energies

Article

Forecasting Electricity Demand by Neural Networks and Definition of Inputs by Multi-Criteria Analysis

Carolina Deina [1], João Lucas Ferreira dos Santos [2], Lucas Henrique Biuk [3], Mauro Lizot [4], Attilio Converti [5,*], Hugo Valadares Siqueira [2,3] and Flavio Trojan [2]

1 Graduate Program in Industrial Engineering (PPGEP), Federal University of Rio Grande do Sul (UFRGS), Av. Paulo Gama, 110, Porto Alegre 90040-060, Brazil
2 Graduate Program in Industrial Engineering (PPGEP), Federal University of Technology-Parana (UTFPR), Rua Doutor Washington Subtil Chueire, 330, Ponta Grossa 84017-220, Brazil
3 Graduate Program in Electrical Engineering (PPGEE), Federal University of Technology-Parana (UTFPR), Rua Doutor Washington Subtil Chueire, 330, Ponta Grossa 84017-220, Brazil
4 Department of General and Applied Administration (DAGA), Federal University of Parana (UFPR), Avenue Prefeito Lothário Meissner, 632, Jardim Botânico 80210-170, Brazil
5 Department of Civil, Chemical and Environmental Engineering, University of Genoa, Pole of Chemical Engineering, Via Opera Pia 15, 15145 Genoa, Italy
* Correspondence: converti@unige.it

Abstract: The planning of efficient policies based on forecasting electricity demand is essential to guarantee the continuity of energy supply for consumers. Some techniques for forecasting electricity demand have used specific procedures to define input variables, which can be particular to each case study. However, the definition of independent and casual variables is still an issue to be explored. There is a lack of models that could help the selection of independent variables, based on correlate criteria and level of importance integrated with artificial networks, which could directly impact the forecasting quality. This work presents a model that integrates a multi-criteria approach which provides the selection of relevant independent variables and artificial neural networks to forecast the electricity demand in countries. It provides to consider the particularities of each application. To demonstrate the applicability of the model a time series of electricity consumption from a southern region of Brazil was used. The dependent inputs used by the neural networks were selected using a traditional method called Wrapper. As a result of this application, with the multi-criteria ELECTRE I method was possible to recognize temperature and average evaporation as explanatory variables. When the variables selected by the multi-criteria approach were included in the predictive models, were observed more consistent results together with artificial neural networks, better than the traditional linear models. The Radial Basis Function Networks and Extreme Learning Machines stood out as potential techniques to be used integrated with a multi-criteria method to better perform the forecasting.

Keywords: electricity demand; multi-criteria forecasting model; dependent variable; artificial neural networks; forecasting models

check for updates

Citation: Deina, C.; dos Santos, J.L.F.; Biuk, L.H.; Lizot, M.; Converti, A.; Siqueira, H.V.; Trojan, F. Forecasting Electricity Demand by Neural Networks and Definition of Inputs by Multi-Criteria Analysis. *Energies* **2023**, *16*, 1712. https://doi.org/10.3390/en16041712

Academic Editors: Andrea Lanzini, Francesco Demetrio Minuto and Sergio Olivero

Received: 17 December 2022
Revised: 27 January 2023
Accepted: 4 February 2023
Published: 8 February 2023

1. Introduction

The electricity demand can be predicted by various mathematical and statistical models; however, choosing a model that provides coherent results is a critical activity in this process [1]. Several external factors can influence electricity consumption, and it needs to be considered as inputs in these models [1]. The use of independent factors can help to reduce errors, and, on the other hand, the forecasting process becomes more complex because it requires a more detailed study of relationships between the variables, which involves recognition of criteria and level of importance [2] since these criteria are different in each case [1,2].

To understand these relationships, different meteorological, socioeconomic, and environmental variables were included in a considerable number of studies related to forecasting electricity demand. However, the literature shows that the selection of the independent variables is mainly performed based on researchers' knowledge, and it does not guarantee that it is the best procedure to solve forecasting problems [2–7]. Thus, the development of an integrated model for selecting independent variables and the importance of them together with artificial neural networks on electricity demand is still a relevant gap.

Countries are faced with challenges in the energy sector, and to remain the system in operation it is necessary to improve decisions already taken and incorporate future targets, such as expanding the energy matrix, guaranteeing energy security, and meeting sustainable development objectives [8].

In this context, the study of forecasting models that integrate a multi-criteria analysis creates the opportunity to incorporate other criteria not considered by the existing models. This step involves economic, social, and environmental dimensions, as well as the efficiency of energy production processes.

Electricity consumption is growing on a global scale. According to the Central Intelligence Agency [9], developed countries consume 43.26% of the total global electricity consumption. Much of this consumption occurs in the residential, commercial, and public sectors, and the industrial sector is the biggest consumer [10].

Electricity cannot be stored, and imbalances between supply and demand must be efficiently modeled to avoid costs that are normally transferred to users [11]. Both excess and reduction in electricity production are harmful to a country's economy [12]. In this case, decisions need to be economically viable, and it is essential to recognize the behavior and impact of all involved variables, which makes it possible to identify present and future trends to allocate investments in the electricity sector correctly.

The present work presents a model to improve the forecasting of electricity demand with a selection of external factors that influence the demand by a multi-criteria approach, forecasts of electricity demand by linear models and Artificial Neural Networks (ANNs), data collected from a real case, and comparisons between the performance of ANNs and linear models.

2. Theoretical Foundations

Table 1 summarizes important studies that used some development to include independent (causal) variables and the main problems related to the theme.

A lot of studies have been using approaches with independent variables included in some moments of the forecast process. However, these inclusions are accomplished following the decision-makers' knowledge, without a structured process to guarantee an efficient selection of independent variables.

