IEEE*, Jamshid Aghaei, *Senior Member, IEEE*, Miadreza Shafie-khah, *Senior Member, IEEE*, and João P. S. Catalão, *Senior Member, IEEE*

Abstract—Competitive transactions resulting from recent restructuring of the electricity market, have made achieving a precise and reliable load forecasting, especially probabilistic load forecasting, an important topic. Hence, this paper presents a novel hybrid method of probabilistic electricity load forecasting, including generalized extreme learning machine (GELM) for training an improved wavelet neural network (IWNN), wavelet preprocessing and bootstrapping. In the proposed method, the forecasting model and data noise uncertainties are taken into account while the output of the model is the load probabilistic interval. In order to validate the method, it is implemented on the Ontario and Australian electricity markets data. Also, in order to remove the influence of model parameters and data on performance validation, Friedman and post-hoc tests, which are non-parametric tests, are applied to the proposed method. The results demonstrate the high performance, accuracy and reliability of the proposed method.

Index Terms—Probabilistic Forecasting, Improved Wavelet Neural Network, Generalized Extreme Learning Machine, Bootstrapping, Wavelet Processing.

I. INTRODUCTION

IN the last decades, the structure of the electricity market has changed a lot, forming the restructured market. In this market, it is essential to access a reliable and accurate load forecasting considering some other activities such as economic dispatch, bidding strategies and unit commitment. However, demand is becoming significantly active and less predictable due to the various demand response programs, distributed energy sources and emerging technologies [1]. As a result, load forecasting plays a very significant role in decision-making activities for market participants.

The majority of the studies in the field of load forecasting focus on the point forecasting techniques.

J.P.S. Catalão acknowledges the support by FEDER funds through COMPETE 2020 and by Portuguese funds through FCT, under Projects SAICT-PAC/0004/2015 - POCI-01-0145-FEDER-016434, POCI-01-0145-FEDER-006961, UID/EEA/50014/2013, UID/CEC/50021/2013, UID/EMS/00151/2013, and 02/SAICT/2017 - 029803, and also funding from the EU 7th Framework Programme FP7/2007-2013 under GA no. 309048.

M. Rafiei and T. Niknam are with the Department of Electrical and Electronics Engineering, Shiraz University of Technology, Shiraz, Iran (e-mails: m_rafiei@ymail.com; niknam@sutech.ac.ir).

J. Aghaei is with the Department of Electrical and Electronics Engineering, Shiraz University of Technology, Shiraz, Iran, also with the Department of Electric Power Engineering, Norwegian University of Science and Technology (NTNU), Trondheim NO-7491, Norway (e-mail: aghaei@sutech.ac.ir).

M. Shafie-khah is with C-MAST, University of Beira Interior, Covilhã 6201-001, Portugal (e-mail: miadreza@gmail.com).

J.P.S. Catalão is with INESC TEC and the Faculty of Engineering of the University of Porto, Porto 4200-465, Portugal, also with C-MAST, University of Beira Interior, Covilhã 6201-001, Portugal, and also with INESC-ID, Instituto Superior Técnico, University of Lisbon, Lisbon 1049-001, Portugal (e-mail: catalao@ubi.pt).

However, the results are not reliable because of the fluctuations' existence in load and structure of the electricity market. Point forecasting has done some statistical techniques such as exponential smoothing models [2], regression [3] and time series [4].

Also, the forecasting type has implemented some artificial intelligence techniques such as neural networks [5], support vector machines [6] and fuzzy systems [7]. In [8], a multiple time series equation model, based on frequent use of first-order least squares, is represented to forecast the load, and its superiority is compared with non-linear and non-parametric methods.

Recurrent extreme learning machine with high accuracy is proposed in [5] as a new method for forecasting the load. This method has been used to train the single-layer recurrent neural network. In [9], a new design based on type II fuzzy logic system is proposed and applied to the active learning theory to forecast the electrical load.

Recently, based on the increased market competition, aging infrastructure, renewable integration requirements and the more active and less predictive electricity market, the market participants are interested in using the probabilistic load forecasting, which provides additional information on the variability and uncertainty of electricity load series in comparison with point forecasting technique. Also, the probabilistic load forecasting is needed in some processes such as stochastic unit commitment [10,11], reliability planning [12] and probabilistic load flow [13].

One form of achieving the probabilistic load forecasting is the prediction intervals (PIs). In this method, the forecasting is based on the point forecasting and the error obtained by uncertainties [14]. In [15], semi-parametric regression model, multivariate time series simulation model and resampling strategy are used for point load forecasting, temperature forecasting and building probabilistic forecasting, respectively. In [16], the quantile regression averaging method on a set of forecasted points is presented to predict the probability. The advantage of this method is the ability to use point forecasting. In [17], a diffused heteroscedastic forecasting method based on the Gaussian process in daily load forecasting is proposed.

Another method which is defined and used in this area is deep learning, which is presented in [18, 19]. The aim of such researches on deep learning theory is to show its ability in different areas. But, still there are two common issues about the theory, which are overfitting and computation time.

The strategy of [20] consists of producing temperature scenarios for probabilistic load forecasting which is performed by fitting a quantile generalized additive model. More articles on this topic are discussed in [14].

It is logical to select the neural network for load forecasting, due to the wide use of neural networks in load forecasting and their extrapolation and estimation ability compared to other methods. In this article, a combination of generalized extreme learning machine (GELM), wavelet processing and bootstrapping is used to forecast the electricity demand.

Indeed, GELM is utilized to train the improved wavelet neural networks (IWNNs), where the benefits of this neural network are more profitable, compared to the common neural networks. This method utilizes the high-speed advantages of ELM-based methods compared to other methods.

Moreover, the wavelet preprocessing is applied to the load data, dividing it into the well-behaved subseries. Then, the forecasting method is individually applied to each subset according to its own characteristics which leads to an improved forecasting model. The bootstrapping technique is used to obtain the uncertainties and PIs. The data noise and model forecasting uncertainties are considered as well. As a result, a hybrid method is achieved which leads to a load forecasting with high reliability, accuracy and speed. The high accuracy and speed of the proposed method lead to its practical applications in the electricity market. Finally, the proposed method is tested on the Ontario and Australian markets while its performance is acceptable and validated as well.

The contributions of this paper are summarized as follows:

- (1) the usage of IWNN and GELM in the probabilistic load forecasting problem;
- (2) considering data noise uncertainties which are a result of non-iterative training machine;
- (3) proposing the combination of GELM, wavelet preprocessing and bootstrapping;
- (4) the usage of Friedman and post-hoc tests to validate the approach.

The rest of this paper is organized as follows: Section II explains the Wavelet preprocessing, Improved WNN and generalized ELM. In section III, the concept of PI, the way to consider uncertainties and calculate PIs, and PIs validation indexes are presented. The implementation of the proposed method is shown in section IV. Section V discusses applying the proposed method on a case study based on real data, while Section VI provides the conclusions.

II. METHODOLOGY

A. Wavelet Preprocessing

One of the most significant tools for frequency component analyzing is the wavelet transform. Indeed, wavelet functions divide information into different parts with different frequency attributes, and each part will be analyzed with similar resolution scale. Here, the wavelet transform is utilized to divide electricity demand series into a set of well-behaved productive subseries.

This method is used for probabilistic electricity price forecasting in [21], and its performance is proved. Forecasting is done separately on productive subseries.

