

Short-term price forecasting of Nordic power market by combination Levenberg–Marquardt and Cuckoo search algorithms

Mun Kyeom Kim ✉

Department of Energy System Engineering, Chung-Ang University, 84 Heukseok-Ro, Dongjak-gu, Seoul, 604-714, Republic of Korea
 ✉ E-mail: mkim@cau.ac.kr

ISSN 1751-8687

Received on 16th March 2014

Revised on 12th March 2015

Accepted on 7th April 2015

doi: 10.1049/iet-gtd.2014.0957

www.ietdl.org

Abstract: This study proposes a new forecasting method for short-term spot prices in the Nordic power market. It proposes a Cuckoo search Levenberg–Marquardt (CSLM)-trained, CSLM feed-forward neural network (CSLM-FFNN) for the solving process that combines the improved Levenberg Marquardt and Cuckoo search algorithms. The proposed model considers actual power generation and system load as input sets to facilitate the efficient use of both transmission and power generation resources by direct market participants. During the training, the proposed CSLM-FFNN model generalises the relationship between the area prices and the system price for the same period. The model can be updated to track online the variation trend of the electricity price and to maintain accuracy because of the rapid training speed in CSLM learning algorithm. The developed model is tested with publicly available data acquired from the Nord Pool, and the model's performance is compared with state-of-the-art artificial neural networks and time-series models. Besides, the proposed approach is applied to forecast market-clearing price in the Spanish electricity market, to further assess the validity of the approach. The results show that the proposed CSLM-FFNN exhibits superior performance than other methods in terms of forecasting accuracy and training efficiency.

Nomenclature

N	number of forecasted hours
D	demand in the spot market (MW)
X	volume difference between power generation and system load
P_{system}	system price
P_{area}	area price
$P_t^{\text{Act}}, P_t^{\text{For}}$	actual and forecasted prices at t -hour, respectively
ΔP	price difference between bidding areas
ΔP_{area_t}	AP difference for each of bidding area at t -hour in the Nord Pool
P_{area_t}	AP at t -hour in the Nord Pool
E_t	forecast error at t -hour
E_t^{Ave}	average error of the forecasting period ($t = 1, 2, \dots, N$)
P_{t-i}	price of i hours ago
L_{t-i}	system load of i hours ago (MW)
G_{t-i}	power generation of i hours ago (MW)
ω	scalar parameter
α	parameter vector
π_1, π_2	multiplication factor and division factor
l	length of the solution vector
n	population size
γ	step size
$E(\alpha)$	error vector
$J(\alpha)$	Jacobian matrix
$\nabla f(\alpha)$	gradient vector
$\nabla^2 f(\alpha)$	Hessian matrix
M_{Iter}	total number of iterations
$\oplus$	entry-wise product

1 Introduction

In a deregulated electricity market, accurate electricity price forecasting represents a key source of information. Market participants and transmission system operators (TSOs) rely on

price forecasting to set their bidding strategies during market business activity. The most basic pricing concept in the electricity market is the market-clearing price [1]. In an unconstrained market-clearing process, generator companies are dispatched based on their bidding prices. However, when transmission constraints exist, the energy for both day-ahead and real-time markets is priced by locational marginal prices (LMPs) [2]. The LMPs can reflect the price of economically dispatched energy and the overhead of energy delivery to locations while keeping power safely flowing across transmission and other power system facilities without violating their physical limits during actual operating conditions. Therefore the LMPs provide important information to market participants with respect to bidding strategies. It is also a crucial indicator that facilitates re-dispatch for congestion management by TSOs [3]. Accurate LMP forecasting, however, is complex because LMPs are affected by market behaviours and depend heavily on transmission congestion. Moreover, because electrical imbalances may typically occur from transmission bottlenecks, these issues reinforce extreme electricity price volatility – or even price spikes [4]. This volatility indicates the need of the electric power industry for practical models that adequately represent electricity prices.

Several methods have recently been reported to predict electricity market prices. Time-series forecasting is traditionally based on linear models such as auto-regressive (AR) and integrated moving average (ARIMA) models [5]. In [6], the authors assessed the forecasting abilities of univariate models including AR, ARIMA and unobserved component models. Time-series models with exogenous variables perform better than univariate models [7]. These models are straightforward to implement; however, they fail to provide satisfactory results when utilising non-linear, non-stationary time series. In [8], the authors suggested generalised auto-regressive conditional heteroskedasticity (GARCH) for electricity price forecasting. A simple technique using weighted nearest neighbours was used to forecast hourly prices in deregulated electricity markets [9]. The majority of time-series models are accurate. However, they represent linear prediction

models that trace patterns in historical input data, whereas electricity prices represent a non-linear function of the input features.

Among various forecasting methods, artificial neural networks (ANNs) have been considered an effective approach. ANNs are a simple yet powerful and flexible forecasting tool. ANNs provide superior solutions for the modelling of complex non-linear relationships compared to traditional linear models [10]. However, traditional ANNs are mostly trained by gradient-based learning algorithms, such as back-propagation (BP), which usually suffers from local optima, over-fitting problems and extended training periods when predicting complicated signals including electricity prices [11]. To solve this problem, the Levenberg–Marquardt (LM) algorithm was developed for layer-by-layer ANN topology only, which is far from optimal [12]. The algorithm represents a satisfactory combination of Newton’s method and steepest descent, but it still cannot avoid local minima in favour of the global minimum. Recently, the cooperative co-evolutionary approach using neural networks weight coefficients has been proposed to solve these problems for short-term load and price forecasting [13]. In [14], an adaptive wavelet neural network combined the localisation property of wavelet and learning capability of the ANN was used to predict LMP in mainland Spain and the PJM electricity markets, respectively. A hybrid model has been proposed for short term electricity price forecasting based on the combination of particle swarm optimisation (PSO) and the adaptive-network-based fuzzy inference system [15]. However, despite the existing research performed in this area, there remains a need for more accurate and robust price forecasting methods in electricity markets.

The main goal of this paper is to develop a model that leads to smaller prediction errors, and to obtain the appropriate length of time to use for day-ahead forecasts. This study combines the LM and Cuckoo search (CS) algorithms into a CSLM-trained, CSLM feed-forward neural network (CSLM-FFNN) to build a forecasting model for short-term electricity prices. In the proposed model, CS is used at the beginning stage to generate the optimal weight parameters, and LM continues the training by taking the best weight selected in the CS algorithm to minimise the training error. Publicly available data acquired from the Nord Pool is used for training and testing the ANN to demonstrate the superiority of the proposed CSLM-FFNN. An input feature selection based on correlation analysis is also conducted to select appropriate input data from the Nord Pool spot market. Two types of spot prices exist in the Nord Pool spot: the system price (SP) and the area price (AP). The SP is a spot price that disregards bottlenecks in

the day-ahead market. However, when the power transfer between bidding areas exceeds trading capacity and transmission congestion is predicted, the bidding areas may exhibit different prices. The AP calculation is iterated so that the available trading capacity between the areas is used to the maximum during every hour of operation. After the addition of flow between the areas, the AP is balanced for a new equilibrium point considering the transfer capacity, and the AP forecast can then be obtained.

The remainder of this paper is organised as follows. Section 2 describes the spot price mechanism, the selection of available input variables and price forecasting strategy in the Nordic power market. Section 3 presents the ANN forecast engine; that is, the proposed CSLM-FFNN model. The testing results are presented in Section 4, and the conclusions are presented in Section 5.

2 Nord Pool

2.1 Spot market structure

In the Nord Pool, Elspot is a spot auction-based day-ahead energy market in which market participants submit offers to sell or bids to buy, physical electricity for delivery in each hour of the following day [16]. The spot price represents the price, independent of any transmission constraint (i.e. the unconstrained market-clearing price) because trading capacity between bidding areas has not been taken into account in determining the price. This price is called the SP for the Nordic power market.

