Electricity Price Forecast Using Combinatorial Neural Network Trained by a New Stochastic Search Method

O. Abedinia[a], N. Amjady[a], M. Shafie-khah[b], J.P.S. Catalão[1 b,c,d]

[a] *Department of Electrical Engineering, Semnan University, Semnan, Iran*
[b] *University of Beira Interior, R. Fonte do Lameiro, 6201-001 Covilha, Portugal*
[c] *INESC-ID, R. Alves Redol, 9, 1000-029 Lisbon, Portugal*
[d] *IST, University of Lisbon, Av. Rovisco Pais, 1, 1049-001 Lisbon, Portugal*

Abstract: Electricity price forecast is key information for successful operation of electricity market participants. However, the time series of electricity price has nonlinear, non-stationary and volatile behaviour and so its forecast method should have high learning capability to extract the complex input/output mapping function of electricity price. In this paper, a Combinatorial Neural Network (CNN) based forecasting engine is proposed to predict the future values of price data. The CNN-based forecasting engine is equipped with a new training mechanism for optimizing the weights of the CNN. This training mechanism is based on an efficient stochastic search method, which is a modified version of chemical reaction optimization algorithm, giving high learning ability to the CNN. The proposed price forecast strategy is tested on the real-world electricity markets of Pennsylvania-New Jersey-Maryland (PJM) and mainland Spain and its obtained results are extensively compared with the results obtained from several other forecast methods. These comparisons illustrate effectiveness of the proposed strategy.

Keywords–Price Forecast, Modified Chemical Reaction Optimization, Feature Selection

1. Introduction

With the restructuring of electric power industry, electricity price has become the focus of all activities in the power market [1]. In order to plan efficient operation and economical capital expansion of deregulated electricity markets, the Generating Companies (GENCO) should be able to adjust their bidding strategies to achieve the maximum benefit and hedge themselves against the financial risks. Also, the consumers should have a plan to purchase electricity from the pool market, or use their own generation facilities when high prices occur. For these purposes, the key information for both GENCOs and consumers is electricity price forecast. However, electricity has distinct characteristics compared to other commodities. Electricity generation and consumption must be continuously matched, while electricity storage is still expensive. Moreover, transmission constraints may limit electricity exchange between power systems. Thus, the time series of electricity price can exhibit a major volatility and the application of forecasting techniques used in other markets can lead to large errors in electricity price forecasting [2].

[1] Corresponding Author;

The key role of electricity price forecasting for the market participants and its complexity have been the driver of many recent research works. In [2], combination of Adaptive Wavelet Neural Network (AWNN) and Generalized Auto-Regressive Conditional Heteroscedastic (GARCH) time series is presented for electricity price prediction. In [3], a day-ahead forecasting approach for prediction of electricity price and load considering their correlation has been presented. Their approach is composed of wavelet packet transform, generalized mutual information and least square support vector machine. Probabilistic electricity price forecasting to provide prediction intervals for electricity price has been presented in [4]. The model of [4] consists of active learning method and variational heteroscedastic Gaussian process. A bat-neural network multi-agent system is proposed in [5] for forecasting stock price. Combination of Support Vector Regression (SVR) and Auto-Regressive Integrated Moving Average (ARIMA) models is introduced in [6] for electricity price forecasting. In [7], an Enhanced Probability Neural Network (EPNN), composed of Probability Neural Network (PNN) and Orthogonal Experimental Design (OED), is presented as a price forecasting system for electricity market participants. The OED is used to smooth parameters in the EPNN to improve the forecasting error. A hybrid electricity price forecast method, composed of WT, ARIMA and Radial Basis Function Neural Network (RBFNN), is proposed in [8]. Also, particle swarm optimization is used to optimize the RBFNN structure in [8]. In [9], it has been discussed that electricity price prediction by only one model is a hard task. Then, the authors of [9] propose a forecasting model that detaches high volatility and daily seasonality of electricity price based on empirical mode decomposition, seasonal adjustment and ARIMA. In [10], one dimensional Discrete Cosine Transform (DCT) input featured Feed-Forward Neural Network (FFNN) is proposed for electricity price forecasting. The authors of [10] have tested their price forecast approach on the mainland Spain and New York electricity markets.

An electricity price forecast method composed of wavelet transform and a hybrid prediction technique is proposed in [11]. The hybrid technique includes a set of cascaded forecasters such that each forecaster is a combination of NN and an evolutionary algorithm. In [12], a feature selection algorithm composed of Mutual Information (MI) and Information Gain (IG) criteria is introduced to refine the input features of electricity price spike forecast process. Also, combination of probabilistic NN and hybrid neuro-evolutionary system is presented as the price spike forecast method in [12]. Combination of modified version of relief algorithm for feature selection and a hybrid NN for electricity price forecast is presented in [13]. In [14], a prediction method for next-day electricity prices based on ARIMA methodology is proposed. In [15], a three-layered feed-forward NN is presented for prediction of next-week electricity prices over mainland Spain and California electricity markets. The NN forecast method of [15] is trained by the Levenberg-Marquardt algorithm. An electricity price forecast method based on Fuzzy NN (FNN) is proposed in [16]. In this reference, combination of fuzzy logic and an efficient learning algorithm is presented for capturing the non-stationary behaviour and outliers of the price time series. A price forecast technique based on WT and ARIMA model is presented in [17]. In this method, the price data is first decomposed by WT and then the ARIMA model predicts the future values of each WT component. An electricity

price prediction method based on the combination of NN and Similar Days (SD) techniques is proposed in [18]. A review of different electricity price forecast methods can be found in [19-21].

Despite the performed research works in the area, more accurate price forecast methods are still demanded. This paper focuses on electricity price prediction for the next day, which is the price forecast process frequently used by the electricity market participants to prepare their bids. The new contributions of this research work can be summarized as follows:

1) A new stochastic search method, which is a modified version of chemical reaction optimization algorithm, is proposed. It is shown that this method can benefit from uni-individual and multi-individual as well as local and global search operators arranged in a coordinated manner. Distinct features of the proposed stochastic search method enhances its exploration capability for finding optimum solutions in complex search spaces.
2) The proposed modified chemical reaction optimization algorithm is adapted as the training mechanism of a Combinatorial Neural Network (CNN) based forecasting engine. The CNN is composed of different Neural Networks (NNs) arranged in a cascaded manner. The proposed training mechanism can search the solution space of the CNN's training problem in different directions in parallel and avoid being trapped in local optima as well as overfitting problem. Moreover, each NN of the CNN is trained to improve the price forecast of the previous NN.

The remaining parts of the paper are organized as follows. In the second section, the proposed CNN-based forecasting engine and its training mechanism based on the modified chemical reaction optimization algorithm are introduced. The numerical results obtained from the proposed price forecast strategy for real-world electricity markets are presented in section three and compared with the results of several other recently published price forecast methods. Finally, section four concludes the paper.

2. Proposed Combinatorial Neural Network (CNN) based Forecasting Engine

Architecture of the suggested forecasting engine, i.e. CNN, is shown in Fig. 1(a), which is composed of cascaded NNs. For simplicity only three NNs are shown in Fig. 1, but in general CNN can have any number of cascaded NNs. Each cascaded NN has multi-layer perceptron structure, which is an efficient structure for estimating neural networks. The first NN, i.e. NN_1 in Fig. 1(a), is fed by the inputs selected by the double-filter feature selection method of [12], which is based on the information theoretic criteria of MI and IG. This feature selection method, denoted by MI-IG, has two cascaded filters to filter out irrelevant candidate features (i.e. the inputs that have low mutual information with the output variable) and redundant candidate features (i.e. the inputs that have high mutual information with the other inputs), respectively. Only the relevant non-redundant candidate inputs, constituting a minimum subset of the most informative features for predicting the output variable, are selected by the feature selection method. As this feature selection method is not the focus of this paper, it is not further discussed here. The interested reader can refer to [12] for details of this technique. The output layer of NN_1 has one neuron. Multi-period forecast, e.g. prediction of electricity price for the next 24 hours, is reached through

recursion, i.e. by feeding the input variables with the output feature of the forecasting engine. For instance, predicted electricity price for the first hour is used as $P(t-1)$ for the price forecast of the second hour if $P(t-1)$ is among the selected candidate inputs of NN_1. The MLP structure of the next NNs of CNN, e.g. NN_2 and NN_3 in Fig. 1(a), is similar to the structure of NN_1, except for an additional input devoted to receive the forecast of the previous NN. In other words, NN_1 is trained to learn the mapping function between the selected inputs by the feature selection method and electricity price of the next hour. Each of the next NNs is trained to learn the mapping function between the selected inputs + price forecast of the previous NN and electricity price of the next hour. In this way, each of the next NNs can employ the forecast of the previous NN, as an input close to the output, to better learn the variations and patterns of the price signal and improve its forecast. The output of the last NN constitutes the price forecast of the CNN.

Fig. 1(b) illustrates the historical data used for training the NNs of the CNN. In this figure, it is assumed that each NN requires 50 days historical data including 50×24=1200 hourly training samples [13,22]. Thus, the whole CNN of Fig. 1(b) requires 3×50=150 days historical data. At first, NN_1 is trained by the historical data from 150 days ago to 101 days ago, denoted by –150 to –101, included in the sliding window 1. After this training phase, NN_1 predicts 24 hourly electricity prices of the day –100. Subsequently, the sliding window 1 proceeds by one day such that NN_1 is trained by the historical data of the days –149 to –100 and predicts the electricity prices of the day –99. This cycle of train-forecast is repeated until NN_1 predicts the electricity prices of the days –100 to –51. These price forecasts plus the selected features are included in the sliding window 2, shown in Fig. 1(b), by which NN_2 is trained and predicts the electricity prices of the day –50. The cycle of train-forecast is repeated for NN_2 until it predicts the electricity prices of the days –50 to –1. By means of the predicted prices and selected features for the days –50 to –1, included in sliding window 3, NN_3 is trained and predicts the electricity prices of the forecast day or day 0 in Fig. 1(b).

To train each NN of the CNN, a new stochastic search technique is proposed in this paper, which optimizes the weights of the NN to learn its associated input/output mapping function. The stochastic search method is a modified version of chemical reaction optimization algorithm. In the following, the chemical reaction optimization algorithm is first introduced briefly. Then, the proposed modified version is presented and its different operators are detailed. Finally, the modified chemical reaction optimization algorithm is adapted as the training mechanism of the CNN-based forecasting engine.

2.1. Chemical Reaction Optimization Algorithm

Chemical Reaction Optimization (CRO) algorithm is a population-based stochastic search technique recently presented by Lam [23]. The underlying idea of CRO is taken from the natural chemical reactions. This algorithm mimics the procedure of high-energy molecules taking part in various types of elementary reactions to make the final products with stable low energy states. Without loss of generality, we focus on a minimization problem to introduce CRO algorithm (here, minimization of the training phase error for the CNN-based forecasting engine).