Then, by inverse wavelet transform, the final forecasted load is produced. Dividing coefficients for load series are regulated according to [22] as follows:

$$p_{RL}^W = 2^{-\left(\frac{R}{2}\right)} \sum_{t=0}^{T-1} p_t W\left(\frac{t-L \cdot 2^R}{2^R}\right) = 2^{-\left(\frac{R}{2}\right)} \sum_{t=0}^{T-1} p_t W_{RL}(t) \quad (1)$$

where $W(\cdot)$ is the picked wavelet function, p_t is the amount of load at t , T is the length of series, and p_{RL}^W is the coefficient for resolution level R and position L .

Preferably, orthogonal wavelet function is selected due to its appropriate mathematical property. Hence, the approximation sets, $A_R(R=1,2,\dots,R^*)$, and details sets, $D_R(R=1,2,\dots,R^*)$, are defined by orthogonal wavelet functions.

Finally, the basic load series $p_t(t=1,2,\dots,T)$ can be reproduced as follows:

$$p_t = D_1 + \dots + D_R + A_R \quad (2)$$

In such problems, the two most common and appropriate wavelet functions are Daubechies and Morlet. The performance of these two functions in probabilistic load forecasting is compared and the related results are shown in section V. Based on the results, Daubechies wavelet function has better efficiency and is used here. Daubechies wavelet offers an appropriate trade-off between wavelength and smoothness, resulting in an appropriate behavior for the load forecast and has been used in lots of excellent research in load and price forecasting [23].

B. Improved Wavelet Neural Networks (IWNNs)

One appropriate tool for load forecasting is wavelet which is the activation function of neural network. Indeed, in comparison with classic feed-forward neural networks, the coefficients and implementation of activation function on inputs have a different structure. Compared to conventional activation functions, WNNs have a high capability of generalization [24]. In this paper, Morlet wavelet is utilized as the hidden layer of the neurons' activation function in WNN. In addition to the hidden layer neurons, inputs are connected directly to outputs by some coefficients to ameliorate the WNN's performance and structure. Consequently, the structure can profit from the capabilities of wavelet functions and can also take the load signal's tendencies.

Fig. 1 shows an IWNN, where x_i and y are inputs and target variables, respectively.

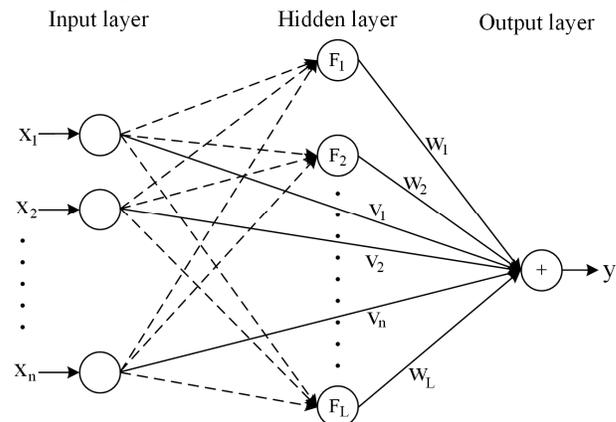


Fig 1: Structure of an IWNN

In the hidden layer, neurons' activation functions are determined as follows:

$$F_i(x_1, x_2, \dots, x_n) = \prod_{j=1}^n \tau_{a_i, b_i}(x_j) \quad \forall i=1, 2, \dots, L \quad (3)$$

$$\tau_{a_i, b_i}(x_j) = \tau\left(\frac{x_j - b_i}{a_i}\right) \quad (4)$$

where L is the number of hidden layer neurons, and $\tau(x)$ is Morlet function, which is defined as follows:

$$\tau(x) = e^{-0.5x^2} \cos(5x) \quad (5)$$

where a_i and b_i are Morlet function's shift and scale coefficients, respectively.

Finally, the neural network's output is represented as follows:

$$y = \sum_{i=1}^L w_i F_i(x_1, x_2, \dots, x_n) + \sum_{j=1}^n v_j x_j \quad (6)$$

in which, w_i is the i^{th} hidden neuron's coefficient, and v_j is the j^{th} input's coefficient.

Based on previous equations, it's obvious that the number of the IWNN's free parameters is equal to $n_p=3L+n$ and their vector can be shown as follows:

$$Z = [v_1, \dots, v_n, w_1, \dots, w_L, a_1, \dots, a_L, b_1, \dots, b_L] \quad (7)$$

C. Generalized ELM (GELM)

As a fast training method for the single layer feed-forward neural networks (SLFNs), the extreme learning machine is presented in [25, 26].

In ELM, biases of hidden layer and input's coefficients are picked randomly, and output's coefficients are achieved by matrix generalized inverse transformation. Here, GELM is used to train the IWNNs.

Based on the generalization, Morlet functions' shift and scale coefficients are picked out randomly, and then the hidden layer's coefficients and also inputs' coefficients are taken by matrix inverse.

It's assumed that there are N different training data cases $\{(x_i, t_i)\}_{i=1}^N$, in which inputs and outputs vectors are $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]$, respectively.

An IWNN with activation function $F(\cdot)$ and L hidden neurons should be considered, which can estimate N cases with zero error, so:

$$Y_j(x_j) = \sum_{i=1}^L (F_i(x_j) w_i) + \sum_{i=1}^n (x_j v_i) = t_j \quad j = 1, 2, \dots, N \quad (8)$$

In the above-mentioned equation, $w_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T$ is the i^{th} hidden neuron's coefficients vector and $v_i = [v_{i1}, v_{i2}, \dots, v_{im}]^T$ is the i^{th} input's coefficients vector. Also, $F_i(x_j)$ is the i^{th} hidden neuron's output by applying the j^{th} input. The equation (8) can be rewritten in matrix form as follows:

$$FW + VX = T \quad (9)$$

where:

$$F = \begin{bmatrix} F_1(a_1, b_1, x_1) & F_2(a_2, b_2, x_1) & \cdots & F_L(a_L, b_L, x_1) \\ F_1(a_1, b_1, x_2) & F_2(a_2, b_2, x_2) & \cdots & F_L(a_L, b_L, x_2) \\ \vdots & \vdots & \ddots & \vdots \\ F_1(a_1, b_1, x_N) & F_2(a_2, b_2, x_N) & \cdots & F_L(a_L, b_L, x_N) \end{bmatrix}_{N \times L} \quad (10)$$

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{L1} & w_{L2} & \cdots & w_{Lm} \end{bmatrix}_{L \times m} = \begin{bmatrix} w_1^T \\ w_2^T \\ \vdots \\ w_L^T \end{bmatrix} \quad (11)$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nn} \end{bmatrix}_{N \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad (12)$$

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nm} \end{bmatrix}_{n \times m} = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} \quad (13)$$

$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1m} \\ t_{21} & t_{22} & \cdots & t_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ t_{N1} & t_{N2} & \cdots & t_{Nm} \end{bmatrix}_{N \times m} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix} \quad (14)$$

Matrix F 's i^{th} row represents the outputs of hidden neuron, while the i^{th} training data pair is applied as input. By defining equations (15), (9) can be rewritten as (16).

$$H = [F \parallel X]_{N \times (L+n)}, \quad \beta = \begin{bmatrix} W \\ V \end{bmatrix}_{(L+n) \times m} \quad (15)$$

$$H\beta = T \quad (16)$$

In the training process, coefficients a_i and b_i are taken randomly. Consequently, matrices F and H are defined, and β will be the matrix of under training variables. Also, matrix T is a definite matrix, then by a generalized inverse, matrix β can be easily calculated as follows:

$$\beta^* = H^\dagger T \quad (17)$$

where H^\dagger is the Moore-Penrose generalized inverse of matrix H .

