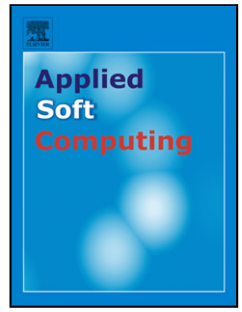
Accepted Manuscript



Title: Empirical Mode Decomposition based Ensemble Deep Learning for Load Demand Time Series Forecasting

Author: id="aut0005" author-id="S1568494617300273-6c3995605dce92efba8bf69f8a61caa6" biographyid="vt0005"> Xueheng Qiu id="aut0010" author-id="S1568494617300273-5d58496303658b4798877c8fabd58081" biographyid="vt0010"> Ye Ren id="aut0015" orcid="0000-0003-0901-5105" author-id="S1568494617300273-16e5d600d6f497dd629cf6fb535467bd" biographyid="vt0015"> Ponnuthurai Nagaratnam Suganthan id="aut0020" author-id="S1568494617300273-bf2cf1500b1013b37aa7e55bfd72fb66" biographyid="vt0020"> Gehan A.J. Amaratunga

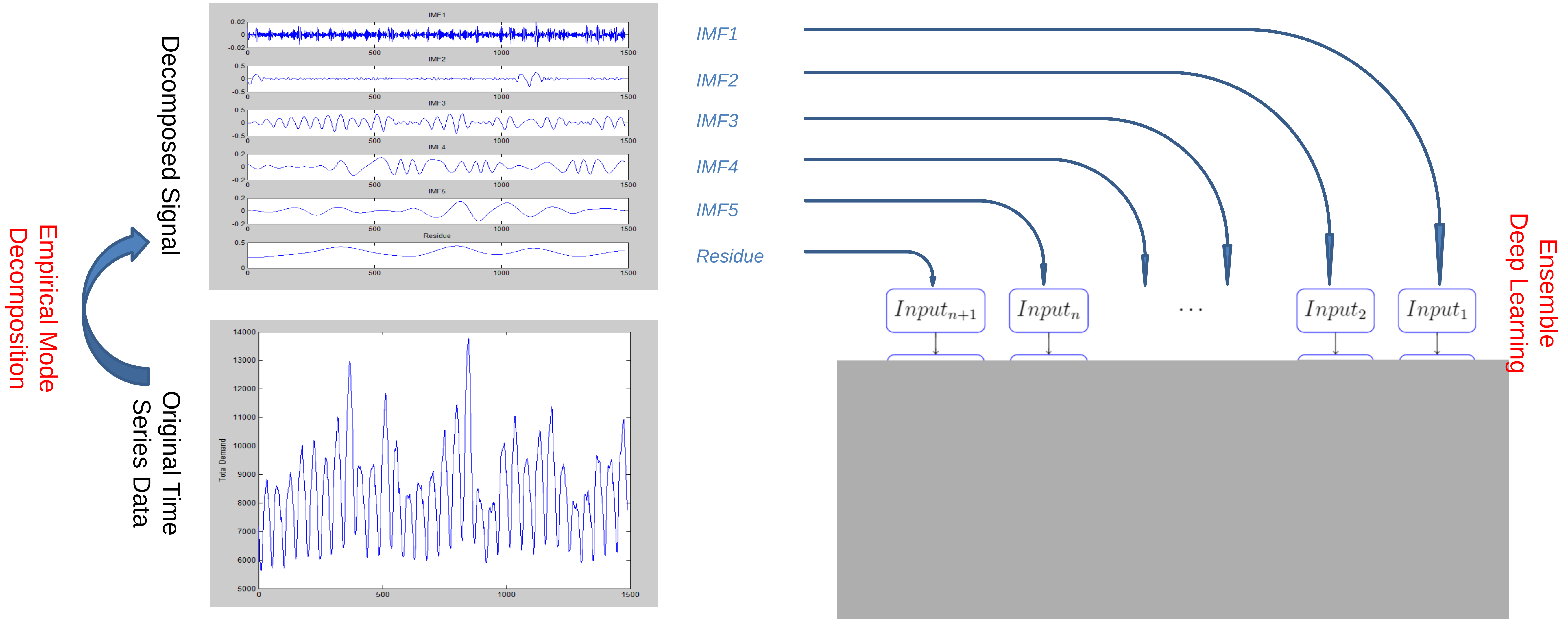
PII:	S1568-4946(17)30027-3
DOI:	http://dx.doi.org/doi:10.1016/j.asoc.2017.01.015
Reference:	ASOC 4009
To appear in:	*Applied Soft Computing*
Received date:	6-7-2016
Revised date:	2-10-2016
Accepted date:	3-1-2017

Please cite this article as: Xueheng Qiu, Ye Ren, Ponnuthurai Nagaratnam Suganthan, Gehan A.J. Amaratunga, Empirical Mode Decomposition based Ensemble Deep Learning for Load Demand Time Series Forecasting, *<![CDATA[Applied Soft Computing Journal]]>* (2017), http://dx.doi.org/10.1016/j.asoc.2017.01.015

Research highlights:

1. An ensemble deep learning method has been proposed for load demand forecasting.
2. The hybrid method composes of Empirical Mode Decomposition and Deep Belief Network.
3. Empirical Mode Decomposition based methods outperform the single structure models.
4. Deep learning shows more advantages when the forecasting horizon increases.

*Graphical abstract (for review)



Empirical Mode Decomposition based Ensemble Deep Learning for Load Demand Time Series Forecasting

Xueheng Qiu[a], Ye Ren[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]*Centre for Advanced Photonics and Electronics, Electrical Engineering Division, Engineering Department, University of Cambridge, Cambridge CB3 0FA, UK*

Abstract

Load demand forecasting is a critical process in the planning of electric utilities. An ensemble method composed of Empirical Mode Decomposition (EMD) algorithm and deep learning approach is presented in this work. For this purpose, the load demand series were first decomposed into several intrinsic mode functions (IMFs). Then a Deep Belief Network (DBN) including two Restricted Boltzmann Machines (RBMs) was used to model each of the extracted IMFs, so that the tendencies of these IMFs can be accurately predicted. Finally, the prediction results of all IMFs can be combined by either unbiased or weighted summation to obtain an aggregated output for load demand. The electricity load demand data sets from Australian Energy Market Operator (AEMO) are used to test the effectiveness of the proposed EMD-based DBN approach. Simulation results demonstrated attractiveness of the proposed method compared with nine forecasting methods.

*Corresponding Author
Email addresses: qiux0004@e.ntu.edu.sg (Xueheng Qiu), re0003ye@e.ntu.edu.sg (Ye Ren), epnsugan@e.ntu.edu.sg (Ponnuthurai Nagaratnam Suganthan), gaja1@hermes.cam.ac.uk (Gehan A. J. Amaratunga)

Preprint submitted to Applied Soft Computing *October 2, 2016*

Keywords: Empirical mode decomposition, Deep learning, Ensemble method, Time series forecasting, Load demand forecasting, Neural networks, Support vector regression, Random forests

Table 1: Nomenclature

RF	**R**andom **F**orest
ANN	**A**rtificial **N**eural **N**etwork
SVM	**S**upport **V**ector **M**achine
SVR	**S**upport **V**ector **R**egression
DBN	**D**eep **B**elief **N**etwork
RBM	**R**estricted **B**oltzmann **M**achine
EMD	**E**mpirical **M**ode **D**ecomposition
IMF	**I**ntrinsic **M**ode **F**unction
ACF	**A**uto**c**orrelation **F**unction
SLFN	**S**ingle-hidden **L**ayer **F**eedforward **N**eural network
MAPE	**M**ean **A**bsolute **P**ercentage **E**rror
RMSE	**R**oot **M**ean **S**quare **E**rror
ARMA	**A**uto **R**egressive **M**oving **A**verage
ARIMA	**A**uto **R**egressive **I**ntegrated **M**oving **A**verage

1. Introduction

Electricity load demand forecasting is one of the most important tasks in the management of modern power systems. Improving the accuracy and efficiency of load demand forecasting can help power companies develop reasonable grid construction planning which will lead to the improvement of economic and social benefits of the systems. Moreover, the forecasting results with high accuracy can also be effective to predict the potential faults in the power systems, and thus provide a reliable safety basis for the grid operation. In another word, the goal of load demand forecasting is to provide reliable power supply while keeping the operational costs as low as possible.

