

Spot Energy Price Forecasting Using Wavelet Transform and Extreme Learning Machine

Lucas Barth da Silva
Industrial and Systems Engineering
Graduate Program (PPGEPS)
Pontifical Catholic University of Paraná
(PUCPR)
Energy Company of Paraná (COPEL)
Curitiba, Brazil
lucas.barth@copel.com

Roberto Zanetti Freire
Industrial and Systems Engineering
Graduate Program (PPGEPS)
Pontifical Catholic University of Paraná
(PUCPR)
Curitiba, Brazil
roberto.freire@copel.com

Osiris Canciglieri Junior
Industrial and Systems Engineering
Graduate Program (PPGEPS)
Pontifical Catholic University of Paraná
(PUCPR)
Curitiba, Brazil
osiris.canciglieri@pucpr.br

Abstract— Given the social importance of energy, there is a concern to promote the sustainable development of the sector. Aiming at this evolution, from the 90s onwards, a wave of liberalization in the sector began to emerge in various parts of the world. These measures promoted an increase in the dynamism of commercial transactions and the transformation of electricity into a commodity. Consequently, futures, short-term, and spot markets were created. In this context, and due to the volatility of energy prices, the forecast of monetary values has become strategic for traders. This work aims to apply a computational intelligence model using Wavelet Transform on input values and the Extreme Machine Learning algorithm for training and prediction (W-ELM). The macro parameters were optimized using the Particle Swarm Optimization algorithm and for the selection of the input variables, a model based on Mutual Information (MI) was used. In the end, the methodology was compared with the traditional methods: Autoregressive Moving Averages (ARIMA) and General Autoregressive Conditional Heteroskedasticity (GARCH) models. Results showed that the W-ELM had better performance for forecasting 1 to 4 weeks of when compared to ARIMA. When the GARCH model results were considered, the proposed method provided worse performance only for 1 step ahead forecasting.

Keywords—Price, Electricity, Forecasting, Wavelet, Extreme Machine Learning; Neural Networks.

I. INTRODUCTION

Electric energy is a basic source of various social benefits, in addition to being associated with a country's industrial and technological advances. The Gross Domestic Product (GDP) growth is intrinsically correlated with the increase in energy consumption, generating the need to increase production as there is economic development. In the 70s and 80s, mainly, problems in the sector's planning arose in several countries (both developed and underdeveloped), either due to poorly sizing capacity or poor management of state monopolies, generating crises due to high levels of indebtedness and inefficiency operating concessionaires. In this context, several countries had to create conditions to increase the sector's liquidity and the development of greater profit-generating opportunities by changing the competitive model [1]. With the liberalization of energy markets, price forecasting becomes essential for the success of sales agents.

The value of energy depends on several factors, depending on the form of relationship between players in the trading environment, determined by local legislation and the composition of the country's energy matrix. In Brazil, which has spot prices defined by computer models, the operation depends on future scenarios, mainly composed of

the following factors: plant reservoir storage levels, river inflows, load forecast, risk aversion curve, planning of generation and transmission expansion, inter-regional transmission limits, availability of generation and transmission equipment, cost of thermal generation, cost of possible energy deficit and a discount rate [2]. The objective of this work is to carry out a forecast of energy prices in the Brazilian Free Market for up to 4 weeks ahead. For this, the price curve and exogenous variables that are most correlated were used. The model used was the Extreme Learning Machine algorithm, which is composed of a single-layer feedforward neural network, with inputs decomposed by Wavelet's transform.

This decomposition tool has been used successfully due to its ability to transform series into their high and low-frequency elements, enabling the capture of breaks, discontinuities, and particularities of the original data [3]. Since there is great volatility in energy prices, the analysis under the aspect of frequency decomposition makes the tool useful for learning patterns that are not easily observable through traditional models that adapt to more linear or smoothed series. In the energy market, several works have been published using the combination of Wavelet Transform (WT) and Neural Networks, as in [4] that proposed an adaptive wavelet neural network (AWNN), in [5] which developed a model combining discrete wavelet transform (DWT) and stochastic recurrent wavelet neural network (SRWNN) and in [6] that used a scheme combining WT, stacked autoencoder (SAE) and long short-term memory neural network (LSTM). In addition to these, there are several examples, some of which will be cited in the next paragraphs.

Analyzing the general literature on energy price forecasting, there has been an increase in publications from 2000 onwards. Initially, there was a considerable number of works focusing on regressive models. In the work of [7], the Autoregressive Conditional Heteroskedasticity (GARCH) method was presented. [8] and [9] used ARIMA models. In the work of [10], hybrid models were introduced, making predictions through dynamic regression combined with Wavelet functions. The first proposition of a tool with computational intelligence method was elaborated by [11], where fuzzy neural networks were applied to predict the energy price in the Canadian market.