In some cases, the independent variables are used without relevance, as highlighted by Sadownik and Barbosa [3], in which the GDP was selected by the author but was not relevant. On the other hand, Darbellay and Slama [4] related that in the multivariate forecast, temperature added as an external variable allowed more information to be integrated into the ANNs and showed improvement over linear models.

Pao [2] found that population and national income are significant variables that directly influence the forecasts. In the study of Adam et al. [13], the input variables were selected by a non-homogeneous Gompertz diffusion process (NHGDP) based on a Genetic Algorithm (GA) approach.

Different meteorological variables were used by Hernández et al. [14] to forecast energy demand. The authors performed an autocorrelation and observed that temperature, relative humidity, and solar radiation presented a correlation with electricity demand. Cui et al. [5], studied the relationship between electricity load and daily temperature using an improved ARIMAX forecast model, and the relative error was smaller than using AR, ARMA.

Table 1. Articles related to the research topic.

Authors (Year)/[Ref]	Country Case Study	Methodology Used	Independent Variables Used
Sadownik and Barbosa (1999) [3]	Brazil	Nonlinear dynamic model and econometric model	Gross Domestic Product (GDP)
Darbellay and Slama (2000) [4]	Czechoslovakia	ANNs and the autoregressive integrated moving average model (ARIMA)	Temperature
Pao (2006) [2]	Taiwan	ANNS, multiple log-linear regression (LNREG), response surface regression (RSREG), and Autoregressive Moving Average	National income (NI), population, GDP, and consumer price index (CPI)
Adam et al. (2011) [13]	Mauritius Islands	ANNs and non-homogeneous Gompertz diffusion process (NHGDP) based on a Genetic Algorithm (GA)	GDP, temperature, hours of sunshine, and humidity
Hernández et al. (2013) [14]	Spain	Multilayer perceptron (MLP)	Rainfall, temperature, average wind speed, average wind direction, relative humidity, pressure, and solar radiation
Cui and Peng, (2015) [5]	China	Autoregressive model (AR), ARMA, ARIMAX, and Sigmoid Function ANN models (MLP)	Temperature
Vu et al. (2015) [15]	Australia	Multiple regression model	Population, GDP, CPI, temperature, wind speed, humidity, evaporation, rainfall, rainy days, solar exposure, and sunshine hours
Torrini et al., (2016) [6]	Brazil	Fuzzy logic methodology, Holt model with Two Parameters	Long-term Population and GDP
Suganthi and Samuel (2016) [16]	general	Econometric models	GDP, CPI, and population
Wang et al. (2016) [7]	China	Artificial Bee Colony (ABC) algorithm which combined with multivariate linear regression (MLR), ANN, and Quadratic Regression Model	GDP, fixed asset investment (FAI), foreign direct investment (FDI), population, urbanization level, carbon emission
Mohammed, (2018) [17]	Iraq	Logarithmic linear regression model and ANNs	Population, GDP, CPI, temperature, and war affect
Wu et al., (2018) [1]	China	Multivariable gray forecasting model GMC (1, N) with fractional order accumulation and the traditional gray forecasting model	GDP, NI, population, industrial output value, and FAI
Sahay et al. (2016) [18]	Canada	MLP	Temperature, day of the week, holiday indicator
Ahamad and Chen (2018) [19]	USA	artificial neural network with nonlinear autoregressive exogenous multivariable Inputs (ANN-NAEMI)	Relative humidity, outdoor air temperature, and global solar radiation
Kim et al. (2019) [20]	Korea	LSTM	temperature, humidity, season
Albuquerque et al. (2022) [21]	Brazil	Random Forest	Dates, weather variables, price, economic variables
Mehmood et al. (2022) [22]	France, Turley, Pakistan	A pool of 10 ML modes including the MLP	Energy generation, temperature
Raza et al. (2022) [23]	Pakistan	Energy modeling tool (LEAP)	GDP (total and sector-wise), population, household size
Rick and Berton (2022) [24]	Brazil	Convolutional neural network, LSTM, Autoencoders	Dates, distribution of energy data
Rao et al. (2023) [25]	China	Support Vector Regression SVR	GDP, population, industrial output, energy generation, among others

Vu et al. [15] used climatic and socioeconomic variables selected via statistical analysis. The monthly demand forecast was closely matched with the real electricity demand. Torrini et al. [6] addressed a fuzzy logic methodology approach to forecast long-term annual electricity demand using population and GDP additional value as variables. In Suganthi and Samuel [16] econometric models were developed to predict Indian energy consumption. The relationship between independent variables and energy consumption was verified using statistical models.

According to Mohammed [17], the comparison with other studies revealed that the selected influential factors have different impacts on the prediction for different countries.

Besides those investigations, recent studies have presented the importance of using neural-based and machine learning (ML) models to predict electricity demand, considering external variables. Traditional models such as MLP are addressed by Sahay et al. [18], Ahamad and Chen [19], and Mehmood et al. [22]. The Random Forest and Support Vector Machines are used by Albuquerque et al. [21] and Rao et al. [25], respectively. More complex models, such as deep learning approaches, are addressed by Raza et al. [23]. Kim et al. [20] and Rick and Berton [24] expose how crucial the nonlinear approaches are for solving this task.

Thus, as possible to perceive as presented in this section, different meteorological, socioeconomic, and environmental variables were included in most of the studies related to forecasting electricity demand. However, the selection of these independent variables was chosen by researchers and experts, which does not guarantee that this procedure is the best option for solving variable selection problems. Even though the authors considered their results satisfactory in some studies, they may have been improved by withdrawing or obtaining information about the model. In fact, as suggested by Vu et al. [15], using a few variables leads to a weak model, while increasing the number of variables leads to the construction of more robust models. Thus, it is necessary to develop procedures to help in the selection of the best set of inputs to increase de model's prediction capability.