Whenever congestion takes place in transmission grids, each Nordic country is separated into several bidding areas. Available transmission capacity may vary and, thus, APs are established for each transmission-constrained area. The APs are published within 2 h of the Elspot market closing [17]. An example of the supply and demand curves for two areas (surplus area and deficit area) is shown in Fig. 1. A lower price in the surplus area will lead to greater purchases and fewer sales, which can provide a parallel shift in the demand curve. However, by increasing the price in the deficit area, the area participants sell more and purchase less, and the sale can provide a parallel shift in the supply curve. Here, P_L and P_H represent the low and high prices when there is full utilisation of trading capacity. $P_{Cap=0}$ is a price in an area with an isolated price calculation. As shown in Fig. 1, the available capacity is included in the AP calculation by moving the supply curve in the deficit area and the demand curve in the surplus area as indicated by the solid curves. The AP calculation is iterated so

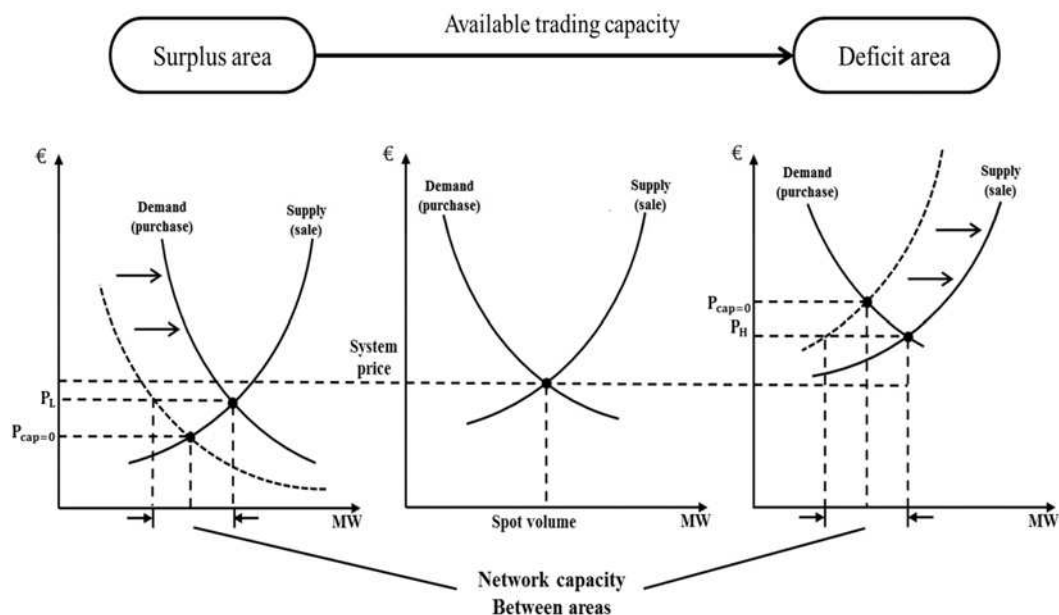


Fig. 1 APs in the Nord Pool

that the available trading capacity between the high price and the low price area is used to the maximum during every hour of operation to ensure that power flows from the low-price area towards the high-price area. Accordingly, the APs in the surplus and deficit areas are the new equilibrium points following the addition of power flow between the areas of purchase and sale. In this situation, the APs are relatively low in the surplus area (P_L) and relatively high in the deficit area (P_H).

2.2 Input feature selection

A selection of input variables is important to achieve high forecasting accuracy. It is a process generally used in machine learning wherein a subset of features available from data is selected for learning algorithm application. Adequate feature selection can enhance the generalisation capability of unseen data and simplify the learning process of the forecasting tool. The candidate set of input variables usually contain different lags of historical prices, electricity load, available generation, generation outages, operational reserves, maintenance schedule, bidding strategies, weather conditions, time indicators and hydro generation availability. An ideal forecasting method should include all possible variables that affect the spot price. However, in reality, it is impossible to include all variables when forecasting spot price. Certain variables are more significant and, in practice, only these can be considered. The unit outage information, although significant, was not considered in the study because it is typically proprietary and not available to all market participants in real time. The amount of operational reserves and the maintenance schedule do not improve the forecast. Moreover, there are some variables, such as bidding strategies and unethical competitive behaviour that are not easily represented in mathematical form.

Another important group of exogenous variables for price forecasting consists of weather conditions (e.g. temperature, relative humidity, rainfall, wind speed etc.) and especially temperature variable. Extreme weather conditions may cause an increase in electricity load, which in turn may lead to a higher spot price because more expensive generation sources must be activated. However, in this paper, these exogenous variables have not been used for price forecast because of the following reasons. Difficult weather conditions are typical for the Nordic region, especially during winter time. Electricity load is higher when the atmospheric temperature rises or falls from a base comfortable level; temperature dependent load variations are more extreme if the humidity is higher, since moisture increases the heat retention capability of air. Atmospheric pressure variations generally cause air temperature variations, and as a consequence, load variations. The correlation of temperature and price is found to be very much similar to the load-price correlation. The effect of temperature and other weather related variables can be incorporated in the electricity demand, and therefore they were not used in the price forecasting model to avoid variable collinearity. Another variable that drives the price is the hour of the day; however, the impact is also reflected in electricity load [18]. Finally, hydropower covers half of the power requirements for the Nordic system. The majority of Norway's electricity is supplied by hydroelectric power, and hydroelectric power accounts for approximately 50% of Sweden's total power capacity. Together, the two countries jointly contribute a substantial proportion of hydroelectric power to the Nord Pool. Thus, reservoir-filling percentages in these two countries have a significant effect on spot prices, and relevant information is collected as an exogenous variable [19]. However, for simplicity, no additional exogenous variables were included in this paper.

A correlation analysis based on Pearson's correlation coefficient is conducted in this paper. The extent of correlated linearity between two variables is tested using the correlation coefficient, which can be calculated as follows

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Here, x and y represent the two studied variable sets, and $\bar{x}$ and $\bar{y}$ represent the average value of the two sets, respectively. The correlation coefficient implies the resemblance between the two variables. If there is no relationship between the two variables, the correlation coefficient is 0 or significantly low, whereas a greater absolute value of R represents a stronger linear correlation between the two variables. A perfect fit gives a coefficient of 1. Thus, the higher the correlation coefficient, the larger the merit order to use as the ANN input.

Similarly, the calculation of the correlation coefficient of determination between the two variables is performed, which is used to validate the adequacy of the regression model. The correlation coefficient of determination (R^2) is defined as follows

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

Table 1 shows the relationship between price and load, which illustrates the correlation between spot price, system load and demand in the Nord Pool spot for 10 July 2009. The correlation coefficient (R) value of 0.971 is obtained for the system load, and 0.978 is obtained for the demand in the spot market. The correlation coefficient of determination (R^2) of 0.793 represents the system load, and 0.865 represents the demand in the spot market. This implies that each variation in the regression is 79.3 and 86.5% and indicates a high goodness of fit. This variable is a natural choice for a parameter to be used in the prediction of the spot price because the spot price is strongly correlated with electricity load.

The spot price can also vary according to the power generation sources because the generating volume moves the supply curve depending on the specific power mix in the bidding area. Table 2 shows the relationship between the spot price and volume of overall demand and regional generation in the Nord Pool for the period 1 22:00–3 23:00 May 2009. Although the demand has a similar volume during this period, different spot prices may occur in the Nordic power market. For example, the spot price rises when the generating volume decreases in Norway, whereas the spot price falls when the generating volume increases in Denmark. This is because electricity is generated from different energy sources in Nordic countries. In Norway, hydropower production is

Table 1 Relationship between spot price and electricity load in the Nord Pool for 10 July 2009

Hour	Spot price, EUR/MWh	System load, MWh/h	Demand in the spot market, MWh/h
0–1	41.83	33 426	24932.3
1–2	36.41	32 078	24082.5
2–3	34.31	31 602	23668.3
3–4	33.41	31 113	23348.3
4–5	33.28	30 763	23257.8
5–6	34.81	31 808	24285.1
6–7	43.54	34 269	26182.7
7–8	49.01	37 085	28690.9
8–9	49.61	39 017	30031.4
9–10	56.3	40 098	31024.2
10–11	58.15	40 818	31487.1
11–12	58.87	40 868	31562.7
12–13	55.65	40 837	31490.6
13–14	54.22	40 662	31327.6
14–15	53.17	40 329	31125.8
15–16	50.99	39 991	30517.3
16–17	49.12	39 496	29678.8
17–18	49.49	39 250	29260.9
18–19	49.2	38 727	28860.4
19–20	47.97	37 933	28293
20–21	47.95	37 165	27948.9
21–22	47.28	36 548	27735.2
22–23	44.89	36 698	27587.5
23–24	42.37	35 096	26378.3
correlation coefficient (R)		0.971	0.978
correlation coefficient of determination (R^2)		0.793	0.865

Table 2 Comparison of spot price with demand and regional generating volumes in the Nord Pool

Period	Spot price, EUR/MWh	Demand, MWh/h	Generating volumes of bidding areas, MWh/h			
			Norway	Denmark	Sweden	Finland
1 May 2009, 22:00–23:00	29.76	36269.3	10478.3	2447.4	17483.6	4864.2
2 May 2009, 22:00–23:00	30.84	36194.7	10403.8	2508.8	17582.9	4822.5
3 May 2009, 22:00–23:00	33.05	36 147	10056.1	2821.1	17 617	4976.2

easily regulated and the price is cheaper than in Denmark, which mainly generates thermal power. Accordingly, the spot price responds to the power mix in bidding areas compared to demand in the Nordic power market.

