IET Generation, Transmission & Distribution

Research Article

IET Journals
The Institution of Engineering and Technology

Point and interval forecasting of real-time and day-ahead electricity prices by a novel hybrid approach

ISSN 1751-8687
Received on 3rd September 2016
Revised 4th February 2017
Accepted on 24th February 2017
E-First on 19th June 2017
doi: 10.1049/iet-gtd.2016.1396
www.ietdl.org

Reza Tahmasebifar[1], Mohammad Kazem Sheikh-El-Eslami[1] ✉, Reza Kheirollahi[1]

[1]*Electrical and Computer Engineering Department, Tarbiat Modares University, Tehran, Iran*

✉ *E-mail: aleslam@modares.ac.ir*

Abstract: Accurate forecasting of electricity market prices presents important information to market participants. This provides forward planning of their bidding strategies in order to maximise revenue, profit, and utility perspectives. Nevertheless, due to the non-stationarities involved in market clearing price, an accurate forecasting of these prices is very complex. In this case, transformation from traditional point forecasts to probabilistic interval ones is of great importance to quantify the uncertainties of potential forecasts. In this study, interval forecasting of market clearing prices is conducted based on a novel approach within two consecutive steps. In the first step, a new hybrid method is proposed to estimate point forecasts: combination of wavelet transformation (Wt), feature selection based on Mutual Information (MI), extreme learning machine (ELM), and bootstrap approaches in an ensemble structure is employed. The second step consists of the following stepwise parts: calculating the variance of the model uncertainties based on the extracted data from the ensemble structure, estimating the noise variance by using the maximum-likelihood estimation (MLE), and improving the accuracy of interval forecasting by using particle swarm optimisation (PSO) algorithm. The effectiveness of the proposed approach termed as Wt–mutual information–ELM–MLE–PSO is validated through electricity market real data of Australian electricity network from real-time and day-ahead market viewpoints.

1 Introduction

For many years, electricity market was dominated by a few companies controlled mainly rules. Unlike regulated markets which the companies determine prices independently, electricity prices are controlled based on a supply–demand relationship in a deregulated market. On the one hand, getting more information to identify key contributors to the electricity market prices supply–demand relationship seems necessary before real-time utilisation; on the other hand, electricity prices are changed during the day, so forecasting of electricity market prices are of interest to researchers, production companies, investors, independent market operators, and large industrial consumers [1]. Over last two decades, different tools and models have been proposed by researchers all over the world; these are: time-series regression models, transfer function, dynamic regression [2, 3], autoregressive integrated moving average [4, 5], and generalised autoregressive conditional heteroscedasticity (GARCH) [6]. To improve the performance of the forecast strategy, a combination of these methods was reported by Liu and Shi [7]. Artificial neural networks (ANNs) are also used to overcome the limitations of linear time-series models [8–13]. By regarding the results of ANNs, an improvement on the electricity prices forecasting was carried out by combining the ANNs with aforementioned methods [14–16].

Owing to non-stationary electricity prices series, forecasting errors are significant and notable. By identifying the aforementioned inherent limitations of traditional point forecasting methods, probabilistic forecasting has attracted the attention of many researchers in the case of risk management or bidding of a marginally safe [17]. In this way, prediction intervals (PIs) are presented to determine the uncertainties related to the point forecasting of the electricity clearing price. With quantified PIs, the electricity market participants should be prepared for the best and worst conditions. Therefore, reliable and effective probabilistic forecasting of the electricity prices is necessarily needed in order to facilitate the variety of decision-making process.

Recursive dynamic factor analysis along with Kalman filter was implemented to determine the intervals of forecasting [18]. Shrivastava *et al.* [19] presented ANNs and wavelet transform hybrid models to estimate electricity prices and its intervals. Khosravi *et al.* [20] used bootstrap moving block of ANNs and GARCH models to predict electricity prices and variance estimation. To construct forecasting intervals, quantile regression and different time-series models have been employed by Nowotarski and Weron [21]. Kou *et al.* [22] recommended models based on active learning and the variational heteroscedastic Gaussian process to forecast the electricity prices in a probabilistic way. A novel method to generate PIs using multi-objective method based on differential evolution has been proposed by Shrivastava *et al.* [23]. Wan *et al.* [24] proposed a novel Pareto optimal PI construction approach for electricity price; that is, it combines extreme learning machine (ELM) and non-dominated sorting genetic algorithm II. A quick forecasting of electricity prices based on ELM combined with the bootstrap approach, termed as ELM-bootstrap herein, have been presented by Chen *et al.* [25]. However, the uncertainties of data noise were not considered. In [17], forecasting intervals were estimated by taking model uncertainties and noise into account. Firefly algorithm has been employed to optimise PI-based cost function and the regulation of ANN parameters, and seven synthetic and real-world test models have been studied to estimate PIs [26]; in this case, by using PI-based cost function and optimising that considerable improvements can be seen in comparison with recent methods. Owing to low accuracy of the point forecasting results in the two recent mentioned methods, interval forecasting did not present accurate outcomes. The reported methods have been evaluated by different indexes to show their novelty and correctness.

In this paper, a novel hybrid approach including data preprocessing methods, ELM, bootstrap approach, maximum-likelihood estimation (MLE), and particle swarm optimisation (PSO) algorithm is proposed to predict probabilistic interval of electricity prices. ELM, a novel learning algorithm with extremely fast learning speed [27, 28], along with bootstrap approach and an ensemble structure are employed to estimate point forecasting and model uncertainties; besides, data preprocessing methods are used to improve the accuracy of the point forecasting and decrease the model uncertainties. Next, the MLE method is implemented to train a new NN; then, calculated weights and biases are considered as primary population of PSO algorithm. In this case, optimised

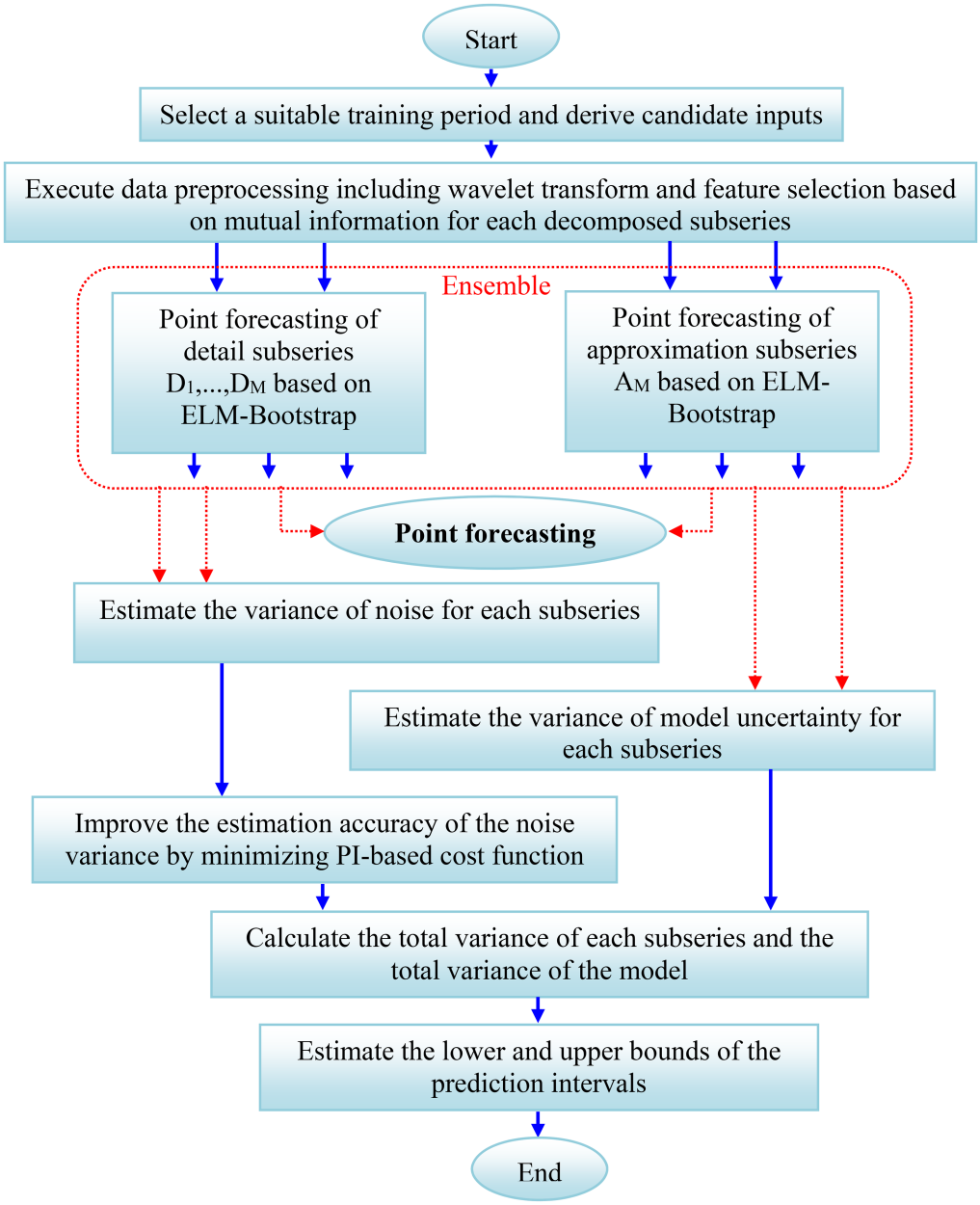

Fig. 1 *Architecture of the proposed forecast strategy*

weights and biases are given by optimising PI-based cost function; consequently, the noise variance of the predicted results can be estimated. In general, Developing point and probabilistic forecasting of the electricity prices for the next hour and the next day in a novel, quick, and better operational structure is the main aim of this paper. This approach is verified based on real data of Australian electricity market. The proposed approach of this paper is comprehensively evaluated by employing all the tools and indexes used in the recent stated research. This will confirm reliable effectiveness of the proposed approach.