Although these studies selected a group of factors with a justification, they did not necessarily include independent "causal" variables because the statistical tests cannot explain whether the relations between them imply the "cause" of a change in demand for electricity.

3. Methodology

The methodology comprised five phases, which are detailed in this section.

3.1. Phase 1—Selection of Variables

The main variables listed in Table 2 were considered independent variables from different studies mentioned in Section 2.

Table 2. Independent variables were selected in the literature review.

Type of Variables	Independent Variables	Alternatives
Socioeconomic variables	National Income (NI)	a1
	Population	a2
	Gross Domestic Product (GDP)	a3
	Consumer Price Index (CPI)	a4
	Fixed Asset Investment (FAI)	a5
	Foreign Direct Investment (FDI)	a6
	Industrial output value	a7
	Urbanization level	a8
Climate Variables	Temperature	a9
	Solar radiation	a10
	Relative humidity	a11
	Rainfall	a12
	Average wind direction	a13
	Average wind speed	a14
	Pressure	a15
	Evaporation	a16
	Rainy days	a17
Environmental variable	Carbon emission	a18

Thus, the independent variables were considered as alternatives to be included in the multi-criteria selection process. These alternatives were filtered in this phase, based on the application of the ELECTRE I method, which selects a core ("kernel") of the alternatives with a better compromise to the objective of the research. The next step was the definition of the criteria (Table 3) corresponding to the main characteristics found before the execution of the forecasting of demands. Still, in this phase, the decision-makers must perform a criteria validation by comparing the suggested literature and/or indicating new relevant criteria, if necessary.

Table 3. Criteria listed in the literature.

Parameters	Criteria	Features
Degree of complexity	g1—Data availability	Amount of information, numerical data, non-numerical data, and numerical analysis
	g2—Correlation with independent variable	Correlation of electricity demand with independent variables
Consumption pattern	g3—Consumption pattern by sector	Industrial, commercial, residential, rural, and other sectors
	g4—Consumption pattern by region	Geographic, economic, and social characteristics of the region
	g5—Consumption pattern by calendar	Holidays, weekends, special days

The criteria weights were defined in order to reflect the importance of each criterion. In this research, the criteria weights were established by the Analytic Hierarchy Process (AHP) method in a pairwise comparison carried out by two specialists on energy demand behavior.

The next stage of this phase was to choose the multi-criteria method for the selection process. This step aimed to select a single group of alternatives (Table 2) based on the collected data regarding each alternative for each criterion. At this point, an over-classification approach, based on the ELECTRE family (*ÉLimination Et Choix Traduisant la REalité*-Elimination and Choice Translating Reality), more specifically ELECTRE I, was chosen to perform the selection of the alternative (independent variables).

ELECTRE I method uses the concepts of outranking comparisons and the index of concordance and discordance (agreement and disagreement). It can be defined by the following concordance index [26]:

$$c(aSb) = \sum_{\{j:g_j(a)\geq g_j(b)\}} w_j \tag{1}$$

where $\{j : g_j(a) \geq g_j(b)\}$ is the set of indices for all the criteria belonging to the concordant coalition with the outranking relation aSb).

The value of the concordance index must be greater than or equal to a given concordance level, s, whose value generally falls within the range $[0.5,\ 1 - \min_{j\in\mathfrak{J}} w_j]$, i.e., $c(aSb) \geq s$.

The discordance index is measured based on a discordance level defined as follows:

$$d(aSb) = \max_{\{j:g_j(a)<g_j(b)\}} \{g_j(b) - g_j(a)\} \tag{2}$$

This level measures in some way the power of the discordant coalition, meaning that if its value surpasses a given level, v, the assertion is no longer valid. Discordant coalition exerts no power whenever $d(aSb) \leq v$. Both concordance and discordance indices have to be computed for every pair of actions (a, b) in the set A, where $a \neq b$.

3.2. Phase 2—Data Pre-Processing

It is consisted of a statistical and graphical time series analysis to identify the level, trend, and seasonality. The presence of a trend can be verified by the Cox-Stuart test, while the Friedman test can attest to seasonality. If the trend is confirmed, it is necessary to perform one or more differentiations until the time series becomes stationary, following Equation (3):

$$x'_t = x_{t+1} - x_t \tag{3}$$

where x'_t is the difference between the current value of the series (x_{t+1}) minus the previous value (x_t).

If a seasonal component is identified, a statistical process known as deseasonalization is applied to the original data x_t, according to Equation (4):

$$z_t = \frac{x_t - \mu_m}{\sigma_m} \tag{4}$$

where z_t is the deseasonalized series, μ_m is the average and σ_m is the standard deviation for each seasonal step m. After performing these initial phases, the demand forecasting stage can be started in the next phases.

3.3. Phase 3—Predictive Models

This phase initially made predictions with linear models and Artificial Neural Networks (ANNs):

- Holt-Winters Exponential Smoothing (HW);
- Autoregressive model (AR);
- Autoregressive integrated moving average model (ARIMA);
- Multilayer perceptron (MLP);
- Radial Basis Function Networks (RBF);
- Extreme Learning Machines (ELM).

The mathematical formalisms of those models are detailed in recent and classical references [27–30].

The endogenous time series corresponds to the energy demand, which is composed of 186 samples, comprising the monthly period from January 2004 to June 2019. For this research, the data were separated into three sets: (a) training (from January 2004 to December 2015, totaling 144 observations); (b) validation (from January 2016 to December 2017, totaling 24 observations); (c) test (from January 2018 to June 2019, totaling 18 observations). The number of data sets used was the same for cases where exogenous variables were included in the forecasting process. Next, another essential step was to estimate the parameters required by each predictive model.