2.3 Price forecasting strategy

The SP is formulated as follows

$$P_{\text{system}} = F(D) \quad (3)$$

Equation (3) indicates that the SP is affected primarily by demand in the spot market. The SP is determined based on hourly bids from both supply- and demand-side participants for the trading of prompt physically delivered electricity. The seller, for example, the owner of a hydroelectric power plant, must decide how much can be delivered and at what price on an hourly basis. The buyer, typically a utility, must assess how much energy is required to meet customer demand in the coming day, and how much should be paid for this volume on an hourly basis.

This paper denotes bidding areas for each Nordic country by an alphanumeric code, such as NO1, NO2, NO3, DK1, DK2, SE and FI. The Norwegian TSO defines the fixed bidding areas in Norway, according to the information concerning the likely pattern of flows in the system for a certain period of time. The number of Norwegian bidding areas can vary. This paper defines three bidding areas as NO1, NO2 and NO3. When necessary, additional price areas are used. Western Denmark (DK1) and eastern Denmark (DK2) are always treated as different bidding areas, and Sweden (SE) and Finland (FI) constitute one bidding area each. For the Nordic bidding area, this can be expressed as a function

$$P_{\text{system}} = F(D_{\text{NO1}}, D_{\text{NO2}}, D_{\text{NO3}}, D_{\text{DK1}}, D_{\text{DK2}}, D_{\text{SE}}, D_{\text{FI}}) \quad (4)$$

When the power flow between the bidding areas is within the limits set by the TSOs, the SP is the only price for that specific hour throughout the entire Nordic market area. However, in the AP calculation aggregated supply and demand curves are created for each area from the bids of the market participants located in that area. The price differences between bidding areas occur when the surplus volume at the SP, in one or more bidding areas, is greater than the total export capacity from these areas. If a bidding area has a power surplus (extra power flows into an adjacent area), a volume corresponding to the transfer capacity on the constrained connection should be considered as a price-independent purchase in the surplus area, and a price-independent sale in the deficit area, using the maximum capacity between the areas, resulting in differing APs. The participants' bids in the bidding areas on each side of the congestion are aggregated into supply and demand curves in the same manner as for the SP calculation.

In the Nord Pool spot's price mechanism, the trading capacity determines the SP and APs, because its direction and volume actually depend on the differences of APs in the bidding areas. In this study, the differences in the generation and system load of each area are used to predict the price differences, which generalises the relationship between the respective APs and the SP during the same period. For this, the following price difference in

the Nordic area can be readily obtained

$$\Delta P_{\text{area}} = F(X_{\text{NO1}}, X_{\text{NO2}}, X_{\text{NO3}}, X_{\text{DK1}}, X_{\text{DK2}}, X_{\text{SE}}, X_{\text{FI}}) \quad (5)$$

Then, the AP is calculated as follows

$$P_{\text{area}} = P_{\text{system}} + \Delta P_{\text{area}} \quad (6)$$

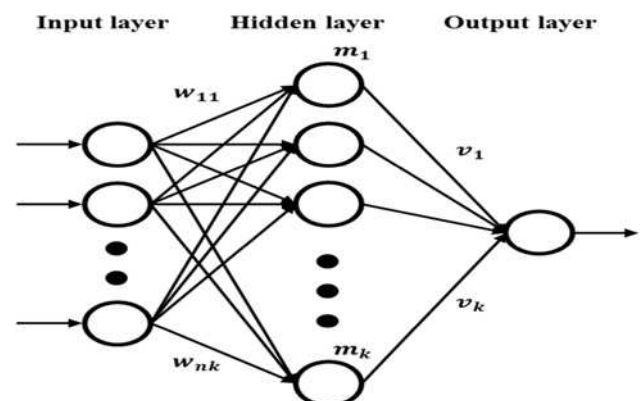
The AP calculation is iterated so that the capacity between bidding areas is used to the maximum; this mechanism itself helps to relieve grid congestion. After the addition of flow between the bidding areas, the AP is found at the new equilibrium point, which is balanced taking the transfer capacity into account.

3 Forecast engine

3.1 LM-FFNN model

Among numerous ANN models, the multi-layer FFNN model has primarily been used because of the well-known universal estimation proficiencies. A three-layer FFNN is particularly suited to forecasting. It implements a non-linear, hyperbolic-tangent sigmoid activation function for the hidden layer and a pure linear transfer function for the output layer. Fig. 2 shows a three-layer FFNN with a single output node, k hidden nodes and n input nodes. w_{nk} represents the connection weight from the n th input node to the k th hidden node, and v_k represents the connecting weight from the k th hidden node to the output node.

Forecasting with neural networks requires two steps: the training step and the learning step. With respect to the training step, the selection of training data is a significant part of the design of the neural network. The adequate selection of inputs is influential to the success of training. During training, a suitable training set with patterns of input and corresponding output pairs is identified and is trained as the FFNN for price forecasting. In the learning step, a learning algorithm entails an optimisation process, which is the minimisation of some error measure between the output produced and the desired output. The error minimisation process is repeated until an acceptable criterion for convergence is reached. In the

**Fig. 2** Three-layer feedforward neural network model

learning process, the BP algorithm is widely recognised as a powerful tool for the training of the FFNN [10]. However, because the standard BP algorithm applies the steepest descent method to update the weights, it converges slowly and often yields suboptimal solutions. For fast and efficient training, second-order learning algorithms are required. The most effective method is the LM algorithm [12], which is a derivative of the Newton method. The mathematical details of the quasi-Newton minimisation technique can be found in the Appendix. The LM algorithm for the Gauss–Newton method is defined as

$$\Delta\alpha = -[J^T(\alpha)J(\alpha) + \omega I]^{-1}J^T(\alpha)E(\alpha) \quad (7)$$

where ω represents a scalar that could be modified following each iteration. Note that when ω is large, the LM algorithm becomes the steepest descent, whereas when ω is small, the algorithm becomes Gauss–Newton, which should provide faster convergence. An update of ω that is too big or too small will cause the neural network to take longer to train. The disadvantage is similar with the learning rate in the standard BP algorithm. An appropriate update of ω is more efficient for convergence. The rule for the adjustment of ω is as follows

$$\omega(t) = \begin{cases} \omega(t-1) \times \pi_1, & \text{if } E(t) > E(t-1) \\ \omega(t) = \omega(t), & \text{if } E(t) = E(t-1) \\ \frac{\omega(t-1)}{\pi_2}, & \text{if } E(t) < E(t-1) \end{cases} \quad (8)$$

To get better results and less time consumption with the LM training algorithm, the ω multiplication factor (π_1) is often set larger than the ω division factor (π_2). In the algorithm, the rate of change in ω is exponential. The LM algorithm provides the best performance between the speed of Newton's method and the guaranteed convergence of the steepest descent, but it still cannot avoid local minimum.

