A bat optimized neural network and wavelet transform approach for short-term price forecasting

Article *in* Applied Energy · January 2018

DOI: 10.1016/j.apenergy.2017.10.058

4 authors, including:

P.M.R. Bento
Universidade da Beira Interior
7 PUBLICATIONS **25** CITATIONS
SEE PROFILE

José Pombo
Universidade da Beira Interior
29 PUBLICATIONS **92** CITATIONS
SEE PROFILE

Silvio Mariano
Universidade da Beira Interior
100 PUBLICATIONS **993** CITATIONS
SEE PROFILE

Applied Energy 210 (2018) 88–97

Contents lists available at ScienceDirect

Applied Energy

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

A bat optimized neural network and wavelet transform approach for short-term price forecasting

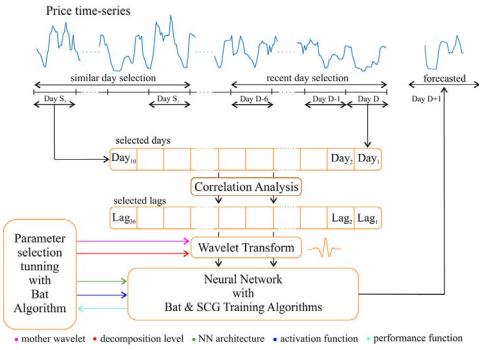
P.M.R. Bento, J.A.N. Pombo, M.R.A. Calado, S.J.P.S. Mariano*

University of Beira interior and Instituto de Telecomunicações, Covilhã, Portugal

HIGHLIGHTS

- We propose a new method for short-term price forecasting (STPF).
- The new method is based on Bat Algorithm, Wavelet Transform and Artificial Neural Networks.
- The method has the capability to auto-tune the best simulation parameters.
- We compare the proposed method in Spanish and Pennsylvania-New Jersey-Maryland (PJM) electricity markets.
- The proposed approach exhibits a better forecasting accuracy.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Artificial neural networks
Bat algorithm
Scaled conjugate gradient
Short-term price forecasting
Similar day selection
Wavelet transform

ABSTRACT

In the competitive power industry environment, electricity price forecasting is a fundamental task when market participants decide upon bidding strategies. This has led researchers in the last years to intensely search for accurate forecasting methods, contributing to better risk assessment, with significant financial repercussions. This paper presents a hybrid method that combines similar and recent day-based selection, correlation and wavelet analysis in a pre-processing stage. Afterwards a feedforward neural network is used alongside Bat and Scaled Conjugate Gradient Algorithms to improve the traditional neural network learning capability. Another feature is the method's capacity to fine-tune neural network architecture and wavelet decomposition, for which there is no optimal paradigm. Numerical testing was applied in a day-ahead framework to historical data pertaining to Spanish and Pennsylvania-New Jersey-Maryland (PJM) electricity markets, revealing positive forecasting results in comparison with other state-of-the-art methods.

1. Introduction

As a widely traded commodity, electricity is sold and bought by producers and consumers who submit their bids, under spot or derivative contracts in a pool-based market [1]. Bids are analyzed by the market operator who then determines the clearing price. But unlike other commodities, electricity cannot be queued and stored economically, with the exception of pumped-storage hydro plants in certain conditions [2].

Therefore, the power industry operates in a very competitive framework, largely due to deregulation and the search for competition policies with the clear intent to reduce marginal costs (and consequently obtain lower consumer electricity prices). In contrast with the previous monopolistic and government-controlled context [1], this deregulated environment creates an additional degree of uncertainty in electricity prices (price volatility). Therefore, the search for reliable and

* Corresponding author.
E-mail addresses: pedrobento194@outlook.com (P.M.R. Bento), Jose_p@portugalmail.com (J.A.N. Pombo), rc@ubi.pt (M.R.A. Calado), sm@ubi.pt (S.J.P.S. Mariano).

http://dx.doi.org/10.1016/j.apenergy.2017.10.058
Received 14 July 2017; Received in revised form 11 October 2017; Accepted 14 October 2017

Nomenclature

Label	Description (Unit)
$a_{j,k}(k)$	wavelet decomposition approximation coefficients
A_i^t	loudness parameter at iteration t
b_j	input signal (bias) for neuron j
$d_{j,k}(k)$	wavelet decomposition detailed coefficients
$\mathrm{E}(w_k)$	error function
f_i^t	frequency assigned to the ith bat at iteration t
$f_{\max}$	bat frequency upper bound
$f_{\min}$	bat frequency lower bound
g_j	activation function value for neuron j
m	number of neurons in a generic layer $k-1$
M	number of wavelet coefficients
n	number of neurons in a generic layer k
N	signal length (number of samples)
p_i	energy probability distribution for wavelet coefficient i
P_t	actual (real) price at time instant t (€/MWh or $/MWh)
$\widehat{P}_t$	forecasted price at time instant t (€/MWh or $/MWh)
P_{Week}	average (mean) weekly price (€/MWh or $/MWh)
q_k	search direction vector at iteration k
r_0	initial pulse emission parameter
r_i^t	pulse emission parameter at iteration t
s_k	Hessian matrix approximation at iteration k
u_i	input signal for neuron j
v_i^t	velocity assigned to the i^{th} bat at iteration t
w_k	weight vector at iteration k
$wt_x(j,k)$	discrete wavelet coefficients (DWT)
$\mathrm{WE}(s)$	wavelet energy per scale
$\mathrm{WE}_{\mathrm{entropy}}(s)$	wavelet entropy per scale
x_i^t	position assigned to the i^{th} bat at iteration t
x^*	current global best solution (bat)
y_j	output signal for neuron j in a generic layer k
z_i	target output response for training sample i
α	loudness constant value
β	random number from a uniform distribution
γ	pulse rate constant value
λ_k, σ_k	scaling factors (SCG)
$\upsilon(t)$	price time-series
$\phi(t)$	scale function (MRA)
$\psi(t)$	mother (base) wavelet function
ω_{ij}	weight of the synaptic connection between neurons i and j

accurate price forecasting techniques, which allow market players to best derive their pool bidding strategies and to optimally schedule energy resources [2,3], can be a differentiating factor between competitors, allowing producers to maximize their profits and consumers to maximize their utilities [4]. As a result, forecasting electricity demand and prices has emerged as one of the major fields of research in electrical engineering [5,6].

A set of details about the price time-series makes the task of forecasting prices far from trivial [7], including high frequency, non-constant mean and variance, multiple seasonality, calendar effect, high level of volatility, and high percentage of unusual price movements [5]. Another set of aspects affecting forecasting accuracy is the uncertainty related with fuel prices, future additions of generation and transmission capacity, regulatory structure and rules, future demand growth, plant operations, and climate changes [2,8].

Presently, some markets with strong penetration of renewable energy sources, particularly wind power (as the Iberian market), have special regulatory regimes (feed-in-tariff scheme) imposing new challenges, mainly due to the fluctuating feed-in of wind power. For example, a day with prices close to zero can be followed by a day of maximum prices, which together with transmission congestion, contributes to price volatility.

A variety of methods have been developed for electricity price forecasting [1] and can also be used for load forecasting [5]. The majority of them, including the present method, focus on lead times ranging from one hour to week-ahead forecasting, with a particular emphasis on day-ahead forecasting, the time horizon corresponding to short-term price forecasting (STPF).

Most methods are based on time-series models, focusing on the past behaviour of price time-series, and can be complemented with some exogenous variables. The first major approach are Parsimonious stochastic methods, for example autoregressive integrated moving average (ARIMA) [9,10] and generalized autoregressive conditional heteroskedastic (GARCH) [11,12] models. A recent approach, proposed in [13], combined ARIMA models and stochastic programming with good results. Other time-series models based on regression are presented in [14,15]. A disadvantage of these methods is their high computational cost [2].

