Knowledge-Based Systems 145 (2018) 182–196

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Ensemble incremental learning Random Vector Functional Link network for short-term electric load forecasting

Xueheng Qiu[a], Ponnuthurai Nagaratnam Suganthan[a,*], Gehan A.J. Amaratunga[b]

[a] School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore
[b] Department of Electrical Engineering, University of Cambridge, Cambridge CB2 1TN, UK

A R T I C L E I N F O

Article history:
Received 1 August 2017
Revised 15 January 2018
Accepted 18 January 2018
Available online 1 February 2018

Keywords:
Empirical Mode Decomposition
Discrete wavelet transform
Random Vector Functional Link network
Incremental learning
Time series forecasting
Electric load forecasting
Neural networks
Random forests

A B S T R A C T

Short-term electric load forecasting plays an important role in the management of modern power systems. Improving the accuracy and efficiency of electric load forecasting can help power utilities design reasonable operational planning which will lead to the improvement of economic and social benefits of the systems. A hybrid incremental learning approach composed of Discrete Wavelet Transform (DWT), Empirical Mode Decomposition (EMD) and Random Vector Functional Link network (RVFL) is presented in this work. RVFL network is a universal approximator with good efficiency because of the randomly generated weights between input and hidden layers and the close form solution for parameter computation. By introducing incremental learning, along with ensemble approach via DWT and EMD into RVFL network, the forecasting performance can be significantly improved with respect to both efficiency and accuracy. The electric load datasets from Australian Energy Market Operator (AEMO) were used to evaluate the effectiveness of the proposed incremental DWT-EMD based RVFL network. Moreover, the attractiveness of the proposed method can be demonstrated by the comparison with eight benchmark forecasting methods.

1. Introduction

It is well known that electricity power supply planning plays an important role in the management of modern power system. Accurate forecasting is beneficial for unit commitment, power system security, as well as energy transfer scheduling [1]. As stated in [2], for short-term electric load forecasting, the ballpark saving from 1% reduction in forecast error for a utility with 1 GW peak is roughly $300,000 per year. Therefore, the goal of load forecasting can be concluded as providing reliable power supply while keeping the operation costs and energy wastage as low as possible.

Electric load forecasting belongs to time series (TS) forecasting paradigm, which can be categorized into three types based on the forecasting horizon: short-term (minutes to one day ahead), medium-term (weeks to months ahead), long-term (years ahead) [3]. In this paper, we mainly focus on one day ahead load forecasting, which belongs to short-term load forecasting category. Due to various exogenous factors (e.g. special occasions, weather conditions, economic fluctuations, etc.), load data often possesses highly nonlinear and complicated patterns, which causes electric load forecasting a difficult task [4].

Since the 1940s, various statistical based linear time series forecasting approaches have been proposed. The common goal of these linear models is to use time series analysis for extrapolating the future energy requirement. The most successful methods are based on Holt-Winters exponential smoothing [5] and Autoregressive Integrated Moving Average (ARIMA) [6], as well as Linear Regression [7]. These statistical learning models are often treated as the baseline for electric load forecasting research nowadays [8–10]. For example, Fan and Hyndman [11] proposed a semi-parametric additive model for short term load forecasting, which is in the regression framework but with some nonlinear relationships and with serially correlated errors.

Artificial neural network (ANN) has been a popular machine learning approach with universal approximation capability for both classification and regression problems in a variety of research fields, such as control, biomedical, manufacturing, and power system management, etc. [12]. For example, in [7], Hong provided a comprehensive study on short term load forecasting using ANNs, as well as regression analysis and fuzzy regression models. One of the most popular way to train the ANN is the back-propagation (BP) supervised learning algorithm [13]. However, the BP algorithm

* Corresponding author.
E-mail addresses: qiux0004@e.ntu.edu.sg (X. Qiu), epnsugan@e.ntu.edu.sg (P.N. Suganthan), gaja1@cam.ac.uk (G.A.J. Amaratunga).

https://doi.org/10.1016/j.knosys.2018.01.015
0950-7051/

has several drawbacks, such as slow convergence and often being trapped in a local minimum [14].

A randomized version of neural network was reported in [15,16], which has random weight assignment and functional link between input and output layers. This variant of neural network is named Random Vector Functional Link (RVFL) network, which improves the efficiency of neural network training by randomly generating the weights from the input layer to hidden layer [16,17]. Another independently developed method, single hidden layer neural network with random weights (RWSLFN), was reported in [18], which is different from RVFL by excluding the functional link. However, some research works have proved that the functional link can significantly improve the performance of RVFL, in particular for time series forecasting [17,19].

Ensemble learning methods, which are also known as hybrid methods, aim to further improve the performance of forecasting methods by strategically combining multiple algorithms. There are three fundamental reasons for advantages of ensemble methods: statistical, computational and representational [20]. According to the way of combination, ensemble learning can be classified into two categories: parallel and sequential [21]. In a parallel combined ensemble method, the training signal is first decomposed into a collection of sub-datasets, which are analyzed individually and combined to generate the final forecasting result; while for sequentially combined ensemble methods, the outputs from one model are treated as the inputs for another forecasting model. There are many parallel ensemble methods in the literature, such as wavelet decomposition [22–24] and Empirical Mode Decomposition (EMD) [25].

Incremental learning, or online learning, is a machine learning paradigm where the machine learning model is updated according to the new examples whenever they emerge [26]. Therefore, incremental learning is different from traditional machine learning in the requirement of training dataset. Incremental learning does not totally depend on a sufficient training set in the training phase, but always update its model as new training examples appear over time. In fact, incremental learning is part of the natural learning and quite common in reality. For example, in [27], a recurrent neural network was constructed for grammar learning, and the author found that “the network fails to learn the task when the entire data set is presented all at once, but succeeds when the data are presented incrementally”. For electric load forecasting, electricity loads can be seen as a stream of incoming data, thereby it is necessary to focus on adaptive methods that are able to learn incrementally. Several incremental learning methods have been proposed for electric load forecasting in the literature. For example, Gabriela Grmanová *et al.* proposed an incremental heterogeneous ensemble model for time series prediction [28]. Moreover, in [29], an incremental electric load forecasting model based on support vector regression was introduced.

“Concept drift”, which means that the statistical properties of the target variable change over time in unforeseen ways, is one of the prominent problems and challenges for incremental learning [30]. It can be gradual or abrupt. Mere concept shift is often addressed by so-called passive methods, i.e. learning technologies which smoothly adapt model parameters such that the current distribution is reliably represented by the model. On the other hand, rapid concept changes often require active methods, which detect concept drift and react accordingly. In this work, load datasets from five states of Australia are used for simulations, which have relatively stable patterns due to the state-wide area, and thus only suffer from mere concept shift. Therefore, in incremental learning phase of the proposed method, if the exogenous factors change, the incoming input data will reflect the changes and help updating our model, which belongs to passive methods category.

There are many EMD based ensemble methods proposed for various research fields in the literature. For example, in [31], EMD and its improved versions are hybridized with SVR and ANN to tackle wind speed forecasting problems. According to the experimental results, it is concluded that EMD based models generally perform better than the corresponding single learning models. In our previous wor [32], an ensemble deep learning approach is proposed for electric load forecasting, which combines EMD and Deep Belief Networks. Some other EMD based ensemble forecasting models include: EMD-gene expression programming (GEP) method for electric load forecasting [33], EMD-SVR for wind power prediction [34], and EMD-RF for acute hypotensive episode prediction [35]. Wavelet transform is also widely applied for time series forecasting such as wavelet-ARIMA [36] and wavelet-particle swarm optimization (PSO)-adaptive neuro-fuzzy inference system (ANFIS) [37]. Interested readers are recommended to read the survey paper of ensemble methods [21].

Motivated by the good performance of EMD and wavelet transform based ensemble forecasting models mentioned above, this work presents a novel decomposition method for electric load TS signal, which combines DWT and EMD sequentially. EMD has a major drawback named mode mixing problem, whereas DWT can decompose TS signal into its frequency components. Therefore, the proposed EMD-DWT decomposition method can perform a better decomposition by using DWT to solve the frequency mixing problem of EMD [38]. Based on this decomposition method, an incremental ensemble learning approach is proposed for electric load forecasting, which is composed of DWT and EMD, along with incremental RVFL network. The proposed incremental DWT-EMD based RVFL network outperforms all of the eight benchmark methods. The effects of functional links, number of hidden neurons and incremental learning of RVFL are also discussed with persuasive comparative experiments.

The remaining part of this paper is organized as follows. Section 2 explains the theoretical background on forecasting models and ensemble approaches. Section 3 presents the algorithm of proposed incremental DWT-EMD based RVFL network. The details of experimental setup are shown in Section 4, followed by the discussion on experiment results in Section 5. In Section 6, two comparative experiments are implemented to evaluate the performance of the proposed method. Finally, the conclusion and future works are stated in Section 7.

2. Background

2.1. Artificial neural network

ANN, as a machine learning model inspired by human brain and central nervous system [39], can be classified into two categories according to the connectionism of network structure: feedforward neural network (FNN) and recurrent neural network (RNN). Single hidden layer feedforward neural network (SLFN) is the simplest version of the FNN, which has three fundamental layers: an input layer with the same number of neurons as the dimension of input features; a hidden layer having neurons with nonlinear activation function; and an output layer which aggregates the outputs from the hidden layer neurons. The direct connections between neurons in the adjacent layers have adjustable weights, which represent the strengths of these links. However, there is no interconnection of neurons within the same layer. The outputs from hidden neurons are calculated by:

$$v_j = f\left(\sum_{i=1}^{n} w_{ij}x_i + b_i\right) \tag{1}$$

and the final output value is:

$$y = g\left(\sum_{j=1}^{h} w_{jo} v_j + b_j\right) \quad (2)$$

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

The weights in SLFN (w_{ij} and w_{jo}) are traditionally tuned by the back propagation (BP) algorithm. As we have mentioned above, BP-based methods are usually very time consuming due to the iteration of BP. Moreover the gradient descent is often trapped in a local minimum. These disadvantages often limit the performance of SLFN.

2.2. Random Vector Functional Link network

RVFL network is a randomized version of the FNN, which has direct connections between input and output neurons (functional link), and uses fixed random weights and closed form least square estimation instead of the BP to tune the weights [15,16].

It is worth noting that all hidden layer weights $\mathbf{W}_h$ in an RVFL are generated with uniformly distributed random values within the interval $[-S, +S]$, where S is a scale factor to be determined during the parameter tuning stage [19]. Therefore, the output $\mathbf{H}$ from hidden neurons can be calculated based on the activation function. Here logistic sigmoid function is used as an example:

$$\mathbf{H} = \mathrm{logsig}(\mathbf{W}_h \cdot \mathbf{X}) \quad (3)$$

where $\mathbf{X}$ is the training data.