The rest of this paper is organised as follows: Section 2 presents the proposed novel hybrid approach; Section 3 deals with the proposed point forecasting strategy; Section 4 illustrates estimating of the PIs and realising uncertainties; comprehensive numerical analysis is provided in Section 5; and finally, conclusions are drawn in Section 6.

2 Proposed novel hybrid approach

The proposed forecast strategy is shown graphically in Fig. 1, and it is explained stepwise for better understanding as follows: (i) by combining data preprocessing methods, ELM, and bootstrap approach, point predictions are estimated according to a proposed ensemble structure; (ii) to calculate the variance of the total prediction error, several components should be given. In the first, the variance of the model uncertainty should be computed; second, the variance of noise should be given by training NNs on minimising PI-based cost function. Finally, the lower and the upper bounds of the PIs can be quantified.

The novelty of the forecast strategy depicted explicitly in Fig. 1 includes two parts: the first part is point forecasting algorithm and the second part is realising prediction uncertainties. In the proposed point forecasting strategy, based on a novel ensemble structure with E iterations, a combination of next approaches are employed: wavelet transformation, feature selection, ELM, and bootstrap approach. The motivations for combining these methods are explained in the following: with respect to the limitations of the traditional NNs such as time-consuming training process and local minimum [17], and due to the better generalisation capability of ELM in comparison with traditional NNs, the ELM [17, 25] is preferable; ELM is combined with data preprocessing methods including wavelet transformation [9, 19, 29, 30] and mutual information (MI) [8, 31] in order to achieve following enhancements: improving the point and interval forecasting, reducing complexity, and decreasing the time required for training NNs; owing to the random initialisation feature of the ELM, the bootstrap approach and an ensemble structure are employed to

achieve predictions which are less deviated from real regression average and attain variance estimation of the model uncertainties.

The implementation of the proposed point forecasting algorithm is briefly explained stepwise as follows:

Step 1: Selecting adequate training period and realising training and testing data.
Step 2: Wavelet transform is completely executed on the price and load time series until the desired resolution level is achieved; training and testing datasets are generated for each wavelet decomposed subseries. The training and testing sets of each subseries are normalised.
Step 3: Feature selection based on MI is implemented on each wavelet decomposed subseries, and the qualified candidate inputs of each subseries are achieved by considering relevancy and redundancy thresholds.
Step 4: Each subseries is dedicated to an ELM-bootstrap; next points for each decomposed subseries are forecasted. Each predicted point is considered as an average of B bootstrap.
Step 5: Reconstructing wavelet transform in order to calculate electricity prices of the predefined hours; and estimating the bootstrap iterations of each approximation and detail subseries of the ensemble prediction.
Step 6: To achieve ensemble prediction, steps 4–5 in E iterations are repeated.

In the field of realising prediction uncertainties as the second novel part in this paper, prediction uncertainties are determined for PIs estimate in a new proposed structure. First, the variance of the model uncertainties is calculated based on the data extracted from the ensemble structure in the first step, and then the noise variance is achieved by training a new NN. Second, to improve the accuracy of interval forecasting, PSO algorithm is employed to optimise the PI-based objective function. In this case, various indexes are used to evaluate the point forecasting results and the accuracy and reliability of the given PIs.

Thanks to the novel aspects of the proposed hybrid point and interval forecasting approach, it is termed as Wt–MI–ELM–MLE–PSO. Comprehensive description and key roles of each step are presented in the subsequent sections.

3 Step 1: estimating of the point forecasting

In this section, all employed methods are described and their roles in the proposed point forecasting strategy are illustrated. Detail information is referenced.

3.1 Wavelet transformation

Wavelet transforms have attracted the attention of researchers in the field of artificial intelligence, machine learning, time-series forecasting, and pattern recognition. In fact, wavelets theory is a generalisation of the transformation theory and Fourier series, and it can compensate the limitations of the Fourier analysis in the case of partial operation and the short-term behavioural modelling. It is considered as an important tool to analyse the signals frequency components and overcome the constraints related to the short-time Fourier transformation. This transformation realises the relevant time–frequency information from non-periodic and transient signals. Wavelets functions decompose the data into different frequency components and each component is examined with a resolution matched to its scale. In this paper, wavelet transform is used to decompose the electricity price and load series into a set of better-behaved constitutive series. These series present better behaviour compared with original price series owing to the fact that they show more stable variance and no outliers, so they can be predicted more accurately. Filtering effect of the wavelet transform is the main reason for better behaviour of the constitutive series. Predictions for the constitutive series are made separately, and reverse wavelet transform is executed to generate the actual predicted prices. The decomposed coefficients of the wavelet transform of the hourly price series are determined as follows [29]:

$$p_{mn}^{L} = 2^{-(m/2)} \sum_{t=0}^{T-1} p_t L\left(\frac{t - n \times 2^m}{2^m}\right) = 2^{-(m/2)} \sum_{t=0}^{T-1} p_t L_{mn}(t) \quad (1)$$

where $L(.)$ is the chosen wavelet function; p_t shows the price in hour t; T defines the series length; and p_{mn}^{L} portrays the decomposition factor related to resolution m and position n.

An effective way of applying wavelet functions is the multi-resolution technique by using a father wavelet function along with its complementary the mother wavelet function. The father wavelet function is employed to realise low-frequency components; while, the mother wavelet function allows extracting the high-frequency components of the series. Orthogonal wavelet functions are preferably chosen because of their mathematical properties. In this way, the approximation series A_m ($m = 1, \ldots, M$) and the detail series D_m ($m = 1, \ldots, M$) are defined [29, 30]

$$A_m = \sum_n p_{mn}^{\phi} \varphi_{mn}(t); \quad m = 1, \ldots, M \quad (2)$$

$$D_m = \sum_n p_{mn}^{\Psi} \psi_{mn}(t); \quad m = 1, \ldots, M \quad (3)$$

where $\varphi_{mn}(t)$ and $\Psi_{mn}(t)$ are father and mother wavelet functions, respectively. p_{mn}^{Φ} and p_{mn}^{Ψ} are the factors given from relation (1).

The expression of the original price series $p_t(t = 1, \ldots, T)$ can now be reconstructed which is the denominated multi-resolution decomposition of the price series

$$p_t = D_1 + \cdots + D_M + A_M \quad (4)$$

Daubechies wavelets, the most appropriate ones for treating as a non-stationary series, are employed in this paper. In the case of Daubechies, by increasing the order of the functions, the smoothness increases consequently; however, the support intervals also increase which leads to deteriorate the predictions. As a result, low-order wavelet functions are preferable. In this case study, a wavelet function of order-four and decomposition level three is used which is an appropriate trade-off between wavelength and smoothness [29].

3.2 Feature selection

Feature selection is a process used in machine learning where subsets of the existing features of the data are employed in a learning algorithm. The best set includes the minimum number of the key features which are effective on output accuracy; on the contrary, the less important features are neglected. This plays a significant role in preprocessing method, and it prevents problem to be more large and complex. One of the developed techniques having been employed in feature selection is MI, which is based on entropy concept [31]. Two analyses can be used for feature selection: correlation analysis and MI analysis. The former is a linear feature selection technique; while, the latter is a non-linear criterion. On account of the fact that electricity price shows a non-linear and multivariate input/output mapping function, evaluating the dependencies of the price signal on its input features based on MI technique is better [8, 31]. MI is commonly used in clustering problems such as pattern recognising and image processing.