The wavelet transform (WT) is another commonly used feature to better deal with a non-constant mean and variance and a significant number of outliers (also known as spikes). The transformed time-series presents typically a better behaviour (more stable variance and less outliers) than the original price series, thus resulting in a better performance [4,16]. A forecasting model integrating DWT and back-propagation neural networks for financial time-series revealed that the low-frequency components coupled with high-frequency components resulted in higher accuracy performance [17]. For instance in [4,18] the authors combine WT with the classical ARIMA and GARCH models.

Given the limitations presented by the first set of methods, authors focused their efforts on artificial intelligence (AI) methods, based on data-driven structures where input–output mapping is learned from historical samples [19], which constitute the second major group of methods. These are better suited to deal with hard non-linear relationships typical of price time-series, and are thus computationally more efficient [2]. A large portion of these AI methods are centered on Artificial Neural Networks (ANNs). While authors in [2,16,20] used feed forward, radial basis and Elman networks in classical approaches, in [21] authors propose an improved training process for the ANN.

In order to conjugate synergies from different features, the current approaches focus on hybrid methods (combining classical and intelligent computing methods). For example, the use of WT and a fused version of neural networks and fuzzy logic [22]. In [23] the authors used a cascaded neural network (CNN) combined with the Chemical reaction optimization (CRO) algorithm to properly train the CNN. Furthermore an hybrid approach, followed in [24], combined fuzzy ARTMAP, wavelet transform and the firefly algorithm to perform the STPF. A hybrid approach based on the MIMO- LSSVM model, complemented with GMI input selection and WT, and optimized by Quasi-Oppositional Artificial Bee Colony (QOABC) was proposed in [19].

This paper introduces an enhanced method to accurately forecast prices with 24 h lead times, combining synergies from different techniques, in brief: (i) a data selection process, enabling a higher predictive performance, combining both similar and recent days to form the training set, then extracting the most relevant lags using correlation analysis; (ii) a pre-processing phase, using wavelet decomposition, whereby a quantitative method is used to select the best mother wavelet candidates; (iii) an ANN training task, blending the bat algorithm and scaled conjugate gradient algorithm, thus improving neural network learning accuracy; (iv) an optimization strategy to determine parameters involving the ANN architecture and wavelet decomposition.

This paper is organized in the following manner: Section 2 provides a succinct theoretical review of artificial neural networks, scaled

conjugate gradient algorithm, bat algorithm and wavelet transform; Section 3 presents the proposed forecasting methodology assembling the efforts of different techniques; Section 4 presents the case studies used to test and validate the proposed method; Section 5 presents and discusses the numerical results of the proposed method in two different markets; Section 6 presents the main conclusions of this work.

2. Background

2.1. Artificial neural networks

The concept of an ANN emerged inspired by the behaviour of the biological nervous system [25], and since its initial development the field of ANNs has evolved exponentially. Today they are among the best AI methods because of their capacity to deal with complex problems, with hard nonlinear characteristics.

However, a non-trivial set of sensitive aspects have to be decided upon when implementing an ANN. These aspects include the choice of input variables (prices, load, weather variables, etc.), the size of the training vectors, the number of hidden layers, the number of neurons in each hidden layer, the choice of training method, the activation functions and the stop criterion, and the network structure itself, the most common being the multi-layer feed forward neural network (FFNN) [5] as showed in Fig. 1.

In each layer, every neuron response is given by the activation function, with a cost given by a biased weighted sum, which works as a threshold. For two consecutive layers $[k-1,k]$ this can be expressed mathematically as follows:

$$y_j = g_j\left(\sum_{i=1}^{n} \omega_{ij} \times u_i + b_j\right), i \in [0,m] \quad j \in [0,n] \tag{1}$$

where m is the number of neurons in layer $k-1$ and n is the number of neurons in layer k; y_j is the output for neuron j given by the activation function value, g_j, u_i and b_j are input signals; and finally, ω_{ij} is the weight of the synaptic connection between neurons i and j.

A procedure known as supervised learning was carried out, learning is accomplished by updating the weight vector in order to minimize the error, i.e. the difference between the desired response and the network output (through a performance function). Commonly, like in this work, the chosen performance function is the mean squared error (MSE), given by the following expression:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - z_i)^2 \tag{2}$$

where N is the number of training samples, y_i is the neuron output response (in the output layer), and z_i the target output response.

2.2. Scaled conjugate gradient

When training neural networks, ordinary numerical optimization algorithms can be used to minimize error, represented by a performance function. Some traditional methods can behave poorly in large-scale problems, and are sensitive to the choice of some parameters, which can lead to a poor convergence rate [26].

A good choice to train large networks, due to its step size scaling mechanism, is scaled conjugate gradient (SCG) algorithm. SCG avoids a time-consuming line-search per learning epoch, requiring less memory usage, thus making the algorithm faster than other second order algorithms. SCG can train any network as long as its weights, net input and transfer functions have derivative functions. This algorithm combines the model-trust region approach from LM algorithm with the CG approach [26].

The SCG procedure to avoid the time-consuming line-search task is described in Eq. (3), where s_k is the Hessian matrix approximation (second order information), E is the error function (to minimize), E′ is the gradient of E and q_k is the search direction. Scaling factors λ_k and σ_k are introduced to approximate the Hessian matrix.

$$s_k \approx \frac{E(w_k + \sigma_k q_k) - E(w_k)}{\sigma_k} + \lambda_k q_k, 0 < \sigma_k \ll 1 \tag{3}$$

A full description of SCG algorithm implementation can be found in [26].

2.3. Metaheuristic algorithms

These stochastic algorithms are used to find acceptable solutions for hard optimization problems. They can be classified as single-solution when the focus is on a single candidate solution and its trajectory, or population-based when the focus is the collective behaviour of the swarm/population. The latter are better suited to ensure an appropriate search space scan and avoid local minima problems. When using these population based algorithms, we need to guarantee that they stay inside the domain search space, to ensure that we used hyperbolic confinement [27].

The Bat Algorithm (BA) is a bio-inspired metaheuristic algorithm, which has been applied with promising results in [28,29]. With respect to optimizing neural network weights, BA outperformed the traditional BP (back propagation) and LM (Levenberg–Marquardt algorithm), and also PSO (Particle Swarm Optimization) and others bio inspirited algorithms [30,31]. The general idea of BA is to apply bat hunting behaviour to optimization problems, where a swarm of bats (several solutions) try to catch prey (optimal acceptable solution).

First, each bat is assigned a frequency between $[f_{\min} f_{\max}]$ according to the search domain, where $\beta \in [0,1]$ is a random vector obtained from a uniform distribution.

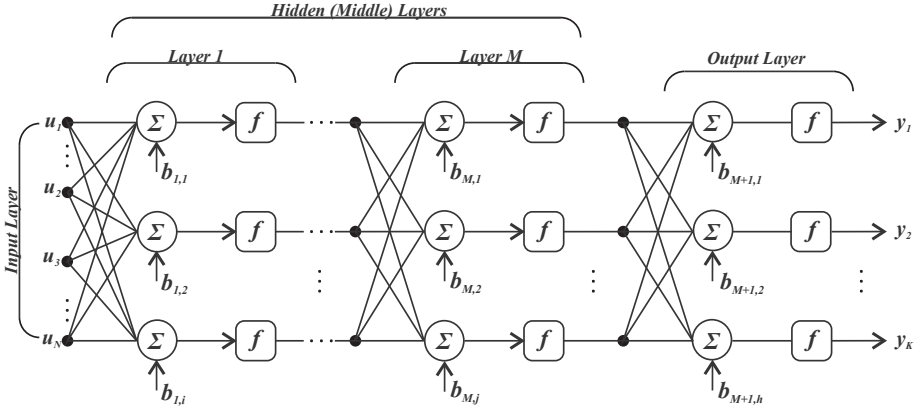


Fig. 1. Multilayer feedforward neural network architecture (FFNN).

