Electric Power Systems Research 168 (2019) 184–199

Contents lists available at ScienceDirect

Electric Power Systems Research

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

A combination approach based on a novel data clustering method and Bayesian recurrent neural network for day-ahead price forecasting of electricity markets

M. Ghayekhloo[a], R. Azimi[b,*], M. Ghofrani[c], M.B. Menhaj[d], E. Shekari[e]

[a] Department of Computer & Information Technology Engineering, Amirkabir University of Technology, Tehran, Iran
[b] Young Researchers and Elite Club, Qazvin Branch, Islamic Azad University, Qazvin, Iran
[c] School of Science, Technology, Engineering and Mathematics (STEM), University of Washington, Bothell, USA
[d] Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran
[e] Department of Decision Science and Knowledge Engineering, University of Economics, Tehran, Iran

ARTICLE INFO

Keywords:
Bayesian approach
Data mining
Forecasting
Fourier analysis
Recurrent neural network
Time series analysis

ABSTRACT

Electricity price forecasting is an ancillary service that plays a key role for market participants in a deregulated structure. Compared to other commodities, electricity price exhibits higher levels of volatility and uncertainty. This imposes more restrictions on the accuracy of the forecast and leads to significant errors. This paper proposes a hybrid electricity price-forecasting framework with a novel time series data mining method to enhance the feature selection. The proposed method includes clustering, preprocessing and training stages. The proposed data clustering method uses both an enhanced game theoretic approach and neural gas in combination with competitive Hebbian learning to provide a better vector quantization. Six strategies are proposed to enable the non-winning neurons to participate in the learning phase and resolve the shortcomings of the original self-organizing map, where the dead neurons are far from the input patterns without having any chance to compete with the winning neurons. The price-load input data are clustered into a proper number of subsets using the proposed data mining method. A novel cluster selection method based on the persistence approach is applied to select the most appropriate cluster as the input to the BRNN. The selected data set is filtered by the harmonic analysis time series, and is time-series processed to provide the proper inputs for training neural networks. Bayesian approach is used to train a recurrent neural network, and forecast the electricity price. The performance of the proposed clustering algorithm is evaluated using different electricity market data. Our results demonstrate the efficiency of the proposed clustering algorithm as compared to K-means, neural gas and self-organizing map clustering methods Our proposed clustering provides 16.7%, 28.6%, and 13% more accurate results than K-means, neural gas, and self-organizing map for the NYISO. CAPITL data. For the NYISO. CENTRL data, the developed clustering outperforms the K-means, neural gas, and self-organizing map by 21.4%, 21.4%, and 8.3%, respectively. The clustering accuracy of the proposed method for NYISO. DUNWOD data is 5.5%, 19%, and 5.5% better than that of the K-means, neural gas, and self-organizing map methods Lastly, for the NYISO. GENESE data, the mean square error value for the proposed clustering is 13.7%, 14.9%, and 12.5% less than that of the K-means, neural gas, and self-organizing map, respectively. The developed forecasting method is also compared with the existing state-of-the-art forecasting algorithms. The comparison results show an improvement in the forecast accuracy of the developed method over other forecasting approaches.

Abbreviations: ACO, ant colony optimization; ANFIS, adaptive-network-based fuzzy inference system; ANN, artificial neural network; AR, autoregressive; ARIMA, autoregressive integrated moving average; ARMA, autoregressive moving average; ARMAX, autoregressive moving average with exogenous inputs; AWNN, adaptive wavelet neural network; BBB, Bayes by Backprop; BMU, best matching unit; BRNNs, Bayesian recurrent neural networks; CHL, competitive Hebbian learning; CNEA, cascaded neuro-evolutionary algorithm; DFT, discrete Fourier transform; DG, distributed generation; DM, Diebold–Mariano (test); FA, firefly algorithm; FNNs, fuzzy neural networks; GARCH, generalized autoregressive conditional heteroskedastic; GP, Gaussian process; HANTS, harmonic (Fourier) analysis of time series; HIS, hybrid intelligent system; IVM, informative vector machine; KELM, kernel extreme learning machine; LAI, leaf area index; LMA, load management agent; LS-SVM, least squares support vector machines; LST, land surface temperature; MA, moving average; MAPE, mean absolute percentage error; MAS, multi-agent systems; MSE, mean square error; NDVI, normalized difference vegetation index; NG, neural gas; NNs, neural networks; NNWT, neural networks and wavelet transform; NYISO, New York independent system operator; PDBT, polarization difference brightness temperature; PHEVs, plug-in hybrid electric vehicles; PSO, particle swarm optimization; QOABC, quasi-oppositional artificial bee colony; RBFN, radial basis function neural networks; RMSE, root mean square error; RNNs, recurrent neural networks; SOM, self-organizing map; SVR, support vector regression; VQ, vector quantization; WNNs, weighted nearest neighbors

* Corresponding author at: Islamic Azad University, Qazvin Branch, Qazvin Province, Qazvin, Barajin, Barajin Rd, Iran.
E-mail address: r.azimi@qiau.ac.ir (R. Azimi).

https://doi.org/10.1016/j.epsr.2018.11.021
Received 4 June 2017; Received in revised form 21 August 2018; Accepted 30 November 2018
Available online 12 December 2018
0378-7796/

Nomenclature

h_{ci} The neighborhood function
$\hat{P}(n)$ The price forecast for hour n
$P_{Actual}(n)$ The actual price for hour n
r_c The positions of the winning neuron
$r_i \in R2$ The positions of neuron i in the two-dimensional lattice
$\tilde{y}$ Reconstructed and error time series
$\text{testTarget}^*_{\text{Load}}$ m latest data in Train Load
$\text{testTarget}^*_{\text{Price}}$ m latest data in Train Price
$\text{Train}^*_{\text{Load}}$ Train Load data − {m latest data in Train Load data}
$\text{Train}^*_{\text{Price}}$ Train Price data − {m latest data in Train Price data}
ClusterData The 2-dimensional clustered data provided by EGT-Cluster
F = {f1, …, fm} Payoff functions for each strategy
K Number of clusters
Load-ClusterData The 1-dimensional clustered load data provided by EGT-Cluster
m Number of hours ahead for electricity price forecasting
MergedData The 2-dimensional data provided by combining 1-dimensional $\text{Train}^*_{\text{Price}}$ and $\text{Train}^*_{\text{Load}}$ data
mi The weight vector associated with neuron i
N Number of price/load training data
P = {p1, …, pn} A set of players
P Total number of price/load data
Price-ClusterData The 1-dimensional clustered price data provided by EGT-Cluster
S = {s1, …, sm} A set of actions for each player
x Parbitrary input pattern
α The learning rate
nf The number of frequencies in the time series
y The original time series
ε Error time series
σ The neighborhood size
$\hat{p}_i^h(n)$ The i-th competition of the h-step ahead forecasting
$e_i^h(n)$ The forecasting errors from the i-th competing models
$e_i^h(n)$ The h-step forecasting errors
H_0 The null hypothesis
H_1 The alternative hypothesis
$d(n)$ The differential costs function
$\bar{d}$ The sample mean differential cost

1. Introduction

Accurate forecasting of electricity price plays an important role for participants in a deregulated electricity market. Electricity price exhibits high levels of volatility and uncertainty, which deteriorate forecasting accuracy and results [1]. Today, data mining methods have been widely used in predicting renewable energy and electricity price data analysis. The complexity of training neural networks for price data with highly irregular patterns needs utilizing a comprehensive data mining and deep and complex approaches to enhance forecasting accuracy [2]. This paper proposes a complex data-mining framework based on a novel data clustering algorithm and Bayesian recurrent neural network for electricity price forecasting.

1.1. Motivation

Ancillary services (AS) are commonly known in the electric power industry as a collection of secondary services to help ensure the reliability, availability, and quality of electricity supply to consumers. Electricity price forecasting is as an ancillary service that plays a key role for market participants in a deregulated structure. As compared to other commodities, electricity price time series data exhibits higher levels of volatility and uncertainty. The uncertainties enforces more restrictions on the accuracy of the forecast and leads to significant errors [3].

1.2. Literature survey

Price forecasting algorithms are classified into three major categories: game theory models [4]; time series models such as autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedastic (GARCH), dynamic regression and linear transfer function models [5–9]; and intelligent methods such as neural networks (NNs) [10], fuzzy neural networks (FNNs) [11] and weighted nearest neighbors (WNN) [12]. The game theory models involve the mathematical solution of the market participants' strategies to forecast the electricity prices. Bahrami and Parniani [13] proposed a new game theory based framework for managing the plug-in hybrid electric vehicles (PHEVs) charging load to reduce peak load. This approach considers the PHEVs charging as a game among all users, to minimize all car owners' cost of charging. The Nash equilibrium point is also applied to solve the optimization problem of the load management.

The time series methods are usually linear models with limited or even no capability to characterize the nonlinear behavior of the price signals. In addition, they are more appropriate for the data with low-frequency changes, and therefore encounter large prediction errors for the price data whose variations are rapid and changes are of high-frequency. Furthermore, the stationary process considered for most of these models cannot capture the non-stationary characteristics of the price series. Intelligent methods provide the non-linear mapping of the non-stationary processes such as price series. However, factors such as insufficient input-output data points or too many tunable parameters may limit their capabilities to extract a global model for the price data. Hybrid of different techniques is proposed as a solution to suppress the limitations of the individual forecasting methods and enhance the performance and accuracy of the price forecast.

1.2.1. Hybrid approaches

Hybrid algorithms have been studied in Refs. [14–25]. Cai et al. [14] used the ant colony optimization algorithm (ACO) to provide a good partition of the universe of discourse, and auto-regression was introduced to better use historical information and improve the forecasting performance. A hybrid of wavelet transform and ARIMA model was used in Ref. [15] to decompose the price series into different frequency components and provide the forecast for each constitutive series. Panapakidis and Dagoumas [16] proposed a combination of neural networks and wavelet transform (NNWT) technique for short-term price forecasting. A short-term forecasting approach based on the wavelet transform and a hybrid of neural networks and fuzzy logic (WNF) was proposed in Catalão et al. [17]. Wavelet transform, ARIMA models and radial basis function neural networks (RBFN) were combined in Shafie-Khah et al. [18] to forecast the day-ahead prices in a competitive electricity market. Amini et al. [19] proposes an ARIMA-based decoupled time series forecasting approach for electric vehicle charging points. The proposed approach improves the accuracy of ARIMA forecasting by tuning the autoregressive parameters. Catalão et al. [20] presented a forecasting technique that combines WT, particle swarm optimization (PSO) and adaptive-network-based fuzzy inference system (ANFIS). A hybrid ARMA with exogenous variables (ARMAX)-GARCH and adaptive wavelet neural network (AWNN) model was proposed in Wu and Shahidehpour [21]. The model used one-period continuously compounded return series instead of the original series to better capture statistical properties of electricity prices such as spikes, mean reversion and seasonality. Amini et al. [22] proposed a self-decision making method for load management. This method applied the Multi-Agent Systems (MAS) by combining the upstream grid, Distributed Generation (DG) and Demand Response Resources (DRR) for peak demand reduction in distribution network feeders. In this method, ANN was used to provide a short term load forecasting. The forecasting output is then used in Load Management Agent (LMA) decision making in constructing commitment protocol between several resources.

A combinatorial neural network-based forecasting engine was proposed in Abedinia et al. [23] that is equipped with a new training mechanism based on modified Chemical Reaction optimization algorithm. Jin et al. [24] developed a novel method for day-ahead electricity price forecasting based on clustering and next symbol prediction. Self-organizing map (SOM) clustering was applied to cluster pattern sequences and obtain corresponding topology relations. In the next symbol prediction step, artificial neural network was applied to predict the topological coordinates of the next day by training the relationship between previous daily pattern sequences and its next day pattern.

Shayeghi et al. [25] proposed a new hybrid method for load and electricity price forecasting. The proposed algorithm consists of three stages; a pre-processing stage to detect change of seasons, tagging days, and to search for similar days; the second stage which is a multi-input multi-output least squares support vector machines (LS-SVM) based hourly predictor; and finally the third stage to optimize the LSSVM parameters using the Quasi-Oppositional Artificial Bee Colony (QOABC) algorithm.

1.2.2. Clustering techniques and their applications for forecasting

Data clustering of electricity price time series can significantly enhance the forecast accuracy. Clustering techniques and their application to identify patterns of price behavior have been studied in [26–34]. Several partitioning-clustering techniques such as K-means, C-means, and expectation maximization (EM) were applied to electricity price time series [26–29]. Feijoo et al. [30] proposed a novel method, called K-SVR, that combines K-means clustering algorithm, support vector machine, and support vector regression to obtain better forecast accuracy and improve the computational speed. MartínezÁlvarez et al. [31] demonstrated that partitioning-clustering techniques better classify the price data than fuzzy clustering methods such as fuzzy C-means. Subtractive clustering technique was used in Zhou et al. [32] for data classification to decrease the sample size. The cluster centers were then utilized as initial rules for an adaptive neuro-fuzzy interface system. Bompard et al. [33] presented a data clustering technique based on similar load profiles for dynamic price forecasting. The SOM algorithm [40] was employed in Fan et al. [34] to cluster the input data set into several subsets in an unsupervised manner. Limitations of SOM for non-convex or discontinuous input distributions and possible existence of dead neurons reduce the accuracy of the map and forecast. Game theory techniques have been used for SOM training to overcome these limitations and provide a global optimization for the map [35,36]. Herbert and Yao [35] proposed a training algorithm where the neurons compete as players with strategy sets and utility functions. The neuron whose action better enhances the overall quality of the map is selected as the winning neuron. However, the feasibility and accuracy of the proposed method is in question, as no experimental result was provided in the paper. In addition, the method requires defining the utility function of each player based on all players' actions and solving the game for each iteration which makes it computationally excessive. Unlike the original SOM algorithm where the weight vectors of all neighboring neurons are updated, the method developed in Neme et al. [36] enables the winning neuron to have different strategies and select between its neighbors to adapt their weights. However, the strategies considered were non-cooperative in topographic map formation without the intention of minimizing the error measure. Wang et al. [37] proposed a two-layer decomposition technique to develop a hybrid model based on fast ensemble empirical mode decomposition, variational mode decomposition and back propagation neural network algorithm optimized by firefly algorithm (FA). Yang et al. [38] proposed a hybrid approach for day-ahead electricity price forecasting by combining the wavelet transform to analyze complex features of prices series, the kernel extreme learning machine (KELM) based on self-adapting particle swarm optimization and an ARMA model.

