

Article

Price Forecasting in the Day-Ahead Energy Market by an Iterative Method with Separate Normal Price and Price Spike Frameworks

Sergey Voronin * and Jarmo Partanen

LUT Energy, Laboratory of Electricity Markets and Power Systems, Lappeenranta University of Technology, P.O. Box 20, Lappeenranta 53851, Finland; E-Mail: jarmo.partanen@lut.fi

* Author to whom correspondence should be addressed; E-Mail: sergey.voronin@lut.fi;
Tel.: +358-40-689-4380; Fax: +358-5-621-2350.

Received: 12 September 2013; in revised form: 25 October 2013 / Accepted: 5 November 2013 /
Published: 12 November 2013

Abstract: A forecasting methodology for prediction of both normal prices and price spikes in the day-ahead energy market is proposed. The method is based on an iterative strategy implemented as a combination of two modules separately applied for normal price and price spike predictions. The normal price module is a mixture of wavelet transform, linear AutoRegressive Integrated Moving Average (ARIMA) and nonlinear neural network models. The probability of a price spike occurrence is produced by a compound classifier in which three single classification techniques are used jointly to make a decision. Combined with the spike value prediction technique, the output from the price spike module aims to provide a comprehensive price spike forecast. The overall electricity price forecast is formed as combined normal price and price spike forecasts. The forecast accuracy of the proposed method is evaluated with real data from the Finnish Nord Pool Spot day-ahead energy market. The proposed method provides significant improvement in both normal price and price spike prediction accuracy compared with some of the most popular forecast techniques applied for case studies of energy markets.

Keywords: electricity price forecasts; price spike forecasts; compound classifier; hybrid methodology; input feature selection

1. Introduction

Electricity price forecasting has become an important area of research in the aftermath of the worldwide deregulation of the power industry. Unlike electricity demand series, electricity price series can exhibit variable means, major volatility and significant spikes [1].

Based on the needs of the energy market, a variety of approaches for electricity price forecasting have been proposed in the last decades, among them, models based on simulation of power system equipment and related cost information [2], game-theory based models which focus on the impact of bidder strategic behavior on electricity prices [3], models based on stochastic modeling of finance [4], regression models [5] and artificial intelligence models [6–9]. In recent years, hybrid approaches have become popular since it is almost universally agreed in the forecasting literature that no single method is best in every situation [10,11].

While most existing approaches to forecasting electricity prices are reasonably effective for normal prices, they cannot deal with the price spikes accurately. In early research, price spikes were truncated before application of the forecasting model to reduce the influence of such observations on the estimation of the model parameters [12,13]. Electricity price spikes, however, are significant for energy market participants to stay competitive in a competitive market.

The Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) process has been tested to simulate price spikes in original price series and has not been able incorporate spikes with a height usually observed in the original prices [14]. Spikes have been incorporated into a Markov-switching model and diffusion models with the addition of a Poisson jump component [15]. Data mining techniques have been applied to the spike forecasting problem and have achieved promising results [16–19].

An analysis of price forecasts along with knowledge of upcoming price spikes is important for market participants to estimate their potential and stay competitive in a competitive market. However, most of the work on electricity market price forecasting is concentrated on improving forecast accuracy rather than the effects of price forecast inaccuracy on market participants. Only a few approaches have been reported in the literature to deal with the problem of future price uncertainty in operation planning in competitive environments [20–22].

The methodology presented in this paper uses an iterative hybrid approach to separately predict normal prices and price spikes in the Finnish day-ahead energy market. Such a strategy provides an opportunity to train the forecasting models more effectively while the non-separate forecasting methods should learn the behaviors of both normal prices and price spikes. The proposed approach uses a wavelet transform (WT) combined with ARIMA, a neural network (NN), a compound classifier and a k-nearest neighbor model (k-NN) to separately implement normal price and price spikes forecasting processes. WT deals with non-stationarity by decomposing the price series into less volatile components. The ARIMA model captures cyclicity of the series clearly exhibiting hourly and weekly patterns. The compound classifier discriminates normal price and price spike processes to separately predict values of those processes by different forecasting engines. The combined ARIMA and NN frameworks produce normal price forecasts by capturing linear and non linear patterns between target and exogenous variables. The k-NN is applied for the price spike value prediction. Time-varying model parameters allow capturing of localized trending of the series.

The methodology is evaluated with real data of Finnish day-ahead energy market. It can, however, be considered to have applicability for the entire Nordic region, as well as deregulated markets in other countries or even for financial markets, since the methodology addresses common statistical features of the price series.

2. Nordic Energy Market

The Nordic region has considerable experience with deregulated electricity markets. The Nordic electricity market was formed in 1993 in conjunction with deregulation of electricity markets in the region. The derivatives and energy markets were separated in 2002 to establish Nord Pool Spot, which currently operates in Norway, Denmark, Sweden, Finland, Estonia, Latvia and Lithuania.

The main goal of Nord Pool Spot is to balance the generation of electricity with the electricity demand, precisely and at an optimal price, so-called equilibrium point trading. The optimal price represents the cost of producing one kilowatt hour of power from the most expensive source needing to be employed in order to balance the system. All the employed generators are paid the same market price.

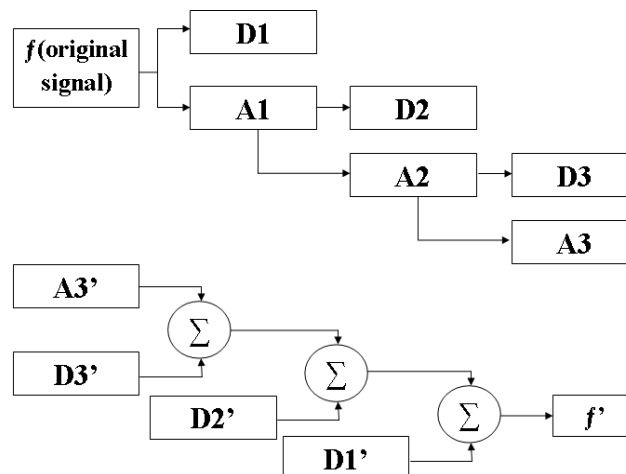
Two different physical operation markets are organized in Nord Pool Spot: Elspot and Elbas. Elspot is a day-ahead energy market in which market participants submit offers to sell, or bids to buy, physical electricity for the next day. Elbas is an intra-day energy market where trades are adjusted in the day-ahead market until one hour prior to delivery time.

3. Mathematical Framework

Before the prediction strategy is described, key features of WT, ARIMA, NN, feature selection technique and different classification frameworks are first introduced.

3.1. WT

When using classical statistical techniques, a stationary process is assumed for the data. For electricity price time series, the assumption of stationarity usually has to be rejected. One way to capture localized trending in the series is to apply models with time-varying parameters [23]. Another way to deal with non-stationarity is the use of mathematical transformations of an initial series. In many cases, information that cannot be readily seen in the time domain can be obtained in the frequency domain. Fourier transform (FT) and short time FT (STFT) are probably the most popular transforms and are used in many different areas, including many branches of engineering. However, these transforms provides poor time or frequency resolution. The WT was developed as an alternative approach to FT/STFT to overcome the resolution problem [24]. Wavelet analysis begins with selection of a proper wavelet (mother wavelet) and analysis of its translated and dilated versions [25]. It is advantageous to scale and translate the mother wavelet using defined scales and positions usually based on powers of two [26]. This technique is known as the discrete WT. An algorithm to implement discrete WT using filters has been developed by Mallat. Multiresolution via Mallat's algorithm is a procedure to obtain approximations (A) and details (D) from a given signal f [27]. In the reconstruction stage, these components can be assembled back into the original signal f (see Figure 1).

Figure 1. Multilevel decomposition (top) and reconstruction (bottom) processes.

In this paper, a Daubechies wavelet of order 5 is used as the mother wavelet to transform the price series into several wavelet subseries. This wavelet offers an appropriate trade-off between wavelength and smoothness, resulting in an appropriate behavior for the price forecast [19,26,28]. Three decomposition levels are considered, since this describes the price series in a more thorough and meaningful way [26].

3.2. Seasonal ARIMA

AutoRegressive Moving Average (ARMA) models form a class of time series models that are widely applicable in the field of time series forecasting [29]. In the case of linear trends and/or seasonal behavior, non-stationary time series processes can be transformed by differentiation of the series to make them stationary. The ARMA model, therefore, is transformed to an autoregressive integrated moving average (ARIMA) model [30]. To capture a linear trend and seasonality (diurnal cycle) in the time series, one-hour (regular) and 24-hour (seasonal) differencing is used.

The Box-Jenkins approach is utilized to build the ARIMA. The approach uses an iterative model building strategy consisting of four stages. In the first stage, the structure of the model is identified. Utilization of the autocorrelation function (ACF) and the partial ACF (PACF) of the sample data is a basic tool to identify the order of the ARIMA best model, which is then estimated by maximum likelihood in the second step. The parameters of the model are estimated such that an overall measure of errors is minimized. Goodness-of-fit is tested on the estimated model residuals in the third step. If the model is not adequate, a new tentative model should be identified. Forecast future outcomes are obtained in the fourth step [29].