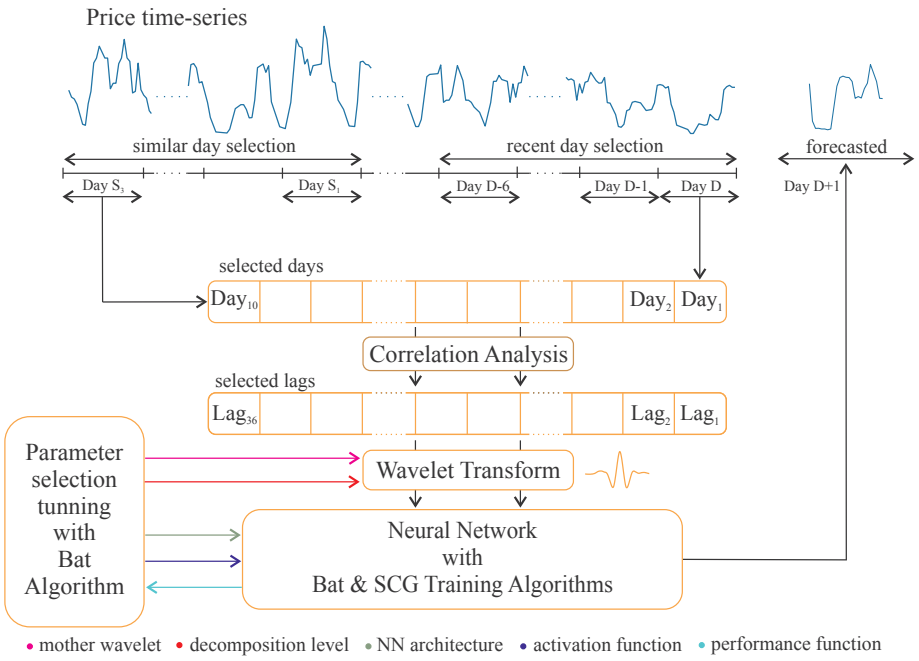


Fig. 2. Proposed methodology (Flowchart).

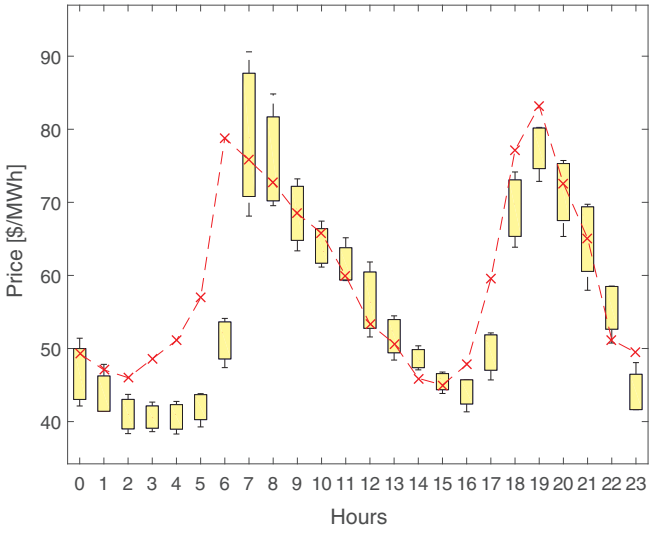


Fig. 3. Similar day based selection representation.

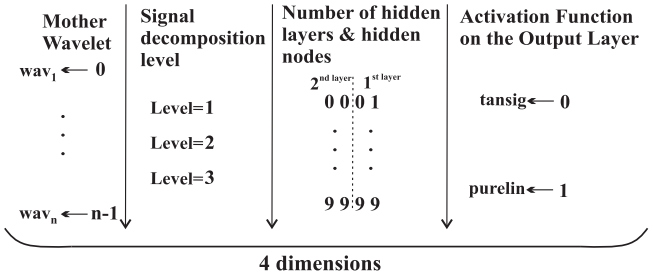


Fig. 4. Problem domain space codification.

$$f_i^t = f_{\min} + (f_{\max} - f_{\min})\beta \tag{4}$$

Second, we define the rules of how positions x_i and velocities v_i are updated inside the multidimensional search domain. New solutions x_i^t e v_i^t are updated each time step t, according to Eqs. (5) and (6):

$$v_i^t = v_i^{t-1} + (x_i^t - x^*)f_i \tag{5}$$

where x^* represents the location of the current global best solution (bat), so the new location is updated accordingly (global walk).

$$x_i^t = x_i^{t-1} + v_i^t \tag{6}$$

A local search procedure is imbedded in the algorithm. As such, at certain times a new solution is generated locally through a random walk process, computing the equation: $x_{\text{new}} = x_{\text{old}} + \varepsilon A^t$, where ε represents a random number in the interval $[-1, 1]$ and A_i^t represents the loudness parameter at each time step t. If the bat is closer to the optimal solution and we accept the new locations as better, then loudness is adjusted as: $A_i^t = \alpha A_i^t$, $\alpha > 0$. Analogously the pulse emission rate r_i^t, is adjusted as: $r_i^t = r_0[1 - exp(-\gamma t)]$, $\gamma > 0$.

This standard BA version can be improved in terms of exploration and exploitation capabilities. A decreasing inertia weight factor and an adaptive frequency modification were implemented in our BA to mitigate the gradually loss of exploitation capability and the loss of local search capacity, respectively, as proposed in [32].

2.4. Wavelet transform

The discrete wavelet transform (DWT) is an implementation of the continuous wavelet transform (CWT) using a discrete set of wavelet scales and translations (dyadic discretization). Compared with CWT, DWT is just as accurate and more efficient [33], and is computed as follows:

$$wt_x(j,k) = \frac{1}{\sqrt{2^j}} \sum_{t=0}^{N-1} v(t) \times \psi^*\left(\frac{t - k2^j}{2^j}\right) \tag{7}$$

where $wt_x(j,k)$ are the wavelet coefficients; j and k are the scaling and translation variables respectively; $v(t)$ is the price time-series; $\psi^*_{j,k}(t)$ is the discretized complex conjugate of the scaled and shifted wavelet function; N is the length of the signal; and t is the discrete time index.

The idea behind DWT is based on multiresolution analysis (MRA), where the original time domain is decomposed into several other scales with different levels of resolution, in a process known as multiresolution decomposition. Two related functions have to be mentioned: first, the mother wavelet function $\psi(t)$ and, secondly, the corresponding scale function $\phi(t)$. Thus, the original $v(t)$ signal can be reconstructed as:

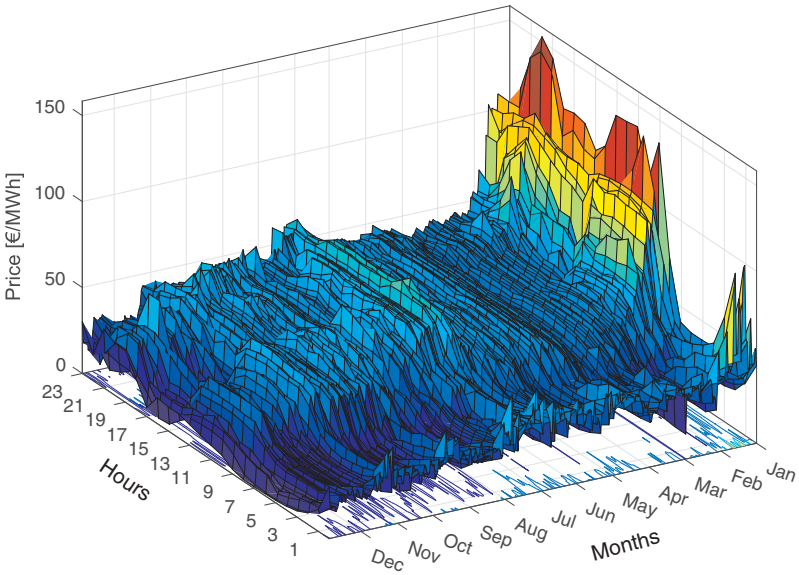


Fig. 5. Electricity prices in the Spanish market at 2002.