Regarding results extracted from references [8, 31], a new formulating of the MI is employed for feature selection of the electricity price forecasting. To achieve this goal, following steps should be implemented. First, all the candidate inputs and target variable are linearly normalised in range [0, 1]. Next, the median of each normalised variable is calculated. The half of the normalised variables which are greater than median are rounded to one, and another half which are less than median are rounded to zero. After that, a binominal distribution is given for each candidate. Then, by defining an auxiliary variable, individual probability distributions of the candidate inputs $P(Y_m)$, target variable $P(X)$, and joint probability distribution $P(X, Y_m)$ are calculated. In this paper, feature selection is employed to choose the best set of candidate

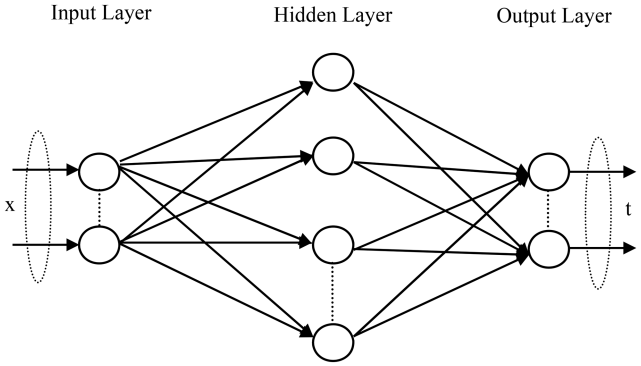

Fig. 2 *Common structure of an ELM [17]*

inputs, e.g. data prices and loads 200 h before forecasting hours. This is considered as a significant stage of preprocessing; moreover, it prevents problem to be multi-dimension. MI between candidate inputs Y_m and target variable X called MI(X, Y_m) is computed as follows:

$$\mathrm{MI}(X, Y_m) = \sum_{i=1}^{2}\sum_{j=1}^{2} P(X_i, Y_{m,j})\log_2\left[\frac{P(X_i, Y_{m,j})}{P(X_i)\times P(Y_{m,j})}\right] \quad (5)$$

where $P(Y_m)$ stands for the individual probability distribution of candidate inputs; $P(X_m)$ introduces the individual probability distribution of target variable; $P(X, Y_m)$ shows the joint probability distribution. Candidate inputs are ranked on their MI. Those of them with values MI(X, Y_m) greater than a relevancy threshold TH1 are retained as the relevant features; nevertheless, the other candidate inputs are filtered out [8]. In fact, features of a forecasting process may consist of extra information. One redundant set of candidate inputs not only increases the computation burden of the education step, but makes defining of the mapping input/output function complex as well. Regarding the limited historical data used for electricity price forecasting, the redundant data should be removed. $S_1 = \{x_1, x_2, \ldots, x_n\}$ is supposed as a set of candidate inputs including redundant features. After calculating MI between each pair of features from S_1 set, the higher values mean more common information between candidate inputs x_i and x_j, so these candidate inputs comprise higher level of redundancy. Maximum redundancy of each feature from S_1 set is computed as follows:

$$\max_{1\le j\le n}\ (\mathrm{MI}(x_i, x_j)) \quad (6)$$

On condition that the value of maximum redundancy criteria is increased more than a predefined redundancy threshold TH2, it is considered as a redundant candidate input, and between this feature and its pair, one of them with lower relevancy factor is removed. Consequently, the redundancy filtering process for all the features in S_1 set is repeated till no redundancy criteria greater than TH2 exists [31].

3.3 Extreme LM

ELM is a new algorithm employed to educate single hidden layer feed-forward NNs. A predefined dataset with N separate samples $\{(x_i, t_i)\}_{i=1}^{N}$ includes $x_i \in R^n$ as inputs and $t_i \in R^m$ as outputs; an ELM with L hidden nodes and activation function $\psi(.)$ can be represented as follows [17, 27, 28]:

$$f_L(x_j) = \sum_{i=1}^{L}\beta_i\psi(a_i\cdot x_j + b_i) = t_j, \quad j = 1, \ldots, N \quad (7)$$

where $\boldsymbol{a}_i = [a_{i1}, a_{i2}, \ldots, a_{in}]^{\mathrm{T}}$ portrays weights vector between the ith hidden node and the input nodes; $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{in}]^{\mathrm{T}}$ represents weights vector between the ith hidden node and the output nodes; b_i shows threshold of the ith hidden node; $\psi(a_i\cdot x_j + b_i)$ is the output of the ith hidden node related to the inputs x_j. Equation (7) can be rewritten compactly as follows:

$$\boldsymbol{H\beta} = \boldsymbol{T} \quad (8)$$

where $\boldsymbol{H}$ signifies hidden layer output matrix of the ELM, and it can be estimated as

$$\boldsymbol{H} = \begin{bmatrix} h(a_1\cdot x_1 + b_1) & \cdots & h(a_L\cdot x_1 + b_L) \\ \vdots & \cdots & \vdots \\ h(a_1\cdot x_N + b_1) & \cdots & h(a_L\cdot x_N + b_L) \end{bmatrix}_{N\times L} \quad (9)$$

The Lth column of the $\boldsymbol{H}$ matrix represents the output vector of the Lth hidden node related to the inputs x_j. In (8), $\boldsymbol{\beta}$ portrays the output weights, and $\boldsymbol{T}$ depicts the target matrix. In the ELM learning process, after initialising the input weight $a_i$and hidden biases b_i such as a_i^* and b_i^*, the output matrix of the hidden layer can be determined exclusively. Then, predicted output weights $\boldsymbol{\beta}^*$ with given target can be calculated based on the following formula:

$$\begin{aligned} &\| H(a_1^*, \ldots, a_L^*, \quad b_1^*, \ldots, b_L^*)\beta^* - \boldsymbol{T} \| \\ &\quad = \min \| H(a_1, \ldots, a_L, \quad b_1, \ldots, b_L)\boldsymbol{\beta} - \boldsymbol{T} \| \end{aligned} \quad (10)$$

Theoretically, ELM objective function defines as minimising the cost function of the traditional gradient-based learning algorithms. For example, it can be used in propagation algorithms:

$$C = \sum_{j=1}^{N}\left(\sum_{i=1}^{L}\beta_i\psi(a_i\cdot x_j + b_i) - t_j\right)^2 \quad (11)$$

Next, the output approximated weights can be computed from simple generalised inverse operation on the hidden layer output matrix; it can be stated as follows:

$$\boldsymbol{\beta} = H^{\dagger}\boldsymbol{T} \quad (12)$$

where $H^{\dagger}$ represents the Moore–Penrose generalised inverse of the hidden layer output matrix $\boldsymbol{H}$. In general, ELM can successfully overcome the limitations of the gradient-based learning algorithms, e.g. overtraining, local minima, and so forth [17, 27, 28]. Common structure of an ELM is depicted in Fig. 2. As a result, in this paper, this approach is taken into account to educate the employed NNs.

3.4 Bootstrap approach

In the proposed ensemble structure, due to random initialising of the ELM, bootstrap approach is used to estimate forecasts which have been less deviated from real regression average. The process employed for residuals bootstrap is described step-by-step as follows: (i) gather training samples $\{(x_i, t_i)\}_{i=1}^{N}$ and estimating $\hat{g}(x_i)$ by using ELM; (ii) on results from the first step, the residuals of primary approximation are computed $\hat{\varepsilon}_i = t_i - \hat{g}(x_i), \quad i = 1, \ldots, N$; (iii) the residuals resulted from the second step are centred around zero $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - (1/N)\sum_{i=1}^{N}\hat{\varepsilon}_i$; (iv) by replacing the recentrated residuals $\{\tilde{\varepsilon}_i\}_{i=1}^{N}$ afresh, bootstrap data $\{\tilde{\varepsilon}_i^*\}_{i=1}^{N}$ are generated; (v) generate $\{t_i^*\}_{i=1}^{N}$ from $t_i^* = \hat{g}(x_i) + \tilde{\varepsilon}_i^*$ and conclude the bootstrapped training data $\{(x_i, t_i^*)\}_{i=1}^{N}$; (vi) estimate $\hat{g}_q^*(x_i)$ based on ELM from the qth bootstrapped samples $\{(x_i, t_i^*)\}_{i=1}^{N}$; (vii) repeat steps 4–6 until finishing B bootstrap replicates.

Indeed, the point forecasting of each subset $[\hat{g}(x_i)]$ can be computed in a simply same manner as reported by Wan *et al.* [17]. In each bootstrap process, bootstrap ELM is educated based on the bootstrapped samples.

3.5 Point forecasting assessment

Point predictions are evaluated by different statistical measures which are based on the absolute error (AE), i.e. absolute value of the differences between the actual price P_h and predicted price $\hat{P}_h$ for a given hour h. In this paper, the mean AE (MAE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE) are used to assess point predictions; in summary, they take following forms:

$$\text{MAE} = \frac{1}{T}\sum_{i=1}^{T}\left|P_h - \hat{P}_h\right| \tag{13}$$

$$\text{MAPE} = \frac{1}{T}\sum_{i=1}^{T}\frac{\left|P_h - \hat{P}_h\right|}{P_h} \tag{14}$$

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{h=1}^{T}\left(P_h - \hat{P}_h\right)^2} \tag{15}$$

where T represents forecasting period. After point forecasting based on the proposed structure, predictions uncertainties should be determined for PIs estimate.