3.2 CS algorithm

CS is a population-based optimisation algorithm based on cuckoo bird reproductive behaviour [20]. This algorithm is stimulated by the forceful parasitic behaviour of some cuckoo species that lay their eggs in the nests of other birds. In particular, CS can be modified to give a relatively high convergence rate to the true global minimum [21]. The basic CS is defined by the effort to survive among cuckoos. Each nest simulates an existing solution when the cuckoo egg simulates a new promising solution for the investigated problem. The main idea of the CS algorithm is to replace the existing solutions in the nests with the better one that is generated by a cuckoo. For simplicity, the CS algorithm follows three basic rules:

- Each cuckoo lays one egg at a time, which represents a solution, and leaves the egg in a random nest.
- The best nests with high quality eggs or solutions will pass to the next generation.
- The number of host nests is fixed, and the egg laid by a cuckoo is discovered by the host bird with the probability $pa \in [0, 1]$. Either the egg is destroyed or the nest is abandoned if the cuckoo's egg is discovered. This will result in the construction of a new nest with random solutions.

According to wide research on the birds, the flying movement of many animals and insects is random, which can be simulated as Levy flights. A Levy flight is a random walk for which the step lengths are distributed according to a heavy-tailed probability distribution [21]. The random walk via Levy flight is more effective in exploring the search space because its step length is longer in the long run.

Initially, n is the number of nests is generated randomly as follows

$$Z^t = [z'_{i,1}, z'_{i,2}, \dots, z'_{i,l}]; \quad i = 1, 2, \dots, n; \quad t = 1, 2, \dots, M_{\text{iter}} \quad (9)$$

When generating new solutions z^{t+1} for a cuckoo i , the Levy flight is defined as

$$z^{t+1} = z'_i + \gamma \oplus \text{Levy}(\theta) \quad (10)$$

Here, step size $\gamma > 0$ should be related to the scales of the problem of interest. The Levy flight essentially provides a random walk, whereas the random step length is drawn from a Levy distribution as shown in the following equation

$$\text{Levy}(\theta) \sim u = t^{-\theta}; \quad (1 < \theta \leq 3) \quad (11)$$

This has an infinite variance with an infinite mean. The steps essentially construct a random walk process with a power-law, step-length distribution and a heavy tail. Some of the new solutions should be generated by Levy walk around the best solution obtained so far, which will speed up the local search. Notably, based on adequate computation, the result of CS always gives the best optimum weight. The optimal solutions obtained by the CS are far better than the finest solutions found by PSO or genetic algorithm (GA) [22]. Specially, CS can benefit not only to avoid being trapped in local minima, but also to simultaneously find all optima in a search space with the help of Levy flights.

3.3 Proposed CSLM algorithm

This section presents a novel learning algorithm that combines the CS and LM algorithms to train parameters of neural networks to minimise the error between the forecast and the actual values and improve the performance by escaping from local minima. The main theory behind this combined algorithm is that, in the first stage, the CS algorithm completes its training. The LM algorithm starts training with the weights generated by the CS algorithm, and the LM trains the network until the stopped condition is satisfied. The LM algorithm incorporates the Newton method and the gradient descent method.

In the proposed CSLM algorithm, each cycle of the search consists of several initialisation steps of the best nest or possible solution (i.e. the weight space and the corresponding biases for FFNN optimisation in this study). The weight optimisation problem and the size of population represent the quality of the solution. In the first epoch, the best weights and biases are initialised with CS and these weights are passed to the LM-FFNN. The training process for searching the optimal weight parameters is continued until the last cycle/epoch of the network is converged. The flowchart of the work is depicted in Fig. 3.

3.4 Forecasting accuracy evaluation

Forecasting error is the main concern for TSOs; a lower error indicates a superior result for price prediction. The most widely used criteria for measuring forecasting error are the mean absolute percentage error (MAPE), the error variance (σ^2), the sum square error (SSE), and the standard deviation of error (SDE) [11]. This accuracy is computed as a function of the actual prices that occurred. The daily MAPE and the error variance (σ^2) can be defined as

$$\text{MAPE} = \frac{100}{N} \sum_{t=1}^N \frac{|P_t^{\text{For}} - P_t^{\text{Act}}|}{P_t^{\text{Act}}} \quad (12)$$

$$\sigma^2 = \frac{1}{N} \left[\sum_{t=1}^N \frac{|P_t^{\text{Act}} - P_t^{\text{For}}|}{P_t^{\text{Act}}} - \text{MAPE} \right]^2 \quad (13)$$

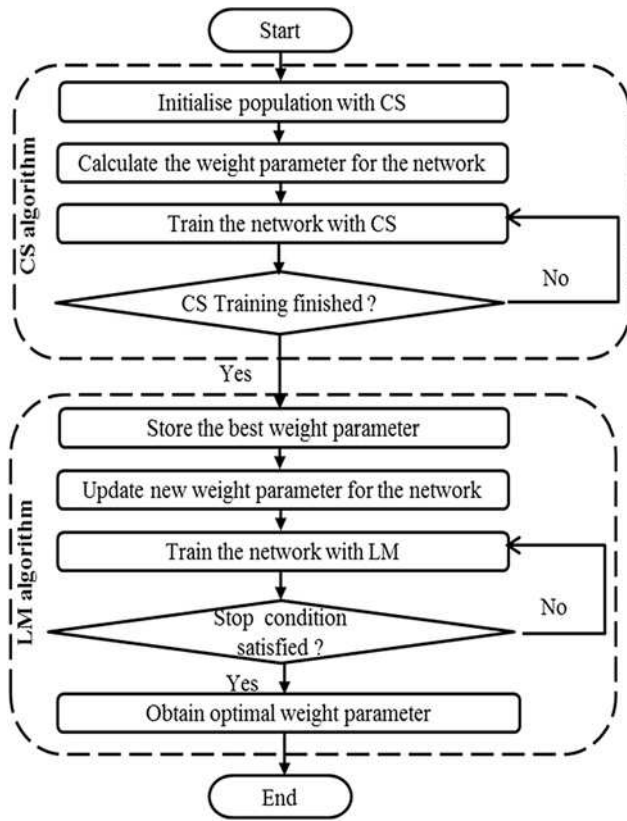


Fig. 3 Flowchart of proposed CSLM algorithm

Moreover, the SSE criterion is computed as

$$SSE = \sum_{t=1}^N (P_t^{\text{For}} - P_t^{\text{Act}})^2 \quad (14)$$

The SDE criterion is given by

$$SDE = \sqrt{\frac{1}{N} \sum_{t=1}^N (E_t - E_t^{\text{Ave}})^2} \quad (15)$$

where

$$E_t = P_t^{\text{For}} - P_t^{\text{Act}} \quad (16)$$

$$E_t^{\text{Ave}} = \frac{1}{N} \sum_{t=1}^N E_t \quad (17)$$

4 Case study

The proposed forecasting method was applied to forecast spot prices in the Nord Pool. Historical data from the Nord Pool [17] were used to train and test the proposed CSLM-FFNN model. To satisfy performance, the proposed CSLM-FFNN model was compared with eight state-of-the-art ANNs, BP-FFNN, GABP-FFNN, PSOBP-FFNN, an artificial bee colony BP (ABCP)-FFNN, LM-FFNN, GALM-FFNN, PSOLM-FFNN, ABCLM-FFNN and other time-series models [5, 8] were also built for tests.