As a result, ELM has significant benefits compared to the traditional approaches. This is mainly because traditional approaches include several iterations with high computational cost, while considering simple matrix calculations, ELM is a high speed approach.

In addition, it is shown that ELM has a better performance in comparison with traditional approaches [26]. The benefits of GELM compared to the basic ELM are presented in section V as well.

III. PREDICTION INTERVALS

A. PIs Formulation

Uncertainties associated with the forecasting model are the main agents that cause uncertainty in load forecasting. The uncertainty befalls because of the incorrect structure and the parameters of neural networks.

Another reason for the forecasting uncertainty is the noise of training data. Indeed, the random behavior of data regression causes this type of uncertainty.

If a set of different data pairs $\{(x_i, t_i)\}_{i=1}^N$ is considered, the forecasting target would be expressed as follows:

$$t_i = r(x_i) + \varepsilon(x_i) \quad (18)$$

In the equation above, t_i is the i^{th} forecasting target, x_i is the vector of inputs, $\varepsilon(x_i)$ represents noise, which has zero average, and $r(x_i)$ is the true regression average.

Practically, the learned neural network $\hat{r}(x_i)$ can be considered as an estimator of true regression $r(x_i)$. Therefore, the forecasting error can be described as follows:

$$t_i - \hat{r}(x_i) = [r(x_i) - \hat{r}(x_i)] + \varepsilon(x_i) \quad (19)$$

term $t_i - \hat{r}(x_i)$ shows the sum of forecasting errors, and $r(x_i) - \hat{r}(x_i)$ represents neural network's estimation error based on true regression.

Taking the assumption that the noise and the forecasting errors are statistically independent, the total forecasting error variance $\hat{\sigma}_t^2$ is equal to the summation of variances of the uncertainty model and data noise as the following equation:

$$\hat{\sigma}_t^2(x_i) = \hat{\sigma}_r^2(x_i) + \hat{\sigma}_\varepsilon^2(x_i) \quad (20)$$

Also, assuming that the time series $\{(x_i, t_i)\}$ and $\hat{\sigma}_t^2(x_i)$ as the total variance of uncertainty, a PI of t_i with $100(1-\delta)\%$ confidence level is given as $[L_r^{(\delta)}(x_i), U_r^{(\delta)}(x_i)]$, where $L_r^{(\delta)}(x_i)$ and $U_r^{(\delta)}(x_i)$ are lower and upper limits of PI, which are described by following equations:

$$\begin{cases} L_r^{(\delta)}(x_i) = \hat{r}(x_i) - z_{1-\delta/2} \sqrt{\hat{\sigma}_t^2(x_i)} \\ U_r^{(\delta)}(x_i) = \hat{r}(x_i) + z_{1-\delta/2} \sqrt{\hat{\sigma}_t^2(x_i)} \end{cases} \quad (21)$$

where $z_{1-\delta/2}$ is the standard normal distribution critical value and it depends on a given confidence level. It is expected that t_i is in the structure of PI with $100(1-\delta)\%$ nominal possibility, so that equation (22) can be concluded as follows:

$$P\left(t_i \in [L_r^{(\delta)}(x_i), U_r^{(\delta)}(x_i)]\right) = 100(1-\delta)\% \quad (22)$$

B. Bootstrapping technique

Bootstrapping technique is employed to model the forecasting model uncertainties. In fact, bootstrapping is a powerful tool to analyze statistical inference processes [27]. Therefore, the model uncertainty is estimated by bootstrapping that can be illustrated as follows:

Step 1: for training the data $\{(x_i, t_i)\}$, $\hat{r}(\cdot)$ is obtained. (Using GELM)

Step 2: the error between the forecasting result and the main target is calculated $\hat{e} = t_i - \hat{r}(x_i)$.

Step 3: the center of obtained error is transferred to zero $\tilde{e}_i = \hat{e}_i - \left(\frac{\sum_i \hat{e}_i}{N}\right)$.

Step 4: new error replaces the previous error. New targets are obtained based on $\tilde{t}_i = \hat{r}(x_i) + \tilde{e}(x_i)$, and new training data is produced in the form of $\{(x_i, \tilde{t}_i)\}$.

Step 5: for new data, $\hat{r}_q(x_i)$ is forecasted in q^{th} bootstrapping iteration.

Step 6: steps 2 to 5 are repeated for B replacements and generation of the new data.

In all of the above equations, $i=1,2,\dots,N$. Finally, new learning data is produced B times, and each time, neural networks are trained, and the demand is forecasted. Afterward, using (23) the average of neural network outputs is calculated. In this equation, $\hat{r}_q(x_i)$ represents the forecasted value corresponding to the q^{th} iteration.

$$\hat{r}(x_i) = \left(\sum_q \hat{r}_q(x_i)\right) / B \quad (23)$$

Eventually, the variance of the model uncertainty is attained as the following.

$$\hat{\sigma}_r^2(x_i) = \left(\sum_q (\hat{r}_q(x_i) - \hat{r}(x_i))^2\right) / (B-1) \quad (24)$$

C. Noise variance

In traditional neural networks, there was no concern about data noise due to the disappearing effect of the noise by repetition. However, in non-repetitive methods such as GELM, there is a concern about data noise. To solve the problem, noise is considered as an uncertainty. Therefore, it is essential to obtain the variance of noise data to model the noise of the data. Thus, (20) can be taken into account as follows:

$$\hat{\sigma}_\varepsilon^2 = E[(t - \hat{r})^2] - \hat{\sigma}_r^2 \quad (25)$$

To estimate the forecasted model with the fitting extra residual goal, the remaining squares can be defined in the following form:

$$R^2(x_i) = \max([t_i - \hat{r}(x_i)]^2 - \hat{\sigma}_r^2(x_i), 0) \quad (26)$$

Now, new sets of trained data with the input x_i and remaining squares output are defined as $\{(x_i, R^2(x_i))\}$, $i=1,2,\dots,N$ while the output of the trained neural network is always positive. Therefore, the variance noise data can be achieved by these trained data. As a result, for considering the uncertainty model, which depends on data noise, a separate neural network with the new data should be trained (according to (20)) while its output should be added to the related variance of the uncertainty forecasting model.

D. PIs evaluation

In this study, to systematically validate the PI, the main indexes for PI evaluation are offered in both reliability and resolution aspects.

Reliability: The key index for the estimated PI validation is reliability. According to the definition of PI, it is expected that t_i is in the PI's limits with the nominal probability $100(1-\delta)\%$, which is called prediction intervals' nominal confidence (PINC). Based on N_t test samples, the real coverage probability of the relating PI is illustrated by the prediction intervals coverage probability (PICP), and its definition is as follows:

$$PICP = \frac{1}{N_t} \sum_{i=1}^{N_t} I_i^{(\delta)} \quad (27)$$

where $I_i^{(\delta)}$ is the PICP's index and is obtained as follows:

$$I_i^{(\delta)} = \begin{cases} 1, & t_i \in [L_r^{(\delta)}(x_i), U_r^{(\delta)}(x_i)] \\ 0, & t_i \notin [L_r^{(\delta)}(x_i), U_r^{(\delta)}(x_i)] \end{cases} \quad (28)$$

The PI's reliability degree is directly represented by PICP. To have a high reliability, the obtained PIs should asymptotically reach the related PINC. Then, the average coverage error (ACE), which is defined as the difference between PICP and PINC, can be calculated and used for evaluating the reliability of PI.

$$ACE = PICP - PINC \quad (29)$$

As it mentioned, the value of ACE should be close to zero as much as possible.