CRO algorithm consists of two main parts including molecules and elementary reactions. In nature, a molecule is composed of several atoms with different characteristics. In CRO algorithm, the structure of a molecule is considered as a solution in the search space. In nature, change of a molecular structure involves with Potential Energy (*PE*) and Kinetic Energy (*KE*). The *PE* quantifies the molecular structure in terms of energy. In the CRO algorithm, it is modelled as the objective function *OF*(.) of the associated solution, denoted by ω:

$$OF(\omega)=PE_{\omega} \tag{1}$$

Physically, a molecule ω can change to another form ω', if $PE_{\omega}\geq PE_{\omega'}$. Otherwise, if $PE_{\omega}+KE_{\omega}\geq PE_{\omega'}$, this change can still be happened, i.e., *KE* allows the molecules to move to higher potential states. In the mathematical modelling of CRO, *KE* of a molecule (or similarly *KE* of a solution of CRO) characterizes its ability for escaping from local minima [24].

As in any population-based stochastic search technique, the individuals of CRO population are evolved to obtain better solutions, i.e. solutions with lower objective function values for the minimization problem. In nature, molecules are changed through elementary reactions toward more stable states with lower *PE* values. While molecules constitute the individuals of CRO population, the elementary reactions are modelled as the evolutionary operators of the individuals. A chemical reaction process includes a set of elementary reactions caused by the collisions among the molecules or between the molecules and the walls of the container. Four types of the elementary reactions are considered in CRO algorithm including on-wall ineffective collision, decomposition, inter-molecular ineffective collision, and synthesis, which are described below.

1) *On-Wall Ineffective Collision*: In this elementary reaction, a molecule hits the wall of the container and returns back. This reaction is not vigorous and usually there is a small change in the molecular structure ω and its *PE*. This reaction is modelled in CRO algorithm as follows:

$$\omega' = N(\omega) \tag{2}$$

where *N*(.) is a neighbourhood search operator, which returns the new solution ω' in the vicinity of the original one ω. *PE* of the new solutions is obtained through the objective function of the optimization problem as shown in equation (1). If $PE_{\omega}+KE_{\omega}\geq PE_{\omega'}$, the original solution ω and its *PE* (i.e. PE_{ω}) are replaced by the new solution ω' and its *PE* (i.e. $PE_{\omega'}$). As the molecule hits the wall, some of its *KE* is lost. Thus, *KE* of the new solution is updated as:

$$KE_{\omega'} = (PE_{\omega} - PE_{\omega'} + KE_{\omega}) \times Rand\left[KELossRate,\ 1\right] \tag{3}$$

where $Rand[KELossRate,1]$ is a random number with uniform distribution in the interval $[KELossRate,1]$; $KELossRate \in (0,1)$ is a parameter of CRO limiting maximum percentage of the *KE* loss. Note that $(1–Rand[KELossRate,1])$ represents the fraction of *KE* lost to the environment. Otherwise, if the condition $PE_{\omega}+ KE_{\omega}\geq PE_{\omega'}$ does not hold, the change is prohibited and the molecule retains its original ω, *PE* and *KE*.

2) *Decomposition*: Decomposition happens when a molecule hits the container and decomposes into two or more molecules (two is assumed here). Despite the previous one, this change is vigorous. If the original molecular structure is represented by ω and the resultant ones are indicated by ω'_1 and ω'_2, the decomposition is shown by:

$[\omega'_1,\omega'_2]=D(\omega)$ (4)

This change occurs if equation (5) holds:

$PE_\omega+KE_\omega \geq PE_{\omega'_1}+PE_{\omega'_2}$ (5)

In this case, considering $PE_\omega+KE_\omega-PE_{\omega'_1}-PE_{\omega'_2}\geq 0$, KE of the newly generated molecules are obtained as follows:

$$KE_{\omega'_1}=\left(PE_\omega+KE_\omega-PE_{\omega'_1}-PE_{\omega'_2}\right)\times Rand[0,1] \quad (6)$$

$$KE_{\omega'_2}=\left(PE_\omega+KE_\omega-PE_{\omega'_1}-PE_{\omega'_2}\right)\times\left(1-Rand[0,1]\right) \quad (7)$$

where $Rand[0,1]$ is a random number with uniform distribution in the interval [0,1]. If the condition of equation (5) does not hold, the decomposition is prohibited and the original ω, PE_ω, and KE_ω are retained. If this condition frequently fails, the concept of buffer can be used to enhance its success rate [25].

3) *Intermolecular Ineffective Collision*: This elementary reaction consists of sudden collision between two or more molecules. Energy change of the molecules in intermolecular ineffective collision is similar to that in on-wall ineffective collision, but intermolecular ineffective collision involves more than one molecule (assume two molecules here). Thus, intermolecular ineffective collision is modelled as follows in the CRO algorithm:

$\omega'_1=N(\omega_1)$ (8)

$\omega'_2=N(\omega_2)$ (9)

where $N(.)$ is the neighbourhood search operator of equation (2); ω_1 and ω_2 are the original molecular structures; ω'_1 and ω'_2 are generated molecules through intermolecular ineffective collision. The changes of the intermolecular ineffective collision are accepted if

$PE_{\omega_1}+PE_{\omega_2}+KE_{\omega_1}+KE_{\omega_2}\geq PE_{\omega'_1}+PE_{\omega'_2}$ (10)

In this case, KE of the new individuals ω'_1 and ω'_2 is determined as follows:

$$KE_{\omega'_1}=\left(PE_{\omega_1}+PE_{\omega_2}+KE_{\omega_1}+KE_{\omega_2}-PE_{\omega'_1}-PE_{\omega'_2}\right)\times Rand[0,1] \quad (11)$$

$$KE_{\omega'_1}=\left(PE_{\omega_1}+PE_{\omega_2}+KE_{\omega_1}+KE_{\omega_2}-PE_{\omega'_1}-PE_{\omega'_2}\right)\times\left(1-Rand[0,1]\right) \quad (12)$$

Otherwise, if the condition of equation (10) does not hold, the changes are prohibited and the original molecules ω_1 and ω_2 as well as their PE and KE are restored.

4) *Synthesis*: In this elementary reaction, two molecules ω_1 and ω_2 collide with each other and combine to construct a new molecular structure ω'. In CRO, synthesis is described as follows:

$\omega'=S(\omega_1,\omega_2)$ (13)

The changes of the synthesis are accepted if

$PE_{\omega_1}+PE_{\omega_2}+KE_{\omega_1}+KE_{\omega_2}\geq PE_{\omega'}$ (14)

In this case KE of the new solution ω' is set as

$$KE_{\omega'} = PE_{\omega_1} + PE_{\omega_2} + KE_{\omega_1} + KE_{\omega_2} - PE_{\omega'} \quad (15)$$

Otherwise, if the condition of equation (14) does not hold, the changes of the synthesis are not performed and the original solutions ω_1 and ω_2 as well as their *PE* and *KE* are retained. $KE_{\omega'}$ is usually higher than KE_{ω_1} and KE_{ω_2} as $PE_{\omega'}$ is expected to have a value similar to PE_{ω_1} and PE_{ω_2}. Thus, the newly produced individual ω' will have higher ability to escape from local minima.

Based on the above explanations, the reasons of selecting CRO algorithm as the design framework of the training mechanism for the NNs of the CNN can be described as follows:

1) CRO includes both uni-individual search operators including on-wall ineffective collision and decomposition and multi-individual search mechanisms including intermolecular ineffective collision and synthesis. Thus, CRO can employ both the personal information of the individuals and the information generated from their interactions.

2) CRO can simultaneously benefit from local and global search operators. While on-wall ineffective collision and intermolecular ineffective collision implement local searches around the current solutions, decomposition and synthesis apply great changes to the molecules. Thus, the individuals take the chance to jump to different areas of the solution space and so explore the other areas enhancing the exploration capability of the algorithm. Additionally, CRO with the aid of its decomposition and synthesis can implement a population-based stochastic search with variable size. Note that decomposition/synthesis increases/decreases the population size.

3) CRO presents an open optimization framework that provides important degrees of freedom for the users to design their own specific models. While it is indicated that *N*(.), *D*(.) and *S*(.,.) are neighbourhood search, decomposition and synthesis operators, their specific functions are not specified in the original CRO algorithm. Indeed, these are considered as the changeable components of CRO [23-25].

In the modified CRO, which will be presented in the next section, effective operators for the changeable components of CRO are first introduced. Then, a mechanism for coordinating the elementary reactions in the form of an evolution procedure is presented.

2.2. Modified CRO

The changeable components of CRO are designed in the modified CRO as follows:

2.2.1) Neighbourhood search operator. This operator of the modified CRO, inspired from Particle Swarm Optimization (PSO) algorithm, is as below:

$$N(\omega_i) = \omega_i + r_1.(\omega_{i,best} - \omega_i) + r_2.(\omega_{gbest} - \omega_i) \quad (16)$$

where $\omega_{i,best}$ is the best structure of the molecule ω_i obtained so far and ω_{gbest} is the best structure of all molecules obtained so far; r_1 and r_2 are two random numbers in the range of (0,1). Thus, the proposed $N(\omega_i)$ searches around ω_i using two difference vectors measuring the difference between the current position ω_i and its previous best position in the solution

space as well as the difference between ω_i and previous best solution of the whole population. In this way, the proposed $N(\omega_i)$ can benefit from the information content of the past experiences of both ω_i and the other individuals. Finally, it should be noted that although $N(\omega_i)$ in equation (16) is inspired from PSO, it is not the same as the evolution strategy of PSO. While cognitive and social behaviours of individuals are used in PSO to update the velocity vectors, these information contents are directly used here, as the neighbourhood search operator, to update position of the individuals in the solution space. Additionally, to further enhance the search ability of $N(.)$ operator, the random coefficients r_1 and r_2 are changed in each iteration according to chaotic variation patterns based on logistic map as follows:

$$r_1^k = 4 \times r_1^{k-1} \times (1 - r_1^{k-1}) \tag{17}$$

$$r_2^k = 4 \times r_2^{k-1} \times (1 - r_2^{k-1}) \tag{18}$$

where k indicates iteration number of the modified CRO. To initialize the chaotic variation patterns of equations (17) and (18), r_1^0 and r_2^0 are randomly selected within the interval (0,1) such that $r_1^0 \notin \{0.25, 0.5, 0.75\}$ and $r_2^0 \notin \{0.25, 0.5, 0.75\}$. For every individual ω_i, the initial seeds r_1^0 and r_2^0 are separately initialized and then the next r_1 and r_2 values in the chain are generated according to equations (17) and (18). In this way, these coefficients can adopt more diverse values and so the proposed $N(.)$ operator can better cover the neighbourhood area around every individual.

