

The price prediction for the energy market based on a new method

Homayoun Ebrahimian, Saeed Barmayoon, Mohsen Mohammadi & Noradin Ghadimi

To cite this article: Homayoun Ebrahimian, Saeed Barmayoon, Mohsen Mohammadi & Noradin Ghadimi (2018) The price prediction for the energy market based on a new method, Economic Research-Ekonomiska Istraživanja, 31:1, 313-337, DOI: [10.1080/1331677X.2018.1429291](https://doi.org/10.1080/1331677X.2018.1429291)

To link to this article: <https://doi.org/10.1080/1331677X.2018.1429291>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 09 Feb 2018.



Submit your article to this journal [↗](#)



Article views: 241



View related articles [↗](#)



View Crossmark data [↗](#)

The price prediction for the energy market based on a new method

Homayoun Ebrahimian^a, Saeed Barmayoon^b, Mohsen Mohammadi^c and Noradin Ghadimi^b

^aDepartment of Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran; ^bYoung Researchers and Elite Club, Ardabil Branch, Islamic Azad University, Ardabil, Iran; ^cDepartment of Electrical Engineering, Payame Noor University (PNU), Tehran, Iran

ABSTRACT

Regarding the complex behaviour of price signalling, its prediction is difficult, where an accurate forecasting can play an important role in electricity markets. In this paper, a feature selection based on mutual information is implemented for day ahead prediction of electricity prices, which are so valuable for determining the redundancy and relevancy of selected features. A combination of wavelet transform (WT) and a hybrid forecast method is presented based on a neural network (NN). Furthermore, an intelligent algorithm is considered for a prediction process to set the proposed forecast engine free parameters based NN. This optimisation process improved the accuracy of the proposed model. To demonstrate the validity of this model, the Pennsylvania-New Jersey-Maryland (PJM) electricity market is considered as a test case and compared with some of the most recent price forecast methods. These comparisons illustrate the effectiveness of the proposed strategy.

ARTICLE HISTORY

Received 2 November 2015
Accepted 7 June 2017

KEYWORDS

Feature selection; hybrid forecast engine; price forecast

JEL CLASSIFICATIONS

C53; E17; E47

1. Introduction

As developing countries seek to improve their economic prospects, electricity reform has been widely viewed as a central part of this effort. While the focus of most research to date has been at economy or utility level; there has been much less research on regional outcomes. Over the past few decades, as developing economies have struggled to pull their populations out of poverty, electricity sector restructuring has been regarded as a crucial facilitating factor for higher levels of economic development (Kalirajan, Rao, & Shand, 1998).

Electricity demand has increased faster than overall energy use in recent years (Esfahani & Ramirez, 2003). Total worldwide electricity consumption grew by 40% during 2000–2010, while overall energy use grew by 26% during the same period. The proportion of electricity consumption in total global energy consumption increased steadily from 16.3% to 17.7% between 2005 and 2010, while the share of oil consumption decreased by 2.2% during the same period. The recent rise in electricity consumption can be credited to (i) a global

CONTACT Noradin Ghadimi  noradin.ghadimi@iauardabil.ac.ir

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

prevalence of electronic devices such as television, personal computers and mobile telephones, and (ii) surging demand for electricity in emerging countries, where electricity use increased by 66% during 2005–2010 (Ferguson, Hill, & Wilkinson, 2000).

The importance of electricity price forecasting on one hand and its complexity on the other hand motivates many researches in this area, especially in the recent years. However, most of time series models are linear predictors, while electricity price is generally a non-linear function of its input features. So, the behaviour of price signal may not be completely captured by the time series techniques (Amjady & Keynia, 2009; Ferguson et al., 2000).

1.1. Literature review

Different prediction methods have been presented in recent years for price forecasting. Non-stationary time series models like Generalised Auto-Regressive Conditional Heteroskedastic (Ferguson et al., 2000), stationary time series models such as Auto-Regressive (AR) (Che & Wang, 2010), dynamic regression and transfer function (Lin, Gow, & Tsai, 2010; Nogales & Conejo, 2006; Nogales, Contreras, Conejo, & Espinola, 2002), and Auto-Regressive Integrated Moving Average (ARIMA) (Contreras, Espinola, Nogales, & Conejo, 2003) have been proposed for this purpose. To solve this problem, some other research works proposed neural networks (NNs) and fuzzy neural networks (FNNs) for price forecasting (Esfahani & Ramirez, 2003). In Ghadimi and Firouz (2015), a combination of NN with chemical reaction optimisation has been proposed for solving the price forecast problem. Confidence interval estimation with a cascaded NN has been presented in Jalili and Ghadimi (2015) for solving the price forecast. In Guo and Luh (2004) a new version of NN named committee has been presented based on the market clearing issue in the price prediction procedure. Application of Hidden Markov Models is presented in Gonzalez, San Roque, and Gonzalez (2005) for solving the mentioned problem. A wavelet Transform signal processing technique is shown in Szkuta, Sanabria, and Dillon (1999) and Dynamic Regression and Transfer Function (TF) models in Zhang and Luh (2005). In Mandal, Senjyu, Urasaki, Funabashi, and Srivastava (2007), a combination of Similar Day (SD) and NN techniques is proposed for locational marginal price (LMP) prediction in the PJM electricity market. NNs and FNNs have the capability of modeling the non-linear input/output mapping functions. However, electricity price is a time variant signal and its functional relationships rapidly vary with time (Esfahani & Ramirez, 2003; Gonzalez et al., 2005). So, derived information or extracted features of the NN or FNN rapidly lose value. While it seems that the NN or FNN learns well the training data, they may encounter large prediction errors in the test phase. Pedregal and Trapero (2007) propose the use of dynamic harmonic regression for forecasting the hourly price time series of electricity markets. This model belongs to the class of the unobserved components models set up as a state space system.

One of the key components in liberalised power sectors is the short-term electricity market, where hourly energy prices are set (Amjady & Keynia, 2009). The market can be settled by two main settlement mechanisms, namely the pay-as-offer (also referred to as pay-as-bid) mechanism, where each selected supplier is paid at its offer price, and the pay-at-MCP mechanism, where all selected suppliers are paid at a uniform market clearing price (MCP), usually the price of the most expensive selected offer. However, in practice, the pay-at-MCP settlement mechanism is widely accepted and used for payments (Che & Wang, 2010; Lin et al., 2010).

1.2. Contribution

Another useful technique for price forecast, proposed in recent years, is wavelet decomposition. A certain regularity of the data is an important pre-condition for the successful application of NNs (Ghadimi, 2015). When using classical statistical techniques, a stationary process is assumed for the data. For electricity price time series, an assumption of stationarity has to be discarded most of the time. Besides, one has to bear in mind that different kinds of non-stationarities may exist. In order to tackle the problem of non-stationarity, wavelets have been utilised because they can produce a good local representation of the signal in both time and frequency domains (Abedinia & Amjady, 2015). For instance, in Conejo, Plazas, Espinola, and Molina (2005), the original price series has been decomposed into four components by the wavelet transform and each sub-series has been separately predicted by the ARIMA time series. The final forecast is obtained by returning to the original domain (inverse transform).

In this paper, a new prediction strategy is proposed for day ahead price forecasting of electricity markets. Contribution of the paper can be summarised as:

- (1) The original price signal is decomposed via discrete wavelet transform and a new data modelling based on the features of both wavelet and time domains is presented.
- (2) An intelligent algorithm has been used for improving the NN based forecast engine training. In this process, the proposed algorithm finds the optimal value of NN weights which have a direct impact in prediction accuracy.
- (3) The whole structure of the proposed technique, including discrete wavelet transform, mixed data model, feature selection, cascaded forecasters, combination of NN and an intelligent algorithm are arranged in a coordinated manner to construct the proposed price forecast strategy.

The remaining parts of the paper are organised as follows. In the second section, the proposed forecast model has been presented. The proposed feature selection model is presented in section 3. The numerical results obtained from the proposed price forecast strategy for real-world electricity markets are presented in section 4 and compared with the results of several other recently published price forecast methods. Finally, section 5 concludes the paper.

2. The proposed price forecast strategy

2.1. The proposed data model

The Short Time Fourier Transform (STFT) provides the time information by computing different FTs for consecutive time intervals and putting them together. STFT gives a fixed resolution at all times. Once the window is chosen, the resolution is set for both time and frequency. A wide analysis window gives poor time resolution and good frequency resolution and vice versa. On the other hand, the electricity price series contains several non-stationary features such as trends, changes in level and slope, and seasonalities, to name a few. So, the price characteristics challenge the traditional Fourier analysis.

The basic concept in wavelet analysis begins with the selection of a proper wavelet (mother wavelet) and then performing an analysis on its translated and dilated versions. A wavelet can be defined as a function $\psi_{(x)}$ with a zero mean;

$$\int_{-\infty}^{+\infty} \psi_{(t)} dt = 0 \tag{1}$$

A signal can be decomposed into many series of wavelets with different scales a and translation b :

$$\psi_{(a,b)(x)} = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \tag{2}$$

So, the continuous wavelet transform $W_{(a,b)}$ of signal $f_{(x)}$ with respect to a wavelet $\psi_{(x)}$ is given by Ghadimi (2015) and Pedregal and Trapero (2007):

$$W_{(a,b)} = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f_{(x)} \psi\left(\frac{x-b}{a}\right) dx \tag{3}$$

where scale parameter a controls the spread of the wavelet and translation factor b determines its central position. $\psi_{(x)}$ is also called the mother wavelet. A $W_{(a,b)}$ coefficient represents how well the original signal $f_{(x)}$ and the scaled/translated mother wavelet match.