1.3. Contributions of manuscript

This paper proposes a hybrid method to forecast the electricity price. The proposed method includes clustering, preprocessing and training stages. An enhanced data clustering method based on six game theories is developed to assign several different strategies to the non-winning neurons to provide a competitive game. This resolves the major problem of the previous studies, such as Ref. [39], where the weight vectors of dead neurons are far from the input patterns without having any chance to contribute in the learning phase.

In the proposed method, NG with CHL is used to define the neighborhood based on the distance of neurons in the input space. This accelerates the learning process and improves the vector quantization. The developed clustering method, named enhanced game theoretic clustering (EGT-Cluster), clusters the input price-load data into appropriate number of subsets which are then preprocessed by the HANTS method and time-lagged analysis. The proposed preprocessing extracts the irregular price information and better characterizes the price behavior by decomposing the price subsets into proper levels of resolution. The preprocessed data provide the input for the training stage where the Bayesian recurrent neural network (BRNN) method is used to forecast the electricity price.

The contributions of the paper are outlined as follows:

1. An enhanced data clustering method, named EGT-Cluster, is proposed based on six new game-theoretic strategies to enhance the clustering performance.
2. An embedded 2-dimensional input selection is introduced to the proposed EGT-Cluster to include both load and price time series in the clustering.
3. A novel method based on persistence approach is developed to select the cluster that provides the most accurate forecast for *m* hours ahead.
4. A hybrid method using data clustering, time-lagged signal analysis and BRNN is proposed for day-ahead electricity price forecasting.

1.4. Organization of manuscript

The rest of the paper is organized as follows. Section 2 explains the basic concepts of game theories, and the proposed clustering method. It also describes the hybrid method developed for the price forecasting. Vector quantization errors of the proposed clustering algorithm and their comparisons with those of the K-means, the original SOM and NG methods are given in Section 3. Accuracy results of the developed and state-of-the-art forecasting methods are also provided in this section. In Section 4, the performance of the proposed method is evaluated and compared with other methods. The conclusions and future work are provided in Section 5.

2. Methodology

2.1. Game theory

Game theory is the study of situations where agents with conflicting interests are involved [41]. Aspects of game theory are applicable to any problem or application that can be expressed as a game. Game theory provides an advanced problem solving technique with applications in many domains. In this approach, each agent is a player with a strategy set and a pay-off function. Players' actions in every stage of the game are determined by the strategies. The pay-off for each player is affected by both his and other players' actions. In a simple game put into formulation, a set of players P = {p_1, …, p_n}, a set of actions S = {s_1, …, s_m} for each player, and the respective payoff functions for each strategy F = {f_1, …, f_m} are observed from the governing rules of the game. The players are rational in a sense that they want to maximize their payoffs. Thus, each player chooses strategy from S to be performed according to the expected payoff from P.

2.2. The proposed clustering method

An enhanced data clustering method, named EGT-Cluster is developed and used in this paper to cluster the price-load data. Players are involved in a non-cooperative game and compete to obtain more input patterns.

Each neuron in the network is a player and can have his strategy and it can takes action by looking at each input pattern under its current strategy. Given this view, the neurons have the ability to decide. That way, by applying any input patterns to the network, the neurons can decide for their direction and amount of their movement based on a set of criteria.

The ideas that are proposed in this paper to determine the movement of neurons and the direction of their movement are simple and interesting. Suppose a pattern is applied to the input space and a neuron is identified as the winner, so that this neuron is closer to the vector. If this neuron decides to go toward the vector, it moves to that direction, and the other neurons that see it look for input patterns in a different direction than the winning neuron and moves on that direction.

Generally, six strategies are proposed to enhance the clustering performance.

Strategy A: winning neuron and its neighbors adapt their weights to approach the input pattern and minimize the Euclidean distance. Non-winning neurons are assigned different strategies and actions depending on their situations and the current iteration number. The neurons situations are determined based on their locations in the vector space.

Strategy O: the neurons' weight vectors are often initialized with values close to each other in the center of the input space. In addition, the patterns are scattered within the input space with an equal probability. At early iterations, the non-winning neurons may choose to move in the opposite direction of the winning neuron. This strategy increases their chance of reaching a cluster that exists in a direction different from the winning neuron's moving direction.

Strategy F: in this strategy, all non-winning neurons move towards a plane that is orthogonal to the winning neurons. This strategy allows other neurons to be the winners in future competitions.

Strategy S: another strategy for the non-winning neurons is to stay in their current positions. This is applicable for the recently winning neurons or the neurons that have won many times as they most likely approached regions with sufficient input patterns. In fact, neurons that are placed in the proper place of space and there are large number of patterns around them, do not require the strategy of moving in the opposite direction of the winning neuron. The strategy of these neurons must stay in place and wait for the arrival of the new patterns. In this regard, another case seems to be when the neuron has recently been won. It seems that choosing the strategy of moving in the opposite direction of the winning neuron is not suitable for the corresponding neuron. Because this neuron may be approaching an area in space, in which there are parts of input patterns. Therefore, it is better for related neuron to prove a little bit in this place and wait to see if another pattern appears nearby.

Strategy R: this strategy is appropriate for the non-winning neurons that are wandering in regions with insufficient or no input pattern, to move randomly in the input space. The random moves increase the neurons' chances to approach regions with sufficient input patterns. This strategy is allowed at early stages of iteration.

Strategy B: the last strategy for the non-winning neurons is to approach the neuron defined as the best player. The best player is identified by an error variable (E_c) calculated as:

$$E_c(t)=E_c(t-1)+||x-m_c|| \tag{1}$$

The Euclidean distance between the input pattern and BMU is added to the cumulative error of the neuron to calculate the error variable. The number of times a neuron wins the competition and becomes the BMU is then calculated as a counter. The error variable of (1) is divided by the counter variable to calculate the average cumulative error for each neuron. The neuron with the minimum average cumulative errors is the best player. Table 1 summarizes the proposed strategies considered for each neuron.

The SOM approach provides a topographic map that preserves the topology of the input data. This means that the two input patterns, which are close in the input space, will remain close in the trained map [40]. The topology preserving property of the map requires defining the neighborhood function based on the distance between neurons in the 2-D lattice. However, the proposed strategies are not consistent with the topology preserving property of SOM. Unlike the SOM method, NG poses no explicit constraints on the lattice. CHL technique is a proper tool to generate the lattice and constantly update it [42,43]. Therefore, the NG algorithm is first used to distribute a specified number of centers. CHL is then used to generate the topology. These techniques can be also applied concurrently [44]. This application requires removing obsolete edges, which could be generated earlier due to the motion of the centers. An edge aging scheme is used to remove such edges [42]. The NG is adapted based on the distances in the input space and not based on the network topology. Therefore, the outcome of the NG method is not affected by the CHL algorithm. On the other hand, NG moves the centers and therefore influences the topology generated by CHL. A combination of NG and CHL, as described, is used in this paper as an effective method for topology learning.

The NG algorithm adapts the weights and sorts the neurons in an order (i_1, i_2, …, i_m) which is based on their distance from the input values. The weight adaptation is related to the position Rank(i) of the i-th neuron as follows:

$$w_i(t+1)=w_i(t)+\alpha(t).\ h_{\sigma(t)}(Rank(\mathrm{i}))\cdot[x(t)-w_i(t)] \tag{2}$$

where the values of $\alpha(t)$ and $\sigma(t)$ are uniformly reduced over time.

$$\alpha(t)=\alpha_0\exp\left(-\frac{t}{\tau_1}\right) \tag{3}$$

Table 1
Proposed strategies for neurons.

Neuron	Strategy
Winning neuron and its close neighbors	A (Approach): winning neuron and its close neighbors move towards the input pattern
Non-winning neurons	O (Opposite): neurons move in opposite direction of the winning neuron
	F (Follow): all non-winning neurons move towards a plane that is orthogonal to the winning neuron
	S (Stay): neurons stay in their current positions
	R (Random): neurons move to random positions in the input space
	B (Best player to approach): neurons approach a neuron surrounded with sufficient input patterns

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\tau_2}\right) \tag{4}$$

where τ_2 and τ_2 are predefined by the user. $h_{\sigma(t)}$ is a Unimodal function with variance $\sigma(t)$. Eq. (2) updates the weight vector of each neuron.

The algorithm uses the gradient descent technique as a first-order iterative optimization algorithm to the error function:

$$E_{NG} = \frac{1}{2C(\sigma)} \sum_{j=1}^{m} \int P(x).\ h_\sigma(Rank(x)).\ (x(t) - w_i(t))^2 d^n x \tag{5}$$

where

$$C(\sigma) = \sum_{k=0}^{m-1} h_\sigma(k) \tag{6}$$

And P(x) is the pattern probability distribution over a set of input patterns x.

Using (5), the neurons move towards the input patterns, according to their ranking based on the proximity to the input patterns. However, since the neuron position is no longer in the grid, the data topology is not preserved. Performing vector quantization only, this does not create a problem, but if the maintenance of the topology is also important, the CHL rule can be used. As a result, a combination of CHL and NG are used to find the neighboring relationships between neurons [42].

To create a topology in CHL method, with the arrival of each input pattern, a neuron is established between the closest neuron and the second nearest neuron to the input pattern. An age variable is then assigned to that neuron which is set at zero. The age of other previously formed neurons for the closest neuron is then increased by one unit. The neurons whose their age exceeds a predefined boundary range will be eliminated.

The clustering stage includes the proposed EGT-Cluster with the following algorithm:

1) Initialize the neurons' weight vectors and place the close values in the center of the input space.
2) Randomly select and apply an input pattern to determine the winning neuron.
3) Apply strategy A to the winning neuron.
4) Apply strategies O, F, S, R and B to the non-winning neurons based on their positions with the input patterns, the state of the corresponding neuron in previous iteration (In terms of winning and non-winning) and the current iteration number.
5) Update the neurons' weight vectors.
6) Check the termination criterion and repeat steps 2–6 if the criterion is not satisfied.
7) End the process if the termination criterion is satisfied.

The termination criterion is determined using the maximum number of iteration predefined by the user. However, if the weight vectors do not change in two consecutive iterations, the algorithm converges, regardless of the predefined maximum number of iteration.

Some of the parameters of the SOM-based method and its variants such as the proposed EGT-Cluster need to be chosen before running the algorithm. These parameters include the size and shape of the map, the distance measure used to determine the similarity of the nodes to each

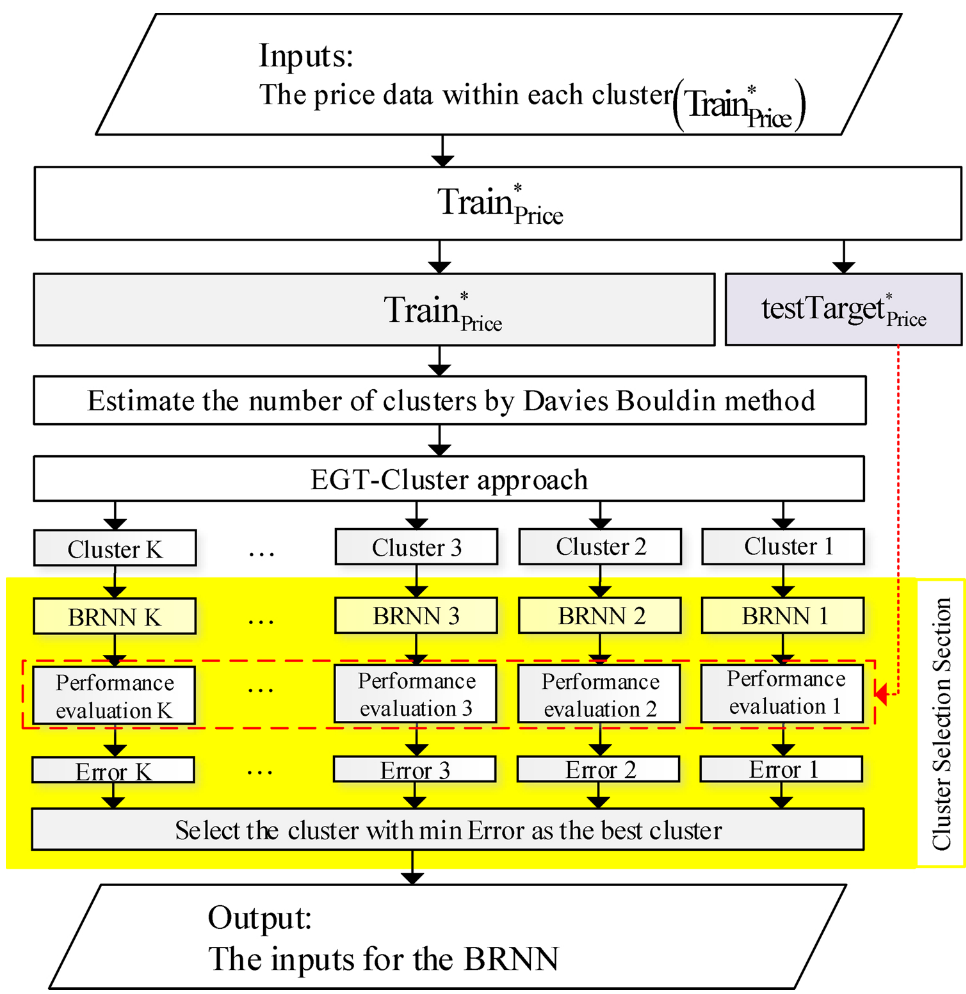

Fig. 1. Proposed cluster selection algorithm for an m-hour ahead price forecasting. The main part of proposed cluster selection method is highlighted in yellow.

other and to the input feature vectors, as well as the kernel function in use for the map training. This paper uses the Kohonen recommended values of Ref. [40] for these parameters.