2.2.2) Decomposition Operator. Suppose that the decomposition operator is shown by:

$$D(\omega_i) = [\omega'_{i1}, \omega'_{i2}] \tag{19}$$

where the argument of $D(.)$ function, i.e. the individual ω_i, is a vector of decision variables of the optimization problem by the number of ND. The proposed $D(\omega_i)$ randomly selects a position j within the vector of ω_i ($0<j<ND$) and decomposes ω_i to two sub-vectors including the decision variables of ω_i from 1 to j and from j+1 to ND. The first and second sub-vectors are assigned to the resultant molecules ω'_{i1} and ω'_{i2}, respectively. The remaining part of ω'_{i1} (i.e. its decision variables from j+1 to ND) and the remaining part of ω'_{i2} (i.e. its decision variables from 1 to j) are randomly selected within their allowable ranges, similar to the initialization of the molecules. The performance of the proposed decomposition operator is graphically shown in Fig. 2. Thus, the proposed $D(\omega_i)$ operator can benefit from both the information content of ω_i and randomization technique. Moreover, the decomposition point is randomly and separately selected for each ω_i, which enhances the search diversity of this operator.

2.2.3) Synthesis operator. The proposed synthesis operator, inspired from uniform crossover of Genetic Algorithm (GA), is as follows:

$$\omega'_i = S(\omega_{i1}, \omega_{i2}) \tag{20}$$

This operator randomly selects a position j ($0<j<ND$) and decomposes each of ω_{i1} and ω_{i2} to two parts including the decision variables from 1 to j and from j+1 to ND. Then, the second parts of ω_{i1} and ω_{i2} are swapped as shown in Fig. 2. In this way, two molecules are generated, among which the individual with lower PE value is selected as ω'_i as illustrated in

Fig. 2. Through this operator, two molecules can share their information in a stochastic manner with the aid of generating a better individual.

Another important aspect of the proposed modified CRO is coordination of its elementary reactions (including the changeable components designed as mentioned above) in the form of a comprehensive evolution strategy to produce better individuals as the algorithm proceeds.

Here, the roulette wheel mechanism is proposed for this purpose. In this mechanism, the whole probability domain, i.e. the interval [0,1], is divided to four segments such that each segment indicates chance of selection of one elementary reaction as shown in Fig. 3. If a priori knowledge about the search characteristics required for solving the optimization problem is available, higher probability can be considered for the associated elementary reactions. For instance, if we know the approximate position of the optimal solution such that higher local search ability is required, larger segments can be given to the *On-Wall Ineffective Collision* and *Intermolecular Ineffective Collision* so that these elementary reactions with neighbourhood search behaviour are selected more. Otherwise, equal segments, i.e. equal chance of selection, can be considered for all elementary reactions. Then, a random number with uniform distribution in the interval [0,1] is generated. The segment that the random number falls in it indicates the selected elementary reaction as shown in Fig. 3. In each iteration, the proposed roulette wheel mechanism is separately implemented for every molecule so that the elementary reaction that should be performed on it is selected. If an elementary reaction with two arguments (i.e. *Intermolecular Ineffective Collision* or *Synthesis*) is selected, the other argument is randomly chosen from the remaining population. Pseudo code of the proposed modified CRO is shown in Fig. 4. In this pseudo code various modules of the modified CRO are specified.

2.3. Application of the proposed modified CRO as the training mechanism of the CNN

To effectively train each NN of CNN by the modified CRO, its generalization capability should be carefully monitored along the training process to avoid overfitting, which is a serious problem for NN training [21,26]. Generalization is a measure of how well the NN performs on the actual problem once training is complete [27]. When overfitting occurs in training phase, the NN training error continues to decrease and it seems that the training process progresses, while indeed the generalization capability of the NN degrades and it loses its prediction ability for unseen forecast samples. However, as the forecast error is not available in the training phase, error of validation samples or validation error is used as an approximation of it to measure the generalization performance of the NN along its training process. Validation samples are a subset of training samples, which are not used for the optimization of the NN's weights and retained unseen for the NN. Thus, error of validation samples can give an estimate of the NN error for unseen forecast samples (here, electricity prices of the next day). Validation error is a better tool for measuring generalization capability of a NN and coping with overfitting problem, compared to training error [27]. To enhance the effectiveness of the validation error, validation samples should be as similar as possible to forecast samples so that the validation error can give a true estimate of the prediction error.

Considering short-run trend characteristic of electricity price time series [11,12], the samples of the day before the forecast day (i.e. the closest samples to the forecast samples) are taken into account as the validation samples in this paper. Some other choices such as the same day in the previous week may be considered as the validation set instead of the previous day, but the short-run trend characteristic of electricity price time series is usually stronger than its weekly periodicity [21]. Thus, the 50 days historical data of each NN of CNN, shown in Fig. 1(b), are divided to a training subset including the first 49 days or 49×24=1176 hourly samples used as the training samples and a validation subset including 24 hourly samples of the last day employed as the validation samples.

Using the produced training and validation samples, the proposed modified CRO can train each NN of the forecasting engine or CNN. To do this, training phase of the NN is modelled as an optimization problem for the modified CRO, in which the objective function is the error of the constructed training samples or training error. Also, decision variables of the optimization problem are the weights of the NN. At the beginning of the training phase, the modified CRO population is initialized by randomly generating the decision variables (i.e. the weights) of each individual within the range [-1,+1]. In each iteration, the elementary reactions of the modified CRO, based on the proposed changeable components and coordination mechanism, so changes the individuals that the training error decreases. At the end of the iteration, validation error of the NN is also evaluated. Whenever the validation error starts to increase, the generalization performance of the NN begins to degrade (i.e. overfitting problem starts to affect the training process), and thus the NN's training phase should be terminated at the corresponding iteration. The weights of the NN are set according to the decision variables of the best individual of the modified CRO in this iteration. In this way, every NN of CNN can be trained. Afterward, the trained CNN can predict the future values of the electricity price for the next day.

3. Numerical Results

The suggested price forecast strategy is tested on the well-known Ackley's benchmark function and the real-world day-ahead electricity markets of PJM in US and mainland Spain in Europe. The results obtained from these numerical experiments are presented in the following sections, respectively.

3.1 Comprehensive Example: Ackley's Benchmark Function

Ackley's benchmark is a well-known test function, which is frequently used for evaluating the performance of stochastic search methods. This test function is as below [28]:

$$f(x_1,x_2) = -a.e^{-b\sqrt{\frac{\sum_{i=1}^{2} x_i^2}{2}}} - e^{\frac{\sum_{i=1}^{2}\cos(c.x_i)}{2}} + a + e^1 \qquad (21)$$
$$a = 20,\ b = 0.2,\ c = 2\times\pi,\ -20 \le x_i \le 20,\ i = 1,2$$

The aim of this benchmark is finding the minimum value of $f(x_1,x_2)$ given above. The shape of Ackley's benchmark function is shown in Fig. 5. It is seen that Ackley's benchmark is a multi-modal test function including many local minima. Thus,

obtaining its global minimum, which is zero as shown in Fig. 5, is a complex optimization task. The results obtained from the proposed modified CRO for this test function are shown in Table 1 and compared with the results of five other well-known stochastic search methods including Evolutionary Algorithm (EA) [29], GA [30], Differential Evolution (DE) [31], Ant Colony (ACO) [32] and PSO [33]. These five stochastic search methods have been implemented according to the procedures given in their corresponding references and tested on the Ackley's benchmark function. For the sake of a fair comparison, all methods of Table 1 have 10 individuals and 50 iterations. Additionally, the settings of each method are fine-tuned based on 10 trial runs. Subsequently, the best, average and worst results, i.e. the objective function values, obtained among 30 other trial runs are reported for every method in Table 1. Table 1 shows that the best, average and worst results of the modified CRO are better than the best, average, and worst results of all other methods. Additionally, it is seen that only the modified CRO finds the global optimum of Ackley's benchmark function, i.e. zero, as its best result among all methods of Table 1. These comparisons clearly illustrate the optimization capability of the proposed modified CRO.

Fig. 6 shows the distribution of the 10 individuals in the first and last iterations for the proposed modified CRO (A_1 and B_1), PSO (A_2 and B_2), DE (A_3 and B_3), ACO (A_4 and B_4), GA (A_5 and B_5), and EA (A_6 and B_6). The best result of each method among the 30 trial runs is selected for this figure. The subplots A_1 to A_6 illustrate that all stochastic search methods begin from random initial points, i.e. their 10 individuals are randomly distributed in the solution space in the first iteration. On the other hand, the subplots B_1 to B_6 show that at the end of the search process, the 10 individuals of the proposed modified CRO concentrate on the global optimum solution of this benchmark function (i.e. zero) much better than the 10 individuals of the five other methods. Fig. 6 clearly illustrates higher search ability of the proposed modified CRO for finding global optimum solution. In Fig. 7, convergence graph of the proposed modified CRO is demonstrated. In this figure, the evolution of the best individual of the modified CRO is shown. Fig. 7 clearly shows high convergence rate of the proposed algorithm.

3.2. PJM Market

As PJM electricity market is well recognized in the U.S. and beyond, the proposed price forecast strategy is tested using the real data of day-ahead energy market of PJM. This market is one of the Regional Transmission Organizations, which plays a vital role in the U.S. electric system. Data of PJM electricity market is obtained from [34].

To evaluate the impact of number of cascaded NNs on the performance of the forecasting engine, CNN with different NNs ranging from 1 to 6 is implemented for price forecast of the sample day of November 20, 2006 in PJM electricity market. The obtained validation errors, measured in terms of mean squared error (MSE), and training times are shown in Fig. 8(a) and 8(b), respectively. Fig. 8(a) shows that the validation error first decreases for CNN with 1, 2 and 3 NNs, but then increases for CNN with 4, 5 and 6 NNs. After CNN with 3 NNs, leading to the minimum validation error, overfitting problem with respect to the cascaded NNs occurs such that further NNs cannot learn more the input/output mapping function of the electricity price. Similar results for the number of NNs are obtained for the other test cases of the paper.

Thus, three NNs are considered for the CNN in the next numerical experiments to avoid this kind of overfitting (although the optimum number of NNs may slightly change in different test cases). The CNNs of Fig. 8 begin from the same validation error, i.e. 4.3e-1. Thus, the CNN with 3 NNs, leading to minimum validation error of 4.8e-6, decreases the validation error by about 10^5 times, indicating effectiveness of the training process of the CNN. From Fig. 8, it is seen that the training time of the CNN increases by the number of NNs, but with three NNs, the training time is just 8.21×10^2 s, i.e. 13m and 41s. This training time, measured on a 64-bit windows-based server with 16 GB of RAM and 24 Intel Xenon processors clocking at 3.33 GHz, is completely reasonable within a day-ahead decision making framework.

In the next numerical experiments of this section, the prediction accuracy of the proposed price forecast strategy is evaluated.