In the last 5 years, there has been a trend towards research into hybrid models. As in [12], which presented a model with inputs decomposed by Wavelets, Neural Networks with a descending stochastic gradient, combined with a bat algorithm for learning. Or in [13], in which the

authors used the Wavelet transform to extract characteristics from the input data, performing the inverse transform for the high and low frequencies components. So, the prediction was performed by a generic neural network. For the adjustment of parameters, an algorithm called Improved Method of Adapting to the Environment (IEAM) was used. In [14], the authors propose a model based on Swarm Particle Optimization (PSO), Variational Mode Decomposition, Seasonal Auto-Regressive Integrated Moving Average (SARIMA), and deep belief network (DBN). Also, [15] presents a neural network algorithm called Long Short-Term Memory (LSTM) with Differential Evolution optimization for tuning the model.

The main contribution of this work is the development of a hybrid method using Wavelet transform and Extreme Learning Machine, with parameter improvement through Particle Swarm Optimization and variable selection using the Mutual Information tool. The next section of this paper presents the tools used in the work to propose the prediction method. In section III, the model proposed in this study is detailed. Then, in section IV, the results of both optimization and predictions are presented. Finally, in sections V and VI, the analysis of results and conclusions were addressed.

II. MATERIALS AND METHODS

In this section, the tools used in all steps for the elaboration of the proposed model will be explained.

A. Variables selection based on Mutual Information (MI)

To differentiate the variables that are relevant to the price and those that are irrelevant or redundant, a Mutual Information based methodology was used. First, it is necessary to understand the concept of MI. Its definition is given in terms of the probability density functions of the variables:

$$I(x; y) = \iint p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy \quad (1)$$

For the selection of variables, the two-stage algorithm was described by Peng, Long & Ding (2005). The first finds a set of candidate variables through the selection method called maximum relevance - minimum redundancy:

$$\max D(S, y), D = \frac{1}{S} \sum_{x_i \in S} I(x_i; y) \quad (2)$$

$$\min R(S), R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \quad (3)$$

$$\max \Phi(D, R), \Phi = D - R \quad (4)$$

where S is the set of variables, x_i and x_j are observations of variables i and j , and y is the output of the supervised learning model. Then, a wrapper is used that wraps the subsets of variables with a classifier (Naive Bayes for example), whose objective is to minimize the categorization error.

B. Wavelet Transform

Each time series forecasting methodology must, within its work context, deal with the different characteristics of the time series. Thus, data pre-processing, especially models with data dissociation at different frequencies, has the function of filtering noise and non-linearities [17].

The Wavelet Transform is a signal processing method that results in a useful decomposition in terms of frequency and time, which makes this methodology suitable for non-stationary processes such as the energy price series [18].

To do so, it uses a function as a base (Mother-Wavelet) and performs the transform through dilated (or contracted) and translated versions of this equation, which are used as signal filters [19].

According to [20], the origin of this process is the Fourier analysis, in which the periodic function of the signal is generated by the superposition of complex exponentials. The extension of this idea is carried out through the generation of a space in $L^2(\mathbb{R})$ arising from the superposition of contracted (or dilated) shapes and translated from the mother function ψ by

$$\Psi_{a,b}(x) = |a|^{\frac{1}{2}} \psi\left(\frac{x-b}{a}\right), b \in \mathbb{R}, a > 0 \quad (5)$$

where ψ is the Mother-Wavelet, and a and b are dilation (contracted and expanded waveforms) and translation factors, respectively.

To expand the data to different resolutions, a scale equation is used according to Eq. (6):

$$W(x) = \sum_{k=0}^{N-1} (-1)^k c_{k+1} \psi(2x + k) \quad (6)$$

where N is the number of observations in the series, and c_k are the Wavelets coefficients, which should be thought of as filters that are applied to the original data [21].

C. Mother Wavelet Selection

An important aspect of model construction is choosing the ideal Mother-Wavelet. Therefore, a method based on Shannon's Entropy was used. According to [22], entropy is a measure of the uncertainty of a random variable. The metric also serves as a measure of joint information between two random variables [23]. Considering that, according to Gao e Yan (2010), the Mother-Wavelet that minimizes the joint entropy between the original signal and the Wavelets coefficients is the most suitable for decomposition, the selection of the Mother-Wavelet was performed using the Minimum Entropy criterion of Shannon. According to [12], given the Wavelet energy per scale (WE):

$$WE(x) = \sum_{k=1}^N |W(x)|^2 \quad (7)$$

where N is the number of Wavelet coefficients and $W(x)$ is the scalar coefficient. The energy probability distribution can be calculated by:

$$p_i = \frac{|W(x)|^2}{WE(x)} \quad (8)$$

Then, the minimum Shannon entropy is calculated according to the equation:

$$WE(x) = \sum_{i=1}^N p_i \log_2 p_i \quad (9)$$

D. Neural Networks

According to [24], computational learning processes have different characteristics and have the power to create other algorithms. Instead of the direct logic of performing the processing and generating results, these machines receive the inputs and the respective observed outputs from a system to understand and reproduce them.