The output layer weights $\mathbf{W}_o$ need to be determined by certain optimization method. Due to the efficiency of closed-form solution, RVFL employs least square estimation to calculate the output layer weights:

$$\mathbf{W}_o = (H^T H)^{-1} H^T \mathbf{Y} \quad (4)$$

where $\mathbf{Y}$ is the training target.

The predicted values can thus be calculated by applying the obtained $\mathbf{W}_o$ and $\mathbf{W}_h$ to testing data:

$$\hat{Y}_s = \mathbf{W}_o \cdot \mathrm{logsig}(\mathbf{W}_h \cdot \mathbf{X}_s) \quad (5)$$

where $\hat{Y}_s$ is the predicted testing values and $\mathbf{X}_s$ is the testing data [40].

2.3. Empirical Mode Decomposition

EMD [25], also known as Hilbert–Huang transform (HHT), is a method to decompose a signal into several intrinsic mode functions (IMF) along with a residue which stands for the trend. EMD is an empirical approach to obtain instantaneous frequency data from non-stationary and nonlinear data sets.

The system load is a random non-stationary process composed of thousands of individual components. Thus, EMD algorithm can be very effective for load forecasting. An IMF is a function that has only one extreme between zero crossings, along with a mean value of zero. After EMD, the original time series signal is decomposed into a set of functions: $x(t) = \sum_{i=1}^{n}(c_i) + r_n$, where c_i is the i^{th} IMF, r_n is the residue, and the number of functions n in the set depends on the original signal.

2.4. Wavelet transform

Wavelet transform is a popular mathematical tool to decompose TS signal into its frequency components, which are beneficial for signal processing and analysis [41]. Wavelet transform is similar to the Fourier transform with a main difference in the merit function. The wavelet transform uses a collection of wavelet functions to represent the original signal, while traditional Fourier transform decomposes the signal into sines and cosines [42].

Maximal Overlap Discrete Wavelet Transform (MODWT) [43], as a variant of DWT, applies low-pass and high-pass filters to the input signal at each level, which has some advantages over standard DWT. The main difference between MODWT and traditional DWT is that MODWT is highly redundant and non-orthogonal. The redundancy increases the effective degrees of freedom (EDOF) on each scale and thus decreases the variance of statistical estimates, which allows better comparison of TS signal with its decomposition [44]. Moreover, the MODWT does not decimate the coefficients in order to keep the number of wavelet and scaling coefficients the same as the number of input data samples at every level of the transform. Therefore, the MODWT is well-defined for all sample sizes N. However, for a complete decomposition of J levels, the DWT requires N to be a multiple of 2^J [42,44].

2.5. Incremental learning with RVFL

Incremental learning RVFL is suitable for real time applications since the learning model needs to be updated whenever the new input patterns are available. As introduced in [45], by taking the pseudoinverse of a partitioned matrix, the stepwise updating of the weight in RVFL can be achieved easily due to the advantages of flat structure of RVFL.

Denote $\mathbf{A}_n$ as the $n \times m$ pattern matrix representing the input matrix consisting of all input vectors combined with enhancement components. a' is the $m \times 1$ new sample pattern which should be added to update the RVFL. Thus, the new pattern matrix A_{n+1} is shown as:

$$A_{n+1} \triangleq \begin{bmatrix} A_n \\ a' \end{bmatrix} \quad (6)$$

where n is the number of samples, as well as the discrete time instance, m is the dimension of the input pattern. Therefore, the pseudoinverse of A_{n+1} can be updated based only on the pseudoinverse of A_n and the imported new input vector a'. The procedure of calculation is shown as follows:

$$A_{n+1}^{+} = [A_n^{+} - db' | d] \quad (7)$$

where

$$\begin{aligned} b' &= a' A_n^{+} \\ d &= (1 + b'b)^{-1} A_n^{+} b \end{aligned} \quad (8)$$

Same with the pattern matrix A_n, we denote the output vector as Y_n, weight matrix as W_n, and the new output as y'. According to the least square solution, we can obtain:

$$\begin{aligned} W_{n+1} &= A_{n+1}^{+} Y_{n+1} \\ W_n &= A_n^{+} Y_n \end{aligned} \quad (9)$$

Therefore, the new weight matrix W_{n+1} can be calculated by:

$$W_{n+1} = W_n + (y' - a' W_n) d \quad (10)$$

where d is the obtained optimal learning rate of weight updating.

3. Proposed incremental RVFL based ensemble method

"Divide and conquer" is a parallel ensemble method which works by decomposing the original TS signal into a collection of

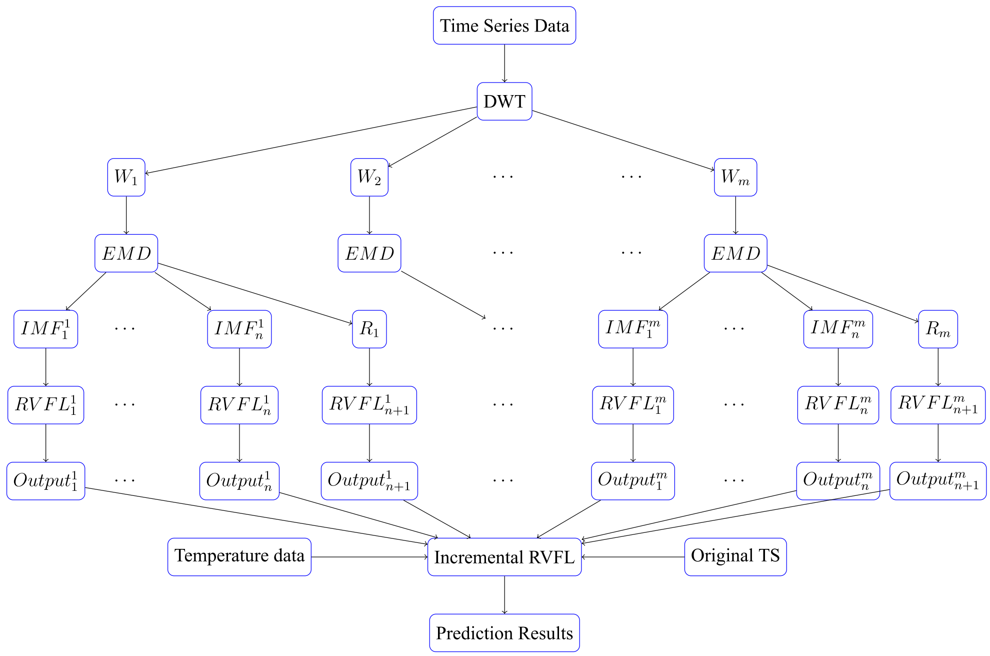

Fig. 1. Schematic diagram of the proposed DWT-EMD based incremental RVFL network.

sub-series until they are simple enough to be solved directly. EMD, as an efficient ensemble method to perform TS signal decomposition, has a major drawback called mode mixing problem: one IMF may consist of signal spanning a wide band of frequency, or more than one IMFs contain signals in a similar frequency band [38]. In the literature, this problem is normally solved by ensemble of EMD (EEMD) [46], which works by applying EMD to uncorrelated Gaussian noise added TS signal repetitively and combining the results to remove the noise. In this paper, the MODWT is employed to deal with the frequency issue, followed by EMD to perform better decomposition. Then an RVFL network is trained for each IMF and residue. The outputs of all sub-series are combined and analyzed by another RVFL to formulate the final prediction results. Fig. 1 is the schematic diagram of the pre-training phase of this proposed ensemble method, whose procedures can be concluded as follows:

1. Apply MODWT to decompose the original TS into several frequency components.
2. Apply EMD to decompose each frequency component into several IMFs and one residue.
3. Construct the training matrix as the input of each RVFL network for each obtained sub-series. Then train an RVFL network for each of the extracted IMF and residue.
4. Construct a new input matrix by combining three aspects: the prediction results of all the sub-series, the original TS signal and the temperature data. Then an incremental RVFL is trained using this new training matrix to formulate the final prediction results.

Incremental learning allows the learning model updating itself whenever the new input patterns are available, which guarantee the effectiveness of the proposed method for long term application. Therefore, after the pre-training phase, the learned ensemble model including the pseudoinverse of A_n and the weight matrix W_n for the incremental RVFL can be updated by the incremental learning phase:

1. When a new input sample is given to the network, MODWT and EMD can be applied to decompose the signal and obtain new input pattern A_{n+1} by combining all the new outputs from all sub-series. Then the new weight matrix W_{n+1} can be updated using Eq. (10).
2. The validation data is applied to check the error level. If the error decreases, we keep the updates, otherwise the weight matrix is not changed.
3. Repeat steps 1 and 2 whenever new samples are presented to the network.

Table 1
Summary of AEMO load datasets.

Dataset	Year	Length	Min	Median	Mean	Max	Std
QLD	2013	17520	4148.7	5752.1	5703.7	8278.4	747.0
	2014	17520	4073.0	5726.0	5745.7	8445.3	794.0
	2015	17520	4281.4	6005.6	6035.4	8808.7	777.2
NSW	2013	17520	5113.0	8045.0	7981.6	13788	1190.9
	2014	17520	5138.1	7987.4	7917.8	11846	1170.1
	2015	17520	5337.4	7990.4	7979.8	12602	1232.7
TAS	2013	17520	659.5	1109.0	1129.3	1650.3	142.3
	2014	17520	569.1	1088.7	1109.7	1630.1	139.0
	2015	17520	479.4	1112.3	1138.2	1667.2	145.3
SA	2013	17520	728.6	1389.3	1426.6	2991.3	301.7
	2014	17520	682.5	1360.8	1403.3	3245.9	312.8
	2015	17520	696.3	1352.7	1398.5	2870.4	306.0
VIC	2013	17520	3551.6	5458.1	5511.8	9587.5	895.9
	2014	17520	3272.9	5307.8	5324.4	10240	921.4
	2015	17520	3369.1	5186.5	5194.6	8579.9	864.7

4. Experiment setup

4.1. Datasets

In this paper, the electric load datasets from Australian Energy Market Operator (AEMO) [47] were used for evaluating the performance of benchmark learning models. There are totally fifteen electric load datasets of year 2013–2015 from five states of Australia: New South Wales (NSW), Tasmania (TAS), Queensland (QLD), South Australia (SA) and Victoria (VIC). For each dataset, the first nine months were used for training, while the remaining three months were used for testing. The statistics of the datasets are summarized in Table 1.