For the PJM market, four test weeks corresponding to four seasons of year 2006 are considered to provide representative results for the whole year. The four considered weeks are Feb. 15 to Feb. 21, May 15 to May 21, August 15 to August 21 and Nov. 15 to Nov. 21 (months 2, 5, 8, and 11). Price prediction results of the proposed forecasting engine for these test weeks are presented in Tables 2 and 3 and compared with the results of six other forecast methods including ARIMA time series, three NNs trained by LM (Levenberg–Marquardt), BFGS (Broyden, Fletcher, Goldfarb, Shanno) and BR (Bayesian Regularization) learning algorithms denoted by NN-LM, NN-BFGS and NN-BR, Mutual Information (MI) feature selection plus composite NN (MI + composite NN) [35] and two-stage MI plus composite NN (MI-MI + composite NN) [35]. These six benchmark methods are frequently used price forecast techniques, which their results for this test case are quoted from [35]. In Tables 2 and 3, Weekly Mean Error (WME) and Weekly Peak Error (WPE), as measures of price forecast accuracy and stability, are given, respectively. The last row of Tables 2 and 3 presents the average results of the four test weeks. WME and WPE are defined as follows:

$$WME\ (\%) = \frac{1}{N_w}\sum_{i=1}^{N_w}\frac{\left|P_i^{true} - P_i^{forecast}\right|}{P_i^{true}}\times 100 \tag{22}$$

$$WPE\ (\%) = \underset{1\le i\le N_w}{Max}\left(\frac{\left|P_i^{true} - P_i^{forecast}\right|}{P_i^{true}}\right)\times 100 \tag{23}$$

where P_i^{true} and $P_i^{forecast}$ represent true and forecast value of electricity price of i^{th} hour, respectively; N_W=168, which is number of hours of a week. It is noted that this research work focuses on day-ahead price forecast in which the forecast horizon is one day or 24 hours. At the end of each day, the historical data is updated by the latest available data and then price forecast is performed for the next day. However, to more accurately evaluate price forecast performance of each method, its mean and peak errors for the longer period of one week, including different weekdays and weekend, are presented, which are obtained from seven day-ahead price forecasts. From Tables 2 and 3 it can be seen that not only average WME and WPE, but also WME and WPE of each test week of the proposed strategy are better than those of all other methods, illustrating better price forecast accuracy and stability of the proposed CNN-based forecasting engine.

The proposed CNN-based forecasting engine is compared with eight other price forecast methods including NN [36], naïve predictor [36], ARIMA [36], Wavelet Transform (WL) [36], Transfer Function (TF) [36], Dynamic Regression (DR) [36], MI + composite NN [35] and MI-MI + composite NN [35] in Tables 4 and 5. The results of the eight benchmark methods are quoted from the corresponding references [35,36]. The error criteria of Tables 4 and 5, i.e. e_{week} and $\sigma^2_{e,week}$, are defined as follows:

$$e_{week}(\%) = \frac{1}{N_w}\sum_{i=1}^{N_w}\frac{\left|P_i^{true} - P_i^{forecast}\right|}{P^{ave-true}} \times 100 \tag{24}$$

$$\sigma^2_{e,week} = \frac{1}{N_w}\sum_{i=1}^{N_w}\left[\frac{\left|P_i^{true} - P_i^{forecast}\right|}{P^{ave-true}} - (e_{week})\right]^2 \tag{25}$$

where

$$P^{ave-true} = \frac{1}{N_w}\sum_{i=1}^{N_w} P_i^{true} \tag{26}$$

e_{week} is a weekly error measure similar to WME defined in equation (22). The only difference between these two error criteria is that the weekly average of true prices, denoted by $P^{ave-true}$ and defined in (26), is used in the denominator of e_{week} instead of P_i^{true} to avoid the adverse effect of the prices close to zero [35,36]. Also, $\sigma^2_{e,week}$ in (25) is error variance as a measure of the forecast stability. As e_{week} and $\sigma^2_{e,week}$ are used in [35,36], they are also used here for the sake of a fair comparison. Moreover, the same test periods of [35,36], i.e. the last week of all months of year 2002, in PJM market are also considered for the proposed price forecast strategy. The $\sigma^2_{e,week}$ values of MI + composite NN and MI-MI + composite NN are not presented in their corresponding reference [35] and so cannot be compared in Table 5. Tables 4 and 5 show that the proposed strategy outperforms all other methods. The proposed method has both the lowest e_{week} value and lowest $\sigma^2_{e,week}$ value among all methods of Tables 4 and 5 in all test periods. Additionally, the average results for the 12 test weeks are reported in the last row of these tables, which again illustrate superiority of the CNN. Thus, the proposed strategy has both more accurate and more stable price forecasts than all eight other methods. Real hourly prices, predicted prices by the CNN and its forecast error for the last week of February are shown in Fig. 9 to graphically illustrate detailed price forecast performance of the proposed method for a typical test week. Closeness of the forecast and real curves (except for some minor deviations) as well as low values of the error curve can be observed from this figure.

In Tables 6 and 7, the proposed strategy is compared with five other price forecast methods, including combination of Similar Day and NN techniques (NNSD) [18], combination of Wavelet Transform, Firefly Algorithm and Fuzzy ARTMAP (WT+FF+FA) [37], Hybrid Neuro-Evolutionary System (HNES) [38], AWNN plus GARCH times series (AWNN+GARCH) [2] and Cascaded Neuro-Evolutionary Algorithm (CNEA) [39] on five test days and two test weeks.

Since these test periods are considered in the mentioned references, they are also used here for the proposed price forecast method. For the test days in Tables 6 and 7, the error criteria of e_{day} and $\sigma^2_{e,day}$ are used instead of e_{week} and $\sigma^2_{e,week}$. The error criteria e_{day} and $\sigma^2_{e,day}$ are similar to e_{week} and $\sigma^2_{e,week}$ in equations (24) and (25) except that N_W=168 hours for a test week should be replaced by 24 hours for a test day. The last row of Tables 6 and 7 shows that the average results of the proposed forecasting engine are better than all other methods. Additionally, except for slightly higher e_{day} than AWNN+GARCH in the test day of Feb. 10th (Table 5) and slightly higher $\sigma^2_{e,week}$ than HNES in the test week of February 22-28 (Table 6), all results of the proposed forecasting engine are better than the results of five other methods.

3.3. Spanish Electricity Market

In this numerical experiment, performed on the Spanish electricity market, four weeks corresponding to four seasons of year 2002 (including the fourth week of February, May, August, and November) are considered. Although this test period seems relatively old, it is a very well-known test case and several other recently published price forecast methods have used it, which include ARIMA [14], mixed model [40], MLP [15], wavelet-ARIMA [17], Weighted Nearest Neighbors (WNN) technique [41], Fuzzy Neural Network (FNN) [16], Hybrid Intelligent System (HIS) [42], Adaptive Wavelet Neural Network (AWNN) [43], Neural Network with Wavelet Transform (NNWT) [44], Simple Recurrent Network (SRN) [45], WT with ARIMA and Radial Basis Function NN (RBFN) [8], Cascaded Neuro-Evolutionary Algorithm (CNEA) [39], Modified Relief (MR)-Mutual Information (MI) feature selection and neural network (MR-MI+NN) [46], HNES [38], MI + composite NN [35], Wavelet-PSO-ANFIS (WPA) [47], and MI-MI + composite NN [35]. The forecast accuracy and stability of the proposed forecasting engine, in terms of e_{week} and $\sigma^2_{e,week}$, are compared with these 17 methods in Tables 8 and 9, respectively. The results of the other methods are directly quoted from their corresponding references. For the methods of mixed model, WNN and SRN, only e_{week} values are presented in their references, while $\sigma^2_{e,week}$ values are not given and so cannot be compared in Table 9.

The results of the methods in Tables 8 and 9 are sorted in terms of average e_{week} and average $\sigma^2_{e,week}$, respectively. It can be seen that the proposed forecasting engine has considerably lower average e_{week} and lower average $\sigma^2_{e,week}$ than all other methods, leading to its higher forecast accuracy and stability considering all test weeks. Considering the test weeks individually, there are 68 comparative cases in Table 8:

4 (number of test weeks) × 17 (number of comparative methods) = 68 (comparative cases)

and 56 comparative cases in Table 9:

4 (number of test weeks) × 14 (number of comparative methods) = 56 (comparative cases)

Among the 68 comparative cases of Table 8, e_{week} of the proposed method is slightly higher than e_{week} of MI-MI + composite NN in the fall test week. However, the proposed strategy has lower e_{week} in all remaining 67 comparative cases.

Also, $\sigma^2_{e,week}$ of the proposed method is slightly higher than $\sigma^2_{e,week}$ of NNWT and WPA in the winter test week. However, the proposed CNN-based forecasting engine has lower $\sigma^2_{e,week}$ in all remaining 54 comparative cases of Table 9. These comparisons illustrate that the proposed forecast strategy outperforms the other methods of Tables 8 and 9 based on all test periods. Real hourly prices, predicted prices by the proposed strategy and its forecast error for the winter test week are shown in Fig. 10 to also give a graphical view about the price forecast performance of the proposed method in this test case. It is seen that the forecast curve accurately follows the trend and ramps of the real curve and only small deviations from the real curve are seen in it. Also, consistently low values of the error curve throughout the forecast horizon illustrate good price forecast accuracy and stability of the proposed method.

4. Conclusion

In this paper, an effective CNN-based forecasting engine is proposed to predict the price of electricity markets. The forecasting engine is equipped with a new training mechanism based on modified CRO. The traditional NN learning algorithms usually search the solution space of NN's weights in a specific direction (such as the steepest descent direction in gradient-based learning algorithms) and so may be trapped in local optima. On the other hand, the proposed modified CRO can search the solution space in various directions in parallel with high exploration capability and search diversity avoiding being trapped in local optima as much as possible. Moreover, the modified CRO has both uni-individual and multi-individual as well as local and global search operators, which enhances the chance of finding optimal solution for the weights of the CNN-based forecasting engine. The proposed price prediction strategy is extensively tested on the well-known Ackley's benchmark function and real-world electricity markets of PJM in US and mainland Spain in Europe. Four performance evaluation criteria including WME, WPE, e_{week} and $\sigma^2_{e,week}$ are mathematically described. These performance evaluation criteria are frequently used in the literature to compare different price forecast methods. The obtained results clearly show that the proposed method outperforms all 36 other price forecast methods on both the electricity markets in terms of WME, WPE, e_{week} and $\sigma^2_{e,week}$. Moreover, the proposed modified CRO has been compared with five other methods on the Ackley's benchmark function in terms of three performance evaluation criteria, including best result, average result and worst result for the objective function. The attained results show that the proposed modified CRO outperforms all five other methods in terms of all three evaluation criteria. These exhaustive evaluations and comparisons clearly confirm the validity of the proposed method and its results.

5. Acknowledgements

The work of M. Shafie-khah and J.P.S. Catalão was supported by FEDER funds (European Union) through COMPETE, and by Portuguese funds through FCT, under Projects FCOMP-01-0124-FEDER-020282 (Ref. PTDC/EEA-EEL/118519/2010)

and UID/CEC/50021/2013. Also, the research leading to these results has received funding from the EU Seventh Framework Programme FP7/2007-2013 under grant agreement no. 309048.