The original signal $f_{(x)}$ can be reconstructed by inverse wavelet transform:

$$f(x) = \int_0^{\infty} \int_{-\infty}^{+\infty} \frac{1}{a^2} W f_{(b,a)} \psi_{b,a(x)} db da \tag{4}$$

Thus, the set of all wavelet coefficients $W_{(a,b)}$, associated to a particular signal, is the wavelet representation of the signal with respect to the mother wavelet. Since the Continuous wavelet transform (CWT) is achieved by continuously scaling and translating the mother wavelet, substantial redundant information is generated. Therefore, instead of doing that, the mother wavelet can be scaled and translated using certain scales and positions, usually based on powers of two (Zhang & Luh, 2005). This scheme is more efficient and just as accurate as the CWT. It is known as the Discrete Wavelet Transform (DWT):

$$W_{(m,n)} = 2^{-(m/2)} \sum_{t=0}^{T-1} f_{(t)} \psi\left(\frac{t-n.2^m}{2^m}\right) \tag{5}$$

where T is the length of the signal $f_{(t)}$. The scaling and translation parameters are functions of the integer variables m and n ($a = 2^m$ and $b = n.2^m$); t is the discrete time index.

Actually, the Mallat strategy is used to implement DWT using filters (Abedinia & Amjadi, 2015; Ghadimi, 2015; Mandal et al., 2007; Pedregal & Trapero, 2007; Zhang & Luh, 2005) which has two stages: decomposition and reconstruction. Figure 1 shows the wavelet decomposition and reconstruction process. The proposed forecast model is divided into three main stages. In Stage one, the data used in the pre-processing are historical electricity load and pricing data. They are fed to the model as time-series signals. Depending on the

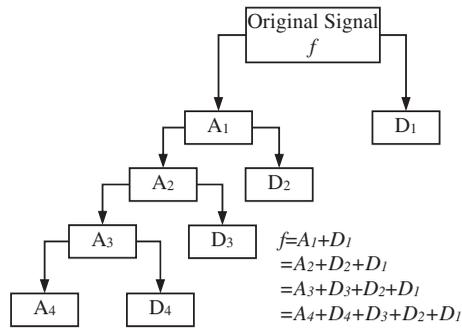


Figure 1. Multilevel reconstruction process. Source: Authors.

selected resolution levels, the time-series signals are decomposed into a number of wavelet coefficients. If the resolution level is defined as n , after decomposing the signal, there will be one approximation coefficient series with n number of detail coefficient series. In stage two, NNs are used for data prediction in the forecast model. The number of NNs needed for the model is determined by the number of wavelet coefficient signals at the output of the pre-processor. For each wavelet coefficient signal, one NN is required to perform the corresponding prediction. Also, in this stage the proposed intelligent method is applied over the weights parameters of NN. At the end, stage three is applied for data post-processing, which uses the same wavelet technique and resolution level as data pre-processing. The outputs from the NNs are recombined to form the final predicted output. Figure 2 presents the main structure of the proposed method.

This paper used three level decompositions as three Details (D) and one Approximate (A) signal. As decomposition involves filtering (high pass and low pass filter) and down-sampling, the wavelet reconstruction involves three steps of up-sampling and filtering.

A wavelet function of type Daubechies of order 4 (db4) is used in this paper as the mother wavelet and was selected. This wavelet offers an appropriate tradeoff between wavelength (for evaluation of local behaviour of signal) and smoothness, resulting in an appropriate behaviour for price forecast. As shown in Figure 1, the stages of the proposed technique are presented. Stage three will be described in the next part.

2.2. Proposed forecast engine

In this paper, a three stage cascade NN (CNN) has been proposed as a forecast engine. To train each NN of the CNN, an optimisation technique is proposed in this paper, which optimises the weights of the NN to learn its associated input/output mapping function. The stochastic search method is a modified version of particle swarm optimisation (Zhang & Luh, 2005). According to the fact that this algorithm is not the novelty of this paper, it was proposed (Mohammadi & Ghadimi, 2015) for readers who want to have further information of this algorithm. To effectively train each NN of CNN by the modified Particle swarm optimization (PSO), its generalisation capability should be carefully monitored along the training process to avoid overfitting, which is a serious problem for NN training (Abedinia & Amjady, 2015; Zhang & Luh, 2005). Generalisation is a measure of how well the NN performs on the actual problem once training is complete (Abedinia & Amjady, 2015).

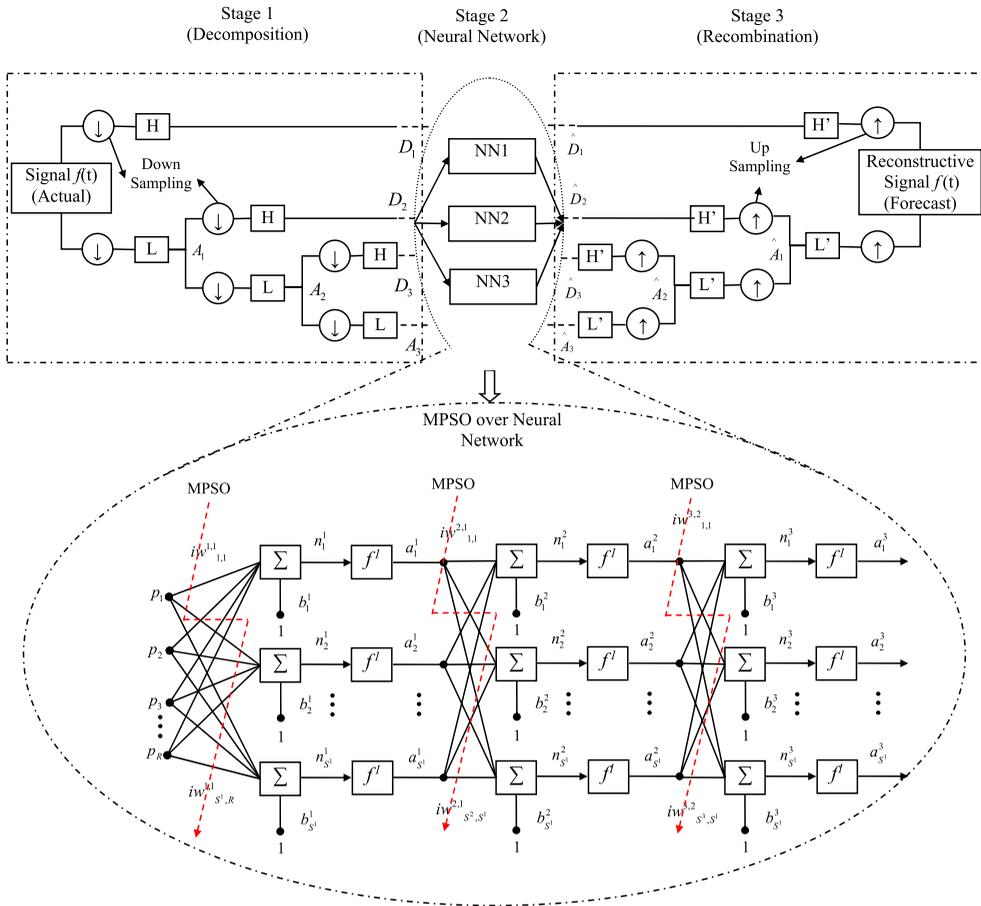


Figure 2. The main structure of the proposed model. Source: Authors.

When overfitting occurs in the training phase, the NN training error continues to decrease and it seems that the training process progresses, while indeed the generalisation capability of the NN degrades and it loses its prediction ability for unseen forecast samples.

Using the produced training and validation samples, the proposed modified PSO can train each NN of the forecasting engine or CNN. To do this, the training phase of the NN is modelled as an optimisation problem for the modified PSO, in which the objective function is the error of the constructed training samples or training error. Also, decision variables of the optimisation problem are the weights of the NN. The architecture of the suggested forecasting engine, i.e. CNN, is shown in Figure 3a, which is composed of cascaded NNs. For simplicity, only three NNs are shown in Figure 3, but in general CNN can have any number of cascaded NNs. Each cascaded NN has a multi-layer perceptron structure, which is an efficient structure for estimating NNs. The first NN, i.e. NN_1 in Figure 3a, is fed by the inputs selected by the double-filter of the proposed feature selection method.

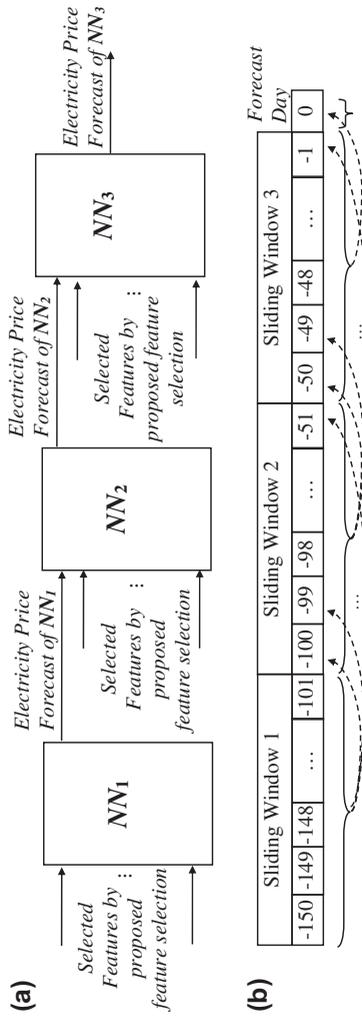


Figure 3. Representation of the proposed price forecasting engine (CNN): (a) architecture and (b) training windows. Source: Authors.

3. Proposed feature selection

The set of candidate inputs is so large and cannot be directly used for any forecast. Besides, irrelevant and redundant candidate inputs may exist in it, which can be misleading for the forecast engine. Additionally, with more input features, forecasters usually require more historical data to extract the mapping function between the inputs and output, where it can create some limitations in forecasting problems.