Based on the human nervous system, which has the capacity for plasticity, that is, to adapt to the environment, neural networks are mainly composed of three types of layers: input, hidden layers, and output. These levels can have several neurons, and the synapses are represented by the mathematical relationships between these neurons. These links occur simply, in practice, as follows: a weighted sum is calculated through an input vector x , multiplied by the weight w of each neuron. The result is passed to an activation function “ φ ”, which determines the reaction force of the output. The solution of this layer is then sent to the next level, where the scheme is repeated until the output, as shown in Fig.1. In each step, it is possible to add a random value called bias, which aims to translate the function activation to facilitate the adaptation of the network to the data [25].

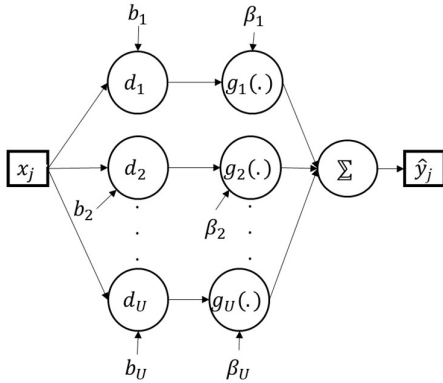


Fig 1. Representation of a Single Layer Feedforward Neural Network.

Mathematically, a single-layer neural network with U neurons can be described based on the following equation:

$$\sum_{i=1}^U \beta_i g(d_i x_j + b_i) = \hat{y}_j \quad (10)$$

where β_i is the weight value of the connection of hidden neuron i to the output layer, g is the activation function of the intermediate level, d_i is the weight of the connection between the input and the respective node, x_j is the value of the input of pattern j , b_i is the value of bias in neuron i and \hat{y}_j is the expected output for pattern x_j [26].

The purpose of the network is to approximate \hat{y}_j with the real output y_j in a set \tilde{N} of sample patterns so that

$$\sum_{j=1}^{\tilde{N}} |\hat{y}_j - y_j| = 0 \quad (11)$$

by modifying the values of β_i , d_i e b_i .

These equations can be written compactly in matrix form:

$$H\beta = Y \quad (12)$$

where H is an $\tilde{N} \times U$ matrix containing the values of $g(d_i x_j + b_i)$, β and Y are the $U \times m$ matrices with the numbers of β and $\tilde{N} \times m$ of y .

E. Extreme Learning Machine

According to [27], there are three types of network architecture: single-layer feedforward, multilayer feedforward, and recurrent networks. The first is the simplest, where there is an input layer that is projected onto an output layer of neurons, but not the other way around and is therefore acyclic.

The second is distinguished from the previous one by the presence of one or more hidden layers. The last type is distinguished from feedforward forms in that it has at least one feedback loop. The ELM network is part of the second group and has only a hidden layer with U neurons, without feedback.

The advantage of this algorithm is that it is not necessary to adjust many parameters, making it faster and more generalized compared to traditional descending gradient models. The ELM system applied in this work was described in [26]. To formulate the model, the authors proved two theorems:

1) In a feedforward network with U hidden neurons and activation function indefinitely differentiable for any interval, with U sampled pairs (x_j, y_j) and β_i , d_i e b_i chosen randomly, according to any continuous probability distribution (thus the probability is 1), the matrix H will be invertible and $\|H\beta + Y\| = 0$.

2) Given any small value for $\varepsilon \geq 0$, under the same conditions as the theorem above, there is $\tilde{N} < U$ so that for any U sampled pairs distinct from (x_j, y_j) , $\|H_{\tilde{N} \times U} \beta_{U \times m} + Y_{\tilde{N} \times m}\| = 0$.

Therefore, the system solution lies in the minimization of the norm $\|H_{\tilde{N} \times U} \beta_{U \times m} + Y_{\tilde{N} \times m}\|$, which, in this model, is performed by the matrix multiplication solution Generalized inverse of Moore–Penrose, so that

$$\|H\hat{\beta} - Y\| = \min_{\beta} \|H\beta - Y\| \quad (13)$$

ELM Algorithm: Given a training set $T = \{(x_j, y_j) | x_j\}$, activation function $g(x)$, and number of hidden neurons U ,

Step 1: Set randomly the input weights d_i and the bias b_i , with $i = 1, \dots, U$.

Step 2: Compute the output of hidden layer H .

Step 3: Calculate the weight of the output β , according to (13)

F. Particle Swarm Optimization

The macro parameters choice (Mother-Wavelet, number of neurons in the hidden layer, activation function, and number of decompositions in the Wavelet transform), a Particle Swarm Optimization algorithm was used.

The model is part of the meta-heuristic optimization systems, which are based on patterns in nature, more

specifically on the observation of behavior in groups. To explain more simply, given a problem defined in a domain of dimension d , a population of N candidate solutions called $\{i_i\}$, $i \in \mathbb{R}$ is randomly determined where each i_i moves in the search space at a speed v_i [28].