3.3. NN

The NN, also called a multilayer perceptron (MLP), is a semi-parametric model and has been developed based on study of the brain functions and the nervous system. Perceptrons are arranged in layers with no connections inside a layer, and each layer is fully connected to preceding and following layers without loops. The first and last layers are called input and output layers, respectively. Other layers are hidden layers. Each layer, therefore, consists of a specific number of computational

elements, called neurons, which are connected to neurons in adjacent layers and capture complex non-linear phenomena. A sigmoid function is used in a hidden layer [31].

The procedure for developing NNs is as follows: data pre-processing; definition of the architecture and parameters; weights initialization; training until the stopping criterion is reached (number of iterations, sum of squares of error is lower than a pre-determined value); finding the network with the minimum error; and forecasting the future outcome.

The NN toolbox of MATLAB was selected for NN model building due to its flexibility and simplicity. The Levenberg-Marquardt (LM) algorithm was used in this study, which is an advanced optimization algorithm and one of the more efficient for training NNs. According to Kolmogorov's theorem, NN can solve a problem by using one hidden layer provided that it has a proper number of hidden neurons (N_h) [32]. Therefore, one hidden layer has been considered in the structure of all NNs utilized in this study.

3.4. Compound Classifier

The problem of the price spike occurrence prediction is stated as a classification problem that can be solved by a pattern recognition framework. The ultimate goal of pattern recognition is to discriminate the class membership of the observed novel objects with the minimum misclassification rate. It had been observed that even if one of the designs would yield the best performance, the sets of patterns misclassified by the different classifiers would not necessarily overlap. This suggests that different classifier designs potentially offer complementary information about the patterns to be classified which could be harnessed to improve the performance of the selected classifier. The idea behind use of the compound classifier presented in this paper is to avoid reliance on a single classifier. Various classifier combination schemes have been devised and it has been experimentally demonstrated that some of them consistently outperform a single best classifier [33]. The majority vote rule is applied to get an overall output (spike or non-spike) from the compound classifier.

The three individual classifiers used together in the compound classifier are a relevance vector machine (RVM), ensemble of bagged decision trees (DT) and probabilistic neural network (PNN). These methods are chosen because they provide probabilistic output (probability of class membership, e.g., probability of spike occurrence). The methods have been previously applied to several other applications with promising results [19,34,35].

3.4.1. RVM

RVM is a statistical learning technique based on Bayesian theory. It was developed for regression and classification problems. In RVM, the method to deal with non-linear data is to use a map function to map the training data from the input space into some high dimensional feature space, so the training data become linearly separable in the feature space. The related kernel function is used to avoid explicit knowledge of the high dimensional mapping [36].

Consider a set of example of input vectors $\{x_i\}_{i=1}^N$ along with a corresponding set of targets $t = \{t_i\}_{i=1}^N$. For classification problem, t_i should be 0 for class C_1 and +1 for class C_2 . The RVM constructs a logistic regression model based on a set of sequence features derived from the input patterns, *i.e.*:

$$p(C_1 | x) \approx \sigma\{y(x, w)\} \quad (1)$$

$$y(x, w) = \sum_{i=1}^N w_i F_i(x) + w_0 \quad (2)$$

where the basis function $F(x) = (F_1(x), F_1(x), \dots, F_N(x))^T = [1, K(x_i, x_1), K(x_i, x_2), \dots, K(x_i, x_N)]^T$, $w = (w_0, \dots, w_N)^T$ is a vector of weights, $\sigma\{y\} = (1 + \exp\{-y\})^{-1}$ is the logistic sigmoid link function and $K(x_i, x_j)_{j=1}^N$ are kernels terms. Assuming a Bernoulli distribution for $P(t | x)$, the likelihood can be written as:

$$P(t | w) = \prod_{n=1}^N \sigma\{y(x_n; w)\}^{t_n} [1 - \sigma\{y(x_n; w)\}]^{1-t_n} \quad (3)$$

To form a Bayesian training criterion, a prior distribution over the vector of model parameters or weights, $p(w)$ must be imposed. The RVM adopts a separable Gaussian prior, with a distinct hyper-parameter, $\alpha(w)$, for each weight:

$$p(w | \alpha) = \prod_{i=1}^N N(w_i | 0, \alpha_i^{-1}) \quad (4)$$

The optimal parameters of the model are then given by the minimiser of the penalized negative log-likelihood:

$$\log\{P(t | w)p(w | \alpha)\} = \sum_{i=1}^N [t_i \log y_i + (1 - t_i) \log(1 - y_i)] - \frac{1}{2} w^T A w \quad (5)$$

where $y_i = \sigma\{y(x_i, w)\}$ and $A = \text{diag}(\alpha)$ is a diagonal matrix with non-zero elements given by the vector of hyper-parameters. A detailed mathematical description of RVM is given in [37]. Here, we select a Gaussian radial basis function (RBF) kernel with its specific value of spread σ_{RVM} for application of RVM [29].

3.4.2. DT

Ensemble of DT creates a forest of a specific number of decision trees whose outputs are combined to make the overall output for the ensemble:

$$p(C | v) = \frac{1}{N} \sum_i^N p_i(C | v) \quad (6)$$

where N_{tree} is a number of trees, C is a class label, v is a feature vector, $p_i(C | v)$ is posterior probability generated by i^{th} tree.

Bagging is a method to develop improved estimating class probabilities from DT classification algorithm. Mathematical description of DT classifier and bagging method can be found in [38,39].

3.4.3. PNN

PNN is a radial basis network that is suitable for classification problems. PNNs are closely related to the Parzen window probability density function estimator [40]. The particular estimator used in this study is:

$$f_C(X) = \frac{1}{(2\pi)^{p/2} \sigma_{PNN}^p} \frac{1}{m} \sum_{i=1}^m \exp \left[-\frac{(X - X_{C_i})^T (X - X_{C_i})}{2\sigma_{PNN}^2} \right] \quad (7)$$

where i is a pattern number; m is a total number of training patterns; X_{C_i} is i^{th} training pattern of class C ; σ_{PNN} is a spread parameter; p is a dimensionality of measurement. Function $f_C(X)$ is simply the sum of small multivariate Gaussian distributions centered at each training sample. However, the sum is not limited to being Gaussian. It can, in fact, approximate any smooth density function.

A PNN is organized into a multilayered feed-forward network with four layers: the input layer (set of measurements), pattern layer (Gaussian functions), summation layer (average operation of the outputs from the second layer for each class) and output layer (a vote, selecting the largest value). Mathematical details of PNN can be found in [41]. Spread of the Gaussian RBF σ_{PNN} is an adjustable parameter of PNN. If spread is near zero, the network acts as a nearest neighbor classifier. As spread becomes larger, the designed network takes into account several nearby design vectors.

3.4.4. Probability Threshold

Prediction of price spike occurrence is a serious imbalanced classification problem (*i.e.*, the non-spike class has many more samples than the spike class). The probabilities of spike occurrence obtained from the above mentioned single classifiers are calculated for every input vector and then compared with a probability threshold V_0 . If the probability is larger than the threshold, a spike is predicted to occur, regardless of whether this probability is less than the probability of non-spikes. This modification is performed because many spikes occur when their occurrence probabilities are smaller than 50% [17].

3.5. k -NN

After having determined the probability of price spike occurrence, it is of considerable interest for market participants to be able to further predict the actual value of the price spike. A k -NN approach has been used for this task in this paper as in [18,42].

In the k -NN, from the training data samples, k -neighboring samples closest to the unknown sample are selected. Then the sum of weighted values of the k -closest samples is computed as the unknown sample's value. The Euclidean distance measure is employed here to calculate the closeness between two instances of the training data set. If $k = 1$ the instance is assigned to the class of its nearest neighbor.

3.6. Feature Selection Technique

Suitable selection of input variables plays a large role in the success of any forecast method. For price forecasts, apart from lagged values, many other input variables can be considered: available generation, fuel costs *etc.* [10,43,44]. The set of potential inputs may be too large for further use in a model. Thus, it is necessary to refine the initial set of potential inputs such that a subset of the most effective inputs is selected to be applied to the forecast engine [43].

The two-step feature selection algorithm consisting of both relevance and redundancy filters is used [43]. The ability to filter out redundant information from the set of the candidate features is the benefit of such a procedure *versus* a simple calculation of relevance value between target and

independent variables. A mutual information (MI) criterion is used within the feature selection algorithm to capture non-linearity of the original price signal [45].