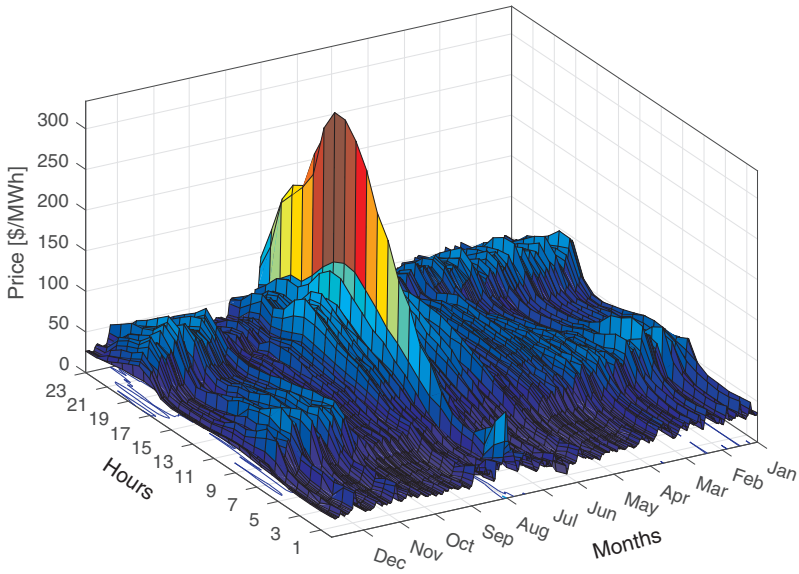


Fig. 6. Electricity prices in the PJM market at 2006.

Table 1
Weekly forecasting errors for the four test weeks (Spain 2002).

	Winter	Spring	Summer	Fall	Average
MAPE (%)	0.61	1.22	1.30	1.81	1.24
$MAPE_{Week}$ (%)	0.57	1.11	1.08	1.46	1.06
EV	1.7E−04	2.7E−04	8.5E−04	6.8E−04	4.9E−04

$$v(t) = \sum_{j=-\infty}^{+\infty} \frac{1}{\sqrt{2^j}} \sum_{k=-\infty}^{+\infty} d_{j,k}(k)\psi_{j,k}(2^j t - k) + \frac{1}{\sqrt{2^{j_0}}} \sum_{k=-\infty}^{+\infty} a_{j_0,k}(k)\phi_{j,k}(2^{j_0} t - k) \quad (8)$$

where j_0 is a predetermined scale, $d_{j,k}(k)$ and $a_{j,k}(k)$ represent respectively the detail (wavelet coefficients) and approximation coefficients.

This process of decomposition is executed iteratively on the approximation coefficients of the previous level, to calculate both detail and approximation coefficients of the next level. Therefore, for a given electricity price series, the approximation coefficients will describe the deterministic components, whereas detail coefficients will uncover the influential fluctuation components (fast varying characteristics). The approximate and detail coefficients thus obtained are afterwards back transformed Eq. (8).

The MRA via WT splits up the price time-series into one series of approximation coefficients (low-frequency) and some subseries of detail coefficients (high-frequency) in the wavelet domain, bringing a filtering effect [4], by decomposing an ill-behaved (original) price series into a set of better-behaved constitutive series (less data variance and outliers). Therefore, price forecasting will have better error improvement [24].

2.5. Mother wavelet selection

When using WT a suitable mother wavelet has to be selected. Nonetheless, choosing a single mother wavelet to decompose large temporal data sets, specifically the price time-series, can be difficult. The common approach is based on qualitative criteria or trial and error tests. Regarding STPF, the choice of small order Daubechies is typically a good choice [4], because of their smoothness and support interval offer a proper trade-off, with special emphasis on order four Daubechies (db4) [18,22,24,34].

An alternative approach was followed, where the “best” mother

Table 2
Comparative $MAPE_{Week}$ (%) for the STPF of the Spanish electricity market.

	Winter	Spring	Summer	Fall	Average
Naive [4]	7.68	7.27	27.30	19.98	15.56[a]
ARIMA [4]	6.32	6.36	13.39	13.78	9.96[a]
Wavelet-ARIMA [4]	4.78	5.69	10.70	11.27	8.11[a]
Mixed-model [48]	6.15	4.46	14.90	11.68	9.30
NN [2]	5.23	5.36	11.40	13.65	8.91
WNN [49]	5.15	4.34	10.89	11.83	8.05
FNN [50]	4.62	5.30	9.84	10.32	7.52
HIS [51]	6.06	7.07	7.47	7.30	6.97
AWNN [52]	3.43	4.67	9.64	9.29	6.75
NNWT [53]	3.61	4.22	9.50	9.28	6.65
CNEA [54]	4.88	4.65	5.79	5.96	5.32
WPA [34]	3.37	3.91	6.50	6.51	5.07
RBFN-PSO [16]	4.27	4.58	6.76	7.35	5.74[a]
Neuro-fuzzy [22]	3.38	4.01	9.47	9.27	6.53[a]
CNN [55]	4.21	4.76	6.01	5.88	5.22
Mixed-Model 2 [56]	4.22	4.39	5.55	5.66	4.95
MRMRMS + RBFNN [57]	3.56	3.90	6.24	6.06	4.94
MRMRMS + MLP [57]	3.60	3.94	6.12	5.83	4.87
MRMRMS + WNN [57]	3.36	3.84	5.96	5.74	4.72
CNN + SSM [23]	3.28	3.62	5.32	5.03	4.31
ARIMA/GARCH + WT [18]	0.63	0.65	1.19	2.18	1.16
Proposed	0.57	1.11	1.08	1.46	1.06

[a] Value determined using the MAPE values from the respective reference.

Table 3
Comparative EV (10^{-4}) for the STPF of the Spanish electricity market.

	Winter	Spring	Summer	Fall	Average
Naive [4]	3.40	4.30	73.8	33.8	28.8[a]
ARIMA [4]	3.40	2.00	15.8	15.7	9.23[a]
Wavelet-ARIMA [4]	1.90	2.50	10.8	10.3	6.38[a]
Mixed-model [48]	NA	NA	NA	NA	NA
NN [2]	1.70	1.80	10.9	13.6	7.00
WNN [49]	NA	NA	NA	NA	NA
FNN [50]	1.80	1.90	9.20	8.80	5.40
HIS [51]	3.40	4.90	2.90	3.10	3.60
AWNN [52]	1.20	3.10	7.40	7.50	4.80
NNWT [53]	0.90	1.70	7.40	4.90	3.70
CNEA [54]	3.60	2.70	4.30	3.90	3.60
WPA [34]	0.80	1.30	5.60	3.30	2.70
RBFN-PSO [16]	1.50	1.90	4.70	4.90	3.25[a]
Neuro-fuzzy [22]	NA	NA	NA	NA	NA
CNN [55]	1.40	3.30	4.50	4.80	3.40
Mixed-Model 2 [56]	1.50	2.90	3.90	5.20	3.38
MRMRMS + RBFNN [57]	NA	NA	NA	NA	NA
MRMRMS + MLP [57]	NA	NA	NA	NA	NA
MRMRMS + WNN [57]	NA	NA	NA	NA	NA
CNN + SSM [23]	1.20	1.20	2.80	1.80	1.70
ARIMA/GARCH + WT [18]	0.20	0.20	0.90	0.80	0.53[a]
Proposed	0.17	0.27	0.85	0.68	0.49

NA – Not available.
[a] Value determined using the EV values from the respective reference.

wavelet candidates were preselected using Minimum Shannon entropy criterion (quantitative criterion). This criterion states that the base (mother) wavelet that minimizes the Shannon entropy of the wavelet coefficients represents the most appropriate base wavelet [35].

First, we determine the wavelet energy per scale $WE(s)$, as follows:

$$WE(s) = \sum_{i=1}^{M} |wt(s,i)|^2 \tag{9}$$

where M is the number of wavelet coefficients and $wt(s,i)$ is the ith wavelet coefficient at scale s. Then, we calculate the energy probability distribution for each wavelet coefficient, defined as:

$$p_i = \frac{|wt(s,i)|^2}{WE(s)} \tag{10}$$

Finally, minimum Shannon entropy, also known as wavelet entropy, is determined as follows:

$$WE_{entropy}(s) = -\sum_{i=1}^{M} p_i \log_2 p_i \tag{11}$$

3. Proposed methodology

The flowchart of the proposed methodology is presented in Fig. 2. This scheme generically illustrates the various steps starting from the price time series and ending in the 24 h price forecast.