More input features from the trend or different seasonality may contain more information content than the selected set. However, for more input features, the CSLM-FFNN model require more training set, and so, a longer training period should be considered. If appropriate input features are selected exactly then the proposed

forecasting model can be more sustainable. For this purpose, different kinds of the sensitivity analysis, which is performed to evaluate the relative importance of the input variables on the accuracy of the forecast, can be utilised as suggested in [23]. First, the most effective lags are selected by correlation analysis. Typically, hourly price usually has a high correlation with its short-run trend, daily periodicity and weakly periodicity. Instead of a single, the forecast can be a linear combination or a regression procedure that can include several similar days. The selection of similar days is part of the training in the CSLM-FFNN. The training data are classified as weekdays from Monday to Friday and weekends as Saturday and Sunday. For example, eight similar days are selected for training to predict the price on Monday or Sunday and so on. One day is taken from the similar days as test data. A similar day is characterised as follows. A Monday is similar to the Monday of the previous week and the same rule applies for Saturdays and Sundays; analogously, a Tuesday is similar to the Monday of the same week, and the same rule applies for Wednesdays, Thursdays and Fridays [24]. With consideration of these characteristics of the price series, the following set of input features was considered to forecast the price P_t at t -hour

$$\{P_{t-1}, P_{t-2}, P_{t-3}, P_{t-4}, P_{t-23}, P_{t-24}, P_{t-25}, P_{t-47}, P_{t-48}, P_{t-49}, P_{t-71}, P_{t-72}, P_{t-73}, P_{t-95}, P_{t-96}, P_{t-97}, P_{t-119}, P_{t-120}, P_{t-121}, P_{t-143}, P_{t-144}, P_{t-145}, P_{t-167}, P_{t-168}, P_{t-169}, P_{t-335}, P_{t-336}, P_{t-337}, P_{t-503}, P_{t-504}, P_{t-505}, P_{t-671}, P_{t-672}, P_{t-673}\} \quad (18)$$

In (18), the first four terms are composed of price information concerning trends in the price signal. The subsequent 18 terms contain price information concerning daily seasonality (up to six days ago), whereas the latter 12 terms relate to weekly seasonality (one to four weeks ago). For daily seasonality, in addition to P_{t-24} (price of 24 h ago), P_{t-23} and P_{t-25} can be considered and, similarly, for the other periods based on correlation.

In order to also consider the price-based variations of system load pattern in the Nordic power market, lagged system load values are also included in the candidate set of inputs. Similarly, the system load signal has the characteristics of short-run trend, daily and weekly periodicities. However, the spot price and system load have generally a different nature by the non-linear signal. For solving this problem, the normalised values of the prices and system loads are calculated, and then correlation coefficients are achieved. The results of the sensitivity analysis between the normalised values of prices and system loads imply that the following set of lagged system load values are highly correlated with price P_t at each hour to be forecasted

$$\{L_{t-1}, L_{t-2}, L_{t-3}, L_{t-24}, L_{t-25}, L_{t-48}, L_{t-49}, L_{t-72}, L_{t-73}, L_{t-95}, L_{t-96}, L_{t-97}, L_{t-119}, L_{t-120}, L_{t-144}, L_{t-145}, L_{t-168}, L_{t-336}, L_{t-337}, L_{t-504}, L_{t-672}\} \quad (19)$$

The different sets of lagged power generation can be also considered as input features, owning a high correlation with the price of each hour

$$\{G_{t-1}, G_{t-2}, G_{t-3}, G_{t-24}, G_{t-25}, G_{t-48}, G_{t-49}, G_{t-72}, G_{t-73}, G_{t-95}, G_{t-96}, G_{t-97}, G_{t-119}, G_{t-120}, G_{t-144}, G_{t-145}, G_{t-168}, G_{t-336}, G_{t-337}, G_{t-504}, G_{t-672}\} \quad (20)$$

The candidate set of inputs for the proposed price forecast strategy includes 34 lagged prices plus 21 lagged system loads and 21

lagged power generations, totally $34 + 21 + 21 = 76$ selected features among total 2019 candidate inputs.

The available data set is divided into training, validation and testing subsets. The training set was used for training and updating the biases and learning rates of the network. In this study, the training set was defined as a three-month period from 13 April to 12 July 2009. When P_t is forecasted, it is used as P_{t-1} for the price prediction of the next hour, and this cycle is repeated until the AP of the training period (here, the next 24 h) is predicted. Thus, the training data includes $91 \times 24 = 2184$ learning patterns. The validation set was utilised to monitor the training process, and the test set was used to evaluate the generalisation capacity after training. The testing set started on 15 July 2009, after the training period, because the Nord Pool spot closes at 12 am, and there are at least 12 h and at most 36 h between the time of the trade and the time of delivery. At the training step, the number of neurons in the hidden layer varies for different applications and usually depends on the size of the training set and the number of input variables. When the various number of neurons in the hidden layer were tested, the best results were produced with five hidden. The

output layer had one unit, which was set to output the spot prices. In addition, after the tests were conducted through the different parameters, the training parameters were as follows: $\pi_1 = 3$ and $\pi_2 = 1.12$. In CA, the value of γ is a constant of 1, and population size is 15 with 100 iterations. All the simulations were implemented with MATLAB on an Intel Xeon processor E5620 with eight quad-core processors having a clock speed of 2.4 GHz and 12 GB of RAM memory.

4.1 Simulation results

Fig. 4 shows the results of the actual and forecasted SP in the Nord Pool. The MAPE of the proposed approach is 2.35%. The forecasting results can accurately track the actual SP. The CSLM-FFNN model generalised relationships between each AP and SP during the same period. The AP forecasting is illustrated in Figs. 5–7. The results showed overall good performance for the AP forecasting during the test period. It is observed that the peak, because of its high volatility and price spikes, is the most difficult