According to [29], to find an optimal solution, the algorithm is structured based on the following steps:

Step 1: Initializes the first population of particles and their corresponding velocity. Then, the error concerning the particles is calculated and the best solution among the candidates is found.

Step 2: For each interaction, the speed depends on two factors: the best Global (obtained from all interactions) and the best Location (calculated on the current interaction), given by the following equation:

$$v_t = v_{t-1} + ci * rand * (p_{best} - p_{actual}) + cg * rand * (g_{best} - p_{actual}) \quad (14)$$

where v_t is the updated velocity and v_{t-1} is the current one. ci and cg are scalars in which the first is an individual cognitive value of the particle and the second is a "social." p_{best} is the best location reached by the individual particle, p_{actual} is the current state and g_{best} is the best position found, considering the entire population.

Step 3: After calculating the new velocity, the particle moves in the domain with the new value, following the formula:

$$p_{new\ actual} = p_{actual} + v_t \quad (15)$$

Step 4: The best Global and the best Local are updated.

Step 5: Check the completion criterion (maximum number of interactions or minimum expected error).

III. METHODOLOGY

In this stage of the work, the specifications of the model and the methodology applied to the dataset will be described.

A. Input Variables

According to [30], despite the numerous benefits of renewable energy sources, there are several challenges due to their intermittence. Since Brazil has a predominantly hydraulic energy matrix, in phases of water scarcity, energy must be supplied by other sources, mainly thermal, which are more expensive and more polluting. That said, it appears that the price of energy is inversely proportional to the behavior of hydrology. Therefore, it is necessary to add exogenous variables that contribute to the construction of dispatch scenarios for price forecasting.

Given this context, and that the study focused on forecasting energy prices for the Southeast/Midwest region of Brazil, the following variables were analyzed to verify the relevance of the composition of monetary values: energy exchange with the northeast region, affluent natural energy of the region, amount of hydraulic generation, amount of thermal generation and level of the Capivara river, energy stored in the region's reservoirs, affluent natural energy of the Paran basin, affluent natural energy of the Paranapanema River. Since the short-term energy operation planning in Brazil considers, as assumptions, hydroelectric

generation scenarios (inflows and levels of reservoirs) and system thermal, as well as energy exchanges between subsystems, the above variables are the main ones that influence the variation in energy prices [31].

The inputs variables choice was made according to the tool described in section II, A of this article, by comparing the price series with each exogenous variable. The components that had resulted in similar or small mutual dependence (redundancy and low relevance) were rejected, reducing the amount of data and the computational cost for the model training. Then, it was concluded that the variables to be used as input are the first five, plus the price history.

B. Input Vector

Initially, in choosing which exogenous variables would be chosen for input, the MI tool took into account the entire history of the series.

After this stage, the price series was divided into sub-series of 6-week periods and then selected the most similar with the series consisting of the study week (S) and its 5 previous ones. The choice of 6-week periods was based on the work of [12], which obtained good results in a similar work carried out in the energy market in Spain. The choice was made through the ordering of the smallest average deviations associating the period S to $S - 5$. At this stage, 11 data sets were selected, totaling 66 observations.

Then, the calculation of autocorrelations (ACF) of the sampled prices was performed and the 36 dates most correlated with week S were chosen. Then, the exogenous variables referring to these days were selected, resulting in 180 data in the input layer.

C. Parameters

In the first step regarding the selection of the parameters, the analysis of the input data concerning the Mother Wavelets was performed according to the methodology described in II.C, and then the 10 functions with the lowest final entropy were selected.

The candidate activation functions were determined based on the fit conditions described in 2.5, namely: sigmoid, sine, radial basis, hard limit, symmetric hard limit, symmetric saturating linear, hyperbolic tangent sigmoid, and linear. The others to be determined are the number of neurons in the hidden layer, between 1 and 99, and the decompositions by Mother-Wavelet, those selected among Daubechies, Least Asymmetric, Best Localized, and Coiflet functions.

For Particle Swarming Optimization the "metaheuristicOpt" package was used [32]. It was implemented in R language and configured according to Table I.

TABLE I. PSO PARAMETERS

Population Number	Max Iter. Number	Max. Velocity	ci	cg
11	35	4	2	2

Data for parameter adjustment were divided as follows: 70% for training and testing and 30% for validation. The first part was performed using the repeated cross-validation method, in which the data were divided into 200 groups of standards, training, and tests were performed through

interactive resampling between sets. Then, a prediction is performed with the data separated for validation and verified if there is a proximity between the errors generated in the tests and the validation. This procedure guarantees the generalization of the model.

The statistical method used to calculate the error to compare different parameters was the Root Mean Squared Error (*RMSE*), given by:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (16)$$

where n is the number of values sampled, y_j is the observed output of the system, and \hat{y}_j is the forecast value.