6. References

[1]. M. Shahidehpour, H. Yamin, and Z. Li, "Market Operations in Electric Power Systems", New York: Wiley, April 2002.

[2]. L. Wu, M. Shahidehpour, "A Hybrid Model for Day-Ahead Price Forecasting, Power Systems", IEEE Transactions on Power System, vol. 25, no. 3, pp. 1519-1530, August 2010.

[3]. H. Shayeghi, A. Ghasemi, M. Moradzadeh, M. Nooshyar, Simultaneous day-ahead forecasting of electricity price and load in smart grids, Energy Conversion and Management, vol. 95, pp. 371–384, May 2015.

[4]. P. Kou, D. Liang, L. Gao, J. Lou, Probabilistic electricity price forecasting with variational heteroscedastic Gaussian process and active learning, Energy Conversion and Management, vol. 89, pp. 298–308, January 2015.

[5]. R. Hafezia, J. Shahrabi, E. Hadavandi, A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price, Applied Soft Computing, vol. 29, pp. 196–210, April 2015.

[6]. J. Che, J. Wang, Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling, Energy Conversion and Management, vol. 51, no. 10, pp. 1911-1917, October 2010.

[7]. W. M. Lin, H. J. Gowa, M. T. Tsai, Electricity price forecasting using Enhanced Probability Neural Network, Energy Conversion and Management, vol. 51, no. 12, pp. 2707–2714, December 2010.

[8]. M. Shafie-khah, M. Parsa Moghaddam, M.K. Sheikh-El-Eslami, Price forecasting of day-ahead electricity markets using a hybrid forecast method, Energy Conversion and Management, vol. 52, no. 5, pp. 2165-2169, May 2011.

[9]. Y. Dong, J. Wang, H. Jiang, J. Wu, Short-term electricity price forecast based on the improved hybrid model, Energy Conversion and Management, vol. 52, no. 8-9, pp. 2987–2995, August 2011.

[10]. S. Anbazhagan, N. Kumarappan, Day-ahead deregulated electricity market price forecasting using neuralnetwork input featured by DCT, Energy Conversion and Management, vol. 78, pp. 711–719, February 2014.

[11]. N. Amjady and F. Keynia, "Day-ahead Price Forecasting of Electricity Markets by a Mixed Data Model and Hybrid Forecast Method", International Journal of Electrical Power & Energy Systems, vol. 30, no. 9, pp. 533-546, November 2008.

[12]. N. Amjady, F. Keynia, "A New Prediction Strategy for Price Spike Forecasting of Day-Ahead Electricity Markets," Applied Soft Computing, vol. 11, no. 6, pp. 4246-4256, September 2011.

[13]. N. Amjady, A. Daraeepour, F. Keynia, "Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network", IET Generation, Transmission & Distribution, vol. 4, no. 3, pp. 432-444, March 2010.

[14]. J. Contreras, R. Espinola, F.J. Nogales, A.J. Conejo, "ARIMA models to predict next-day electricity prices". IEEE Transaction on Power System, vol. 18, no. 3, pp. 1014-1020, August 2003.

[15]. 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", Electric Power System Research, vol. 21, no. 10, pp. 1297-1304, August 2007.

[16]. N. Amjady, "Day-ahead price forecasting of electricity markets by a new fuzzy neural network", IEEE Transaction on Power System, vol. 21, no. 2, pp. 887-896, May 2006.

[17]. A. J. Conejo, M. A. Plazas, R. Espinola, A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models", IEEE Transaction on Power System, vol. 20, no. 2, pp. 1035-1042, May 2005.

[18]. P. Mandal, T. Senjyu, N. Urasaki, T. Fundabashi, and A. K. Srivastava, "A novel approach to forecast electricity price for PJM using neural network and similar days method", IEEE Transaction on Power System,, vol. 22, no. 4, pp. 2058-2065, November 2007.

[19]. S. Kumar Aggarwal, L. Mohan Saini, A. Kumar, "Electricity price forecasting in deregulated markets: A review and evaluation, " Electrical Power and Energy Systems, vol. 31, no. 1, pp. 13–22, January 2009.

[20]. R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future, International Journal of Forecasting, " vol. 30, pp. 1030–1081, October-December 2014.

[21]. N. Amjady, "Electric Power Systems: Advanced Forecasting Techniques and Optimal Generation Scheduling", Chapter 4, CRC Press, Taylor & Francis, 2012.

[22]. G. J. Tsekouras, N. D. Hatziargyriou, E. N. Dialynas, "An Optimized Adaptive Neural Network for Annual Midterm Energy Forecasting", IEEE Transactions on Power Systems, vol. 21, no. 1, pp. 385-391, February 2006.

[23]. A. Y. S. Lam and V. O. K. Li, "Chemical-reaction-inspired metaheuristic for optimization", IEEE Transaction on Evolutionary Computation, vol. 14, no. 3, pp. 381-399, June 2010.

[24]. J. J. Q. Yu, A. Y. S. Lam, and V. O. K. Li, "Evolutionary artificial neural network based on chemical reaction optimization", in Proc. IEEE CEC, pp. 2083-2090, June 2011.

[25]. A. Y. S. Lam, V. O. K. Li, and J. J. Q. Yu, "Real-Coded Chemical Reaction Optimization", IEEE Transactions on Evolutionary Computation, vol. 16, no. 3, pp. 339-353, June 2012.

[26]. H. Mao, X. J. Zeng, G. Leng, Y. J. Zhai, J. A. Keane, "Short-Term and Midterm Load Forecasting Using a Bilevel Optimization Model", IEEE Transaction on Power System, vol. 24, no. 2, pp. 1080-1090, May 2009.

[27]. D. R. Hush, B. G. Horne, "Progress in supervised Neural Networks", IEEE Signal Processing Magazine, vol. 10, no. 1, pp. 8-39, January 1993.

[28] H. Narasimhan, "Parallel Artificial Bee Colony (PABC) Algorithm," World Congress on Nature & Biologically Inspired Computing. NaBIC 2009, pp. 306-311, December 2009.

[29] G. S. Hornby and J. B. Pollack, "Creating high-level components with a generative representation for body-brain evolution," Artificial Life, vol. 8, no. 3, pp. 223-246, August 2002.

[30] X. J. Zeng, J. Tao, P. Zhang, H. Pan and Y. Y. Wang, "Reactive Power Optimization of Wind Farm based on Improved Genetic Algorithm," 2nd International Conference on Advances in Energy Engineering, vol. 14, pp. 1362–1367, August 2011.

[31] K. V. Price, S. Rainer, and L. Jouni, "Differential Evolution: A Practical Approach to Global Optimization," Springer-Verlag, Berlin, December 2005.

[32] R. S. Parpinelli, H. S. Lopes and A. A. Freitas, "Data mining with an ant colony optimization algorithm," IEEE Transaction on Evolutionary Computation, vol. 6, no. 4, pp. 321-332, August 2002.

[33] O. Abedinia, N. Amjady, A. Ghasemi, Z. Hejrat, "Solution of economic load dispatch problem via hybrid particle swarm optimization with time-varying acceleration coefficients and bacteria foraging algorithm techniques," European Transactions on Electrical Power, vol. 23, no. 8, pp. 1504-1522, November 2013.

[34]. PJM Electricity Market Data. [Online]. Available: http://www.pjm.com/.

[35]. F. Keynia, "A new feature selection algorithm and composite neural network for electricity price forecasting", Engineering Applications of Artificial Intelligence, vol. 25, no. 8, pp. 1687-1697, December 2012.

[36]. A. J. Conejo, J. Contreras, R. Espinola, M. A. Plazas, "Forecasting electricity prices for a day-ahead pool-based electric energy market", International Journal of Forecast, vol. 21, no. 3, pp. 435-462, July–September 2005.

[37] P. Mandal, AU. Haque, J. Meng, A. Srivastava, R. Martinez, A novel hybrid approach using wavelet, firefly algorithm, and fuzzy ARTMAP for day-ahead electricity price forecasting. IEEE Trans Power Syst, vol. 28, no. 2, pp. 1041–51, 2013.

[38] N. Amjady, F. Keynia, Application of a new hybrid neuro-evolutionary system for day-ahead price forecasting of electricity markets. Appl Soft Comput, vol. 10, no. 3, pp. 784–92, 2010.

[39]. N. Amjady, F. Keynia, "Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm", IEEE Transaction on Power System, vol. 24, no. 1, pp. 306-318, February 2009.

[40]. C. Garcia-Martos, J. Rodriguez, and M. J. Sanchez, "Mixed models for short-run forecasting of electricity prices: Application for the Spanish market," IEEE Transaction on Power System, vol. 22, no. 2, pp. 544-552, May 2007.

[41]. A. T. Lora, J. M. R. Santos, A. G. Expósito, J. L. M. Ramos, and J. C. R. Santos, "Electricity market price forecasting based on weighted nearest neighbors techniques", IEEE Transaction on Power System, vol. 22, no. 3, pp. 1294-1301, August 2007.

[42]. N. Amjady, M. Hemmati, "Day-ahead price forecasting of electricity markets by a hybrid intelligent system", European Transactions on Electric Power. vol. 19, no. 1, pp. 89-102, January 2009.

[43]. N. M. Pindoriya, S. N. Singh, & S. K. Singh, "An adaptive wavelet neural network-based energy price forecasting in electricity markets", IEEE Transaction on Power Systems, vol. 23, no. 3, pp. 1423-1432, August 2008.

[44]. J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, “Neural networks and wavelet transform for short-term electricity prices forecasting”, in Proc. 15th Int. Conf. Intelligent System Applications to Power Systems, Curitiba, Brazil, pp. 1-5, November 2009.

[45]. S. Anbazhagan, N. Kumarappan, Day-ahead deregulated electricity market price forecasting using recurrent neural network. IEEE Syst Journal, vol. 7, no. 4, pp.866–72, 2013.

[46]. N. Amjady, A. Daraeepour, Design of input vector for day-ahead price forecasting of electricity markets. Exp Syst Appl., vol. 36, no. 10, pp. 12281–94, 2009.

[47]. J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, “Hybrid Wavelet-PSO-ANFIS Approachfor Short-Term Electricity Prices Forecasting,” IEEE Transactions on Power Systems, vol. 26, no. 1, pp. 137-144, February 2011.