In the feature selection technique applied, $SET_1 = \{x_1, x_2, \dots, x_t\}$ is supposed as a set of candidate inputs. For each feature $x_i \in SET_1$, its MI value with the target feature y (continuous or binary) is computed as $MI(x_i, y)$. If the $MI(x_i, y)$ value between a candidate variable and a target is greater than a pre-specified value V_1 , then this candidate is retained for further processing; otherwise it is filtered out:

$$MI(x_i, y) > V_1, 1 \leq i \leq t \quad (8)$$

In the second step, the set of the retained candidates is supposed as $SET_1 \subset SET_1$. For any two retained candidates $x_a, x_b \in SET_1$, their MI value supposed as the redundancy measure is computed. If the MI value between any two candidate variables (x_a and x_b) is smaller than a prespecified value V_2 , both variables are retained; otherwise, only the variable having the largest MI value with respect to the target [$MI(x_a, y)$ or $MI(x_b, y)$] is retained. For instance, for $x_a, x_b \in SET_1$:

$$MI(x_a, x_b) > V_2, 1 \leq a, b \leq t, a \neq b \quad (9)$$

The redundancy filtering process is repeated for all candidate inputs of SET_1 until no redundancy measure becomes greater than V_2 . The subset of candidate variables $SET_2 \subset SET_1$ that passed the redundancy filter is finally selected as best inputs by the proposed two-step feature selection algorithm.

4. Electricity Price Spike Definition

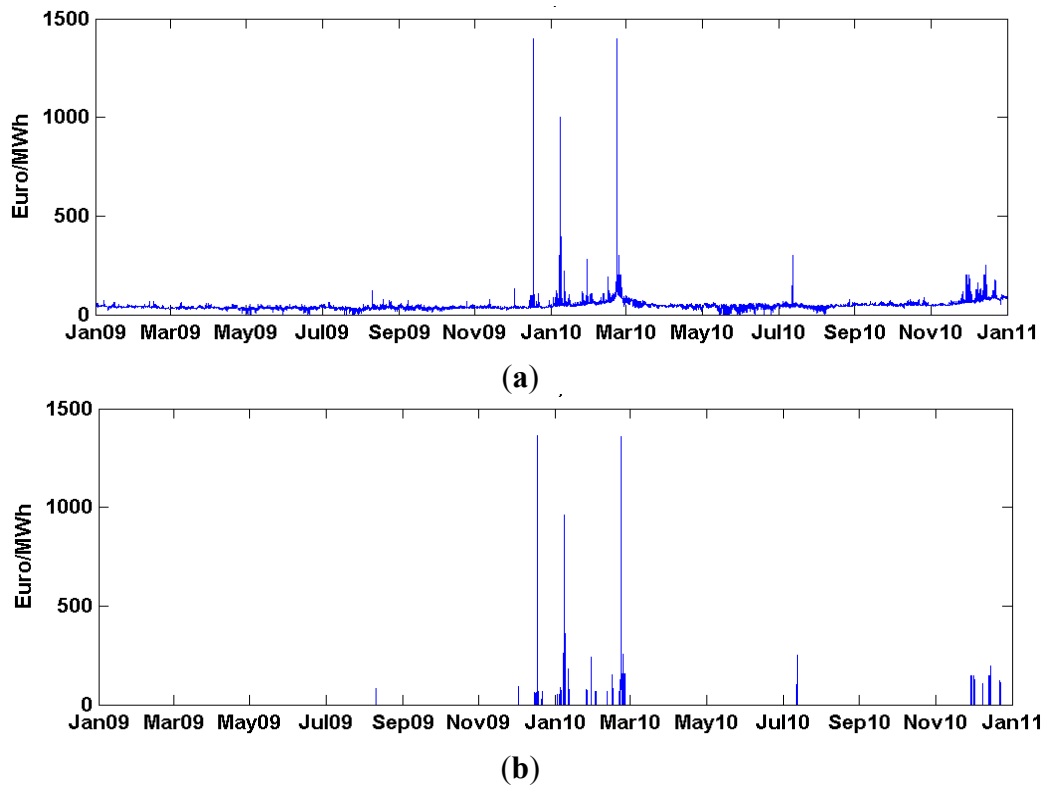
A spike is defined as a price that surpasses a specified threshold. Some authors suggest the use of fixed log-price change thresholds [46], a varying log-price range threshold [47], or a fixed threshold value for a whole price time series under consideration [19]. Here, the statistical method employed in [16] is used, where the spike threshold is calculated as $\mu + 3\sigma$. Notations μ and σ indicate the mean and standard deviation of the considered market price data. The threshold is time-varying and calculated on the basis of half-a-year price data before each day of the considered period, *i.e.*, of years 2009–2010, to capture evolving conditions of the market. All the prices exceeding this threshold are considered as spikes and extracted from the original price series of the Finnish day-ahead energy market of Nord Pool Spot over the period from 1 January 2009 to 31 December 2010 (see Figure 2). Table 1 shows the basic distribution parameters for normal prices and spikes in the Finnish day-ahead energy market of years 2009–2010.

It can be seen from Table 1 that spikes constitute about 1.0% of all the prices. However, their magnitude and unexpectedness cause them to have disproportionate significance in the energy markets.

Table 1. Basic statistics for normal spot prices and price spikes in terms of (euro/MWh).

Type	Number of observations	Mean	Std	Skewness	Kurtosis
Normal prices	1,7324	44.62	15.77	9.65	1.88
Spikes	196	240.98	256.68	13.61	3.30

Figure 2. (a) Original price data of the Finnish day-ahead energy market in years 2009–2010; (b) Extracted price spikes.

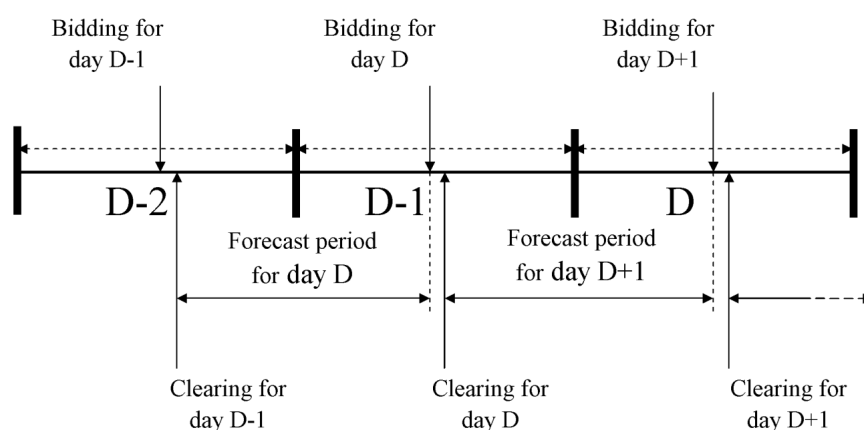


5. Proposed Method

5.1. Forecasting Framework

The time framework to forecast hourly day-ahead market prices in the Nordic day-ahead energy market is illustrated in Figure 3. The market spot price forecasts for day D are required on day $D-1$ (bidding hour: 12 a.m. CET on day $D-1$). The market spot price data for day $D-1$ are announced by the system operator and are available on day $D-2$ (clearing hour: around 1 p.m. CET on day $D-2$). The actual forecasting of day-ahead prices for day D can take place between the clearing hour for day $D-1$ of day $D-2$ and the bidding hour for day D of day $D-1$.

Figure 3. Time framework to forecast market prices for the Nordic day-ahead energy market.



In multistep ahead prediction, the predicted price value of the current step is used to determine its value in the next step, and this cycle is repeated until the price values of the whole forecast horizon are predicted.

5.2. Price Spike Module: Compound Classifier

First, the set of candidate inputs for the compound classifier is constructed. Values of original price series lagged up to 200 h before a forecast hour are considered among the candidate inputs. If the period of the study is extended further, the results are not affected seriously, that is, the relation of the current price with the price of much more than one week ago is very small [48]. Thus, lagged hours take into account short-run trend, daily and weekly periodicity of the electricity time series itself and external explanatory time series [19,43,49].

Electricity demand and supply are among the candidate inputs for the compound classifier since the relations of these variables are known to drive the movement in the price spikes to a large extent [17]. Therefore, total electricity generation (*i.e.*, internal supply (*sup*)) and electricity demand (*d*) in Finland, both lagged up to 200 h before a forecast hour, are selected. Despite its general importance, demand and supply forecasts will not be the focus of the present paper. Here, a WT + ARIMA model [26] is implemented to predict supply. The effect of weather variation is incorporated in a WT + ARIMAX model to predict demand. Atmospheric temperature is chosen as an indicator of weather variability.

Approximation ($A3_p$) and detail ($D1_p$) price wavelet components of a price series, both lagged up to 200 h before a forecast hour, are also candidate inputs for the compound classifier. In [19], a high correlation of a spiky price series with these wavelet components has been described. Hourly (ind_h), daily (ind_d), and seasonal (ind_s) indices are considered as candidate inputs to indicate the temporal effect.

The ARIMA is used as the initial price forecasting model and produces preliminary day-ahead predictions for all price wavelet subseries over the forecast period. For more clarity, prices and the wavelet components predicted by the ARIMA are additionally indexed as “arima” in the paper. For instance, price value predicted by the initial forecasting model at hour *h* is notated as $p_{arima,h}$ and used in the candidate input set for the compound classifier.

Finally, the initial set of candidate inputs for the compound classifier, *i.e.*, for each single classifier, includes both historical and forecasted features of both wavelet and time domains. For instance, the 1008 candidate inputs to predict possibility of spike occurrence at hour *h* are $\{p_{arima,h}, p_{h-1}, \dots, p_{h-200}; d_h, \dots, d_{h-200}; sup_h, \dots, sup_{h-200}; A3_{p_arima,h}, A3_{p,h-1}, \dots, A3_{p,h-200}; D1_{p_arima,h}, D1_{p,h-1}, \dots, D1_{p,h-200}; ind_h; ind_d; ind_s\}$.