The measure used to compare the final result between models was the Mean Absolute Percentage Error (*MAPE*), which is calculated using the following function:

$$MAPE = \frac{100}{n} \sum_{j=1}^n \frac{(y_j - \hat{y}_j)}{y_j} \quad (17)$$

D. Framework

After the formation of the input vectors, each series of variables are decomposed using the Wavelet transform described in II.B. Then, the neural network is trained, and the energy price forecast is performed. The graphical representation of the construction of the model resulting from this work is shown in Fig. 2.

IV. FORECASTING

From this stage on, the results of predictions for 1 to 4 steps ahead will be presented and compared with the traditional ARIMA and GARCH models. The first model was parameterized using the Akaike information criterion (AIC), resulting in a model without autoregressive order, 1 degree of differentiation, and 5 moving averages (ARIMA(0,1,5)). Using the same criterion, a GARCH (1,1) model was chosen.

A. Database

The information used for training and testing the model and parameterization of the autoregressive systems were extracted from the period between January 7, 2012, and May 16, 2015. The input data for model validation are the price results observed in the period between May 23, 2015, until May 12, 2018.

All values are publicly available on the website of the Chamber of Energy Trading in Brazil (CCEE) and the National System Operator (ONS).

B. Model Results

The comparison of the final result between the models, performed by Mean Absolute Percentage Error, is shown in Table I.

TABLE II. COMPARISON AMONG MODELS USING THE MAPE METRIC

	1-week ahead	2-weeks ahead	3-weeks ahead	4-weeks ahead
ARIMA	23.57	39.36	60.63	72.01
GARCH	9.14	45.00	35.20	51.82
W-ELM	21.36	31.27	29.48	30.33

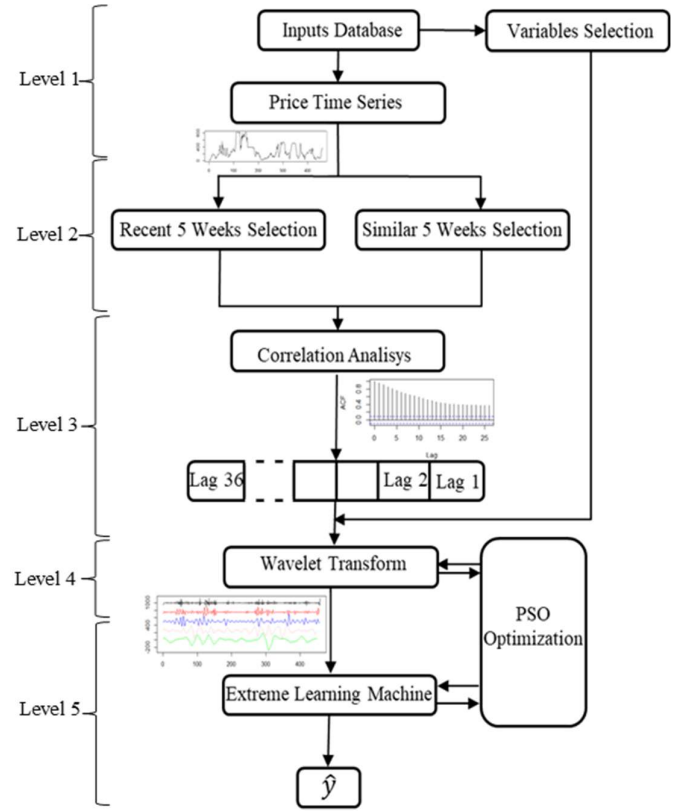


Fig 2. Graphical Representation of the Proposed Model.

Level 1) Separation of the database into price series and exogenous series;

Level 2) Selection of current 6-week series and 10 of similar historical subseries and formation of new series with 36 elements;

Level 3) Carrying out the auto-correlation analysis in this series and choosing the 36 most correlated Lags. Then the values of exogenous variables from the selected periods are added to the model;

Level 4) Performing Wavelet Transformation;

Level 5) Forecasting with Extreme Learning Machine.

After PSO optimization, the parameters of each model were chosen (1 to 4 steps ahead): The best Mother-Wavelet, the level of decomposition, the activation function used in the Extreme Learning Machine algorithm, and the number of neurons in the hidden layer. The result can be seen in Table II, with the respective Root Mean Squared Error values.