The Davies Bouldin method [45] is first used to select the most appropriate number of sub-trains (K). The proposed EGT-Cluster method then splits the training data into K sub-trains. Several methods have been proposed to determine the number of clusters. These methods include, but are not limited to Elbow method [46], average silhouette method [47], Calinski–Harabasz method [48], Gap method [49] and the Davies Bouldin method. Among these methods, Davies–Bouldin is the least complex method with accurate results [50].

Splitting the training data into multiple sub-trains and selecting the most appropriate sub-dataset (the best cluster) as input to BRNN, significantly accelerates the forecast process. This is because the most appropriate portion of the data rather than the whole data is used for the BRNN training. The proposed technique is particularly important for very short-term forecasting where the forecast horizon can be as short as a few seconds ahead.

2.3. Proposed cluster selection method

A novel cluster selection method is proposed to select the most appropriate cluster as input to BRNNs.

Fig. 1 shows the flowchart for the proposed cluster selection method. The proposed cluster selection method is based on the persistence forecasting which considers the day before as the most important day for predicting the variable of interest at a specific day [65]. In this paper, the persistence method is considered in a different way where we assume the cluster that based on its data, the best prediction was obtained for today using the BRNN, is the most appropriate cluster for predicting tomorrow.

According to Fig. 1, the cluster selection algorithm for an *m*-hour ahead price forecasting is as follows:

i. The price data within each cluster are used as the inputs to individual BRNN. *K* processors are used in parallel to speed up the procedure.
ii. The $\text{testTarget}^{*}_{\text{Price}}$ data are used to evaluate the performance of each BRNN, as they are the most relevant data to the forecasts.
iii. An error is calculated for each BRNN by comparing its forecast outputs with the$\text{testTarget}^{*}_{\text{Price}}$. The cluster with the lowest prediction error is selected as the best cluster.

The data of the best cluster are used in the pre-processing stage (HANTS method and time series analysis) to provide the inputs for the BRNN.

2.4. Bayesian recurrent neural networks

An implementation of an RNN, called BRNN [45], which is trained using variation Bayes for RNNs [51], is used in this paper to estimate the electricity price and provide the forecast.

RNNs are widely used for sequence prediction tasks due to their capability to selectively pass information through sequence steps and process sequential data one by one [52,53]. Therefore, they provide a powerful tool to model input and/or output containing sequences of elements that are dependent. They can also model time and sequential dependencies at the same time [55].

Information is carried across neurons while taking in input through RNN loops. Fig. 2 shows a simple diagram of an unrolled RNN.

where x_t is the input, A is part of the RNN and h_t is the output. h_t is then compared to the test data, which is usually a small subset of the original data.

Bayesian methods provide a tool for RNNs to express their uncertainty and regularize during training [54]. Variational inference via the Bayes-by-Backprop (BBB) algorithm is used to provide an efficient Bayesian training for RNNs [56]. Fig. 3 shows the application of BBB to RNN.

Results of Ref. [51] demonstrated that BRNNs have better performance over RNNs by better regulizing the network and providing superior uncertainty properties, particularly for out-of-distrubtion data.

2.5. Harmonic (Fourier) analysis of time series

Harmonic analysis represents functions or signals as the superposition of basic waves using the concept of Fourier series, Fourier integrals, Fourier transforms and the generalized function. The HANTS method determines the most important expected frequency of the time series, and uses the harmonic components for least squares curve fitting. The HANTS method is based on the DFT as given by:

$$\begin{aligned} \tilde{y}(t_j) &= a_0 + \sum_{i=1}^{nf} [a_i \cos(2\pi f_i t_j) + b_i \sin(2\pi f_i t_j)] \\ y(t_j) &= \tilde{y}(t_j) + \varepsilon(t_j) \end{aligned} \tag{7}$$

where y, $\tilde{y}$ and ε are the original, reconstructed and error time series; $t_j$is the time when y is observed; j = 1, 2, …, N is the number of samples in a time series; nf is the number of frequencies in the time series; and a_i and b_i are the coefficients of the i-th harmonic with frequency f_i. The curve fitting is done by optimizing (7) using the linear least square method.

The HANTS algorithm has been widely used to reconstruct time series of NDVI [57], LAI, and LST [58] as well as the PDBT [59] to remove random noise or eliminate cloud/snow contamination [60]. The HANTS method is useful for the noise suppression of volatile time series data such as electricity price data [61,62]. The HANTS method filters the time series electricity price data to better characterize the price behavior and provides more appropriate learning for BRNN to enhance the accuracy of the forecast results.

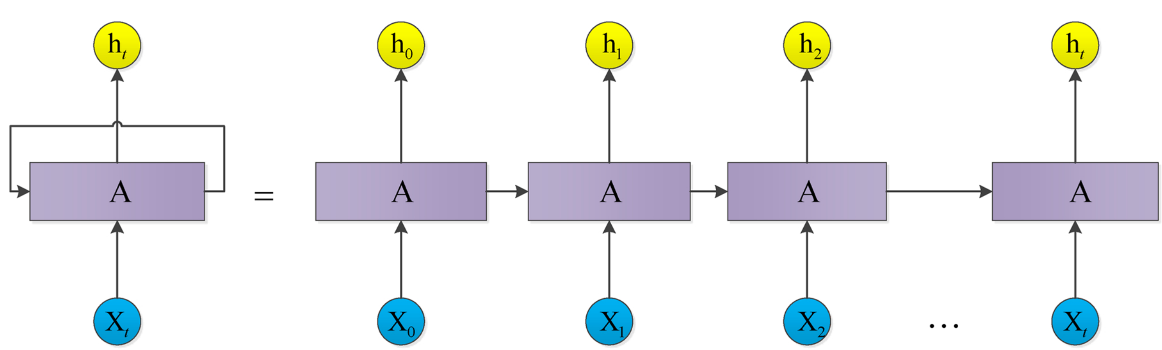

Fig. 2. A simple diagram of an unrolled RNN.

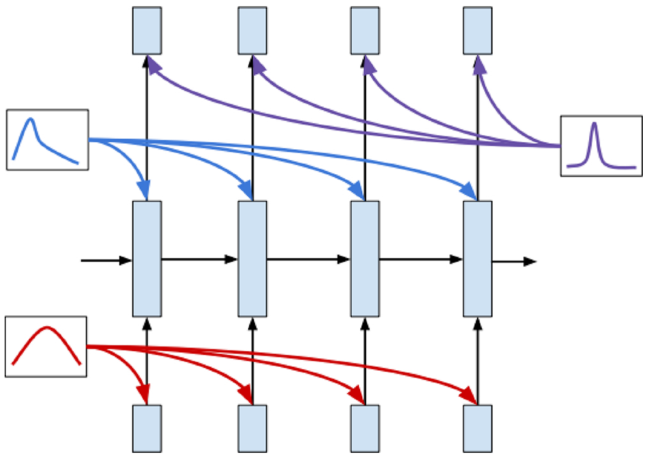

Fig. 3. Illustration of BBB applied to a RNN to provide an efficient Bayesian training for RNNs.

2.6. Time-lagged representations

Fig. 4 shows the structure of the time series for training and testing periods. This structure contains m' rows of training data where $P(t)$ is the current data point. The testing data includes m rows of inputs with the forecasts as outputs. For both training and testing inputs, N represents the length of the lagging window in time. Each set of N points in the training and testing data has a corresponding output and is shifted one step back to form the successive row of the associated data.

2.7. The proposed hybrid forecasting method

A hybrid forecasting method is proposed which consists of three stages: clustering; pre-processing; and training. Table 2 provides a description of the parameters used in the proposed hybrid forecasting.

Fig. 5 shows the proposed forecasting method. The following procedure is applied for an m-hour ahead prediction ($1 \le m \le 24$):

1) The hourly price and load data are used with 80% allocated for training and the remaining 20% for testing.

Table 2
Parameters and terminologies used in the proposed forecasting method, and their descriptions.

Parameter	Description
P	Total number of price/load data
N	Number of price/load training data
m	Number of hours ahead for electricity price forecasting
K	Number of clusters
$\mathrm{Train}^*_{\mathrm{Price}}$	$\mathrm{Train}_{\mathrm{Price}}$ data – {m latest data in $\mathrm{Train}_{\mathrm{Price}}$ data}
$\mathrm{Train}^*_{\mathrm{Load}}$	$\mathrm{Train}_{\mathrm{Load}}$ data – {m latest data in $\mathrm{Train}_{\mathrm{Load}}$ data}
MergedData	The 2-dimensional data provided by combining 1-dimensional $\mathrm{Train}^*_{\mathrm{Price}}$ and $\mathrm{Train}^*_{\mathrm{Load}}$ data
$\mathrm{testTarget}^*_{\mathrm{Price}}$	m latest data in $\mathrm{Train}_{\mathrm{Price}}$
$\mathrm{testTarget}^*_{\mathrm{Load}}$	m latest data in $\mathrm{Train}_{\mathrm{Load}}$
ClusterData	The 2-dimensional clustered data provided by EGT-Cluster
Price-ClusterData	The 1-dimensional clustered price data provided by EGT-Cluster
Load-ClusterData	The 1-dimensional clustered load data provided by EGT-Cluster

2) The price and load training data are divided to $\mathrm{Train}^*_{\mathrm{Price}}$, $\mathrm{Train}^*_{\mathrm{Load}}$ and $\mathrm{testTarget}^*_{\mathrm{Price}}$, $\mathrm{testTarget}^*_{\mathrm{Load}}$ sets.
3) The one-dimensional hourly price and load data in $\mathrm{Train}^*_{\mathrm{Price}}$ and $\mathrm{Train}^*_{\mathrm{Load}}$ are combined to form a two-dimensional data:

$$\mathrm{MergedData}(:,1) = \mathrm{Train}^*_{\mathrm{Price}}$$
$$\mathrm{MergedData}(:,2) = \mathrm{Train}^*_{\mathrm{Load}} \quad (8)$$

4) The proposed EGT-Cluster groups the MergedData into a number of clusters (K) determined by Davies Bouldin method.
5) The data in each cluster (1 to k) are sorted based on their time order.
6) The sorted data in each cluster (1 to k) are separated to one dimensional price and load data.

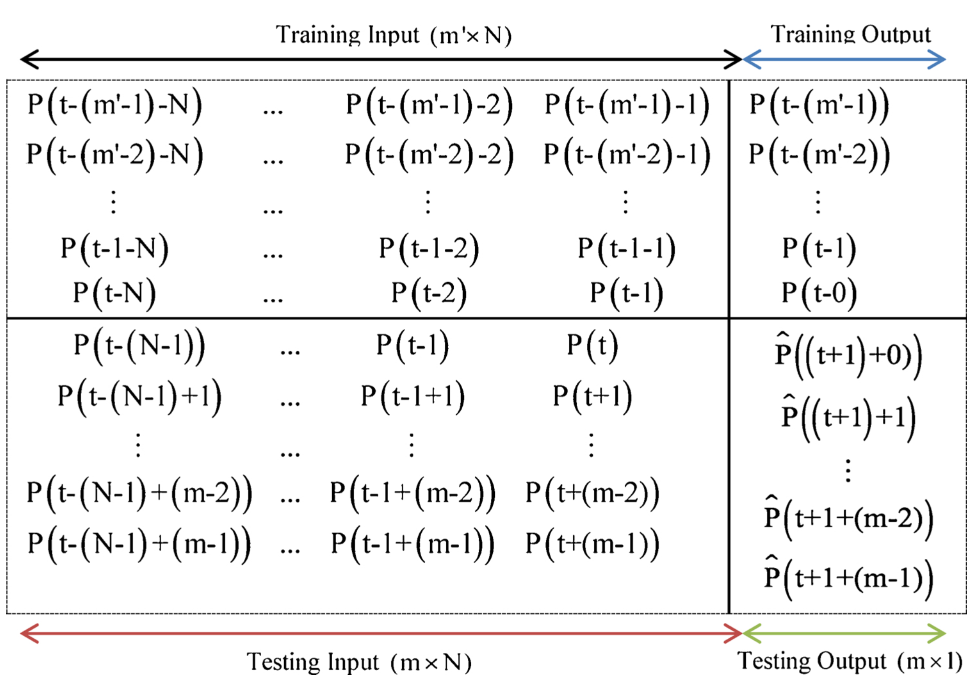

Fig. 4. Time-lagged structure for training and testing.