5.3. Normal Price Module

If the forecast sample is classified as a non-spike, the normal price module is activated. All electricity price spikes are extracted from the original training price series and replaced by the corresponding mean price value to form new normal price series.

Next, the set of candidate inputs for the normal price module is constructed. Firstly, the new normal price series is decomposed into four wavelet components of normal price series. Although the wavelet components are obtained by decomposition of the normal price signal, historical values of the original normal price series are considered among the candidate inputs of each price wavelet component, since it is still possible that some characteristics of the price are better highlighted in the original time domain [45]. Historical and forecasted electricity demand data are also considered within the candidate

input set. The ARIMA produces preliminary day-ahead forecasts for all wavelet subseries of the normal price series. Finally, the candidate input set for each wavelet subseries of normal price includes forecasted and lagged price and demand values of these subseries (e.g., $A3_d$ is an approximation wavelet subseries of demand to predict $A3_p$) plus original normal prices (p) lagged up to 200 h before a forecast day. For instance, the 602 candidate inputs to predict approximation normal price wavelet component at hour h ($A3_{p,h}$) are $\{A3_{p_arima,h}, A3_{p,h-1}, \dots, A3_{p,h-200}; A3_{d,h}, \dots, A3_{d,h-200}; p_{h-1}, \dots, p_{h-200}\}$.

5.4. Price Spike Module: k -NN

If the forecast sample is classified as a spike, the price spike module is activated. The target set to train a k -NN is formed by the price spike samples extracted from the original training price series. The k -NN uses the set of candidate inputs similar to the one utilized for the compound classifier.

5.5. Search Procedure to Tune Model Parameters

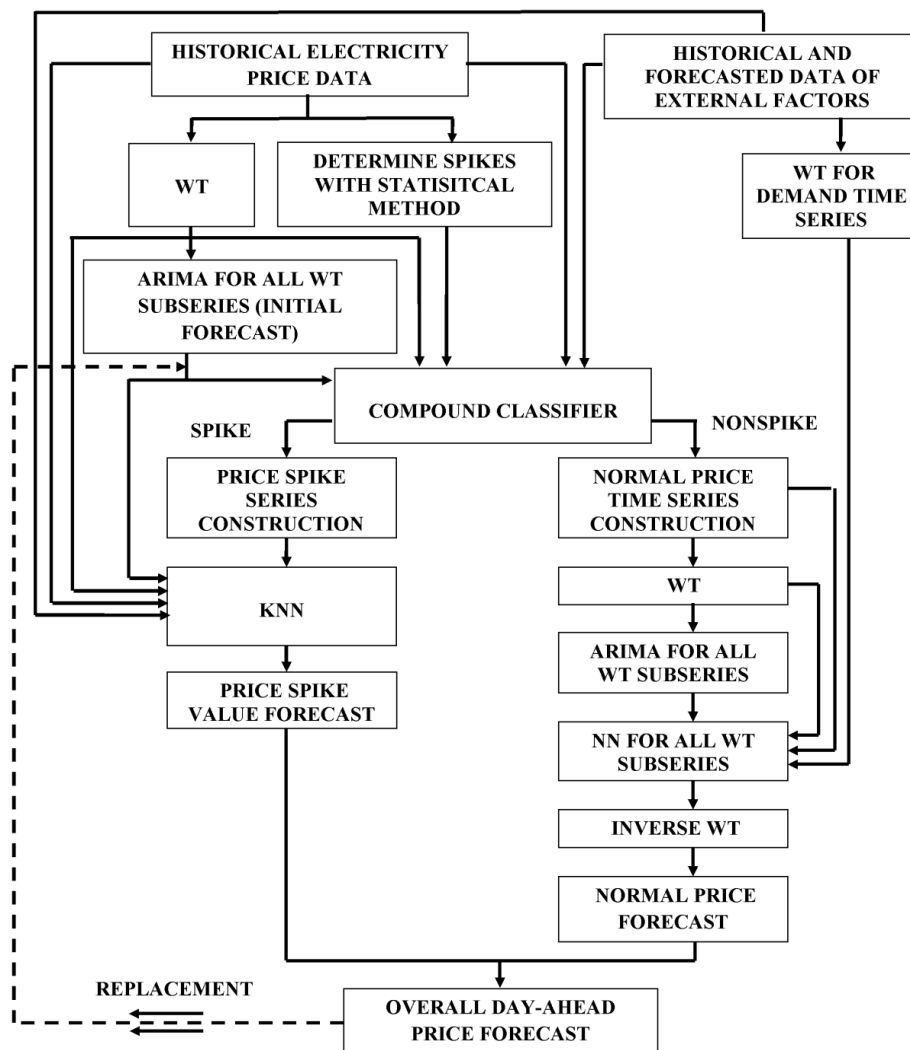
Usually, the adjustable parameters of a forecasting model are selected based on past experience in the study. However, as each energy market has characteristics of its own, selection of optimal model parameters is an open area of a research. In this work, an analytical iterative search procedure is realized. It can automatically adjust the parameters of a forecasting model on a selected validation set with minimum reliance on the heuristics. The procedure is used to select optimal set of inputs by defining the threshold values V_1 , V_2 and parameter settings separately for NN (N_h), k -NN (k), RVM (σ_{RVM} and V_0), DT (N_{tree} and V_0) and PNN (σ_{PNN} and V_0).

There are four adjustable parameters when the proposed search procedure is applied for RVM. The procedure is outlined below:

- A. Initial values for V_0 , V_1 , V_2 and σ_{RVM} for RVM are set.
- B. Using the selected inputs, training samples are constructed. The classifier is trained and produces forecast on the validation set. The corresponding validation error is evaluated and stored.
- C. Each adjustable parameter is varied by turn at a neighborhood around its previously selected value, while three remaining parameters are kept constant. A fixed radius of neighborhood ($\pm 25\%$ of the previously selected value) is considered in the local search. For each value of the varied parameter in the neighborhood, training of the classifier is repeated and validation error is evaluated and stored. The value of the varied parameter resulting in the least validation error is selected and fixed. When only the first cycle described in the current step is executed, this cycle is repeated again. This modification is made to avoid a local minimum trap in the search procedure. Therefore, if the procedure misses the optimum solution in one cycle, it may find the optimum point in the next cycle.
- D. If the selected values of the adjustable parameters in the current cycle are the same as their previous values, the search procedure is terminated. Otherwise go to step 3.

5.6. Forecast Strategy

The proposed forecast strategy can be summarized by the following step-by-step algorithm, shown also in Figure 4.

Figure 4. Procedure of the proposed method.

- A. Original electricity price series is decomposed into four subseries by the wavelet transform.
- B. ARIMA models are built to predict the future values of the price wavelet subseries 24 h ahead.
- C. The compound classifier is activated:
 - C.1. The set of candidate inputs for each single classifier is constructed.
 - C.2. The threshold values for the proposed two-step feature selection algorithm and parameter settings for each single classifier are fine-tuned.
 - C.3. Each classifier is trained and predicts the spike occurrence possibility 24 h ahead.
 - C.4. Final output from the compound classifier is formed in a majority voting scheme.
- D. For all test samples forecasted as non-spikes, the normal price module is activated.
 - D.1. All spike samples are extracted from the original training price series. The new adjusted normal price series is decomposed into four wavelet subseries.
 - D.2. ARIMA models are built to predict the future values of the normal price wavelet subseries 24 h ahead.
 - D.3. The set of candidate inputs to predict each normal price wavelet subseries by NN is constructed.

- D.4. V_1 , V_2 and N_h for NN's are fine-tuned.
- D.5. With the selected parameters, NNs are trained and predict normal price wavelet subseries 24 h ahead.
- E. For all test samples forecasted as spikes the price spike module is activated.
 - E.1. All spike samples previously extracted from the original training price series are formed into spike series and used as targets to train the k-NN model.
 - E.2. The set of candidate inputs for the k-NN model is constructed.
 - E.3. V_1 , V_2 and k for the k-NN are fine-tuned.
 - E.4. With the selected parameters, the k-NN is trained and predicts price spike value.
- F. The overall electricity price forecast is formed as a joint output from the normal price and price spike modules.
- G. The overall price forecast (original and transformed into price wavelet components) replaces the predictions produced by the initial forecasting model for the current forecast day, since it is expected that electricity prices predicted by the separate forecasting frameworks have more accuracy. After replacement, the forecasting cycle is repeated as shown in Figure 4 until no difference in the overall electricity price forecast output of two successive iteration steps is observed.

6. Case Study

For examination of the proposed method, the real hourly data of the Finnish day-ahead energy market are considered. The electricity price, demand and supply historical data over the period from November 2008 to December 2009 are used to establish the initial training data sample set. The data over the period from January 2010 to December 2010 are used as the test set.