Time series (TS) analysis is a hot research field, which aims to extract meaningful statistics and other characteristics of TS data by analyzing the data itself. Methods of time series analysis can be divided into two categories: univariate and multivariate. For example, Raza has developed an exponentially weighted moving average (EWMA) based shift-detection methods for detecting covariate shifts in non-stationary environments [1]. In the testing stage, Kolmogorov-Smirnov statistical hypothesis test is applied for univariate TS, and the Hotelling T-Squared multivariate statistical hypothesis test is used in the case of multivariate TS. Moreover, many TS datasets have cyclical or seasonal characteristic, which influences TS analysis. Many models have been designed to deal with the cyclic characteristics. For example, Gharehbaghi has developed a pattern recognition framework for detecting dynamic changes on cyclic time series, which combines the discriminant analysis and k-means clustering method [2].

As load demand forecasting belongs to time series forecasting paradigm, there are four types based on the forecasting horizon: long-term (years ahead), medium-term (months to a year ahead), short-term (a day to weeks ahead) and very short-term (minutes to hours ahead) [3]. In this paper, we mainly focus on short term as well as very short term load forecasting. Electricity load demand forecasting with high accuracy is a challenging task. There are many external factors such as climate change and social activities which cause the data to be highly nonlinear and unpredictable [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. Bargur and Mandel have examined the energy consumption and economic growth using trend analysis for Israel [5]. Moreover, the most successful methods are based on Holt-Winters exponential smoothing [6] and Autoregressive Integrated Moving Average (ARIMA) [7], as well as Linear Regression [8].

In the recent years, with the rapid development of computational intelligence, artificial neural network (ANN) [9], fuzzy comprehensive evaluation [10] and support vector machine (SVM) methods [11] have been widely used for short-term load forecasting. Luis Hernández presented a solution for short-term load forecasting in micro-grids. The proposed system includes pattern recognition, a k-means clustering algorithm, and demand forecasting using ANN [12]. Wavelet analysis also can be used for short term load forecasting [13].

ANN has been successfully applied in the fields of classification and regression, but still fell out of fashion as it is often trapped in a local minimum [14]. In 2006, Geoffrey Hinton *et al.* [15] rekindled interest in neural networks by showing substantially better performance by a "deep" neural network. Since then, deep learning has become popular in machine learning field. Takashi Kuremoto proposed a time series forecasting predictor model using Deep Belief Network (DBN) with multiple restricted Boltzmann machines (RBMs) [16]. The CATS benchmark data has been used in the form of 5 blocks with 20 missing and 980 known in each block. The model was then optimized by particle swarm optimization (PSO) algorithm. This work has shown DBN's superiority over conventional multilayer perceptron (MLP) neural network model and statistical model ARIMA. Busseti also conducted simulations to compare deep learning methods with traditional shallow neural networks [17]. The work successfully showed the advantages of deep learning architectures to the problems of electricity load demand forecasting. Moreover, since 2009, Juergen Schmidhuber and his deep learning team have designed effective deep neural networks for image classification [18], handwriting recognition [19] and so on. He has also summarized the deep learning related works in his survey paper [20].

Ensemble learning methods, which obtain better forecasting performance by strategically combining multiple learning algorithms, has been widely applied in various research fields including pattern classification, regression and time series forecasting. Dietterich has concluded three fundamental reasons for the success of ensemble methods: statistical, computational and representational [21]. In addition, Bias-variance decomposition [22] and strength-correlation also explain why ensemble methods have better performance than their non-ensemble counterparts.

Fast algorithms are commonly used with ensembles, such as decision trees. The ensemble of decision trees is called "random forests" introduced by [23]. RF increases the variance of base learning models by combining the concept of bagging and random subspaces [24], thereby improving the performance of this learning model. Manuel Fernández-Delgado et al. have compared 179 learning models from 17 families using 121 classification datasets in their survey paper, among which RF has achieved the best performance [25].

Among the various ensemble methods [26, 27, 28, 29, 30], divide and conquer [31] is a concept which is often applied in time series forecasting. Wavelet transform is a commonly used time series decomposition algorithm. It decomposes the original time series into certain orthonormal sub series by looking at the time frequency domain. [32] use wavelet based nonlinear multistate decomposition model for electricity load forecasting. Adaptive wavelet neural network model is used for forecasting short term electric load with feed forward neurons. Empirical mode decomposition (EMD) [33] is another decomposition method suitable for time series forecasting, which is a part of Hilbert-Huang Transform (HHT). EMD is different from wavelet transform as EMD processes the time series data in time domain. A comparative study on different variations of EMD for wind speed forecasting was reported in [34].

In [35], a survey paper explores the application of machine learning methods to energy-based time series forecasting with two main objectives: (i) providing a compact mathematical formulation of the mainly used techniques; (ii) reviewing the latest works of time series forecasting related to electricity price and demand markets. A wide variety of data mining approaches are discussed in this work, including linear models, non-linear machine learning methods, and ensemble models. Several common points are concluded, such as the horizon of prediction normally equals to one day, and the accuracy measures mainly used are MAPE and RMSE, which are consistent with the experiments in our paper. Moreover, the survey work states that: "the current trend in electricity forecasting points to the development of ensembles, thus highlighting single strengths of every method". Therefore, interested readers should view this survey paper as a good guidance for time series forecasting.

In [27], we have proposed an ensemble deep learning algorithm for regression and time series forecasting, which composed of DBNs trained using different number of BP epochs and an SVR applied to analyze the relationship between these outputs and target values. It is worth noting that the input to each DBN is the original time series data. To further improve the ensemble learning architecture, in this paper, we adopt the concept of "divide and conquer", and construct a novel electricity load demand forecasting method based on EMD and deep learning algorithms. The advantages of the proposed method are demonstrated on real world datasets compared with nine benchmark learning algorithms: Persistence, SVR, ANN, DBN, RF, EMD based SVR, EMD based ANN, EMD based RF, as well as the ensemble DBN proposed in the previous work.

The remaining of this paper is organized as follows: Section 2 explains the theoretical background on forecasting methods. Section 3 presents the proposed EMD based ensemble deep learning method. Section 4 shows the procedures for experiment setup, followed by the discussion about experiment results in Section 5. In Section 6, two comparative experiments are implemented to evaluate the performance of the proposed method. Finally in Section 7, the conclusions and future works are stated.

2. Theoretical Background on Forecasting Models

2.1. Artificial Neural Network

ANN is a learning model inspired by human brain, especially the central nervous system [36]. The simplest model of ANN is called single-hidden layer feedforward neural network (SLFN). There are three fundamental layers in an SLFN: an input layer with the same number of neurons as the dimension of input features; a hidden layer comprised of neurons with nonlinear activation function; and an output layer which aggregates the outputs from the hidden layer neurons. The output from SLFN is:

$$y = g(\sum_{j=1}^{h} w_{jo} v_j + b_j) \tag{1}$$

$$v_j = f(\sum_{i=1}^{n} w_{ij}x_i + b_i) \tag{2}$$

where x_i is the input to the neuron; $f()$ and $g()$ are nonlinear activation functions; v_j is the output of hidden layer neuron j; 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.

To train an SLFN, random values are assigned to the weights, then the weights are tuned by certain method such as back-propagation (BP) [37] or using a closed form solution [38, 39].

2.2. *Support Vector Regression*

The Support Vector Machine (SVM) is a statistical learning theory based machine learning method which proposed by [11], using structural risk minimization as its fundamental concept. The Support Vector Regression (SVR) is a regression method which shared the same theoretical background as SVM. It has been widely applied in time series prediction such as load demand forecasting.