TABLE III. SELECTED PARAMETERS FOR EACH ELM MODEL

Prediction Horizon	1-week	2-weeks	3-weeks	4-weeks
Wavelet Function	Daubechies	Coiflet	Coiflet	Daubechies
Function Length	6	12	12	2
Decomp. Level	3	1	1	1
Neurons Number	42	47	83	76
RMSE (Validation)	52.47	71.63	65.3	70.17

V. RESULTS

In this part, the analysis of the results and points subject to improvement observed in the implementation of

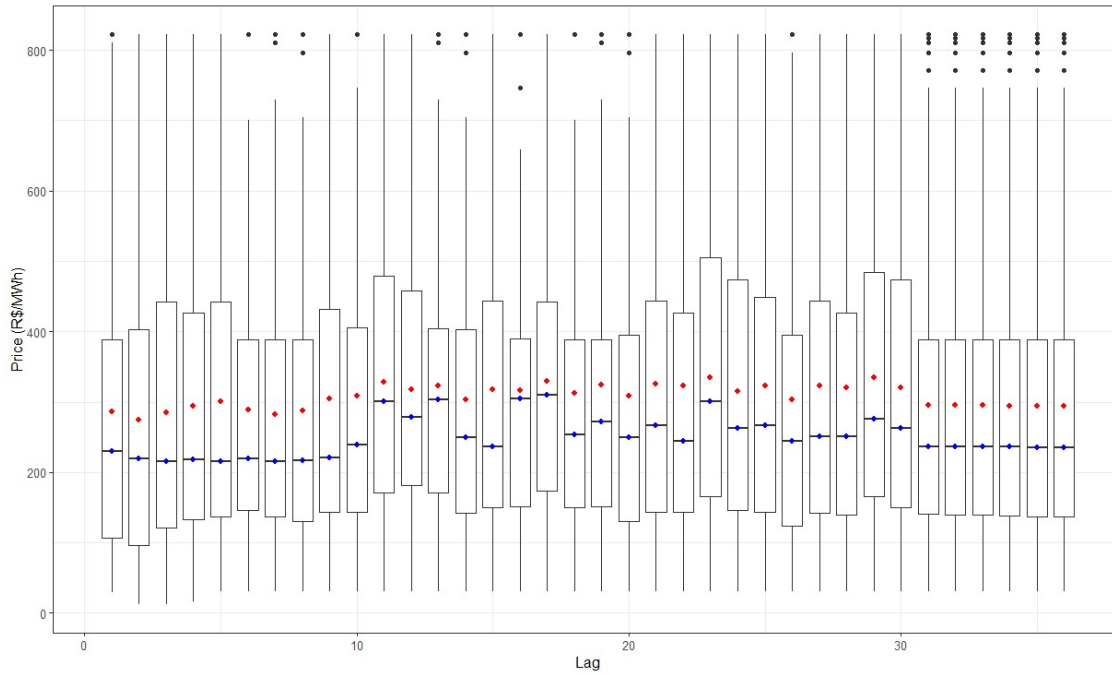


Fig 3. Price distribution for each lag of all model inputs. The red dot refers to the mean and the blue dot to the median.

the volume available in reservoirs and dispatching thermoelectric plants in greater the amount. Therefore, the model failed to detect these anomalies, which culminated in spikes in system errors. Regarding the comparison with the ARIMA and GARCH models, the W-ELM methodology proved to be better in all step-ahead scenarios, except for the one-step-ahead forecast of the GARCH model, which had higher accuracy in detecting volatilities.

Due to the aforementioned problem, the model is subject to improvement in the choice of parameters and input variables. Each interaction lasted, on average, 12 minutes. Therefore, a reassessment is possible that can include more levels of Wavelets decomposition, the

the model will be carried out. The first discussion to be carried out at this point is about the application of correlation analysis, where 36 lags were selected as input to the model. As shown in Appendix A - Fig. 3, it is possible to observe that there was standardization in the distribution of values of delayed inputs, which helps in learning the neural network.

As can be seen in Appendix A - Fig. 4, the forecast of the model followed the actual values. Larger errors can be observed at the end of 2017 and the beginning of 2018. This can be explained by atypical price volatilities in this period due to the severe drought that occurred in 2017, reducing

addition of other inputs that correct the aforementioned anomalies, the inclusion of hidden layers in the neural network, the use of other methods for parameter optimization, the use of other pre-processing methods, such as time series clustering, or experimenting with other numbers of correlated lags as input.

It is concluded, then, that the model has a good performance in predicting energy prices in the Brazilian free market since the forecast curve follows the observed values, however, it can still be improved, given the aforementioned error peaks and the high computational cost, by experimenting with other pre-processing models, parameters and optimization methods.