6.1. Training Phase

Training periods for the forecasting models of the normal price and price spike modules are different. As recommended in [26,43], a 50 days training period preceding the forecast day is considered for the NNs of the normal price module. It should be borne in mind that the price series have local trends since market conditions evolve with time, and, hence, use of a long training period may result in significant inaccuracies.

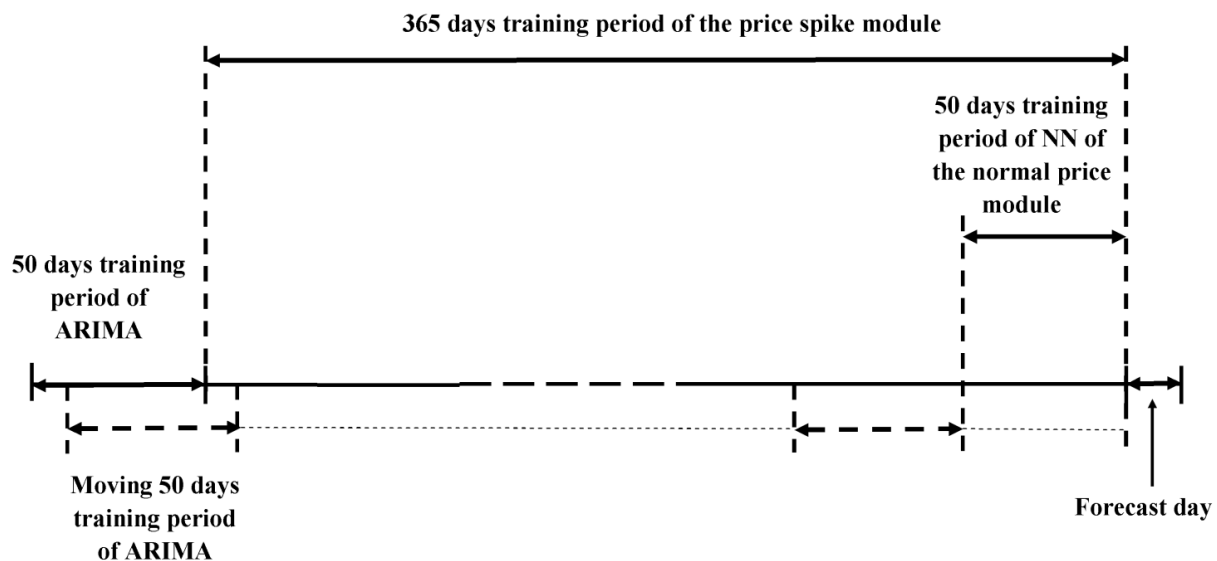
However, there are only few price spike samples in the whole data set (see Table 1). Unlike normal price prediction, in order to get a sufficient number of spike samples to train the model, a longer price series period is required. Hence, 365 days preceeding the forecast day are considered for the price spike module (the compound classifier and the k-NN model).

Since the forecasting models of the normal price and price spike modules have the inputs preliminarily predicted by other models their training periods are extended to comprise two consecutive periods: a moving training period for the preliminary model and the training period of the main model.

As a result, to predict normal prices or price spikes, a day denoted by D is considered in the corresponding second training period. Values of prices for this day are assumed to be unknown. The preliminary ARIMA models are trained by the historical data of the 50 days proceeding hour 1 of

day D and predict price wavelet subseries of day D . To improve the performance of the ARIMA forecast process for each day of the second training period ($D = 1, \dots, 50$ for NNs or $D = 1, \dots, 365$ for the price spike module), the ARIMA models are trained by the immediately previous 50-days period. This process is repeated until forecasts from the ARIMA models are obtained for all days of the corresponding second training period (see Figure 5).

Figure 5. Historical data required for the training of the normal price and price spike modules to produce overall price forecast on a single forecast day.



6.2. Validation Phase

The 24 h before the forecast day are removed from the training set of the NNs of the normal price module and used as the validation data set. Then, the NNs are trained by the remaining training samples. Adjusted parameters are fine-tuned on the validation data set. For the price spike module, all adjustable parameters of the classification approaches are fine-tuned by a 10-fold cross-validation technique applied for a whole training data set.

6.3. Numerical Results

The obtained results of the two-step feature selection algorithm implemented for the compound classifier, the k-NN and the NN to predict prices in the Finnish day-ahead energy market for a single forecast day, 5 January 2010, are presented in Tables 2 and 3. Since electricity price spikes have a very volatile stochastic nature with respect to the normal price time series, regular and periodic behaviour of price spikes are not so obvious (see Table 2). Otherwise, variables of the short-run trend (e.g., $A3_{p,h-1}$, $D3_{p,h-2}$), daily periodicity (e.g., $A3_{p,h-25}$, $D3_{p,h-24}$) and weekly periodicity (e.g., $A3_{p,h-169}$, $A3_{d,h-169}$) are among the selected input features to forecast normal price wavelet components (see Table 3).

Table 2. Selected inputs for the three classification approaches of the compound classifier and the k-NN for a single forecast day, 5 January 2010.

Engine	$V_0/V_1/V_2$	Parameter	Selected candidates
RVM	0.43/0.46/0.64	$\sigma_{RVM} = 0.13$	$A3_{p_arima,h}, A3_{p,h-1}, A3_{p,h-2}, A3_{p,h-4}, A3_{p,h-5}, A3_{p,h-6}, A3_{p,h-7},$ $D1_{p_arima,h}, D1_{p,h-1}, D1_{p,h-2}, D1_{p,h-3}, p_{h-1}, p_{h-2}, p_{h-3}, p_{h-4}, d_h, d_{h-2},$ $d_{h-46}, d_{h-72}, sup_h, ind_h, ind_d$
PNN	0.47/0.50/0.78	$\sigma_{PNN} = 0.03$	$A3_{p_arima,h}, A3_{p,h-1}, A3_{p,h-2}, A3_{p,h-3}, A3_{p,h-4}, A3_{p,h-5}, A3_{p,h-6}, A3_{p,h-22},$ $D1_{p_arima,h}, D1_{p,h-1}, D1_{p,h-2}, D1_{p,h-3}, D1_{p,h-4}, D1_{p,h-5}, p_{h-2}, p_{h-3}, p_{h-4},$ $d_{h-2}, d_{h-21}, d_{h-22}, sup_{h-2}, ind_h, ind_d, ind_s$
DT	0.42/0.48/0.61	$N_{tree} = 100$	$A3_{p_arima,h}, A3_{p,h-1}, A3_{p,h-2}, A3_{p,h-4}, A3_{p,h-5}, A3_{p,h-6}, A3_{p,h-7}, D1_{p,h-1},$ $D1_{p,h-2}, D1_{p,h-3}, p_{h-1}, p_{h-2}, p_{h-3}, p_{h-4}, p_{h-5}, d_h, d_{h-4}, d_{h-19}, d_{h-69}, d_{h-73},$ ind_h, ind_d
k-NN	-/0.45/0.56	$k = 3$	$A3_{p_arima,h}, A3_{p,h-2}, A3_{p,h-9}, A3_{p,h-15}, A3_{p,h-21}, D1_{p,h-2}, D1_{p,h-5},$ $D1_{p,h-7}, D1_{p,h-8}, D1_{p,h-16}, p_{h-1}, p_{h-3}, p_{h-7}, p_{h-52}, d_{h-190}, ind_h, ind_d, ind_s$

Table 3. Inputs selected by the two-step feature selection to predict the normal price wavelet components for the NN for a single forecast day, 5 January 2010.

Subseries	$N_h/V_1/V_2$	Selected candidates
$A3_p$	4/0.52/0.71	$A3_{p_arima,h}, A3_{p,h-1}, A3_{p,h-3}, A3_{p,h-4}, A3_{p,h-16}, A3_{p,h-21}, A3_{p,h-25}, A3_{p,h-72}, A3_{p,h-97},$ $A3_{p,h-121}, A3_{p,h-144}, A3_{p,h-169}, A3_{d,h-8}, A3_{d,h-10}, A3_{d,h-11}, A3_{d,h-42}, A3_{d,h-91}, A3_{d,h-98},$ $A3_{d,h-141}, A3_{d,h-169}, p_{h-72}, p_{h-95}, p_{h-97}, p_{h-120}$
$D3_p$	7/0.47/0.81	$D3_{p_arima,h}, D3_{p,h-1}, D3_{p,h-2}, D3_{p,h-11}, D3_{p,h-24}, D3_{p,h-48}, D3_{p,h-60}, D3_{p,h-96}, D3_{d,h-12},$ $D3_{d,h-47}, D3_{d,h-71}, D3_{d,h-143}$
$D2_p$	4/0.41/0.74	$D2_{p_arima,h}, D2_{p,h-1}, D2_{p,h-7}, D2_{p,h-8}, D2_{p,h-24}$
$D1_p$	6/0.15/0.85	$D1_{p_arima,h}, D1_{p,h-6}, D1_{p,h-24}, D1_{p,h-30}, D1_{p,h-48}, D1_{p,h-72}, D1_{p,h-94}, D1_{p,h-120}, D1_{p,h-157}$

In this paper, Adapted Mean Average Percentage Error (AMAPE) proposed in [10] was considered to evaluate the forecast results:

$$AMAPE = \left(\sum_{t=1}^T \left| P_{iACTUAL} - P_{iFOREC} \right| / \left(\sum_{t=1}^T P_{iACTUAL} / T \right) \right) / T \cdot 100\% \quad (10)$$

where $P_{iACTUAL}$ and P_{iFOREC} are actual and forecast values of hour i , respectively; and T is the number of predictions.