Resolution: High reliability PIs can easily be obtained by increasing wideness of intervals. However, it's not a logical strategy for practical applications. Hence, the evaluation criterion of resolution (ECR), which represents the sharpness of intervals, is applied to calculate the average of PIs. The criterion expresses the ability of the method in focusing the uncertainties. ECR is obtained as follows:

$$ECR = \left(\sum_{i=1}^N [U_r^{(\delta)}(x_i) - L_r^{(\delta)}(x_i)] \right) / N_i \quad (30)$$

Smaller ECR shows narrower intervals and PIs with better performance. An efficient probabilistic forecasting has to provide both the desired reliability and acceptable resolution.

IV. IMPLEMENTATION

Based on (26), to obtain the data noise uncertainty, first, the load forecasting and the variance of the model uncertainty should be determined.

Fig. 2 shows the flowchart of the probabilistic load forecasting method. According to the flowchart, in the beginning, the training data should be selected based on the forecasting period. In this paper, the objective is forecasting the demand for a period of one week as an hourly prediction. The hourly load data for a year before the under forecasting week are used as training data. In order to test and validate the system performance, the one-year data are divided into three subseries as training, validation and test data. These subseries are used as neural network training, parameter regulation and performance evaluation of the model.

Using all of the one-year data, in addition to the complexity of the required neural network, can lead to increasing the computational load and forecasting time, while reducing the accuracy as well. The reasons for the accuracy reduction are the data that have no mutual relation with load forecasting data. Therefore, the auto correlation function (ACF) is used to reduce the complexity and computational burden and increase the accuracy of forecasting models by reducing irrelevant data.

In the next step, the data is applied to the wavelet transform by Daubechies function and divides them into an approximate subseries A1 and three detail subseries D1, D2 and D3. These four subseries are used to train neural networks, as they have been forecasted. Consequently, the four outputs of the neural network are returned to the original load state by the inverse wavelet transform.

As mentioned before, to consider the uncertainty of the forecasting model, a bootstrapping is used which requires B neural networks.

Therefore, based on these two issues, 4*B neural networks are needed to obtain the forecasting model uncertainty. The neural network related to the data noise is trained, using the initial data and the output of the forecasting model uncertainty section.

According to the Fig. 3, overall, 4B+1 neural networks are needed for the probabilistic hourly load forecasting.

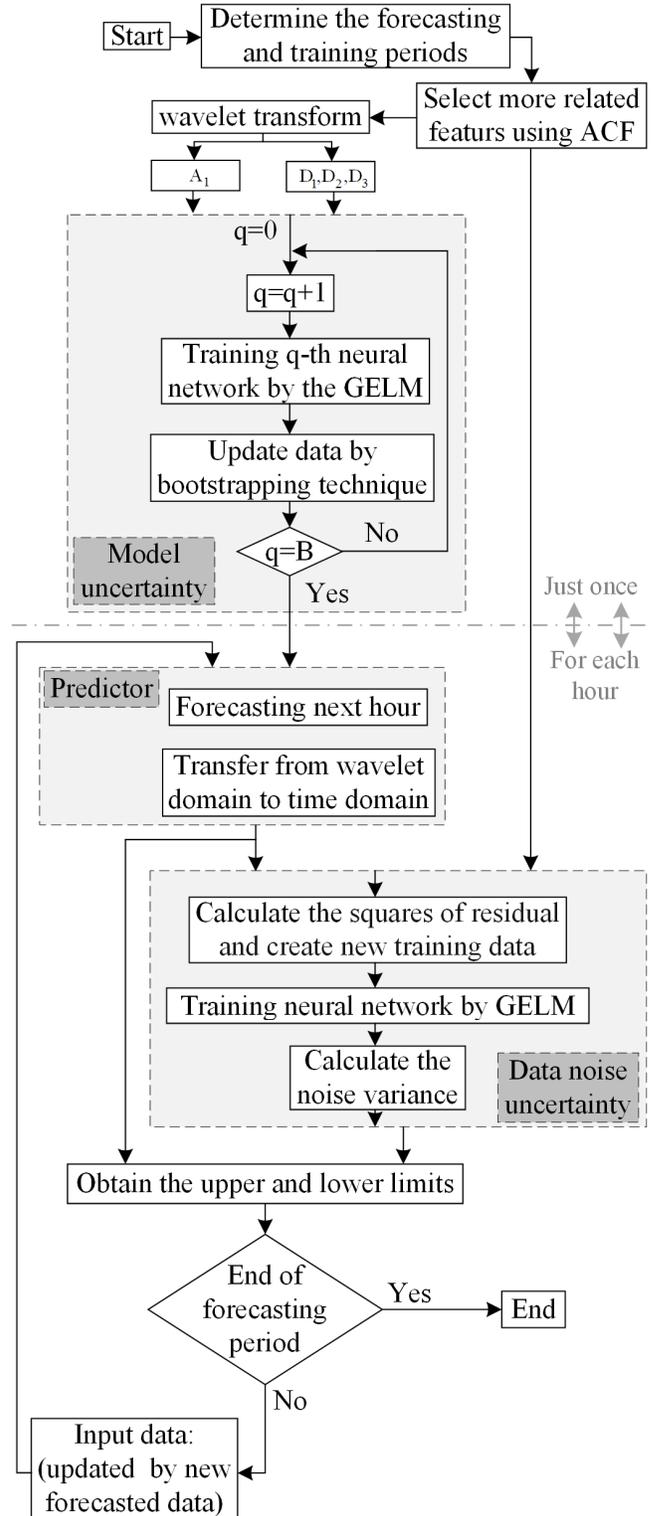


Fig. 2. Flowchart of the proposed method.

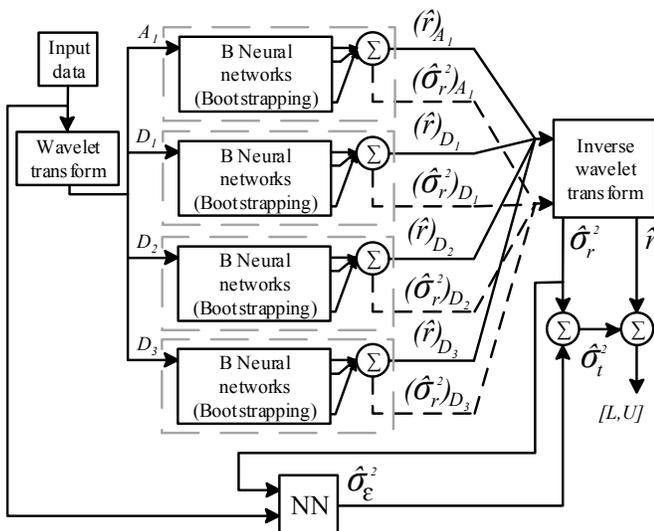


Fig. 3. The performance of all 4B+1 N

V. CASE STUDY AND RESULTS

To evaluate the performance of the proposed method, it has been tested in two different cases. In both cases, the real market data are considered. In the first case, the objective is to analyze the performance and feasibility of the tools which are represented in the proposed method. The goal is to analyze the feasibility of GELM, wavelet preprocessing and Daubechies function. Then, the proposed method is compared with three other methods which in the first one, the wavelet processing is not applied (GELM + bootstrapping) while in the second one; instead of GELM, common ELM and NN are employed (Wavelet preprocessing + ELM + bootstrapping), and in the last one, the Morlet function is used as the wavelet function in wavelet preprocessing. Table I shows the results of this case. In the second case, the advantages of the proposed method are compared with four different methods. The features of these four methods are as follows:

A. First method

In this method, as proposed in [28], the conventional two-layer neural network is trained with back propagation (BP) for forecasting the load. Both model and data noise uncertainties are considered, while bootstrapping is used to obtain the PIs.