Regarding the HW model, the parameters, such as the smoothing constants α, β, and γ, needed to be determined. According to Montgomery et al. [31], no exact method for defining these values exists. Therefore, for this research, empirical parametric tests were performed until values that exhibited the best adjustments for the forecasts were found. The AR and ARIMA models had their coefficients (p) and (q) determined from the partial autocorrelation graph (PACF) between the data series and its lags.

The most relevant lags used in the inputs to address neural networks, MLP, RBF, and ELM, were determined using the Wrapper method, with consideration of up to 6 lags. Regarding the architectures of the neural models, they were all built with one intermediate layer. A grid search starting at 5, with increments of 5 until 200 neurons, was performed to determine the number of neurons in that layer.

The MLP was trained using the backpropagation algorithm [30]. The non-supervised step of the RBF was adjusted using the k-means algorithm, and the supervised one was the backpropagation [27,30]. The stopping criterion addressed for both models was 2000 iterations. The ELM was trained using the Moore-Penrose Inverse Operation [27], a closed-form solution. In all cases, the activation function addressed was the hyperbolic tangent [30].

Having the configurations defined for all models, they were executed 30 times, and the best execution (lowest mean square error on the validation set) was selected. As the data were monthly, forecasts were made considering the following horizons: $h = 1$, $h = 3$, and $h = 6$.

3.4. Phase 4—Post-Processing of Data

In the fourth phase, post-processing of data was completed so that they are denormalized and de-standardized. Afterward, the data had to return to the original domain so that the performance resulting from the predictions could be compared. Thus, Equation (5) was applied to return to the seasonal pattern and Equation (6) to return to the trend behavior:

$$x_i = (z_i.\sigma_m) + \mu_m \tag{5}$$

$$x_t = x_{t-1} + x'_t \tag{6}$$

where the variables are the same in Equations (1) and (2).

Having the predictions placed in the original domain, a comparison was made among the performance obtained by the forecasting models, which corresponds to the fifth phase of the methodology.

3.5. Phase 5—Performance of Predictive Models

With the data already post-processed, the performances of predictions made were compared using three different error measures, namely Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), which were expressed by Equations (7), (8) and (9), respectively:

$$\text{MSE} = \frac{1}{N_S}\sum_{t=1}^{N_S}(d_t - y_t)^2 \tag{7}$$

$$\text{MAE} = \frac{1}{N_S}\sum_{t=1}^{N_S}|d_t - y_i| \tag{8}$$

$$\text{MAPE} = \frac{1}{N_S}\sum_{t=1}^{N_S}\left|\frac{d_t - y_i}{d_t}\right| \times 100 \tag{9}$$

where N_S is the number of samples, d_t is the desired result, and y_i is the output given by the models. The resulting performances of each model are shown in the following sections.

4. Results

This section discusses a real application of linear predictive models and ANNs, reaching the phases after the selection of independent variables by the multi-criteria analysis. To carry out this study, raw data on electricity consumption from the State of Paraná-Brazil was collected, available in the database of the Federal Government's Energy Research Company website [32]. These data refer to the monthly electricity consumption (in MWh) from January 2004 to June 2019, totaling 186 monthly samples of electricity consumption. The data were plotted in Figure 1, and components of trend (growth) and seasonality were visually identified, which were verified using the Cox-Stuart and Friedman non-parametric tests (Table 4).

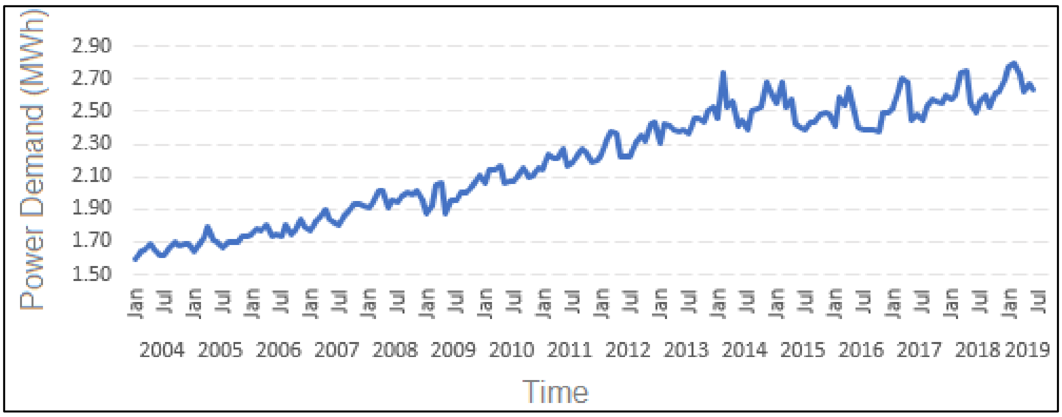

Figure 1. Paraná state energy consumption data for the period from January 2004 to June 2019.

Table 4. Non-parametric tests for trend and seasonality analysis.

Tests	Original Series p-Value < 5%	Differentiated and Seasonally Adjusted Series p-Value > 5%
Cox-Stuart	0	0.251322103
Friedman	3.49×10^{-15}	0.994761543

Considering the p-values < 0.05 in both tests, the presence of trend and seasonality in the series was statistically proven, as evidenced in Table 4 and by the visual analysis in Figure 1. Therefore, these components were removed through differentiation (Equation (3)) and deseasonalization (Equation (4)). The values of the preprocessed series in Figure 2 varied between +3 and −3, centered at zero, indicating the stationary condition.