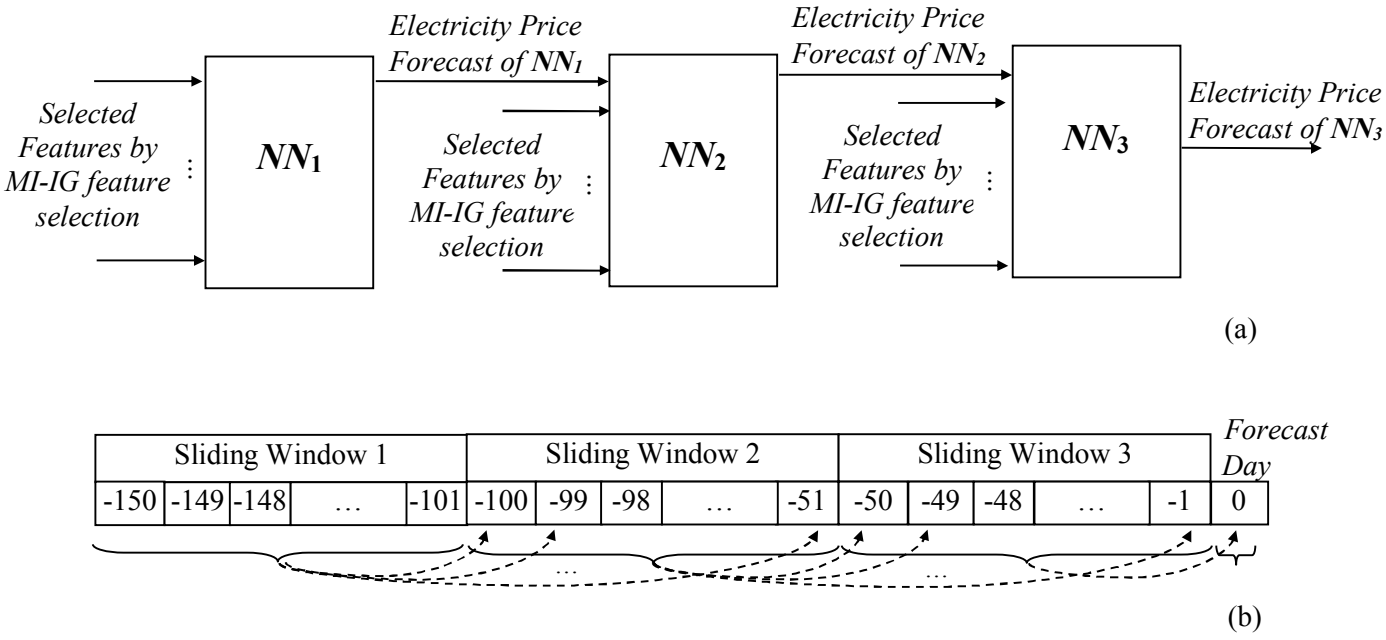


Fig. 1) Representation of the proposed price forecasting engine (CNN): (a) Architecture, and (b) training windows

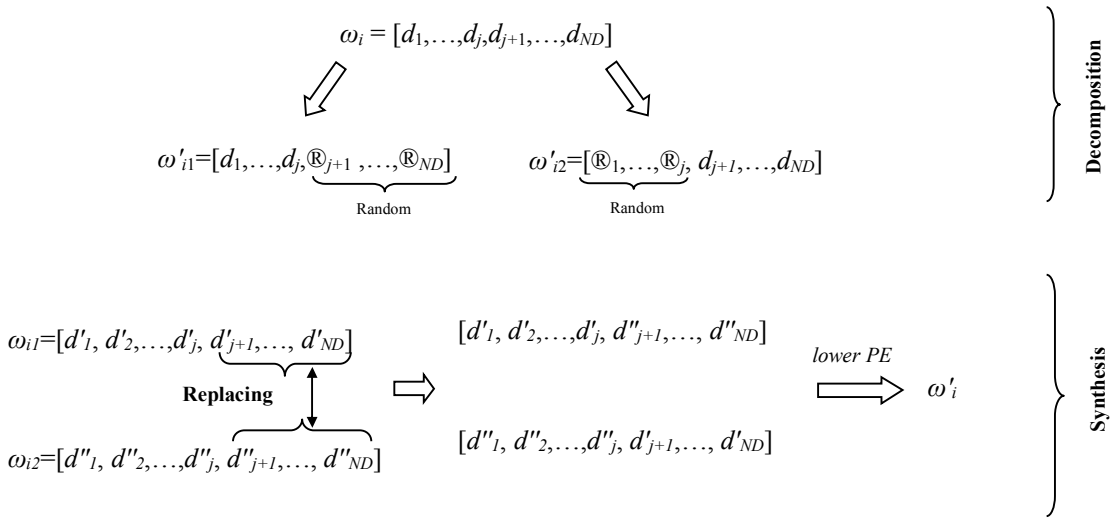


Fig. 2) Representation of the proposed decomposition and synthesis operators for the modified CRO

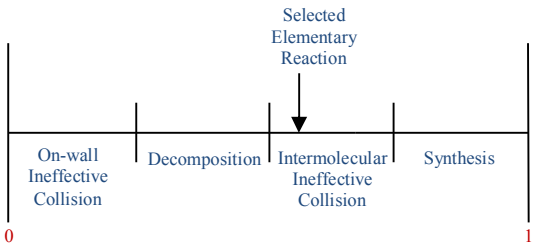


Fig. 3) Roulette wheel mechanism used for selection of the elementary reactions in modified CRO

Pseudo code of Modified CRO

/* Input of the modified CRO algorithm */
1. Input objective function $OF(.)$ and decision variables of the optimization problem
/* Initialization of the modified CRO algorithm */
2. Set the settings including *PopSize* and *KELossRate*
3. **for** each molecule ω **do**
4. Randomly generate the elements of ω within the allowable ranges of the decision variables
5. Calculate $OF(.)$ of ω and assign as PE_ω
6. Assign *InitialKE* of ω
7. **end for**
/* Evolution of the modified CRO algorithm */
8. **while** the stopping criteria is not satisfied **do**
9. **for** each molecule ω **do**
10. Generate a random number with uniform distribution in the interval [0,1]
/* The module of *On-Wall Ineffective Collision* */
11. **if** the generated random number is within the segment of *On-Wall Ineffective Collision* **then**
12. $\omega' = N(\omega)$
13. **if** $PE_\omega + KE_\omega \geq PE_{\omega'}$ **then**
14. The new molecule ω' replaces the original molecule ω
15. **end if**
16. **else**
17. The original molecule ω is restored
18. **end if**
/* The module of *Decomposition* */
19. **if** the generated random number is within the segment of *Decomposition* **then**
20. $[\omega'_1,\omega'_2] = D(\omega)$
21. **if** $PE_\omega + KE_\omega \geq PE_{\omega'_1} + PE_{\omega'_2}$ **then**
22. The new molecules ω'_1 and ω'_2 replace the original molecule ω
23. **end if**
24. **else**
25. The original molecule ω is restored
26. **end if**
/* The module of *Intermolecular Ineffective Collision* */
27. **if** the generated random number is within the segment of *Intermolecular Ineffective Collision* **then**
28. $\omega' = N(\omega)$
29. Randomly select a molecule v from the remaining population
30. $v' = N(v)$
30. **if** $PE_\omega + PE_v + KE_\omega + KE_v \geq PE_{\omega'} + PE_{v'}$ **then**
31. The new molecules ω' and v' replace the original molecules ω and v
32. **end if**
33. **else**
34. The original molecules ω and v are restored
35. **end if**
/* The module of *Synthesis* */
36. **if** the generated random number is within the segment of *Synthesis* **then**
37. Randomly select a molecule v from the remaining population
38. $\omega' = S(\omega,v)$
39. **if** $PE_\omega + PE_v + KE_\omega + KE_v \geq PE_{\omega'}$ **then**
40. The new molecule ω' replaces the original molecules ω and v
41. **end if**
42. **else**
43. The original molecules ω and v are restored
44. **end if**
45. **end for**
46. Specify the best solution in terms of $OF(.)$ value
47. **end while**
/* Output of the modified CRO algorithm */
48. Output the best solution of the last iteration
49. **end**

Fig. 4) Pseudo code of the proposed modified CRO

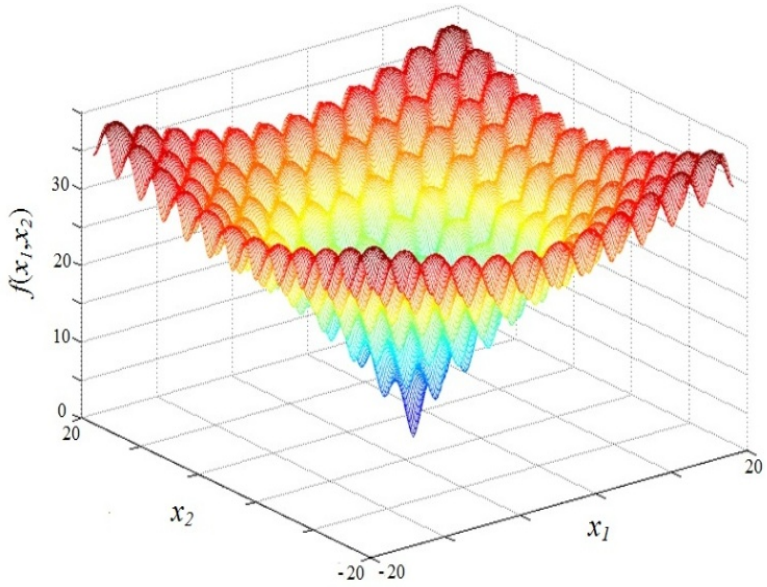


Fig. 5) The shape of Ackley benchmark function

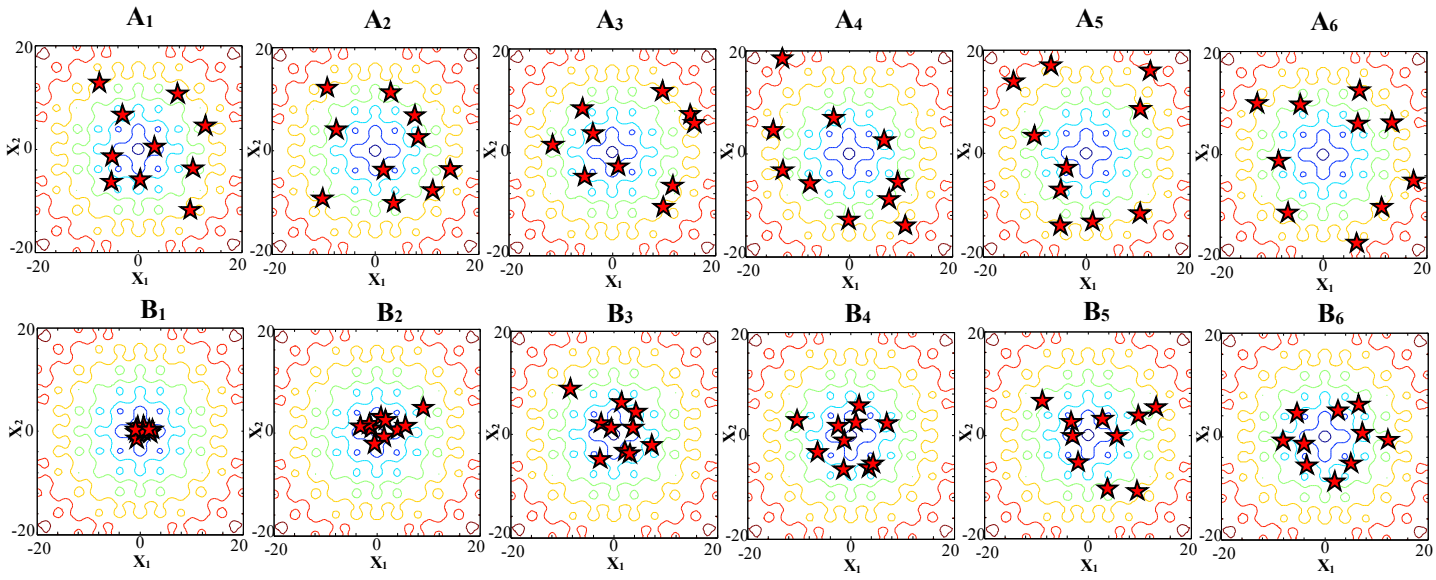


Fig. 6) Distribution of the 10 individuals in the first iteration for the proposed modified CRO (A_1), PSO (A_2), DE (A_3), ACO (A_4), GA (A_5), and EA (A_6); distribution of the 10 individuals in the last iteration for the proposed modified CRO (B_1), PSO (B_2), DE (B_3), ACO (B_4), GA (B_5), and EA (B_6)

