Electricity energy price forecasting based on hybrid multi-stage heterogeneous ensemble: Brazilian commercial and residential cases

Matheus Henrique Dal Molin Ribeiro*†, Ramon Gomes da Silva*, Cristiane Canton ‡, Naylene Fraccanabbia§, Viviana Cocco Mariani§¶ and Leandro dos Santos Coelho*¶

*Industrial & Systems Engineering Graduate Program (PPGEPS), Pontifical Catholic University of Parana (PUCPR), Curitiba, Parana, Brazil. 80215–901

†Department of Mathematics, Federal Technological University of Parana (UTFPR), Pato Branco, Parana, Brazil. 85503–390

‡Pato Branco University Center (UNIDEP), Pato Branco, Parana, Brazil. 85503–378

§Mechanical Engineering Graduate Program (PPGEM), Pontifical Catholic University of Parana (PUCPR), Curitiba, Parana, Brazil. 80215–901

¶Department of Electrical Engineering, Federal University of Parana (UFPR), Curitiba, Parana, Brazil. 81530–000

Emails: matheus.dalmolinribeiro@gmail.com, gomes.ramon@pucpr.edu.br, canton.cristiane@gmail.com, nayfraccanabbia@outlook.com, viviana.mariani@pucpr.br, leandro.coelho@pucpr.br

***Abstract*—The development of accurate models to forecast electricity energy prices is a challenge due to the number of factors which can affect this commodity. In this paper, a hybrid multi-stage approach is proposed to forecast multi-step-ahead (one, two and three-month-ahead) Brazilian commercial and residential electricity energy prices. The proposed data analysis combines the pre-processing named complementary ensemble empirical mode decomposition (CEEMD) in the first stage coupled with the coyote optimization algorithm (COA) to define the CEEMD's hyperparameters, aiming to deal with time series non-linearities and enhance the model's performance. On the next stage, four machine learning models named extreme learning machine, Gaussian process, gradient boosting machine, and relevance vector machine are employed to train and predict the CEEMD's components. Finally, in the final stage, the results of the previous step are directly integrated to compose a heterogeneous ensemble learning of components to obtain the final forecasts. In this case, a grid of models is obtained. The best model is one that has better generalization out-of-sample. Through developed comparisons, results showed that combining COA-CEEMD with a heterogeneous ensemble learning can develop accurate forecasts. The modeling developed in this paper is promising and can support decision making in electricity energy price forecasting.**

***Index Terms*—Decomposition, electricity prices, ensemble learning, forecasting, optimization.**

I. Introduction

The electricity power systems plays a key role in the citizen's life as well as in the economy of society. The analysis of the energy market, especially electricity price forecasting plays a crucial role in families and commerce's strategic planning. This analysis allows families and managers to use the forecasting information to adjust their finances. Indeed, the developing of high-accurate forecast models for electricity energy prices are important, but hard due to these data presenting high frequency, volatility, non-linearity, and seasonality [1]. Moreover, other factors such as weather and energy demand, and the impact of renewable energy sources [2] make the forecasting process a challenge.

Considering the developing of efficient forecasting models, the most common strategies are adopting ensemble of models as well as to hybridize several approaches such as pre-processing (decomposition methods), optimization (single and multi-objective algorithms), and artificial intelligence models (nonlinear/machine learning models) [3]–[6]. An ensemble of models works on the divide-and-conquer scheme. In this structure, a set of weak models are used to solve the same task and when combined (average rule for regression problems) they generate a strong model [7]. Indeed, pre-processing approaches such as decomposition methods aim to extract the data noise and non-linearities, where the original signal is decomposed into certain signals (namely components) with different frequencies [8]. In this respect, in the forecasting field, different models (heterogeneous ensemble of components) or the same model (homogeneous ensemble of components) can be used to train and predict each decomposed component. The final forecast can be obtained by aggregation (directly aggregation) of the results of the previous step [9]. Alongside this, evolutionary algorithms can be used to obtain the hyperparameters of machine learning models [10] or hyperparameters of decomposition methods [11], aiming to

The authors would like to thank the National Council of Scientific and Technologic Development of Brazil – CNPq (Grants number: 307958/2019-1-PQ, 307966/2019-4-PQ, 404659/2016-0-Univ, 405101/2016-3-Univ), PRONEX '*Fundação Araucária*' 042/2018, and *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* - Brasil (CAPES) - Finance Code 001 for financial support of this work.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

make the model more accurate.

In the field of electricity energy price forecasting, to obtain efficient models, much of the attention in previous research has been given to hybrid forecasting models. Firstly, [12] combined the decomposition methods variational mode decomposition (VMD) and improved complementary ensemble empirical mode decomposition with adaptive noise. Subsequently, each component was trained and predicted by partially recurrent Elman neural network optimized by a multi-objective grey wolf optimizer. In the same way, [13] used VMD with hyperparameters defined by self-adaptive particle swarm optimization. To forecast the modes in a single horizon, seasonal autoregressive integrated moving average and deep belief network were adopted to forecast regular and irregular modes, respectively. Adjacent to the previous studies, [14] adopted wavelet transform coupled with stacked autoencoder and long short-term memory to forecast residential, commercial and industrial electricity prices. Many other studies such as [15]–[17] discussed the feasibility of hybrid models to forecast electricity energy prices. As limitations of most of the previous studies can be highlighted: the definition of decomposition methods' hyperparameters by trial and error, and the use of a homogeneous ensemble of components.