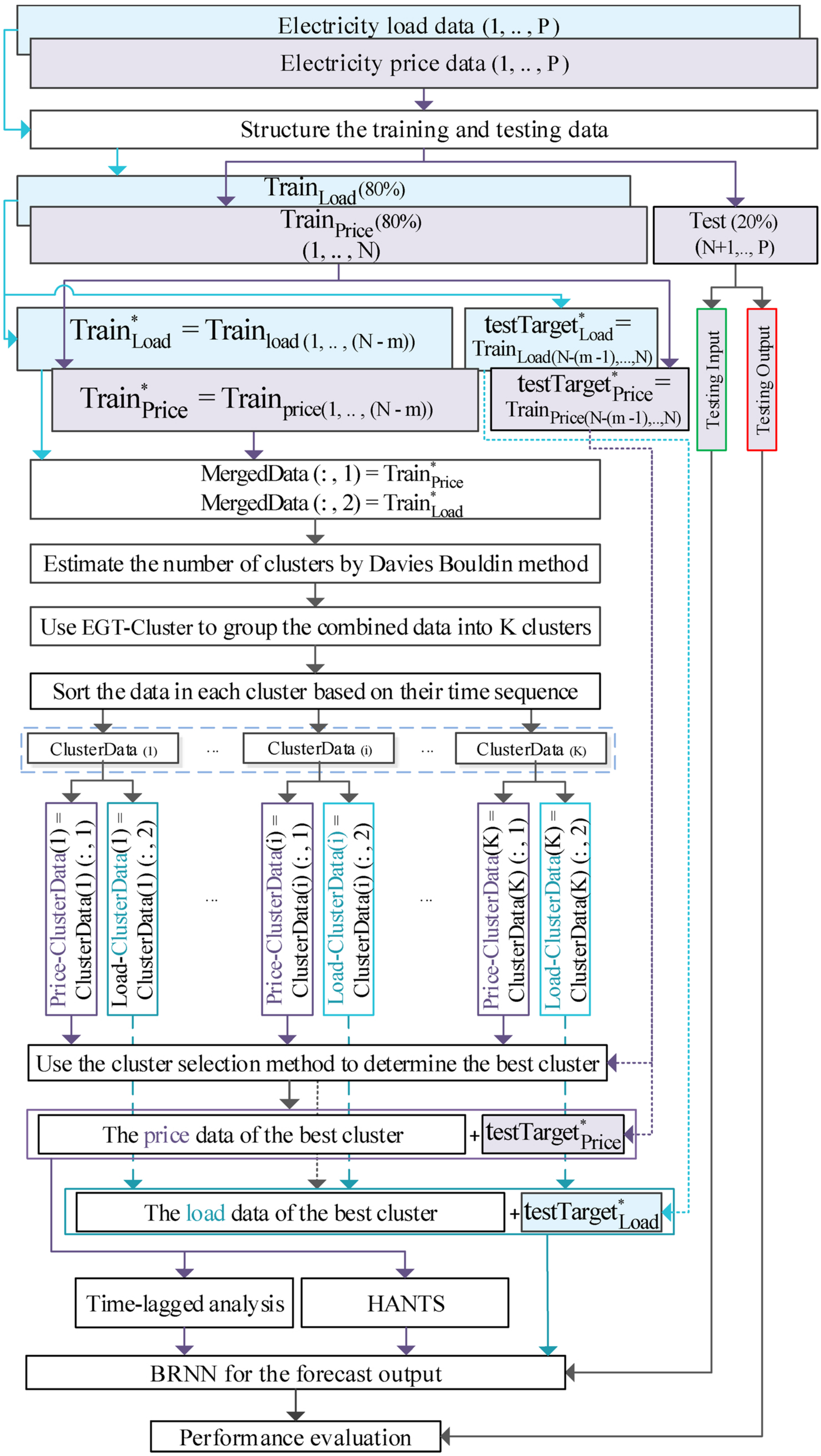

Fig. 5. Proposed forecasting method for an m-hour ahead electricity price prediction.

$$\text{Price-ClusterData}_{(K)} = \text{ClusterData}_{(K)}(:, 1)$$
$$\text{Load-ClusterData}_{(K)} = \text{ClusterData}_{(K)}(:, 2) \tag{9}$$

7) A cluster selection method is used to determine the cluster that provides the most accurate forecast for *m* hours ahead. The method is discussed later in this section.
8) The price data for the best cluster as well as the *m* latest data points of the price training data are processed by the HANTS method for identifying and removing outliers and provides Fourier coefficients of the filtered price time series as an effective input to the BRNN. These data are also time- lagged analyzed provide the other inputs to the BRNN.
9) The load data for the best cluster, which is already selected by the proposed cluster selection method based on electricity price forecasting, as well as the *m* latest data points of the load training data is directly used as other inputs for BRNN.

More information regarding the training parameters are as follows:

a) Number of input: 7 (lagged electricity price) + 1(electricity load) + 1(HANTS);
b) Number of hidden layers and neurons: 1 hidden layer with 8 neurons, 1 output layer;
c) Transfer function of the hidden layer: Tansing and purline;
d) Learning algorithm: Bayesian-Recurrent;
e) Comparison functions: MAPE, RMSE and Forecast skill;
f) Data distribution (train − test) = 80% train, 20% test.

2.8. Evaluation criteria

2.8.1. Data clustering evaluation

Mean squared error (MSE) criterion is used to evaluate the performance of the proposed EGT-Cluster method as follows:

$$MSE = \frac{1}{K.\ n_k} \sum\nolimits_{k=1}^{K} \sum\nolimits_{i=1}^{n_k} ||X_i^{(k)} - C_k||^2 \tag{10}$$

where K is the number of clusters, n_k is the number of data elements in cluster k; $X_i^{(k)}$ is a data point in the k-th cluster and C_k is the centroid for the k-th cluster.

The MSE function is the most commonly used cost function in clustering. The lower MSE values indicate better clustering results.

2.8.2. Time-series forecasting evaluation

Mean absolute percentage error (MAPE), root mean square error (RMSE), forecast-skill along with Diebold–Mariano Test [63,64] are used as the performance indicators in this paper. The equations for the three methods MAPE, RMSE and forecast-skill are as follows:

$$\text{MAPE}(\%) = \frac{1}{N}\sum_{n=1}^{N} \frac{|\hat{P}(n) - P_{Actual}(n)|}{P_{Actual}(n)} \tag{11}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (\hat{P}(n) - P_{Actual}(n))^2} \tag{12}$$

$$\text{Forecast Skill}(\%) = 100 \times \left(1 - \frac{\text{RMSE}}{\text{RMSE(Persistence)}}\right) \tag{13}$$

where N is the total number of hours, $\hat{P}(n)$ and $P_{Actual}(n)$ are the price forecast and the actual price for hour n. The forecasting skill is used as a comparison with the commonly used persistence forecasting method [65].

Finally, in order to perform a statistical comparative analysis (null hypothesis testing) of the best result selected with the RMSE, MAPE criterions, the Diebold–Mariano (DM) test function [26] is computed to provide a peer to peer comparison of the proposed methods with other forecasting models. The DM test compares the forecast accuracy of two forecasting methods based on testing the null hypothesis H_0 (where the two competing forecasts have the same level of accuracy) against the alternative hypothesis H_1 (where forecasting method 2 is less accurate than method 1).

The DM test function is as follows:

Let $\{P_{Actual}(n)\}$ denote the actual price time series, and $\{\widehat{p}_i^{\,h}(n)\}$, denote the i-th competition of the h-step ahead price forecasting time series. By considering $e_i^h(n)$ as the forecasting errors from the i-th competing models, the h-step forecasting errors $e_i^h(n)$ is:

$$e_i^h(n) = p^h(n) - \widehat{p}_i^{\,h}(n)(i = 1, 2, 3, \ldots m) \tag{14}$$

where m is the number of the forecasting models.

The accuracy of each forecast is calculated by the cost function:

$$L(p^h(n) - \widehat{p}_i^{\,h}(n)) = L(e_i^h(n)) \tag{15}$$

By considering h = 1, the superscript h is omitted in the following context. The squared-error loss function and the absolute-error loss function are the most commonly used cost functions. The squared-error cost function is given by:

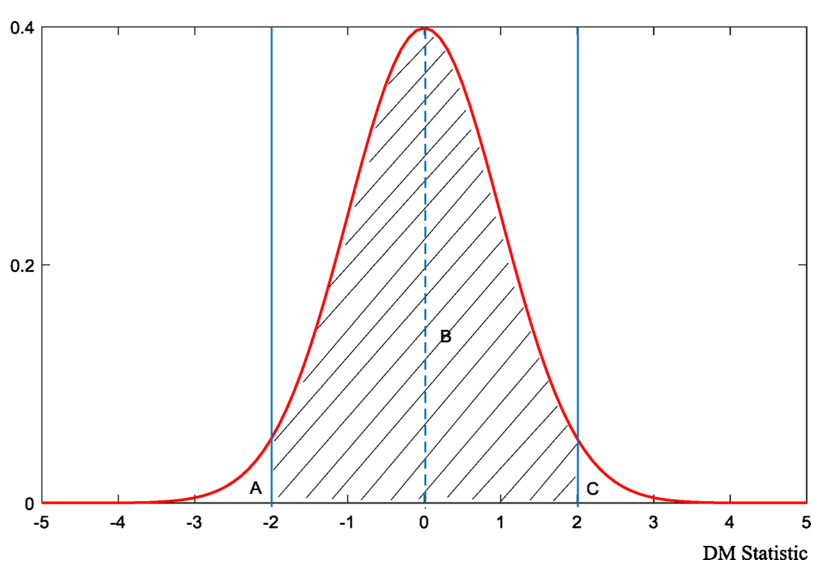

Fig. 6. The normal distribution.

$$L(p^h(n)-\widehat{p}_i^h(n)) = L(e_i^h(n)) = \sum_{t=1}^{T} (e_i(n))^2 \tag{16}$$

Absolute-error cost function is calculated by:

$$L(p^h(n)-\widehat{p}_i^h(n)) = L(e_i^h(n)) = \sum_{t=1}^{T} |e_i(n)| \tag{17}$$

To evaluate whether a prediction model (e.g. model A) has performed better than another predictive model (e.g. model B), we test the equal accuracy hypothesis. The null hypothesis is as follows:

$$\mathrm{H}_0\text{: } E[L(e_1(n))]=E[L(e_2(n))] \tag{18}$$

The alternative hypothesis with the assumption that model A is better than the model B is given as:

$$\mathrm{H}_1\text{: } E[L(e_1(n))] \neq E[L(e_2(n))] \tag{19}$$

The DM test is based on the differential costs $d(n)$:

$$d(n) = L(e_1(n)) - L(e_2(n)) \tag{20}$$

Equivalently, the null hypothesis of equal forecast accuracy is shown by H_0: $E[d(n)] = 0$. The sample mean differential cost, $\bar{d}$, is then define as follows:

$$\bar{d} = \frac{1}{N}\sum_{n=1}^{N} d(n) = \frac{1}{N}\sum_{n=1}^{N} [L(e_1(n)) - L(e_2(n))] \tag{21}$$

Given that in large samples, the sample mean differential cost, $\bar{d}$ is approximately normally distributed with mean π and variance $(2\pi\widehat{f_d}(0))/N$, The DM test statistic for testing the null hypothesis of equal forecast accuracy is as follows:

$$DM = \frac{\bar{d}}{\sqrt{(2\pi\widehat{f_d}(0))/N}} \xrightarrow{d} N(0,1) \tag{22}$$

where $2\pi\widehat{f_d}(0)$ is a asymptotically consistent estimator of the asymptotic variance of $\sqrt{N}\bar{d}$. Note that because the sample of differential costs *d(n)* are serially correlated for $h > 1$, the variance is used in the DM test statistic. Since the DM statistics converges towards a normal distribution, the null hypothesis can be rejected at the 5% level if $|DM| > 1.96$; According to Fig. 1, this condition is related to the zone A and zone C. Otherwise, if $|DM| \le 1.96$, the null hypothesis H0 cannot be rejected, and this case is related to zone B in Fig. 6.

Table 3
MSE values and processing times for the proposed EGT-Cluster as well as that of the K-means [61], SOM [62] and NG [63].

Price time series dataset	Evaluation criteria	Clustering methods			
		K-means	NG	SOM	EGT-Cluster
NYISO.	MSE	0.0024	0.0028	0.0023	0.0020
CAPITL	Time(s)	0.5153	4.5273	1.0445	0.5014
NYISO.	MSE	0.0028	0.0028	0.0024	0.0022
CENTRL	Time(s)	0.6948	4.4571	0.7124	0.5620
NYISO.	MSE	0.0018	0.0021	0.0018	0.0017
DUNWOD	Time(s)	1.0078	4.4491	1.0479	0.582
NYISO.	MSE	0.0073	0.0074	0.0072	0.0063
GENESE	Time(s)	0.4012	4.4628	0.9568	0.6014

3. Case studies

The MATLAB tool on a PC with Intel Core i7-6700T, 2.8 GHz, and 8GB RAM processor is used for training, testing and forecasting the electricity price time series data.