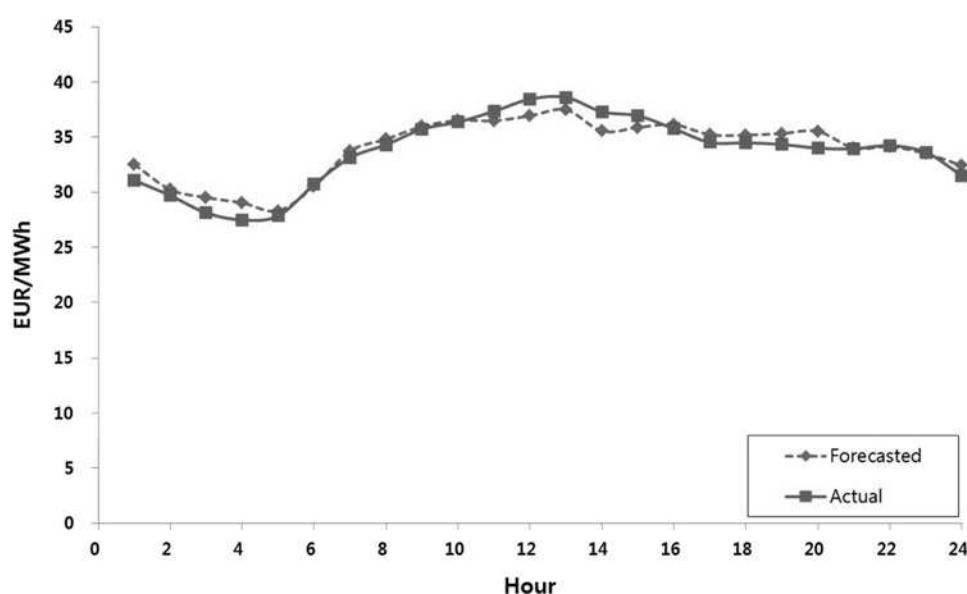


Fig. 4 Actual and forecast SP in the Nord Pool Spot

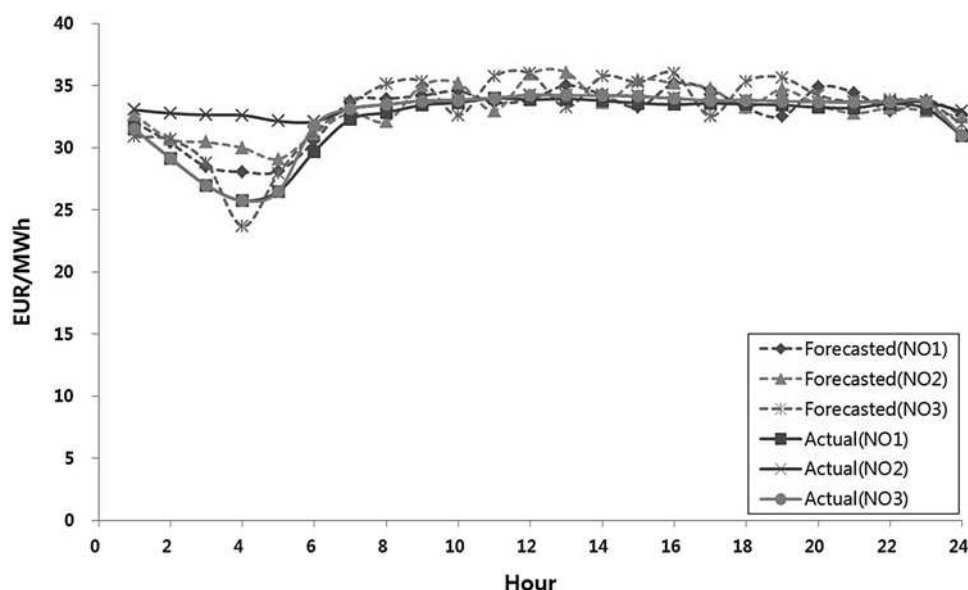


Fig. 5 Actual and forecast AP of NO1, NO2 and NO3

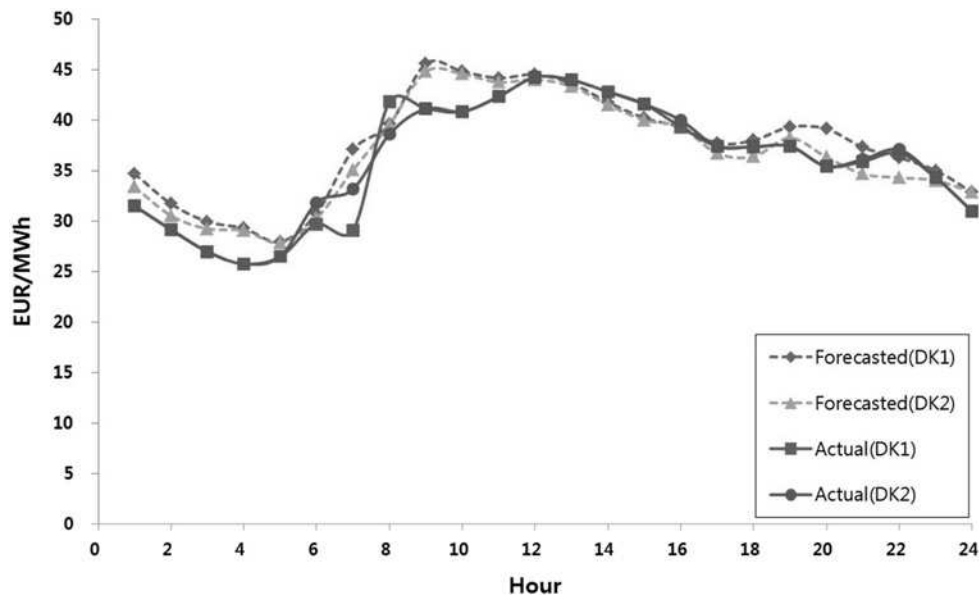


Fig. 6 Actual and forecast AP of DK1 and DK2

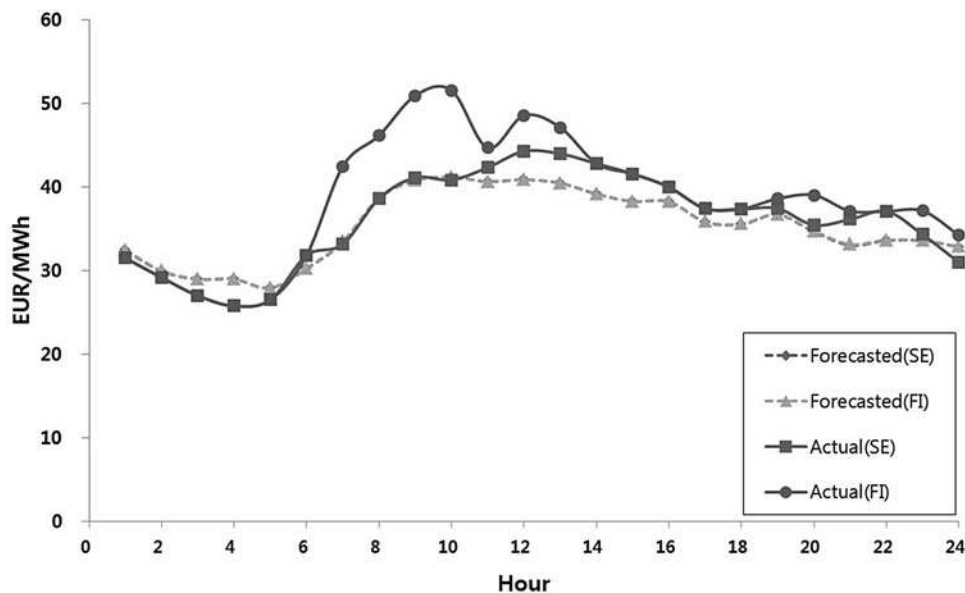


Fig. 7 Actual and forecast AP of SE and FI

feature to predict. During peak hours, most generating units are running under high and even full capacity, usually away from their most economic operating point. To solve this problem, some research has suggested adding pre-processing actions, such as limiting the magnitude of spikes or excluding days with price spikes from the training data to improve training and test performance. However, price spikes are indicative of abnormalities in the system and are natural price signals. Accordingly, the proposed method did not seek to improve the forecast results by implementing such pre-processing methods; however, in contrast to previous work, accurate price predictions were obtained during peak hours for all Nordic countries. This offers valuable information for market participants and facilitates sound decisions because price spikes have the capacity to significantly affect the profitability of both suppliers and customers.

The APs respond highly to the power mix in bidding areas. Fig. 5 shows that a hydropower area, such as Norway (NO1, NO2 and NO3) has an AP that is lower than the SP. Hydropower is easily regulated and can show substantial differences during the day. For

this reason, the transmission requirements can vary greatly. There can be daily and hourly patterns with less price variation in hydropower-dominant areas because of the high degree of controllability. Contrastingly, in a predominantly thermal bidding area, the APs are greater than the SP. Finland has a similar power mix to Sweden, but with a higher share of thermal and nuclear power. Denmark has mainly thermal power generation, with an increasing share of wind power. Figs. 6 and 7 show that the APs for DK1, DK2 and FI are above the SP. This has implications for power trading between the Nordic countries. There were slight differences between the forecasted and actual prices because of higher price volatility in DK1, and particularly in FI. Moreover, a comparison of SE and FI shows that the APs are of vital importance for the efficient use of hydropower in the mixed hydrothermal Nordic system.

Table 3 presents the error values for each Nordic area to assess the prediction capacity. The first column indicates the area, the second column presents the MAPE, the third column presents the square root of the SSE and the fourth column presents the SDE. The

Table 3 Statistical analysis of forecasting error

Area	MAPE, %	$\sqrt{SSE}$	SDE
NO1	3.28	5.65	0.82
NO2	3.59	6.84	1.36
NO3	3.71	6.53	1.18
DK1	5.49	11.92	2.05
DK2	3.91	7.68	1.50
SE	2.87	6.00	1.20
FI	7.04	17.98	3.01

Table 4 Generalisation performance of state-of-the-art ANNs

ANN model	MAPE, %	Error variance (σ^2)	Training time, s
BP-FFNN	6.81	0.2280	189.76
GABP-FFNN	6.02	0.1836	762.38
PSOBP-FFNN	5.16	0.0954	472.12
ABCBP-FFNN	5.07	0.0876	595.74
LM-FFNN	5.63	0.1258	87.85
GALM-FFNN	5.34	0.1147	538.10
PSOLM-FFNN	4.93	0.0808	241.97
ABCLM-FFNN	4.85	0.0762	372.79
CSLM-FFNN	4.27	0.0443	28.33

lower values of MAPE, SSE and SDE demonstrate that the proposed forecasting model is efficient and provides a reliable prediction for AP forecasting.

4.2 Performance test

This section tests the generalisation performance of proposed CSLM-FFNN. Eight state-of-the-art ANNs, BP-FFNN, GABP-FFNN, PSOBP-FFNN, ABCBP-FFNN, LM-FFNN, GALM-FFNN, PSOLM-FFNN and ABCLM-FFNN were tested for comparison. For each model, the tests are conducted with the same training and testing data sets. For GA settings, single point crossover operation with the rate of 0.8 was employed. Mutation rate and chosen generation gap value were 0.01 and 0.9, respectively. For ABC settings, colony size was 50 and the number of food sources was 25, which is equal to the number of employed bees. The upper and lower bounds were $[-0.5, 0.5]$. For PSO settings, the initial number of particles and generations were supposed to be 20 and 100, respectively. Also, the social components and the inertia weight were set as 2 and 0.8 experimentally.