Suppose a time series data set is given as

$$D = \{(X_i, y_i)\}, 1 \leqslant i \leqslant N \tag{3}$$

where X_i is the input vector at time i with m elements and y_i is the corresponding output data. The regression function can be defined as

$$f(X_i) = W^T\phi(X_i) + b \tag{4}$$

where W is the weight vector, b is the bias, and $\phi(X)$ maps the input vector X to a higher dimensional feature space. W and b can be obtained by solving the following optimization problem:

$$\text{Min}\ \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}(\varepsilon_i + \varepsilon_i^*) \tag{5}$$

Subject to:

$$\begin{aligned} & y_i - W^T(\phi(x)) - b \leq \xi + \varepsilon_i \\ & W^T(\phi(x)) + b - y_i \leq \xi + \varepsilon_i^* \\ & \varepsilon_i, \varepsilon_i^* \geq 0 \end{aligned} \tag{6}$$

where C is a predefined positive trade-off parameter between model simplicity and generalization ability, ξ is a free parameter that serves as a threshold, ε_i and ε_i^* are the slack variables measuring the cost of the errors.

For nonlinear input data set, kernel functions can be used to map from original space onto a higher dimensional feature space in which a linear regression model can be built. The most frequently used kernel function is the Gaussian radial function (RBF) with a width of σ

$$K(X_i, X_j) = exp(-\|X_i - X_j\|^2/(2\sigma^2)) \tag{7}$$

2.3. Random Forests

Random forests, or random decision forests [40, 41], proposed by [23], is an ensemble learning method for both of classification and regression problems. Random forests combine bagging and random subspace method (RSM) by conducting random feature subspace at each node of the classification and regression tree (CART) [24]. Bagging (bootstrap aggregating), developed by [42], is a widely used ensemble method. In bagging ensemble method,one trains each weak learning machine on bootstrap samples of the original training samples, then aggregating the outputs. RSM is a combining method which trains the learning machines on randomly chosen subspaces of the original input space, and combines the outputs by a majority vote or median [41]. More specifically, at each node of the decision tree in random forest, m features from totally n input features are randomly selected. Then according to an impurity criterion, one of these features is used to perform a partition along the feature axis [24]. The algorithm of RF is presented in Table 2 [43, 44].

Table 2: Random Forests Algorithm

Random Forests Algorithm:
Given:
X is the training dataset with dimension $N \times n$, where N is the number of observations, n is the number of input features.
Y is the target values of the training dataset with dimension $N \times 1$.
L is the number of trees in RF.
T_i refers to each decision tree in RF, where $i = 1...L$.
m is the number of features randomly selected in each node of decision tree.
1). In each decision tree T_i in RF, generate the training set by sampling N times from all observations with replacement.
2). In each node of one decision tree, m randomly selected features are used to calculate the best split criterion for T_i.
3). Repeat step 2 until the decision tree T_i is fully grown.
4). Aggregate the outputs given by all the decision trees to obtain final result. For classification, the output value is determined by majority vote. For regression, the mean or median of all the outputs is treated as the predicted value.

2.4. *Deep Belief Network (DBN)*

Deep learning is a branch of machine learning algorithms that attempts to model high-level abstractions in data by using model architectures, with complex structures with multiple non-linear transformations [45]. Deep learning algorithms are fundamentally based on distributed representations, which means that observed data can be represented by interactions of many different factors on different levels. The main promise of deep learning is replacing handcrafted features with efficient algorithms for unsupervised feature extraction [46]. In another word, deep learning attempts to abstract important features in input data set by deep architecture in an unsupervised way.

The DBN proposed by [15] provides a new way to train deep generative models, which is called layer-wise greedy pre-training algorithm. Figure 1 shows the flowchart of a DBN. There is no inter-connection between units in each layer. An restricted Boltzmann machine (RBM) is a neural network which can learn the probability distribution over the input dataset. The DBN pre-training procedure treats each consecutive pair of layers in the MLP as a restricted Boltzmann machine (RBM) [47] whose joint probability is defined as

$$P_{h,v}(h,v) = \frac{1}{Z_{h,v}} \cdot e^{(v^T W h + v^T b + a^T h)} \tag{8}$$

for the Bernoulli-Bernoulli RBM applied to binary v with a second bias vector b and normalization term $Z_{h,v}$, and

$$P_{h,v}(h,v) = \frac{1}{Z_{h,v}} \cdot e^{(v^T W h + (v-b)^T (v-b) + a^T h)} \tag{9}$$

for the Gaussion-Bernoulli RBM applied to continuous variable v [48]. In both cases the conditional probability $P_{h|v}(h|v)$ has the same form as that in an MLP layer.

The RBM parameters can be efficiently trained in an unsupervised fashion by maximizing the likelihood $\mathcal{L} = \prod_t \sum_h P_{h,v}(h, v(t))$ over training samples $v(t)$ with the approximate contrastive divergence algorithm [49].

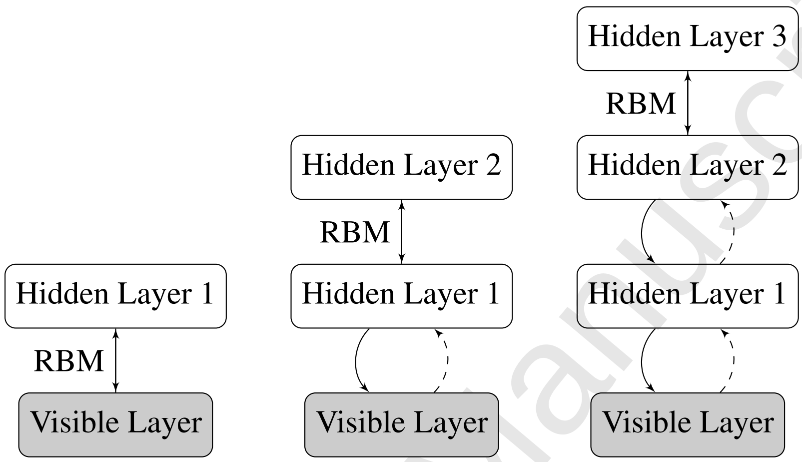


Figure 1: Flowchart of a Deep Belief Network (DBN)

To train multiple layers, one trains the first layer, freezes it, and uses the conditional expectation of the output as the input to the next layer and continues training next layers. Hinton and many others have found that initializing MLPs with pretrained parameters never hurts and often helps [15, 50].

2.5. *Empirical Mode Decomposition*

EMD [33], also known as 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. The system load behavior is influenced by a number of factors, which can be classified as: economic factors, time, day, season, weather and random effects. Thus, EMD algorithm can be very effective for load demand forecasting.

An IMF is a function that has only one extreme between zero crossings, along with a mean value of zero. The procedures of algorithm inside EMD are shown as follows:

1. With a given time series signal $x(t)$, create its upper and lower envelopes by a cubic-spline interpolation of local maxima and minima.
2. Calculate the mean of the upper and lower envelopes m_1.
3. Subtract the mean from the original time series to obtain the first component $h(t) = x(t) - m(t)$.
4. Repeat steps 1 to 3 by considering $h(t)$ as new $x(t)$ until one of the following stopping criteria is satisfied: i) $m(t)$ approaches zero, ii) the numbers of zero-crossings and extrema of $h(t)$ differs at most by one, or iii) the predefined maximum iteration is reached.
5. Treat $h(t)$ as an IMF and compute residue signal: $r(t) = x(t) - h(t)$.
6. Use the residual signal $r(t)$ as new $x(t)$ to find next IMF. Repeat steps 1 to 5 until all IMFs are obtained.

Finally the original TS signal is decomposed as:

$$x(t) = \sum_{i=1}^{n} (c_i) + r_n \tag{10}$$

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

End data points extending problem is an important issue which should be considered in EMD algorithm. In usual case, the end points are not the local extrema, so it is necessary to extend two maximum extrema and two minimum extrema to get the envelope of extrema. In [33], the characteristic waves were used to get the maximum extrema and the minimum extrema. However, different characteristic waves will cause different results, and it is difficult to choose the proper waves for every iteration. In this work, the Matlab package for EMD, which is implemented by G. Rilling, was used to decompose the TS signal. According to [51], good results can be obtained by just mirroring the extrema close to the edges (or mirror extending method). Several publications in the literature present some improved methods for mitigating of end effect in EMD. For example, in [52], end mirror extending is used in high frequency, while least square polynomial extending is used in low frequency. However, no complete solution is in sight currently, which means that there is room for improving the solution of the endpoint effect of EMD method.