In respect to the gaps in the above-mentioned studies, this paper focuses on the development of a hybrid multi-stage model. Are coupled the methods: complementary ensemble empirical mode decomposition (CEEMD) [18], a recent proposed metaheuristic coyote optimization algorithm (COA) [19] and machine learning models to develop a multi-stage heterogeneous ensemble model, to perform multi-step-ahead (one, two and three months-ahead) forecasting of commercial and residential electricity prices in Brazil. Firstly, the COA optimizer is applied to define the CEEMD's hyperparameters, and subsequently, CEEMD decompose the series of electricity energy prices (commercial and residential). Thereafter, the components obtained in the previous step (intrinsic mode functions - IMF and one residue) are trained using extreme learning machines (ELM) [20], relevance vector machines (RVM) [21], Gaussian process (GP) [22] and gradient boosting machines (GBM) [23]. The hyperparameters of each model are obtained by grid-search during leave-one-out time slice cross-validation. Finally, the prediction results of different components are directly integrated to generate the final electricity price. Afterward, by the grid of models, the most adequate multi-stage model is the one with the best generalization out-of-sample capacity in terms of mean absolute error (MAE) and mean squared error (RRMSE).

The contribution of this work to the literature is threefold, which are described as follows: (i) Firstly, this paper contributes to field of time series pre-processing by coupling the CEEMD with recent proposed bio-inspired metaheuristic named COA to extract CEEMD's hyperparameters (number of ensembles, noise amplitude and number of components); (ii) Second, with the combination of the different nonlinear models (ELM, RVM, GP, and GBM), to train and predict each component of the decomposed stage, the heterogeneous developed model can learn the data patterns and reflect the high-frequency of electricity price data; and (iii) Finally, this paper contributes to the literature of models used to forecasting electricity prices by investigating the performance of multi-stage models coupled with optimization and heterogeneous ensemble of components over multi-stage homogeneous ensemble models associated with optimization as well as with individual models. Alongside this, the performance of the developed framework was examined based on the Brazilian's electricity commercial and residential price data into three different forecasting horizons. The criteria RRMSE and MAE, as well as the Diebold-Mariano test, were adopted.

The organization of the present paper is as follows: Section II-A presents the datasets adopted in this paper. Section II-B brings a brief description of the adopted methods. Section III details the procedures of the research methodology. Section IV describes the results and discussions. Finally, Section V concludes the paper with final considerations, limitations of the study and proposals of future works.

II. Material & Methods

This Section presents the description of the data (subsection II-A) and methods applied in this paper (subection II-B).

A. Material

The datasets analyzed in this paper refers to Brazil's commercial and residential electricity prices (Brazilian currency - Real - R$) by megawatt-hora (MWh). The datasets consist of 306 monthly observations from April 1994 to September 2019. These data were obtained from the website of the Institute of Applied Economics Research (IPEA) (*Instituto de Pesquisa Econômica Aplicada*, in Portuguese) available in http://www.ipeadata.gov.br/Default.aspx. The data was split into Training and Testing sets in the proportion of 70% and 30%, respectively. In Table I is presented a summary of the statistical indicators of commercial and residential electricity prices.

TABLE I
DESCRIPTIVE MEASURES FOR BRAZILIAN COMMERCIAL AND RESIDENTIAL ELECTRICITY PRICE (R$)

Dataset	Set	# of Samples	Statistical Indicator				
			Minimum	Median	Mean	Maximum	Standard Deviation
Commercial	All set	306	32.89	271.05	258.47	570.80	126.86
	Training set	214	83.00	204.27	199.58	302.14	79.74
	Test set	92	254.58	434.42	399.09	570.80	95.76
Residential	All set	306	28.50	286.56	273.35	803.11	131.31
	Training set	214	68.40	226.34	213.53	333.15	84.26
	Test set	92	275.42	445.31	416.91	803.11	99.79

The electricity energy prices series are illustrated in Figure 1. The high value presented by residential electricity prices in 2013 October can be seen as an outlier and it can be attributed to climate, political factors or measurement errors.

B. Methods

This subsection describes the methods employed in this paper.

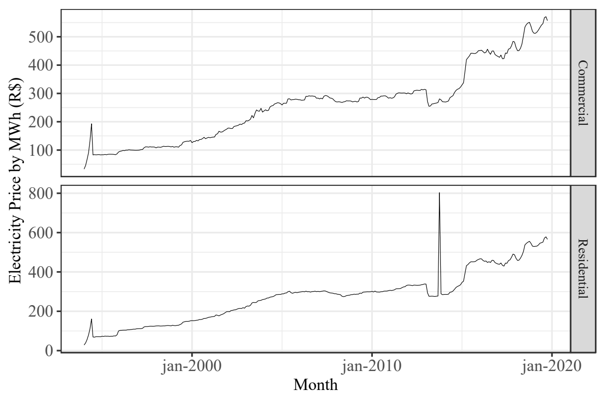


Fig. 1. Brazilian commercial and residential electricity prices (R$)

1) Coyote Optimization Algorithm: The COA optimizer is a recent proposed evolutionary algorithm that considers the social relations inside the packs of the *Canis latrans* species proposed by [19]. In the COA approach, the population is divided into N_p packs with N_c coyotes each. The optimization process starts when the global population of coyotes is defined (all candidate solutions). In the sequence, the coyotes' adaptation in the current social condition (set of decision variables, or in this case, current hyperparameters value) is evaluated using the objective function (in this paper the inverse of orthogonality index - described in subsection II-B2). In it is turning, the alpha coyote of the pack is defined and the social tendency is stated. For each coyote, based on social tendency, the social condition is updated, evaluated and its adaptation verified. Taken into account biological events of life, the birth and death of coyotes are stated. Subsequently, the transition between packs and coyotes' age (in years) is updated. The process ends when the best coyote (solution) is selected.