4 Step 2: estimating of the PIs

4.1 PIs formulation

Uncertainties in ANNs forecasting are defined by considering following points: model uncertainty due to the mis-specification of ANN structure and its parameters, the noise of training data caused by accidental characteristics of the regression data [17]. On the basis of a definite set of distinct pairs $\{(x_i, t_i)\}_{i=1}^{N}$ the prediction target can be given as

$$t_i = g(x_i) + \varepsilon(x_i) \tag{16}$$

where t_i shows the ith prediction target; $\boldsymbol{x}_i$ is the input vector including historical market clearing prices and demands for the electricity price forecasting; $\varepsilon(x_i)$ portrays the noise with zero mean; and $g(x_i)$ signifies the true regression mean. It is assumed that the noise is approximately distributed with mean zero and variance $\hat{\sigma}_\varepsilon^2$ that may depend on the input variables [17]. Owing to the fact that trained NN $\hat{g}(x_i)$ can be considered as a prediction of the true regression from practical applications points of view, the forecasting error for each subseries can be expressed as follows:

$$t_i - \hat{g}(x_i) = \left[g(x_i) - \hat{g}(x_i)\right] + \varepsilon(x_i) \tag{17}$$

where $t_i - \hat{g}(x_i)$ represents the total forecasting error and $g(x_i) - \hat{g}(x_i)$ illustrates the estimation error of the NN related to the true regression. Since estimation error and noise are statistically independent, the variance of the total forecasting errors for each subseries $\hat{\sigma}^2_{t\_\text{subseries}}(x_i)$ can be given by adding the variance of the model uncertainty $\hat{\sigma}^2_{g\_\text{subseries}}(x_i)$ to the variance of noise related to each subseries $\hat{\sigma}^2_{\varepsilon\_\text{subseries}}(x_i)$; it takes the following form:

$$\hat{\sigma}^2_{t\_\text{subseries}}(x_i) = \hat{\sigma}^2_{g\_\text{subseries}}(x_i) + \hat{\sigma}^2_{\varepsilon\_\text{subseries}}(x_i) \tag{18}$$

On the basis of this assumption that the estimation error and the subseries noise are statistically independent, the variance of the total prediction error can be computed through the resulting formula

$$\hat{\sigma}^2_t(x_i) = \hat{\sigma}^2_{t\_A_M}(x_i) + \sum_{m=1}^{M}\hat{\sigma}^2_{t\_D_m}(x_i) \tag{19}$$

where $\hat{\sigma}^2_{t-A_M}(x_i)$ characterises the variance of the total prediction error related to the approximation series; $\hat{\sigma}^2_{t-D_m}(x_i)$ represents the variance of the total prediction error of the detail series; $m = 1, \ldots, M$ defines the decomposition levels of the wavelet transform. For a given time series $\{(x_i, t_i)\}$, the PI t_i along with confident level $100(1-\alpha)\%$ are expressed as $\left[L_\text{t}^{(\alpha)}(x_i), U_\text{t}^{(\alpha)}(x_i)\right]$; in this case, $L_\text{t}^{(\alpha)}(x_i)$ and $U_\text{t}^{(\alpha)}(x_i)$ are introduced as lower and upper bounds of the PIs, respectively. They are formulated as

$$L_\text{t}^{(\alpha)}(x_i) = \hat{g}(x_i) - z_{1-\alpha/2}\sqrt{\hat{\sigma}^2_\text{t}(x_i)} \tag{20}$$

$$U_\text{t}^{(\alpha)}(x_i) = \hat{g}(x_i) + z_{1-\alpha/2}\sqrt{\hat{\sigma}^2_\text{t}(x_i)} \tag{21}$$

where $z_{1-\alpha/2}$ portrays the critical value of the standard normal distribution, and it depends on the expected confident level $100(1-\alpha)\%$. It is estimated that the future target is placed within the constructed PI with the following nominal probability:

$$P\left(t_i \in \left[L_\text{t}^{(\alpha)}(x_i), U_\text{t}^{(\alpha)}(x_i)\right]\right) = 100(1-\alpha)\% \tag{22}$$

A general framework of the proposed point and interval forecasting approach is shown in Fig. 3.

4.2 Variance of model uncertainty

In the case of NN regression, the random initialisation of the network parameters and model structure mis-specification lead to model uncertainty [17]. The uncertainty in the NN estimation $\hat{g}(x_i)$ of the true regression $g(x_i)$ can be quantified by the ensemble structure explained in the foregoing sections. By applying the ELM-bootstrap to each subseries and by considering ensemble structure for all the forecasts, a less biased estimation of true regression of the future targets can be derived. The mean E outputs of the proposed ensemble structure are taken into account as the approximation of the true regression. It takes the form

$$\hat{g}(x_i) = \frac{1}{E}\sum_{e=1}^{E}\hat{g}_e(x_i) \tag{23}$$

where $\hat{g}_e(x_i)$ represents the estimation in the eth iteration of each subseries achieved by the following equation:

$$\hat{g}_e(x_i) = \frac{1}{B}\sum_{q=1}^{B}\hat{g}_q(x_i) \tag{24}$$

where $\hat{g}_q(x_i)$ indicates the qth bootstrap iteration of each approximation and detail subseries, thanks to ELM-bootstrap. The variance of model uncertainty for each subseries can be estimated by regarding the variance of outputs related to the E outputs ensemble structure

$$\hat{\sigma}^2_g(x_i) = \frac{1}{E-1}\sum_{e=1}^{E}\left(\hat{g}_e(x_i) - \hat{g}(x_i)\right)^2 \tag{25}$$

4.3 Variance of noise

After separately determining the model uncertainty for each subseries, to construct PIs, it is needed to estimate the variance of the noise $\hat{\sigma}^2_\varepsilon$ for each subseries. According to (18), the variance of the noise can be given in the following equation [17]:

$$\hat{\sigma}^2_\varepsilon \sim E\left[\left(t - \hat{g}\right)^2\right] - \hat{\sigma}^2_g \tag{26}$$

A set of squared residuals is computed in order that the model is estimated to fit the remaining residuals

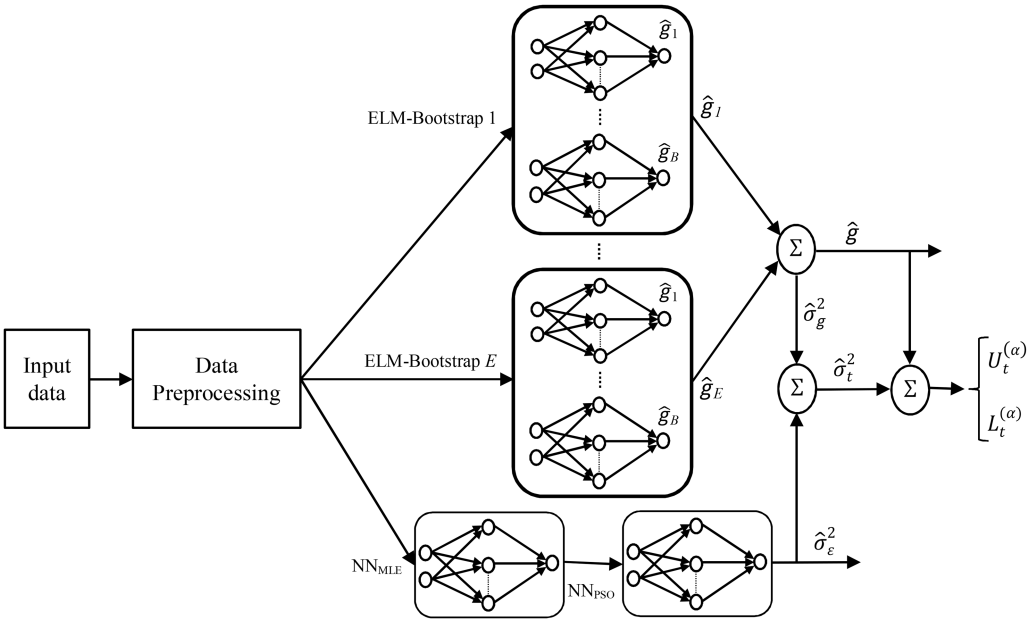

Fig. 3 *General framework of the proposed point and interval forecasting approach*

$$r^2(x_i) = \max\left(\left[t_i - \hat{g}(x_i)\right]^2 - \hat{\sigma}_g^2(x_i),\ 0\right) \tag{27}$$

where $\hat{g}(x_i)$ and $\hat{\sigma}_g^2(x_i)$ are achieved based on (23) and (25), respectively. A separate dataset is produced by considering the residuals and related inputs

$$D_{r^2} = \left\{\left(x_i,\ r^2(x_i)\right)\right\}_{i=1}^{N} \tag{28}$$

A separate model NN_ε can be indirectly trained for the purpose that the variance of the noise $\hat{\sigma}_\varepsilon^2$ is estimated based on the following cost function. To estimate the noise variance of the residuals, the MLE is used to educate the new NN. The probability of observing samples is maximised in D_{r^2} [17]

$$C_N = \frac{1}{2}\sum_{i=1}^{N}\left[\frac{r^2(x_i)}{\hat{\sigma}_\varepsilon^2(x_i)} + \ln\left(\hat{\sigma}_\varepsilon^2(x_i)\right)\right] \tag{29}$$

The output activation function of the new NN_ε is exponentially determined in the hope that the estimated variance will be positive at all times. The minimisation of the cost function can be achieved based on a traditional gradient decent method.