Relevancy: Individual relevance of the candidate inputs with the target variable is an important factor for feature selection. Mutual Information (MI) is a good criterion to measure relevancy between two random variables. The MI between two random variables x and y can be interpreted as the information about y that we obtain by studying x and vice versa. The entropy $H(x)$ of a continuous random variable x with probability distribution $P(x)$ is defined as follows (Amjady & Keynia, 2009):

$$H(x) = - \int P(x) \log_2 (P(x)) dx \quad (6)$$

The joint entropy $H(x,y)$ of two discrete random variables x and y with a joint probability distribution $P(x,y)$ is defined by:

$$H(x, y) = - \sum_{i=1}^n \sum_{j=1}^m P(x_i, y_j) \log_2 (P(x_i, y_j)) \quad (7)$$

where $H(x,y)$ indicates the total entropy of random variables x and y (Mohammadi & Ghadimi, 2015). The MI of continuous variables x and y , denoted by $MI(x;y)$, is defined based on their joint probability distribution $P(x,y)$ and the respective individual probability distributions $P(x)$ and $P(y)$ as follows (Mohammadi & Ghadimi, 2015):

$$MI(x;y) = \iint P(x, y) \log_2 \left(\frac{P(x, y)}{P(x)P(y)} \right) \quad (8)$$

MI between two discrete random variables x and y with n and m discrete values, respectively, becomes as follows:

$$MI(x;y) = \sum_{i=1}^n \sum_{j=1}^m P(x_i, y_j) \log_2 \left(\frac{P(x_i, y_j)}{P(x_i)P(y_j)} \right) \quad (9)$$

If the mutual information between two random variables is large, the two variables are closely related and vice versa. If the mutual information becomes zero, the two random variables are totally unrelated or the two variables are independent. Entropy and MI are often illustrated in the form of a Venn diagram, as shown in Figure 4. $MI(x;y)$ measures dependence between X and Y . If $MI(x;y)$ is large (small), it means x and y are closely (not closely) related. If x and y are independent then $P(x; y) = P(x)P(y)$, and so $MI(x;y) = 0$.

In MI, the approximation criterion based on the binomial distribution is used, which can be computed by a reasonable amount of historical data and low computation burden. However, this technique does not consider redundancy of candidate features. It has been recognised that the combinations of individually good features do not necessarily lead to

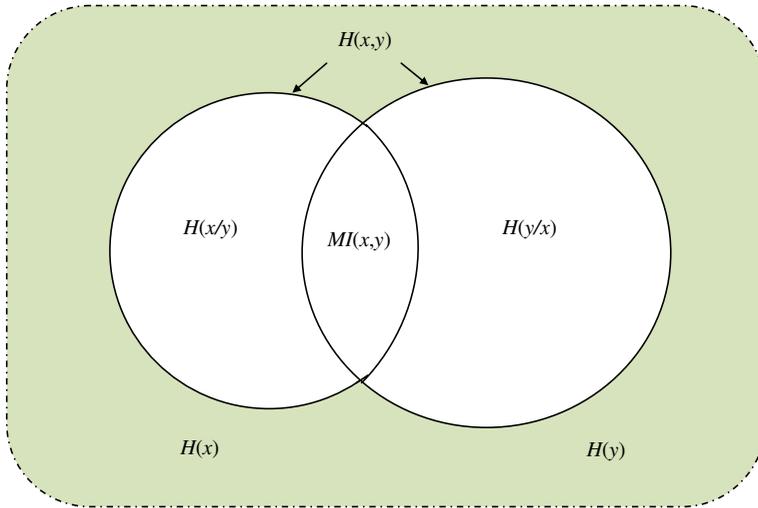


Figure 4. Representation of relationships between the entropies of two random variables, X and Y, and their mutual information. Source: Authors.

good classification performance. Accordingly, an effective method to remove redundant features is presented in this paper, which consists of two stage feature selection and is called MIMI. If the $S_1 = \{x_1, x_2, \dots, x_n\}$ is a set of candidate inputs including redundant features, the MI between each pair of features of S_1 can be calculated as:

$$MI(x_i, x_j), 1 < i, j < n \tag{10}$$

A higher value of $MI(x_i, x_j)$ means more common information between the candidate inputs, which means the high level of redundancy. So, for each feature $x_i \in S_1$, its maximum redundancy in the set S_1 is computed:

$$Max(MI(x_i, x_j)), 1 < j \leq n \tag{11}$$

By combining the feature selection technique and the proposed redundancy filter, we reach the following two-stage feature selection as:

$$Max(MI(x_k, x_j)) = MI(x_k, x_m) > TH2, 1 < m \leq n \tag{12}$$

So, between x_k and its partner x_m , one variable should be eliminated. Accordingly, the relevancy factors of these features, i.e. $MI(x_k, y)$ and $MI(x_m, y)$, are considered and the feature with a lower relevancy factor (less relevant feature) is filtered out. It can be feasible to more than two redundant features such that only one of them is eligible to be retained.

Accordingly, the redundancy filtering procedure is repeated for all features of S_1 , until no redundancy measure becomes greater than $TH2$. The redundancy filter is the second stage of the proposed two-stage feature selection algorithm.

The sub-set of features $S_2 \subset S_1$ that pass the redundancy filter are finally selected candidate inputs by the proposed two-stage feature selection algorithm. So, the proposed feature selection algorithm can be described as Figure 5. The proposed strategy for price forecast can be summarised in the literature as:

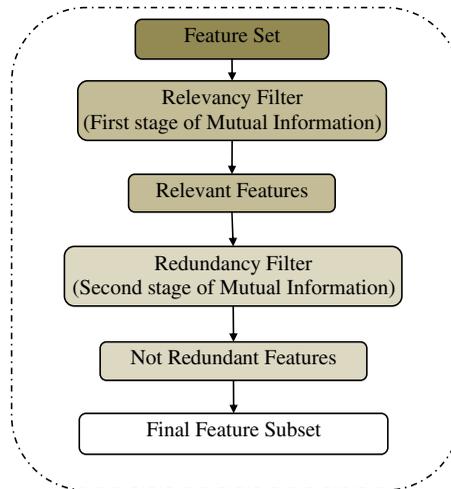


Figure 5. Structure of the proposed feature selection algorithm. Source: Authors.

- (1) Decompose the original price signal via the DWT into four components: A_3 , D_3 , D_2 and D_1 .
- (2) Select initial values for modified particle swarm optimization (MPSO) as a start point of the intelligent method.
- (3) Consider the weights of NN as decision variables through the minimisation of the objective function.
- (4) In order to give the algorithm freedom to select weights from an unlimited range of values, mutation in integer variables produces a progressive 'shift' of the centre of the range of positive/negative values provided for selection. At each iteration the range of integer values offered to the algorithm is randomly selected. As the error progresses towards the target level, the rate of progress is controlled by progressively adapting the numerical range within which the mutation shifts are applied.
- (5) The candidate features are selected by the proposed two-stage MIMI feature selection.
- (6) The individual forecast value of the decomposed approximation (\hat{A}_3) and detail signals (\hat{D}_1, \hat{D}_2 and \hat{D}_3) are then obtained from the NN and the desired day-ahead price forecasts are obtained after the wavelet reconstruction.

The proposed flowchart of the wavelet neural network (WNN) and proposed feature selection technique is presented in Figure 6.

4. Numerical results

In this paper, the Pennsylvania-New Jersey-Maryland (PJM) market is well recognised in the U.S. and beyond; the proposed forecast strategy is tested using the data from the day-ahead energy market and system operations of PJM (Ghadimi, 2015), with the 4 weeks corresponding to the four seasons of year 2006 for the examination of the proposed method considered in this test case being 15–21 February, 15–21 May, 15–21 August and 15–21 November (months 2, 5, 8 and 11).