This stochastic population-based algorithm has recently been applied to define the hyperparameters of three-diode photovoltaic models [24], heavy-duty gas turbine [25], and photovoltaic cells [26]. Also, it has reached better results than the genetic algorithm, differential evolution, and particle swarm optimization approaches in the aforementioned studies. Due to the promising potentials results, a search of the literature reveals that the COA has not yet to be applied for the CEEMD's hyperparameters definition, then it is adopted.

2) Complementary Ensemble Empirical Mode Decomposition: Over the last decades, empirical mode decomposition (EMD) [27] and its improvements such as ensemble empirical mode decomposition (EEMD) and CEEMD [18] were proposed to deal with non-linearity and non-stationarity of time series.

The EMD separates the original signal into IMF and one residual component. The main drawback of this decomposition is named mode mixed problem (MMP). The MMP is characterized by the fact that disparate scales could appear in one IMF. Next, to overcome this disadvantage EEMD was proposed, and in the sequence CEEMD. In spite of the fact of EEMD has effectively resolved the MMP, the residue noise in the signal reconstruction has been raised, and the noise is independent and identically distributed [28]. To improve EEMD, [18] proposed the CEEMD, in which the paired noises are perfectly anti-correlated and have an exact cancellation of the residual noise in the reconstruction of the signal. Due to the effectiveness of CEEMD in several domains of knowledge [29], [30], this paper will employ this decomposition approach to preprocess the electricity energy prices.

The CEEMD has three main hyperparameters named number of trials or number o ensembles, number of components and noise amplitude. Especially, the noise amplitude is designed to be some percentage of the data standard deviation. In most of the cases, these hyperparameters are defined by trial and error [29], [30]. To solve this question, this paper proposes the use of the COA approach to minimize the inverse of the orthogonal index (OI) [11]. The OI is used to measure the orthogonality of the EMD numerically, and a value close of zero is desirable. A smaller OI indicates the best decomposition result [27].

The OI can be computed as follows:

$$\mathrm{OI} = \sum_{t=0}^{T} \left(\sum_{i=1}^{k} \sum_{j=1}^{k} \mathrm{IMF}_i(t)\, \mathrm{IMF}_j(t)/x^2(t) \right), \qquad (1)$$

in which T is the number of time series observations, IMF_i and IMF_j are the i-th and j-th components, k is the number of components, $x(t)$ is the original signal at time $t = 0, \ldots, T$.

3) Extreme Learning Machine: The ELM is a learning algorithm proposed by [20] designed for single-hidden layer feedforward neural networks. In this approach, hidden nodes are chosen randomly and outputs are obtained analytically. Good generalization and learning speed are the main advantages of ELM. The input weights and hidden biases are specified arbitrarily and then are fixed. The output weights are obtained by solving the multiplication of the Moore-Penrose Generalized inverse matrix and output variable matrix [10].

4) Gradient Boosting Machine: The GBM is an ensemble approach that employs a sequential learning process to build an efficient classification or regression model [23]. A regression tree is initially fitted to the data and, on this basis, predictions and the initial residue are computed. A new model is fitted to the previous residuals, a new prediction, to which the initial forecast is added, and then a new residue is obtained. This process is repeated iteratively until a convergence criterion been met. In each iteration, a new model is fitted to the data, aiming to compensate for the weaknesses of the previous model [5].

5) Gaussian Process: A GP is a stochastic process, in which every set of the random variable is multivariate normally distributed. In this respect, a GP is entirely specified by its statistical orders mean and covariance or kernel function. Through kernel function, it is possible to maps the similarity between points of the training set with the purpose of to predict new observations [22].

6) Relevance Vector Machines: The RVM is a probabilistic version of support vector machines [31], [32] that uses Bayesian inference. It has some advantages regarding the SVM framework, such as: avoid the set of free hyperparameters to be defined, the predictions are probabilistic, reduce the computational complexity, and the kernel function does not need to satisfy the Merce's condition. In this framework, probabilistic modeling is conducted and a priori is introduced over the weights. Sparsity is achieved is this approach, once that in the practice the posterior distributions of weights are close of zero [21].

C. Performance Indicators

To check the models' performance, the MAE and RRMSE criteria are used. These measures are described in Table II.

TABLE II
SUMMARY OF PERFORMANCE INDICATORS

Metric	Equation	Definition
MAE	$\sum_{i=1}^{n} \left\| \frac{y_i - \hat{y}_i}{n} \right\|$	The average absolute forecast error of n times forecast results
RRMSE	$\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\frac{1}{n}\sum_{i=1}^{n} y_i}$	The relative root mean squared error
IP	$100 \times \frac{M_p - M_c}{M_c}$	Improvement percentage index

The n represents the number of observations, y_i and $\hat{y}_i$ are the *i*-th observed and predicted values, respectively. Also, the M_c and M_p represent the performance measure of compared and proposed model, respectively.

D. Diebold-Mariano Test

The DM test [33] is applied in this paper to compare the forecasting errors of proposed versus compared models. The hypotheses and DM test statistic are given by (2),

$$H : \begin{cases} H_0 : \epsilon_i^p = \epsilon_i^c \\ H_1 : \epsilon_i^p < \epsilon_i^c, \end{cases} \qquad \mathrm{DM} = \frac{\frac{\sum_{i=1}^{n} [L(\epsilon_i^p) - L(\epsilon_i^c)]}{n}}{\sqrt{\frac{s^2}{n}}} s^2, \tag{2}$$

in which L is a loss function that can estimate the accuracy of each model, ϵ_i^p is the error of the proposed model, ϵ_i^c is the error of the compared model, and s^2 is an estimate for the variance of $d_i = L(\epsilon_i^p) - L(\epsilon_i^c)$. The null hypothesis is rejected if DM $<$ -z_α, being z_α the percentile of normal distribution and α the significance level.