B. Second method

In this method, the improved Dolphins optimization algorithm in [29] is used for learning two-layer neural network while both model and data noise uncertainties are considered. Also, bootstrapping is used to obtain the PIs.

C. Third method

In this method, a learning technique based on RELM in [5] is used to train the recurrent neural network. The uncertainty and PIs are as same as other methods.

D. Fourth method

In this method, the quantile regression averaging method [16] on a set of family forecasted points is used to calculate the PIs. Several neural network based forecasts with different variables are used for point forecasts, and their results are used in the regression.

All these four methods have been simulated and applied on real data. As mentioned before, the first two methods and the fourth one use multi-layer neural networks.

However, neural network methods which are employed in the proposed and the third methods are single-layered.

In general, the multi-layer networks have high capabilities of modeling the systems, compared to the single layer ones. However, the proposed and the third methods have much higher speed. Therefore, it is more comfortable and unrestricted to use them on probability calculations problems which need to run frequently. The objective has overcome the limitations of single-layer network with the capabilities of the wavelet neural networks, its improvements and the performance of GELM. Also, it has tried to obtain high forecasting performance and probabilistic calculations as well. The comparisons of the methods are given in Tables II to V.

In previous cases, the results obtained based on limited data sets and comparisons were parametric. Accordingly, it is normal that even a better approach shows a weaker result in some especial situations in these cases.

TABLE I:
VALIDATION OF THE TOOLS (ONTARIO-DEC. 2014)

Method	PINC	PICP	ACE	ECR	Time/Time of proposed method
Without wavelet preprocessing	90%	88.43%	-1.57%	1032.3	0.3
	95%	92.37%	-2.63%	1434.7	
Basic ELM and SLFN	90%	90.72%	0.72%	910.8	~1
	95%	93.97%	-1.03%	1193.6	
With Morlet wavelet function	90%	88.61%	-1.39%	981.4	~1
	95%	94.07%	-0.93%	11205.4	
Proposed method	90%	90.48%	0.48%	869.8	1
	95%	95.83%	0.83%	1159.8	

TABLE II:
PERFORMANCE COMPARISON OF THE METHODS (ONTARIO-Aug. 2014)

Method	PINC	PICP	ACE	ECR
BP + Bootstrapping [28]	90%	86.75%	-3.25%	830.6
Dolphin + Bootstrapping [29]		91.54%	1.45%	613.1
RELM + Bootstrapping [5]		89.17%	-0.83%	650
Quantile Regression [16]		87.76%	-2.24%	663.7
Proposed method		88.69%	-1.31%	614.9
BP + Bootstrapping [28]	95%	93.91%	-1.09%	983.8
Dolphin + Bootstrapping [29]		94.24%	-0.76%	828.2
RELM + Bootstrapping [5]		95.81%	0.81%	845.5
Quantile Regression [16]		94.45%	-0.55%	832.8
Proposed method		95.24%	0.24%	785.7

TABLE III:
PERFORMANCE COMPARISON OF THE METHODS (ONTARIO-Dec. 2014)

Method	PINC	PICP	ACE	ECR
BP + Bootstrapping [28]	90%	91.03%	1.03%	1023.5
Dolphin + Bootstrapping [29]		90.78%	0.78%	934.6
RELM + Bootstrapping [5]		89.21%	-0.79%	944.1
Quantile Regression [16]		89.51%	-0.49%	938.7
Proposed method		90.48%	0.48%	869.8
BP + Bootstrapping [28]	95%	93.57%	-1.43%	1242.8
Dolphin + Bootstrapping [29]		93.89%	-1.11%	1183.7
RELM + Bootstrapping [5]		94.04%	-0.96%	1212.5
Quantile Regression [16]		93.89%	-1.11%	1168.6
Proposed method		95.83%	0.83%	1159.8

TABLE IV:
PERFORMANCE COMPARISON OF THE METHODS (AUSTRALIA-May. 2014)

Method	PINC	PICP	ACE	ECR
BP + Bootstrapping [28]	90%	88.27%	-1.73%	326.1
Dolphin + Bootstrapping [29]		90.76%	0.76%	340.7
RELM + Bootstrapping [5]		91.03%	1.03%	313.6
Quantile Regression [16]		90.83%	0.83%	306.5
Proposed method		89.29%	-0.71%	292.5
BP + Bootstrapping [28]	95%	96.43%	1.43%	397.0
Dolphin + Bootstrapping [29]		94.14%	-0.86%	381.5
RELM + Bootstrapping [5]		96.17%	1.17%	357.6
Quantile Regression [16]		95.83%	0.83%	395.1
Proposed method		95.83%	0.83%	377.8

TABLE V:
PERFORMANCE COMPARISON OF THE METHODS (AUSTRALIA-Oct. 2014)

Method	PINC	PICP	ACE	ECR
BP + Bootstrapping [28]	90%	91.02%	1.02%	250.6
Dolphin + Bootstrapping [29]		90.57%	0.57%	203.0
RELM + Bootstrapping [5]		89.67%	-0.33%	195.5
Quantile Regression [16]		90.86%	0.86%	199.3
Proposed method		89.88%	-0.12%	193.1
BP + Bootstrapping [28]	95%	93.41%	-1.59%	295.4
Dolphin + Bootstrapping [29]		96.01%	1.01%	224.6
RELM + Bootstrapping [5]		95.87%	0.87%	265.3
Quantile Regression [16]		93.83%	-1.17%	231.3
Proposed method		94.05%	-0.95%	227.7

TABLE VI:
COMPARISON OF ACE OF THE METHODS BASED ON FRIEDMAN TEST

Method Data set (2014)	BP + Bootstrapping [28]		Dolphin algorithm + Bootstrapping [29]		RELM + Bootstrapping [5]		Quantile Regression [16]		Proposed method	
	ACE	rank	ACE	rank	ACE	rank	ACE	rank	ACE	rank
Ontario-Feb.	2.46%	(5)	1.31%	(3)	1.42%	(4)	1.22%	(2)	1.03%	(1)
Ontario-May.	1.25%	(2)	1.54%	(4)	1.65%	(5)	1.32%	(3)	0.97%	(1)
Ontario-Aug.	1.18%	(5)	0.54%	(1)	0.98%	(3)	1.05%	(4)	0.63%	(2)
Ontario-Oct.	1.21%	(4)	0.65%	(2)	2.02%	(5)	1.12%	(3)	0.37%	(1)
Ontario-Dec.	0.73%	(4)	1.73%	(5)	0.45%	(1.5)	0.45%	(1.5)	0.67%	(3)
Australia-Feb.	1.65%	(3)	0.48%	(1)	2.37%	(4)	2.43%	(5)	1.02%	(2)
Australia-May.	1.65%	(2.5)	1.65%	(2.5)	1.82%	(4)	1.91%	(5)	0.87%	(1)
Australia-Aug.	1.05%	(2)	1.24%	(5)	0.93%	(1)	1.21%	(4)	1.17%	(3)
Australia-Oct.	1.32%	(3,5)	2.22%	(5)	1.32%	(3.5)	1.05%	(2)	0.54%	(1)
Australia-Dec.	1.7%	(5)	1.47%	(4)	1.32%	(3)	1.03%	(1.5)	1.03%	(1.5)
Total rank		3.6		3.25		3.4		3.1		1.65

Hence, three non-parametric test methods, like Friedman, Games-Howell and Tukey HSD methods were used while the Games-Howell and Tukey HSD methods were Post-hoc tests. As Friedman test, the neural networks are trained by ten various data sets and all of the five methods in the second case and ACE and ECR criterions are calculated as well. Indeed, for each data set and training method, both criterions achieved

three different PINCs and five different initial parameters, which have contributed to using the total average. Then the utilized methods are ranked for each data set while the best method obtains the rank of 1, and the worst one obtains rank 5.