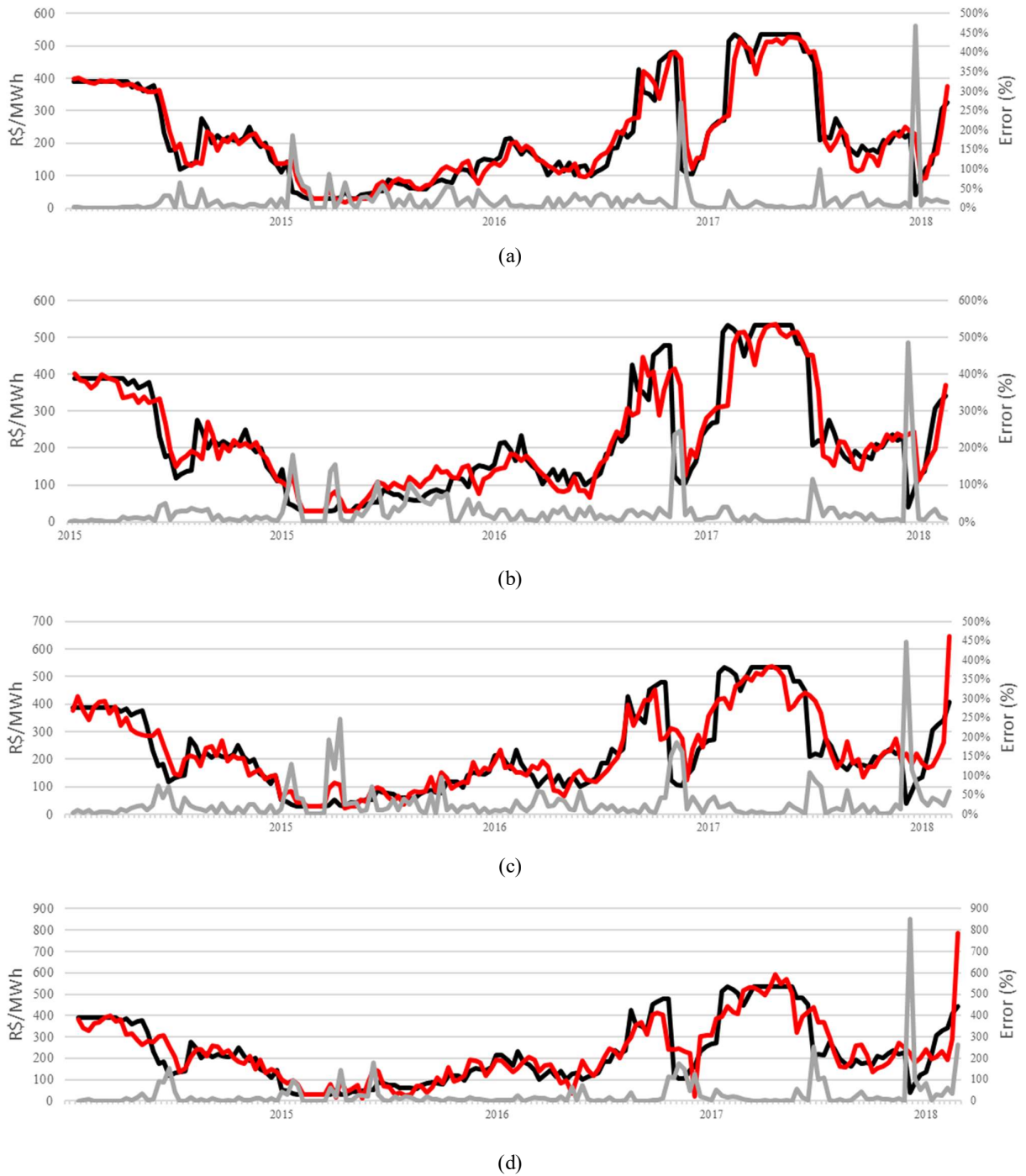


Fig 4. Price prediction (Brazilian Reais per Megawatts/hour) 1-4 steps ahead (a, b, c and d). Model outputs in red, actual observations in black, and errors in percentage on the secondary axis, in gray.

VI. CONCLUSION

This work presented a new model for forecasting energy prices in the Brazilian free market. The scheme used a method based on Mutual Information for the selection of input variables, autocorrelation for the choice of lags used in the composition of the input vector, Wavelet transform for the decomposition of the signal into its referent

frequencies as a function of time and the algorithm of Extreme Learning Machine for learning and prediction.

The parameters referring to Mother-Wavelet, number of signal decomposition levels, number of neurons, and activation function were selected after applying the Particle Swarm Optimization model. One model was used for each level of prediction, 1 to 4 steps ahead.

The model showed good results, since the forecast curve followed the real values, doing better than traditional methods, such as ARIMA and GARCH, except for the 1 step ahead forecast, where the latter obtained better results. However, at some points, there were peaks of errors, verified in times of hydrological anomalies.

For future works, the exercise of finding methodologies that recognize times of environmental anomalies and better adjust to these periods is recommended. Another point to be studied in addition to the study of other changes in the model parameters that were not included in this work, such as the prediction without Wavelet.