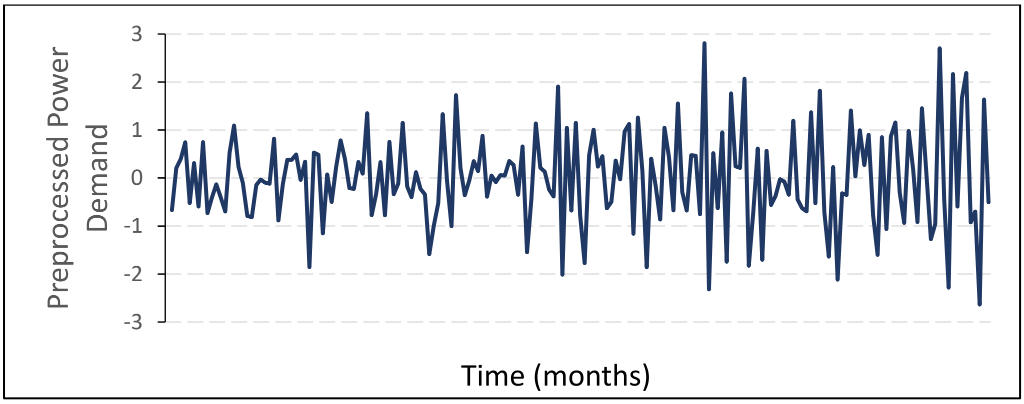

Figure 2. Paraná State energy consumption data for the period from January 2004 to June 2019, differentiated and seasonally adjusted.

Next, the parameters of each model were defined. The HW model was used for predicting series that present components of level, trend, and seasonality. It presents as parameters the smooth components (α, β, and γ). The additive model that provided the best results used $\alpha = 0.63$, $\beta = 0.12$, and $\gamma = 0.45$. The parameters of the AR (p) and ARIMA (p, d, q) models were estimated from the analysis of the partial autocorrelation function (PACF) for the stationary series (Figure 3). The parameter d was set to 1, corresponding to the number of differentiations necessary to withdraw from the trend. From the analysis of Figure 3, it can be seen that a series has a significant partial autocorrelation coefficient up to the fifth lag, with a 95% confidence level. From 1 to 5 lags for p and q, the AR (2) and ARIMA (5, 1, 5) models presented the smallest forecasting errors. To adjust the models, maximum likelihood estimators were addressed.

For ANNs, the number of neurons in the intermediate layer was empirically chosen to consider a grid search, starting with 5 neurons and using increments of 5, until reaching 200 neurons.

The method wrapper was applied for each variation, which made a progressive scan considering up to six lags. The configurations of the ANNs that gave the best prediction results were the MLP of 5 neurons in the middle layer and with the third lag as an input for prediction; RBF with 10 neurons and 6 lags; ELM with 25 neurons and lags 1 and 6. After the forecasting, one step ahead ($h = 1$) for the energy consumption series was performed.

Table 5 lists the values of MSE in the original domain of the series, MSE (d) that represents the error of considering the series after the preprocessed adjusted values, MAE, MAE (d), and MAPE (as it is a percentage, the measurement is not sensitive to the treatment of the data), as well as their respective performance rankings. The comparatives used data from the test portion. The MSE results are in $(\text{MWh})^2$ and the MAE in MWh.

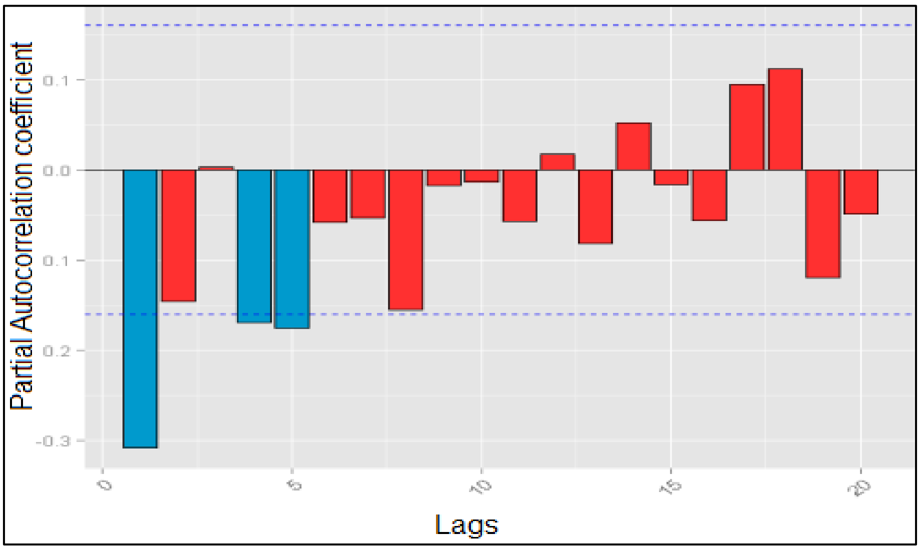

Figure 3. Graph of the partial autocorrelation function (PACF) for the differentiated and seasonally adjusted time series of electricity consumption in Paraná.

Table 5. Forecast results for one step ahead, compared by the error measures MSE, MAE, MAPE, and their respective performance rankings.