In addition, two performance measures that are spike prediction accuracy and spike prediction confidence proposed in [17] are used to reliably assess the performance of the compound classifier.

Spike prediction accuracy is a ratio of the number of correctly classified spikes (N_{corr}) to the number of actual spikes (N_{sp}):

$$Spike \text{ prediction accuracy} = (N_{corr} / N_{sp}) \cdot 100\% \quad (11)$$

This measure was introduced because the ability to correctly predict spike occurrence is the subject of greatest concern.

Spike prediction confidence aims to account for the uncertainties and risks carried within the forecast. Spike prediction confidence is described as:

$$\text{Spike prediction confidence} = (N_{corr} / N_{as_sp}) \cdot 100\% \quad (12)$$

where N_{corr} is the number of correctly classified spikes and N_{as_sp} is the number of observations classified as spikes. As the classifier may misclassify some nonspikes as spikes, this definition is used to assess the percentile in which the classifier makes this kind of a mistake.

Only few research works have considered price forecasting in the Finnish day-ahead energy market and it was not possible to find price forecast methods considering the above-mentioned test period for price forecast. Therefore, the overall accuracy of the proposed method is compared with some of the most popular price forecast techniques applied for case studies of energy markets of other countries: seasonal ARIMA [5,44,50]; WT+ARIMA [26,28]; NN [6,50]; WT + NN [11]. Additionally, WT + ARIMA + NN, which has not been found in the literature is among competitive techniques. To demonstrate the efficiency of the proposed methodology, its obtained results for the Finnish day-ahead energy market in year 2010 are shown in Table 4 with corresponding results obtained from five other prediction techniques.

In the WT + NN and WT + ARIMA + NN models separate NNs with the LM algorithm are applied for each price wavelet component. For a fair comparison, NN, WT + NN and WT + ARIMA + NN have historical and forecasted demand data among the candidate inputs. Feature selection analysis based on the proposed two-step feature selection is utilized for all examined models. The adjustable parameters of the competing models are fine-tuned by the proposed search procedure. It should be noted that among the competing examined models, only the WT + ARIMA + NN has preliminarily predicted price values in its set of candidate inputs *i.e.*, the NN uses predictions from ARIMA as the candidate input.

As seen from Table 4, the AMAPE values corresponding to the proposed strategy are lower than the values obtained from other examined methods. The accuracy improvement of the proposed method with respect to seasonal ARIMA, WT + ARIMA, NN, WT + NN, and WT + ARIMA + NN in terms of AMAPE is 45.88% $[(1 - 8.08/14.93) \times 100\%]$, 19.44% $[(1 - 8.08/10.03) \times 100\%]$, 35.46% $[(1 - 8.08/12.52) \times 100\%]$, 32.55% $[(1 - 8.08/11.98) \times 100\%]$, and 16.36% $[(1 - 8.08/9.66) \times 100\%]$, respectively. It can also be seen that the use of WT results in an improvement in the model accuracy. This improvement in ARIMA in comparison with WT+ARIMA in terms of AMAPE is 32.82% $[(1 - 10.03/14.93) \times 100\%]$. For the NN in comparison with WT+NN, this value is 4.31% $[(1 - 11.98/12.52) \times 100\%]$. The results also confirm the efficiency of the hybrid methodology with linear and nonlinear modeling capabilities (WT + NN *versus* WT + ARIMA + NN) where the improvement is 19.37% $[(1 - 9.66/11.98) \times 100\%]$.

Table 4. AMAPE (%) obtained from different techniques for price forecasts in the Finnish day-ahead energy market of year 2010.

Type	Seasonal ARIMA	WT + ARIMA	NN	WT + NN	WT + ARIMA + NN	Proposed method
Normal	10.53	7.53	8.17	8.01	7.18	5.89
Spikes	55.76	40.51	46.33	44.22	37.72	32.91
Overall	14.93	10.03	12.52	11.98	9.66	8.08

It is expected that implementation of the proposed iteration strategy increases the accuracy of the overall price prediction. Detailed results of the proposed iteration strategy for the four test weeks of the

Finnish day-ahead energy market of year 2010 are shown in Table 5. These test weeks are related to dates 1–7 January 2010, 8–14 January 2010, 29 January–4 February 2010, 5–11 February 2010, respectively, and indicate periods of high volatility in energy price series. Iteration 0 in Table 5 represents the obtained results from the initial forecasting model.

Table 5. Accuracy of the proposed iteration procedure in terms of AMAPE (%) for the four test weeks of year 2010.

Iteration	Week1	Week2	Week5	Week7
0	17.46	37.27	13.49	10.87
1	12.56	26.24	7.96	6.93
2	9.50	25.16	7.24	6.81
3	9.41	-	-	6.59

As seen from Table 5, the iteration procedure converges in at most three cycles and the prediction error for the four test weeks at the end of the iterative forecast process with respect to Iteration 1 is improved by 13% on average. In addition, the performance of the proposed compound classifier is compared with selected single classifiers and other techniques: Naïve Bayesian [17], SVM [17], PNN [19], RVM [34], and DT [35]. N_{corr} and N_{as_sp} for the Finnish day-ahead energy market of year 2010 are presented in the second and third columns of Table 6, respectively. Corresponding spike prediction accuracy and confidence in terms of percentage are given in the fourth and fifth columns of Table 6. Candidate inputs of all alternative classifiers are similar to the candidate input set of the compound classifier and refined by the proposed two-step feature selection. All preliminarily predicted price variables which are among the input sets of each competing classifier are predicted by ARIMA model. This action is similar to the case when spike occurrence is predicted using forecasts from the initial forecasting model.

Table 6. N_{corr} , N_{as_sp} , occurrence prediction accuracy and confidence in terms of percentage (%) for price spike classification in the Finnish day-ahead energy market of year 2010.

Engine	ARIMA as a preliminary forecasting model				Final iteration step of the proposed methodology			
	Predicted		Accuracy	Confidence	Predicted		Accuracy	Confidence
	N_{corr}	N_{as_sp}			N_{corr}	N_{as_sp}		
Bayes	124	247	68.13	50.20	-	-	-	-
SVM	120	174	65.93	68.79	-	-	-	-
PNN	112	155	61.54	72.26	147	161	80.77	91.30
RVM	119	168	65.38	70.83	163	190	89.56	85.79
DT	122	166	67.03	71.08	152	179	83.51	84.92
Comp.	122	152	67.03	77.63	162	174	89.01	93.10

To justify the proposed iteration strategy particularly for the price spike occurrence forecast, N_{corr} and N_{as_sp} , accuracy and confidence measures obtained from the compound classifier on the final iteration step of the proposed methodology are shown in the sixth, seventh, eighth and ninth columns of Table 6, respectively. Total number of actual spike samples in the testing data set is 182. From the

obtained results given in Table 6, it can be seen that the use of the iteration strategy results in a notable accuracy improvement of price spike occurrence prediction. Only RVM has slightly better spike prediction accuracy than the compound classifier, while the compound classifier has considerably better spike prediction confidence than RVM.

Table 7 shows the results obtained from each single classifier and the compound classifier itself on the final iteration step. The set of actual price spike test samples of year 2010 are divided according to their price value intervals (see the second column of Table 7). Large price spikes with values varying between 300 and 1500 euro/MWh constitute around 15% of all the spike samples. Because of their values and stochastic character, they are extremely important for all market participants. All the classifiers presented in Table 7 are able to correctly discriminate all the large spike samples of the test period. The accuracy of the examined classifiers varies in the prediction of price spike samples with values between 85 and 300 euro/MWh.

Table 7. Results obtained by the compound classifier for different price spike intervals.

Price interval, euro/MWh	Number of Actual Spikes	Number of predicted spikes			
		PNN	RVM	DT	Compound
85–150	66	50	55	53	55
150–300	87	68	79	70	78
300–500	18	18	18	18	18
500–1000	1	1	1	1	1
1000–1500	10	10	10	10	10
TOTAL	182	147	163	152	162

For a more detailed representation of the performance of the proposed forecast strategy and separately for price spike occurrence on the whole test year, their results for all the weeks of year 2010 are shown in Table 8. There are six measures given for all test weeks of the day-ahead Finnish energy market of 2010: AMAPE, N_{sp} , N_{corr} , N_{as_sp} , accuracy and confidence of the spike forecast.