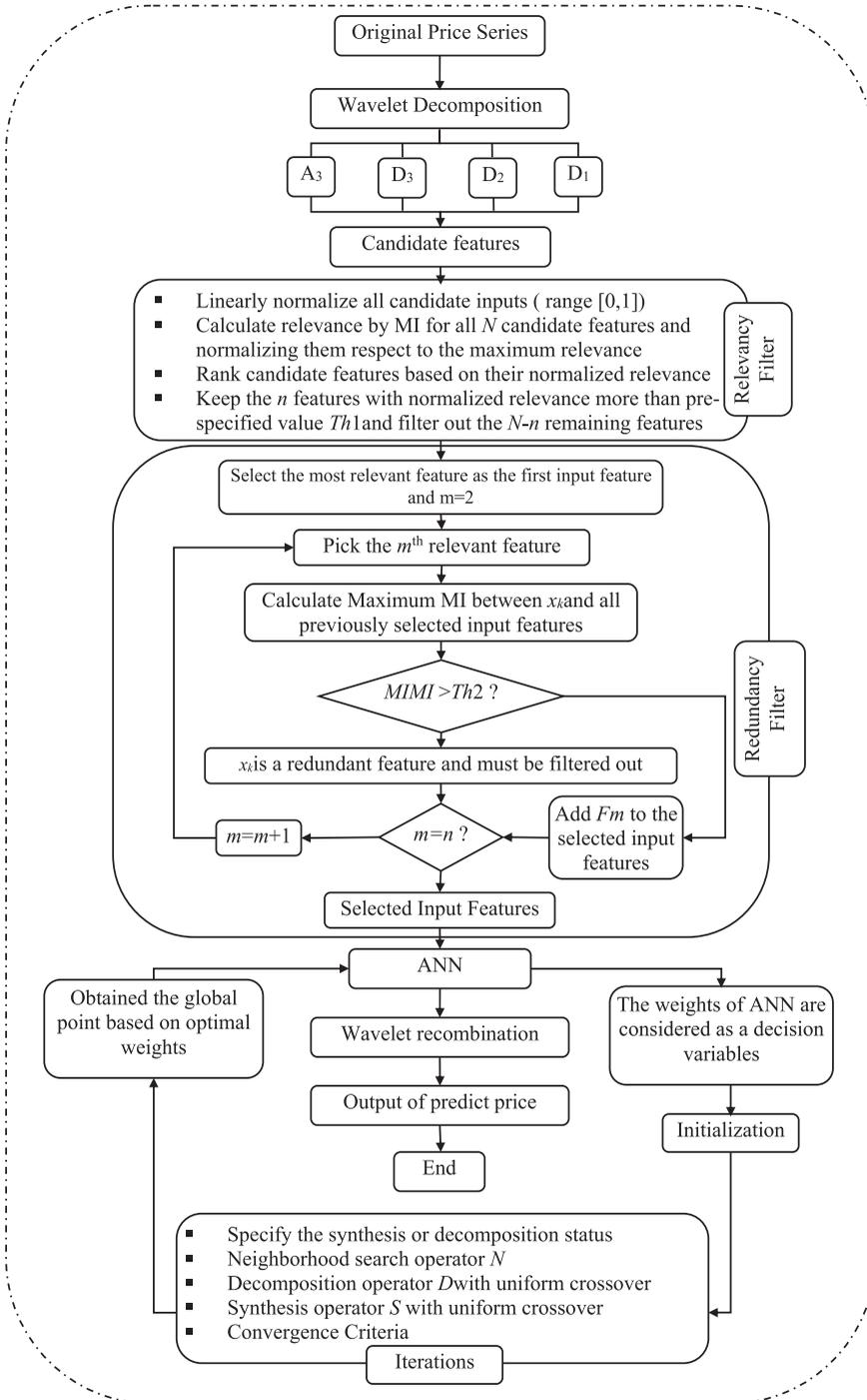


Figure 6. The total flowchart of the proposed method. Source: Authors.

4.1. PJM market

At first, sample results of the feature selection technique (two step correlation analysis) are presented. The candidate inputs for each sub-series $A3$, $D3$, $D2$ and $D1$ include lagged values of that sub-series, price (P), load (L) and available generation (AVG) up to 200 h age, respectively (in total 800 candidates). For instance, the 800 candidate inputs for the $A3$ sub-series are $\{A3_{(h-1)}, A3_{(h-2)}, \dots, A3_{(h-200)}, P_{(h-1)}, P_{(h-2)}, \dots, P_{(h-200)}, L_{(h-1)}, L_{(h-2)}, \dots, L_{(h-200)}, AVG_{(h-1)}, AVG_{(h-2)}, \dots, AVG_{(h-200)}\}$. The first 200 candidates (lagged values of $A3$) for $D3$, $D2$ and $D1$ are replaced by the lagged values of the related sub-series. Obtained results from the proposed model for the first day of the Autumn test week are shown in Table 1. Similar results to those shown in Table 1 have been obtained for the other days of the four test weeks.

$A3$ is the low frequency component of the original price signal and follows the trend of the signal. Usually the most useful information of the original signal can be found in the $A3$ sub-series. On the other hand, detail components contain high frequency components of the original signal, such that $D1$ is the highest frequency component. In Mohammadi and Ghadimi (2015) it has been described that $D1$ is more related to the noisy part of the original signal. $D2$ is a lower frequency component than $D1$ and $D3$ contains lower frequency information than $D1$ and $D2$. More information about the wavelet results has been presented in Appendix A. In Amjady, Daraeepour, and Keynia (2010) and Garcia and Kirschen (2006) it has been explained that the approximation sub-series constitutes the main component of the DWT, while the three details sub-series provide small adjustments. Table 1 shows that $A3$ is dependent on its own lagged values and also lagged values of price, load and available generation. For the prediction of the $A3$ sub-series, 11 inputs from the auto-regression part (candidate inputs from 1–200) and 15 inputs from the lagged values of the exogenous variables (candidate inputs from 201–800) are selected. In other words, $A3$ is dependent on both the auto-regression part and exogenous variables (like the original signal), since it is the most similar component to the original series. On the other hand, Table 1 shows that the details are mainly auto-regressive signals. Also, their dependency on the exogenous variables decreases from $D3$ to $D1$. Among 18 selected candidates of the $D3$ sub-series, 10 and eight inputs are from the auto-regression and cross-regression parts, respectively. For the prediction of $D2$ and $D1$ components, no lagged value of the exogenous variables is selected. For this forecast process, the auto-regression part is in the wavelet domain and all exogenous variables are from the time domain. Therefore, the above-mentioned divisions also indicate the proportion of the wavelet and time domain features of the selected candidate inputs in the proposed mixed data model.

The first forecaster from the set of 24 cascaded forecasters of $A3$ uses the 26 input variables shown in Table 1 and predicts $A3_{(h)}$, i.e., the value of $A3$ for the first hour of the next day. Then the 26 input variables of Table 1 are shifted 1 hour ahead and used by the second forecaster to predict $A3_{(h+1)}$, i.e., the value of $A3$ for the second hour of the next day. This cycle is repeated until $A3_{(h+23)}$, i.e., the value of $A3$ for the last hour of the next day, is predicted by the 24th forecaster of the set of cascaded forecasters of $A3$ (this cycle is similarly repeated to predict the next 24 hourly values of the other wavelet components $D3$, $D2$ and $D1$). Each forecaster of $A3$ (except the first one) uses the prediction of its previous forecaster in the chain, since $A3_{(h-1)}$ is among the selected input variables of $A3$ (Table 1). Besides, $P_{(h-1)}$ is also in the selected inputs for $A3$ according to the results of Table 1. The price forecast of the whole proposed method for hour h is used as $P_{(h-1)}$ for the next forecaster of the

Table 1. Selected inputs by the proposed feature selection for the four DWT components in the first day of the Autumn test week.

Sub-series	Number of selected candidates	Selected candidates
A3	26	$A3_{(h-11)'}^{(h-24)'}$, $A3_{(h-26)'}$, $A3_{(h-48)'}$, $A3_{(h-53)'}$, $A3_{(h-66)'}$, $A3_{(h-73)'}$, $A3_{(h-98)'}$, $A3_{(h-120)'}$, $A3_{(h-140)'}$, $A3_{(h-168)'}$, $P_{(h-1)'}$, $P_{(h-24)'}$, $L_{(h-1)'}$, $L_{(h-24)'}$, $L_{(h-28)'}$, $L_{(h-60)'}$, $L_{(h-72)'}$, $L_{(h-95)'}$ $L_{(h-107)'}$, $L_{(h-110)'}$, $L_{(h-110)'}$, $L_{(h-168)'}$, $AVG_{(h-24)'}$, $AVG_{(h-48)'}$, $AVG_{(h-168)'}$
D3	18	$D3_{(h-1)'}$, $D3_{(h-9)'}$, $D3_{(h-10)'}$, $D3_{(h-14)'}$, $D3_{(h-20)'}$, $D3_{(h-24)'}$, $D3_{(h-35)'}$, $D3_{(h-37)'}$, $D3_{(h-168)'}$, $D3_{(h-172)'}$, $L_{(h-8)'}$, $L_{(h-122)'}$, $L_{(h-200)'}$, $AVG_{(h-2)'}$, $AVG_{(h-48)'}$, $AVG_{(h-120)'}$ $AVG_{(h-192)'}$
D2	24	$D2_{(h-1)'}$, $D2_{(h-4)'}$, $D2_{(h-7)'}$, $D2_{(h-18)'}$, $D2_{(h-22)'}$, $D2_{(h-24)'}$, $D2_{(h-25)'}$, $D2_{(h-30)'}$, $D2_{(h-66)'}$, $D2_{(h-71)'}$, $D2_{(h-73)'}$, $D2_{(h-75)'}$, $D2_{(h-120)'}$, $D2_{(h-121)'}$, $D2_{(h-139)'}$, $D2_{(h-140)'}$ $D2_{(h-144)'}$, $D2_{(h-160)'}$, $D2_{(h-168)'}$, $D2_{(h-169)'}$, $D2_{(h-191)'}$, $D2_{(h-192)'}$, $D2_{(h-193)'}$
D1	23	$D1_{(h-3)'}$, $D1_{(h-4)'}$, $D1_{(h-20)'}$, $D1_{(h-24)'}$, $D1_{(h-28)'}$, $D1_{(h-45)'}$, $D1_{(h-48)'}$, $D1_{(h-51)'}$, $D1_{(h-52)'}$, $D1_{(h-72)'}$, $D1_{(h-78)'}$, $D1_{(h-99)'}$, $D1_{(h-118)'}$, $D1_{(h-120)'}$, $D1_{(h-123)'}$, $D1_{(h-140)'}$, $D1_{(h-144)'}$ $D1_{(h-148)'}$, $D1_{(h-167)'}$, $D1_{(h-168)'}$, $D1_{(h-171)'}$, $D1_{(h-188)'}$, $D1_{(h-192)'}$

Notes: P, Price; L, Load; AVG, Available Generation; h, Hour index.
 Source: Authors.

chain of forecasters of $A3$. From the results of Table 1, it is seen that $L_{(h-1)}$ and $AVG_{(h-1)}$ are also among the selected inputs for $A3$. Actual value of load and available generation are used for $L_{(h-1)}$ and $AVG_{(h-1)}$ in the cascaded forecasters of $A3$, since the forecasts of load and available generation are not accessible for us and also are not the focus of our paper. Similarly, for $L_{(h-8)}$ and $AVG_{(h-2)}$ selected for $D3$ (Table 1), the actual values are used for the set of cascaded forecasters of $D3$ when predicting the price values of the forecast day. Each forecaster from the set of 24 cascaded forecasters of each wavelet component receives 50 data points, including 49 training samples and one validation sample. The results for the filtering ratio (number of whole candidates/number of selected inputs) of the proposed feature selection technique can be obtained from Table 1:

$$A3:800/26 = 30.77, \quad D3:800/18 = 44.44, \quad D2:800/24 = 33.33, \quad D1:800/23 = 34.78 \tag{13}$$

These great filtering ratios significantly simplify the learning process of the proposed forecast method (learning the effect of 800 inputs on the output is a hard task or even impossible for any forecast method).