One Step Ahead						
Model	**Model's Parameters**	**MSE**	**MSE(d)**	**MAE**	**MAE(d)**	**MAPE**
HW	$\alpha = 0.63$, $\beta = 0.12$, $\gamma = 0.45$	5679573386.83	0.2566	62124.83	0.4286	2.3265
AR	$p = 2$	5247636249.96	0.2371	60064.66	0.4144	2.2765
ARIMA	$p = 5, d = 1, q = 5$	5860003521.64	0.2647	65994.23	0.4553	2.5035
MLP	neurons = 5, lag = 3	6557391893.34	0.3152	70231.17	0.4796	2.6589
RBF	neurons = 10, lag = 6	4634913071.81	0.2094	55409.00	0.3823	2.0994
ELM	neurons = 25, lags = 1, 6	4599769468.03	0.1945	48604.73	0.3463	1.8538

Performance ranking					
Model	**MSE**	**MSE (d)**	**MAE**	**MAE (d)**	**MAPE**
HW	4	4	4	4	4
AR	3	3	3	3	3
ARIMA	5	5	5	5	5
MLP	6	6	6	6	6
RBF	2	2	2	2	2
ELM	1	1	1	1	1

The results listed in Table 5 show that the best performance was obtained by the ELM, followed by the RBF, for all the error measures analyzed, whereas the MLP exhibited the worst results. Figure 4 shows the data given by the models and the actual data separated for testing.

Although visually the results in Figure 4 seem close, the series under analysis has monthly consumption in the order of millions of Wh. The difference between the actual and estimated value, which is represented in thousands of Wh, when squared and divided by the number of forecasts (MSE calculation), can reach billions of Wh per month for the entire state since the MSE penalizes greater errors compared to other performance measures. In the sequence, Table 6 presents the error metrics and the ranking considering three steps ahead, while Figure 5 shows the forecasting made by the models and the actual data separated for testing.

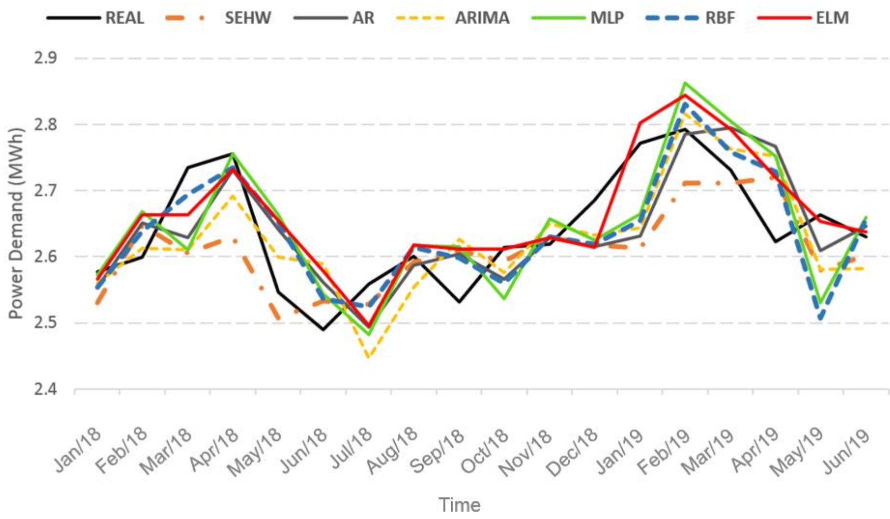

Figure 4. Graph comparing actual data with data predicted for 1 step ahead, using HW, AR, ARIMA, MLP, RBF, and ELM models.

Table 6. Forecast results for three steps ahead, compared by the error measures MSE, MAE, MAPE, and their respective performance rankings.

Three Steps Ahead						
Model	**Model's Parameters**	**MSE**	**MSE(d)**	**MAE**	**MAE(d)**	**MAPE**
HW	$\alpha = 0.63$, $\beta = 0.12$, $\gamma = 0.45$	7485077440.82	0.3292	72273.26	0.4987	2.7328
AR	$p = 2$	7013687861.15	0.3247	70897.07	0.4836	2.6770
ARIMA	$p = 5$, $d = 1$, $q = 5$	8924931364.90	0.4853	80795.54	0.5770	3.0485
MLP	neurons = 5, lag = 3	6573580787.03	0.3162	70331.85	0.4806	2.6625
RBF	neurons = 10, lag = 6	6528243517.99	0.2871	62256.53	0.4296	2.3556
ELM	neurons = 25, lags = 1, 6	6366837641.56	0.2740	64652.19	0.4296	2.4487

Performance ranking					
Model	**MSE**	**MSE (d)**	**MAE**	**MAE (d)**	**MAPE**
HW	5	5	5	5	5
AR	4	4	4	4	4
ARIMA	6	6	6	6	6
MLP	3	3	3	3	3
RBF	2	2	1	1	1
ELM	1	1	2	2	2

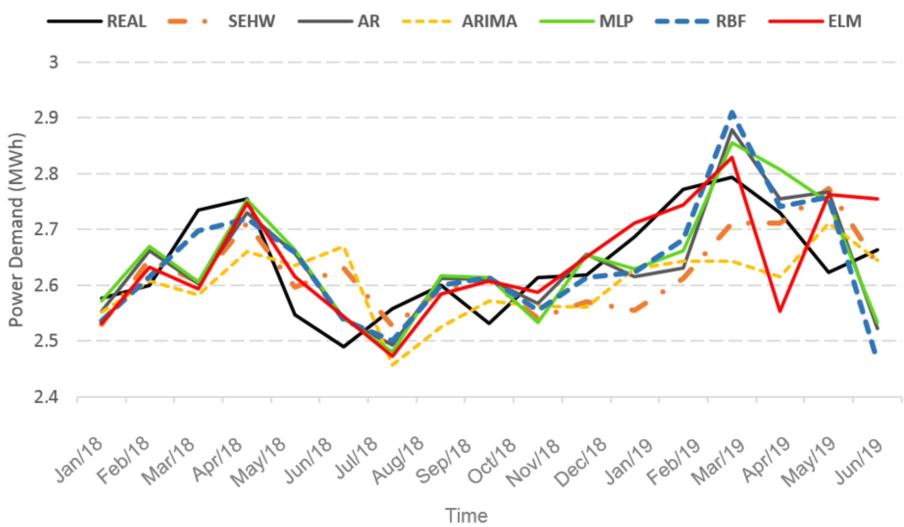

Figure 5. Graph comparing actual data with data predicted for three steps ahead, using HW, AR, ARIMA, MLP, RBF, and ELM models.