III. THE PROPOSED HYBRID MULTI-STAGE FORECASTING SYSTEM

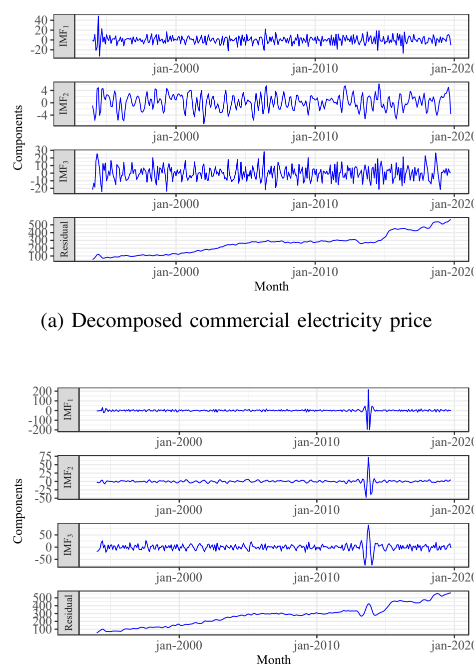
This section presents the steps adopted to develop the proposed hybrid multi-stage forecasting model.

Step 1: Firstly, the COA is coupled with CEEMD to define the CEEMD's hyperparameters. For COA optimizer, the number of coyotes and packs are defined as 5 and 10, respectively. These values are selected by the trial and error, once that there is no guideline for the definition of COA parameters [19]. Moreover, if an increase in the number of packs, and/or coyotes is considered, the optimization time will also increase due to the greater number of evaluations to be carried out. However, for this problem, it was observed that the accuracy does not improve significantly. In this way, the initial values adopted for these parameters are fixed, for both problems. Table III shows the CEEMD's hyperparameters defined by COA. Finally, the original electricity price is decomposed.

TABLE III
SEARCH BOUNDARIES OF THE CEEMD PARAMETERS.

Hyperparameter	Boundaries		Select Hyperparameters	
	Lower bound	Upper Bound	Commercial	Residential
Number of Ensembles	50	100	50	52
Number of Components	2	5	4	4
Noise amplitude	0.2	0.5	0.4918	0.4923

The decomposed commercial and residential electricity prices are illustrated in Figures 2a and 2b, respectively.



(a) Decomposed commercial electricity price

(b) Decomposed residential electricity price

Fig. 2. Decomposed series

The behavior observed in the residual component of both series is due to the growing trend in prices.

Step 2: Second, each component obtained in step III (three IMFs and one residue) is trained using ELM, GBM, GP, and RVM. In the training stage, leave-one-out cross-validation with time slice is adopted. The inputs are defined by autocorrelation and partial autocorrelation analysis. The data are centered by

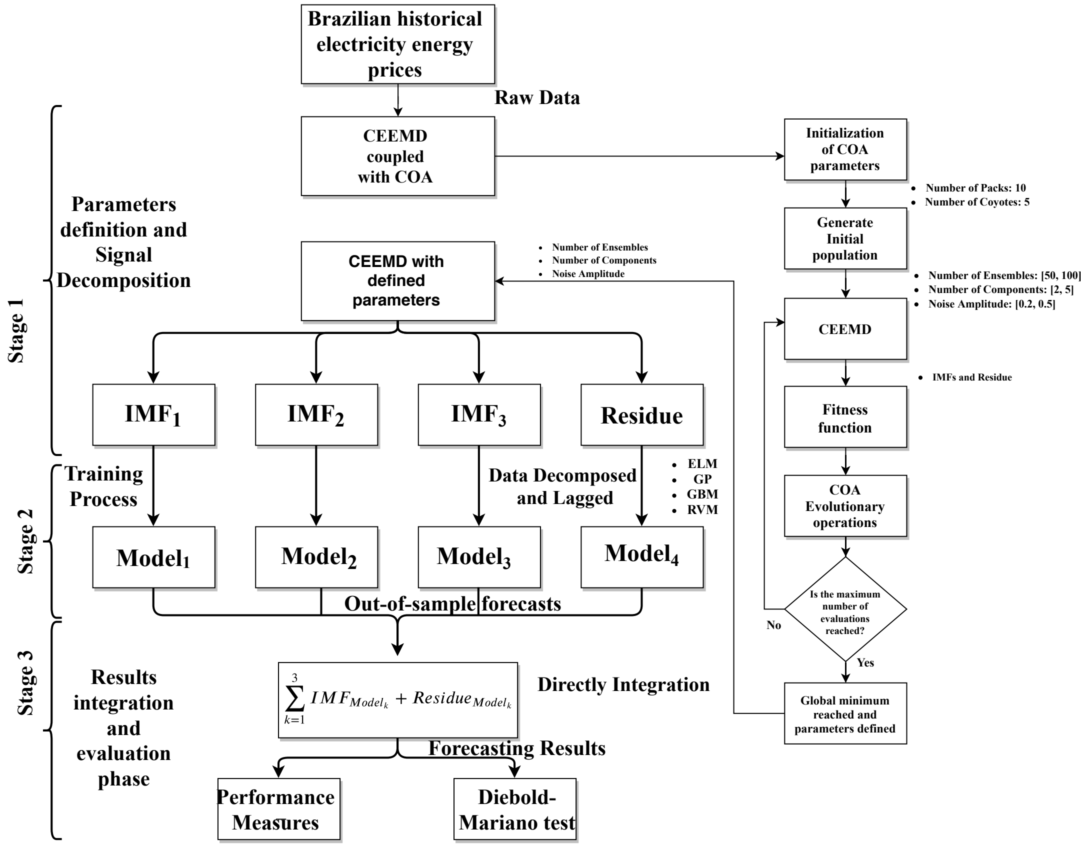


Fig. 3. Framework of proposed approach

its mean value and divided by its standard deviation. The training structure is stated as follows:

