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 Cantoni‡, 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
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 (RMSE).

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 (subsection 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-hour (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>
<thead>
<tr>
<th>Dataset</th>
<th>Set</th>
<th># of Samples</th>
<th>Minimum</th>
<th>Median</th>
<th>Mean</th>
<th>Maximum</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial</td>
<td>All set</td>
<td>306</td>
<td>32.89</td>
<td>271.90</td>
<td>258.47</td>
<td>570.80</td>
<td>126.86</td>
</tr>
<tr>
<td></td>
<td>Training set</td>
<td>214</td>
<td>83.00</td>
<td>204.27</td>
<td>199.58</td>
<td>302.14</td>
<td>79.74</td>
</tr>
<tr>
<td></td>
<td>Test set</td>
<td>92</td>
<td>25.58</td>
<td>434.42</td>
<td>416.91</td>
<td>99.79</td>
<td></td>
</tr>
<tr>
<td>Residential</td>
<td>All set</td>
<td>306</td>
<td>28.90</td>
<td>290.58</td>
<td>213.53</td>
<td>803.11</td>
<td>183.11</td>
</tr>
<tr>
<td></td>
<td>Training set</td>
<td>214</td>
<td>60.40</td>
<td>226.54</td>
<td>213.53</td>
<td>333.15</td>
<td>84.26</td>
</tr>
<tr>
<td></td>
<td>Test set</td>
<td>92</td>
<td>275.42</td>
<td>445.31</td>
<td>416.91</td>
<td>99.79</td>
<td></td>
</tr>
</tbody>
</table>

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.
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:

$$OI = \frac{1}{T} \sum_{t=0}^{T} \left( \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{\text{IMF}_i(t) \cdot \text{IMF}_j(t)}{x(t)} \right),$$

in which $T$ is the number of time series observations, $\text{IMF}_i$ and $\text{IMF}_j$ 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 Mercé’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**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Equation</th>
<th>Definition</th>
</tr>
</thead>
</table>
| MAE        | \[
\sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| \] | The average absolute forecast error of \( n \) times forecast results |
| RRMSE      | \[\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2\] | The relative root mean squared error |
| IP         | \[100 \times \frac{M_p - M_c}{M_p}\] | 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),

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

in which \( L \) is a loss function that can estimate the accuracy of each model, \( \epsilon_i^p \) is the error of the proposed model, \( \epsilon_i^c \) 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 \( DM < -z_\alpha \), being \( z_\alpha \) the percentile of normal distribution and \( \alpha \) 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**

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>Boundaries</th>
<th>Select Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Ensembles</td>
<td>90</td>
<td>TR1</td>
</tr>
<tr>
<td>Number of Components</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Noise amplitude</td>
<td>0.2</td>
<td>0.5</td>
</tr>
</tbody>
</table>

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

\[ y(t+1,k) = f \{ y(t,k), \ldots , y(t-n_y,k) \} + \epsilon \sim N(0,\sigma^2) \quad (3) \]

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

\[
\begin{align*}
\hat{y}(t+h,k) &= f \{ \hat{y}(t+(h-1),k), \hat{y}(t+(h-2),k), y(t+(h-3),k) \} \quad (4) \\
\hat{y}(t+h,k) &= f \{ y(t+(h-1),k), \hat{y}(t+(h-2),k), y(t+(h-3),k) \} \quad (5) \\
\hat{y}(t+h,k) &= f \{ \hat{y}(t+(h-1),k), \hat{y}(t+(h-2),k), y(t+(h-3),k) \} \quad (6)
\end{align*}
\]

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 \( (k = 1, \ldots , 4) \) on time \( t \) and forecast horizon \( h (h = 1, 2, 3) \), \( y(t+(h-n_y),k) \) are the previously prices lagged in \( n_y = 1, \ldots , 3 \) and \( \epsilon \) is the random error.

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

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.
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 models with and without 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 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 of the developed model and multi-stage homogeneous ensembles named COA-CEEMD-GP, COA-CEEMD-ELM, COA-CEEMD-RVM, and COA-CEEMD-GBM.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>Forecasting Horizon</th>
<th>MAE</th>
<th>RRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed</td>
<td>One-month-ahead</td>
<td>9.1191</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Two-months-ahead</td>
<td>12.3772</td>
<td>0.0415</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Three-months-ahead</td>
<td>14.0000</td>
<td>0.0510</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>Forecasting Horizon</th>
<th>MAE</th>
<th>RRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed</td>
<td>One-month-ahead</td>
<td>9.1191</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Two-months-ahead</td>
<td>12.3772</td>
<td>0.0415</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Three-months-ahead</td>
<td>14.0000</td>
<td>0.0510</td>
</tr>
</tbody>
</table>

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

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 forecasting, 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>
<thead>
<tr>
<th>Model</th>
<th>One-month-ahead</th>
<th>Two-months-ahead</th>
<th>Three-months-ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Commercial</td>
<td>Residential</td>
<td>Commercial</td>
</tr>
<tr>
<td>COA-CEEMD-ELM</td>
<td>-5.93***</td>
<td>-5.95***</td>
<td>-4.83***</td>
</tr>
<tr>
<td>COA-CEEMD-RVM</td>
<td>-11.78***</td>
<td>-11.73***</td>
<td>-7.25***</td>
</tr>
<tr>
<td>COA-CEEMD-GP</td>
<td>-1.99***</td>
<td>-1.63***</td>
<td>-0.19</td>
</tr>
<tr>
<td>COA-CEEMD-GBM</td>
<td>-4.79***</td>
<td>-8.87***</td>
<td>-5.17***</td>
</tr>
<tr>
<td>ELM</td>
<td>-7.05***</td>
<td>-7.01***</td>
<td>-4.96***</td>
</tr>
<tr>
<td>GP</td>
<td>-1.91*</td>
<td>-1.12*</td>
<td>-1.45*</td>
</tr>
<tr>
<td>GBM</td>
<td>-8.98***</td>
<td>-8.99***</td>
<td>-8.13***</td>
</tr>
</tbody>
</table>

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 5a 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-
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