3.1. Input selection and features extraction

As illustrated in Fig. 2, the first task was to handpick the training data for the ANN, serving two objectives: to assemble data containing information that enables a good forecast and is reasonably small sized to avoid excessive time consumption. When forecasting prices for a day $D + 1$, price data up until hour 24 of day D were considered to be known.

Generally, when assembling the training data, a significant price data series is necessary to provide a comprehensive set of input patterns to the ANN, therefore ensuring proper mapping between the inputs fed to the ANN and its response. On the other hand, with this sizeable historical price data series comes considerable computational cost and the risk of ANN overtraining. So an alternative is to use similar price days as input patterns, e.g. authors in [36] using the Euclidean norm criteria as a metric for similarity obtained good forecasting results in the PJM market.

To serve this purpose, the process is as follows: given the current (day D) prices, find historical days with a similar price profile (with the lower average of absolute deviations). Assuming given repetitive patterns in the price time series, the targets for these similar days should be reasonably close to the current day's target (day $D + 1$). In addition, price data from recent days, prior to the current day, is also included, providing information about the price recent behaviour.

To build the forecasting model, the input selection is formed with the three most similar days (S_1, S_2 and S_3), day D, and the preceding hourly historical price data of the six days prior to the current day (day D), without information from exogenous variables (simplicity and comparison purposes), for a total of 10 selected days, revealing an acceptable computational cost, as illustrated in Fig. 2.

The reason behind the use of similar day-based selection is illustrated in Fig. 3: the box-and-whisker plot reveals the hourly price distribution of the similar days and the curve in red represents the current day. As we can see for some hours the boxes overlap the curve and for others they are distant from the curve.

The next step is to define the ANN input vector. To achieve this, the most important price values were extracted using correlation analysis (CA), which was performed for each training set, as selected lags may differ. Therefore, the ANN input vector was formed by the 36 most correlated lags, an input with the respective hour, another with the correspondent weekday, and a binary input indicating the presence of a holiday, for a total of 39 input entries.

3.2. Mother wavelet pre-selection

As seen previously, when deciding upon the most suited mother wavelet, the usual approach is to test different ones in the context of the implemented method. This experimentation and evaluation is time consuming, but culminates in a single wavelet, the remaining being abandoned [37]. Alternatively, we could rely on qualitative methods, which are basically supported by past experiences. However, the time gained may not compensate the risk of choosing a non-suitable mother wavelet, therefore not allowing a good extraction of features from our

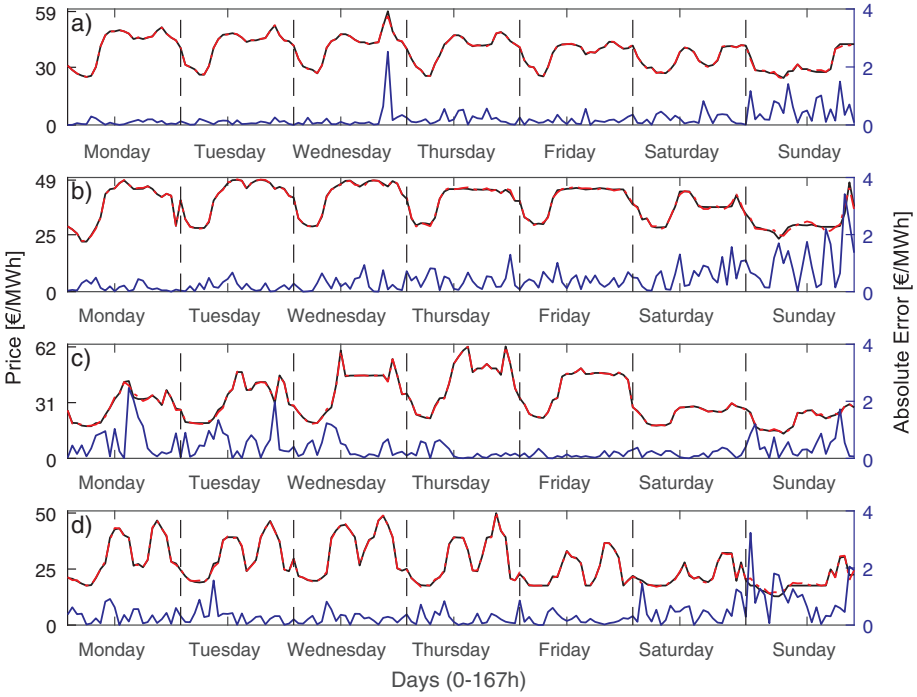


Fig. 7. Spanish market test weeks: (a) winter; (b) spring; (c) summer; and (d) fall. Actual and forecasted prices, respectively, black solid line and red dashed line (left axis) and absolute error, blue solid line (scaled right axis). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Weekly forecasting errors for the four test weeks (PJM 2006).

	Winter	Spring	Summer	Fall	Average
MAPE (%)	0.79	0.80	0.81	0.54	0.75
$MAPE_{Week}$ (%)	0.74	0.72	0.68	0.45	0.65
EV	7.20E−05	4.52E−05	1.54E−05	1.53E−05	3.70E−05

Table 5
Comparative MAPE (%) for the STPF of the PJM Electricity Market.

	Winter	Spring	Summer	Fall	Average
NSA [57]	8.45	8.23	10.09	9.68	9.11
PCA [57]	7.31	8.12	10.28	9.59	8.83
CA [57]	11.10	8.07	9.94	6.82	8.98
MI [57]	6.46	6.34	6.87	6.93	6.38
MR [57]	6.20	6.41	6.38	6.35	6.33
CA + CA [57]	6.11	6.31	6.45	6.29	6.29
MI + CA [57]	5.65	6.10	6.23	5.95	5.98
MR + CA [57]	6.49	5.63	6.16	5.39	5.92
CA + MI [57]	5.59	5.14	5.76	5.62	5.53
MI + MI [57]	4.94	4.99	4.81	4.75	4.87
MR + MI [57]	4.83	4.79	4.81	4.88	4.83
MI + IG [57]	4.71	4.47	4.70	4.43	4.58
MRMRMS [57]	4.62	4.33	4.62	4.38	4.49
NN [21]	6.35	4.74	4.76	7.00	5.71
MLP + LM [18]	9.82	8.87	10.43	9.54	9.67
PCA + CNN [18]	8.61	7.34	8.17	8.36	8.12
WT + NN [18]	4.44	4.31	4.78	4.75	4.57
ARIMA/GARCH + WT [18]	0.73	0.60	0.75	0.86	0.74
Proposed	0.79	0.80	0.81	0.54	0.74

time series.

To speed up the process, we used the minimum Shannon entropy criterion, which reveals the mother wavelets with higher energy concentrations, therefore with low entropy magnitudes, enabling the effective extraction of features by the wavelet transform.

The nine best orthogonal mother wavelets selected were: db2, db3, db4, coif1, fk4, sym2, sym3, sym4 and bior1.1 and as expected there is a prevalence of Daubechies and Symlets mother wavelets.

3.3. ANN training

An important task when using ANN's is its training process, i.e. extraction of existing dependencies and relations from the input training data, without losing its generalization capacity (over-fitting). To reach a proper learning state, traditional approaches regarding supervised networks use gradient descent algorithms (such as SCG) to adjust the weights and minimize the error function Eq. (2). However there are some issues affecting ANN performance, such as sensitivity to the initial positioning, getting trapped in a local minima, and slowness [38].

To address these issues evolutionary BA was used in combination with the SCG algorithm as a training algorithm for the ANN. BA alone can suffer from stagnation [39] and premature convergence in high dimensional problems [40]. Due to this fact the two methods were hybridized.