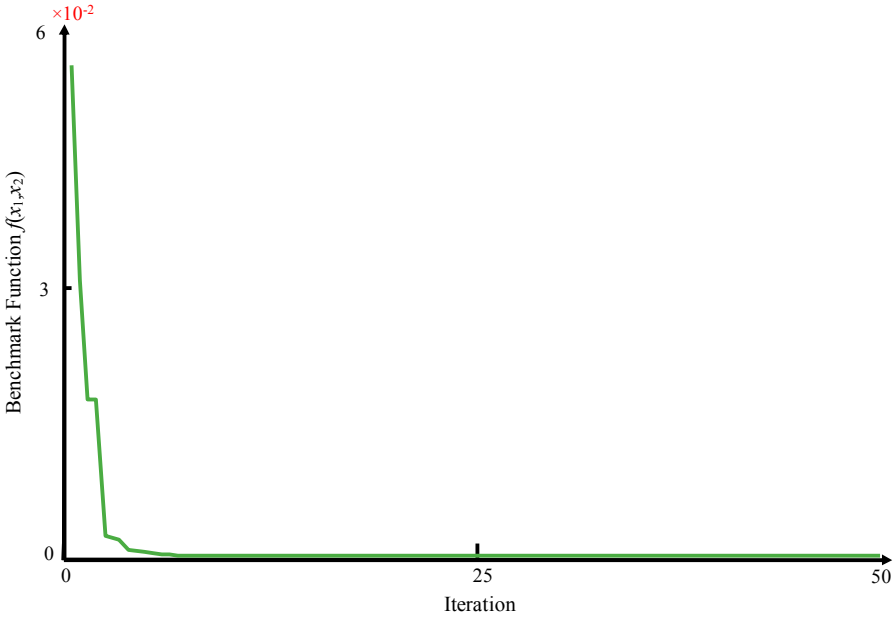
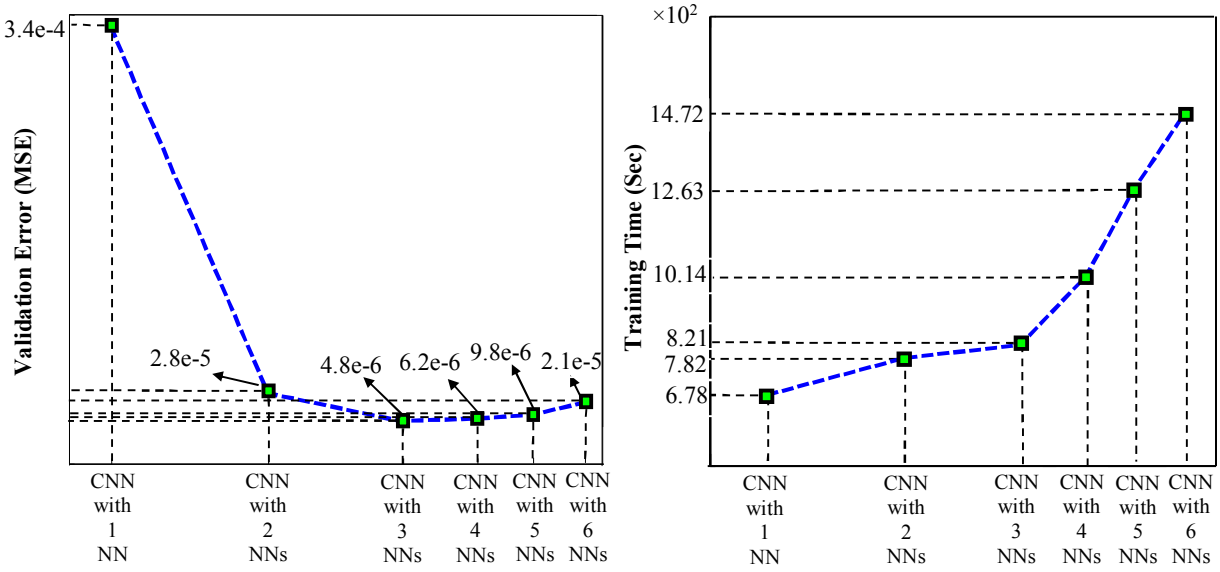


Fig. 7) Convergence trend of the proposed modified CRO for Ackley benchmark function



(a) (b)

Fig. 8) Validation error (a) and training time (b) of CNN with 1 to 6 NNs for November 20, 2006 of PJM electricity market

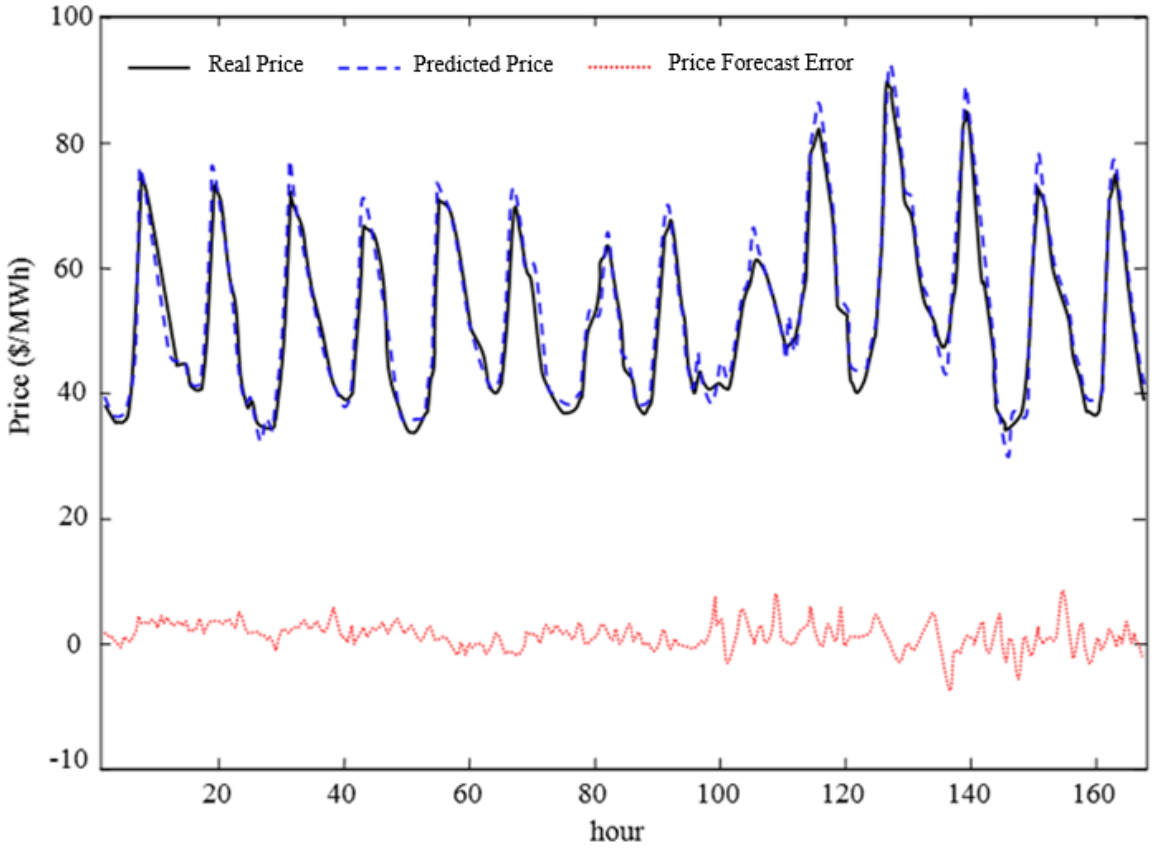


Fig. 9) Real hourly prices, predicted prices by the proposed method and price forecast error for the last week of February in year 2006 of the PJM test case