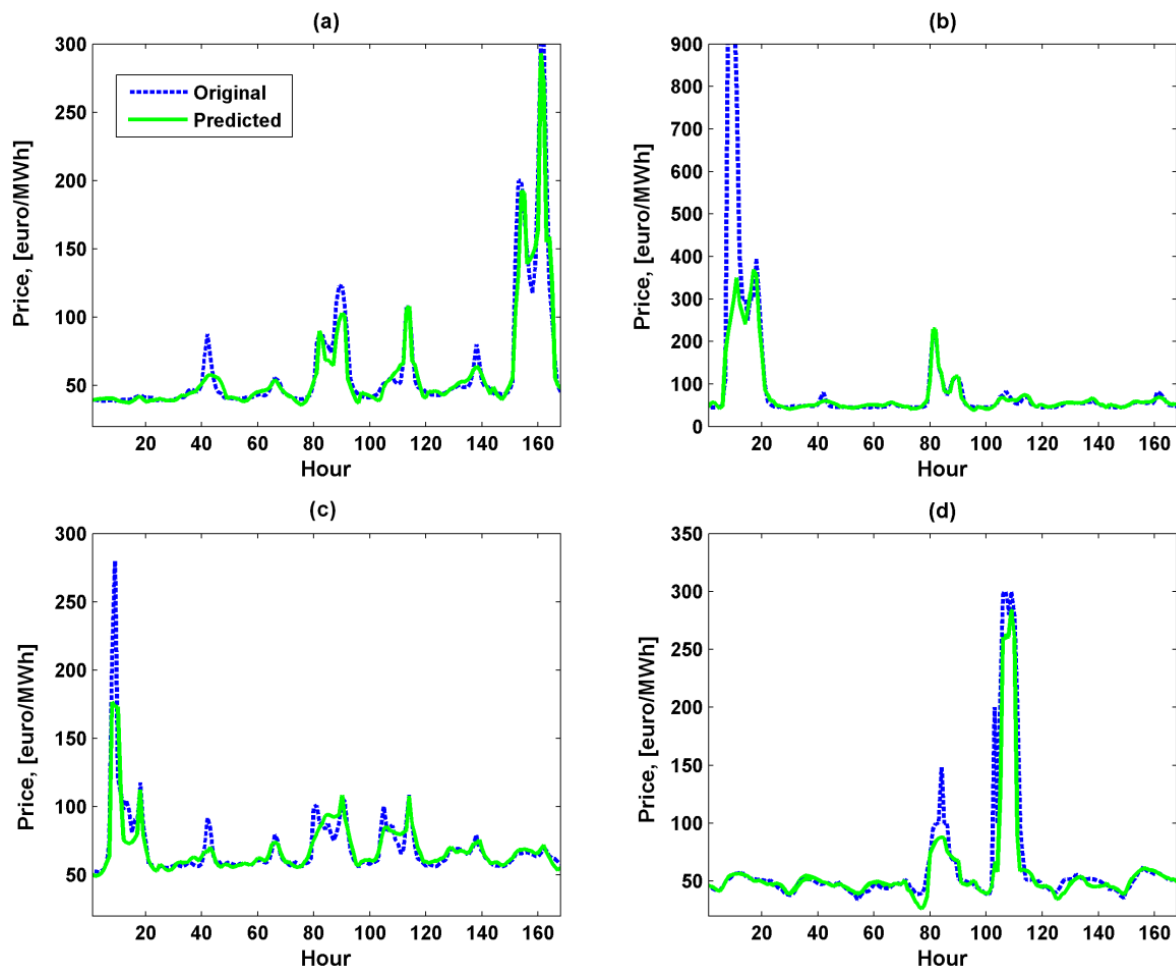
As can be seen from Table 8, price forecasts of the weeks related to a winter season (December–February), *i.e.*, weeks 1–8 and 48–52 of year 2010, have higher prediction error with respect to price forecasts related to other yearly seasons. The performance of the forecasting model is worse during the winter season, due to extreme price volatility, reflected in price spikes, which is caused by a number of complex factors and exists during periods of market stress. These stressed market situations are generally associated with extreme meteorological events and unusually high demand. However, in light of the fact that price spike values are highly stochastic, the achieved forecast accuracy level is fairly good and provides market participants with an ability to analyze spikes and thus manage their risks. Moreover, as can be seen from Table 8, occurrence of price spikes generally existing in the winter period is predicted by the proposed methodology with high accuracy and confidence. In this context, price spike prediction can be considered as a forecasting of a price volatility rather than exact price value. In order to graphically illustrate the price forecast performance of the proposed methodology and emphasize its ability to capture spikes, the forecasted and actual signals for the four selected spiky weeks (1,2,5 and 28 in Table 8) of the Finnish day-ahead energy market of year 2010 are shown in Figure 6.

Table 8. Obtained results from the proposed forecasting methodology for each week of 2010.

Week	1	2	3	4	5	6
MAPE	9.41	25.16	6.31	5.75	7.24	4.51
$N_{sp}/N_{corr}/N_{as_sp}$	23/21/22	22/22/22	0/0/0	9/8/8	7/7/7	1/0/0
Accur./Conf.	91.30/95.45	100/100	-	88.89/100	100/100	0/-
Week	7	8	9	10	11	12
MAPE	6.59	30.75	6.49	6.11	4.76	3.55
$N_{sp}/N_{corr}/N_{as_sp}$	5/5/5	44/39/39	2/2/2	0/0/0	0/0/1	0/0/1
Accur./Conf.	100/100	77.27/97.14	100/100	-	-/0	-
Week	13	14	15	16	17	18
MAPE	2.84	2.99	3.31	5.07	6.11	6.88
$N_{sp}/N_{corr}/N_{as_sp}$	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0
Accur./Conf.	-	-	-	-	-	-
Week	19	20	21	22	23	24
MAPE	7.43	15.51	6.35	8.03	7.23	6.46
$N_{sp}/N_{corr}/N_{as_sp}$	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0
Accur./Conf.	-	-	-	-	-	-
Week	25	26	27	28	29	30
MAPE	6.15	7.23	4.26	12.82	4.56	5.38
$N_{sp}/N_{corr}/N_{as_sp}$	0/0/0	0/0/0	0/0/0	9/7/7	0/0/2	0/0/0
Accur./Conf.	-	-	-	77.78/100	-/0	-
Week	31	32	33	34	35	36
MAPE	7.58	6.34	3.06	3.14	4.99	2.19
$N_{sp}/N_{corr}/N_{as_sp}$	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0
Accur./Conf.	-	-	-	-	-	-
Week	37	38	39	40	41	42
MAPE	3.64	2.65	3.64	2.43	3.83	4.09
$N_{sp}/N_{corr}/N_{as_sp}$	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0	0/0/0
Accur./Conf.	-	-	-	-	-	-
Week	43	44	45	46	47	48
MAPE	5.91	3.14	2.26	2.83	4.03	17.12
$N_{sp}/N_{corr}/N_{as_sp}$	0/0/1	0/0/0	0/0/0	0/0/0	0/0/0	12/7/8
Accur./Conf.	-	-	-	-	-	58.33/87.50
Week	49	50	51	52		
MAPE	8.37	11.02	5.70	4.22		
$N_{sp}/N_{corr}/N_{as_sp}$	3/3/3	35/33/34	9/8/11	0/0/1		
Accur./Conf.	100/100	82.86/96.67	88.89/72.73	-/0		

As can be seen in Figure 6, all the forecasted price curves acceptably follow the actual curves. The proposed methodology based on a hybrid iterative strategy is able to capture essential features of the given price time series: non-constant mean, cyclicity, exhibiting daily and weekly patterns, major volatility and significant outliers. This ability results in the superiority of the proposed methodology over all the examined alternative techniques.

Figure 6. Real and predicted prices for the four weeks with prominent spikes of the Finnish energy market of year 2010: (a) Week 1; (b) Week 2; (c) Week 5; (d) Week 28.



The total running time to set up the proposed separate forecasting strategy including its normal price module, price spike module, and iterative prediction process for the first forecast day is about 42 h since price predictions produced by the initial forecasting model are required over the period up to 365 days. The running time of the training and prediction procedures for the next forecast days after the first one is significantly lower (about 50 min) and considered suitable for day-ahead energy market operation. All the competitive non-separate forecasting approaches examined for price prediction have lower computation costs than the proposed separate forecasting strategy but are outperformed by the proposed strategy in terms of forecasting accuracy. The prediction accuracy is a crucial concern for a forecasting method (as far as the computation time is reasonable). The PNN and RVM classifiers of the compound classifier have relatively lower computational costs than the alternative back-propagation NN and SVM, respectively. The training process of the PNN is carried out through one run of each training sample unlike the back-propagation algorithm. The RVM is faster than the SVM in decision speed, as the RVM has a much sparser structure (the number of relevant vectors *versus* the number of support vectors). The computation times to set up the proposed and competitive forecasting strategies are measured on a hardware platform comprising an Intel Core i5 2.40 GHz processor (Intel Corporation, Santa Clara, CA, USA) and 3.24 GB RAM. All computer codes are provided by the MATLAB (MathWorks, Natick, MA, USA) and R (R Development Core Team, Auckland, New Zealand) software packages.