So, BA and SCG algorithms were both employed in an offline and supervised learning process. The solutions correspond to the neural network weights. The performance function (subject to optimization) is given by Eq. (2). At the end of each BA iteration, the best "bat" is evaluated by the SCG algorithm, which by default (initially) only performs one training epoch. If after this process the error does not decrease, then in the next iteration of the SCG algorithm, the number of learning epochs is doubled. The training process is completed when a minimum number of iterations, and a threshold on the training and validation performance are reached.

In addition, to tackle the initial guess problem, the initial weights were positioned according to the well referenced "Nguyen-Widrow" algorithm [41], to prevent overfitting and ameliorate the initial weights guess problem. A cross-validation technique [42] was also used. Therefore, network performance was evaluated k_{fold} times, alternating the validation subset and training subsets (k_{fold}−1 subsets) in a circular form.

3.4. ANN architecture and wavelet analysis tuning

As mentioned earlier, there is a set of parameters involving neural networks architecture and wavelet analysis for which there is no consensus concerning optimal arrangement [17]. Thereby an optimization

Table 6
Comparative $MAPE_{Week}$ (%) for the STPF of the PJM electricity market.

	Winter	Spring	Summer	Fall	Average
AC + NN [58]	11.29	7.39	8.27	6.89	8.46
PCA + NN [58]	8.13	7.43	8.68	9.57	8.45
RFE + NN [58]	5.97	5.36	5.95	5.73	5.75
GC + NN [58]	5.76	5.24	5.61	5.46	5.52
MR + NN [58]	5.21	5.13	5.57	4.79	5.18
CNN + SSM [23]	3.97	4.02	4.04	4.12	4.04
Proposed	0.74	0.72	0.68	0.45	0.65

algorithm can be implemented to find the best settings, e.g. [43]. In wavelet analysis, the parameters concern the definition of the appropriate decomposition level and the choice of the best mother wavelet, and, in ANNs, the choice of the optimal architecture (number of layers and number of hidden neurons).

To carry out this optimization a BA was used, positioning eleven bats operating over four dimensions (parameters) within an acceptable framework. First, it chooses the mother wavelet from the best candidates, already presented. Second, it determines the best decomposition level for the wavelet analysis, from one to three levels, since more levels of decomposition lead to unnecessary complexity for the ANN. Third, it decides between one or two hidden layers. Although one hidden layer is almost always sufficient it can lead to a huge number of hidden neurons, and therefore a network with two-hidden layers may enable a more reasonable configuration [44]. Each layer contains up to 99 neurons. Fourth and last, it decides upon the activation function on the output layer, to maintain the network's ability to represent linear sections. A linear activation function (*purelin*) is preferable over the common *tansig* [7]. The codification of the domain search space for the BA is illustrated in the Fig. 4.

4. Forecasting framework

Typically, daily price profiles are classified as weekdays (business days), from Monday to Friday, and weekend days, Saturday and Sunday, which are different. Another consideration besides weekends are public holidays, known as the calendar effect, since price profiles on non-holidays are particularly different from those on public holidays [2]. By the same token, this difference in price profiles must be considered when addressing model forecasting accuracy, since price patterns regarding weekdays are fed to the model through the recent days input selection mechanism. Thus, the forecasting performance for weekends and holidays is normally inferior due to weaker cross-correlation with those weekdays patterns, together with a relatively scarce number of holidays and their non-uniform distribution across the calendar [45].

To measure the method's predictive capacity we used three common error measures (criterions) [4,18]. The first is the mean absolute percentage error (MAPE), which is computed as follows:

$$\text{MAPE}\ (\%) = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{P_t - \widehat{P}_t}{P_t}\right| \tag{12}$$

where P_t and $\widehat{P}_t$ are the actual and the forecasted price values at hour t respectively, and $N = 24$ is the number of hourly samples in a day. The second metric is the useful weekly MAPE, which is better to assess problems related with close to zero prices or spiking prices, and is calculated as:

$$\text{MAPE}_{\text{Week}}\ (\%) = \frac{100}{168}\sum_{t=1}^{N}\left|\frac{P_t - \widehat{P}_t}{P_{\text{Week}}}\right| \tag{13}$$

where $P_{\text{Week}} = \sum_{t=1}^{168} P_t$, that is the average weekly price. Finally, the last error measure is the weekly error variance, which is as an important measure of the model uncertainty (variability), and is expressed as:

$$\text{EV} = \frac{1}{168}\sum_{t=1}^{N}\left|\frac{P_t - \widehat{P}_t}{P_{\text{Week}}} - \text{MAPE}_{\text{Week}}\right|^2 \tag{14}$$

As mentioned before, a cross-validation technique was used. Therefore, the data set is partitioned into mutually exclusive subsets ($k_{\text{fold}} = 4$), where 25% of the data set forms the validation set and 75% the training set. The method's effectiveness was tested on two different case studies. Numerical results were obtained using MATLAB® running on a Windows® 10 operative system with 64-bit support with 2.4-GHz-based processor and 8 GB of RAM. The running time was less than 30 min for each forecast. This running time is heavily accounted for the computational time needed to perform the domain space (Fig. 4)

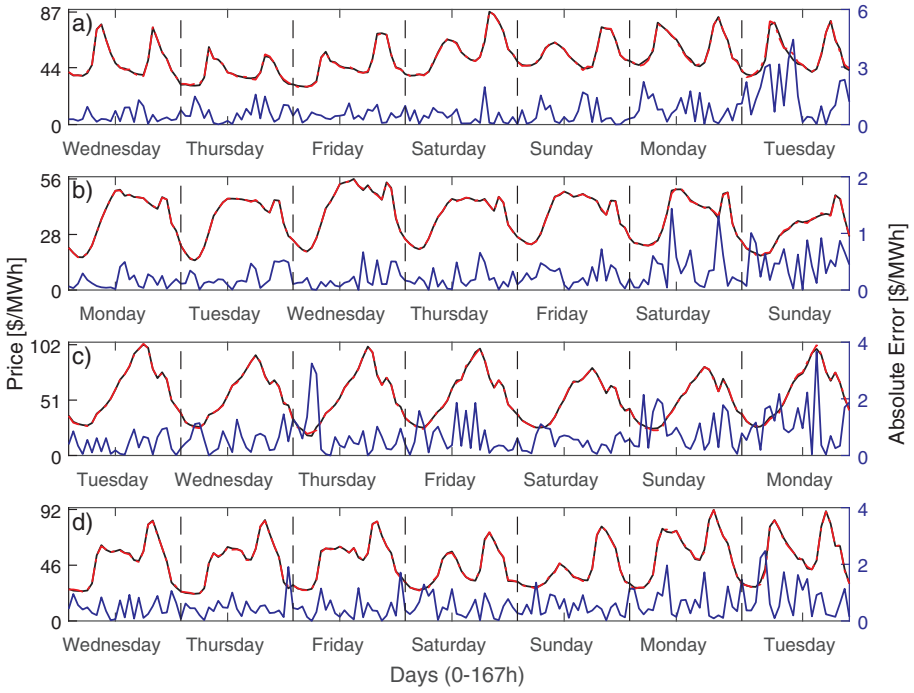


Fig. 8. PJM market test weeks: (a) winter; (b) spring; (c) summer; and (d) fall. Actual and forecasted prices, respectively, black solid line and red dashed line (left axis) and absolute error, blue solid line (scaled right axis). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

optimization. Below we present the two case studies used to validate the proposed method.

4.1. Case I

The first case study was the duopoly Spanish market with a dominant player, resulting in price changes related to the strategic behaviour of the dominant player, which are hard to predict [4]. For comparison purposes the following reference weeks were used [2]: the winter week from February 18 to February 24, 2002; the spring week from May 20 to May 26, 2002; the summer week from August 19 to August 25, 2002 and lastly the fall week from November 18 to November 24, 2002. This data was provided by the Red Electrica de España (REE) and is available in [46].

For each hour and month, prices are represented by a surface plot in Fig. 5. First, a brief statistical analysis was made of the prices in the electricity market of mainland Spain in 2002. The average price and variance for the whole year was 37.4 €/MWh and 263.1 $€^2$/MWh^2 respectively, the maximum recorded price was 158.4 €/MWh on the 10th of January. In contrast, on 2nd of April, prices reached the minimum of 0 €/MWh. A distinctive trait in the winter months, especially in January, is the high average (62.0 €/MWh) and variance (787.4 $€^2$/MWh^2), illustrating a typical seasonal effect in the Spain price time series.