The results listed in Table 6 show that the ELM once again had higher forecasting performance than the others considering the MSE, whereas the RBF network was the best one considering MAE and MAPE. It is noteworthy that MAE and MAPE do not penalize large errors as the MSE does. Concerning the MSE in the real domain, the error of the RBF compared to the ELM in percentage terms was 2.54% higher, while from the MLP was 3.25% higher. On the other hand, the AR, HW, and ARIMA models performed much lower compared to the neural models.

Finally, forecasts were obtained considering the six steps ahead, whose results are listed in Table 7, while Figure 6 brings the predictions made by the models and the actual data separated for testing.

Table 7. Forecast results for six steps ahead, compared by the error measures MSE, MAE, MAPE, and their respective performance rankings.

Six Steps Ahead						
Model	**Model's Parameters**	**MSE**	**MSE(d)**	**MAE**	**MAE(d)**	**MAPE**
HW	$\alpha = 0.63$, $\beta = 0.12$, $\gamma = 0.45$	8283400542.03	0.3916	76488.21	0.5268	2.8931
AR	$p = 2$	6840287879.73	0.3175	71062.50	0.4815	2.6849
ARIMA	$p = 5$, $d = 1$, $q = 5$	8174710719.13	0.3865	73181.97	0.5041	2.7387
MLP	neurons = 5, lag = 3	6686402960.82	0.3168	69819.32	0.4766	2.6426
RBF	neurons = 10, lag = 6	6595147324.83	0.3118	68586.05	0.4724	2.5950
ELM	neurons = 25, lags = 1, 6	6406990346.37	0.3020	63890.27	0.4487	2.4310

Performance ranking					
Model	**MSE**	**MSE (d)**	**MAE**	**MAE (d)**	**MAPE**
HW	6	6	6	6	6
AR	4	4	4	4	4
ARIMA	5	5	5	5	5
MLP	3	3	3	3	3
RBF	2	2	2	2	2
ELM	1	1	1	1	1

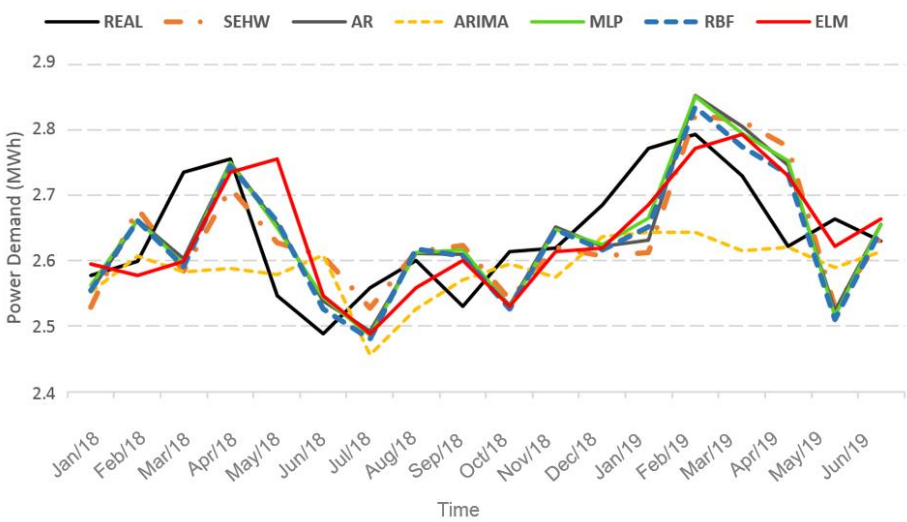

Figure 6. Graph comparing actual data with data predicted for six steps ahead, using HW, AR, ARIMA, MLP, RBF, and ELM models.

It is necessary to note that the ELM and RBF maintained the position in the ranking with the best performances for all horizons, while the HW, for six steps ahead, had the worst performance. Comparing the MSE in the real domain, the ELM in relation to the RBF, MLP, AR, ARIMA, and HW models allowed a reduction in the percentage error by 3.94, 4.36, 6.76, 27.59, and 29.29%, respectively.

After the predictions for one, three, and six steps ahead considering the independent variables in the univariate forecasting, the next stage was to evaluate the use of these independent variables in multivariate forecasting.

4.1. Selection of Independent Variables

The data collection of the exogenous variables was selected from some bases, such as the System for Estimating Greenhouse Gas Emissions (SEEG); National Meteorological Institute (INMET); Institute of Applied Economic Research (IPEA); Brazilian Institute of Geography and Statistics (IBGE); and Chamber of Commercialization of Electric Energy (CCEE). Those are official organisms of the Brazilian government that make available the data to researchers.

4.1.1. Criteria

The criteria of both qualitative and quantitative types were validated by the decision-makers. The measurement of qualitative ones was carried out based on the perception and experience of the decision-makers and experts in the theme. In sequence, are presented details related to the criteria adopted:

Criterion g1—Data availability: This criterion is quantitative and is related to the availability of data for training models. The score attributed to the alternatives is represented by Equation (10):

$$gd = \frac{qntvi}{qntvd} \times 100 \tag{10}$$

where gd represents the degree of data availability, $qntvi$ corresponds to the amount of data available for the independent variable under analysis, and $qntvd$ refers to the amount of data available for training and validation of the dependent series (consumption of electricity in Paraná).