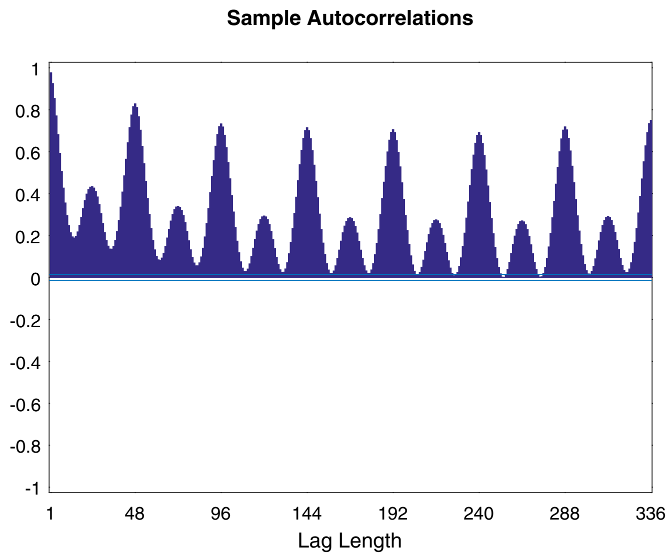

Fig. 2. Autocorrelation function for electricity load data in TAS.

The electric load data is sampled every half an hour, therefore 48 data points are recorded for one day. To identify cycles and patterns in load TS, we employ autocorrelation function (ACF) as a guidance for informative feature subset selection. Suppose a time series data set is given as $X = \{X_t : t \in T\}$, where T is the index set. The lag k autocorrelation coefficient r_k can be computed by:

$$r_k = r(X_t, X_{t-k}) = \frac{\sum_{t=k+1}^{n}(X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \tag{11}$$

where $\bar{X}$ is the mean value of all X in the given time series, r_k measures the linear correlation of the time series at times t and $t-k$.

From Fig. 2, which shows the ACF for electricity load data with a time window of one week, three strongest dependent lag variables can be identified: the value of previous half-an-hour (X_{t-1}), the value at the same time in the previous week (X_{t-336}), as well as the value at the same time in the previous day (X_{t-48}). Therefore, in order to take all the most informative lag variables into consideration, we select the data points of the whole previous day (X_{t-48} to X_{t-96}) and the same day in the previous week (X_{t-336} to X_{t-384}) to construct the input feature set for one day ahead load forecasting in this work.

After DWT-EMD decomposition, we get $m \times n$ IMFs and m residues. Let's mark the IMFs and residues as $\mathbf{X}^{i,j}$, where $1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n+1$. For each IMF and residue, when we want to predict the value at time t (marked as $X_t^{i,j}$), the corresponding input features include the data points of the whole previous day ($X_{t-48}^{i,j}$ to $X_{t-96}^{i,j}$) and the same day in the previous week ($X_{t-336}^{i,j}$ to $X_{t-384}^{i,j}$). This is the content of the input matrix for the RVFLs analyzing the sub-series. As a result, the predicted values of all the sub-series for time t are:

$$\{\hat{X}_t^{i,j} : 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n+1\} \tag{12}$$

As mentioned above, to construct the input matrix for the incremental RVFL, we combine the predicted values with original input signal (X_{t-48} to X_{t-96}, and X_{t-336} to X_{t-384}), and the recently observed temperature data (T_{t-1}).

Temperature data of above five states in Australia is also considered for electric load forecasting in this study, which is provided by Australian Bureau of Meteorology [48]. Specifically, the temperature data from high demand areas in each state is considered: Sydney and Canberra for NSW, Horbat and Launceston airport for TAS, Brisbane for QLD, Adelaide for SA, and Melbourne for VIC. The daily maximum temperature and daily minimum temperature datasets are used as additional features to help improve the performance of the proposed electric load forecasting method. The influence of temperature data will be discussed in Section 5.3.

4.2. Variations of RVFL network

There are eight different RVFL network configurations by varying the network components: input layer and hidden layer biases, along with input-output connections (functional links). In [17,19], the authors compared the performance of all the variations of RVFL network for both regression and classification problems. According to the conclusions made, the direct input-output connections can improve the performance of RVFL network significantly. Moreover, although the input and hidden layer biases have little influence on the performance, it is still necessary to retain biases to ensure the neural networks function properly as a universal approximator. Hence, in this work, the RVFL network with direct input-output connections and input/hidden layer biases was selected for our proposed EMD-RVFL network. Moreover, RWSLFN (the RVFL variant without functional links) is also included to perform comparison and verify the results reported in [17,19].

4.3. Methodology

For the time series load datasets, all the training and testing values are linearly scaled to [0, 1]. The scaling is done separately and independently for each year. The scaling formula is:

$$\bar{y}_i = \frac{y_{max} - y_i}{y_{max} - y_{min}} \tag{13}$$

where y_i is the value of the ith point of the dataset, $\bar{y}_i$ is the corresponding scaled value, and y_{max} and y_{min} are the maximum and minimum values of the dataset, respectively.

To implement the simulation, the deep learning toolbox was used for neural networks, including SLFN and EMD based SLFN (EMD-SLFN) [49]. RF and EMD-RF were developed from the function "TreeBagger" in Matlab. We set the parameter "NumPredictorsToSample" as one third of the number of input features to invoke RF algorithm. RVFL, EMD-RVFL and the proposed incremental DWT-EMD-RVFL were developed by the authors in Matlab based on the work in [19].

For SLFN and EMD-SLFN, the size of neural networks is determined by the size of input vector. Based on the experience in our previous work [32], the number of iterations for back propagation is set as 500 to avoid overfitting. For RF and EMD based RF, because of the same reasons as above, the number of decision trees is set as 500. For RVFL, EMD-RVFL and the proposed method, according to suggestion in [17,19], the randomization used a uniform distribution in $[-1, 1]$, the number of hidden neurons is selected over 1000: 10000 with a step-size of 1000.

4.4. Error measurement

There are two error measures being used to evaluate the performance of learning models in this work: Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), which are defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i' - y_i)^2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i' - y_i}{y_i}\right| \tag{14}$$

Table 2
Performance comparison between RVFL variants with and without direct links.

Dataset	Year	Metrics	Number of hidden neurons 100 +link	100 −link	1000 +link	1000 −link	10,000 +link	10,000 −link
QLD	2013	RMSE	**325.761**	330.520	**318.559**	318.747	**314.144**	314.150
		MAPE	**4.069%**	4.150%	3.968%	**3.967%**	3.926%	3.926%
	2014	RMSE	**414.871**	424.211	403.132	**402.781**	396.435	**396.394**
		MAPE	**4.989%**	5.125%	4.810%	**4.804%**	4.735%	**4.734%**
	2015	RMSE	**378.435**	384.532	368.445	**368.360**	**362.576**	362.613
		MAPE	**4.461%**	4.549%	4.309%	**4.307%**	**4.249%**	4.250%
NSW	2013	RMSE	**641.722**	648.763	**627.046**	627.223	**617.595**	617.609
		MAPE	**6.605%**	6.707%	**6.318%**	6.323%	**6.198%**	6.199%
	2014	RMSE	**624.748**	627.883	621.408	**620.368**	624.098	**624.058**
		MAPE	**6.159%**	6.237%	6.067%	**6.054%**	6.062%	6.062%
	2015	RMSE	728.598	**725.640**	715.637	**712.831**	711.597	**711.240**
		MAPE	**6.861%**	6.927%	6.666%	**6.639%**	6.629%	**6.626%**
TAS	2013	RMSE	**75.911**	77.032	**75.589**	75.744	**74.974**	74.992
		MAPE	**5.509%**	5.618%	**5.469%**	5.484%	**5.395%**	5.397%
	2014	RMSE	**72.788**	72.957	**72.219**	73.725	73.115	**72.232**
		MAPE	**5.507%**	5.588%	**5.505%**	5.518%	**5.450%**	5.451%
	2015	RMSE	**69.146**	70.172	**69.717**	69.917	**68.882**	68.908
		MAPE	**4.934%**	5.029%	**4.969%**	4.985%	**4.894%**	4.896%
SA	2013	RMSE	**198.108**	200.202	**194.991**	195.031	**193.207**	193.230
		MAPE	**11.961%**	12.099%	**11.759%**	11.763%	**11.652%**	11.653%
	2014	RMSE	**177.654**	178.835	**177.390**	177.447	**177.201**	177.206
		MAPE	**11.215%**	11.316%	**11.095%**	11.098%	**11.030%**	11.031%
	2015	RMSE	**245.944**	249.006	**243.512**	243.669	243.061	**243.054**
		MAPE	**12.838%**	12.987%	**12.685%**	12.693%	12.486%	12.486%
VIC	2013	RMSE	**564.909**	568.801	**555.698**	555.947	**548.845**	548.892
		MAPE	**8.734%**	8.853%	**8.473%**	8.479%	8.314%	8.314%
	2014	RMSE	**563.379**	566.304	560.036	**559.695**	**558.722**	558.729
		MAPE	**9.735%**	9.780%	9.603%	**9.593%**	9.535%	9.535%
	2015	RMSE	**575.090**	581.875	**562.418**	562.751	**556.468**	556.501
		MAPE	**9.245%**	9.370%	**8.940%**	8.946%	**8.816%**	8.817%
Win-tie-lose			**29-0-1**		**19-0-11**		**18-5-7**	

where y_i' is the predicted value of corresponding y_i, and n is the number of data points in the testing TS dataset.

Except for above traditional error measures, readers are also suggested to try some new error measures. For example, in [50], a new error measure is proposed based on finding a restricted permutation of the original forecast that minimizes the point-wise error, which may be able to reduce the so-called "double penalty" effect.

5. Results and discussion

5.1. Effect of functional links and number of hidden neurons

We focus on the effect of the direct connections in RVFL network in this section. In order to make a reasonable comparison, the other issues (e.g. parameters, activation functions, range of randomization, etc.) in RVFL with and without functional links (RWSLFN) were set the same, except for the number of hidden neurons. Table 2 shows the prediction results using RVFL and RWSLFN with different number of hidden neurons. The row "win-tie-lose" in the bottom means the number of times that RVFL wins, ties, and loses to RWSLFN, respectively. Several observations can be concluded by this comparison:

1. The number of hidden neurons have important influence on the performance of RVFL variants. Sufficient number of hidden neurons can ensure the completeness of information obtained.
2. The superiority of functional links can also be observed, which may due to the reason that the direct links can serve as a regularization for the randomization [17].
3. However, the advantage of functional links decreases as the number of hidden neurons increasing, which means that the functional links and number of hidden neurons complement each other for information gaining.

Therefore, taking model complexity into consideration, RVFL with direct links and reasonable number of hidden neurons is recommended.

In this study, several hundreds or thousands hidden neurons are sufficient for load forecasting. However, for practical application, the number of hidden neurons should be chosen based on the characteristics of different datasets, as well as the requirement of accuracy and efficiency. We recommend readers try different RVFL models with various numbers of hidden neurons, and choose the most suitable one according to the actual requirements.