$$y_{(t+1,k)} = f\left\{y_{(t,k)}, \dots, y_{(t-n_y,k)}\right\} + \epsilon \quad \epsilon \sim N(\mathbf{0}, \sigma^2), \quad (3)$$

and forecast electricity energy prices one-month-ahead (4), two-months-ahead (5), and three-months-ahead (6) according to:

$$\hat{y}_{(t+h,k)} = f\left\{y_{(t+h-1,k)}, y_{(t+h-2,k)}, y_{(t+h-3,k)}\right\} \quad (4)$$

$$\hat{y}_{(t+h,k)} = f\left\{\hat{y}_{(t+h-1,k)}, y_{(t+h-2,k)}, y_{(t+h-3,k)}\right\} \quad (5)$$

$$\hat{y}_{(t+h,k)} = f\left\{\hat{y}_{(t+h-1,k)}, \hat{y}_{(t+h-2,k)}, y_{(t+h-3,k)}\right\} \quad (6)$$

in which f is a function related to the adopted model in the training stage, $\hat{y}_{(t+h,k)}$ is the forecast value for k-th component obtained in the decomposition stage ($k = 1,\dots,4$) on time t and forecast horizon h ($h = 1, 2, 3$), $y_{(t+h-n_y,k)}$ are the previously prices lagged in $n_y = 1, \dots, 3$ and ϵ is the random error.

Table IV presents the models' hyperparameters obtained by grid-search.

TABLE IV
HYPERPARAMETERS OBTAINED BY GRID-SEARCH FOR EACH MODEL

Dataset	Component	Model					
		ELM			RVM	GBM	
		Active Function	Neurons	Weights Initialization	Sigma	Boosting Iteractions	Max Tree Depth
Commercial	IMF_1	Hardlin	17	Normal Gaussian	0.6250	50	3
	IMF_2	Hardlin	13	Uniform Negative	0.6845	100	1
	IMF_3	Hardlin	13	Uniform Negative	0.5682	50	1
	Residue	Sigmoide	17	Normal Gaussian	10.3476	250	5
	Non-decomposed	Sigmoide	9	Normal Gaussian	19.6809	150	3
Residential	IMF_1	Sine	13	Uniform Positive	0.6071	150	1
	IMF_2	Rectifier Linear Unit	12	Uniform Negative	0.6641	250	1
	IMF_3	Rectifier Linear Unit	13	Uniform Negative	0.4929	100	2
	Residue	Sigmoide	13	Normal Gaussian	10.3089	250	5
	Non-decomposed	Sigmoide	14	Normal Gaussian	22.9609	150	3

Also, for the GP approach, there are no hyperparameters for tuning and the linear kernel function is used. Moreover, for GBM the shrinkage and minimum number of terminal node size are held constant equal 0.1 and 10, respectively. Finally, for RVM the radial basis kernel function is adopted. The kernels of GP and RVM were defined by grid-search.

Step 3: Finally, the forecasts of different models used for each component are directly integrated (simple sum) to generate final electricity price values. Afterward, by the grid of models, the most adequate multi-stage model is the one with the best generalization out-of-sample capacity in terms of MAE and RRMSE. Table V describes the models used for each component.

TABLE V
MODELS ADOPTED BY EACH COMPONENT IN EACH DATASET

Dataset	Component	Forecasting Horizon		
		One-month-ahead	Two-months-ahead	Three-months-ahead
Commercial	IMF_1	ELM	ELM	ELM
	IMF_2	GP	ELM	RVM
	IMF_3	GBM	GP	GBM
	Residue	GP	GP	GP
Residential	IMF_1	GP	GP	RVM
	IMF_2	GP	ELM	GBM
	IMF_3	ELM	ELM	RVM
	Residue	GP	GP	GP

Step 4: Obtaining forecasts out-of-sample (test set), performance indicators defined in II-C are computed and two kinds of comparisons are conducted. The first is the comparison of multi-stage heterogeneous ensemble model and multi-stage homogeneous ensemble model. Second, a comparison of multi-stage heterogeneous ensemble model and models without consider decomposition is developed. Finally, the DM test described in subsection II-D is employed.

Figure 3 summarize the main steps used in the data analysis. The results presented in Section IV are generated using the `R` software [34].

IV. Results

This section describes the results of the developed experiments in three ways in forecasts out-of-sample (test set). First, subsection IV-A is designed to compare the results of the proposed multi-stage heterogeneous ensemble model and a multi-stage homogeneous ensemble model. In the sequence, subsection IV-B is used to compare the performance of developed approaches and models without previous decomposition. To finish, subsection IV-C presents DM test to statistically evaluate the errors of the proposed approach versus other models. Additionally, Figures 4 and 5 illustrate the relation between observed and predicted values, and the magnitude of the sum of standardized squared errors, respectively. In Table VI and VII, the best results are presented in bold

A. Comparison of multi-stage heterogeneous ensemble and multi-stage homogeneous ensemble models

Table VI illustrates the performance or developed model and multi-stage homogeneous ensembles named COA-CEEMD-GP, COA-CEEMD-ELM, COA-CEEMD-RVM, and COA-CEEMD-GBM.