Table 4 lists the results of the training performances. In terms of forecasting error, the eight ANN models are approximate with a MAPE from ~ 4.85 to 6.81%, whereas the proposed CSLM-FFNN model is the lowest. The error variance, a significant performance criterion, was also calculated to measure the robustness of the proposed model. The smaller the variance, the more precise the prediction of APs. This indicates that CSLM-FFNN is capable of representing the non-linear function better than eight state-of-the-art ANNs. In terms of training speed, the CSLM-FFNN is faster than the eight compared ANNs. Specifically, it is 21 times and 13 times faster than ABCBP-FFNN and ABCLM-FFNN, respectively. The proposed CSLM-FFNN shows overall superior performance to the eight compared ANNs.

Table 5 shows a comparison between the proposed CSLM-FFNN and time series methods (ARIMA and GARCH) with respect to the MAPE criterion. The MAPE showed that the proposed CSLM-FFNN was superior for each area during all periods. Consequently, the proposed CSLM-FFNN has high forecasting accuracy and its performance is less affected by volatility, which can be of the utmost importance in practical application.

4.3 Comparison with other recent methods

The forecasting method is compared with some of the most recently published works in this area, which gives a better insight about the

Table 5 Comparison of MAPE (%) with time series methods

Area	ARIMA [5]	GARCH [8]	CSLM-FFNN
NO1	5.63	5.32	3.28
NO2	4.17	4.01	3.59
NO3	5.98	5.57	3.71
DK1	11.42	9.46	5.49
DK2	10.74	8.75	3.91
SE	8.45	7.23	2.87
FI	21.79	15.27	7.04

Table 6 Comparative MAPE (%) results for the Spanish electricity market, 2002

Method	Winter	Spring	Summer	Fall	Average
mixed-model [26]	6.15	4.46	14.90	11.68	9.30
LMFFNN [27]	5.23	5.36	11.40	13.65	8.91
wavelet-ARIMA [28]	4.78	5.69	10.70	11.27	8.11
WNN [9]	5.15	4.34	10.89	11.83	8.05
HIS [29]	6.06	7.07	7.47	7.30	6.97
AWNN [14]	3.43	4.67	9.64	9.29	6.75
CNEA [30]	4.88	4.65	5.79	5.96	5.32
HPA [15]	3.65	4.19	6.76	6.53	5.28
CSLM-FFNN	2.72	3.36	4.58	4.73	3.85

forecast capability of the method. For this, the proposed CSLM-FFNN was performed to forecast next-week prices in the electricity market of mainland Spain, commonly used as the test case in several price forecasting researches. Price forecasting was obtained using historical data of year 2002 for the Spanish electricity market [25]. For the sake of simplicity and clear comparison, no exogenous variables are also considered. Besides, to make the fair comparison with other methods reported in [9, 14, 15, 26–30], the same test weeks are selected, which correspond to the four seasons of year 2002. Hourly historical price data of the 42 days previous to the day of the week whose prices are to be forecasted were considered to build the forecasting model.

Table 6 shows a comparison between the proposed CSLM-FFNN and the other methods, regarding the MAPE criterion. As the test results reveal, the proposed method presents an overall better forecasting accuracy than all other considered methods. The MAPE for the Spanish market has an average value of 3.85%. In summer and fall seasons, volatility of the Spanish electricity market increases [28] and so all other methods encounter considerably larger forecasting errors. However, accuracy of proposed CSLM-FFNN has smaller variations. Furthermore, the average running time is less than 11 s using MATLAB on an Intel Xeon processor E5620 with eight quad-core processors having a clock speed of 2.4 GHz and 12 GB of RAM memory. Recently, an average value of 2.58% for the MAPE has been reported using a hybrid method based on WT, CLSSVM and EGARCH model [31]. However, this approach has a major drawback: the running time of about 10 min. Therefore the proposed CSLM-FFNN provides the best trade-off between forecasting accuracy and computation time, which can be of the utmost importance for real-life applications.

5 Concluding remarks

This paper proposed a new forecasting approach for short-term electricity prices by achieving more accurate results and improved convergence speed. The proposed CSLM-FFNN model was examined using the Nord Pool, which is considered a successful electrical market. The simulation results revealed the forecasting capabilities of the CSLM-FFNN and its superiority over the eight state-of-the-art ANNs, BP-FFNN, GABP-FFNN, PSOBP-FFNN, ABCBP-FFNN, LM-FFNN, GALM-FFNN, PSOLM-FFNN, ABCLM-FFNN and time-series models, ARIMA, GARCH in both training efficiency and forecasting accuracy. The forecasting error of the CSLM-FFNN is acceptably small for practical use and

satisfactory performance with a close spike tracking capability is observed. The rapid fast training speed of the CSLM-FFNN is another advantage because it enables efficient on-line model updating to maintain the forecasting accuracy during practical use. In order to demonstrate the effectiveness of the proposed method over the other methods in the area, day-ahead price prediction of the Spanish electricity market is also considered. The MAPE results from the comparisons showed that the CSLM-FFNN was more effective than other recent methods. For summer and fall seasons, all other methods demonstrated additional prediction errors because of high volatility, price spikes and the non-linear behaviour of price signals, whereas the proposed method demonstrated relatively accurate predictions. This factor provides effective information to market participants and facilitates sound, profitable decisions. Selection of the best input features for nonlinear behaviour of electricity price signal, dependent on the structure of the electricity market, is a challenging task, which can be a matter for future research.

6 Acknowledgments

This research was supported by the Chung-Ang University Research Grants in 2014.