REFERENCES

- [1] S. C. Bhattacharyya, *Energy economics: Concepts, issues, markets and governance*. Londres: Springer-Verlag London Ltd., 2019.
- [2] R. Mayo, *Derivados de eletricidade & gerenciamento de risco: Roberto Mayo*. Synergia, 2009.
- [3] S. Lahmiri, "Wavelet low- and high-frequency components as features for predicting stock prices with backpropagation neural networks," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 26, no. 2, pp. 218–227, 2014.
- [4] N. M. Pindoriya, S. N. Singh, and S. K. Singh, "An adaptive wavelet neural network-based energy price forecasting in electricity markets," *IEEE Trans. Power Syst.*, 2008.
- [5] L. Huang and J. Wang, "Forecasting energy fluctuation model by wavelet decomposition and stochastic recurrent wavelet neural network," *Neurocomputing*, vol. 309, pp. 70–82, 2018.
- [6] W. Qiao and Z. Yang, "Forecast the electricity price of U.S. using a wavelet transform-based hybrid model," *Energy*, vol. 193, p. 116704, 2020.
- [7] R. C. Garcia, J. Contreras, M. van Akkeren, and J. B. C. Garcia, "A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 867–874, 2003.
- [8] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Trans. Power Syst.*, vol. 18, pp. 1014–2020, 2003.
- [9] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Trans. Power Syst.*, vol. 20, pp. 1035–1042, 2005.
- [10] A. J. Conejo, F. J. Nogales, R. Espinola, and J. Contreras, "Forecasting next-day electricity prices by time series models," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 342–348, 2002.
- [11] C. P. Rodriguez and G. J. Anders, "Energy Price Forecasting in the Ontario Competitive Power System Market," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 366–374, 2004.
- [12] P. M. R. Bento, J. A. N. Pombo, M. R. A. Calado, and S. J. P. S. Mariano, "A bat optimized neural network and wavelet transform approach for short-term price forecasting," *Appl. Energy*, vol. 210, pp. 88–97, 2017.
- [13] N. Singh, S. Ranjan Mohanty, and R. Dev Shukla, "Short term electricity price forecast based on environmentally adapted generalized neuron," *Energy*, vol. 125, pp. 127–139, 2017.
- [14] J. Zhang, Z. Tan, and Y. Wei, "An adaptive hybrid model for short term electricity price forecasting," 2019.
- [15] L. Peng, S. Liu, R. Liu, and L. Wang, "Effective long short-term memory with differential evolution algorithm for electricity price prediction," *Energy*, vol. 162, pp. 1301–1314, Nov. 2018.
- [16] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [17] M. H. D. M. Ribeiro, S. F. Stefenon, J. D. de Lima, A. Nied, V. C. Mariani, and L. dos S. Coelho, "Electricity Price Forecasting Based on Self-Adaptive Decomposition and Heterogeneous Ensemble Learning," *Energies*, vol. 13, no. 19, 2020.
- [18] Z. Yang, L. Ce, and L. Lian, "Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods," *Appl. Energy*, vol. 190, pp. 291–305, Mar. 2017.
- [19] R. Gençay, F. Selçuk, and B. Whitcher, "DISCRETE WAVELET TRANSFORMS," in *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, R. Gençay, F. Selçuk, and B. Whitcher, Eds. San Diego: Academic Press, 2002, pp. 96–160.
- [20] P. A. Morettin, *Ondas e ondaletas: da análise de Fourier à análise de ondaletas*. Edusp, 1999.
- [21] A. Graps, "An Introduction to Wavelets," *IEEE Comput. Sci. Eng.*, 1995.
- [22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2^o. Wiley, 2006.
- [23] R. X. Gao and R. Yan, *Wavelets: Theory and Applications for Manufacturing*. Springer US, 2010.
- [24] P. Domingos, *O algoritmo mestre - Como a busca pelo algoritmo de machine learning definitivo recriará nosso mundo*. São Paulo, BR: Novatec Editora Ltda., 2015.
- [25] J. Weissmann, G. Mazzola, and G. Milmeister, *Comprehensive Mathematics for Computer Scientists 2: Calculus and ODEs, Splines, Probability, Fourier and Wavelet Theory, Fractals and Neural Networks, Categories and Lambda Calculus*. Berlin: Springer-Verlag Berlin Heidelberg, 2005.
- [26] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.
- [27] S. S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. New Jersey, US: Prentice Hall, 1999.
- [28] D. Simon, *Evolutionary Optimization Algorithms*. Wiley, 2013.
- [29] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, 1995, vol. 4, pp. 1942–1948 vol.4.
- [30] M. Alipour, J. Aghaei, M. Norouzi, T. Niknam, S. Hashemi, and M. Lehtonen, "A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration," *Energy*, vol. 205, p. 118106, 2020.
- [31] C. de P. de E. E.- Cepel, "Planejamento de Curto Prazo e Programação da Operação Energética." [Online]. Available: http://www.cepel.br/pt_br/linhas-de-pesquisa/menu/planejamento-de-curto-prazo-e-programacao-da-operacao-energetica.htm. [Accessed: 21-Aug-2021].
- [32] L. Septem Riza, Iip, E. Prasetyo Nugroho, M. B. Adi Prabowo, E. Junaeti, and A. G. Abdullah, "metaheuristicOpt: Metaheuristic for Optimization." 2019.