Several price series data from different markets are used to evaluate the accuracy of the proposed clustering method (EGT-Cluster) and its comparison with the K-means, original SOM and NG algorithms.

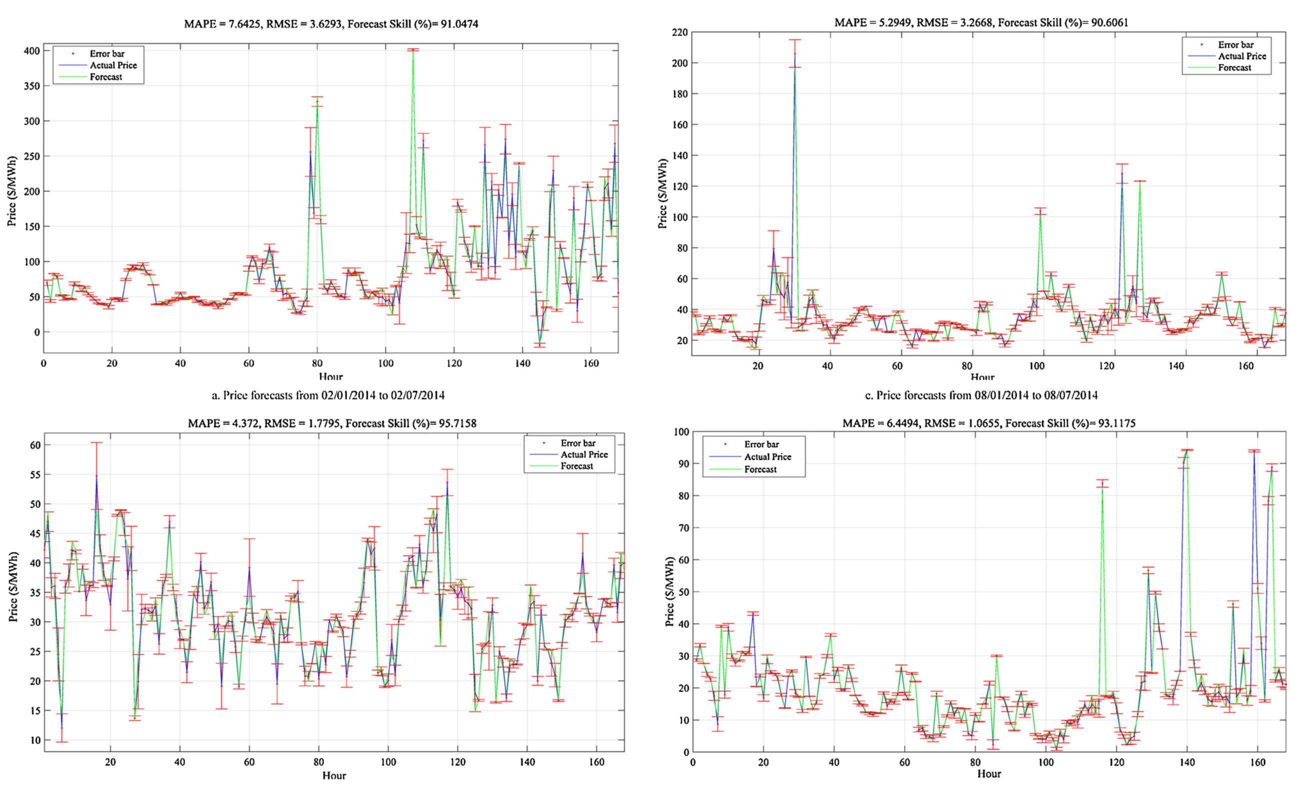

Fig. 7. (a) Price forecasts from 02/01/2014 to 02/07/2014. (b) Price forecasts from 05/01/2014 to 05/07/2014. (c) Price forecasts from 08/01/2014 to 08/07/2014. (d) Price forecasts from 10/01/2014 to 10/07/2014. The price forecasts along with error bars for four different weeks include: a) Winter 2014, for Capitol zone, b) Spring, for Central zone, c) Summer, for Dunwod zone and d) fall 2014 in Dunwod zone.

Detailed information for NYISO market zones are shown in Fig. 7.

Table 3 provides the MSE values and processing times for the proposed EGT-Cluster as well as that of the K-means [66], SOM [67] and NG [68]. The benchmark methods of the table are configured using the parameters provided in the associated reference. The processing time is the time required for the convergence of the algorithms.

The same number of clusters is used for all the algorithms in order to provide a fair comparison. The objective function is the MSE values which are calculated for different clustering methods with the same cluster number. This is a standard approach which has been widely used in literature [69–72].

The MSE values for the proposed EGT-Cluster are the lowest which demonstrates the enhanced accuracy of the proposed clustering algorithm as compared to the other clustering techniques. The EGT-Cluster clusters the NYISO. CAPITL data 16.7%, 28.6%, and 13% more accurate than K-means, NG, and SOM, respectively. As for clustering the NYISO. CENTRL data, the EGT-Cluster outperforms the K-means, NG, and SOM by 21.4%, 21.4%, and 8.3%, respectively. The clustering accuracy of the EGT-Cluster for NYISO. DUNWOD data is 5.5%, 19%, and 5.5% better than that of the K-means, NG, and SOM methods. Lastly, for the NYISO. GENESE data, the MSE value for the proposed EGT-Cluster is 13.7%, 14.9%, and 12.5% less than that of the K-means, NG, and SOM, respectively. The proposed EGT-Cluster algorithm has a faster processing time compared to the SOM and NG and competes with K-means algorithm.

The proposed hybrid forecasting method is then utilized to predict the hourly electricity price. The NYISO (New York Independent System Operator) market price data from 2008 to 2014 is used to evaluate the performance of the proposed forecasting algorithm. The NYISO data are available at Refs. [73,74]. BRNN with eight layers predict the price for 1–24 h ahead. The hidden and output layers use Tansing and Purline functions, respectively. BRNN inputs include 7 electricity price time-series data, 1 harmonic analyzed price data and 1 load data. The structure of the network consists of 8 hidden and one output neurons.

Table 4 provides the RMSE, MAPE and Forecast skill results to evaluate the performance of the BRNN method with different clustering techniques including the proposed EGT-Cluster, K-means, SOM and NG clustering algorithm. The best results are highlighted based on market data from 1 to 5. The results indicate that the accuracy performance of the BRNN with the proposed EGT-Cluster algorithm is better than that of the BRNN forecasting with other clustering algorithms.

Table 5 shows the hourly forecast and actual prices for February 1, 2015. The average of the forecast results for the 24 h is also calculated and provided in this table.

Table 6 provides the accuracy results for the proposed hybrid forecasting method. The results are the average errors for four seasons of 2014.

Fig. 7 shows the price forecasts along with error bars for four different weeks in winter (a), spring (b), summer (c) and fall 2014 (d) for Capitol, Central, Dunwod and Genese zones, respectively.

Fig. 8 shows the electricity price forecasting results for different inputs into BRNN with and without EGT-Cluster clustering. The main objective is to evaluate the performance of the proposed clustering method and effect of different data input and preprocessing on the forecast accuracy. The results are the average MAPE for year 2014. The inputs are the electricity price, electricity price and load, electricity price pre-processed by HANTS, and electricity price pre-processed by the HANTS as well as the load data. The results demonstrate the improved accuracies for the forecasts with EGT-Cluster clustering. It is

Table 5
Forecast results for 02/01/2015.

Hour	Actual price ($/MWh)	Price forecast ($/MWh) for the proposed method	Number of clusters	MAPE (%)	
				Proposed method	Persistence
1	51.25	52.6355	8	2.7034	23.9805
2	55.18	54.0647	6	2.0213	7.1221
3	44.03	47.2674	9	7.3527	25.3236
4	108.72	102.0716	3	6.1151	59.5015
5	41.51	44.5624	7	7.3533	161.9128
6	81.49	80.5286	8	1.1798	49.0612
7	46.99	51.6405	6	9.8967	73.4199
8	172.65	168.5397	9	2.3807	72.7831
9	62.97	62.0973	3	1.3859	174.1782
10	49.81	50.5278	6	1.4411	26.4204
11	34.09	33.8261	3	0.7743	46.1132
12	35.1	35.2905	9	0.5428	2.8775
13	62.88	61.4138	7	2.3318	44.1794
14	47.87	46.6849	6	2.4757	31.3558
15	43.69	43.6723	5	0.0405	9.5674
16	21.13	22.6839	8	7.3539	106.7676
17	22.1	24.8927	6	12.6367	4.3891
18	23.81	21.8567	3	8.2039	7.1819
19	25.83	27.9299	11	8.1297	7.8204
20	29.76	27.1337	7	8.8249	13.2056
21	30.49	28.0276	9	8.0759	2.3942
22	65.25	65.025	5	0.3448	53.272
23	133.19	129.9268	7	2.45	51.0098
24	63.54	62.2757	4	1.9898	109.616
Ave.	–	–	–	4.4169	48.4772

Table 6
The accuracy results of the proposed forecasting method for 2014.

Evaluation criteria	Winter 2014	Spring 2014	Summer 2014	Fall 2014	Average
MAPE (%)	2.17	4.47	5.91	4.06	4.14
RMSE ($/MWh)	1.01	2.89	3.01	1.52	2.13
Forecast skill (%)	96.8	91.17	88.25	93.73	91.84

Table 4
The Comparative analysis of the BRNN method with different clustering techniques including the proposed EGT-Cluster, K-means, SOM and NG clustering algorithm based on the RMSE, MAPE and Forecast skill. The forecasting results with highest accuracy for each market are highlighted in green.

Method	Evaluation Criteria	Market 1	Market 2	Market 3	Market 4	Market 5	Avg.
BRNN + K-means	MAPE (%)	7.7832	9.667	7.2283	7.6732	10.0621	**8.48276**
	RMSE ($/MWh)	5.2085	5.2675	5.0369	4.7884	4.7093	**5.00212**
	Forecast Skill (%)	89.5099	89.7206	89.7133	90.1709	89.1547	**89.65388**
BRNN + SOM	MAPE (%)	7.5072	9.5706	6.186	7.973	11.1768	**8.48272**
	RMSE ($/MWh)	5.2911	5.0875	4.0342	5.1896	5.0937	**4.93922**
	Forecast Skill (%)	90.387	90.0507	91.5359	89.437	88.9456	**90.07124**
BRNN + NG	MAPE (%)	8.4973	7.1892	6.4891	8.1086	10.7938	**8.2156**
	RMSE ($/MWh)	4.933	5.3449	4.4425	5.2924	4.8015	**4.96286**
	Forecast Skill (%)	91.0592	90.3403	90.7218	89.2411	89.0143	**90.07534**
BRNN + Proposed EGT-Cluster	MAPE (%)	8.1711	6.6489	6.2071	7.3934	9.1975	**7.4836**
	RMSE ($/MWh)	4.4066	5.1623	4.1606	4.4505	3.8629	**4.40858**
	Forecast Skill (%)	91.4801	90.9271	91.5079	90.425	91.0004	**91.0681**

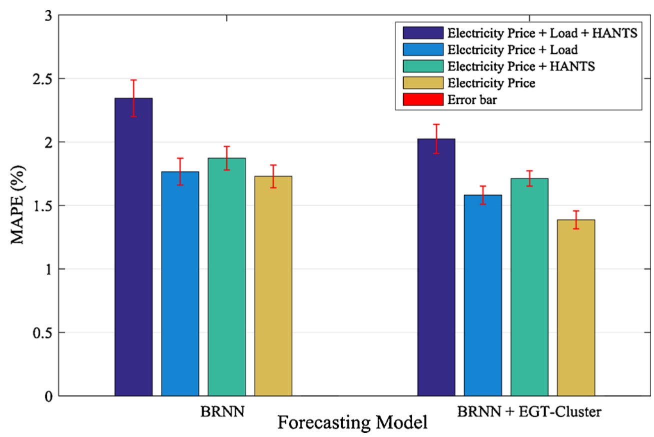

Fig. 8. Root mean squared error (RMSE) for price forecasting with different preprocessing methods. The error bars indicate ± one standard error of the mean.

Table 7
The MAPE (%) results for the proposed hybrid forecasting method against the state-of-the-art forecasting techniques.

Method	Evaluation criteria	Winter	Spring	Summer	Fall	Avg.
ARIMA [7]	MAPE (%)	6.32	6.36	13.39	13.78	9.96
Mixed-model [75]	MAPE (%)	6.15	4.46	14.90	11.68	9.30
NN [10]	MAPE (%)	5.23	5.36	11.40	13.65	8.91
Wavelet-ARIMA [14]	MAPE (%)	4.78	5.69	10.70	11.27	8.11
WNN [12]	MAPE (%)	5.15	4.34	10.89	11.83	8.05
FNN [11]	MAPE (%)	4.62	5.30	9.84	10.32	7.52
HIS [76]	MAPE (%)	6.06	7.07	7.47	7.30	6.97
AWNN [77]	MAPE (%)	3.43	4.67	9.64	9.29	6.75
NNWT [16]	MAPE (%)	3.61	4.22	9.50	9.28	6.65
WNF [17]	MAPE (%)	3.38	4.01	9.47	9.27	6.53
Wavelet-ARIMA-RBFN [18]	MAPE (%)	4.27	4.58	6.76	7.35	5.74
CNEA [78]	MAPE (%)	4.88	4.65	5.79	5.96	5.32
WPA [20]	MAPE (%)	3.37	3.91	6.50	6.51	5.07
Elman network [79]	MAPE (%)	4.11	4.37	9.09	8.66	6.56
Local GP [80]	MAPE (%)	2.75	3.44	5.61	5.80	4.40
Local IVM [80]	MAPE (%)	2.48	3.10	4.90	5.20	3.92
Proposed model	MAPE (%)	2.31	2.83	4.16	4.97	3.56

also shown that the hybrid approach of using the proposed EGT-Cluster method along with the HANTS methods provides the lowest forecasting error.