5.2. Effect of incremental learning

In this part, we concentrate on the effectiveness of incremental learning. As we have mentioned above, for each dataset, the first nine months are used for training, and the last three months are used for testing. For incremental learning, the data points in testing subset are imported to the network one by one, thereby stepwise updating the model obtained in training phase. Meanwhile, for non-incremental learning, the model is fixed during testing phase. The performance comparison of incremental and non-incremental learning is shown in Table 3. All the models have 100 hidden neurons with functional links. Same with Section 5.1, the row "win-tie-lose" in the bottom means the number of times that incremental learning wins, ties, and loses to non-incremental learning, respectively. From the comparison results "$30-0-0$", we can conclude that incremental learning is definitely beneficial for short term electric load TS forecasting with RVFL and its ensemble models. Moreover, it also worth noting that the proposed

Table 3
Performance comparison between incremental and non-incremental learning. I stands for incremental, and N stands for non-incremental.

Dataset	Year	Metrics	Learning models					
			RVFL		EMD-RVFL		DWT-EMD-RVFL	
			I	N	I	N	I	N
QLD	2013	RMSE	**318.217**	325.761	**278.983**	290.260	**267.111**	285.491
		MAPE	**3.968%**	4.069%	**3.580%**	3.733%	**3.326%**	3.545%
	2014	RMSE	**394.353**	414.871	**322.909**	343.685	**328.702**	356.075
		MAPE	**4.778%**	4.989%	**3.988%**	4.193%	**3.932%**	4.238%
	2015	RMSE	**370.099**	378.435	**353.267**	366.844	**308.760**	325.967
		MAPE	**4.331%**	4.461%	**4.150%**	4.325%	**3.656%**	3.861%
NSW	2013	RMSE	**614.786**	641.722	**513.276**	538.169	**516.216**	530.103
		MAPE	**6.228%**	6.605%	**5.177%**	5.497%	**5.075%**	5.539%
	2014	RMSE	**604.592**	624.748	**520.422**	552.439	**496.081**	528.685
		MAPE	**5.843%**	6.159%	**5.136%**	5.484%	**4.674%**	4.967%
	2015	RMSE	**674.971**	728.598	**648.562**	697.413	**553.518**	626.850
		MAPE	**6.311%**	6.861%	**6.031%**	6.482%	**5.060%**	5.681%
TAS	2013	RMSE	**74.955**	75.911	**65.979**	67.173	**62.569**	65.430
		MAPE	**5.440%**	5.509%	**4.744%**	4.841%	**4.535%**	4.739%
	2014	RMSE	**70.873**	72.788	**67.803**	69.545	**61.880**	63.719
		MAPE	**5.346%**	5.507%	**5.068%**	5.217%	**4.608%**	4.755%
	2015	RMSE	**67.156**	69.146	**66.033**	68.196	**61.610**	63.646
		MAPE	**4.767%**	4.934%	**4.736%**	4.911%	**4.371%**	4.531%
SA	2013	RMSE	**191.545**	198.108	**186.902**	195.089	**163.689**	171.889
		MAPE	**11.405%**	11.961%	**11.068%**	11.715%	**9.686%**	10.179%
	2014	RMSE	**173.237**	177.654	**169.683**	174.251	**151.902**	156.496
		MAPE	**10.815%**	11.215%	**10.547%**	10.945%	**9.373%**	9.712%
	2015	RMSE	**241.125**	245.944	**235.117**	240.803	**221.009**	241.887
		MAPE	**12.347%**	12.838%	**12.024%**	12.536%	**11.663%**	12.817%
VIC	2013	RMSE	**535.626**	564.909	**511.024**	538.539	**437.622**	469.076
		MAPE	**8.104%**	8.734%	**7.550%**	7.989%	**6.321%**	6.888%
	2014	RMSE	**522.461**	563.379	**469.280**	506.752	**400.502**	428.211
		MAPE	**8.773%**	9.735%	**7.752%**	8.635%	**6.426%**	6.982%
	2015	RMSE	**555.641**	575.090	**541.444**	562.414	**486.335**	516.758
		MAPE	**8.779%**	9.245%	**8.549%**	9.022%	**7.587%**	8.022%
Win-tie-lose			**30-0-0**		**30-0-0**		**30-0-0**	

DWT-EMD-RVFL outperforms RVFL and EMD-RVFL in every case. The statistical testing is conducted and discussed in Section 5.3.

5.3. Performance comparison with benchmarks

In this section, the performance of the proposed incremental DWT-EMD-RVFL approach is evaluated by comparing with several benchmark methods. First of all, the persistence method, which is the simplest forecasting method, is employed as the baseline for comparing the performance of learning models in this work. For persistence method, the load value at the same hour of last day is used as the forecast for each of the 24 hours of next day, which works well because of the highly periodic characteristic of electric load TS. Moreover, a modified version of another benchmark method GLMLF-B (General Linear Model based Load Forecaster - Benchmark) [51] is also employed, which can offer a higher baseline for the other machine learning algorithms. In this study, since the forecasting horizon is one day (or 48 steps), to predict X_t, we use the corresponding hour, day, month, as well as the load value X_{t-48} at the same time in the previous day, as input features to construct a multiple linear regression (MLR) model, which shares similar ideas with GLMLF-B. Except the persistence method, all the other benchmark models have made use of the temperature data, which can improve the performance of forecasting.

The prediction results for one-day-ahead electric load forecasting are shown in Table 4. The numbers in bold mean that the corresponding method achieves the best performance for this dataset under this performance measurement. The prediction results generated by the proposed method without temperature data are also recorded, which can be compared with the results from the proposed method with temperature data. From Table 4, we can clearly see that the proposed method achieves the best performance for every case except QLD. The MAPE values of the load prediction for QLD are significantly lower than the ones for all the other regions. This phenomenon proves the fact that the pattern of the load data of QLD is much more stable and simpler than the patterns of the load data of other regions, which is easier for the benchmarks to analyze. Therefore, our proposed method cannot show much advantage on the QLD datasets.

Moreover, statistical tests are employed to give a detail analysis about the performance differences among all the learning models. The Friedman test ranks the algorithms for each dataset separately, and then assign average ranks in case of ties. The null-hypothesis states that all the algorithms have the same performance. If the null-hypothesis is rejected, in order to tell whether the performances of two among totally k learning models are significantly different, the Nemenyi post-hoc test is applied to compare all the learning models with each other. The comparison results of statistical test based on RMSE and MAPE are shown in Figs. 3 and 4, respectively. The methods with better ranks are at the top whereas the methods with worse ranks are at the bottom. It is worth noting that the models within a vertical line whose length is less than or equal to a critical distance have statistically the same performance. The critical distance for Nemenyi test is defined as:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{15}$$

where k is the number of algorithms ($k = 9$ in this experiment), N is the number of data sets ($N = 15$ here), and q_α is the critical value based on the studentized range statistic divided by $\sqrt{2}$ [52].

From the results of simulations and statistical tests, several conclusions can be made:

Table 4
Prediction results for one-day-ahead electric load forecasting.

Dataset	Year	Metrics	Prediction model								Proposed	
			Persistence	GLMLF-B [51]	SLFN [39]	RF [53]	RVFL [15,16]	EMD-SLFN [35]	EMD-RF [54]	EMD-RVFL [55]	−T	+T
QLD	2013	RMSE	492.589	355.503	307.892	278.511	281.583	230.600	266.686	244.820	233.221	**218.329**
		MAPE	6.348%	4.323%	3.912%	3.449%	3.518%	2.953%	3.273%	3.160%	2.967%	**2.797%**
	2014	RMSE	588.706	399.927	373.318	334.405	342.874	**256.671**	315.772	270.137	284.766	263.873
		MAPE	7.144%	5.073%	4.655%	4.100%	4.139%	3.243%	3.716%	3.442%	3.440%	**3.215%**
	2015	RMSE	553.077	369.409	368.554	328.787	329.461	318.157	341.356	304.317	277.861	**261.799**
		MAPE	6.633%	4.749%	4.546%	3.933%	3.835%	3.782%	3.806%	3.673%	3.341%	**3.170%**
NSW	2013	RMSE	901.511	643.155	614.006	556.797	608.919	519.295	497.665	438.721	436.890	**403.403**
		MAPE	8.657%	6.426%	6.067%	5.315%	6.139%	5.234%	4.773%	4.386%	4.240%	**4.040%**
	2014	RMSE	878.077	632.275	620.177	585.587	531.664	492.079	527.581	435.748	419.416	**380.410**
		MAPE	8.595%	6.069%	5.834%	5.486%	5.033%	4.758%	4.862%	4.357%	3.906%	**3.629%**
	2015	RMSE	1055.406	713.827	630.819	575.037	549.645	550.728	563.351	520.891	466.842	**400.443**
		MAPE	9.689%	6.669%	5.953%	5.222%	5.303%	5.289%	5.185%	5.033%	4.368%	**3.881%**
TAS	2013	RMSE	97.855	84.095	67.148	68.648	68.650	64.216	65.411	61.119	58.175	**56.725**
		MAPE	6.870%	6.042%	4.771%	5.016%	5.080%	4.823%	4.746%	4.468%	4.228%	**4.157%**
	2014	RMSE	86.863	82.193	69.768	67.517	67.683	62.987	67.395	64.239	58.721	**57.025**
		MAPE	6.512%	6.324%	5.305%	5.160%	5.219%	4.796%	5.142%	4.886%	4.361%	**4.298%**
	2015	RMSE	83.993	75.808	59.962	61.599	64.554	60.767	64.089	60.305	58.687	**55.837**
		MAPE	6.035%	5.452%	4.299%	4.414%	4.561%	4.352%	4.487%	4.353%	4.169%	**3.999%**
SA	2013	RMSE	288.779	206.009	167.367	166.099	166.461	168.514	150.116	163.474	146.936	**139.500**
		MAPE	16.022%	12.617%	10.118%	10.123%	10.377%	10.322%	9.734%	10.118%	8.740%	**8.572%**
	2014	RMSE	242.356	175.991	147.681	154.063	159.824	154.693	152.912	155.543	142.976	**133.605**
		MAPE	14.419%	11.659%	9.540%	9.798%	10.219%	9.993%	9.323%	9.956%	8.769%	**8.351%**
	2015	RMSE	365.141	244.246	196.482	173.032	183.903	176.610	225.926	177.217	192.189	**163.204**
		MAPE	18.395%	12.677%	11.168%	9.374%	10.073%	9.961%	11.938%	9.748%	10.122%	**9.086%**
VIC	2013	RMSE	781.683	584.131	560.457	487.053	537.208	439.463	477.357	451.230	376.096	**360.696**
		MAPE	10.711%	9.159%	8.531%	7.073%	8.269%	6.241%	6.895%	6.697%	5.368%	**5.241%**
	2014	RMSE	719.402	621.930	634.557	530.069	487.838	463.882	498.345	422.682	350.499	**335.468**
		MAPE	11.246%	10.079%	10.172%	8.564%	8.019%	7.807%	8.373%	6.872%	5.528%	**5.328%**
	2015	RMSE	874.691	684.131	485.400	479.175	486.965	518.785	486.611	471.469	430.092	**385.754**
		MAPE	12.886%	10.159%	7.669%	7.475%	7.260%	8.130%	7.230%	7.404%	6.661%	**6.106%**

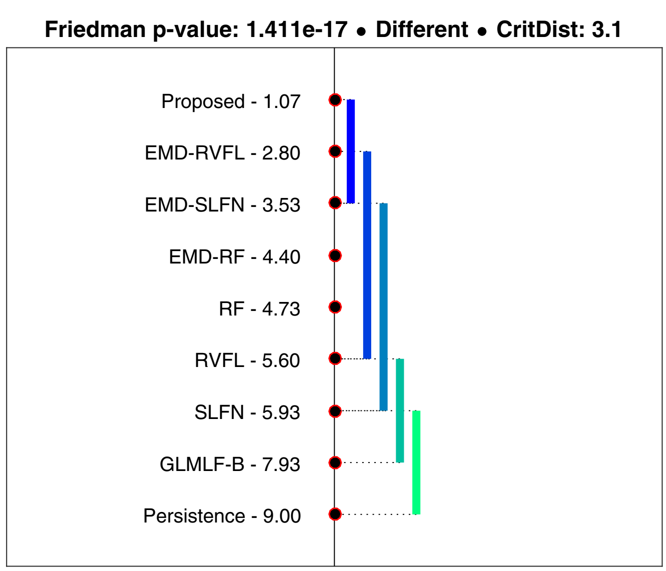

Fig. 3. Nemenyi test for electric load forecasting based on RMSE. The critical distance is 3.1.