Figure 2 shows an example of the decomposed load demand TS signal with a time window of one month.

3. Proposed EMD based Deep Learning Method

A divide and conquer algorithm works by recursively breaking down a problem into two or more sub-problems of the same (or related) type, until these become simple enough to be solved directly. The solutions to the sub-problems are then combined to give a solution to the original problem.

In the proposed method, the load demand data is decomposed into several IMFs and one residue by EMD method. A DBN composed of two RBMs and one

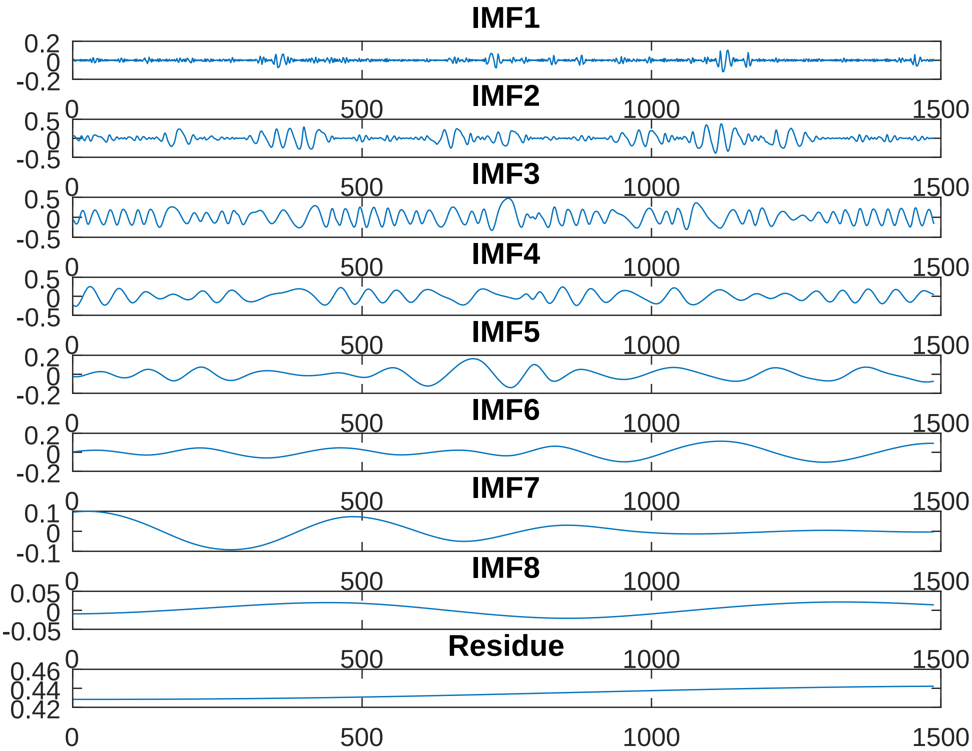


Figure 2: Example of the obtained IMF components after EMD with a time window of one month.

ANN is applied to each IMF including the residue. As the completeness of the forecasting for all sub series, the prediction results can be aggregated by single learning machine or simply summed to obtain the final prediction. The procedure of this proposed method is shown as follows:

1. The time series data is decomposed by EMD into several IMFs and one residue.
2. For each IMF and residue, we construct one training matrix as the input for one DBN.

3. Train DBN to obtain the predicted results for each of the extracted IMF and residue.
4. Combine all the prediction results by summation or with a linear neural network to formulate an ensemble output for TS.

Figure 3 shows the overall schematic of this ensemble method.

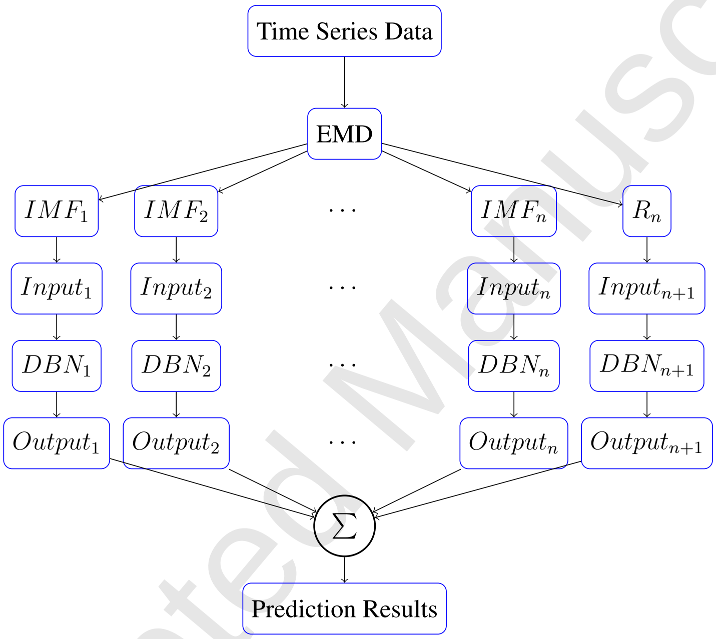


Figure 3: Schematic Diagram of the Proposed EMD based Deep Learning Approach

4. Methodology and Experiments

In this paper, the performance of proposed ensemble method is evaluated by comparing with nine benchmark methods: Persistence, SVR, ANN, DBN, Ensemble DBN (EDBN), EMD based SVR model(EMD-SVR), EMD based ANN model (EMD-SLFN) and EMD based RF (EMD-RF).

4.1. Datasets

The electricity load demand data sets from Australian Energy Market Operator (AEMO) were used for the comparison [53]. Especially, the data sets of year 2013 from New South Wales (NSW), Tasmania (TAS), Queensland (QLD), South Australia (SA) and Victoria (VIC) were chosen to train and test the proposed method. For each area, four months were chosen to reflect the factors of different seasons: January, April, July and October. During the simulation, the first three weeks was used to train the model, and the remaining one week was used for testing. Thus, there are totally 1008 examples for training and 336 examples for testing. Moreover, six fold cross-validation was employed during training to improve the generalization.

4.2. Characteristics of Data

The load demand data is sampled every half an hour, which means there are 48 data points for one day. Due to the influence from climate and social activities, the electricity load data shows three main nest cycles: daily, weekly and yearly. To identify cycles and patterns in load demand time series data, autocorrelation function (ACF) can be applied as a guidance for informative feature subset selection [3]. 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 $t - k$.

Three strongest dependent lag variables can be identified from the ACF of electricity load demand TS: the value at half-an-hour before (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, for one day head load demand forecasting in this work, the input feature set is composed by 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}), which include all the most informative lag variables.

4.3. Methodology

For the time series load demand datasets, all the training and testing values are linearly scaled to $[0, 1]$. The scaling formula is:

$$\bar{y}_i = \frac{y_{max} - y_i}{y_{max} - y_{min}} \tag{12}$$

To implement the simulation, LIBSVM toolbox was used for the SVR model [54], while deep learning toolbox was used for neural networks, including ANN, DBN, EDBN [27], EMD-ANN and the proposed method [55]. RF and EMD-RF are 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.

For SVR and EMD based SVR, we choose the RBF kernel function with parameters chosen by a grid search. As suggested by the authors of LIBSVM toolbox, exponentially growing sequences of C and σ is used for parameter selection, where the range of C is [$2^{-4}, 2^4$], and the range of σ is [10^{-3},10^{-1}]. For ANN and EMD-ANN, the size of neural networks is determined by the size of input vector. For DBN and the proposed method, two RBMs are stacked for pre-training with the size of [100 100]. The number of iterations for back propagation is set as 500. For RF and EMD based RF, the number of decision trees is set as 500.

4.4. Performance Estimation

In this paper, two error measures are used to examine the accuracy of a prediction model: Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). They are defined as

$$
\begin{aligned}
RMSE &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i' - y_i)^2} \\
MAPE &= \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i' - y_i}{y_i}\right|
\end{aligned}
\tag{13}
$$

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

5. Results and Comparison

In this study, two forecasting horizons are adopted for comparison: half an hour (very short term) and one day ahead (short term).