Fig. 9 provides the electricity price forecasting results and computing times for different number of neurons in the hidden layer of BRNN. As shown, the forecast accuracy increases for few neuron numbers higher than 8. However, the computational time also increases which prolongs the forecasting process. Therefore, number 8 is selected as the neuron number that provides the most appropriate trade-off between the accuracy and computational time. Table 7 provides a comparison of the accuracy results for the proposed hybrid forecasting method with those of the state-of-the-art forecasting techniques including ARIMA[7], Mixed-model [75], NN [8], Wavelet-ARIMA [11], WNN [10], FNN [11], HIS [62], AWNN [77], NNWT [16], WNF [17], Wavelet-ARIMA-RBFN [18], CNEA [64], combination of wavelet transform and particle swarm optimization as well as adaptive-network-based fuzzy inference system (called WPA) [20], Elman network method [79], local GP [80] and local IVM [80].

These methods are all configured using the parameters provided in the associated reference. The results are the weekly average errors for four weeks of 2002. The winter week is from February 18 to February 24; the spring week includes May 20–May 26; the summer week is August 19–25 and the fall week is from November 18 to November 24. The results demonstrate that the proposed model outperforms the state-of-the-art forecasting methods. It should be noted that the results reported in Table 7 were extracted from other publications and that the cited studies relied on the exact same data and same use of those data. Table 8 provides a comparative evaluation of these forecasting methods as well as those of the state-of-the-art forecasting techniques evaluated for five different Spanish electricity markets. The use of other clustering methods, as well as other effective inputs such as wavelet in our proposed forecasting algorithm is evaluated in this table to provide an in-depth analysis. The forecasting results with highest accuracy for each market are highlighted in green. The results demonstrate that the proposed forecasting method, on average, produces better results than other methods. It is also concluded that the proposed EGT-Cluster algorithm has the best performance among all clustering methods. The use of wavelet as an input result in more accurate forecasts for two markets (Markets 1 and 4). The use of HANTS method produces

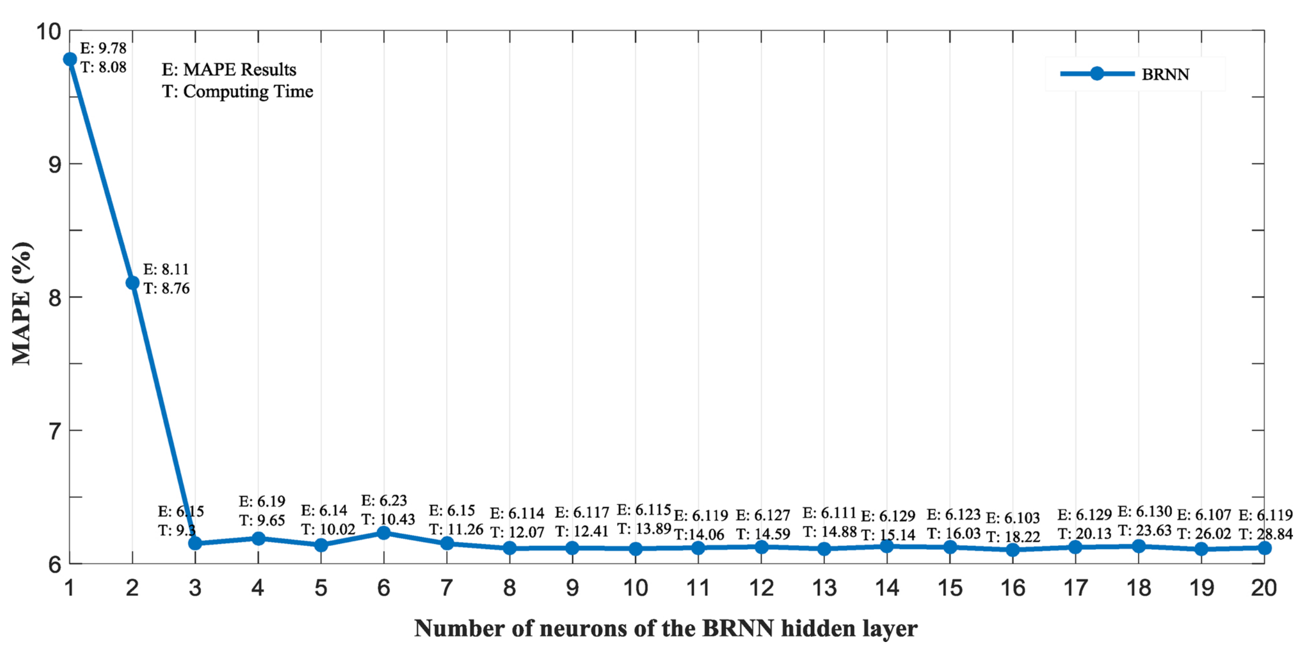

Fig. 9. Price forecasting results and computing times for different number of neurons in the hidden layer of BRNN.

Table 8
Comparative analysis of the proposed hybrid method and several state-of-the-art forecasting approaches as well as BRNN for five different Spanish electricity markets. The forecasting results with the highest accuracy for each market are highlighted in green.

Method	Evaluation Criteria	Market 1	Market 2	Market 3	Market 4	Market 5	Avg.
ARIMA [7]	MAPE (%)	16.8114	39.531	18.2714	14.7472	28.199	**23.512**
	RMSE ($/MWh)	11.8713	25.4299	12.0862	12.5715	20.3001	**16.451**
	Forecast Skill (%)	78.0539	60.7005	78.7531	77.0134	58.7912	**70.662**
Mixed-model [75]	MAPE (%)	12.3268	28.5287	14.9920	11.4804	29.0473	**19.275**
	RMSE ($/MWh)	11.0487	19.2604	10.2327	8.9701	18.4601	**13.594**
	Forecast Skill (%)	80.6538	61.4042	80.3185	82.4188	61.9695	**73.352**
NN [10]	MAPE (%)	11.8708	24.8988	15.3673	10.9018	18.3165	**16.271**
	RMSE ($/MWh)	10.4697	13.8385	9.2026	8.7654	11.8448	**10.824**
	Forecast Skill (%)	82.417	74.8859	82.7715	84.1708	81.1153	**81.072**
Wavelet-ARIMA [14]	MAPE (%)	11.4255	14.7892	8.7296	9.2139	15.2269	**11.877**
	RMSE ($/MWh)	8.2004	10.072	6.2123	7.0953	7.2239	**7.7607**
	Forecast Skill (%)	84.9488	81.1124	88.9805	87.6066	87.2525	**85.980**
WNN [12]	MAPE (%)	9.7954	13.7382	9.5901	9.8259	14.7966	**11.549**
	RMSE ($/MWh)	7.9603	9.866	7.6481	7.452	8.8745	**8.3601**
	Forecast Skill (%)	86.5142	82.9887	86.3576	87.2358	84.2656	**85.472**
FNN [11]	MAPE (%)	9.1262	13.6498	9.3466	9.5965	13.386	**11.0210**
	RMSE ($/MWh)	6.3942	9.2857	6.7753	6.9126	7.2022	**7.314**
	Forecast Skill (%)	88.7917	82.9019	87.897	87.7096	87.6133	**86.982**
HIS [76]	MAPE (%)	9.7631	10.9259	9.6694	11.6659	11.3674	**10.678**
	RMSE ($/MWh)	7.2079	6.1364	7.8092	9.0268	6.0439	**7.244**
	Forecast Skill (%)	87.2553	88.9979	86.268	82.3349	89.5723	**86.88**
AWNN [77]	MAPE (%)	10.2069	11.0211	8.9642	10.0612	12.5155	**10.55**
	RMSE ($/MWh)	8.1991	6.367	6.2038	7.3449	6.7463	**6.972**
	Forecast Skill (%)	86.0131	88.5884	89.1813	86.0114	87.7375	**87.506**
NNWT [16]	MAPE (%)	9.7334	9.6798	8.1674	9.8051	11.3853	**9.754**
	RMSE ($/MWh)	6.5183	5.5541	6.156	7.2148	6.3452	**6.357**
	Forecast Skill (%)	88.5221	90.056	89.7161	86.7392	88.6034	**88.727**
WNF [17]	MAPE (%)	9.3163	10.0314	8.2529	8.9835	11.6038	**9.6375**
	RMSE ($/MWh)	6.3729	5.9937	6.2781	6.0429	6.5831	**6.254**
	Forecast Skill (%)	88.8928	89.4181	88.6865	89.1549	88.0018	**88.83**
Wavelet-ARIMA-RBFN [18]	MAPE (%)	8.2213	9.1729	8.165	8.3564	11.2554	**9.034**
	RMSE ($/MWh)	5.7279	5.6231	5.9809	5.9262	5.9428	**5.840**
	Forecast Skill (%)	89.9199	90.2473	89.9172	89.7478	89.6931	**89.905**
CNEA [78]	MAPE (%)	8.8909	9.0438	8.0754	8.1283	11.0149	**9.0306**
	RMSE ($/MWh)	5.9923	5.112	5.6777	5.7422	5.3988	**5.584**
	Forecast Skill (%)	89.4771	90.8082	90.8005	90.0924	90.0781	**90.251**
WPA [20]	MAPE (%)	10.1498	11.1787	7.0069	8.058	11.5158	**9.5818**
	RMSE ($/MWh)	6.7786	6.7604	5.2125	5.4768	6.0972	**6.0651**
	Forecast Skill (%)	87.9992	87.1695	91.0037	90.3809	89.1744	**89.145**
Elman network [79]	MAPE (%)	9.1485	9.5562	6.8936	7.8067	11.164	**8.9138**
	RMSE ($/MWh)	7.511	6.0699	4.9371	5.0958	5.4295	**5.8086**
	Forecast Skill (%)	88.0543	88.9588	91.8704	90.857	89.9334	**89.934**
Local GP [80]	MAPE (%)	8.5137	9.7351	6.7596	8.1761	10.90826	**8.8185**
	RMSE ($/MWh)	5.5004	6.3755	4.3219	5.3946	5.247	**5.3678**
	Forecast Skill (%)	90.343	87.7554	92.3601	90.307	90.1908	**90.191**
Local IVM [80]	MAPE (%)	8.4696	9.7916	6.5167	8.524	10.7704	**8.8144**
	RMSE ($/MWh)	5.4947	6.1698	4.1955	6.4691	5.0225	**5.4703**
	Forecast Skill (%)	90.4972	88.4833	92.7877	88.483	90.7283	**90.195**
BRNN	MAPE (%)	8.248	8.7163	6.4433	7.5368	10.636	**8.3160**
	RMSE ($/MWh)	5.4649	4.8777	4.1546	4.8049	4.7064	**4.8017**
	Forecast Skill (%)	90.5255	91.2118	92.6007	90.1422	91.1569	**91.127**
Proposed Method + Wavelet Analysis	MAPE (%)	5.7897	6.1394	5.8634	7.0906	9.4563	**6.8678**
	RMSE ($/MWh)	3.7122	4.2913	3.8803	3.9963	4.6484	**4.1057**
	Forecast Skill (%)	93.5544	92.1499	93.2075	92.7343	92.0623	**92.741**
Proposed forecasting + K-means	MAPE (%)	7.235	9.1794	6.9691	7.4091	9.8972	**8.1379**
	RMSE ($/MWh)	4.923	5.1296	4.8828	4.5686	4.5149	**4.8037**
	Forecast Skill (%)	91.4894	90.7969	91.3596	91.2383	91.4912	**91.275**
Proposed forecasting + SOM	MAPE (%)	6.8471	9.0369	5.0696	7.5939	10.5802	**7.8255**
	RMSE ($/MWh)	4.8001	4.7698	3.831	4.6833	4.6835	**4.5535**
	Forecast Skill (%)	91.6409	91.3777	93.086	91.0153	91.1986	**91.663**
Proposed forecasting + NG	MAPE (%)	8.1409	6.6356	5.961	7.5173	10.5491	**7.7607**
	RMSE ($/MWh)	4.5583	4.6550	4.0595	4.7934	4.5757	**4.5283**
	Forecast Skill (%)	91.9662	91.9234	92.923	90.9551	91.2	**91.7935**
Proposed forecasting + HANTS + Proposed EGT-Cluster	MAPE (%)	6.1271	5.5226	4.8759	7.1292	8.0374	**6.3384**
	RMSE ($/MWh)	3.9835	3.6975	3.3582	4.0699	4.2408	**3.8699**
	Forecast Skill (%)	93.0237	94.8648	94.7351	92.2743	95.1904	**94.0176**

competitive results for these two markets, and outperforms the wavelet analysis for three markets (Markets 2, 3, and 5). In addition, HANTS method provides relatively better performance based on the average forecasting results. Finally, Table 9 compares the accuracy and processing time of the proposed forecasting method with those of the BRNN model where the training data are directly given to the network without any pre-processing as well as five different sampling-based approaches. In these methods, instead of employing clustering, several

Table 9
Comparative analysis of the state-of-the-art forecasting methods and the proposed hybrid method.