4.4 PIs evaluation and improving estimation of variance of noise

PIs are usually evaluated from accuracy and reliability aspects. The difference between nominal probability $100(1-\alpha)\%$ termed as PI nominal confidence (PINC) and the actual coverage probability (CP) of the constructed PIs termed as PICP is defined as average coverage error (ACE) which takes the following form:

$$\mathrm{ACE} = \mathrm{PICP} - \mathrm{PINC} \tag{30}$$

For N_t number of the test samples, the PICP is calculated as follows:

$$\mathrm{PICP} = \frac{1}{N_\mathrm{t}}\sum_{i=1}^{N_\mathrm{t}} I_i^{(\alpha)}, \quad I_i^{(\alpha)} = \begin{cases} 1 & t_i \in \left[L_\mathrm{t}^{(\alpha)}(x_i),\ U_\mathrm{t}^{(\alpha)}(x_i)\right] \\ 0 & t_i \notin \left[L_\mathrm{t}^{(\alpha)}(x_i),\ U_\mathrm{t}^{(\alpha)}(x_i)\right] \end{cases} \tag{31}$$

where L_t and U_t are the lower and the upper bounds of the PIs (related to the actual target t_i with inputs x_i), respectively. The PICP directly reveals the degree of PIs reliability. For the PIs of high reliability, the value of PICP should be as close to PINC as possible, i.e. the absolute ACE value is expected to be as small as possible.

Despite the fact that the high reliability can be reached by easily increasing the width of PIs, the sharpness index should be considered in PIs evaluation to determine the uncertainties of electricity price forecasting in an accurate manner. The width of PIs ($\nu_\mathrm{t}^{(\alpha)}$) can be computed through [17]

$$\nu_\mathrm{t}^{(\alpha)}(x_i) = U_\mathrm{t}^{(\alpha)}(x_i) - L_\mathrm{t}^{(\alpha)}(x_i) \tag{32}$$

Interval score of PIs from point forecasting $[S_\mathrm{t}^{(\alpha)}(x_i)]$ is defined as follows:

$$S_\mathrm{t}^{(\alpha)}(x_i) = \begin{cases} -2\alpha\nu_\mathrm{t}^{(\alpha)}(x_i) - 4\left[L_\mathrm{t}^{(\alpha)}(x_i) - t_i\right], & \text{if } t_i < L_\mathrm{t}^{(\alpha)}(x_i) \\ -2\alpha\nu_\mathrm{t}^{(\alpha)}(x_i), & \text{if } t_i \in \left[L_\mathrm{t}^{(\alpha)}(x_i),\ U_\mathrm{t}^{(\alpha)}(x_i)\right] \\ -2\alpha\nu_\mathrm{t}^{(\alpha)}(x_i) - 4\left[t_i - U_\mathrm{t}^{(\alpha)}(x_i)\right], & \text{if } t_i > U_\mathrm{t}^{(\alpha)}(x_i) \end{cases} \tag{33}$$

The overall score value of PIs ($\Omega_\mathrm{t}^{(\alpha)}$) can be achieved according to the averaging scores of the test samples

$$\Omega_\mathrm{t}^{(\alpha)} = \frac{1}{N_\mathrm{t}}\sum_{i=1}^{N_\mathrm{t}} S_\mathrm{t}^{(\alpha)}(x_i) \tag{34}$$

To carry out some improvements on estimation of the variance of noise, the concentration on the NN_ε optimising is of great importance in order to evaluate $\hat{\sigma}_\varepsilon^2(x_i)$ more accurate than the method mentioned in the previous section. The main idea is NN_ε training based on minimising of the PI-based cost function instead of using the one defined in (29).

In [26], PIs were evaluated according to their width by using PI normalised average width (PINAW)

$$\mathrm{PINAW} = \frac{1}{R \times N}\sum_{i=1}^{N}\left(U_i^{(\alpha)} - L_i^{(\alpha)}\right) \tag{35}$$

where R defines difference bounds of main targets, and it is considered as $R = \max\left(t_i\right) - \min\left(t_i\right)$, $i = 1, \ldots, N$. Both PICP

Table 1 Results of MAPE and RMSE indexes related to the point forecasting of the 23rd February for both real-time and day-ahead markets in the case of Daubechies with order-three (db3) and order-four (db4), Symlet with order-two (sym2) and order-three (sym3), and Coiflet with order-two (coif2) and order-three (coif3) transform functions

Wavelet function		db3	db4	sym2	sym3	coif2	coif3
real time	MAPE	4.60	**1.71**	4.12	4.73	3.26	6.03
	RMSE	1.30	**0.55**	1.21	1.32	0.99	1.75
day ahead	MAPE	2.07	**1.36**	2.90	1.98	2.62	3.00
	RMSE	0.61	**0.42**	0.94	0.58	0.77	0.84

Bold values indicate the results of the strategy proposed in this paper.

Table 2 Results of MAPE and RMSE indexes related to the point forecasting of the 23rd February for both real-time and day-ahead markets in the case of Daubechies with order-four (db4) with decomposition levels including levels 2, 3, and 4

Decomposition		Level 2	Level 3	Level 4
real time	MAPE	3.97	**1.71**	2.67
	RMSE	1.28	**0.55**	0.80
day ahead	MAPE	2.27	**1.36**	1.65
	RMSE	0.71	**0.42**	0.50

Bold values indicate the results of the strategy proposed in this paper.

and PINAW are one-dimensional criterion, and they evaluate PIs from different points of views. Coverage width-based criterion (CWC) evaluates PIs from both width and CP aspects simultaneously [26, 32]

$$\mathrm{CWC} = \mathrm{PINAW} + \gamma(\mathrm{PICP})\, \mathrm{e}^{\eta(\mathrm{PINC}-\mathrm{PICP})},$$
$$\gamma(\mathrm{PICP}) = \begin{cases} 0 & \mathrm{PICP} \geq \mathrm{PINC} \\ 1 & \mathrm{PICP} < \mathrm{PINC} \end{cases} \tag{36}$$

where η portrays the control parameter, and it is taken as 100 in this paper.

The CWC illustrates an effective compromise between informativeness and correctness of PIs [26]. The CWC criterion is considered as cost function for training $\mathrm{NN}_{\varepsilon}$ for each subseries. It is expected that setting parameters of the $\mathrm{NN}_{\varepsilon}$ improves the quality of the constructed PIs through CWC minimisation. In this paper, the evolutionary optimisation algorithm is employed in order to minimise the CWC and setting parameters of the $\mathrm{NN}_{\varepsilon}^{\mathrm{opt}}$; the detailed information was reported in [26]. In this paper, PSO [33] is used to minimise the CWC cost function. The employed parameters of the PSO are the same as [9].

Thanks to the aforementioned PIs evaluation indexes, to comprehensively evaluate resulted PIs in this paper, ACE [17–26], PINAW [19, 20, 22, 24, 26], Score (interval score or Winkler score) [17, 18, 23], and CWC indexes [19, 20, 26, 32] all are taken into account. Each index carries some important information, so gathering all evaluation indexes gives better understanding of the effectiveness of the proposed approach.

5 Case study and results

5.1 Electricity market and test data

The effectiveness of the proposed approach is studied on the Australian National Electricity Market (ANEM). The ANEM consists of five regional market jurisdictions [17] which the New South Wales (NSW) is taken into account in this paper. Load data and electricity prices in 2008 are considered to train and test the proposed structure. The data of 50 days before the forecasted day ($50 \times 48 = 2400$ in NSW market) are considered as training data. In this case, electrical loads and prices data of 200 h before the forecasted hour (800 sample data in NSW market) for the proposed approach in real-time market, and prices data of 200 h before the forecasted hour (400 sample data in NSW market) in day-ahead market are considered as input candidates; then, by employing MI, the best inputs of the NNs are chosen based on the TH1 and TH2 bounds. With respect to the test method reported in [8–12, 15], market clearing prices corresponding to 4 weeks of four seasons in 1 year including the periods from 22 to 28 November in spring, 23 to 29 February in summer, 16 to 22 May in autumn, and 23 to 29 August in winter are forecasted.