5.1. Performance Comparison for Half-an-hour ahead Load Forecasting

The simplest forecasting method is persistence method, which assumes that the conditions at the future time of forecast are the same as the current values. The persistence method works well for very short term load demand forecasting since the temperature and human factors change little during a short time period. Therefore, persistence method can be treated as a baseline for evaluating the effectiveness of machine learning models. The one step ahead (half an hour) forecasting results of persistence method are shown in Table 3. We can see that all of the machine-learning algorithms outperform the persistence method for half an hour ahead forecasting.

The original load demand time series data was modeled by SVR, SLFN and RF without decomposition to reveal the advantages of EMD based hybrid approach. Comparing the forecasting results listed in Table 3, we can conclude that the EMD based hybrid approach generally outperforms the single structure machine learning algorithms most of the time. Moreover, EMD-SVR, EMD-SLFN and EMD-RF model has comparable performance with each other. In addition, the proposed EMD-DBN model has the best performance for half-an-hour ahead forecasting in most cases.

The comparison results of Nemenyi test among all the learning methods based on RMSE and MAPE are shown in Figure 4. The Nemenyi test is a post-hoc test which is used when all classifiers are compared to each other [56]. As shown, the methods with better ranks are at the top whereas the methods with worse ranks are at the bottom. The methods within a vertical line whose length is less than or equal to a critical distance have statistically the same performance. The title of the graphs shows Friedman p-value. If it is smaller than 0.05, there exists significant difference among these models. The critical difference is calculated by:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{14}$$

where k is the number of algorithms, N is the number of data sets, and q_α is the critical value based on the studentized range statistic divided by $\sqrt{2}$ [56]. Therefore, we can conclude from the results of statistical testing that our proposed method has the best rank and significantly outperforms the non-ensemble methods with a 95% confidence. It is also worth noting that the proposed EMD-DBN method has better rank compared with EDBN, which shows the advantages of divide and conquer concept.

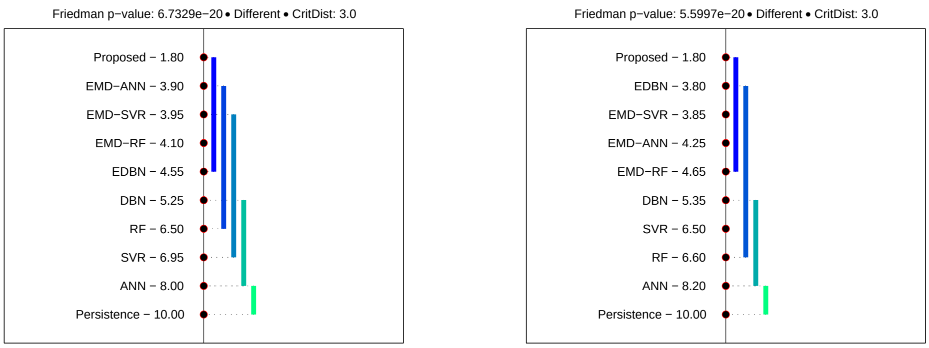


Figure 4: Nemenyi testing results for half-an-hour ahead load forecasting based on RMSE (left) and MAPE (right). The critical distance is 3.0.

5.2. Performance Comparison for One-day ahead Load Forecasting

The prediction results for one day ahead load forecasting are shown in Table 4. Similar to very short term load forecasting, the performance comparison includes the persistence method. In this case, the time horizon is 24 hour, which means that we assume the predicted value is the same as the value 24 hours ago. Due to the daily seasonality of the load demand data, the accuracy of persistence method falls in an acceptable range. Therefore, the effectiveness of the involved machine learning algorithms can be verified by outperforming the persistence methods. Same as the previous experiment, the Nemenyi test is also used to compare the one-day ahead forecasting performances. The results based on RMSE and MAPE are shown in Figure 5.

By analyzing the forecasting outputs of SVR and ANN listed in Table 4, we can find that these two methods have comparable performances. This phenomenon may be due to the reason that both models have similar network structure with one hidden layer [59]. It is also worth noting that the DBN model out-

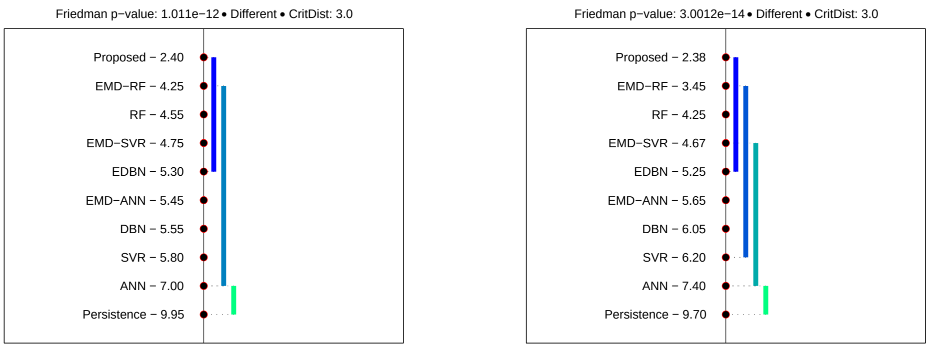


Figure 5: Nemenyi testing results for one-day ahead load forecasting based on RMSE (left) and MAPE (right). The critical distance is 3.0.

performs both SVR and ANN for one day ahead forecasting and half an hour ahead load forecasting. Moreover, the EMD based hybird methods outperform the single structure models, which can confirm the advantages of EMD based ensemble algorithms. Most outstandingly, according to the Nemenyi testing rank, we can conclude that the proposed EMD-based DBN approach has successfully outperformed all the benchmark methods in both experiments on both forecasting horizons.

6. Comparative Experiments

In this section, three comparative experiments were implemented to evaluate the performance of our proposed method. The comparison conditions such as dataset partitioning and cross-validation were kept the same for the reported methods [60, 61, 62] and the proposed method.

6.1. Comparative Experiment with SRSVRCABC Model

The first experiment uses historical monthly electric load demand data of Northeast China to compare with two benchmark methods: seasonal recurrent SVR with chaotic artificial bee colony (SRSVRCABC) model in [61] and TF-ε-SVR-SA model in [60]. According to Hong's paper, there are totally 64 monthly electric load data points from January 2004 to April 2009, which are divided into three parts: the training set (32 data points, December 2004 to July 2007), the validation data set (14 data points, August 2007 to September 2008), and the testing data sets (7 data points, from October 2008 to April 2009). Moreover, based on the same comparison conditions, 25 data points are fed in as input matrix to predict the following monthly load data.

Table 5 shows the actual values and the forecast values obtained using all benchmark methods. Obviously, the proposed method has the smallest MAPE values compared with ARIMA, TF-ε-SVR-SA and SRSVRCABC models.

6.2. Comparative Experiment with AFCM

In the second comparative experiment, the electricity load demand data of May 2007 from New South Wales, Australia is used for the simulations. According to [62], this experiment is divided into two parts: one part with small sample size, and another part with large sample size. In the first part, the data set contains the load demand data from 00:00 on May 2 to 23:30 on May 8 with the same interval of 30 min. The data set is divided into two parts: one is the training set which contains the historical data of first six days; another one is the testing set which contains the remaining one day's data. In the second part with large sample size, 1104 data points from May 2 to May 24 are used to train the model to predict the load demand in the following one week from May 18 to May 24. The adaptive fuzzy combination model (AFCM) from [62] along with SVR are implemented to compare with the proposed model.

From the forecasting results listed in Table 6, some observations can be made from comparison. First of all, all the learning methods are effective for short time load demand forecasting since all of them can give reasonable results. Second, by comparing the differences between small and large sample size parts, the proposed EMD-DBN method can reduce the influence caused by redundant information in the large size data set to give better performance. Finally, it is clear that the EMD based deep learning method has outperformed the benchmark methods in both experiments.