4.2. Case II

The second case study was the Pennsylvania-New Jersey-Maryland (PJM) electricity spot market, one of the Regional Transmission Organizations, which is well-known in the U.S. and beyond, and plays a vital role in the U.S. electric system [23]. For the PJM market, four test weeks corresponding to four seasons of year 2006 were considered for the sake of comparison. The four considered weeks were February 15 to February 21, May 15 to May 21, August 15 to August 21, and November 15 to November 21. Data of PJM electricity market is obtained from [47].

The 2006 prices in the PJM spot market for each hour and month are represented by a surface plot in Fig. 6. As for the first case study, a brief statistical analysis was made. The average price and variance for the whole year were 48.1 $/MWh and 548.6 $\$^2$/MWh^2 respectively. The maximum recorded price was 333.9 $/MWh, on the 3 of August. In contrast, the minimum of 0 $/MWh was reached on several days. A distinctive trait of the summer months of July and August was its high average and variance. For example, in August the variance reached 2242.6 $\$^2$/MWh^2, with an average price value of 67.5 $/MWh.

5. Discussion and numerical results

5.1. Case I

The results obtained for the four weeks referred above are presented in Table 1 for all the selected error measures (MAPE, $MAPE_{Week}$ and EV). These three criterions revealed quite satisfactory results, and as we can see the winter week presents the best results, despite its higher variance, meaning that the data selection and features extraction performed well.

For comparison and validation purposes, results are compared in Tables 2 and 3 with a variety of other price forecasting methods, ranging from the classical Parsimonious methods to the recent hybrid methods (with the same test weeks and criterions). As can be verified, the results of the proposed method are among the best for both the average $MAPE_{Week}$ and EV of all test weeks. Only in the spring week did the ARIMA/GARCH + WT [18] slightly surpass the proposed method. Another peculiarity that can also be observed is that results typically get worse for the summer and fall test weeks [18]. The comparison also revealed that of all the soft computing and hybrid approaches, the proposed method presents a superior accuracy.

The obtained numerical results are also shown in Fig. 7, where each test week has its corresponding subplot. In each subplot, the forecasted prices are represented by a red dashed line and the actual prices are represented by a black solid line (left axis). The absolute error is represented by a blue solid line (scaled right axis). While for weekdays the proposed method presents a small error and the two curves are virtually overlapping, for weekends and holidays this error increases, and the difference between curves becomes more visible, an expected outcome given the reasons presented earlier in the forecasting framework section.

5.2. Case II

Price prediction results of the proposed forecasting method for referred test weeks are presented in Table 4 for all the selected error measures (MAPE, $MAPE_{Week}$ and EV). These three criteria revealed quite satisfactory results, and as we can see the fall week presents the best result. In contrast, the spring week reveals the worst result. As in case I, there was no direct relationship between the variance and the results for the test weeks.

For comparison and validation purposes, criteria are compared in Tables 5 and 6 with a diversity of other price forecasting methods (with the same test weeks and criterions). As shown, the results of the proposed method are among the best for both the average MAPE and $MAPE_{Week}$ of all test weeks. The average MAPE is equal to that presented by ARIMA/GARCH + WT [18] and surpasses the remaining approaches. The comparison also revealed that among all the soft computing and hybrid approaches, the proposed method stands out given the superior accuracy.

The obtained numerical results are also shown in Fig. 8, where each test week has its corresponding subplot. Similarly to the first case study in each subplot, the forecasted prices are represented by a red dashed line and the actual prices are represented by a black solid line (left axis). The absolute error is represented by a blue solid line (scaled right axis). As in case I, the referred figure also show that, for weekdays, the proposed method presents a short error and therefore the two curves are virtually overlapping. However, for weekends and holidays, this error typically increases and the difference between curves is more evident.

6. Conclusion

A novel hybrid method for forecasting one day-ahead electricity price (STPF) is presented. The recent literature reveals a propensity to merge different techniques and algorithms in hybrid approaches. Following this path, the design behind the proposed hybrid method came as a response to a set of constraints concerning STPF, by combining recent/similar day selection (best input features selection), wavelet transform, neural networks, bat and scaled conjugate gradient algorithms in an optimized arrangement. A particular feature of this method is the use of BA to perform the domain space optimization, addressing sensible issues such as neural network architecture and wavelet parameter selection. Furthermore, several key aspects were taken into consideration, including the selection of the most suitable mother wavelets to enhance DWT performance, as well as aspects concerning ANN performance: initial weights guess, overfitting avoidance and computational cost reduction. A quantitative method (minimum Shannon entropy) was responsible for the referred appropriate mother wavelet selection, constituting another highlight of the proposed method. For these reasons, the proposed approach can be applied to different electricity markets without loss of performance.

In both study cases (the commonly used Spanish and Pennsylvania-New Jersey-Maryland electricity markets) the performed optimization favored ANN architectures with 2 hidden layers, a linear activation function (*purelin*) in the output layer, Symlets 3 and 4 and Daubechies 3, all up to 3 decomposition levels. Numerical results and comparative

works were conducted to illustrate the merits of the proposed approach, revealing average MAPE values under and around 1%, which in comparison with other literature revealed the method's superior forecasting accuracy. In addition, results obtained for special days, such as weekends, showed the method's capacity to capture complex price time-series characteristics.

Acknowledgements

The authors wish to thank Instituto de Telecomunicações for funding this work, with the investigation scholarship PEst-OE/EEI/LA0008/2013.