Moreover, in methods with similar performances, the ranking is done in such a way that the average stays 3. In the next step, each method's total rank, which is equal to the average of the method's ranks, is calculated. Then, the method with the lowest total rank has the best performance in comparison with other ones. Comparing the five methods using Friedman test, the amount of critical difference (CD) is equal to 1.21 [30]. Thus, if the difference between two methods' total ranks is more than CD, the lower rank method would have significantly better performance than the other one. Further details of the test are available in [30].

Tables VI and VII illustrate the results of the test based on ACE and ECR criteria, respectively. For Tukey HSD and Games-Howell tests, the p-values of the proposed method and four others have been calculated as well. In these tests, the level of significance (α) is equal to 0.05. It should be noted that, according to the post-hoc tests, two methods are significantly different if the p-value between them is under α . Table VIII represents the results of these tests based on ACE and ECR criteria.

The data of Ontario's electricity market (OEM) and Australian electricity market (AEMO) are obtained from [31, 32]. Based on Table I, using wavelet processing can increase the forecasting period by 3 times while the accuracy of forecasting can significantly increase. However, it can overlook this time based on the very high speed of ELM methods compared to the conventional methods. Also, it is seen that using ELM method and SLFN neural networks has no significant influence on the forecasting time. But, these methods slightly decrease the accuracy.

Finally, the usage of Morlet wavelet function instead of Daubechies is shown that, the Morlet function slightly decreases the accuracy without any perceptible effect on time. Consequently, the results proved that the tools which are used in the proposed method had a positive impact on the accuracy.

TABLE VII:
COMPARISON OF ECR OF THE METHODS BASED ON FRIEDMAN TEST

Method Data set (2014)	BP + Bootstrapping [28]		Dolphin algorithm + Bootstrapping [29]		RELM + Bootstrapping [5]		Quantile Regression [16]		Proposed method	
	ECR	rank	ECR	rank	ECR	rank	ECR	rank	ECR	rank
Ontario-Feb.	892.8	(4)	771.9	(2)	1003.8	(5)	792.3	(3)	683.3	(1)
Ontario-May.	1202.3	(5)	813.8	(2)	934.6	(3)	941.1	(4)	712.0	(1)
Ontario-Aug.	1044.7	(4)	819.2	(1.5)	1074.1	(5)	819.2	(1.5)	823.9	(3)
Ontario-Oct.	833.5	(2)	1192.5	(5)	819.2	(1)	955.1	(4)	928.3	(3)
Ontario-Dec.	885.7	(4)	860.1	(3)	942.8	(5)	820.4	(1)	828.2	(2)
Australia-Feb.	143.1	(1)	295.3	(4)	292.5	(3)	303.6	(5)	250.6	(2)
Australia-May.	312.7	(5)	244.6	(2.5)	297.0	(4)	244.6	(2.5)	203.0	(1)
Australia-Aug.	313.6	(5)	235.3	(2)	306.5	(3)	310.3	(4)	195.5	(1)
Australia-Oct.	347.0	(5)	333.6	(3)	243.2	(2)	337.7	(4)	193.1	(1)
Australia-Dec.	281.5	(2)	304.2	(5)	228.0	(1)	301.9	(4)	295.4	(3)
Total rank		3.7		3		3.2		3.3		1.8

TABLE VIII:
P-VALUES BETWEEN THE METHODS BASED ON POST-HOC TESTS

Proposed method vs.	p-value of Tukey HSD		p-value of Games- Howell	
	ACE	ECR	ACE	ECR
BP + Bootstrapping [28]	0.014	0.008	0.016	0.009
Dolphin algorithm + Bootstrapping [29]	0.027	0.074	0.027	0.081
RELM + Bootstrapping [5]	0.09	0.031	0.092	0.033
Quantile Regression [16]	0.045	0.036	0.044	0.034

Tables II to V demonstrate the high performance of the proposed method, compared to the other methods. Also, in the result of the proposed method, the value of PICP is very close to the value of PINC, and thus, the value of ACE is very small. In all of the tests, the absolute value of ACE is not more than 1.4%. In addition, the value of ECR in the proposed method is less than other methods.

According to Tables VI and VII, compared to BP, RELM and Quantile Regression methods, the difference between the final rankings of the proposed method, is more than CD which can lead to the conclusion that the proposed method is significantly superior to the mentioned methods. Also, the difference between the final ranking of the proposed method and the improved Dolphin algorithm is not more than CD in ACE criteria, however, the difference is quite close to CD.

Table VIII demonstrates the comparison of the BP and Quantile Regression, in which the p-values of the proposed method are less than α in both ACE and ECR criteria. Therefore, the proposed method is more useful than the BP and Quantile Regression methods. Also, in both tests, the p-values of the proposed method are less than other methods such as the improved Dolphin algorithm (from the ACE perspective) and RELM (from the ECR perspective) which is less than α as well. Consequently, the proposed method has rejected the improved Dolphin algorithm from the ACE perspective and RELM from the ECR perspective. However, from the ACE perspective, the p-values of the proposed method and RELM are very close to α .

In addition to what is specified in the tables, the time factor should be checked as well. This is mainly due to the fact that the main goal is to reach probabilistic forecasting and many program executions are required to calculate the variance. The calculation time of the BP, improved Dolphin and Quantile Regression methods are hundred times more than the proposed method. Although, the performance on BP and improved Dolphin methods are good in some cases, in many other cases are impractical and lose their application because of the high computational load.

Therefore, the proposed method has a limitation on the number of layers. However, the high ability of the network structure and training method has led to the superiority and significantly applicability of the proposed method. Overall, by developing ELM methods (such as the suggested GELM method) and using them in multi-layer neural networks, a high speed and efficient method can be achieved.

Figures 4 to 7 show the graphical results of the load probabilistic forecasting. The first two figures demonstrate the Ontario electricity market in December and August 2014. Also, the next two figures show the Australian electricity market in May and October 2014, respectively.

Each figure contains actual and forecasting load information related to the one week hourly demand.

The forecasting results are represented for both PICP=90% and PICP=95%, respectively.

Both the numerical and graphical results have demonstrated higher accuracy and efficiency of the proposed method in forecasting followed by a series of demand with high reliability.

Flexibility, accuracy and reliability of the proposed method proved the effectiveness and acceptable performance in practical applications of the load forecasting.

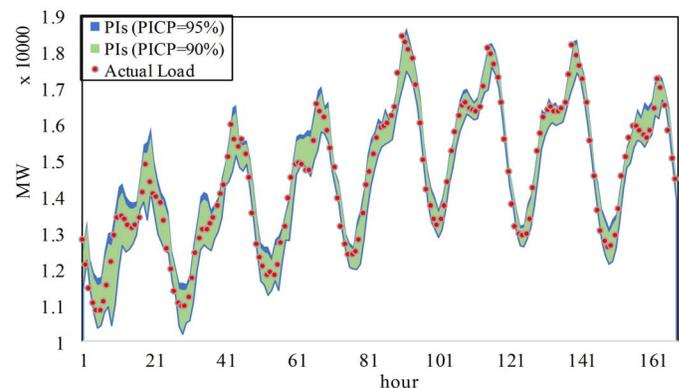


Fig. 4. Load PIs by the proposed method with confidence levels 90% and 95% (Ontario-Dec. 2014).