6.3. Comparative Experiment with PSF-NN

For the third comparative experiment, according to [63], we use electricity load demand data for the state of NSW in Australia for three years: 2009, 2010 and 2011. The data from the first two years is used to train the prediction model, while the remaining data for 2011 is used for testing. The forecasting horizon is still one day. There are four benchmark methods, the pattern sequence-based forecasting (PSF) method and three combined PSF-NN models with three different feature sets. Table 7 shows the comparative results. The proposed EMD based deep learning approach outperforms all the benchmark methods on both error measures.

7. Conclusion

In this paper, we proposed an ensemble deep learning method based on EMD and DBN. The proposed method has been evaluated with three electricity load demand datasets from AEMO. Nine benchmark methods have been compared to verify the effectiveness of the proposed method: Persistence, SVR, ANN, DBN, RF, EDBN, EMD-SVR, EMD-SLFN and EMD-RF. Two error measures (RMSE and MAPE) were used to evaluate the performance of these prediction models. Moreover, two comparative experiments are also implemented to verify the effectiveness of the proposed method. According to the prediction results, several observations can be concluded:

1. EMD based hybrid methods normally outperform the corresponding single structure models for load demand time series forecasting.
2. Deep learning algorithms show their advantages in dealing with nonlinear features when the forecasting horizon increases.
3. Random Forests, as a decision tree based method, is effective for load demand forecasting with the advantage of fast training.
4. The proposed EMD based ensemble deep learning approach has the best performance according to the statistical testing.

For future work, additional nonlinear features such as climate and human activities need to be considered in constructing a more complex model. The potential learning ability of deep learning methods will be well suited to such complex problems. Moreover, since the ensemble deep learning model is more time consuming when compared with the single structure model, optimization techniques can be designed to simplify the structure and increase the efficiency of deep learning model.

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. Raza, G. Prasad, Y. Li, EWMA model based shift-detection methods for detecting covariate shifts in non-stationary environments, Pattern Recognition 48 (3) (2015) 659–669.

[2] A. Gharehbaghi, P. Ask, A. Babic, A pattern recognition framework for detecting dynamic changes on cyclic time series, Pattern Recognition 48 (3) (2015) 696–708.

[3] I. Koprinska, M. Rana, V. G. Agelidis, Correlation and instance based feature selection for electricity load forecasting, Knowledge-Based Systems 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, International Journal of Energy Research 38 (3) (2014) 350–362.

[5] J. Bargur, A. Mandel, M. ha-Tekhniyon le-mehkar u-fituah (Haifa Israel), I. M. ha-energyah veha tashtit, Energy Consumption and Economic Growth in Israel: Trend Analysis (1960-1979), Ministry of Energy and Infrastructure, 1981.

[6] C. C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, International Journal of Forecasting 20 (2004) 5–10.

[7] G. E. P. Box, G. M. Jenkins, Time series analysis: forecasting and control, Holden-Day series in time series analysis and digital processing, Holden-Day, 1976.

[8] A. D. Papalexopoulos, T. C. Hesterberg, A regression-based approach to short-term system load forecasting, IEEE Transactions on Power Systems 5 (1990) 1535–1547.

[9] G. A. Darbellay, M. Slama, Forecasting the short-term demand for electricity: Do neural networks stand a better chance?, International Journal of Forecasting 16 (2000) 71–83.

[10] L. C. Ying, M. C. Pan, Using adaptive network based fuzzy inference system to forecast regional electricity loads, Energy Conversion and Management 49 (2008) 205–211.

[11] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (3) (1995) 273–297.

[12] L. Hernández, C. Baladrón, J. M. Aguiar, B. Carro, A. Sánchez-Esguevillas, J. Lloret, Artificial neural networks for short-term load forecasting in microgrids environment, Energy 75 (2014) 252–264.

[13] M. G. M. Ghayekhloo, M. B. Menhaj, A hybrid short-term load forecasting with a new data preprocessing framework, Electric Power Systems Research 119 (2015) 138–148.

[14] W. C. Hong, Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model, Energy Conversation and Management 50 (2009) 105–117.

[15] G. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural computation 18 (7) (2006) 1527–1554.

[16] T. Kuremoto, S. Kimura, K. Kobayashi, M. Obayashi, Time series forecasting using restricted boltzmann machine, in: Emerging Intelligent Computing Technology and Applications, Springer, 2012, pp. 17–22.

[17] E. Busseti, I. Osband, S. Wong, Deep learning for time series modeling, Tech. Rep. CS229, Stanford University (2012).

[18] D. Ciresan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR2012), 2012, pp. 3642–3649.

[19] D. C. Ciresan, U. Meier, L. M. Gambardella, J. Schmidhuber, Deep, big, simple neural nets for handwritten digit recognition, Neural Computation 22 (12) (2010) 3207–3220.

[20] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks 61 (2015) 85–117.

[21] T. G. Dietterich, Ensemble methods in machine learning, in: Multiple classifier systems, Springer, 2000, pp. 1–15.

[22] S. Geman, E. Bienenstock, R. Doursat, Neural networks and the bias/variance dilemma, Neural computation 4 (1992) 1–58.

[23] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.

[24] L. Breiman, J. H. Freidman, R. A. Olshen, C. J. Stone, Classification and Regression Trees, The Wadsworth and Brooks-Cole statistics-probability series, Taylor & Francis, 1984.

[25] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, Do we need hundreds of classifiers to solve real world classification problems?, Journal of Machine Learning Research 15 (2014) 3133–3181.

[26] L.-Y. Wei, Comprehensive learning particle swarm optimization based memetic algorithm for model selection in short-term load forecasting using support vector regression, Applied Soft Computing 25 (2014) 15–25.

[27] X. Qiu, L. Zhang, Y. Ren, P. N. Suganthan, G. Amaratunga, Ensemble deep learning for regression and time series forecasting, in: Proc. IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL), 2014, pp. 1–6.

[28] Z. Hu, Y. Bao, T. Xiong, A hybrid anfis model based on empirical mode decomposition for stock time series forecasting, Applied Soft Computing 42 (2016) 368–376.

[29] Y. Ren, X. Qiu, P. N. Suganthan, EMD based AdaBoost-BPNN method for wind speed forecasting, in: Proc. IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL2014), Orlando, US, 2014, pp. 1–6.

[30] Y. Ren, P. N. Suganthan, N. Srikanth, A novel empirical mode decomposition with support vector regression for wind speed forecasting, IEEE Transactions on Neural Network and Learning Systems PP (99) (2015) 1.

[31] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, Introduction to Algorithms, MIT Press, 2000.

[32] D. Benaouda, F. Murtagh, J.-L. Starck, O. Renaud, Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting, Neurocomputing 70 (2006) 139–154.

[33] 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.

[34] Y. Ren, P. N. Suganthan, N. Srikanth, A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods, IEEE Transactions on Sustainable Energy 6 (1) (2015) 236–244.

[35] F. Martínez-Álvarez, A. Troncoso, G. Asencio-Cortés, J. C. Riquelme, A survey on data mining techniques applied to electricity-related time series forecasting, Energies 8 (11) (2015) 13162–13193.

[36] S. Haykin, Neural Networks: A Comprehensive Foundation, International edition, Prentice Hall, 1999.

[37] C.-N. Ko, C.-M. Lee, Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended kalman filter, Energy 49 (2013) 413–422.

[38] L. Zhang, P. N. Suganthan, A survey of randomized algorithms for training neural networks, Information Sciences 000 (2016) 1–10.

[39] Y. Ren, P. N. Suganthan, N. Srikanth, G. Amaratunga, Random vector functional link network for short-term electricity load demand forecasting, Information Sciences 000 (2016) 1–16.

[40] T. K. Ho, Random decision forests, in: Proc. the Third International Conference on Document Analysis and Recognition, Montreal, Que, 1995, pp. 278–282.

[41] T. K. Ho, The random subspace method for constructing decision forests, Pattern Analysis and Machine Intelligence 20 (8) (1998) 832–844.

[42] L. Breiman, Bagging predictors, Machine Learning 24 (2) (1996) 123–140.

[43] L. Zhang, P. N. Suganthan, Random forests with ensemble of feature spaces, Pattern Recognition 47 (2014) 3429–3437.