Although the wavelet components are obtained by the decomposition of the price signal, the past values of the price are considered among the candidate inputs of each wavelet component, since it is still possible that some characteristics of the price signal are better highlighted in the original time domain. So, it is better to also include the information content of the price signal in the candidate set of input variables to provide a richer candidate set. As seen from the sample results of the feature selection technique (Table 1), two price variables ($P_{(h-1)}$ and $P_{(h-24)}$) are among the selected candidates for the $A3$ sub-series. However, the price variables are selected less than the other exogenous variables (i.e., load and available generation), since the wavelet components are representatives of the price signal.

The obtained results of the proposed hybrid method and four other price forecast techniques for LMP prediction of the four test weeks of PJM electricity market are shown in Tables 2 and 3. In these tables, WME (Weekly Mean Error) and WPE (Weekly Peak Error) are error indicators:

$$WME = \frac{1}{168} \sum_{i=1}^{168} \frac{|P_{iACT} - P_{iFOR}|}{P_{iACT}} \tag{14}$$

$$WPE = \text{Max}_{1 \leq i \leq 168} \left(\frac{|P_{iACT} - P_{iFOR}|}{P_{iACT}} \right) \tag{15}$$

where P_{iACT} and P_{iFOR} are actual and forecast LMP of hour i , respectively. WME and WPE of the ARIMA time series (which has been proposed in Contreras et al. (2003) for price forecast)

Table 2. WME for the four test weeks of the PJM electricity market in year 2006.

Test week	ARIMA (%)	MLP+LM (%)	MLP+BFGS (%)	MLP+BR (%)	Proposed method (%)
Winter	11.21	9.82	12.90	13.22	4.01
Spring	15.30	8.87	10.12	12.92	4.21
Summer	13.56	10.43	11.46	11.98	4.21
Autumn	12.93	9.54	9.83	12.24	4.43
Average	13.25	9.67	11.08	12.59	4.21

Source: Authors.

are shown in the second column of Tables 2 and 3, respectively. A multivariate ARIMA model is used here in which the polynomials of the back-shift operator are considered for the exogenous variables (load and available generation) like price and error terms. More details about the multivariate ARIMA model can be found in Contreras et al. (2003). The Multi-Layer Perceptron (MLP) with Levenberg–Marquardt algorithm (LM) learning algorithm has been proposed in Catalão, Mariano, Mendes, and Ferreira (2007) for price forecast and its results are shown in the third column of Tables 2 and 3 (MLP+LM). BFGS (Broyden, Fletcher, Goldfarb, Shanno) and BR (Bayesian Regularisation) are also efficient learning algorithms for MLP (Mohammadi & Ghadimi, 2015), which, as seen from columns four and five of Tables 2 and 3 (MLP+BFGS and MLP+BR) produce results close to those of MLP+LM. For the sake of a fair comparison, ARIMA time series, MLP+LM, MLP+BFGS, MLP+BR, and the proposed method all have the same training and test periods.

For a better illustration of this matter, APE (Absolute Percentage Error) values of the proposed price forecast strategy for the spring test week are shown in Figure 7 by a black solid line. The APE value of the i^{th} forecast hour (APE_i) is calculated as follows:

Table 3. WPE for the four test weeks of the PJM electricity market in year 2006.

Test week	ARIMA (%)	MLP+LM (%)	MLP+BFGS (%)	MLP+BR (%)	Proposed method (%)
Winter	45.10	22.31	25.42	23.63	16.11
Spring	34.32	18.45	21.39	20.13	22.53
Summer	44.48	27.34	38.12	30.22	22.64
Autumn	55.81	32.12	29.23	33.11	15.33
Average	44.93	25.05	28.54	26.77	19.15

Source: Authors.

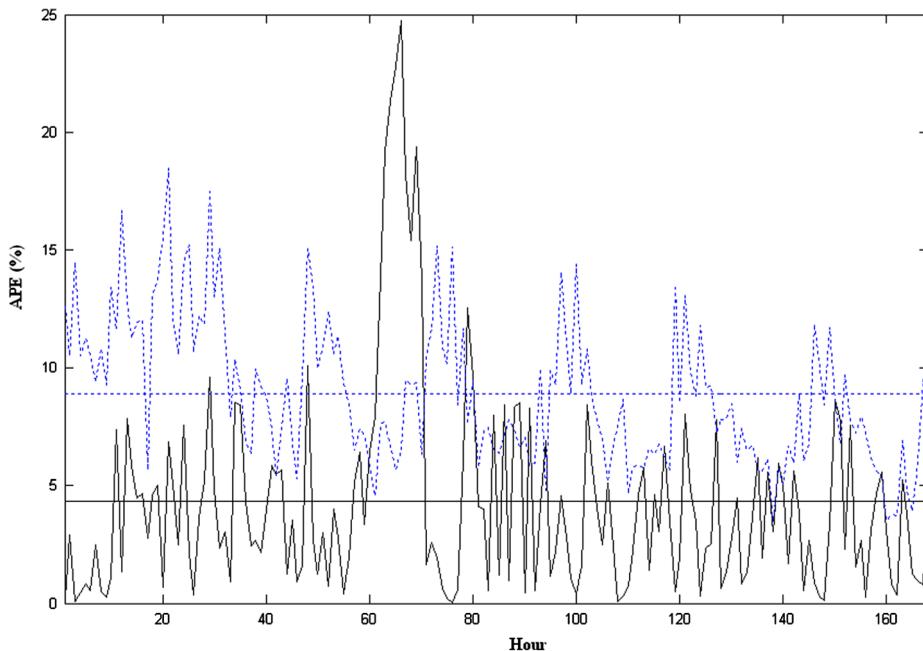


Figure 7. APE values (in terms of percentage) of the proposed price forecast strategy (black solid line) and MLP+LM (blue dotted line). Source: Authors.

$$APE_i = \frac{|P_{iACT} - P_{iFOR}|}{P_{iACT}} \tag{16}$$

The WME and WPE measure average and peak of 168 hourly APE values of a week, respectively. For comparison, APE values of MLP+LM (owning the least value of both WME and WPE among the four other examined methods in the spring test week) are also shown in Figure 7 by a blue dotted line. As seen from Figure 7, the proposed forecast strategy has less APE values than MLP+LM in most of the hours of the spring test week, such that the WME of the proposed strategy is considerably lower than the WME of MLP+LM. However, in a few hours of the spring test week, the APE values of the proposed strategy increase and become more than MLP+LM. These hours are concentrated around the peak error of the proposed strategy for the spring test week.

Moreover, for presenting more details of the proposed prediction model, in a whole year, its results for all weeks of year 2006 are shown in Table 4. From this table it can be seen that WMEs and WPEs for all 52 weeks of the year are close to the WMEs and WPEs of the four considered test weeks. The average WME for the 52 weeks is slightly higher than the average WME of the 4 weeks, while the average WPE of the 52 weeks is lower than the average WPE of the 4 weeks. Hence, it is seen that the proposed forecast strategy can produce acceptable results in the long-run for the whole year. In all examined test cases, no divergence problem of the proposed method has been observed at all, while oscillatory behaviours and trapping in dead bands and local minima have been avoided as much as possible.

Now the hybrid method is compared with some recent price forecast techniques proposed for the PJM electricity market. In Table 5, the hybrid forecast strategy is compared with the combination of similar days (SD) and NN techniques proposed in Mandal et al. (2007) for price forecasting in the PJM electricity Market. In Mandal et al. (2007) the following error measures, e_{day} and e_{week} have been considered for the test days and weeks, respectively, to avoid the adverse effect of prices close to zero:

$$e_{day} = \frac{1}{24} \sum_{i=1}^{24} \frac{|P_{iACT} - P_{iFOR}|}{P_{Ave-Day}}, \quad P_{Ave-Day} = \frac{1}{24} \sum_{i=1}^{24} P_{iACT} \tag{17}$$

Table 4. The results of the proposed price forecast strategy for all weeks of year 2006 in the PJM electricity market (all WME and WPE values are in percentages).

Week	1	2	3	4	5	6	7	8	9	10	11	12
WPE	16.19	12.27	10.40	13.32	11.61	8.18	9.31	17.43	13.30	11.51	13.44	16.17
WME	5.25	4.26	4.42	4.36	4.28	3.49	4.18	4.24	4.60	4.22	4.04	6.49
Week	13	14	15	16	17	18	19	20	21	22	23	24
WPE	21.27	16.6	18.59	13.54	17.5	9.17	14.52	12.22	18.62	13.15	22.78	23.31
WME	4.41	5.25	4.41	4.37	4.38	3.40	4.37	3.34	4.34	4.22	5.58	6.15
Week	25	26	27	28	29	30	31	32	33	34	35	36
WPE	23.69	21.72	20.36	12.72	12.57	12.37	24.28	20.33	21.23	24.11	22.55	17.17
WME	6.19	5.28	5.29	5.71	6.66	5.78	7.29	5.23	4.89	4.83	6.43	5.48
Week	37	38	39	40	41	42	43	44	45	46	47	48
WPE	9.32	18.56	14.70	10.55	14.36	21.47	22.68	20.52	16.69	12.22	16.56	18.31
WME	5.39	5.32	4.48	5.36	6.17	5.43	7.34	5.23	4.31	4.23	4.58	6.40
Week	49	50	51	52								
WPE	10.27	12.81	20.79	10.27								
WME	4.37	5.52	5.52	4.37								

Source: Authors.