7 References

- Kim, M.K., Nam, Y.W., Park, J.K.: 'Market-clearing for pricing system security based on voltage stability criteria', *Energy*, 2011, **36**, pp. 1255–1264
- Aggarwal, S.K., Saini, L.M., Kumar, A.: 'Electricity price forecasting in deregulated markets: a review and evaluation', *Int. J. Electr. Power Energy Syst.*, 2009, **31**, pp. 13–22
- Kim, M.K., Hur, D.: 'Decomposition-coordination strategy to improve power transfer capability of interconnected systems', *Int. J. Electr. Power Energy Syst.*, 2011, **33**, pp. 1638–1647
- Cifter, A.: 'Forecasting electricity price volatility with the Markov-switching GARCH model: evidence from the Nordic electric power market', *Electr. Power Syst. Res.*, 2013, **102**, pp. 61–67
- Contreras, J., Espinola, R., Nogales, F.J., Conejo, A.J.: 'ARIMA models to predict next day electricity prices', *IEEE Trans. Power Syst.*, 2003, **18**, pp. 1014–1020
- Cuadras, J.C., Hlouskova, J., Kossmeier, S., Obersteiner, M.: 'Forecasting electricity spot-prices using linear univariate time-series models', *Appl. Energy*, 2004, **77**, pp. 87–106
- Weron, R., Misorek, A.: 'Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models', *Int. J. Forecast.*, 2008, **24**, pp. 744–763
- Garcia, R.C., Contreras, J., Akkeren, M.V., Garcia, J.B.C.: 'A GARCH forecasting model to predict day-ahead electricity prices', *IEEE Trans. Power Syst.*, 2005, **20**, (2), pp. 867–874
- Lora, A.T., Santos, J.M.R., Expósito, A.G., Ramos, J.L.M., Santos, J.C.R.: 'Electricity market price forecasting based on weighted nearest neighbors techniques', *IEEE Trans. Power Syst.*, 2007, **22**, pp. 1294–1301
- Szkuta, B.R., Sanabria, L.A., Dillon, T.S.: 'Electricity price short-term forecasting using artificial neural networks', *IEEE Trans. Power Syst.*, 1999, **14**, pp. 851–858
- Mandal, P., Senjyu, T., Funabashi, T.: 'Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market', *Energy Convers. Manage.*, 2006, **47**, pp. 2128–2142
- Hagan, M.T., Mehraj, M.B.: 'Training feed forward networks with the Marquardt algorithm', *IEEE Trans. Neural Netw.*, 1994, **5**, pp. 989–993
- Karsaz, A., Mashhadi, H.R., Mirsalehi, M.M.: 'Market clearing price and load forecasting using cooperative co-evolutionary approach', *Electr. Power Energy Syst.*, 2010, **32**, pp. 408–415
- Pindoriya, N.M., Singh, S.N., Singh, S.K.: 'An adaptive wavelet neural network-based energy price forecasting in electricity markets', *IEEE Trans. Power Syst.*, 2008, **23**, pp. 1423–1432
- Pousinho, H.M.I., Mendes, V.M.F., Catalão, J.P.S.: 'Short-term electricity prices forecasting in a competitive market by a hybrid PSO-ANFIS approach', *Electr. Power Energy Syst.*, 2012, **39**, pp. 29–35
- Wangensteen, I.: 'Power system economics: the Nordic electricity market' (Tapir Academic Publishers, 2006)
- Nord Pool spot market. Available at: <http://www.nordpoolspot.com>
- Rodriguez, C.P., Anders, G.J.: 'Energy price forecasting in the Ontario competitive power system market', *IEEE Trans. Power Syst.*, 2004, **19**, pp. 366–374
- Vehviläinen, I., Pykkönen, T.: 'Stochastic factor model for electricity spot price-the case of the Nordic Market', *Energy Econ.*, 2005, **27**, pp. 351–367
- Yang, X.S., Deb, S.: 'Cuckoo search via Lévy flights'. World Congress on Nature & Biologically Inspired Computing, 2009, pp. 210–214
- Walton, S., Hassan, O., Morgan, K., Brown, M.: 'Modified cuckoo search: a new gradient free optimisation algorithm', *Chaos, Solitons Fractals*, 2011, **44**, pp. 710–718
- Rajabioun, R.: 'Cuckoo optimization algorithm', *Appl. Soft Comput.*, 2011, **11**, pp. 5508–5518
- Mandal, P., Senjyu, T., Urasaki, N., Funabashi, T.: 'A neural network based several hour-ahead electric load forecasting using similar days approach', *Int. J. Electr. Power Energy Syst.*, 2006, **28**, pp. 367–373
- Amjady, N., Hemmati, M.: 'Energy price forecasting-problems and proposals for such predictions', *IEEE Power Energy Mag.*, 2006, **4**, pp. 20–29
- Spanish electricity market. Available at: <http://www.omel.es/en/inicio>
- Garcia-Martos, C., Rodriguez, J., Sanchez, M.J.: 'Mixed models for short-run forecasting of electricity prices: application for the Spanish market', *IEEE Trans. Power Syst.*, 2007, **22**, pp. 544–552
- Catalão, J.P.S., Mariano, S.J.P.S., Mendes, V.M.F., Ferreira, L.A.F.M.: 'Short-term electricity prices forecasting in a competitive market: a neural network approach', *Electr. Power Syst. Res.*, 2007, **77**, pp. 1297–1304
- Conejo, A.J., Plazas, M.A., Espinola, R., Molina, A.B.: 'Day-ahead electricity price forecasting using the wavelet transform and ARIMA models', *IEEE Trans. Power Syst.*, 2005, **31**, pp. 1035–1042
- Amjady, N., Hemmati, H.: 'Day-ahead price forecasting of electricity markets by a hybrid intelligent system', *Eur. Trans. Elect. Power*, 2009, **19**, pp. 89–102
- Amjady, N., Keynia, F.: 'Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro evolutionary algorithm', *IEEE Trans. Power Syst.*, 2009, **24**, pp. 306–318
- Zhang, J., Tan, Z.: 'Day-ahead electricity price forecasting using WT, CLSSVM and EGARCH model', *Electr. Power Energy Syst.*, 2012, **45**, pp. 362–368

8 Appendix: LM algorithm

Newton's update for optimising a performance index $f(\alpha)$ is

$$\alpha_{k+1} = \alpha_k - A_k^{-1} \mathbf{g}_k \quad (21)$$

where $A_k \equiv \nabla^2 f(\alpha)|_{\alpha=\alpha_k}$ represents the Hessian matrix and $\mathbf{g}_k \equiv \nabla f(\alpha)|_{\alpha=\alpha_k}$ represents the gradient vector. Assuming that $f(\alpha)$ is a sum of square function, given by

$$f(\alpha) = \sum_{i=1}^N E_i^2(\alpha) \quad (22)$$

Then, it can be shown that

$$\nabla f(\alpha) = J^T(\alpha) \mathbf{E}(\alpha) \quad (23)$$

$$\nabla^2 f(\alpha) = J^T(\alpha) \mathbf{J}(\alpha) + \mathbf{U}(\alpha) \quad (24)$$

where $\mathbf{E}(\alpha)$ is the error vector and $\mathbf{J}(\alpha)$ is the Jacobian matrix given by

$$\mathbf{J}(\alpha) = \begin{bmatrix} \frac{\partial E_1(\alpha)}{\partial \alpha_1} & \frac{\partial E_1(\alpha)}{\partial \alpha_2} & \dots & \frac{\partial E_1(\alpha)}{\partial \alpha_n} \\ \frac{\partial E_2(\alpha)}{\partial \alpha_1} & \frac{\partial E_2(\alpha)}{\partial \alpha_2} & \dots & \frac{\partial E_2(\alpha)}{\partial \alpha_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial E_N(\alpha)}{\partial \alpha_1} & \frac{\partial E_N(\alpha)}{\partial \alpha_2} & \dots & \frac{\partial E_N(\alpha)}{\partial \alpha_n} \end{bmatrix} \quad (25)$$

where $\mathbf{U}(\alpha)$ is given by

$$\mathbf{U}(\alpha) = \sum_{i=1}^N E_i(\alpha) \nabla^2 E_i(\alpha) \quad (26)$$

Assuming that $\mathbf{U}(\alpha) \simeq 0$, the Hessian matrix can be approximated as

$$\nabla^2 f(\alpha) \approx 2J^T(\alpha) \mathbf{J}(\alpha) \quad (27)$$

Substituting (27) and (23) into (21), the Gauss-Newton method is obtained

$$\alpha_{k+1} = \alpha_k - [J^T(\alpha_k) \mathbf{J}(\alpha_k)]^{-1} J^T(\alpha_k) \mathbf{E}(\alpha_k) \quad (28)$$

One problem with the Gauss-Newton method is that the matrix $H = J^T J$ may not be invertible. This can be overcome by

using the following modification to the approximate Hessian matrix

$$\mathbf{M} = \mathbf{H} + \omega \mathbf{I} \quad (29)$$

To make this matrix invertible, suppose that the eigenvalues and eigenvectors of $\mathbf{H}$ are $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ and $\{v_1, v_2, \dots, v_n\}$.

Then

$$\mathbf{M}z_i = [\mathbf{H} + \omega \mathbf{I}]v_i = \mathbf{H}v_i + \omega v_i = \lambda v_i + \omega v_i = (\lambda_i + \omega)v_i \quad (30)$$

Therefore the eigenvectors of $\mathbf{M}$ are the same as the eigenvectors of $\mathbf{H}$, and the eigenvalues of $\mathbf{M}$ are $(\lambda_i + \omega)$. $\mathbf{M}$ can be made positive definite by increasing ω until $(\lambda_i + \omega) > 0$ for all i , and therefore the matrix will be invertible

$$\alpha_{k+1} = \alpha_k - [J^T(\alpha_k)J(\alpha_k) + \omega_k I]^{-1} J^T(\alpha_k)E(\alpha_k)$$

or

$$\Delta\alpha_k = -[J^T(\alpha_k)J(\alpha_k) + \omega_k I]^{-1} J^T(\alpha_k)E(\alpha_k) \quad (31)$$

The LM algorithm has a useful feature that as ω_k is increased it approaches the steepest descent algorithm with small learning rate

$$\begin{aligned} \alpha_{k+1} &\cong \alpha_k - \frac{1}{\omega_k} J^T(\alpha_k)E(\alpha_k) \\ &= \alpha_k - \frac{1}{2\omega_k} \nabla f(\alpha) \text{ for large } \omega_k \end{aligned} \quad (32)$$