TABLE VI
Performance measures of proposed and compared models used to forecasting Brazilian electricity price multi-step-ahead

Dataset	Model	Forecasting Horizon					
		One-month-ahead		Two-months-ahead		Three-months-ahead	
		MAE	RRMSE	MAE	RRMSE	MAE	RRMSE
Commercial	Proposed	**9.1191**	**0.0272**	**12.5722**	**0.0415**	**14.0000**	**0.0501**
	COA-CEEMD-GP	9.8200	0.0295	12.6248	0.0417	14.2700	0.0510
	COA-CEEMD-ELM	38.0209	0.1289	68.6489	0.2489	83.8030	0.2926
	COA-CEEMD-GBM	108.6709	0.3490	109.1811	0.3500	109.6404	0.3500
	COA-CEEMD-RVM	303.1389	0.9202	314.3207	0.9431	322.8980	0.9522
Residential	Proposed	**8.9811**	**0.0305**	**11.0507**	**0.0381**	**13.2598**	**0.0460**
	COA-CEEMD-GP	9.0935	0.0310	11.1217	0.0390	13.2872	0.0473
	COA-CEEMD-ELM	35.0976	0.1146	52.1548	0.1786	60.5261	0.2000
	COA-CEEMD-GBM	103.2843	0.3175	103.5004	0.3174	104.3387	0.3176
	COA-CEEMD-RVM	310.2537	0.9146	327.5563	0.9344	334.7754	0.9444

Through Table VI it should be noted that the use of different models to compose an ensemble of components improves the final accuracy of the model. This is verified for the case of commercial as well as residential electricity prices. The observed superiority of the proposed approach regarding these compared models should be attributed to the fact that diversity of heterogeneous ensembles is higher than homogeneous ensembles, which plays an important role in ensemble learning [35].

By investigating the improvement on the errors of the proposed model regarding compared COA-CEEMD homogeneous ensemble models, it is possible to infer that the improvement in MAE criterion ranged between 7.14% - 96.99%, 0.42% - 96%, and 1.89% - 95.66%, for commercial electricity price on one, two and three-months-ahead forecasting, respectively. In the residential context, the same pattern is observed, since the reduction on RRMSE is ranged between 1.61% - 96.67%, 2.31% - 95.92%, and 2.75% - 95.13% in horizons of one up to three-months-ahead, respectively. In both cases, similar behavior can be observed between adopted criteria.

B. Comparison of multi-stage heterogeneous ensemble and non-decomposed models

Table VII shows the performance of the developed model and models without considering CEEMD as preprocess, named GP, ELM, RVM, and GBM.

TABLE VII
Performance measures of proposed and compared models used to forecasting Brazilian electricity price multi-step-ahead

Dataset	Model	Forecasting Horizon					
		One-month-ahead		Two-months-ahead		Three-months-ahead	
		MAE	RRMSE	MAE	RRMSE	MAE	RRMSE
Commercial	Proposed	**9.1191**	**0.0272**	**12.5722**	**0.0415**	**14.0000**	**0.0501**
	GP	10.1264	0.0366	13.0655	0.0487	15.4955	0.0577
	ELM	61.8013	0.2107	51.7893	0.1774	66.5200	0.2331
	GBM	109.7469	0.3525	110.3789	0.3529	111.0765	0.3531
	RVM	320.8700	0.9463	331.5674	0.9605	334.7101	0.9640
Residential	Proposed	**8.9811**	**0.0305**	**11.0507**	**0.0381**	**13.2598**	**0.0460**
	GP	9.4405	0.0342	12.1558	0.0451	14.4413	0.0552
	ELM	52.9344	0.1721	95.2086	0.3413	91.7472	0.3177
	GBM	105.467	0.3231	106.5075	0.3258	106.9078	0.3259
	RVM	325.8226	0.9275	348.0082	0.9558	360.2490	0.9681

In respect of the results presented in Table VII, the outcomes of this paper reinforce findings presented by [36] and [37], which point out the benefits of using decomposition techniques as a way of pre-processing time series. In particular, the use of the COA-CEEMD approach is important for the development of an effective model for forecasting electricity prices. Second, to [18], the use of decomposition as pre-processing is useful in the time series field because through the use of this technique it is possible to deal with non-stationarity and non-linearity of the data. Also, the results described in this section corroborate the findings of [38], since the ensemble models achieve better accuracy than its members.

Concerning the enhancement of hybrid model regarding non-decomposed models, seen the MAE in one-month-ahead forecasting, the reduction is ranged between 9.95% - 97.16% and 4.87% - 97.24% in commercial and residential electricity prices, respectively. Similarly behavior is observed for the other two time windows. In respect of the reduction of the RRMSE criterion, for the commercial dataset, the improvement is ranged 25.68% - 97.12%, 14.78% - 95.67%, and 13.17% - 94.80% in one, two, and three-months-ahead, respectively. In it is turn, for the residential dataset, the RRMSE improvement is ranged between 10.82% - 96.71%, 15.52% - 96.01%, and 16.67% - 95.25% in one, two and three-months-ahead forecasting, respectively.

C. Statistical tests to compare proposed multi-stage heterogeneous ensemble model and other models

To demonstrate the statistical comparisons between errors of the proposed and compared models described in subsections IV-A and IV-B, in Table VIII can be seen the statistic of DM test, as well as when the comparisons are statistically significant.

TABLE VIII
STATISTIC OF DM TEST FOR STATISTICAL COMPARISON OF PROPOSED APPROACH VERSUS OTHER MODELS

Model	Forecasting Horizon					
	One-month-ahead		Two-monhts-ahead		Three-months-ahead	
	Commercial	Residential	Commercial	Residential	Commercial	Residential
COA-CEEMD-ELM	-5.93***	-5.95***	-4.83***	-4.14***	-3.71***	-3.31***
COA-CEEMD-RVM	-11.78***	-11.73***	-7.27***	-7.49***	-5.87***	-6.06***
COA-CEEMD-GP	-1.99*	-1.63*	-0.19	-0.86	-1.31*	-1.34*
COA-CEEMD-GBM	-8.79***	-8.87***	-5.17***	-5.11***	-4.04***	-3.96***
ELM	-7.05***	-7.01***	-4.90***	-4.87***	-4.49***	-3.84***
RVM	-12.78***	-12.41***	-7.85***	-8.35***	-6.10***	-6.95***
GP	-1.91*	-1.12*	-1.45*	-1.41*	-1.99**	-1.95**
GBM	-8.98***	-8.91***	-5.22***	-5.15***	-4.06***	-4.00***

Note: ***1% significance level; **5% significance level; * 10% significance level.