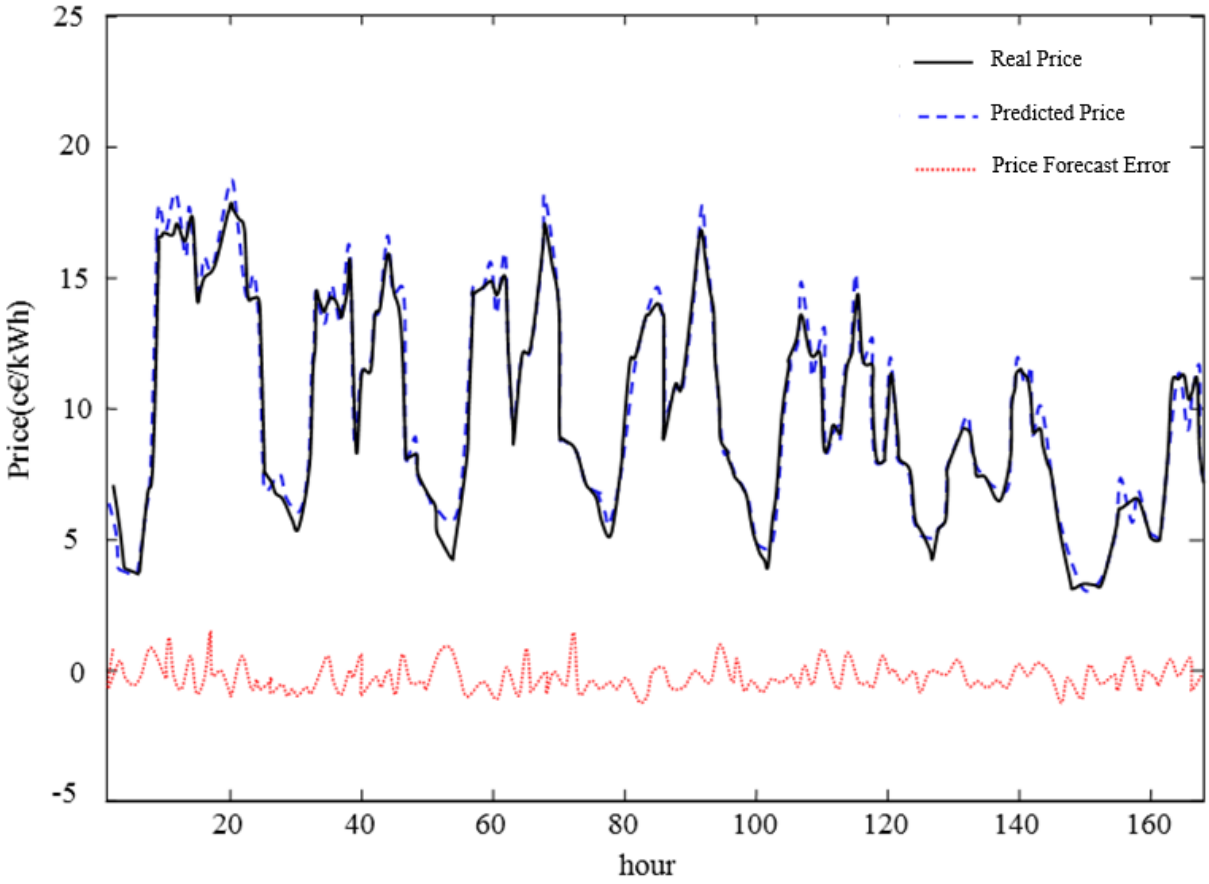


Fig. 10) Real hourly prices, predicted prices by the proposed strategy and price forecast error for the winter test week in the Spanish test case

Table 1) Obtained results for Ackley's benchmark function

Index	**EA [29]**	**GA [30]**	**DE [31]**	**ACO [32]**	**PSO [33]**	**Proposed Modified CRO**
Best	3.14e-8	6.15e-9	5.21e-14	4.87e-14	4.63e-15	0.00
Average	2.031	5.83e-1	5.02e-1	5.09e-1	4.22e-1	1.48e-2
Worst	4.182	4.361	1.572	8.54e-1	1.482	5.73e-2

Table 2) WME (%) for the four test weeks of the PJM electricity market in year 2006

Test Week	ARIMA [35]	NN-LM [35]	NN-BFGS [35]	NN-BR [35]	MI+ composite NN [35]	MI-MI+ composite NN [35]	Proposed
Winter	11.21	9.82	12.90	13.22	4.40	4.33	3.97
Spring	15.30	8.87	10.12	12.92	4.46	4.24	4.02
Summer	13.56	10.43	11.46	11.98	4.81	4.52	4.04
Fall	12.93	9.54	9.83	12.24	4.83	4.61	4.12
Average	13.25	9.67	11.08	12.59	4.63	4.43	4.04

Table 3) WPE (%) for the four test weeks of the PJM electricity market in year 2006

Test Week	ARIMA [35]	NN-LM [35]	NN-BFGS [35]	NN-BR [35]	MI+ composite NN [35]	MI-MI+ composite NN [35]	Proposed
Winter	45.10	22.31	25.42	23.63	9.17	8.13	7.98
Spring	34.32	18.45	21.39	20.13	8.02	7.78	7.56
Summer	44.48	27.34	38.12	30.22	16.10	14.46	13.89
Fall	55.81	32.12	29.23	33.11	19.04	18.33	18.02
Average	44.93	25.05	28.54	26.77	13.08	12.17	11.86

Table 4) e_{week} (%) for the last week of all months of year 2002 in the PJM electricity market

Month	NN [36]	NAÏVE [36]	ARIMA [36]	WL [36]	TF [36]	DR [36]	MI+ composite NN [35]	MI-MI+ composite NN [35]	Proposed
January	9.21	10.6	6.6	10.81	5.48	4.9	4.42	4.38	4.12
February	8.73	9.3	7.71	8.06	4.45	4.54	4.40	4.35	4.14
March	17.98	25.1	15.34	12.10	6.59	6.55	5.19	5.08	4.53
April	32.54	14.22	15.36	14.89	9.49	9.23	5.47	5.38	4.95
May	14.21	12.56	12.51	12.31	6.02	5.53	4.09	4.08	4.01
June	27.33	21.27	17.53	12.94	6.89	6.41	4.93	4.77	4.72
July	27.32	26.11	24.67	16.64	5.64	5.69	5.04	4.92	4.24
August	18.45	25.41	16.14	22.36	6.07	6.03	5.85	5.58	4.16
September	15.9	19.26	16.82	14.74	5.05	4.97	4.37	4.25	4.08
October	12.78	7.45	10.71	10.64	5.72	5.29	4.76	4.62	4.31
November	11.15	10.25	11.19	12.56	5.67	5.58	5.33	5.12	4.96
December	32.53	22.61	27.77	25.70	9.72	7.33	6.11	5.85	5.74
Average	19.01	17.01	15.19	14.48	6.40	6.00	5.00	4.86	4.50

Table 5) $\sigma^2_{e,week}$ for the last week of all months of year 2002 in the PJM electricity market

Month	NN [36]	NAÏVE [36]	ARIMA [36]	WL [36]	TF [36]	DR [36]	Proposed
January	0.009	0.008	0.006	0.008	0.004	0.004	0.003
February	0.007	0.008	0.008	0.007	0.003	0.003	0.003
March	0.027	0.060	0.014	0.010	0.006	0.005	0.003
April	0.103	0.019	0.023	0.014	0.010	0.010	0.006
May	0.013	0.011	0.009	0.008	0.003	0.004	0.002
June	0.057	0.075	0.045	0.018	0.006	0.005	0.003
July	0.130	0.125	0.084	0.023	0.003	0.003	0.003
August	0.035	0.060	0.030	0.061	0.005	0.006	0.004
September	0.030	0.056	0.033	0.018	0.002	0.002	0.002
October	0.015	0.005	0.010	0.006	0.003	0.003	0.002
November	0.010	0.011	0.009	0.010	0.002	0.002	0.002
December	0.116	0.070	0.081	0.046	0.013	0.008	0.006
Average	0.0460	0.0423	0.0293	0.0191	0.0050	0.0046	0.0040

Table 6) e_{day} (%) and e_{week} (%) for the five test days and two test weeks of year 2006 in the PJM electricity market

Test Day	NNSD [18]	WT+FF+FA [37]	HNES [38]	AWNN+GARCH [2]	CNEA [39]	Proposed
Jan. 20th	6.93	5.04	4.98	3.71	4.73	3.35
Feb. 10th	7.96	5.43	4.10	2.85	4.50	2.93
Mar. 5th	7.88	4.82	4.45	5.48	4.92	3.86
Apr. 7th	9.02	6.24	4.67	4.17	4.22	3.68
May. 13th	6.91	4.11	4.05	4.06	3.96	3.52
February 1-7	7.66	6.07	4.62	5.27	4.02	3.31
February 22-28	8.88	6.12	4.66	5.01	4.13	3.29
Average	7.89	5.40	4.50	4.36	4.35	3.42

Table 7) $\sigma^2_{e,day}$ and $\sigma^2_{e,week}$ for the five test days and two test weeks of year 2006 in the PJM electricity market

Test Day	NNSD [18]	CNEA [39]	WT+FF+FA [37]	AWNN+GARCH [2]	HNES [38]	Proposed
Jan. 20th	0.0034	0.0031	0.0016	0.0010	0.0020	0.0010
Feb. 10th	0.0050	0.0036	0.0021	0.0015	0.0012	0.0010
Mar. 5th	0.0061	0.0042	0.0032	0.0033	0.0015	0.0015
Apr. 7th	0.0038	0.0022	0.0019	0.0013	0.0018	0.0011
May. 13th	0.0049	0.0027	0.0016	0.0015	0.0013	0.0012
February 1-7	0.0066	0.0044	0.0023	0.0037	0.0016	0.0014
February 22-28	0.0047	0.0035	0.0024	0.0025	0.0017	0.0018
Average	0.0049	0.0034	0.0022	0.0021	0.0016	0.00128

Table 8) e_{week} (%) for the four test weeks of the Spanish electricity market

Method	Winter	Spring	Summer	Fall	Average
ARIMA [14]	6.32	6.36	13.39	13.78	9.96
Mixed Model [40]	6.15	4.46	14.90	11.68	9.30
MLP [15]	5.23	5.36	11.40	13.65	8.91
Wavelet-ARIMA [17]	4.78	5.69	10.70	11.27	8.11
WNN [41]	5.15	4.34	10.89	11.83	8.05
FNN [16]	4.62	5.30	9.84	10.32	7.52
HIS [42]	6.06	7.07	7.47	7.30	6.97
AWNN [43]	3.43	4.67	9.64	9.29	6.76
NNWT [44]	3.61	4.22	9.50	9.28	6.65
SRN [45]	4.11	4.37	9.09	8.66	6.56
RBFN [8]	4.27	4.58	6.76	7.35	5.74
CNEA [39]	4.88	4.65	5.79	5.96	5.32
MR-MI+NN [46]	4.21	4.76	6.01	5.88	5.22
HNES [38]	4.28	4.39	6.53	5.37	5.14
MI+ composite NN [35]	4.51	4.28	6.47	5.27	5.13
WPA [47]	3.37	3.91	6.50	6.51	5.07
MI-MI+ composite NN [35]	4.29	4.20	6.31	5.01	4.95
Proposed	3.28	3.62	5.32	5.03	4.31

Table 9) $\sigma^2_{e,week}$ for the four test weeks of the Spanish electricity market

Method	Winter	Spring	Summer	Fall	Average
ARIMA [14]	0.0034	0.0020	0.0158	0.0157	0.0092
MLP [15]	0.0017	0.0018	0.0109	0.0136	0.0070
Wavelet-ARIMA [17]	0.0019	0.0025	0.0108	0.0103	0.0064
FNN [16]	0.0018	0.0019	0.0092	0.0088	0.0054
AWNN [43]	0.0012	0.0031	0.0074	0.0075	0.0048
NNWT [44]	0.0009	0.0017	0.0074	0.0049	0.0037
CNEA [39]	0.0036	0.0027	0.0043	0.0039	0.0036
HIS [42]	0.0034	0.0049	0.0029	0.0031	0.0036
MR-MI+NN [46]	0.0014	0.0033	0.0045	0.0048	0.0035
RBFN [8]	0.0015	0.0019	0.0047	0.0049	0.0033
WPA [47]	0.0008	0.0013	0.0056	0.0033	0.0027
MI+ composite NN [35]	0.0014	0.0014	0.0033	0.0022	0.0021
HNES [38]	0.0013	0.0015	0.0033	0.0022	0.0021
MI-MI+ composite NN [35]	0.0014	0.0014	0.0032	0.0023	0.0021
Proposed	0.0012	0.0012	0.0028	0.0018	0.0017