Method	Evaluation criteria	Market 6	Market 7	Market 8	Market 9	Market 10	Avg.
BRNN	MAPE (%)	10.3781	7.3194	9.0481	16.4639	7.348	10.1115
	RMSE ($/MWh)	6.5351	18.1578	5.0652	10.3508	5.3837	9.0985
	Forecast skill (%)	89.2061	90.8887	90.9633	82.6748	90.5955	88.8656
	Processing time (S)	17.163	13.3901	16.7729	13.917	15.0554	15.2596
Proposed forecasting method	MAPE (%)	6.5113	6.6968	7.5605	11.8928	6.4369	7.21966
	RMSE ($/MWh)	4.128	13.0061	4.3638	6.6124	4.057	6.4334
	Forecast skill (%)	93.1687	93.0462	92.8221	89.0997	92.813	92.1899
	Processing time (S)	6.1298	5.963	6.6302	6.0127	6.731	6.29334
Random subsample-based proposed forecasting method: Approach 1	MAPE (%)	9.227	6.9168	7.302	14.1488	34.6878	14.4564
	RMSE ($/MWh)	5.719	16.8717	4.0789	9.7274	21.4991	11.5792
	Forecast skill (%)	90.4997	92.0205	93.1147	86.8372	64.5724	85.4089
	Processing time (S)	5.0674	4.1649	5.3546	5.6688	4.1996	4.89106
Random subsample-based proposed forecasting method: Approach 2	MAPE (%)	8.9654	7.0611	12.075	14.8744	7.0498	10.0051
	RMSE ($/MWh)	5.3006	17.4039	10.5131	10.3752	5.1914	9.75684
	Forecast skill (%)	91.8023	90.9802	82.4172	84.4451	91.0083	88.1306
	Processing time (S)	7.9162	7.1794	8.7314	8.4639	9.0475	8.26768
Random subsample-based proposed forecasting method: Approach 3	MAPE (%)	9.1997	7.391	10.9411	15.7495	7.103	10.0768
	RMSE ($/MWh)	5.8097	18.1606	8.6344	9.7513	5.2791	9.52702
	Forecast skill (%)	90.6253	89.673	84.2429	83.7655	90.9434	87.8500
	Processing time (S)	10.2718	9.6394	11.4666	10.602	11.8225	10.7604
Random subsample-based proposed forecasting method: Approach 4	MAPE (%)	9.3286	8.0797	9.7757	16.1795	7.0398	10.0806
	RMSE ($/MWh)	6.1327	19.3801	8.7892	9.9612	5.1903	9.8907
	Forecast skill (%)	89.3664	88.9087	84.4648	83.1525	91.0715	87.39278
	Processing time (S)	13.6415	15.7198	13.433	12.343	14.0659	13.84064
Random subsample-based proposed forecasting method: Approach 5	MAPE (%)	10.1832	7.7892	8.0343	16.1431	7.8344	9.99684
	RMSE ($/MWh)	6.3964	17.2501	6.5147	9.8093	5.2978	9.05366
	Forecast skill (%)	89.4476	91.6918	86.5496	83.5232	90.8735	88.41714
	Processing time (S)	15.0535	16.6012	14.573	13.5491	16.5375	15.26286

different sampling methods have been used. In Approach 1, 10% of the whole training data was randomly sampled whereas Approaches 2–5 that randomly sample 20%, 30%, 40% and 50% of the training data, respectively. These approaches provide the selected samples for the neural network rather the whole training data. The results are for five different NYISO's electricity markets. The results demonstrate that the accuracy performance of the proposed forecasting method is better than the BRNN model without any pre-processing as well as the five random sampling based methods. In addition, the processing time of the proposed method is significantly lower than the BRNN method. Overall, the performance of the sampling based approaches is better than the BRNN method. This is due to the proper functioning of these methods in withstanding the noise and outlier data.

In this part, the forecasting performance of the three models is compared by the DM test. Using the classical version of the DM test demonstrated in Section 2. H, the forecasting comparison of every two forecasting models is summarized in Table 10. The zero hypotheses, H_0: $E[L(e_1(n))] = E[L(e_2(n))]$ means that the observed differences between the performances of two forecasting models are not significant, while the alternative hypothesis, H_1: $E[L(e_1(n))] \neq E[L(e_2(n))]$ means that the observed difference between the performances of two forecasting models is significant. Here, four models are evaluated: Model A: Proposed forecasting method, Model B: BRNN, Model C: Proposed Method + Wavelet Analysis and Model D: Elman network [79].

From Table 10, the comparison of model A with models B–D are concluded as follows:

Overall, according to the DM test based on the absolute-error and the squared-error loss, the proposed method (Model A) has better performance than models B–D. Excluding the DM-AE results in Market 1 as well as the DM-AE and DM-SE results in Market 4 where the observed difference between the proposed method and model B is insignificant, the proposed method has shown better performance than

Table 10
The DM test results based on MAE (DM-AE) and RMSE (DM-SE) for difference forecasting approaches. The green highlight is related to the best performance of the first model and the yellow highlight is related to the best performance of the second model in each peer-to-peer comparison.

Dataset	Evaluation Criteria	Method					
		A vs B	A vs C	A vs D	B vs C	B vs D	C vs D
Market 1	DM-AE	-2.1243	1.7603	-1.1459	1.7921	0.4093	0.7053
	DM-SE	-2.7029	2.252	-1.8406	1.1648	-0.1403	-2.8078
Market 2	DM-AE	-2.937	-1.7509	-2.3026	2.4809	-0.0762	-2.1965
	DM-SE	-3.0077	-1.2066	-2.0378	1.4812	-1.9652	-3.1126
Market 3	DM-AE	-2.8756	0.5878	-2.4335	2.8127	-3.3106	-1.118
	DM-SE	-2.08	-0.2599	-1.9942	2.5003	-0.8631	-1.1116
Market 4	DM-AE	-1.0398	-2.1923	-2.6121	2.2202	-1.9085	-0.5688
	DM-SE	-1.5277	-2.8447	-2.1069	0.3846	-1.3328	-2.8949
Market 5	DM-AE	-2.6462	-3.0058	-1.6964	1.9116	-2.1575	-1.8843
	DM-SE	-2.0104	-2.7102	-2.4114	1.4368	-1.8103	-1.4156

model B. Model C reaches better results than the proposed method in Market 1, but the observed difference between the proposed method and model C is not significant in Markets 2 and 3. For Markets 4 and 5, the proposed method performs better than model C. Comparing the DM-AE and DM-SE results in Market 1 as well as the DM-AE results in Market 5 demonstrates a comparable performance between the proposed model and model D. However, in other cases, the proposed method has shown better performance than model D.

4. Discussions

In general, based on the calculated error indexes including RMSE, MAPE and Forecast skill, the proposed method showed better forecasting results than the other methods. The results indicated that the BRNN, the Proposed Method + Wavelet Analysis and the Elman network [79] model have better results as compared to other methods. In addition, in comparison to the proposed method, related approach sometimes have the same results, and even in specific cases, more accurate than our proposed method. However, the proposed method provided better results in the majority of cases.

In order to provide a rigorous statistical evaluation of the proposed method, the Diebold–Mariano (DM) test results based on MAE (DM-AE) and RMSE (DM-SE) was used. Based on the DM test results, the proposed method had a similar performance as these methods in some cases. In one case, the hybridization of the proposed method + wavelet Analysis provides better performance, but in most cases, the proposed method was more accurate than the other three methods. A remarkable point in the results of Table 10 is the proximity of the results of the proposed method + wavelet to the general proposed method. This indicates the positive effect of the wave in the proposed method. It is concluded that providing an improved approach to HANTS in accordance with the characteristics and behaviors of electricity price time series can further improve the results of the proposed method.

5. Conclusion

A hybrid electricity price forecasting is proposed that consists of three stages: clustering; pre-processing; and training. The clustering stage includes an EGT-Cluster method with a new algorithm to enable the non-winning neurons to participate in the learning phase and improve the clustering efficiency and accuracy. A combination of HANTS method and time series analysis are used in the pre-processing stage to provide the most appropriate inputs for BRNN learning. Bayesian approach is then used to train the BRNN and forecast the electricity price. The performance of the proposed clustering algorithm and developed forecasting method is evaluated using different electricity market data. Our results demonstrate the enhanced efficiency of the proposed clustering algorithm as compared to the other clustering techniques. This is due to the more competitive game provided by the proposed strategies that resolves the major problem of the existing clustering techniques where the weight vectors of non-winning neurons are far from the input patterns without having any chance to contribute in the learning phase. Accuracy of the proposed forecasting method is compared with that of the existing state-of-the-art forecasting techniques. The comparison demonstrates an improvement in the forecast accuracy. In this paper, a hybrid method was proposed to predict electricity price time series data. Due to the increased complexity of hybrid methods by increasing the data volume, the proposed algorithm may not be compatible with the data scalability and this drawback should be considered for future work. In the future, we plan to extend our proposed method to include stochastic analysis and develop a probabilistic electricity price-forecasting framework with a lower time complexity.

References

[1] U.D.I. Helman, B.F. Hobbs, R.P. O'Neill, Chapter 5 – the design of US wholesale energy and ancillary service auction markets: theory and practice, in: F.P. Sioshansi (Ed.), Competitive Electricity Markets, Elsevier, Oxford, 2008, pp. 179–243.
[2] S. Singaravel, J. Suykens, P. Geyer, Deep-learning neural-network architectures and methods: using component-based models in building-design energy prediction, Adv. Eng. Inform. 38 (2018) 81–90.
[3] N. Amjady, F. Keynia, Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique, Energy Convers. Manage. 50 (2009) 2976–2982.
[4] A.M. Gonzalez, A.M. San Roque, J.G. Gonzalez, Modeling and forecasting electricity prices with input/output hidden Markov models, IEEE Trans. Power Syst. 20 (February (1)) (2005) 13–24.
[5] J.C. Cuaresma, J. Hlouskova, S. Kossmeier, M. Obersteiner, Forecasting electricity spot-prices using linear univariate time-series models, Appl. Energy 77 (January (1)) (2004) 87–106.
[6] O.B. Fosso, A. Gjelsvik, A. Haugstad, B. Mo, I. Wangensteen, Generation scheduling in a deregulated system: the Norwegian case, IEEE Trans. Power Syst. 14 (February (1)) (1999) 75–81.
[7] J. Contreras, R. Espínola, F.J. Nogales, A.J. Conejo, ARIMA models to predict next-day electricity prices, IEEE Trans. Power Syst. 18 (August (3)) (2003) 1014–1020.
[8] J. Garcia, M. van Akkeren, J.B.C. Garcia, A Garch forecasting model to predict day-ahead electricity prices, IEEE Trans. Power Syst. 20 (May (2)) (2005) 867–874.
[9] J. Nogales, Forecasting next-day electricity prices by time series models, IEEE Trans. Power Syst. 17 (May (2)) (2002) 342–348.
[10] J.P.S. Catalão, S.J.P.S. Mariano, V.M.F. Mendes, L.A.F.M. Ferreira, Short-term electricity prices forecasting in a competitive market: a neural network approach, Elect. Power Syst. Res. 77 (August (10)) (2007) 1297–1304.
[11] N. Amjady, Day-ahead price forecasting of electricity markets by a new fuzzy neural network, IEEE Trans. Power Syst. 21 (May (2)) (2006) 887–896.
[12] A.T. Lora, J.M.R. Santos, A.G. Expósito, J.L.M. Ramos, J.C.R. Santos, Electricity market price forecasting based on weighted nearest neighbors techniques, IEEE Trans. Power Syst. 22 (August (3)) (2007) 1294–1301.
[13] S. Bahrami, M. Parniani, Game theoretic based charging strategy for plug-in hybrid electric vehicles, IEEE Trans. Smart Grid 5 (September (5)) (2014) 2368–2375.
[14] Q. Cai, D. Zhang, W. Zheng, S.C.H. Leung, A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression, Knowl. Based Syst. 74 (2015) 61–68.
[15] Z. Yang, L. Ce, L. Lian, Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods, Appl. Energy 190 (2017) 291–305.
[16] I.P. Panapakidis, A.S. Dagoumas, Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model, Energy 118 (2017) 231–245.
[17] J.P.S. Catalão, H.M.I. Pousinho, V.M.F. Mendes, Short-term electricity prices forecasting in a competitive market by a hybrid intelligent approach, Energy Convers. Manage. 52 (February (2)) (2011) 1061–1065.
[18] M. Shafie-Khah, M.P. Moghaddam, M.K. Sheikh-El-Eslami, Price forecasting of day-ahead electricity markets using a hybrid forecast method, Energy Convers. Manage. 52 (May (5)) (2011) 2165–2169.
[19] M.H. Amini, A. Kargarian, O. Karabasoglu, ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation, Electr. Power Syst. Res. 140 (2016) 378–390.
[20] J.P.S. Catalão, H.M.I. Pousinho, V.M.F. Mendes, Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting, IEEE Trans. Power Syst. 26 (February (1)) (2011) 137–144.
[21] L. Wu, M. Shahidehpour, A hybrid model for day-ahead price forecasting, IEEE Trans. Power Syst. 25 (August (3)) (2010) 1519–1530.
[22] M.H. Amini, B. Nabi, M. Haghifam, Load management using multi-agent systems in smart distribution network, 2013 IEEE Power & Energy Society General Meeting, Vancouver, BC, 2013, pp. 1–5.
[23] O. Abedinia, N. Amjady, M. Shafie-khah, J.P.S. Catalão, Electricity price forecast using Combinatorial Neural Network trained by a new stochastic search method, Energy Convers. Manage. 105 (2015) 642–654.
[24] C.H. Jin, G. Pok, Y. Lee, H.-W. Park, K.D. Kim, U. Yun, et al., A SOM clustering pattern sequence-based next symbol prediction method for day-ahead direct electricity load and price forecasting, Energy Convers. Manage. 90 (2015) 84–92.
[25] H. Shayeghi, A. Ghasemi, M. Moradzadeh, M. Nooshyar, Simultaneous day-ahead forecasting of electricity price and load in smart grids, Energy Convers. Manage. 95 (2015) 371–384.
[26] F. Azevedo, Z.A. Vale, Forecasting electricity prices with historical statistical information using neural networks and clustering techniques, Proc. IEEE PES Power Syst. Conf. Exp. November, 2006, pp. 44–50.
[27] L. Xie, H. Zheng, L. Zhang, Electricity price forecasting by clustering-LSSVM, Proc. 8th IEEE Int. Power Eng. Conf. (2007) 697–702.
[28] K. Sokhanvar, A. Karimpour, N. Pariz, Electricity price forecasting using a clustering approach, Proc. 2nd IEEE Int. Conf. Power Ener. December, 2008, pp. 1302–1305.
[29] F. MartínezÁlvarez, A. Troncoso, J.C. Riquelme, J.M. Riquelme, Partitioning-clustering techniques applied to the electricity price time series, Proc. 8th Int. Conf. Intell. Data Eng. Auto. Learn. December, 2007, pp. 990–999.
[30] F. Feijoo, W. Silva, T.K. Das, A computationally efficient electricity price forecasting model for real time energy markets, Energy Convers. Manage. 113 (2016) 27–35.
[31] F. MartínezÁlvarez, A. Troncoso, J.C. Riquelme, J.M. Riquelme, Discovering patterns in electricity price using clustering techniques, Proc. Int. Conf. Renew. Ener. Power Qual. March, 2007.
[32] H. Zhou, X.H. Wu, X.G. Li, An ANFIS model of electricity price forecasting based on subtractive clustering, Proc. IEEE PES General Meeting, July, 2011, pp. 1–5.
[33] E. Bompard, G. Ciwei, R. Napoli, F. Torelli, Dynamic price forecast in a competitive