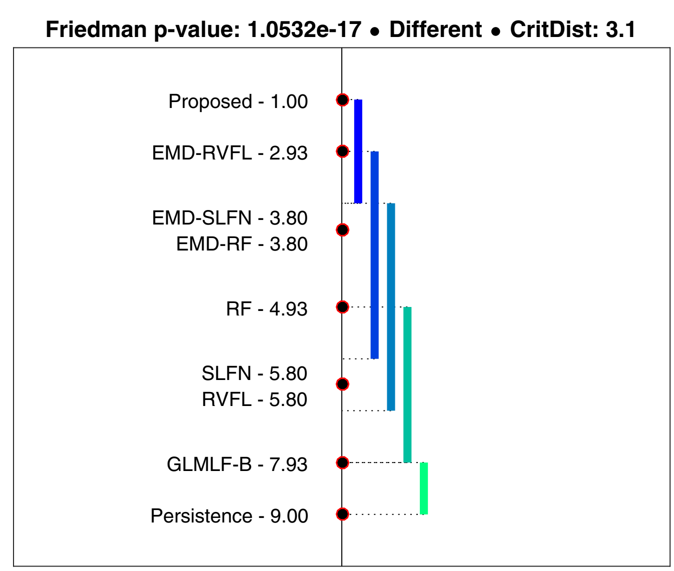

Fig. 4. Nemenyi test for electric load forecasting based on MAPE. The critical distance is 3.1.

1. The original load TS data was modeled by SLFN, RF and RVFL without decomposition. Therefore, the advantages of EMD based ensemble methods can be revealed by performing comparisons.
2. By employing incremental learning, RVFL based models have comparable (or even better) performance with traditional NN and RF based models with less computation time.
3. The proposed incremental DWT-EMD based RVFL approach achieves the best rank and significantly outperform the non-EMD based benchmarks and EMD-RF with a 95% confidence.

In order to find where in the forecast the proposed method offers a key advantage in performance, a comparison between the forecasting results for original RVFL and the proposed method was conducted. Comparisons of predicted values with actual values for the proposed method and RVFL network are shown in Figs. 5 and 6, respectively. This part of load data is selected from the testing

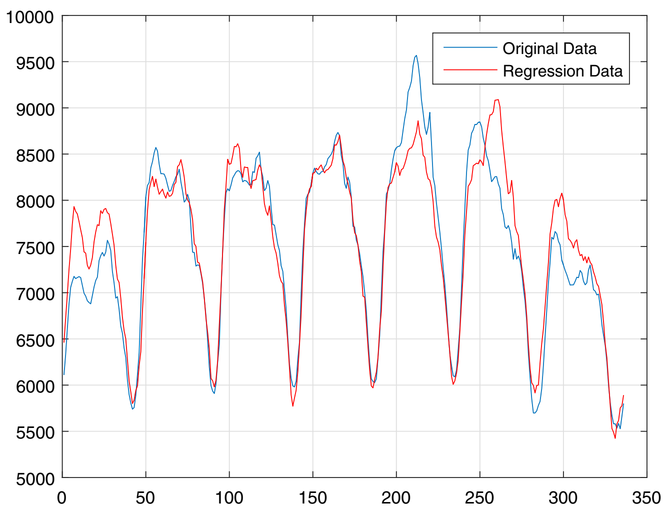

Fig. 5. Comparison of predicted values with actual values for the proposed method.

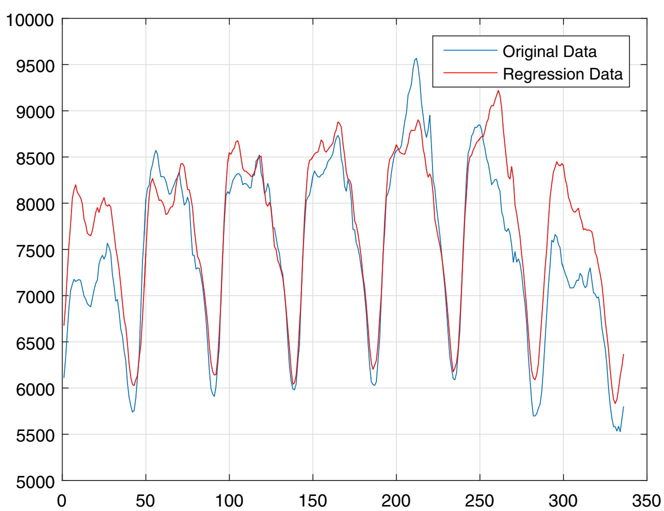

Fig. 6. Comparison of predicted values with actual values for RVFL.

dataset of NSW of the year 2013, with a time window of one week (from Sunday to Saturday).

From the comparison results, we can conclude that the key improvements caused by the proposed method are located on the data points during the weekends. In fact, the difference between the electric load in weekdays and weekends is one of the dominant challenges for prediction methods. Some published works, such as [23], deal with this problem by introducing additional input features of calendar information (e.g. holidays, weekends, etc.). However, in this work, under the help of decomposition algorithms DWT and EMD, as well as the incremental learning, the proposed method can detect the pattern changes caused by weekends, and modify the model by itself.

As we have mentioned in Section 3, a possible reason why the proposed DWT-EMD method performs better than EMD only is that MODWT may help solve the mode mixing problem of EMD. To illustrate our conjecture, we use the electric load data from QLD of the year 2014 as an example. To analyze the frequency spectrum in each sub-signal, Fast Fourier Transform (FFT) is applied to all the IMFs and residue as shown in Fig. 7. Obviously, the first IMF contains the widest band of frequency, which is hard to predict. In fact, in the literature, some works just treat the first IMF component as the noise and simply ignore it. However, in this work, we employ MODWT to help reduce the frequency range in each IMF. Fig. 8 shows the obtained wavelet coefficients and scaling coefficients for the same electric load data. Then we apply EMD on each wavelet based sub-signal and check the frequency spectrum for the first IMF using FFT. Comparing Fig. 9 and the first subfigure in Fig. 7, we can clearly see that the band of frequency for the first IMF obtained from DWT-EMD method is significantly narrower than the one obtained by EMD itself. In other words, after the proposed DWT-EMD decomposition method, the obtained subsignals are much easier to be analyzed and modeled by learning methods compared with signals decomposed by EMD only. Therefore, the proposed DWT-EMD based ensemble learning method can achieve better performance, as demonstrated by the simulation results shown in this section.

In our previously published paper [32], the benchmarks are evaluated using the same load datasets form AEMO of the year 2013. The prediction results for one day ahead load forecasting are shown in Table 3 in that paper. For each area, four months were chosen to reflect the factors of different seasons: January, April, July and October. To show the performance of the proposed method and the benchmark models in different seasons, the same comparison simulations are implemented in this study using load dataset from NSW of the year 2015. The results are shown in Table 5. From the results, by taking the factors of different seasons into consideration, we can tell that the benchmark methods performs relatively stable for the same dataset in different seasons. Moreover, in this case, our proposed method still outperforms all benchmarks models significantly.

5.4. Computation time comparison

The computation time of benchmark methods for electric load forecasting using the datasets of year 2015 is shown in Fig. 10. It is easy to conclude that RVFL is much faster than ANN and RF. ANN needs to be iteratively tuned by BP algorithm to convergence to the optimal weights. RF needs to train a group of decision trees. Different from ANN and RF, RVFL has a closed form solution. Benefit from the good efficiency of RVFL, the proposed DWT-EMD-RVFL model also has a reasonable fast computation speed.

6. Comparative experiment

In this section, three comparative experiments were implemented to test the effectiveness of our proposed method. The comparison conditions such as dataset partition and feature selection were kept the same for the reported methods and the proposed method.

6.1. Comparative experiment with semi-parametric additive model

In [11], Fan and Hyndman proposed a semi-parametric additive model for short-term load forecasting, which has been successfully used by AEMO to forecast the short-term loads of two regions with different characteristics in the Australian National Electricity Market. Specifically, the half-hourly demand datasets from Victoria, Australia of the years from January 2004 to September 2008 were used to train the models. The load datasets from October 2008 to March 2009 were used as out-of-sample test data. Except for the historical demand data, the additional input variables include lagged and future temperatures, and calendar variables. In this work, the simulations were implemented using the same training and testing datasets, thereby offering uniform comparisons. The comparison results are shown in Table 6, which lead to the conclusion that our proposed method has better performance compared with the benchmark models.