5.2 Comparison analysis

To evaluate the probabilistic forecasting performances of the proposed structure termed as Wt–MI–ELM–MLE–PSO, the results are compared with other two benchmarks, ELM-bootstrap proposed in [25] and hybrid method in [17] termed as ELM–MLE herein. Also, a comparison on point forecasting is performed between the proposed structure and the ELM-bootstrap [17, 25]. In the proposed structure, the iterations of the bootstrap approach B and the ensemble structure E are considered as 100 and 10, respectively. The three-level wavelet transforms including Daubechies wavelet function with order-four (db4) are employed to analyse historical subseries. This has been concluded by performing a fair comparison between Daubechies, Symlet, and Coiflet wavelet transforms with different orders, and then the selected transform function (Daubechies with order-four) has been analysed from decomposition level aspect; the detailed results are given in Tables 1 and 2. In this case study, the number of neurons for the day-ahead and real-time NNs are not the same and determined based on the trial-and-error procedure to achieve the optimum results; also, the other parameters of the proposed structure such as relevancy (TH1) and redundancy (TH2) thresholds in the feature selection, and the parameters of the PSO algorithm have been determined based on trial-and-error procedure. The probabilistic forecasts are determined with PINC taken as 95%. To execute the proposed structure, MATLAB R2014a and a laptop with Intel(R) Core(TM) i5-2410M central processing unit (CPU) @2.30 GHz CPU and 8.0 G random access memory specifications are employed.

Table 3 represents the average values of the point forecasting indexes for the four considered weeks of the four determined seasons in 2008. As it can be seen, the proposed structure improves the indexes compared with the traditional methods. Also, the necessity of using feature selection is verified by comparing the analytical results of the MI–MI–ELM method and ELM-bootstrap method. Considering the value of MAPE index, by combining the feature selection method with ELM-bootstrap method following significant improvements are made: 19.6% for February, 13.1% for May, 13.7% for August, and 20.67% for November in the case of real-time market; 17.37% for February, 12.72% for May, 18.61% for August, and 4.28% for November in the case of day-ahead market. This analysis shows that employing the feature selection method improves the accuracy of the point and probabilistic forecasting remarkably. From MAPE average value aspect, in real-time market, the proposed structure has achieved 69.5, 44.2, 41.5, and 64.4% in February, May, August, and November, respectively; Moreover, in day-ahead market, it has obtained 81.2, 58.5, 64.6, and 78.3%. The extracted data explain significant improvements on the output results.

Table 4 shows the variance of the AE of point forecasting ($|P_h - \hat{P}_h|$) in the case of real-time and day-ahead markets; also, variance improvement of the proposed method with respect to ELM-bootstrap, and MI–MI–ELM is presented for four test weeks. As it can be seen in Table 4, the forecasts with lower error variance are related to the proposed strategy of this paper. In this case, the uncertainties of the proposed forecasting approach are less than

other methods which lead to more accurate point and interval forecasting of the electricity prices. Moreover, the effectiveness of the feature selection method is proved by comparing the results of the MI–MI–ELM and ELM-bootstrap. Table 5 represents the variance of the total prediction error $[\hat{\sigma}_t^2(x_i)]$ and its improvement based on the proposed strategy of this paper with respect to ELM-bootstrap method [25] and ELM–MLE method [17] in the case of day-ahead and real-time markets. By regarding (20) and (21), $\hat{\sigma}_t^2(x_i)$ plays a key role in determining the upper and lower bounds of the PIs. As it be concluded in Table 5, proposed strategy effectively improves the value of the $\hat{\sigma}_t^2(x_i)$ including mean, median, and variance compared with other methods. In this case, interval forecasting is executed more reliable and more accurate than other methods.

In Table 6, to analyse the interval forecasting, the indexes including ACE, Score, PINAW, and CWC are calculated for the proposed structure and two traditional methods. Since PICP generated by the proposed structure is greater than PINC, the proposed structure generates larger ACE index and ACE which get the values close to zero; then, it carries a better reliability compared with traditional methods. Narrower PIs generated by the proposed structure decrease the value of the PINAW index significantly. CWC and Score indexes, which represent the overall skill of the PIs, have been improved pointedly by the proposed structure. Decreasing the value of the PICP lower than PINC leads to generate a great value of the CWC in the ELM-bootstrap. In a few cases, it may be observed that one of the ACE and PINAW indexes in the two traditional methods is better than the proposed structure, but the proposed model generates PIs with better overall skills compared with the others. Table 7 presents the improvement on the CWC and Score indexes by the proposed structure in all the pointed weeks in both real-time and day-ahead markets in comparison with the traditional methods. By considering these improvements, it can be concluded that the proposed structure is able to generate a compromise solution between reliability and the accuracy of the PIs. In fact, the proposed model generates the narrower PIs, while the real values of the prices are located in the PIs range with higher probability related to the PINC value.

Figs. 4 and 5 depict real prices, predicted prices, and AE ($|P_h - \hat{P}_h|$) based on the proposed methods for 1 week during May

Table 3 Average value of the point forecasting indexes in the real-time and day-ahead markets

Market	Method	ELM-bootstrap [17, 25]			MI–MI–ELM			Proposed structure		
	Index	MAE, $/MWh	MAPE, %	RMSE, $/MWh	MAE, $/MWh	MAPE, %	RMSE, $/MWh	MAE, $/MWh	MAPE, %	RMSE, $/MWh
real-time	February	1.66	6.98	2.54	1.33	5.61	2.05	0.50	2.13	0.73
	May	3.82	8.42	5.74	3.38	7.32	5.15	1.91	4.70	2.68
	August	3.30	7.74	5.00	2.84	6.68	4.41	1.80	4.53	2.64
	November	1.94	8.03	2.65	1.57	6.37	2.13	0.70	2.86	1.04
day-ahead	February	2.57	10.53	3.76	2.01	8.70	2.97	0.45	1.98	0.74
	May	5.83	12.57	8.71	5.08	10.97	8.11	2.13	5.22	3.39
	August	5.26	12.95	7.67	4.50	10.54	6.62	1.72	4.58	2.55
	November	2.67	11.67	3.47	2.51	11.17	3.38	0.61	2.53	0.87

Table 4 Error variance and the variance improvement of point forecasting for four test weeks in the case of real-time and day-ahead markets

Market	Test period	ELM-bootstrap [17, 25]	MI–MI–ELM	Proposed structure	Variance improvement[a] with respect to ELM-bootstrap, %	Variance improvement[a] with respect to MI–MI–ELM, %
real-time	February	4.67	3.71	0.38	91.76	89.64
	May	19.32	18.93	4.09	78.82	78.38
	August	14.90	25.13	4.06	72.74	83.84
	November	3.81	2.31	0.68	81.97	70.31
day-ahead	February	18.31	6.15	0.59	96.74	90.30
	May	60.84	52.96	5.49	90.98	89.63
	August	72.49	33.05	3.32	95.42	89.95
	November	46.94	5.42	0.40	99.14	92.58

[a] $\text{Improvement} = \frac{\text{value of the previous method} - \text{value of the proposed method}}{\text{value of the previous method}}$.

Table 5 Variance of the total prediction error $[\hat{\sigma}_t^2(x_i)]$ and its improvement for interval forecasting in the case of real-time and day-ahead markets for four test weeks

Market	Test period	ELM-bootstrap [25]			ELM–MLE [17]			Proposed structure			Improvement with respect to ELM-bootstrap, %			Improvement with respect to ELM–MLE, %		
		Mean	Med	Var	Mean	Med	Var	Mean	Med	Var	Mean	Med	Var	Mean	Med	Var
real-time	February	3.31	1.86	36.23	3.15	2.56	3.38	1.06	1.07	0.30	68	42.64	99.15	66.44	58.26	90.86
	May	12.87	6.63	420.4	5.31	5.73	6.61	3.18	2.81	1.79	75.29	57.53	99.57	40.11	50.84	72.85
	August	11.95	4.11	309.2	4.35	3.17	8.52	3.38	2.64	5.40	71.71	35.66	98.25	22.34	16.80	36.55
	November	5.91	3.50	136.5	4.01	3.39	6.35	3.02	2.96	0.18	48.83	15.38	99.87	24.47	12.73	97.13
day-ahead	February	13.86	6.10	408.10	5.22	4.31	15.92	1.80	0.72	14.56	86.97	88.16	96.43	65.39	83.30	8.50
	May	49.92	30.80	$3.4 \times 10^{+3}$	10.6	9.04	48.75	9.04	6.94	46.32	81.89	77.46	98.65	14.67	23.21	4.98
	August	34.54	15.63	$2.6 \times 10^{+3}$	9.71	7.46	54.15	8.09	5.49	51.66	76.57	64.88	98.08	16.64	26.37	4.61
	November	39.98	15.32	$1.9 \times 10^{+3}$	5.23	4.54	8.56	2.01	1.17	5.83	94.96	92.36	99.70	61.50	74.22	32

in the case of real-time market and August in the case of day-ahead market. As it can be seen, the proposed method can acceptably predict the prices of the next hour and next day of the Australian electricity network.