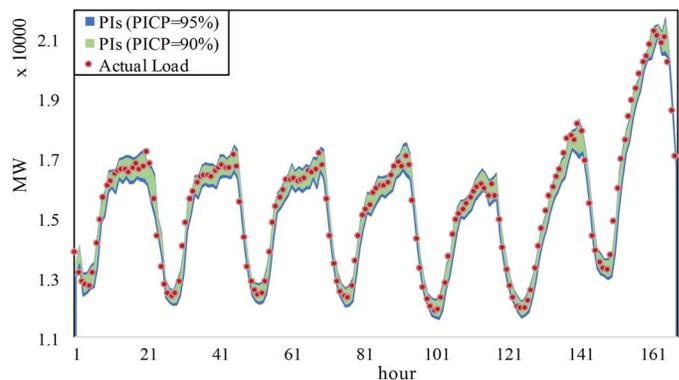


Fig. 5. Load PIs by the proposed method with confidence levels 90% and 95% (Ontario-Aug. 2014).

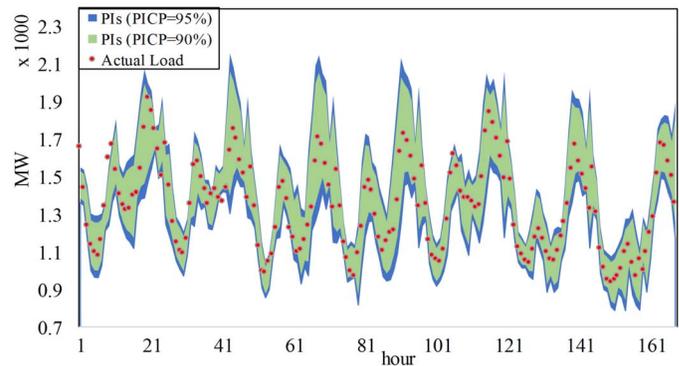


Fig. 6. Load PIs by the proposed method with confidence levels 90% and 95% (Australia-May. 2014).

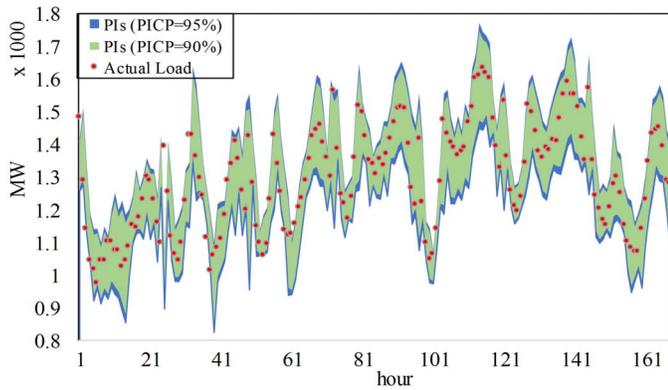


Fig. 7. Load PIs by the proposed method with confidence levels 90% and 95% (Australia-Oct. 2014).

VI. CONCLUSIONS

The nature of time-series electricity demand is highly nonlinear and is dependent on various parameters. Therefore, it is always difficult to be forecasted since it involves numerous uncertainties. An accurate and reliable load forecasting can help market participants and facilitate related activities such as economic dispatch and unit commitment. Therefore, using a new and efficient tool is very important to obtain such a goal. In this paper, a new high performance combined method was proposed, consisting of GELM, IWNNs, wavelet processing and bootstrapping. Additionally, the uncertainties related to the forecasting model and data noise were considered, while forecasting intervals were obtained as well. Based on the used tools, the proposed method had high speed and reliability. Its performance was evaluated using real data from Ontario and Australian markets, and the results were acceptable. Also, based on the results, it can be concluded that the high speed, accuracy and uncertainties' consideration have made this method a highly efficient technique for practical applications of probabilistic load forecasting. Future work can include: i) to consider other data such as weather information as neural network inputs; ii) to make a new generalization on ELM based methods to be used on multi-layer neural networks.

REFERENCES

- [1] J. Xie, T. Hong, T. Laing, and C. Kang, "On normality assumption in residual simulation for probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1046-1053, May 2017.
- [2] W. Christiaanse, "Short-term load forecasting using general exponential smoothing," *IEEE Trans. Power App. Systems*, pp. 900-911, 1971.
- [3] L. Ghelardoni, A. Ghio, and D. Anguita, "Energy load forecasting using empirical mode decomposition and support vector regression," *IEEE Transactions on Smart Grid*, vol. 4, pp. 549-556, 2013.
- [4] M. Chaouch, "Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 411-419, 2014.
- [5] Ö. F. Ertugrul, "Forecasting electricity load by a novel recurrent extreme learning machines approach," *International Journal of Electrical Power & Energy Systems*, vol. 78, pp. 429-435, June 2016.
- [6] E. Ceperic, V. Ceperic, and A. Baric, "A strategy for short-term load forecasting by support vector regression machines," *IEEE Transactions on Power Systems*, vol. 28, pp. 4356-4364, 2013.
- [7] D. Sáez, F. Ávila, D. Olivares, C. Cañazares, and L. Marín, "Fuzzy prediction interval models for forecasting renewable resources and loads in microgrids," *IEEE Trans. Smart Grid*, vol. 6, pp. 548-556, 2015.

- [8] A. Clements, A. Hurn, and Z. Li, "Forecasting day-ahead electricity load using a multiple equation time series approach," *European Journal of Operational Research*, vol. 251, pp. 522-530, 2016.
- [9] S. Hassan, A. Khosravi, J. Jaafar, and M. A. Khanesar, "A systematic design of interval type-2 fuzzy logic system using extreme learning machine for electricity load demand forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 82, pp. 1-10, 2016.
- [10] L. Wu, M. Shahidehpour, and T. Li, "Stochastic security-constrained unit commitment," *IEEE Trans. Power Systems*, vol. 22, no. 2, pp. 800-811, 2007.
- [11] Y. Wang, Q. Xia, and C. Kang, "Unit commitment with volatile node injections by using interval optimization," *IEEE Trans. Power Systems*, vol. 26, no. 3, pp. 1705-1713, Aug. 2011.
- [12] R. Billinton and D. Huang, "Effects of load forecast uncertainty on bulk electric system reliability evaluation," *IEEE Trans. Power Systems*, vol. 23, no. 2, pp. 418-425, May 2008.
- [13] B. Borkowska, "Probabilistic load flow," *IEEE Trans. on Power Apparatus and Systems*, vol. 93, pp. 752-759, 1974.
- [14] T. Hong and S. Fan, "Probabilistic electric load forecasting: a tutorial review," *International Journal of Forecasting*, vol.32, no. 3, pp. 914-938, 2016.
- [15] P. Gaillard, Y. Goude, and R. Nedellec, "Semi-parametric models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1038-1050, 2016.
- [16] B. Liu, J. Nowotarski, T. Hong, and R. Weron, "Probabilistic load forecasting via quantile regression averaging on sister forecasts," *IEEE Transactions on Smart Grid*, vol. 8, pp. 730-737, 2017.
- [17] P. Kou and F. Gao, "A sparse heteroscedastic model for the probabilistic load forecasting in energy-intensive enterprises," *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 144-154, Feb. 2014.
- [18] H. Shi, M. Xu, and R. Li, "Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN," *IEEE Transactions on Smart Grid*, 2017.
- [19] A. Dedinec, S. Filiposka, A. Dedinec, and L. Kocarev, "Deep belief network based electricity load forecasting: An analysis of Macedonian case," *Energy*, vol. 115, pp. 1688-1700, 2016.
- [20] V. Dordonnat, A. Pichavant, and A. Pierrot, "GEFCom2014 probabilistic electric load forecasting using time series and semi-parametric regression models," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1005-1011, 2016.
- [21] M. Rafiei, T. Niknam, and M. H. Khooban, "Probabilistic electricity price forecasting by improved clonal selection algorithm and wavelet preprocessing," *Neural Computing and Applications*, pp. 1-13, 2016.
- [22] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Trans. Power Systems*, vol. 20, pp. 1035-1042, 2005.
- [23] M. Hanmandlu and B. K. Chauhan, "Load forecasting using hybrid models," *IEEE Transactions on Power Systems*, vol. 26, pp. 20-29, 2011.
- [24] A. T. Lora, J. R. Santos, J. R. Santos, J. L. M. Ramos, and A. G. Expósito, "Electricity market price forecasting: Neural networks versus weighted-distance k nearest neighbours," in *International Conference on Database and Expert Systems Applications*, pp. 321-330, 2002.
- [25] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, pp. 985-990, 2004.
- [26] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, pp. 489-501, 2006.
- [27] B. Efron, "Bootstrap methods: another look at the jackknife," in *Breakthroughs in Statistics*, ed: Springer, pp. 569-593, 1992.
- [28] L. Sharma, M. Chakrawarti, A. Dutta, and N. Adhikari, "Neural Network Based Approach for Short-term Load Forecasting," *International Journal of Science, Environment and Technology*, vol. 1, pp. 416-424, 2012.
- [29] M. Rafiei, T. Niknam, and M. H. Khooban, "A novel intelligent strategy for probabilistic electricity price forecasting: Wavelet neural network based modified dolphin optimization algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 31, no. 1, pp. 301-312, 2016.
- [30] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1-30, 2006.
- [31] *Website of Ontario electricity market*. Available: <http://www.ieso.ca>
- [32] *Website of Australian electricity market*. Available: <http://www.aemo.com.au>