[44] L. Zhang, P. N. Suganthan, Oblique decision tree ensemble via multisurface proximal support vector machine, IEEE Transactions on Cybernetics 45 (10) (2015) 2165–2176.

[45] Y. Bengio, Learning deep architectures for ai, Foundations and Trends in Machine Learning 2 (2009) 1–127.

[46] H. A. Song, S.-Y. Lee, Hierarchical representation using nmf, Neural Information Processing 8226 (2013) 466–473.

[47] G. E. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.

[48] T. Yamashita, M. Tanaka, E. Yoshida, Y. Yamauchi, H. Fujiyoshi, To be Bernoulli or to be Gaussian, for a restricted Boltzmann machine, in: 22nd International Conference on Pattern Recognition, 2014, pp. 1520–1525.

[49] G. E. Hinton, Training products of experts by minimizing contrastive divergence, Neural Computation 14 (8) (2002) 1771–1800.

[50] G. E. Hinton, Neural Networks: Tricks of the Trade: Second Edition, Springer Berlin Heidelberg, 2012, Ch. A Practical Guide to Training Restricted Boltzmann Machines, pp. 599–619.

[51] G. Rilling, P. Flandrin, P. Gonçalves, On empirical mode decomposition and its algorithms, in: Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03, Grado (Italy), 2003.

[52] Q. Zhang, H. Zhu, L. Shen, A new method for mitigation of end effect in empirical mode decomposition, in: 2nd International Asia Conference on Informatics in Control, Automation and Robotics, 2010.

[53] AEMO, Australian energy market operator, `http://www.aemo.com.au/` (2013).

[54] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology (TIST) 2 (3) (2011) 27.

[55] R. B. Palm, Prediction as a candidate for learning deep hierarchical models of data, Master's thesis, Technical University of Denmark (2012).

[56] J. Demšar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research 7 (2006) 1–30.

[57] 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.

[58] H. Liu, C. Chen, H. Tian, Y. Li, A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks, Renewable Energy 48 (2012) 545–556.

[59] E. Romero, D. Toppo, Comparing support vector machines and feedforward neural networks with similar hidden-layer weights, IEEE Transactions on Neural Networks 18 (3) (2007) 959–963.

[60] J. Wang, W. Zhu, W. Zhang, D. Sun, A trend fixed on firstly and seasonal adjustment model combined with the ε-SVR for short-term forecasting of electricity demand, Energy Policy 37 (2009) 4901–4909.

[61] W.-C. Hong, Electric load forecasting by seasonal recurrent svr (support vector regression) with chaotic artificial bee colony algorithm, Energy 36 (2011) 5568–5578.

[62] J. Che, J. Wang, G. Wang, An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting, Energy 37 (2012) 657–664.

[63] I. Koprinska, M. Rana, A. Troncoso, F. Martínez-Álvarez, Combining pattern sequence similarity with neural networks for forecasting electricity de-

mand time series, in: International Joint Conference on Neural Networks (IJCNN), 2013.

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.

Ye Ren received the B.E. and the M.E. degrees in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2008 and 2011, respectively. He is currently pursuing the Ph.D. degree at Nanyang Technological University, Singapore. His research interests include classification, regression, and wind speed 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 PhD 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 PhD student) won the IEEE CIS Outstanding PhD 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.

Table 3: Prediction results for half-an-hour ahead load forecasting

Dataset	Month	Metrics	Prediction model									
			Persistence	SVR [11]	ANN [36]	DBN [15]	RF [23]	EDBN [27]	EMD-SVR [57]	EMD-ANN [58]	EMD-RF	Proposed
NSW	Jan	RMSE	164.02	94.24	96.66	79.16	93.36	75.42	78.56	82.11	76.10	**49.86**
		MAPE	1.64%	0.93%	0.98%	0.78%	0.88%	0.70%	0.78%	0.88%	0.76%	**0.53%**
	Apr	RMSE	248.14	162.57	140.74	70.36	142.85	134.47	114.09	87.76	120.16	**69.55**
		MAPE	2.43%	1.88%	1.26%	**0.64%**	1.18%	1.14%	1.07%	0.82%	1.10%	0.65%
	Jul	RMSE	235.66	117.87	165.42	105.63	114.83	78.09	**74.29**	81.22	120.77	75.09
		MAPE	2.31%	1.20%	1.68%	1.09%	1.20%	**0.67%**	**0.67%**	0.81%	1.22%	0.70%
	Oct	RMSE	159.98	58.26	76.64	62.58	69.06	64.36	54.58	66.00	76.58	**51.68**
		MAPE	1.65%	0.64%	0.82%	0.70%	0.75%	0.66%	0.60%	0.74%	0.82%	**0.55%**
TAS	Jan	RMSE	17.80	13.87	13.84	12.90	11.80	13.24	12.54	**11.39**	13.52	11.59
		MAPE	1.27%	1.09%	1.09%	1.01%	1.03%	1.06%	0.97%	0.78%	1.09%	**0.74%**
	Apr	RMSE	37.42	25.26	27.03	19.53	24.82	21.35	21.76	18.03	25.19	**16.23**
		MAPE	2.32%	1.49%	1.63%	1.25%	1.26%	1.21%	1.36%	1.16%	1.45%	**1.07%**
	Jul	RMSE	43.84	34.03	33.97	30.43	33.59	**22.62**	30.90	29.14	32.34	24.44
		MAPE	2.73%	2.10%	2.03%	1.76%	2.07%	**1.36%**	1.91%	1.98%	2.17%	1.54%
	Oct	RMSE	22.80	15.89	18.49	16.80	16.94	20.41	14.94	9.37	13.69	**8.81**
		MAPE	1.63%	1.09%	1.36%	1.19%	1.19%	1.34%	1.06%	0.70%	0.97%	**0.66%**
QLD	Jan	RMSE	109.39	51.31	62.97	44.95	40.30	51.03	42.12	33.88	33.34	**25.39**
		MAPE	1.50%	0.70%	0.85%	0.62%	0.54%	0.63%	0.57%	0.48%	0.44%	**0.34%**
	Apr	RMSE	137.11	71.30	65.84	48.48	57.20	60.27	51.30	56.02	54.78	**48.34**
		MAPE	1.91%	0.94%	0.93%	0.66%	0.86%	0.75%	**0.61%**	0.73%	0.81%	0.67%
	Jul	RMSE	127.23	46.07	51.79	38.45	45.52	42.00	35.53	41.48	45.39	**30.61**
		MAPE	1.92%	0.72%	0.82%	0.54%	0.85%	0.60%	0.53%	0.62%	0.69%	**0.44%**
	Oct	RMSE	110.19	57.90	63.17	61.03	61.37	55.49	54.89	46.78	48.41	**40.46**
		MAPE	1.61%	0.80%	0.92%	0.86%	0.77%	0.71%	0.75%	0.69%	0.67%	**0.56%**
VIC	Jan	RMSE	174.95	117.79	114.73	106.25	120.34	**78.58**	82.96	117.45	115.66	98.75
		MAPE	2.52%	1.54%	1.55%	1.32%	1.58%	**1.05%**	1.09%	1.56%	1.59%	1.35%
	Apr	RMSE	162.18	148.89	149.20	104.48	102.34	75.59	96.24	77.02	68.30	**64.11**
		MAPE	2.15%	1.97%	1.99%	1.49%	1.53%	0.98%	1.33%	0.99%	0.92%	**0.87%**
	Jul	RMSE	171.99	69.41	119.43	86.63	62.57	66.36	119.27	114.34	61.84	**58.70**
		MAPE	2.44%	1.01%	1.74%	1.26%	1.00%	0.91%	1.73%	1.67%	**0.88%**	**0.88%**
	Oct	RMSE	139.47	62.88	96.63	91.20	87.11	68.50	90.49	91.38	**55.19**	57.95
		MAPE	1.97%	0.92%	1.42%	1.35%	1.20%	0.89%	1.23%	1.23%	0.85%	**0.84%**
SA	Jan	RMSE	72.11	50.37	55.65	59.17	53.92	**39.87**	58.29	46.34	42.52	51.91
		MAPE	3.04%	1.92%	2.33%	2.54%	2.23%	1.69%	2.29%	1.73%	**1.55%**	1.69%
	Apr	RMSE	58.53	45.40	39.55	44.85	46.42	35.65	**26.14**	27.98	31.27	33.44
		MAPE	3.03%	1.93%	2.33%	2.54%	2.50%	1.75%	**1.37%**	1.54%	1.83%	1.89%
	Jul	RMSE	75.05	47.68	56.12	48.55	59.99	38.03	37.38	39.16	31.93	**29.18**
		MAPE	4.25%	2.34%	3.53%	3.07%	3.49%	1.98%	2.17%	2.33%	1.86%	**1.70%**
	Oct	RMSE	48.49	42.74	37.94	40.56	42.48	43.59	30.16	**25.14**	26.59	34.17
		MAPE	2.62%	1.77%	2.23%	2.17%	2.30%	1.92%	1.67%	1.69%	1.71%	**1.62%**