Table 5. Comparison of the proposed method with SD+NN for price forecast in the PJM electricity market.

Test period	SD+NN (Mandal et al., 2007) (%)	Proposed (%)
20 January 2006	6.93	4.92
1 February 2006	7.96	3.35
5 March 2006	7.88	4.54
7 April 2006	9.02	5.03
13 May 2006	6.91	3.40
1–7 February 2006	7.66	4.05
22–28 February 2006	8.88	4.33

Source: Authors.

$$e_{week} = \frac{1}{168} \sum_{i=1}^{168} \frac{|P_{iACT} - P_{iFOR}|}{P_{Ave-Week}}, \quad P_{Ave-Week} = \frac{1}{168} \sum_{i=1}^{168} P_{iACT} \quad (18)$$

For the sake of a fair comparison, the same error measures, e_{day} and e_{week} , have also been considered for the proposed method in the calculation of the results of Table 5 (third column). In other words, the forecast errors of the five test days in Table 5 are in terms of e_{day} and the forecast errors of the two test weeks in the table are in terms of e_{week} . The results of SD+NN in Table 5 have been directly quoted from Mandal et al. (2007). As seen from Table 5, the proposed method outperforms SD+NN in all test days and weeks and considerable differences between the forecast errors of the proposed method and SD+NN are observed.

Regarding the limitation of this model, it can be considered that the volatility of the price signal in different markets can be an ineligible problem. So, by this problem premature convergence or a big error can happen in the prediction process. For this purpose, different tools as well as methods can be enhanced to solve the problem. This section can be presented as the future work of this manuscript.

4.2. New England electricity market

As the second case study, the New England electricity market has been considered to test the proposed prediction method. Similar to the previous test case, 50 days prior to the forecasting day are considered as the training and validation samples.

Actually, 50 days of historical data was used to set up the proposed forecast strategy in which the last day is removed and served as the validation period. So, the forecast engine of each wavelet component receives 50 data points, including 49 training samples and one validation sample. The forecast engine part of each forecaster is trained by its own 49 training samples, where the optimal point of the training phase is determined by the validation error related to its validation sample. Then, the proposed meta-heuristic algorithm part of the forecast engine further modifies the obtained weights of the training phase. In this way, all $4 \times 24 = 96$ forecasters are trained. It is noted that all other examined methods of the paper have the same historical data (50 days ago) for the sake of a fair comparison. In the first experiment of this test case, we consider three test months of December, March and August in the year 2005. Obtained results for the second test case have been presented in Table 6 for day-ahead electricity price forecast. In this way, monthly forecast errors in terms



Table 6. Monthly MAPE and MAE for price prediction in the New England electricity market.

Test Month	ISO (http://www.iso-ne.com/markets/hrly_data/index.html)		Single SVM (http://www.iso-ne.com/markets/hrly_data/index.html)		Hybrid network (http://www.iso-ne.com/markets/hrly_data/index.html)		HNIN+MRA (http://www.iso-ne.com/markets/hrly_data/index.html)		Proposed	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
December 2002	5.53	12.66	4.87	11.31	4.26	10.24	4.22	8.32	3.99	7.19
March 2005	—	—	4.72	7.17	4.09	6.35	4.08	6.12	3.80	5.20
August 2005	—	—	8.74	8.85	7.70	7.91	6.71	6.83	5.50	5.37

Source: Authors.

Table 7. Monthly MAPE for the New England electricity market.

Test month	Dynamic price forecast (Bompard et al., 2007)	HNN+MRA (http://www.iso-ne.com/markets/hrly_data/index.html)	Proposed
February	6.84	5.24	5.14
March	6.56	6.11	5.12
April	9.77	7.87	6.10
May	6.96	5.43	4.49
June	8.12	7.23	6.00
July	8.06	7.54	6.12
August	6.19	6.07	4.83
September	6.24	6.20	4.91
October	7.38	6.43	5.27
November	7.94	6.13	4.74
December	7.37	5.91	4.44

Source: Authors.

of mean absolute percentage error (MAPE) and Mean absolute error (MAE) are reported in Table 6 for different forecasting models, such as the Support-Vector Machine (SVM), hybrid network, Hybrid Neural Network (HNN) plus Modified Relief Algorithm (MRA) and electricity price forecasts from New England ISO (http://www.iso-ne.com/markets/hrly_data/index.html) are considered as benchmarks. For the sake of a fair comparison, the same conditions (shown in http://www.iso-ne.com/markets/hrly_data/index.html) have been considered in this paper.

In the second numerical experiment, the proposed method is compared with the dynamic price forecast technique presented in Bompard, Ciwei, Napoli, and Torelli (2007) and the results are presented in Table 7. The same test months shown in http://www.iso-ne.com/markets/hrly_data/index.html have been adopted in this test case, which include February 2004–December 2004 for the New England electricity market. Moreover, the forecasting errors are measured in terms of MAPE in this table. Obtained results in these tables demonstrate the validity of the proposed method. Regarding the achieved numerical results in two case studies in Ontario and New England markets, it can be claimed that the proposed method has good forecasting in these markets in comparison with other methods.

5. Conclusion

In this paper, a new feature selection and forecast engine is presented for day ahead prediction of electricity prices. Regarding to the developing countries which seek to improve their economic prospects, electricity reform has been widely viewed as a central part of this effort. For this purpose, the related researches for electricity price can play an important role in their economy. Therefore, in this paper we looked for a new prediction model for solving the mentioned problem. Over the past few decades, as developing economies have struggled to pull their populations out of poverty, electricity sector restructuring has been regarded as a crucial facilitating factor for higher levels of economic development. Accordingly, the two-step feature selection based on mutual information takes the data characteristics proposed and then selects a sub-set of them using correlation and instance-based feature selection methods, applied in a systematic way. Then, a combination of wavelet transform and a hybrid forecast method is presented based on NN and an optimisation algorithm. The hybrid method is examined for LMP prediction in the PJM electricity market and compared with some of the most recent price forecast techniques. These comparisons reveal the forecast capability of the proposed method. Besides, a structural

analysis is also performed on the method to show the effectiveness of its hybrid structure. The obtained results prove the validity proposed strategy.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Abedinia, O., & Amjady, N. (2015, July). Short-term wind power prediction based on hybrid neural network and chaotic shark smell optimization. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 2(3), 245–254.
- Amjady, N., Daraeepour, A., & Keynia, F. (2010). Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network. *IET Generation, Transmission & Distribution*, 4(3), 432–444.
- Amjady, N., & Keynia, F. (2009). Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique. *Energy Conversion and Management*, 50(12), 2976–2982.
- Bompard, E., Ciwei, G., Napoli, R., & Torelli, F. (2007). Dynamic price forecast in a competitive electricity market. *IET Generation, Transmission & Distribution*, 1(5), 776–783.
- Catalão, J. P. S., Mariano, S. J. P. S., Mendes, V. M. F., & Ferreira, L. A. F. M. (2007). Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power System Research*, 77, 1297–1304.
- Che, J., & Wang, J. (October 2010). Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. *Energy Conversion and Management*, 51(10), 1911–1917.
- Conejo, A. J., Plazas, M. A., Espinola, R., & Molina, A. B. (2005). Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Transactions on Power Systems*, 20, 1035–1042.
- Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems*, 18, 1014–1020.
- Esfahani, H. S., & Ramirez, M. T. (2003). Institutions, infrastructure, and economic growth. *Journal of Development Economics*, 70, 443–477.
- Ferguson, F., Hill, R., & Wilkinson, W. (2000). Electricity use and economic development. *Energy Policy*, 28, 923–934.
- Garcia, M. P., & Kirschen, D. S. (2006). Forecasting system imbalance volumes in competitive electricity markets. *IEEE Transactions on Power Systems*, 21, 240–248.
- Ghadimi, N. (2015). A new hybrid algorithm based on optimal fuzzy controller in multimachine power system. *Complexity*, 21(1), 78–93.
- Ghadimi, N., & Firouz, M. H. (2015). Short-term management of hydro-power systems based on uncertainty model in electricity markets. *Journal of Power Technologies*, 95(4), 265.
- Gonzalez, A. M., San Roque, A. M., & Gonzalez, J. G. (2005). Modeling and forecasting electricity prices with input/output hidden markov models. *IEEE Transactions on Power Systems*, 20, 13–24.
- Guo, J. J., & Luh, P. B. (2004). Improving market clearing price prediction by using a committee machine of neural networks. *IEEE Transactions on Power Systems*, 19, 1867–1876.
- Jalili, A., & Ghadimi, N. (2015). Hybrid harmony search algorithm and fuzzy mechanism for solving congestion management problem in an electricity market. *Complexity*, 21, 90–98.
- Kalirajan, K., Rao, M., & Shand, R. (1998). *The economics of electricity supply in India*. New Delhi: Macmillan India Limited.
- Lin, W. M., Gow, H. J., & Tsai, M. T. (2010, December). Electricity price forecasting using enhanced probability neural network. *Energy Conversion and Management*, 51(12), 2707–2714.