7. Conclusions

In this paper, an iterative forecasting methodology for prediction of prices in the day-ahead energy market is proposed. The proposed method is composed of two modules separately applied for normal prices and price spikes prediction. The forecasting performance of the proposed method is compared with the corresponding values of the most popular frameworks for price prediction and separately for price spike occurrence prediction. The proposed method generally outperforms other competing forecasting methods due to its ability to capture different distinct features of the given price time series and incorporate an iteration strategy to separately predict two processes: normal prices and price spikes.

It should be noted that many other exogenous variables can be considered in candidate sets for feature selection, such as fuel costs and meteorological information, but this is a topic for future research. There is a clear need for a more accurate method for price spike value prediction. Possible methods which could be based on NNs or RVM regression approaches will be considered in future. Moreover, investigation of the energy costs sensitivity to price forecast accuracy across different market participants is a topic for further research.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bunn, D.W. *Modelling Prices in Competitive Electricity Markets*; Wiley: New York, NY, USA, 2004; pp. 1–17.
2. Bastian, J.; Zhu, J.; Banunarayanan, V.; Mukerji, R. Forecasting energy prices in a competitive market. *Comp. App. Power* **1999**, *12*, 40–45.
3. Zhang, H.; Gao, F.; Wu, J.; Liu, K.; Liu, X. Optimal bidding strategies for wind power producers in the day-ahead electricity market. *Energies* **2012**, *5*, 4804–4823.
4. Schneider, S. Power spot price models with negative prices. *J. Energy Mark.* **2012**, *4*, 77–102.
5. Contreras, J.; Espínola, R.; Nogales, F.J.; Conejo, A.J. ARIMA models to predict next-day electricity prices. *IEEE Trans. Power Syst.* **2003**, *18*, 1014–1020.
6. Mandal, P.; Srivastava, A.K.; Senjyu, T.; Negnevitsky, M. A new recursive neural network algorithm to forecast electricity price for PJM day-ahead market. *Int. J. Energy Res.* **2010**, *34*, 507–522.
7. Hong, Y.; Wu, C.-P. Day-ahead electricity price forecasting using a hybrid principal component analysis network. *Energies* **2012**, *5*, 4711–4725.
8. Georgilakis, P.S. Artificial intelligence solution to electricity price forecasting problem. *Appl. Artif. Intell.* **2007**, *21*, 707–727.
9. Georgilakis, P.S. Market Clearing Price Forecasting in Deregulated Electricity Markets Using Adaptively Trained Neural Networks. In Proceedings of the fourth Hellenic Conference on Artificial Intelligence, Heraklion, Greece, 18–20 May 2006; pp. 56–66.
10. Wu, L.; Shahidehpour, M. A hybrid model for day-ahead price forecasting. *IEEE Trans. Power Syst.* **2010**, *25*, 1519–1530.

11. Shafie-khah, M.; Moghaddam, M.P.; Sheikh-El-Eslami, M.K. Price forecasting of day-ahead electricity markets using a hybrid forecast method. *Energy Convers. Manag.* **2011**, *52*, 2165–2169.
12. Yamin, H.Y.; Shahidehpour, S.M.; Li, Z. Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets. *Electr. Power Energy Syst.* **2004**, *26*, 571–581.
13. Weron, R. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*; Wiley: Chichester, UK, 2006; pp. 125–127.
14. Keles, D.; Genoese, M.; Most, D.; Fichtner, W. Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. *Energy Econ.* **2012**, *34*, 1012–1032.
15. Becker, R.; Hurn, S.; Pavlov, V. Modelling spikes in electricity prices. *Econ. Rec.* **2007**, *83*, 371–382.
16. Lu, X.; Dong, Z.Y.; Li, X. Electricity market price spike forecast with data mining techniques. *Electr. Power Syst. Res.* **2005**, *73*, 19–29.
17. Zhao, J.H.; Dong, Z.Y.; Li, X.; Wong, K.P. A framework for electricity price spike analysis with advanced data mining methods. *IEEE Trans. Power Syst.* **2007**, *22*, 376–385.
18. Zhao, J.H.; Dong, Z.Y.; Li, X. Electricity market price spike forecasting and decision making. *IET Gener. Transm. Distrib.* **2007**, *1*, 647–654.
19. Amjady, N.; Keynia, F. Electricity market price spike analysis by a hybrid data model and feature selection technique. *Electr. Power Syst. Res.* **2010**, *80*, 318–327.
20. Plazas, M.A.; Conejo, A.J.; Prieto, F.J. Multimarket optimal bid-ding for a power producer. *IEEE Trans. Power Syst.* **2005**, *20*, 2041–2050.
21. Carrion, M.; Philpott, A.B.; Conejo, A.J.; Arroyo, J.M. A stochastic programming approach to electric energy procurement for large consumers. *IEEE Trans. Power Syst.* **2007**, *22*, 744–754.
22. Dimitroulas, D.K.; Georgilakis, P.S. A new memetic algorithm approach for the price based unit commitment problem. *Appl. Energy* **2011**, *88*, 4687–4699.
23. Granger, C.W.J. Non-linear models: Where do we go next—Time varying parameter models? *Stud. Nonlinear Dyn. Econom.* **2008**, *12*, doi:10.2202/1558-3708.1639.
24. Cervigón, R. In *Biomedical Applications of the Discrete Wavelet Transform, Discrete Wavelet Transforms—Biomedical Applications*; Olkkonen, H., Ed., InTech: Rijeka, Croatia, 2011; pp. 1–16, doi:10.5772/21576.
25. Galli, A.W.; Heydt, G.T.; Ribeiro, P.F. Exploring the power of wavelet analysis. *IEEE Comput. Appl. Power* **1996**, *9*, 37–41.
26. 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**, *20*, 1035–1042.
27. Mallat, S.G. A theory for multiresolution signal decomposition—the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693.
28. Tan, Z.; Zhang, J.; Wang, J.; Xu, J. Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Appl. Energy* **2010**, *87*, 3606–3610.
29. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis. Forecasting and Control*; Holden-Day: San Francisco, CA, USA, 1970; pp. 51–60.
30. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*; Springer: New York, NY, USA, 1996; pp. 180–185.

31. 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*, 1297–1304.
32. Haykin, S. *Neural Networks: A Comprehensive Foundation*, 2nd ed.; Prentice-Hall: Englewood Cliffs, NJ, USA, 1999; pp. 109–110.
33. Kittler, J.; Hatef, M.; Duin, R.P.; Matas, J.G. On combining classifiers. *IEEE Trans. Patt. Anal. Mach. Int.* **1998**, *20*, 226–239.
34. Meng, K.; Dong, Z.; Wang, H.; Wang, Y. Comparisons of machine learning methods for electricity regional reference price forecasting. *Lect. Notes Comp. Sci.* **2009**, *5551*, 827–835.
35. Huang, D.; Zareipour, H.; Rosehart, W.D.; Amjady, N. Data mining for electricity price classification and the application to demand-side management. *IEEE Trans. Smart Grid* **2012**, *3*, 808–817.
36. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995; pp. 138–140.
37. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
38. Ali, J.; Khan, R.; Ahmad, N.; Maqsood, I. Random forests and decision trees. *IJCSI Int. J. Compt. Sci. Issues* **2012**, *9*, 272–278.
39. Provost, F.; Domingos, P. *Well-Trained PETs: Improving Probability Estimation Trees*; Stern School of Business: New York, NY, USA, 2000; pp. 1–26.
40. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; Wiley-Interscience: New York, NY, USA, 2001; pp. 164–166.
41. Specht, D.F. Probabilistic Neural Networks for Classification Mapping, or Associative Memory. In Proceedings of the IEEE International Conference on Neural Networks, San Diego, CA, USA, 24–27 July 1988; pp. 525–532.
42. Lora, A.; Santos, J.; Exposito, A.; Ramos, J.; Santos, J. Electricity market price forecasting based on weighted nearest neighbors techniques. *IEEE Trans. Power Syst.* **2007**, *22*, 1294–1301.
43. Amjady, N.; Keynia, F. Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method. *Int. J. Electr. Power Energy Syst.* **2008**, *30*, 533–546.
44. Nogales, F.J.; Contreras, J.; Conejo, A.J.; Espinola, R. Forecasting next-day electricity prices by time series models. *IEEE Trans. Power Syst.* **2002**, *17*, 342–348.
45. Amjady, N.; Keynia, F. Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. *Trans. Power Syst.* **2009**, *24*, 306–318.
46. Bierbrauer, S.T.; Weron, R. *Lecture Notes in Computer Science*; Springer: Berlin, Germany, 2004; pp. 859–867.
47. Cartea, A.; Figueroa, M. Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. *Appl. Math. Financ.* **2005**, *12*, 313–335.
48. Vahidinasab, V.; Jadid, S.; Kazemi, A. Day-ahead price forecasting in restructured power systems using artificial neural networks. *Electr. Power Syst. Res.* **2008**, *78*, 1332–1342.
49. Amjady, N.; Daraeepour, A. Mixed price and load forecasting of electricity markets by a new iterative prediction method. *Electr. Power Syst. Res.* **2009**, *79*, 1329–1336.

50. Taylor, J.W.; de Menezes, L.M.; McSharry, P.E. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *Int. J. Forecast.* **2006**, *22*, 1–16.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).