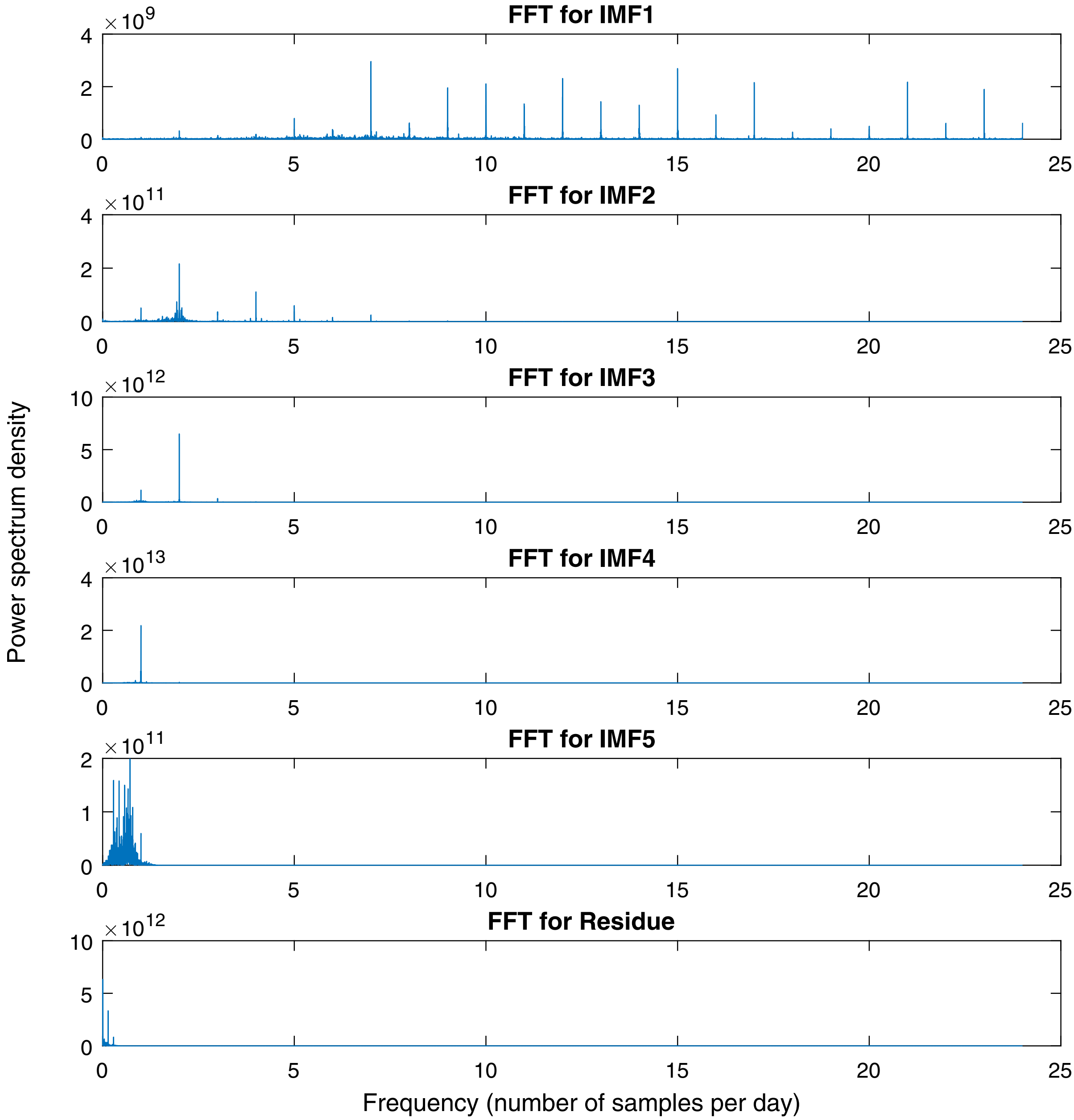

Fig. 7. Frequency spectrum for each sub-signal obtained from EMD.

Table 5
Prediction results for different seasons using the load data from NSW of the year 2015.

Month	Metrics	Prediction model								
		Persistence	GLMLF-B [51]	SLFN [39]	RF [53]	RVFL [15,16]	EMD-SLFN [35]	EMD-RF [54]	EMD-RVFL [55]	Proposed
Jan	RMSE	842.732	612.303	464.061	440.572	428.908	379.871	428.388	403.271	**193.800**
	MAPE	7.393%	5.599%	4.094%	3.527%	3.869%	3.319%	3.328%	3.423%	**1.857%**
Apr	RMSE	769.606	525.145	448.827	437.975	425.228	400.455	441.052	411.820	**212.703**
	MAPE	6.801%	5.259%	4.294%	3.992%	3.936%	4.031%	3.971%	3.861%	**2.030%**
Jul	RMSE	989.372	614.706	501.107	411.863	493.064	440.431	402.973	423.287	**296.743**
	MAPE	9.831%	6.135%	5.145%	4.290%	5.093%	4.719%	4.144%	4.423%	**2.961%**
Oct	RMSE	1620.508	1091.054	1021.127	953.213	1004.394	913.704	911.767	987.330	**659.407**
	MAPE	14.887%	9.404%	8.981%	7.139%	8.863%	7.217%	6.817%	7.035%	**5.934%**

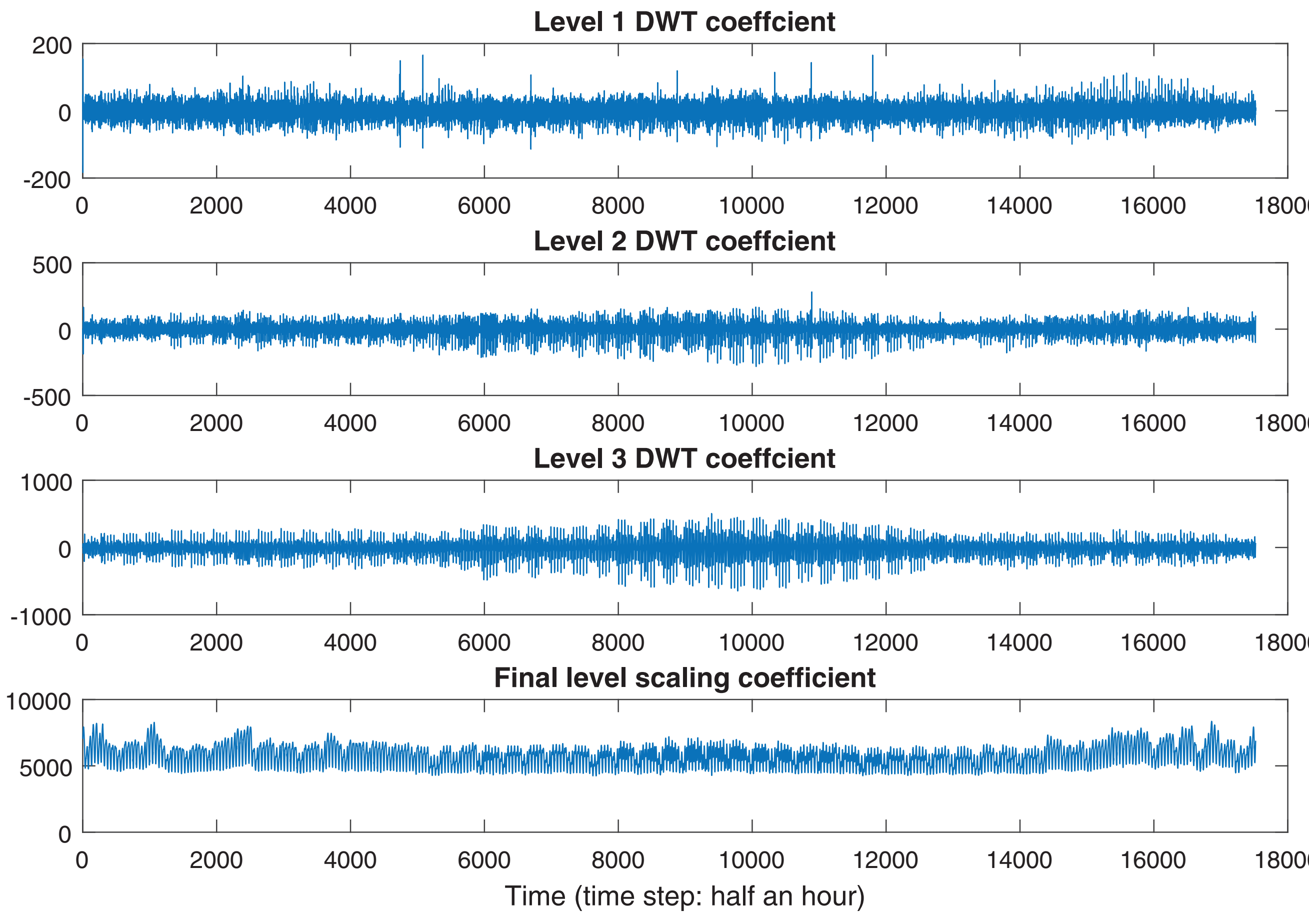

Fig. 8. Wavelet coefficient and scaling coefficient obtained by MODWT.

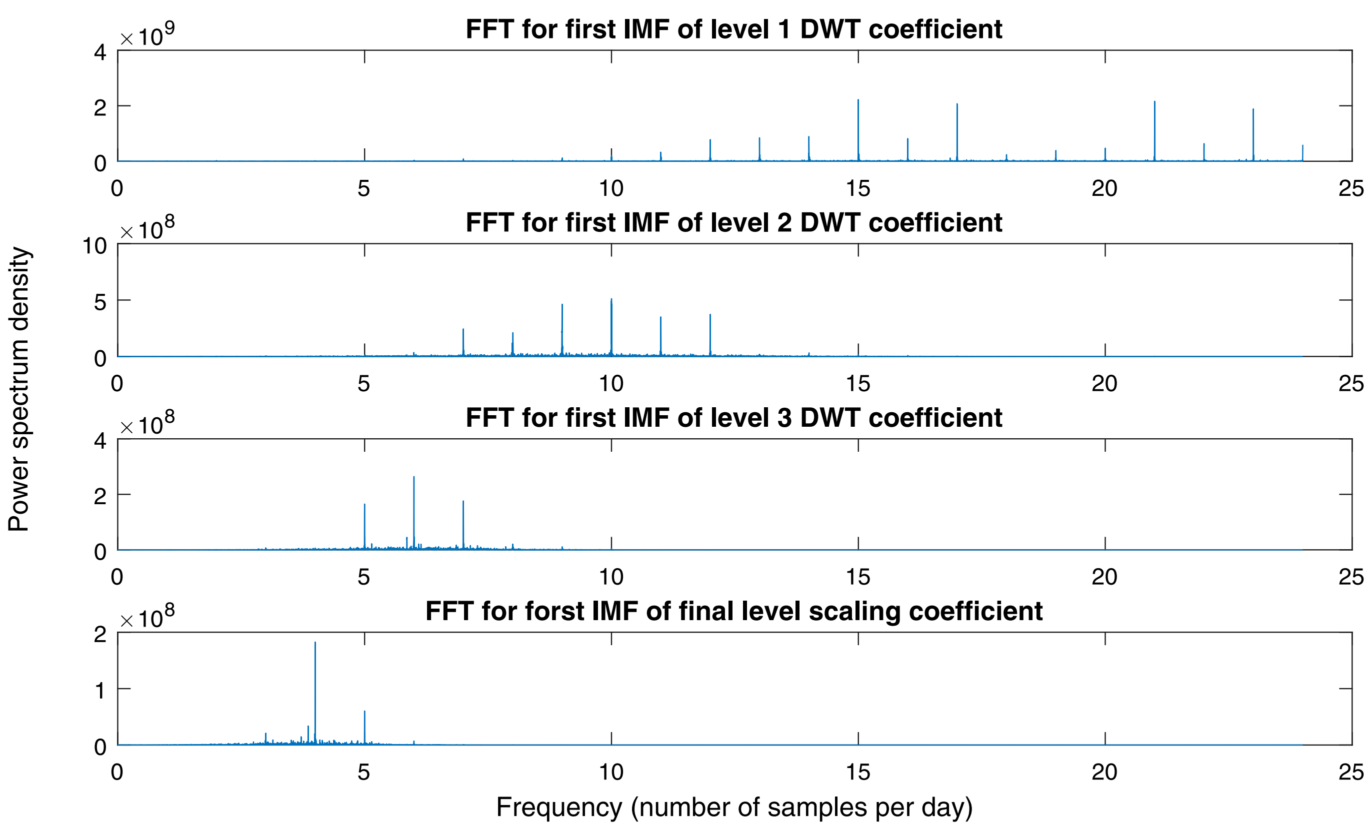

Fig. 9. Frequency spectrum for the first IMF of each wavelet based sub-signal.