electricity market, IET Gener. Transm. Distrib. 1 (5) (2007) 776–783.
[34] S. Fan, C. Mao, L. Chen, Next-day electricity-price forecasting using a hybrid network, IET Gener. Transm. Distrib. 1 (January (1)) (2007) 176–182.
[35] J. Herbert, J. Yao, A game-theoretic approach to competitive learning in self-organizing maps, in: L. Wang, K. Chen, Y.S. Ong (Eds.), Advances in Natural Computation, Springer, 2005, pp. 129–138.
[36] A. Neme, S. Hernandez, O. Neme, L. Hernandez, Self-organizing maps with non-cooperative strategies (SOM-NC), in: J.C. Príncipe, R. Miikkulainen (Eds.), Advances in Self-Organizing Maps, Springer, 2005, pp. 200–208.
[37] D. Wang, H. Luo, O. Grunder, Y. Lin, H. Guo, Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm, Appl. Energy 190 (2017) 390–407.
[38] Z. Yang, L. Ce, L. Lian, Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods, Appl. Energy 190 (2017) 291–305.
[39] G. Spanakis, G. Weiss, AMSOM: Adaptive Moving Self-organizing Map for Clustering and Visualization, eprint arXiv:1605.06047, 2016.
[40] T. Kohonen, Self-Organizing Maps, Springer Series in Information Sciences, third ed., 30, Springer, Berlin, Heidelberg, New York, 1995, 1997, 2001.
[41] L. Pavel, Game Theory for Control of Optical Networks, Springer, 2012.
[42] T. Martinetz, K. Schulten, A neural-gas network learns topologies, in: T. Kohonen, K. Makisara, O. Simula, J. Kangas (Eds.), Proceedings of International Conference on Artificial Neural Networks (ICANN-91), vol. 1, 1991, pp. 397–402.
[43] Self-organizing maps, Springer-Verlag New York, Inc., 1997.
[44] T.M. Martinetz, K.J. Schultcn, A "neural-gas" network learns topologies, in: T. Kohonen, K. Makisara, O. Simula, J. Kangas (Eds.), Artificial Neural Networks, North-Holland. Amsterdam, 1991, pp. 397–402.
[45] D.L. Davies, D.W. Bouldin, Cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. 1 (2) (1979) 95–104.
[46] D.J. Ketchen Jr., C.L. Shook, The application of cluster analysis in Strategic Management Research: an analysis and critique, Strateg. Manage. J. 17 (6) (1996) 441–458.
[47] J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65.
[48] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Commun. Stat. 3 (1) (1974) 1–27.
[49] R. Tibshirani, G. Walther, T. Hastie, Estimating the number of data clusters via the Gap statistic, J. R. Stat. Soc. B 63 (2000) 411–423.
[50] Q. Zhao, M. Xu, P. Fränti, Sum-of-squares based cluster validity index and significance analysis, Adaptive and Natural Computing Algorithms: 9th International Conference, ICANNGA 2009, Kuopio, Finland, April 23–25, 2009 Revised Selected Papers. M. Kolehmainen, P. Toivanen and B. Beliczynski. Berlin, Heidelberg, Springer Berlin Heidelberg: 313-322, 2009.
[51] M. Fortunato, C. Blundell, O. Vinyals. Bayesian recurrent neural networks. arXiv preprint arXiv:1704.02798, 2017.
[52] Y. Wu, M. Schuster, Z. Chen, Q. Le, Quoc, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al. Google's neural machine translation system: bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144, 2016.
[53] D. Amodei, R. Anubhai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, J. Chen, M. Chrzanowski, A. Coates, G. Diamos, et al. Deep speech 2: end-to-end speech recognition in english and mandarin. arXiv preprint arXiv:1512.02595, 2015.
[54] R.M. Neal, Bayesian Learning for Neural Networks, Springer-Verlag, New York, 1996.
[55] Z. Lipton, J. Berkowitz, C. Elkan, A Critical Review of Recurrent Neural Networks for Sequence Learning, eprint arXiv:1506.00019, 2015.
[56] C. Blundell, J. Cornebise, K. Kavukcuoglu, D. Wierstra, Weight Uncertainty in Neural Networks, eprint arXiv:1505.05424, 2015.
[57] C.J. Tucker, Red and photographic infrared linear combinations for monitoring vegetation, Remote Sens. Environ. 8 (2) (1979) 127–150.
[58] C.O. Justice, J.R.G. Townshend, E.F. Vermote, E. Masuoka, R.E. Wolfe, N. Saleous, et al., An overview of MODIS land data processing and product status, Remote Sens. Environ. 83 (2002) 3–15.
[59] H. Shang, L. Jia, M. Menenti, Analyzing the inundation pattern of the Poyang Lake flood plain by passive microwave data, J. Hydrometeorol. 16 (2) (2015) 652–667.
[60] J. Zhou, L. Jia, M. Menenti, Reconstruction of global MODIS NDVI time series: performance of Harmonic ANalysis of Time Series (HANTS), Remote Sens. Environ. 163 (2015) 217–228.
[61] W. Verhoef, Application of Harmonic Analysis of NDVI Time Series (HANTS). in: S. Azzali, M. Menenti (Eds.), Fourier Analysis of Temporal NDVI in the Southern African and American Continents, DLO Winand Staring Centre, Wageningen, The Netherlands, Report 108, 1996, pp. 19–24.
[62] G.J. Roerink, M. Menenti, W. Verhoef, Reconstructing cloudfree NDVI composites using Fourier analysis of time series, Int. J. Remote Sens. 21 (9) (2000) 1911–1917.
[63] F.X. Diebold, Element of Forecasting, 4th ed., Thomson South-Western, Cincinnati, OH, USA, 2007, pp. 257–287.
[64] F.X. Diebold, R. Mariano, Comparing predictive accuracy, J. Bus. Econ. Stat. 13 (1995) 253–265.
[65] University of Illinois at Urtanu-Chiimpaign. Persistence Forecasting: Today Equals Tomorrow. Retrieved on 2007-02 16.
[66] S.P. Lloyd, Least squares quantization in PCM, IEEE Trans. Inf. Theory 28 (2) (1982) 129–136.
[67] T. Kohonen, The self-organizing map, Proc. IEEE 78 (9) (1990) 1464–1480.
[68] T.M. Martinetz, S.G. Berkovich, K.J. Schulten, 'Neural-gas' network for vector quantization and its application to time-series prediction, IEEE Trans. Neural Netw. 4 (4) (1993) 558–569.
[69] R.C. Camargos, M. do Carmo Nicoletti, Three case studies using agglomerative clustering, Paper Presented at the Intelligent Systems Design and Applications, Cham, 2017.
[70] M. Meilă, H. Heckerman, An experimental comparison of model-based clustering methods, Mach. Learn. 42 (1) (2001) 9–29.
[71] R.C. Camargos, P.R. Nietto, M. do Carmo Nicoletti, Agglomerative and divisive approaches to unsupervised learning in gestalt clusters, Paper Presented at the Intelligent Systems Design and Applications, Cham, 2017.
[72] M. Abdel-Mottaleb, S. Krischnamachari, N. Mankovich, Performance evaluation of clustering algorithms for scalable image retrieval, IEEE Computer Society Workshop on Empirical Evaluation of Computer Vision Algorithms (1998).
[73] [Online]. Available: http://www.nyiso.com/public/markets_operations.
[74] [Online]. Available: http://www.energyonline.com.
[75] C. García-Martos, J. Rodríguez, M.J. Sánchez, Mixed models for short-run forecasting of electricity prices: application for the Spanish market, IEEE Trans. Power Syst. 22 (May (2)) (2007) 544–552.
[76] N. Amjady, H. Hemmati, Day-ahead price forecasting of electricity markets by a hybrid intelligent system, Eur. Trans. Electr. Power. 19 (January (1)) (2009) 89–102.
[77] N.M. Pindoriya, S.N. Singh, S.K. Singh, An adaptive wavelet neural network-based energy price forecasting in electricity markets, IEEE Trans. Power Syst. 23 (August (3)) (2008) 1423–1432.
[78] N. Amjady, F. Keynia, Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm, IEEE Trans. Power Syst. 24 (February (1)) (2009) 306–318.
[79] S. Anbazhagan, N. Kumarappan, Day-ahead deregulated electricity market price forecasting using recurrent neural network, IEEE Syst. J. 7 (December (7)) (2013) 866–872.
[80] E.E. Elattar, Day-ahead price forecasting of electricity markets based on local informative vector machine, Gener. Transm. Distrib. IET 7 (10) (2013) 1063–1071.

M. Ghayekhloo received her B.Sc. degree in Computer Engineering from Mazandaran University of Science and Technology, Babol, Iran, and the M.Sc. degree from Science and Research Branch, Islamic Azad University, Qazvin, Iran in 2011 and 2014, respectively. She is currently a Ph.D. student in Computer Engineering at Amirkabir University of Technology, Tehran, Iran. Her research interests include optimization algorithms, artificial neural networks, computational intelligence and their applications in power systems.

R. Azimi received his B.Sc. degree in Software Engineering from Mehrastan University, Guilan, Iran, in 2011 and the M.Sc. degree from Science and Research Branch, Islamic Azad University, Qazvin, Iran in 2014. His research interests include distributed computing, clustering, artificial intelligence and their applications in power systems.

M. Ghofrani received his B.Sc. degree in Electrical Engineering from Amirkabir University of Technology, Tehran, Iran in 2005, the M.Sc. degree from University of Tehran, Tehran, Iran, in 2008, and the Ph.D. degree from the University of Nevada, Reno, in 2014. He is currently an Assistant Professor at the School of Science, Technology, Engineering and Mathematics, University of Washington, Bothell. His research interests include power systems operation and planning, renewable energy systems, smart grids, electric vehicles and electricity market.

M. B. Menhaj received his Ph.D. degree from the School of Electrical and Computer Engineering at Ohio State University (OSU) in 1992. After completing one year with OSU as a post-doctoral fellow, in 1993, he joined Amirkabir University of Technology, Tehran, Iran, where he is currently a Full Professor. From December 2000 to August 2003, he was with the School of Electrical and Computer Engineering and Department of Computer Science at OSU as a visiting faculty member and research scholar. His research interests include theory of computational intelligence, learning automation, adaptive filtering, image processing, pattern recognition, and their applications in control and power systems.

E. Shekari received his B.Sc. degree in Computer Engineering from Ferdowsi University, Mashhad, Iran, and the M.Sc. degree in Decision Science and Knowledge Engineering from University of Economics Science, Tehran, Iran in 2011 and 2014, respectively. His research interests include computational intelligence, machine learning and data mining.