References

[1] Weron R. Electricity price forecasting: a review of the state-of-the-art with a look into the future. Int J Forecast 2014;30(4):1030–81.
[2] Catalão JPS, Mariano SJPS, Mendes VMF, Ferreira LAFM. Short-term electricity prices forecasting in a competitive market: a neural network approach. Electr Power Syst Res 2007;77(10):1297–304.
[3] Mandal P, Srivastava AK, Senjyu T, Negnevitsky M. Electricity price forecasting using neural networks and similar days. 1st ed. John Wiley & Sons; 2017. p. 215–50.
[4] Conejo AJ, Plazas MA, Espinola R, Molina AB. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Trans Power Syst 2005;20(2):1035–42.
[5] Aggarwal SK, Saini LM, Kumar A. Electricity price forecasting in deregulated markets: a review and evaluation. Int J Electr Power Energy Syst 2009;31(1):13–22.
[6] Monteiro C, Fernandez-Jimenez LA, Ramirez-Rosado IJ. Explanatory information analysis for day-ahead price forecasting in the Iberian electricity market. Energies 2015;8(9):10464–86.
[7] Panapakidis IP, Dagoumas AS. Day-ahead electricity price forecasting via the application of artificial neural network based models. Appl Energy 2016;172:132–51.
[8] Singh N, Mohanty SR. A review of price forecasting problem and techniques in deregulated electricity markets. J Power Energy Eng 2015(September):1–19.
[9] Contreras J, Espínola R, Nogales FJ, Conejo AJ. ARIMA models to predict next-day electricity prices. IEEE Trans Power Syst 2003;18(3):1014–20.
[10] Conejo AJ, Contreras J, Espı R, Plazas MA. Forecasting electricity prices for a day-ahead pool-based electric energy market. Int J Forecast 2005;21:435–62.
[11] Liu H, Shi J. Applying ARMA-GARCH approaches to forecasting short-term electricity prices. Energy Econ 2013;37:152–66.
[12] Hua ZHZ, Li XLX, Li-zi ZLZ. Electricity price forecasting based on GARCH model in deregulated market. In: 2005 Int. Power Eng. Conf.; 2005.
[13] Sánchez AA, Nieta D, González V, Contreras J. Portfolio decision of short-term electricity forecasted prices through stochastic programming. Energies 2016:1–19.
[14] Nogales FJ, Contreras J, Conejo AJ, Espínola R. Forecasting next-day electricity prices by time series models. IEEE Trans Power Syst 2002;17(2):342–8.
[15] Crespo Cuaresma J, Hlouskova J, Kossmeier S, Obersteiner M. Forecasting electricity spot-prices using linear univariate time-series models. Appl Energy 2004;77(1):87–106.
[16] Shafie-Khah M, Moghaddam MP, Sheikh-El-Eslami MK. Price forecasting of day-ahead electricity markets using a hybrid forecast method. Energy Convers Manage 2011;52(5):2165–9.
[17] Lahmiri S. Wavelet low- and high-frequency components as features for predicting stock prices with backpropagation neural networks. J King Saud Univ – Comput Inf Sci 2014;26(2):218–27.
[18] 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(11):3606–10.
[19] Shayeghi H, Ghasemi A, Moradzadeh M, Nooshyar M. Simultaneous day-ahead forecasting of electricity price and load in smart grids. Energy Convers Manage 2015;95:371–84.
[20] Chintham V, Vardhan NH. Electricity price forecasting of deregulated market using Elman neural network. In: IEEE INDICON; 2015. p. 1–5.
[21] Gholipour Khajeh M, Maleki A, Rosen MA, Ahmadi MH. Electricity price forecasting using neural networks with an improved iterative training algorithm. Int J Ambient Energy 2017:1–12.
[22] Catalão JPS, Pousinho HMI, Mendes VMF. Short-term electricity prices forecasting in a competitive market by a hybrid intelligent approach. Energy Convers Manage Feb. 2011;52(2):1061–5.
[23] Abedinia O, Amjady N, Shafie-khah M, Catalão JPS. Electricity price forecast using combinatorial neural network trained by a new stochastic search method. Energy Convers Manage 2015;105(2015):642–54.
[24] Mandal P, Haque, Meng J, Srivastava, Martinez R. A novel hybrid approach using wavelet, firefly algorithm, and fuzzy ARTMAP for day-ahead electricity price forecasting. IEEE Trans Power Syst 2013;28(2):1041–51.
[25] Abraham TH. (Physio)logical circuits: the intellectual origins of the Mcculloch-Pitts neural networks. J History Behav Sci 2002;38(1):3–25.
[26] Møller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 1993;6(4):525–33.
[27] Rua D, Montero J, Lu J, Martinez L, Kerre EE. Proceedings of the 8th international FLINS conference. In: Computational intelligence in decision and control; 2008.
[28] Yang XS. A new metaheuristic Bat-inspired algorithm. Stud Comput Intell 2010;284:65–74.
[29] Fister I, Yang X-S, Fong S, Zhuang Y. Bat algorithm: Recent advances. In: CINTI 2014 - 15th IEEE international symposium on computational intelligence and informatics, proceedings; 2014. p. 163–7.
[30] Khan K, Sahai A. A comparison of BA, GA, PSO, BP and LM for training feed forward neural networks in e-learning context. Int J Intell Syst Appl 2012;4(7):23–9.
[31] Khan K, Sahai A, Talal R. Comparative Study between the (BA) Algorithm and (PSO) Algorithm to Train (RBF) network at data classification. Int J Intell Syst Appl 2014;92(5):23–9.
[32] Yilmaz S, Kucuksille EU. Improved Bat Algorithm (IBA) on continuous optimization problems. Lect Notes Softw Eng 2013;1(3):279–83.
[33] Reis AJR, Alves da Silva AP. Feature extraction via multiresolution analysis for short-term load forecasting. IEEE Trans Power Syst 2005;20(1):189–98.
[34] Catalão JPS, Pousinho HMI, Mendes VMF. Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. IEEE Trans Power Syst 2011;26(1):137–44.
[35] Gao RX, Yan R. Wavelets: theory and applications for manufacturing. Wavelets Theory Appl Manuf 2011:1–224.
[36] Mandal P, Senjyu T, Uezato K, Funabashitm T. Several-hours-ahead electricity price and load forecasting using neural networks. In: 2005 IEEE power Eng. Soc. Gen. Meet.; 2005. p. 2205–12.
[37] Li S, Wang P, Goel L. A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection. IEEE Trans Power Syst 2016;31(3):1788–98.
[38] Bashir ZA, El-Hawary ME. Applying wavelets to short-term load forecasting using PSO-based neural networks. IEEE Trans Power Syst 2009;24(1):20–7.
[39] Fister I, Brest J, Yang XS. Modified bat algorithm with quaternion representation. In: 2015 IEEE Congr. Evol. Comput. CEC 2015 - Proc.; 2015. p. 491–8.
[40] Zhu B, Zhu W, Liu Z, Duan Q, Cao L. A novel quantum-behaved bat algorithm with mean best position directed for numerical optimization. Comput Intell Neurosci 2016;2016.
[41] Pavelka A, Procházka A. Algorithms for initialization of neural network weights. In: Sb. Prisp. 12. rocniku Konf. MATLAB 2004, vol. 2; 2004. p. 453–9.
[42] Setiono R. Feedforward neural network construction using cross validation. Neural Comput 2001;2877(1):2865–77.
[43] Jaddi NS, Abdullah S, Hamdan AR. Optimization of neural network model using modified bat-inspired algorithm. Appl Soft Comput J 2015;37:71–86.
[44] Guoqiang Zhang B, Eddy Patuwo MYH. Forecasting with artificial neural networks: the state of the artF. Int J Forecast 1998;14:35–62.
[45] Gil HA, Gómez-Quiles C, Gómez-Expósito A, Santos JR. Forecasting prices in electricity markets: needs, tools and limitations. In: Pardalos PM, Sorokinl A, Rebennackl S, Iliadisl NA, Pereira MVF, editors. Handbook of networks in power systems I. Berlin, Heidelberg; 2012. p. 141–2.
[46] REE. Daily Market Hourly Price [Online]. Available: < http://www.omie.es/files/flash/ResultadosMercado.swf > [accessed: 04-Apr-2017].
[47] PJM. Day-Ahead LMP [Online]. Available: < http://www.pjm.com/markets-and-operations/energy/real-time/monthlylmp.aspx > [accessed: 01-Jan-2017].
[48] García-Martos C, Rodríguez J, Sánchez MJ. Mixed models for short-run forecasting of electricity prices: Application for the Spanish market. IEEE Trans Power Syst 2007;22(2):544–52.
[49] Lora AT, Santos JMR, Exposito AG, Ramos JLM, Santos JCR. Electricity market price forecasting based on weighted nearest neighbors techniques. IEEE Trans Power Syst 2007;22(3):1294–301.
[50] Amjady N. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. IEEE Trans Power Syst 2006;21(2):887–96.
[51] Amjady N, Hemmati M. Day-ahead price forecasting of electricity markets by a hybrid intelligent system. Eur Trans Electr Power 2009;19(1):89–102.
[52] Pindoriya NM, Singh SN, Singh SK. An adaptive wavelet neural network-based energy price forecasting in electricity markets. IEEE Trans Power Syst 2008;23(3):1423–32.
[53] Catalão JPS, Pousinho HMI, Mendes VMF. Neural networks and wavelet transform for short-term electricity prices forecasting. In: 15th International Conference on Intelligent System Applications to Power Systems, ISAP '09, vol. 18; 2009. p. 85–92.
[54] Amjady N, Keynia F. Day-Ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. IEEE Trans Power Syst 2009;24(1):306–18.
[55] Amjady N, Daraeepour A. Design of input vector for day-ahead price forecasting of electricity markets. Expert Syst Appl 2009;36(10):12281–94.
[56] Amjady N, Daraeepour A. Mixed price and load forecasting of electricity markets by a new iterative prediction method. Electr Power Syst Res 2009;79(9):1329–36.
[57] Abedinia O, Amjady N, Zareipour H. A new feature selection technique for load and price forecast of electrical power systems. IEEE Trans Power Syst 2017;32(1):62–74.
[58] Amjady N, Daraeepour A. Day-ahead electricity price forecasting using the relief algorithm and neural networks. In: 5th International Conference on the European Electricity Market; 2008. p. 1–7.