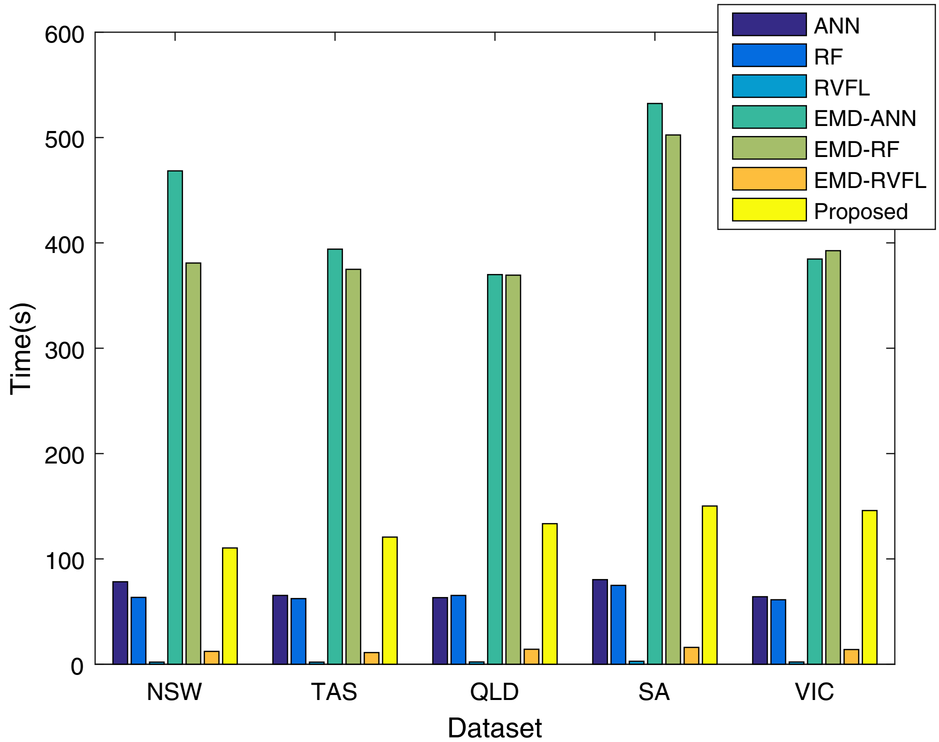

Fig. 10. Computation time of learning models for electric load forecasting.

Table 6
Forecasting results for comparative experiment one. The results of additive model, ANN and Hybrid model are obtained from [11].

Month	Proposed		Additive model [11]		ANN		Hybrid	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
Oct	**77.16**	**1.39%**	88.55	1.66%	134.87	2.57%	121.83	2.15%
Nov	**65.34**	**1.19%**	94.33	1.74%	140.52	2.63%	123.50	2.12%
Dec	**61.76**	**1.18%**	79.89	1.55%	126.39	2.49%	116.34	2.17%
Jan	**90.34**	**1.45%**	110.21	1.88%	168.04	2.81%	126.73	2.14%
Feb	**62.21**	**1.11%**	96.84	1.64%	139.68	2.37%	119.07	1.95%
Mar	**62.58**	**1.14%**	87.45	1.59%	123.21	2.29%	116.49	1.94%
Average	**69.90**	**1.24%**	92.82	1.68%	138.79	2.53%	120.66	2.08%

6.2. Comparative experiment with dART&HS-ARTMAP model

In [56], a neural network based on adaptive resonance theory, named distributed ART (adaptive resonance theory) & HS-ARTMAP (Hyper-spherical ART-MAP network) was applied to the electric load forecasting problem. The distributed ART combines the stable fast learning capabilities of winner-take-all ART systems with the noise tolerance and code compression capabilities of multi-layer perceptions. The HS-ARTMAP, a hybrid of an RBF (Radial Basis Function)-network-like module which uses hyper-sphere basis function substitute the Gaussian basis function and an ART-like module, performs incremental learning capabilities in function approximation problem. The electricity load data in the head days in January from 1999 to 2009 of New South Wales, Australia was used to conduct the experiment. The inputs of the neural network coming from historic load values were composed of two parts: one is the data p steps before the determined instant, and the other is the data in the same time of past q years. For one-day ahead forecasting, $q = 1, p = 18$ was selected for the simulation. The last year data was used for testing, while the remaining part was used for training. Table 7 shows the forecasting results obtained by the proposed method and benchmarks in [56]. Obviously, the proposed method outperforms all the benchmark methods, including Improved BP, HS-ARTMAP and dART&HS-ARTMAP.

6.3. Comparative experiment with DSHW and ARIMA variates

In the second comparative experiment, the transmission level data obtained from British Columbia with an online service was used for comparison. The system has 292 substations, powering the 4.4 M inhabitants of British Columbia. In [8], the data from 2004 throughout 2010 was used for simulation, among which the last year of the TS being divided equally between validation and testing, while the remainder of the TS being used for training. Several benchmark models were tested in [8], including double seasonal Holt-Winters (DSHW) exponential smoothing algorithm, and variations of autoregressive integrated moving average (ARIMA). In this paper, the same TS signal was analyzed by our proposed method with the same comparison conditions. From Table 8, which shows the comparison results, we can conclude that our proposed method outperforms the benchmark DSHW and ARIMA variates significantly.

Table 7
Forecasting results for comparative experiment two.

Metric	Improved BP [39]	HS-ARTMAP [57]	dART&HS-ARTMAP [56]	Proposed
MAPE	3.40%	2.87%	1.91%	**1.88%**

Table 8
Forecasting results for comparative experiment three. The ARIMA variates have weekly preprocessing step presented in [8].

Metric	ARX	ARX-24	ARIMAX	DSHW [58]	Proposed
RMSE	170	192	225	250	**108**
MAPE	1.7%	1.9%	2.2%	2.6%	**1.2%**

7. Conclusion

In this paper, a hybrid incremental learning approach is presented for short-term electric load forecasting, which composed of Discrete Wavelet Transform (DWT), Empirical Mode Decomposition (EMD) and Random Vector Functional Link (RVFL) network. Fifteen electric load datasets from AEMO were used for evaluating the performance of the proposed method by comparing with several benchmarks. Moreover, two comparative experiments were also implemented to verify the effectiveness of the proposed method. Based on the experiment results, the following conclusions are made:

1. Both sufficient number of hidden neurons and functional links can benefit the overall performance of RVFL networks. Taking model complexity into consideration, RVFL with direct links and reasonable number of hidden neurons is recommended.
2. Incremental learning is beneficial for short term electricity load TS forecasting with RVFL and its ensemble models.
3. The proposed DWT-EMD based ensemble approach outperforms EMD based and single structure models.
4. The proposed incremental DWT-EMD based RVFL approach achieves the best rank and significantly outperforms the non-EMD based benchmarks and EMD-RF with a 95% confidence.

For future work, the proposed decomposition approach can be combined with various learning algorithms, such as support vector regression, kernel ridge regression, and even deep learning models. These outcome ensemble models can be tested on various applications, including renewable energy and power system related financial data. For example, electricity price data has a close relationship with electricity load data, thus can be analyzed by the learning models sharing the same concepts of the proposed method. Moreover, probabilistic load forecasting methods, which provides electric load forecasting output in the form of intervals, scenarios, density functions or probabilities, can also been developed by combining the proposed accurate point forecasting models with good classification methods.

Acknowledgment

This project is funded by the National Research Foundation Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