Table 4: Prediction results for one day ahead load forecasting

Dataset	Month	Metrics	Prediction model									
			Persistence	SVR [11]	ANN [36]	DBN [15]	RF [23]	EDBN [27]	EMD-SVR [57]	EMD-ANN [58]	EMD-RF	Proposed
NSW	Jan	RMSE	978.24	703.43	750.53	639.75	**521.14**	636.03	611.20	748.30	544.17	541.53
		MAPE	8.55%	6.23%	7.2%	5.95%	**4.26%**	5.70%	5.19%	6.66%	4.54%	4.62%
	Apr	RMSE	729.50	474.38	578.05	**361.63**	500.70	551.74	569.28	512.59	495.28	377.63
		MAPE	6.71%	4.27%	5.41%	3.36%	4.25%	4.78%	5.27%	4.57%	4.21%	**3.22%**
	Jul	RMSE	609.82	574.30	534.75	415.81	387.15	414.90	402.69	345.90	353.90	**322.04**
		MAPE	6.22%	5.86%	5.38%	4.11%	4.01%	4.07%	3.95%	3.09%	3.67%	**3.08%**
	Oct	RMSE	587.14	393.32	345.07	350.82	296.53	334.12	**272.01**	299.34	333.82	282.34
		MAPE	5.36%	3.74%	3.48%	3.41%	2.78%	3.14%	2.76%	2.90%	3.17%	**2.71%**
TAS	Jan	RMSE	89.82	60.97	69.92	63.96	65.90	60.68	61.73	63.38	58.51	**56.10**
		MAPE	7.24%	4.81%	5.42%	4.98%	4.77%	4.82%	4.49%	4.87%	4.67%	**4.05%**
	Apr	RMSE	157.73	111.89	94.40	93.81	92.64	109.78	104.59	87.41	86.61	**85.13**
		MAPE	10.22%	7.48%	6.3%	6.12%	6.10%	7.28%	6.87%	5.92%	**5.80%**	**5.80%**
	Jul	RMSE	120.47	90.99	89.17	87.30	90.48	85.19	92.54	82.92	81.34	**73.91**
		MAPE	8.11%	5.89%	6.28%	6.04%	6.17%	6.04	6.09%	5.50%	5.54%	**4.93%**
	Oct	RMSE	109.46	79.45	72.86	75.73	69.80	80.81	82.85	80.85	73.86	**68.26**
		MAPE	7.48%	5.55%	5.24%	5.15%	**4.63%**	5.05	5.60%	5.63%	4.88%	4.75%
QLD	Jan	RMSE	461.09	282.07	299.32	228.86	195.85	218.55	196.20	273.70	**178.63**	191.22
		MAPE	5.25%	3.65%	3.61%	2.78%	2.41%	2.69%	2.56%	3.28%	**2.21%**	2.56%
	Apr	RMSE	489.63	266.39	339.93	247.56	231.01	259.34	264.00	237.58	**201.74**	243.68
		MAPE	6.25%	3.53%	3.77%	2.99%	2.78%	3.33%	3.47%	3.11%	**2.44%**	2.93%
	Jul	RMSE	430.46	223.17	203.00	213.20	156.08	159.45	164.68	174.64	150.01	**142.84**
		MAPE	5.90%	3.10%	3.03%	2.95%	2.32%	2.32%	2.46%	2.45%	2.29%	**2.08%**
	Oct	RMSE	417.33	298.76	263.12	251.34	236.50	292.93	**218.71**	248.55	260.94	219.19
		MAPE	5.54%	3.93%	3.46%	3.40%	2.88%	3.53%	**2.82%**	3.27%	3.15%	2.88%
VIC	Jan	RMSE	990.74	**587.98**	811.43	915.21	739.65	762.16	806.29	781.17	783.58	762.57
		MAPE	9.48%	**7.16%**	9.32%	8.79%	8.77%	9.14%	9.48%	9.07%	9.32%	8.86%
	Apr	RMSE	669.87	330.93	359.03	353.02	366.16	343.18	363.50	376.12	393.63	**321.59**
		MAPE	8.40%	4.43%	4.95%	4.55%	4.65%	4.49%	4.67%	4.79%	5.04%	**4.35%**
	Jul	RMSE	721.85	297.07	305.88	**276.25**	302.15	285.14	298.12	386.64	300.65	285.45
		MAPE	9.76%	4.38%	4.29%	**3.72%**	4.29%	3.65%	4.27%	5.29%	4.26%	3.83%
	Oct	RMSE	577.70	391.11	347.91	389.06	364.32	401.02	309.63	332.44	344.06	**322.91**
		MAPE	8.30%	4.50%	4.79%	4.85%	4.16%	4.72%	3.78%	4.15%	3.92%	**3.73%**
SA	Jan	RMSE	433.57	337.10	411.66	401.25	349.87	363.49	280.70	397.66	288.85	**238.09**
		MAPE	14.32%	13.34%	13.72%	13.62%	13.41%	14.43%	11.13%	13.80%	13.03%	**10.46%**
	Apr	RMSE	180.20	124.43	119.4	**117.61**	127.90	105.39	121.60	126.78	124.60	125.31
		MAPE	9.36%	6.71%	**6.42%**	6.67%	6.88%	6.56%	6.78%	6.87%	6.65%	6.76%
	Jul	RMSE	289.94	150.84	151.06	148.23	154.67	148.55	**141.78**	153.22	161.71	160.82
		MAPE	16.84%	8.54%	8.66%	8.50%	8.95%	8.59%	**8.48%**	9.53%	9.12%	9.60%
	Oct	RMSE	240.53	210.72	233.48	204.16	218.30	203.53	203.38	199.77	209.70	**192.74**
		MAPE	11.54%	8.94%	10.03%	9.33%	9.11%	9.32%	8.39%	8.54%	8.22%	**8.11%**

Table 5: Forecasting results for monthly electric load demand in Northeastern China as used in [61, 60]

Time point	Actual	ARIMA	TF-ε-SVR-SA [60]	SRSVRCABC [61]	Proposed
October 2008	181.07	192.9316	184.5035	178.4199	181.1451
November 2008	180.56	191.127	190.3608	188.3091	182.4483
December 2008	189.03	189.9155	202.9795	195.3528	184.2608
January 2009	182.07	191.9947	195.7532	187.0825	184.1379
February 2009	167.35	189.9398	167.5795	166.1220	178.7636
March 2009	189.30	183.9876	185.9358	185.1950	184.8475
April 2009	175.84	189.3480	180.1648	179.5335	175.7244
MAPE(%)		6.044	3.799	2.387	**1.998**

Table 6: Forecasting results for electric load demand in New South Wales in 2007 [62]

Sample Size	Metric	SVR	AFCM [62]	Proposed
Small	MAPE	1.3678%	0.9905%	**0.6695%**
	RMSE	145.865	125.323	**83.570**
Large	MAPE	1.6580%	1.2325%	**0.9187%**
	RMSE	181.617	158.754	**118.492**

Table 7: Comparative Results with PSF-NNs [63]

Prediction method	MAE[MW]	MAPE[%]
Proposed	**266.58**	**3.00**
PSF	352.03	3.96
PSF-NN1	311.94	3.44
PSF-NN2	402.13	4.51
PSF-NN3	300.77	3.37