- Mandal, P., Senjyu, T., Urasaki, N., Funabashi, T., & Srivastava, A. K. (2007, November). A novel approach to forecast electricity price for PJM using neural network and similar days method. *IEEE Transactions on Power Systems*, 22(4), 2058–2065.
- Mohammadi, M., & Ghadimi, N. (2015). Optimal location and optimized parameters for robust power system stabilizer using honeybee mating optimization. *Complexity*, 21(1), 242–258.
- Nogales, F. J., & Conejo, A. J. (2006, April). Electricity price forecasting through transfer function models. *Journal of the Operational Research Society*, 57(4), 350–356.
- Nogales, F. J., Contreras, J., Conejo, A. J., & Espinola, R. (2002). Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, 17, 342–348.
- Pedregal, D. J., & Trapero, J. R. (2007, May). Electricity prices forecasting by automatic dynamic harmonic regression models. *Energy Conversion and Management*, 48(5), 1710–1719.
Retrieved from http://www.iso-ne.com/markets/hrly_data/index.html
- Szkuta, B. R., Sanabria, L. A., & Dillon, T. S. (1999). Electricity price short-term forecasting using artificial neural networks. *IEEE Transactions on Power Systems*, 14(3), 851–857.
- Zhang, L., & Luh, P. B. (2005). Neural network-based market clearing price prediction and confidence interval estimation with an improved extended kalman filter method. *IEEE Transactions on Power Systems*, 20(1), 59–66.

Appendix A.

The plot of price, load and available generation plus the plot of the wavelet components A_3 , D_3 , D_2 and D_1 in the four time intervals used in the research are shown in Figures A1–A8, respectively. Each of the four time intervals consists of a test week and 50 days as the training period (more accurately, 49 days as the training period and 1 day as the validation period).

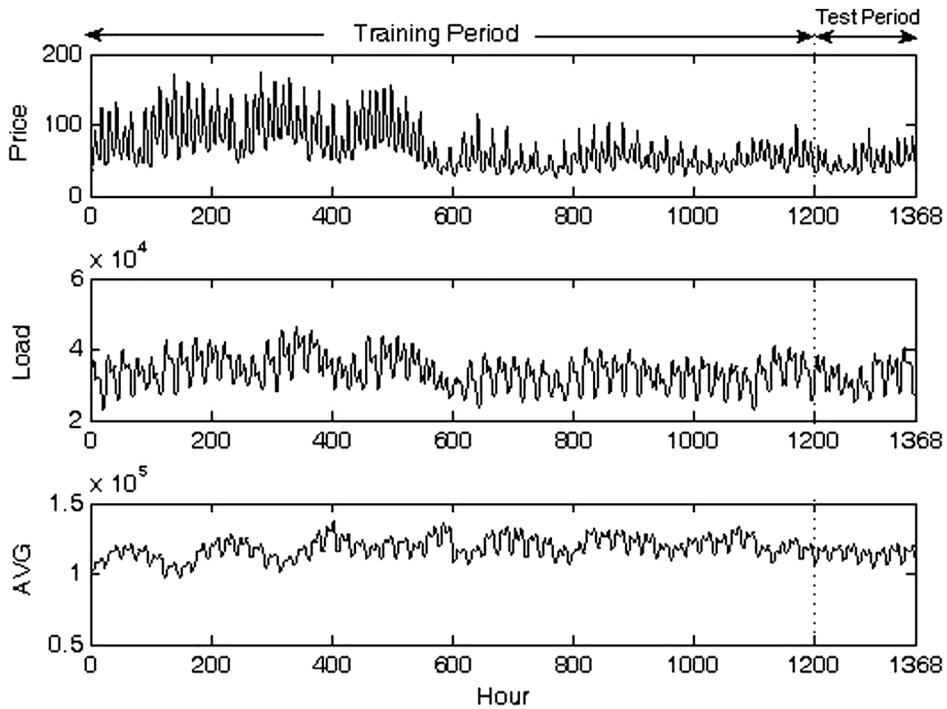


Figure A1. Plot of price, load and available generation (AVG) for the Winter test week (15–21 February) and its corresponding 50 days training period. Source: Authors.

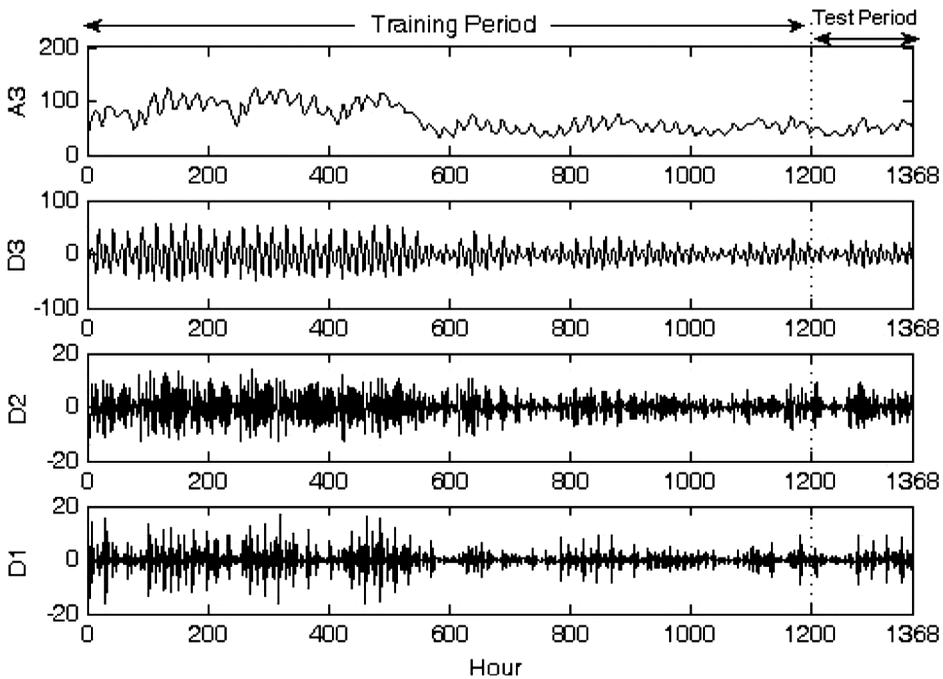


Figure A2. Plot of wavelet components A3, D3, D2 and D1 for the price time series shown in Figure A1. Source: Authors.

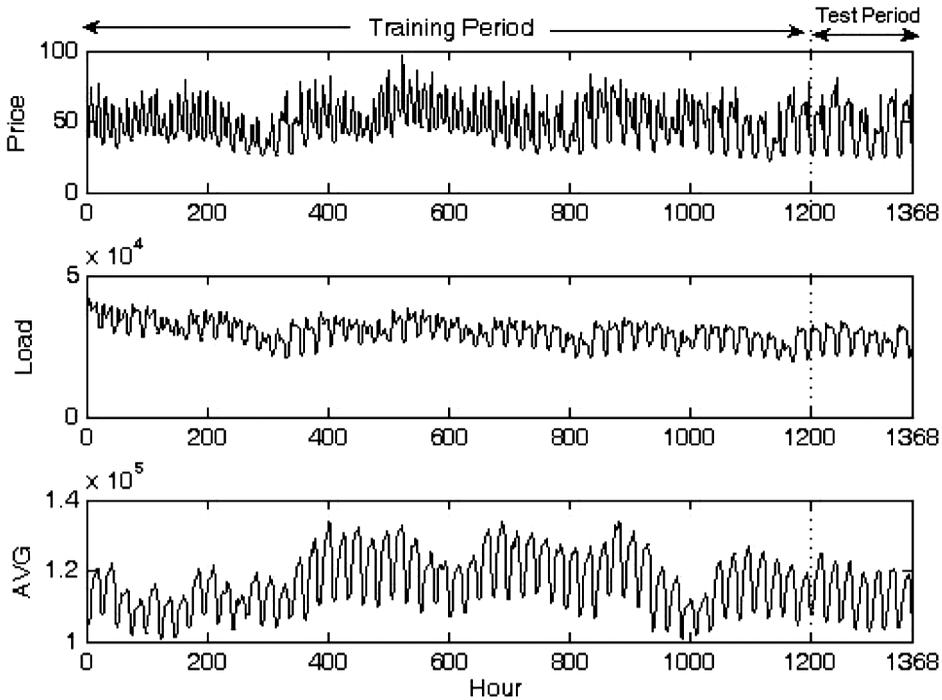


Figure A3. Plot of price, load and AVG for the Spring test week (15–21 May) and its corresponding 50 days training period. Source: Authors.

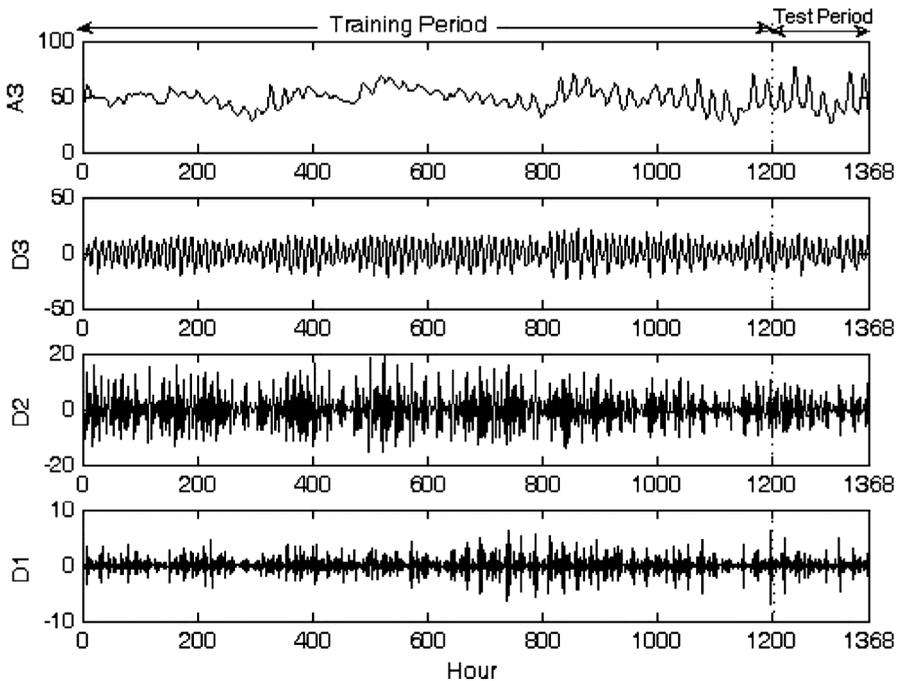


Figure A4. Plot of wavelet components A3, D3, D2 and D1 for the price time series shown in Figure A3. Source: Authors.