[1] H.K. Alfares, M. Nazeeruddin, Electric load forecasting: literature survey and classification of methods, Int. J. Syst. Sci. 33 (2002) 23–34.
[2] T. Hong, Crystal Ball Lessons in Predictive Analytics, EnergyBiz, 2015, pp. 35–37.
[3] I. Koprinska, M. Rana, V.G. Agelidis, Correlation and instance based feature selection for electricity load forecasting, Knowl. Based Syst. 82 (2015) 29–40.
[4] J.C. Sousa, H.M. Jorge, L.P. Neves, Short-term load forecasting based on support vector regression and load profiling, Int. J. Energy Res. 38 (3) (2014) 350–362.
[5] J.W. Taylor, Short-term electricity demand forecasting using double seasonal exponential smoothing, J. Oper. Res. Soc. 54 (8) (2003) 799–805.
[6] C.-M. Lee, C.-N. Ko, Short-term load forecasting using lifting scheme and ARIMA models, Expert Syst. Appl. 38 (2011) 5902–5911.
[7] T. Hong, Short Term Electric Load Forecasting, Ph.D. thesis, Graduate Program of Operation Research and Dept. of Electrical and Computer Engineering, North Carolina State University, 2010.
[8] B.A. Høverstad, A. Tidemann, H. Langseth, P. Öztürk, Short-term load forecasting with seasonal decomposition using evolution for parameter tuning, IEEE Trans. Smart Grid 6 (4) (2015) 1904–1913.
[9] C.-N. Yu, A. Tidemann, P. Mirowski, T.K. Ho, A sparse coding approach to household electricity demand forecasting in smart grids, IEEE Trans. Smart Grid PP (2016) 1–11.
[10] Y. Goude, R. Nedellec, N. Kong, Local short and middle term electricity load forecasting with semi-parametric additive models, IEEE Trans. Smart Grid 5 (1) (2014) 440–446.
[11] S. Fan, R. Hyndman, Short-term load forecasting based on a semi-parametric additive model, IEEE Trans. Power Syst. 27 (1) (2012) 134–141.
[12] M. Leshno1, V.Y. Lin, A. Pinkus, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural Netw. 6 (1993) 861–867.
[13] G.A. Darbellay, M. Slama, Forecasting the short-term demand for electricity: do neural networks stand a better chance? Int. J. Forecast. 16 (2000) 71–83.
[14] G. Zhang, B.E. Patuwo, M.Y. Hu, Forecasting with artificial neural networks: the state of the art, Int. J. Forecast. 14 (1998) 35–62.
[15] Y.-H. Pao, S.M. Phillips, D.J. Sobajic, Neural-net computing and the intelligent control of systems, Int. J. Control 56 (1992) 263–289.
[16] Y.-H. Pao, G.-H. Park, D.J. Sobajic, Learning and generalization characteristics of the random vector functional-link net, Neurocomputing 6 (1994) 163–180.
[17] Y. Ren, P.N. Suganthan, N. Srikanth, G. Amaratunga, Random vector functional link network for short-term electricity load demand forecasting, Inf. Sci. 367–368 (2016) 1078–1093.
[18] W.F. Schmidt, M.A. Kraaijveld, R.P.W. Duin, Feedforward neural networks with random weights, in: Proceedings of the IAPR International Conference on Pattern Recognition Conference B: Pattern Recognition Methodology and Systems, 1992, pp. 1–4.
[19] L. Zhang, P.N. Suganthan, A comprehensive evaluation of random vector functional link networks, Inf. Sci. 367–368 (2016) 1094–1105.
[20] T.G. Dietterich, Ensemble methods in machine learning, in: Multiple Classifier Systems, Springer, 2000, pp. 1–15.
[21] Y. Ren, L. Zhang, P. Suganthan, Ensemble classification and regression-recent developments, applications and future directions [review article], IEEE Comput. Intell. Mag. 11 (1) (2016) 41–53.
[22] C. Guan, P.B. Luh, L.D. Michel, Y. Wang, P.B. Friedland, Very short-term load forecasting: wavelet neural networks with data pre-filtering, IEEE Trans. Power Syst. 28 (1) (2013) 30–41.
[23] R.-A. Hooshmand, H. Amooshahi, M. Parastegari, A hybrid intelligent algorithms based short-term load forecasting approach, Int. J. Electr. Power Energy Syst. 45 (2013) 313–324.
[24] A. Abdoos, M. Hemmati, A.A. Abdoos, Short term load forecasting using a hybrid intelligent method, Knowl. Based Syst. 76 (2015) 139–147.
[25] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, in: Roy. Soc. London A, vol. 454, 1998, pp. 903–995.
[26] X. Geng, K. Smith-Miles, Incremental Learning, Springer US, Boston, MA, pp. 731–735.
[27] J.L. Elman, Learning and development in neural networks: the importance of starting small, Cognition 48 (1993) 71–99.
[28] G. Grmanová, P. Laurinec, V. Rozinajová, A.B. Ezzeddine, M. Lucká, P. Lacko, P. Vrablecová, P. Návrat, Incremental ensemble learning for electricity load forecasting, Acta Polytech. Hungarica 13 (2016) 97–117.
[29] Y. Yang, J. Che, Y. Li, Y. Zhao, S. Zhu, An incremental electric load forecasting model based on support vector regression, Energy 113 (2016) 796–808.
[30] A. Gepperth, B. Hammer, Incremental learning algorithms and applications, in: Proc. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. (ESANN2016), Bruges, Belgium, 2016.
[31] Y. Ren, P.N. Suganthan, N. Srikanth, A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods, IEEE Trans. Sustainable Energy 6 (1) (2015) 236–244.
[32] X. Qiu, Y. Ren, P.N. Suganthan, G.A.J. Amaratumga, Empirical mode decomposition based ensemble deep learning for load demand time series forecasting, Appl. Soft Comput. 54 (2017) 246–255.
[33] X. Fan, Y. Zhu, The application of empirical mode decomposition and gene expression programming to short-term load forecasting, in: Proc. International Conference on Natural Computation (ICNC'10), volume 8, Yantai, China, 2010, pp. 4331–4334.

[34] L. Ye, P. Liu, Combined model based on emd-svm for short-term wind power prediction, in: Proc. Chinese Society for Electrical Engineering (CSEE), vol. 31, 2011, pp. 102–108.
[35] H. Liu, C. Chen, H. Tian, Y. Li, A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks, Renew. Energy 48 (2012) 545–556.
[36] H. Liu, H.-Q. Tian, C. Chen, Y.-F. Li, A hybrid statistical method to predict wind speed and wind power, Renew. Energy 35 (2010) 1857–1861.
[37] J.P.S. Catalao, H.M.I. Pousinho, V.M.F. Mendes, Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting, IEEE Trans. Power Syst. 26 (2011) 137–144.
[38] Y. Chen, M.Q. Feng, A technique to improve the empirical mode decomposition in the hilbert-huang transform, Earthquake Eng. Eng. Vib. 2 (2003) 796–808.
[39] S. Haykin, Neural Networks: A Comprehensive Foundation, International edition, Prentice Hall, 1999.
[40] Y. Ren, X. Qiu, P.N. Suganthan, N. Srikanth, G. Amaratunga, Detecting wind power ramp with random vector functional link (rvfl) network, in: Proc. IEEE Symposium Series on Computational Intelligence (CIEL2015), Cape Town, South Africa, 2015.
[41] A. Grossmann, J. Morlet, Decomposition of hardy functions into square integrable wavelets of constant shape, SIAM J. Math. Anal. 15 (1984) 723–736.
[42] D.C. Kiplangat, K. Asokan, K.S. Kumar, Improved week-ahead predictions of wind speed using simple linear models with wavelet decomposition, Renew. Energy 93 (2016) 38–44.
[43] D. Percival, A. Walden, Wavelet Methods for Time Series Analysis, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2006.
[44] D.B. Percival, M. Wang, J.E. Overland, An introduction to wavelet analysis with applications to vegetation time series, Community Ecol. 5 (1) (2004) 19–30.
[45] C. Chen, J. Wan, A rapid learning and dynamic stepwise updating algorithm for flat neural networks and the application to time-series prediction, IEEE Trans. Syst. Man Cybern. Part B 29 (1) (1999) 62–72.
[46] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, Adv. Adapt. Data Anal. 1 (2009) 1–41.
[47] Australian energy market operator, 2016. URL http://www.aemo.com.au/.
[48] Australian bureau of meteorology, 2017. URL http://www.bom.gov.au/.
[49] R.B. Palm, Prediction as a Candidate for Learning Deep Hierarchical Models of Data, 2012 Master's thesis.
[50] S. Haben, J. Ward, D.V. Greetham, C. Singleton, P. Grindrod, A new error measure for forecasts of household-level, high resolution electrical energy consumption, Int. J. Forecast. 30 (2014) 246–256.
[51] T. Hong, P. Pinson, S. Fan, Global energy forecasting competition 2012, Int. J. Forecast. 30 (2) (2014) 357–363.
[52] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[53] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[54] Z. Fan, Y. Zuo, D. Jiang, X. Cai, Prediction of acute hypotensive episodes using random forest based on genetic programming, in: Proc. IEEE Conference on Evolutionary Computation (CEC2015), 2015.
[55] X. Qiu, P.N. Suganthan, G. Amaratunga, Electricity load demand time series forecasting with empirical mode decomposition based random vector functional link network, in: Proc. IEEE Conference on Systems, Man and Cybernetics (SMC2016), Budapest, Hungary, 2016.
[56] Y. Cai, J.Z. Wang, Y. Tang, Y.C. Yang, An efficient approach for electric load forecasting using distributed ART (adaptive resonance theory) & HS-ARTMAP (hyper-spherical ARTMAP network) neural network, Energy 36 (2011) 1340–1350.
[57] M.-C. Su, J. Lee, K.-L. Hsieh, A new ARTMAP-based neural network for incremental learning, Neurocomputing 69 (2006) 2284–2300.
[58] J. Taylor, P. McSharry, Short-term load forecasting methods: an evaluation based on european data, IEEE Trans. Power Syst. 22 (4) (2007) 2213–2219.

Xueheng Qiu received his B.Eng. degree from Nanyang Technological University in 2012. He is currently working towards the Ph.D. degree in the school of Electrical and Electronic Engineering in Nanyang Technological University, His research interests include various ensemble deep learning algorithms for regression and time series forecasting.

Ponnuthurai Nagaratnam Suganthan received the B.A degree, Postgraduate Certificate and M.A degree in Electrical and Information Engineering from the University of Cambridge, UK in 1990, 1992 and 1994, respectively. After completing his Ph.D. research in 1995, he served as a predoctoral Research Assistant in the Dept. of Electrical Engineering, University of Sydney in 1995–96 and a lecturer in the Dept. of Computer Science and Electrical Engineering, University of Queensland in 1996–99. He moved to NTU in 1999. He is an Editorial Board Member of the Evolutionary Computation Journal, MIT Press. He is an associate editor of the IEEE Trans on Cybernetics, IEEE Trans on Evolutionary Computation, Information Sciences (Elsevier), Pattern Recognition (Elsevier) and Int. J. of Swarm Intelligence Research Journals. He is a founding co-editor-in-chief of Swarm and Evolutionary Computation, an Elsevier Journal. His co-authored SaDE paper published in April 2009 won "IEEE Trans. on Evolutionary Computation" outstanding paper award in 2012. Dr. Jane Jing Liang (his former Ph.D. student) won the IEEE CIS Outstanding Ph.D. dissertation award, 2014. His research interests include evolutionary computation, pattern recognition, multi-objective evolutionary algorithms, applications of evolutionary computation and neural networks. His publications have been well cited according to Googlescholar Citations. His SCI indexed publications attracted over 1000 SCI citations in each calendar year 2013, 2014 and 2015. He is a Fellow of the IEEE and an elected AdCom member of IEEE Computational Intelligence Society (2014–2016).

Gehan A. J. Amaratunga received the B.Sc. degree in electrical engineering from the University of Wales, Cardiff, U.K., in 1979, and the Ph.D. degree from the University of Cambridge, Cambridge, U.K., in 1983. He was at academic and research positions at Southampton University, the University of Liverpool, and Stanford University, California. He has long collaborations with industry partners in Europe, the U.S., and Asia. He is also a Co-Founder of four spin-out companies, including Nanoinstruments, which is now part of Aixtron. He is one of four Founding Advisers to the Sri Lanka Institute of Nanotechnology. He is currently the 1966 Professor of Engineering and Head of Electronics, Power, and Energy Conversion at the University of Cambridge. Since 2009, he has been a Visiting Professor in the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He has authored or coauthored more than 500 papers. He holds 25 patents. His current research interests include novel materials and device structures for nanotechnology-enhanced batteries, supercapacitors, solar cells, and power electronics for optimum grid connection of photovoltaic electricity generation systems. Dr. Amaratunga is a Fellow of the Royal Academy of Engineering. He was the recipient of awards from the Royal Academy of Engineering, the Institution of Engineering and Technology, and the Royal Society.