Mehdi Rafiei was born in Khomeini-shahr, Iran, in 1991. He received the B.S. degree in Electrical Engineering from Yazd University, Iran, in 2013, and M.S. degree from Shiraz University of Technology, Shiraz, Iran, in 2015. He is currently a researcher in Esfahan Regional Electric Company (EREC). His research fields cover: Power Systems, Neural Networks, Optimization Algorithms, Power Market, Reliability and Maintenance.



Taher Niknam (M'14) received the B.Sc. degree from Shiraz University, Shiraz, Iran, in 1998, and the M.Sc. and Ph.D. degrees from the Sharif University of Technology, Tehran, Iran, in 2000 and 2005, respectively, all in power electrical engineering. He is a Full Professor with the Shiraz University of Technology, Shiraz, Iran. His research interests include power system restructuring, impacts of distributed generations on power systems, optimization methods, and evolutionary algorithms.



Jamshid Aghaei (M'12-SM'15) received the B.Sc. degree in electrical engineering from the Power and Water Institute of Technology, Tehran, Iran, in 2003, and the M.Sc. and Ph.D. degrees from the Iran University of Science and Technology, Tehran, in 2005 and 2009, respectively. He is currently an Associate Professor at the Shiraz University of Technology, Shiraz, Iran, and also a Research Fellow at the Norwegian University of Science and Technology, Trondheim, Norway. His research interests include renewable energy systems, smart grids, electricity markets, and power system operation, optimization, and planning. Dr. Aghaei is a Member of the Iranian Association of Electrical and Electronic Engineers and a Guest Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. He was considered one of the outstanding reviewers of the IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, in 2017.



Miadrezha Shafie-khah (M'13-SM'17) received the M.Sc. and Ph.D. degrees in electrical engineering from Tarbiat Modares University, Tehran, Iran, in 2008 and 2012, respectively. He received his first postdoc from the University of Beira Interior (UBI), Covilha, Portugal in 2015, while working on the 5.2-million-euro FP7 project SiNGULAR ("Smart and Sustainable Insular Electricity Grids Under Large-Scale Renewable Integration"). He received his second postdoc from the University of Salerno, Salerno, Italy in 2016. He is currently an Invited Assistant Professor and Senior Researcher at C-MAST/UBI, where he has a major role of co-coordinating a WP in the 2.1- million-euro national project ESGRIDS ("Enhancing Smart GRIDS for Sustainability"), while co-supervising four PhD students and two post-doctoral fellows. He was considered one of the Outstanding Reviewers of the IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, in 2014, and one of the Best Reviewers of the IEEE TRANSACTIONS ON SMART GRID, in 2016. His research interests include power market simulation, market power monitoring, power system optimization, demand response, electric vehicles, price forecasting and smart grids.



João P. S. Catalão (M'04-SM'12) received the M.Sc. degree from the Instituto Superior Técnico (IST), Lisbon, Portugal, in 2003, and the Ph.D. degree and Habilitation for Full Professor ("Agregação") from the University of Beira Interior (UBI), Covilha, Portugal, in 2007 and 2013, respectively.

Currently, he is a Professor at the Faculty of Engineering of the University of Porto (FEUP), Porto, Portugal, and Researcher at INESC TEC, INESC-ID/IST-UL, and C-MAST/UBI. He was the Primary Coordinator of the EU-funded FP7 project SiNGULAR ("Smart and Sustainable Insular Electricity Grids Under Large-Scale Renewable Integration"), a 5.2-million-euro project involving 11 industry partners. He has authored or coauthored more than 600 publications, including 215 journal papers (more than 60 IEEE Transactions/Journal papers), 335 conference proceedings papers, 2 books, 34 book chapters, and 14 technical reports, with an *h*-index of 35, an *i10*-index of 135, and over 5500 citations (according to Google Scholar), having supervised more than 50 post-docs, Ph.D. and M.Sc. students. He is the Editor of the books entitled *Electric Power Systems: Advanced Forecasting Techniques and Optimal Generation Scheduling* and *Smart and Sustainable Power Systems: Operations, Planning and Economics of Insular Electricity Grids* (Boca Raton, FL, USA: CRC Press, 2012 and 2015, respectively). His research interests include power system operations and planning, hydro and thermal scheduling, wind and price forecasting, distributed renewable generation, demand response and smart grids.

Prof. Catalão is an Editor of the IEEE TRANSACTIONS ON SMART GRID, an Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS, and an Associate Editor of the *IET Renewable Power Generation*. From 2011 till 2018 (seven years) he was an Editor of the IEEE TRANSACTIONS ON SUSTAINABLE ENERGY. He was the Guest Editor-in-Chief for the Special Section on "Real-Time Demand Response" of the IEEE TRANSACTIONS ON SMART GRID, published in December 2012, and the Guest Editor-in-Chief for the Special Section on "Reserve and Flexibility for Handling Variability and Uncertainty of Renewable Generation" of the IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, published in April 2016. Since May 2017, he is the Corresponding Guest Editor for the Special Section on "Industrial and Commercial Demand Response" of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. He was the recipient of the 2011 Scientific Merit Award UBI-FE/Santander Universities, the 2012 Scientific Award UTL/Santander Totta, the 2016 FEUP Diploma of Scientific Recognition, and the Best INESC-ID Researcher 2017 Award, in addition to an Honorable Mention in the 2017 Scientific Awards ULisboa/Santander Universities. Moreover, he has won 4 Best Paper Awards at IEEE Conferences.