Through the DM test, it can be stated that in 95.83% of the cases, the proposed approach reached statistically lower errors than the other models. The RVM based approaches reached high differences between the errors, and GP based approaches reach similar errors regarding the proposed framework. These findings are valid for all datasets and forecast horizons.

Figures 4a and 4b expose that the multi-stage heterogeneous ensemble model learns the data behavior, being able to obtain forecast prices similar to observed values. For commercial and residential datasets, the good performance (regarding RRMSE and MAE) in the training set is maintained in the test set. Exceptionally, for the outlier of the test set for the residential case, the proposed model was not able to capture the high variability

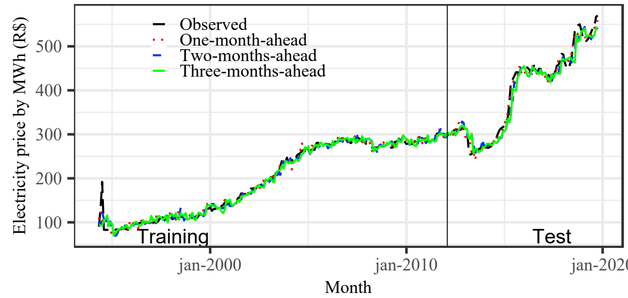
Finally, considering the illustrated by the radars plots (Figures 5b and 5b), the results exposed in previous sections are reinforced. Best models reached the sum of standardized squared errors close to the center of the plot.

V. CONCLUSION

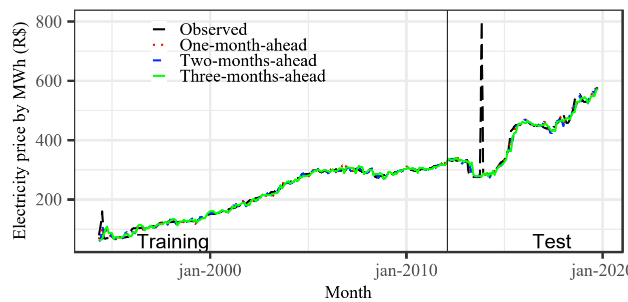
In this paper, a hybrid multi-stage heterogeneous ensemble model was proposed to forecast multi-step-ahead (one, two, and three-months-ahead) Brazilian commercial and residential electric energy prices. In the first stage, the COA optimizer was adopted to define the hyperparameters of pre-processing CEEMD. In the sequence, the four obtained components (three IMF and one residual) by COA-CEEMD were trained and predict by different models (ELM, GBM, GP, and RVM). To choose the most suitable model for each component, the grid-search approach was conducted. The final forecasts were obtained through a heterogeneous ensemble of components directly integrated.

Our findings suggest that: (i) The use of COA-CEEMD improves the final results regarding not applying decomposition; (ii) The use of different models for components allow to improve the final accuracy concerning the use of a homogeneous ensemble model of components; and (iii) The proposed approach reaches better accuracy than the compared models, and the good performance is constant when the forecast horizon is expanded;

Even with good results achieved, this study has the follow-

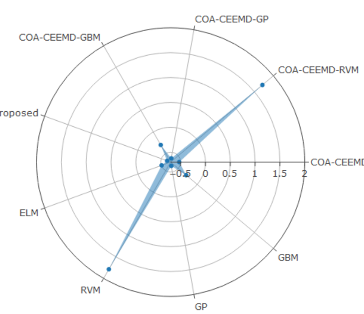


(a) Commercial electricity price

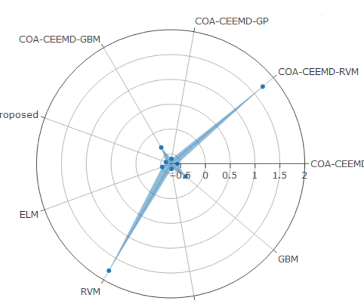


(b) Residential electricity price

Fig. 4. Electricity energy price observed versus forecast multi-step-ahead



(a) Commercial electricity price



(b) Residential electricity price

Fig. 5. Radar plot for standardized sum of squared errors

ing limitations: (i) The political, climatic and demand factors were not taken into account in the data analysis; and (ii) The proposed model was not able to capture the variability of an extreme observation for the set of data referring to the price of residential electricity (iii) The parameters of COA optimizer were selected by trial and error. For future works, it is desirable: (i) the adoption of robust techniques to deal with outliers; (ii) hybridization of decomposition techniques; (iii) reconstruct the decomposed signal through the weighted integration considering the no negative constraint theory; (iv) selection of the models for components through optimization techniques; and (v) developing an adaptive version of COA to define the number of coyotes and packs.

References

[1] 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," *Applied Energy*, vol. 190, pp. 291–305, 2017.

[2] K. Maciejowska, "Assessing the impact of renewable energy sources on the electricity price level and variability – a quantile regression approach," *Energy Economics*, vol. 85, no. 104532, 2020.

[3] W. Puchalsky, G. T. Ribeiro, C. P. da Veiga, R. Z. Freire, and L. S. Coelho, "Agribusiness time series forecasting using wavelet neural networks and metaheuristic optimization: An analysis of the soybean sack price and perishable products demand," *International Journal of Production Economics*, vol. 203, pp. 174–189, 2018.

[4] M. H. D. M. Ribeiro, V. H. A. Ribeiro, G. Reynoso-Meza, and L. S. Coelho, "Multi-objective ensemble model for short-term price forecasting in corn price time series," in *International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, July 2019, pp. 1–8.