Figs. 6 and 7 show the given PIs computed by the proposed structure and the ELM–MLE for real-time and day-ahead markets of the February and November, respectively. As it can be seen in these figures, the proposed structure presents PIs with higher reliability and lower width than ELM–MLE method. In the day-ahead market, uncertainty in the point forecasting of the ELM–MLE method is larger than real-time market because of longer predicted horizon; consequently, in the day-ahead market, PIs with larger width and lower reliability are generated. In this case, the proposed model operates better in both the markets. From the computation burden aspect, execution of the proposed structure

Table 6 Evaluating the results of PIs for 4 week in real-time and day-ahead markets

Market	Method	Index	February	May	August	November
real-time	ELM-Bootstrap [25]	ACE	−10.18	−17.62	−15.83	−9.88
		Score	−1.60	−3.82	−3.17	−1.45
		PINAW	10.27	9.21	10.30	18.06
		CWC	$2.63 \times 10^{+4}$	$4.48 \times 10^{+7}$	$7.52 \times 10^{+6}$	$1.95 \times 10^{+4}$
	ELM–MLE [17]	ACE	1.73	0.24	−5.42	3.51
		Score	−1.57	−2.54	−2.59	−1.45
		PINAW	20.91	16.82	13.70	35.50
		CWC	20.91	16.82	238.83	35.50
	proposed structure	ACE	2.62	2.02	−0.65	3.81
		Score	−0.46	−1.53	−1.24	−0.63
		PINAW	6.87	10.07	9.29	12.98
		CWC	6.87	10.07	11.21	12.98
day-ahead	ELM-Bootstrap [25]	ACE	−20.89	−13.45	−23.27	−32.20
		Score	−3.54	−6.61	−8.70	−6.71
		PINAW	20.19	18.87	18.13	46.92
		CWC	$1.18 \times 10^{+9}$	$6.95 \times 10^{+6}$	$1.3 \times 10^{+10}$	$9.7 \times 10^{+13}$
	ELM–MLE [17]	ACE	0.83	2.02	0.54	1.73
		Score	−2.48	−4.62	−4.73	−2.28
		PINAW	34.59	33.54	35.97	46.02
		CWC	34.59	33.54	35.97	46.02
	proposed structure	ACE	3.21	1.13	1.13	4.11
		Score	−0.52	−1.59	−1.34	−0.51
		PINAW	7.16	11.27	10.68	10.96
		CWC	7.16	11.27	10.68	10.96

Table 7 Realising improvements of PIs indexes based on proposed model in real-time and day-ahead markets

Market	Test period	Score,(%		CWC, %	
		Method [25]	Method [17]	Method [25]	Method [17]
real-time	February	71.25	70.70	99.97	67.15
	May	59.94	39.76	99.99	40.13
	August	60.88	52.12	99.99	95.31
	November	56.55	56.55	99.93	63.44
day-ahead	February	85.31	79.03	99.99	79.30
	May	75.95	65.58	99.99	96.21
	August	84.60	71.67	99.99	70.31
	November	92.40	77.63	99.99	76.18

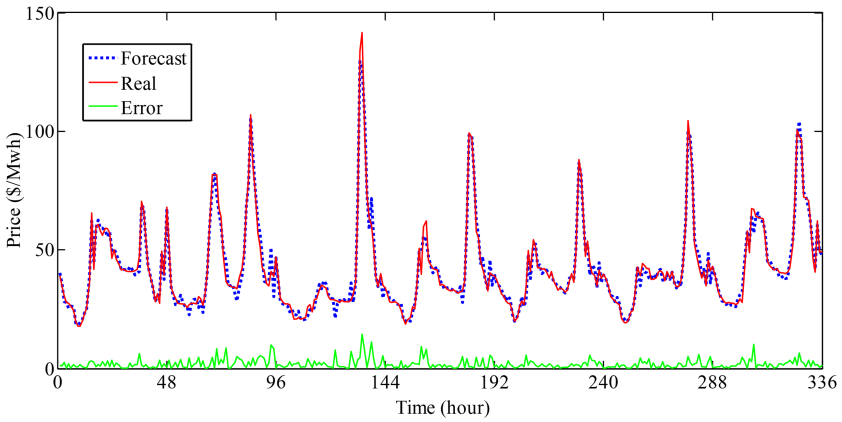

Fig. 4 *Real prices (red), predicted prices (blue), and AE (green) for 1 week during May in the case of real-time market*

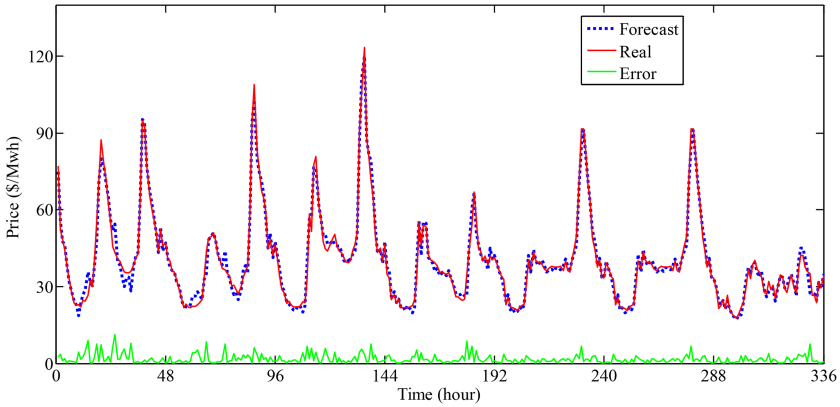

Fig. 5 *Real prices (red), predicted prices (blue), and AE (green) for 1 week during August in the case of day-ahead market*

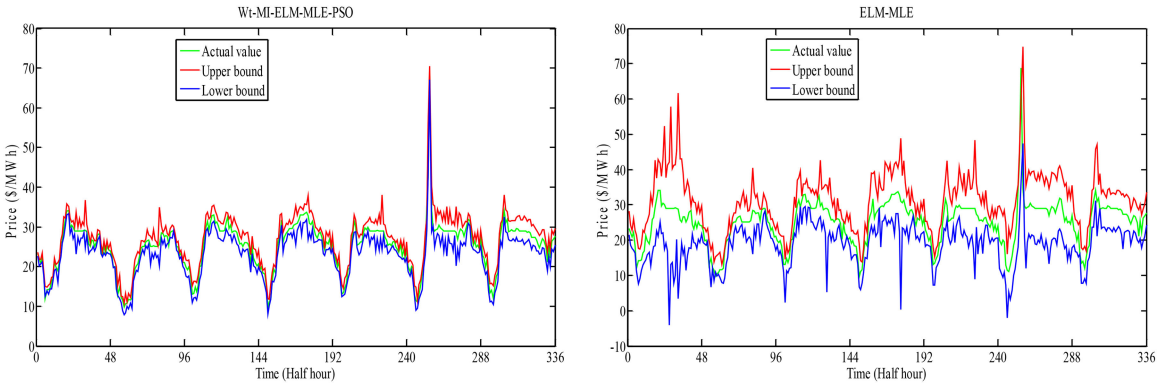

Fig. 6 *Constructed PIs with PINC 95% in February in real-time market by using ELM–MLE [17] and the proposed Wt–MI–ELM–MLE–PSO*

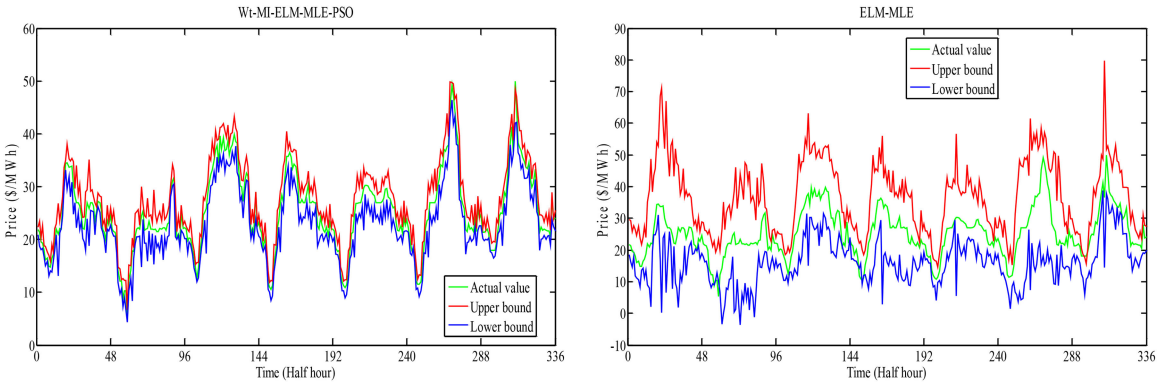

Fig. 7 *Constructed PIs with PINC 95% in November in day-ahead market by using ELM–MLE [17] and the proposed Wt–MI–ELM–MLE–PSO*

takes larger computation burden in comparison with others; however, it is not significant matter in offline calculations.