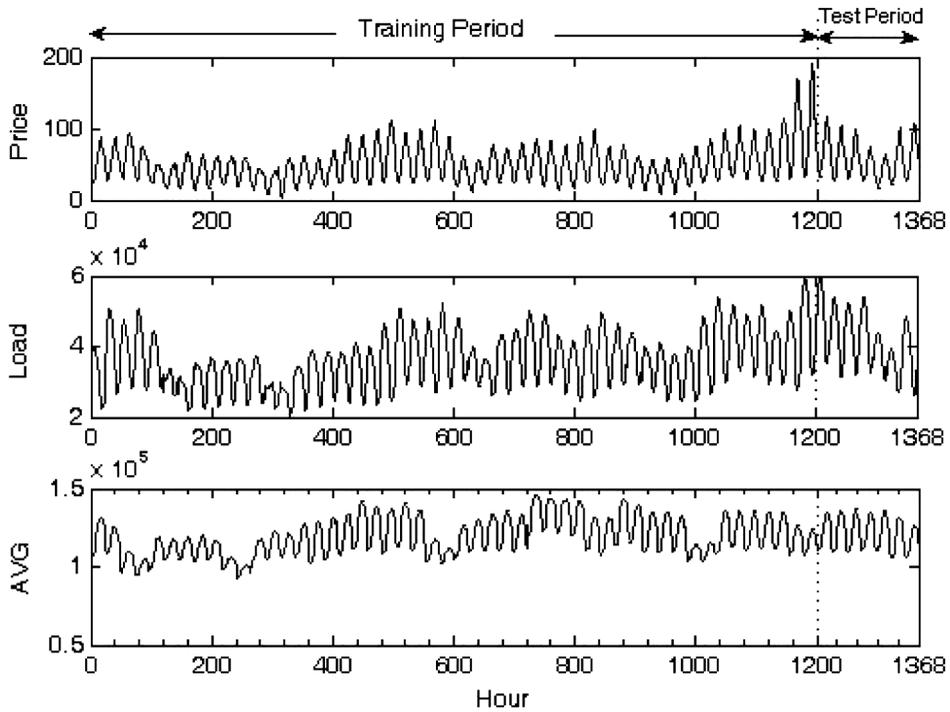


Figure A5. Plot of price, load and AVG for the Summer test week (15–21 August) and its corresponding 50 days training period. Source: Authors.

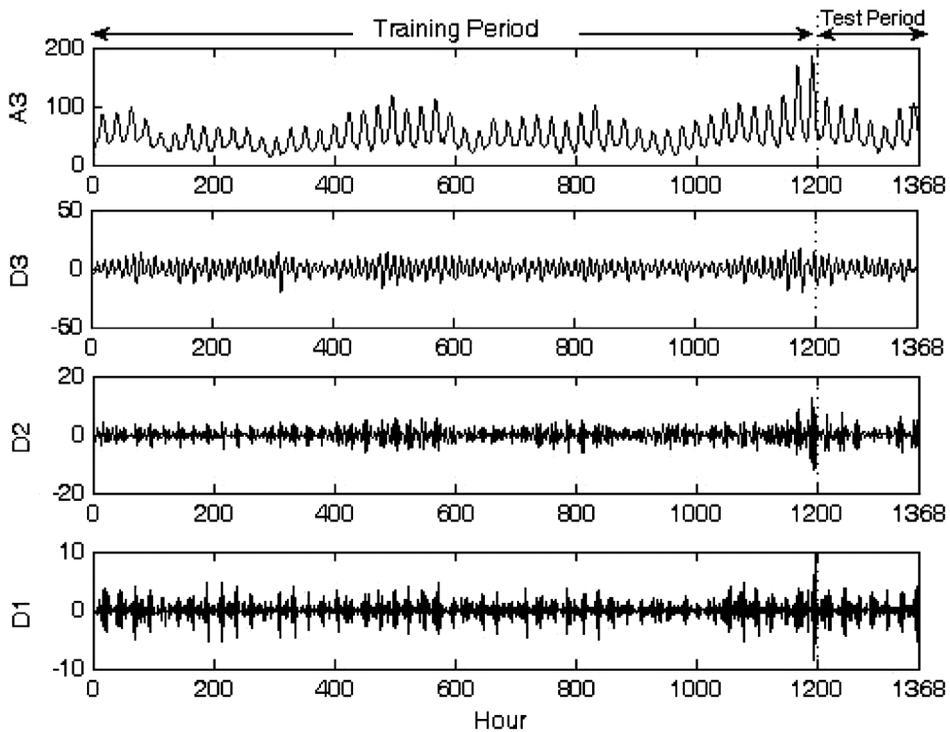


Figure A6. Plot of wavelet components A3, D3, D2 and D1 for the price time series shown in Figure A5. Source: Authors.

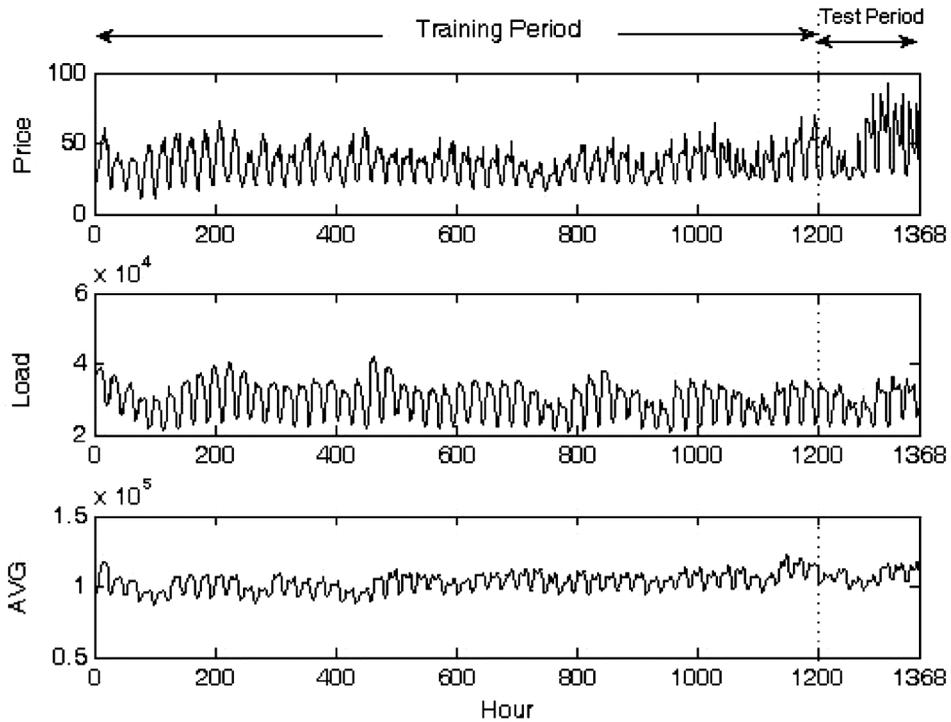


Figure A7. Plot of price, load and AVG for the Autumn test week (15–21 November) and its corresponding 50 days training period. Source: Authors.

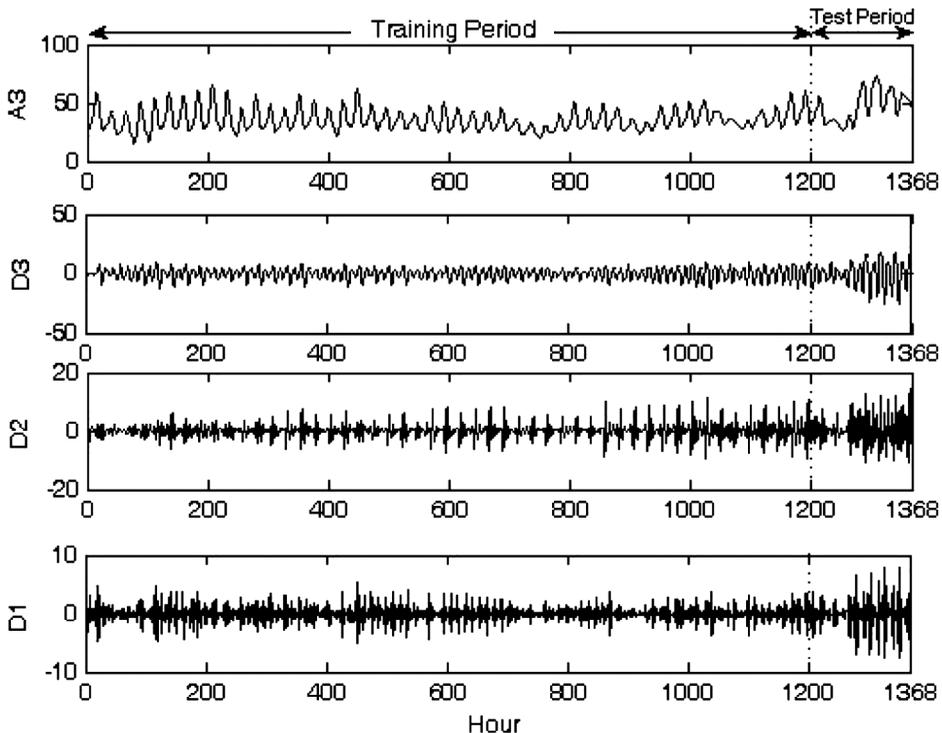


Figure A8. Plot of wavelet components A3, D3, D2 and D1 for the price time series shown in Figure A7. Source: Authors.