[5] M. H. D. M. Ribeiro and L. S. Coelho, "Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series," *Applied Soft Computing*, vol. 86, no. 105837, 2020.

[6] M. H. D. M. Ribeiro, V. C. Mariani, L. Manke, and L. S. Coelho, "Nonlinear system identification based on ensemble learning: The cascaded tanks system case," in *14th Brazilian Symposium on Intelligent Automation (SBAI)*, Ouro Preto, Brazil, October 2019, pp. 415–420.

[7] J. a. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, "Ensemble approaches for regression: A survey," *ACM Computing Surveys*, vol. 45, no. 1, pp. 1–40, 2012.

[8] W. Zhang, Z. Qu, K. Zhang, W. Mao, Y. Ma, and X. Fan, "A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting," *Energy Conversion and Management*, vol. 136, pp. 439–451, 2017.

[9] P. Du, J. Wang, Z. Guo, and W. Yang, "Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting," *Energy Conversion and Management*, vol. 150, pp. 90–107, 2017.

[10] V. C. Mariani, S. H. Och, L. S. Coelho, and E. Domingues, "Pressure prediction of a spark ignition single cylinder engine using optimized extreme learning machine models," *Applied Energy*, vol. 249, pp. 204–221, 2019.

[11] D. Zhang, C. Cai, S. Chen, and L. Ling, "An improved genetic algorithm for optimizing ensemble empirical mode decomposition method," *Systems Science & Control Engineering*, vol. 7, no. 2, pp. 53–63, 2019.

[12] W. Yang, J. Wang, T. Niu, and P. Du, "A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting," *Applied Energy*, vol. 235, pp. 1205–1225, 2019.

[13] J. Zhang, Z. Tan, and Y. Wei, "An adaptive hybrid model for short term electricity price forecasting," *Applied Energy*, vol. 258, no. 114087, 2020.

[14] W. Qiao and Z. Yang, "Forecast the electricity price of U.S. using a wavelet transform-based hybrid model," *Energy*, vol. 193, no. 116704, 2020.

[15] X. Zhang, J. Wang, and Y. Gao, "A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and svm," *Energy Economics*, vol. 81, pp. 899–913, 2019.

[16] W. Yang, J. Wang, T. Niu, and P. Du, "A novel system for multi-step electricity price forecasting for electricity market management," *Applied Soft Computing*, vol. 88, no. 106029, 2020.

[17] J. Wang, W. Yang, P. Du, and T. Niu, "Outlier-robust hybrid electricity price forecasting model for electricity market management," *Journal of Cleaner Production*, vol. 249, no. 119318, 2020.

[18] J.-R. Yeh, J.-S. Shieh, and N. E. Huang, "Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method," *Advances in Adaptive Data Analysis*, vol. 02, no. 02, pp. 135–156, 2010.

[19] J. Pierezan and L. S. Coelho, "Coyote optimization algorithm: A new metaheuristic for global optimization problems," in *IEEE Congress on Evolutionary Computation (CEC)*, Rio de Janeiro, Brazil, 2018, pp. 1–8.

[20] 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, neural Networks.

[21] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of machine learning research*, vol. 1, no. Jun, pp. 211–244, 2001.

[22] C. E. Rasmussen, *Gaussian Processes in Machine Learning*. Heidelberg, Germany: Springer, 2004, pp. 63–71.

[23] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[24] M. H. Qais, H. M. Hasanien, S. Alghuwainem, and A. S. Nouh, "Coyote optimization algorithm for parameters extraction of three-diode photovoltaic models of photovoltaic modules," *Energy*, vol. 187, no. 116001, 2019.

[25] J. Pierezan, G. Maidl, E. M. Yamao, L. S. Coelho, and V. C. Mariani, "Cultural coyote optimization algorithm applied to a heavy duty gas turbine operation," *Energy Conversion and Management*, vol. 199, no. 111932, 2019.

[26] V. J. Chin and Z. Salam, "Coyote optimization algorithm for the parameter extraction of photovoltaic cells," *Solar Energy*, vol. 194, pp. 656–670, 2019.

[27] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[28] T. Li, Y. Zhou, X. Li, J. Wu, and T. He, "Forecasting daily crude oil prices using improved CEEMDAN and Ridge regression-based predictors," *Energies*, vol. 12, 2019.

[29] J. Wu, Y. Chen, T. Zhou, and T. Li, "An adaptive hybrid learning paradigm integrating CEEMD, ARIMA and SBL for crude oil price forecasting," *Energies*, vol. 12, no. 7, 2019.

[30] S. Zhu, X. Wang, N. Shi, and M. Lu, "CEEMD-subset-OASVR-GRNN for ozone forecasting: Xiamen and Harbin as cases," *Atmospheric Pollution Research*, vol. 11, no. 4, pp. 744–754, 2020.

[31] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144–152.

[32] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[33] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995.

[34] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: http://www.R-project.org/

[35] H. G. Zefrehi and H. Altınçay, "Imbalance learning using heterogeneous ensembles," *Expert Systems with Applications*, vol. 142, no. 113005, 2020.

[36] C. Wu, J. Wang, X. Chen, P. Du, and W. Yang, "A novel hybrid system based on multi-objective optimization for wind speed forecasting," *Renewable Energy*, vol. 146, pp. 149–165, 2020.

[37] Y. Hao, C. Tian, and C. Wu, "Modelling of carbon price in two real carbon trading markets," *Journal of Cleaner Production*, vol. 244, no. 118556, 2020.

[38] R. K. Agrawal, F. Muchahary, and M. M. Tripathi, "Ensemble of relevance vector machines and boosted trees for electricity price forecasting," *Applied Energy*, vol. 250, pp. 540–548, 2019.