6 Conclusion

In this paper, a novel hybrid structure termed as Wt–MI–ELM–MLE–PSO is proposed to estimate point and interval forecasting of the electricity prices within two consecutive steps. In the first step, preprocessing data methods including wavelet transform and MI-based feature selection along with ELM, bootstrap approach, and an ensemble structure are employed. In the second step, the variances of model uncertainty and variance of noise are estimated in the hope that PIs are forecasted. Various indexes are used to evaluate the point forecasting results and the accuracy and reliability of the given PIs. Numerical results based on the real data of the Australia NSW electricity market in real-time and day-ahead markets show the effectiveness and correctness of the proposed structure for both point and interval forecasting. By regarding the improvements made on the point forecasting accuracy, the model uncertainty is decreased, and the PIs are further narrowed. Also, pointed indexes are improved by training NN based on optimising PI-based cost function.

7 References

[1] Mazengia, D.H., Tuan, L.: 'Forecasting spot electricity market prices using time series models'. IEEE Int. Conf. on Sustainable Energy Technologies, 2008 ICSET 2008, 2008, pp. 1256–1261

[2] Conejo, A.J., Contreras, J., Espinola, R., *et al.*: 'Forecasting electricity prices for a day-ahead pool-based electric energy market', *Int. J. Forecast.*, 2005, **21**, pp. 435–462

[3] Zareipour, H., Cañizares, C.A., Bhattacharya, K., *et al.*: 'Application of public-domain market information to forecast Ontario's wholesale electricity prices', *IEEE Trans. Power Syst.*, 2006, **21**, pp. 1707–1717

[4] Contreras, J., Espinola, R., Nogales, F.J., *et al.*: 'ARIMA models to predict next-day electricity prices', *IEEE Trans. Power Syst.*, 2003, **18**, pp. 1014–1020

[5] Crespo Cuaresma, J., Hlouskova, J., Kossmeier, S., *et al.*: 'Forecasting electricity spot-prices using linear univariate time-series models', *Appl. Energy*, 2004, **77**, pp. 87–106

[6] Garcia, R.C., Contreras, J., Van Akkeren, M., *et al.*: 'A GARCH forecasting model to predict day-ahead electricity prices', *IEEE Trans. Power Syst.*, 2005, **20**, pp. 867–874

[7] Liu, H., Shi, J.: 'Applying ARMA–GARCH approaches to forecasting short-term electricity prices', *Energy Econ.*, 2013, **37**, pp. 152–166

[8] Amjady, N., Keynia, F.: 'Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm', *IEEE Trans. Power Syst.*, 2009, **24**, pp. 306–318

[9] Catalão, J., Pousinho, H., Mendes, V.: 'Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting', *IEEE Trans. Power Syst.*, 2011, **26**, pp. 137–144
[10] Anbazhagan, S., Kumarappan, N.: 'Day-ahead deregulated electricity market price forecasting using recurrent neural network', *IEEE Syst. J.*, 2013, **7**, pp. 866–872
[11] Anbazhagan, S., Kumarappan, N.: 'Day-ahead deregulated electricity market price forecasting using neural network input featured by DCT', *Energy Convers. Manage.*, 2014, **78**, pp. 711–719
[12] Osório, G., Matias, J., Catalão, J.: 'Electricity prices forecasting by a hybrid evolutionary-adaptive methodology', *Energy Convers. Manage.*, 2014, **80**, pp. 363–373
[13] Panapakidis, I.P., Dagoumas, A.S.: 'Day-ahead electricity price forecasting via the application of artificial neural network based models', *Appl. Energy*, 2016, **172**, pp. 132–151
[14] Mandal, P., Srivastava, A.K., Park, J.W.: 'An effort to optimize similar days parameters for ANN-based electricity price forecasting', *IEEE Trans. Ind. Appl.*, 2009, **45**, pp. 1888–1896
[15] Shafie-Khah, M., Moghaddam, M.P., Sheikh-El-Eslami, M.: 'Price forecasting of day-ahead electricity markets using a hybrid forecast method', *Energy Convers. Manage.*, 2011, **52**, pp. 2165–2169
[16] Li, X., Yu, C., Ren, S.*, et al.*: 'Day-ahead electricity price forecasting based on panel cointegration and particle filter', *Electr. Power Syst. Res.*, 2013, **95**, pp. 66–76
[17] Wan, C., Xu, Z., Wang, Y.*, et al.*: 'A hybrid approach for probabilistic forecasting of electricity price', *IEEE Trans. Smart Grid*, 2014, **5**, pp. 463–470
[18] Wu, H., Chan, S., Tsui, K.*, et al.*: 'A new recursive dynamic factor analysis for point and interval forecast of electricity price', *IEEE Trans. Power Syst.*, 2013, **28**, pp. 2352–2365
[19] Shrivastava, N.A., Panigrahi, B.K.: 'Point and prediction interval estimation for electricity markets with machine learning techniques and wavelet transforms', *Neurocomputing*, 2013, **118**, pp. 301–310
[20] Khosravi, A., Nahavandi, S., Creighton, D.: 'A neural network-GARCH-based method for construction of prediction intervals', *Electr. Power Syst. Res.*, 2013, **96**, pp. 185–193
[21] Nowotarski, J., Weron, R.: 'Merging quantile regression with forecast averaging to obtain more accurate interval forecasts of Nord Pool spot prices', Hugo Steinhaus Center, Wroclaw University of Technology, 2014
[22] Kou, P., Liang, D., Gao, L.*, et al.*: 'Probabilistic electricity price forecasting with variational heteroscedastic Gaussian process and active learning', *Energy Convers. Manage.*, 2015, **89**, pp. 298–308
[23] Shrivastava, N.A., Panigrahi, B.K.: 'Prediction interval estimations for electricity demands and prices: a multi-objective approach', *IET Gener. Transm. Distrib.*, 2015, **9**, pp. 494–502
[24] Wan, C., Niu, M., Song, Y.*, et al.*: 'Pareto optimal prediction intervals of electricity price', *IEEE Trans. Power Syst.*, 2017, **32**, (1), pp. 817–819
[25] Chen, X., Dong, Z.Y., Meng, K.*, et al.*: 'Electricity price forecasting with extreme learning machine and bootstrapping', *IEEE Trans. Power Syst.*, 2012, **27**, pp. 2055–2062
[26] Khosravi, A., Nahavandi, S., Srinivasan, D.*, et al.*: 'Constructing optimal prediction intervals by using neural networks and bootstrap method', *IEEE Trans. Neural Netw. Learn. Syst.*, 2015, **26**, pp. 1810–1815
[27] Huang, G.B., Zhu, Q.Y., Siew, C.K.: 'Extreme learning machine: a new learning scheme of feedforward neural networks'. 2004 IEEE Int. Joint Conf. on Neural Networks, 2004 Proc., 2004, pp. 985–990
[28] Huang, G.B., Zhu, Q.Y., Siew, C.K.: 'Extreme learning machine: theory and applications', *Neurocomputing*, 2006, **70**, pp. 489–501
[29] Conejo, A.J., Plazas, M., Espinola, R.*, et al.*: 'Day-ahead electricity price forecasting using the wavelet transform and ARIMA models', *IEEE Trans. Power Syst.*, 2005, **20**, pp. 1035–1042
[30] Shrivastava, N.A., Panigrahi, B.K.: 'A hybrid wavelet-ELM based short term price forecasting for electricity markets', *Int. J. Electr. Power Energy Syst.*, 2014, **55**, pp. 41–50
[31] Keynia, F.: 'A new feature selection algorithm and composite neural network for electricity price forecasting', *Eng. Appl. Artif. Intell.*, 2012, **25**, pp. 1687–1697
[32] Khosravi, A., Nahavandi, S.: 'Closure to the discussion of 'prediction intervals for short-term wind farm generation forecasts' and 'combined nonparametric prediction intervals for wind power generation' and the discussion of 'combined nonparametric prediction intervals for wind power generation'', *IEEE Trans. Sustain. Energy*, 2014, **5**, pp. 1022–1023
[33] Eberhart, J.K.A.R.: 'Particle swarm optimization'. Proc. IEEE Int. Conf. Neural Networks, 1995, pp. 1942–1948