Criterion g2—Correlation with the dependent variable: This quantitative criterion seeks to correlate the variable "electricity consumption" with exogenous variables (alternatives). Pearson's correlation coefficient (r) was used [33], and for reasons of collinearity, the maximum score (100) was assigned to coefficients with values of 0.8 and −0.8. Equations (11) and (12) express the score attributed to the alternatives, with Equation (11) being used for positive r values and Equation (12) for negative ones:

$$d = \frac{r - 0}{0.8} \times 100 \tag{11}$$

$$d = \frac{r - 0}{-0.8} \times 100 \tag{12}$$

These measures represent the percentages of the distances (d) of correlation coefficient values concerning the zero value, which means that the closer to zero the correlation coefficients, the lower the score attributed to these alternatives, while the values −0.8 and 0.8 present a score equal to 100.

Criterion g3—Consumption pattern by sector: Some variables might influence the demand depending on particular contexts, varying according to the sector. In the time series of electric energy consumption, the percentages corresponding to the consumption of industrial, residential, commercial, and other sectors are 41.68, 23.72, 18.36 and 16.23%, respectively [32].

The scale developed for this criterion was based on this percentage, that is, if the decision-makers consider that the independent variable has an influence on energy consumption only in the industrial and commercial sectors, for example, the score attributed to this variable corresponds to the percentage sum of consumption of these sectors (41.68% + 18.36% = 60.04%). This criterion is qualitative due to the value attributed to each alternative to the analyzed context. Therefore, the ELECTRE I decision matrix can be made based on the perception and experience of the decision-makers. The g4 and g5 criteria outlined below are also qualitative in nature, as there are no parameters for their measurement; so they were assessed based on the 5-point Likert ordinal scale (Table 8).

Table 8. Example of the scale used in criteria g4 and g5.

Value	Description
5	Very high
4	High
3	Average
2	Low
1	Very low

Criterion g4—Consumption patterns by region: (geographical, economic, and social characteristics). It determines the influence that the explanatory variables can have. The same variable presents very high or very low influence in different regions due to the different locations' characteristics.

Criterion g5—Consumption pattern by the calendar: The variables that influence demand may present different frequencies depending on the calendar (vacation, weekdays, weekends, holidays, among others). This criterion determines how recurrent (or not) the influence of an explanatory variable on demand is due to the calendar pattern. With the criteria established for selecting variables, the next step was to apply the AHP method to identify the degree of importance of each criterion that was established *a priori*.

4.1.2. Application of AHP (Weights Definition)

The AHP method proposed by Saaty [34] was used to determine the weights regarding each criterion. The two decision-makers performed the paired comparison (AHP method) concerning the above five criteria. With a peer review, the decision-makers answered which criteria they considered the most important and what intensity of importance one presented in relation to the other.

The weight of the decision-makers' judgments was considered equal, with the same importance. After performing this comparison alongside, it was possible to obtain the judgment matrix (Table 9), which had its data normalized by dividing the value of the judgment obtained for each criterion by the total sum of its respective column (Table 10). These expressed values referring to the preference vectors of each criterion (Table 11) were calculated by the geometric mean of the normalized values.

Table 9. AHP evaluation matrix (Part A).

Criteria	g1	g2	g3	g4	g5
g1	1	1/4	3	5	3
g2	4	1	5	7	5
g3	1/3	1/5	1	2	1
g4	1/5	1/7	$\frac{1}{2}$	1	$\frac{1}{2}$
g5	1/3	1/5	1	2	1
Sum	**5.86**	**1.79**	**10.5**	**17**	**10.5**

Table 10. AHP normalized weights considered for the criteria (Part B), in which GM = Geometric mean CR = 0.0308.

CR	g1	g2	g3	g4	g5	GM
g1	0.17	0.14	0.29	0.29	0.29	0.24
g2	0.66	0.56	0.48	0.41	0.48	0.51
g3	0.06	0.11	0.1	0.12	0.1	0.10
g4	0.03	0.08	0.05	0.06	0.05	0.05
g5	0.06	0.11	0.1	0.12	0.1	0.10

Table 11. Criteria score ranking.

Ranking	Criterion	Score	Description
1	g2	0.51	Correlation with the dependent variable
2	g1	0.24	Data availability
3	g3	0.10	Consumption pattern by sector
4	g4	0.10	Consumption pattern by calendar
5	g5	0.05	Consumption pattern by region

Finally, the ratio of consistency (CR) was calculated based on the AHP theory formulations [34] to verify the coherence in the judgments of the decision-makers. The condition of consistency of acceptable judgments is CR $\leq$ 0.10 [34]. The value of CR calculated for this problem was 0.0308, and therefore it was followed by the stage of variable selection by ELECTRE I.

4.1.3. Application of ELECTRE I (Selection of Independent Variables)

For the ELECTRE I application, an evaluation matrix was built based on the quantitative scales previously defined and the qualitative scales, which were assigned according to the premises established by the decision-makers. The data were also normalized using Equation (13) so that all criteria alternatives had equivalent evaluations:

$$Benefit = \frac{\left(V_{ij} - MinV_j\right)}{\left(MaxV_j - MinV_j\right)} \tag{13}$$

where V_{ij} is the value of the alternative being evaluated, while $MinV_{\mathrm{j}}$ and $MaxV_j$ are the minimum and maximum values of the alternatives for a given criterion.

As a result, the standardized decision matrix shown in Table 12 was obtained, in which the variables were previously defined in Tables 2 and 11.

Table 12. Ranking of criteria scores.

ai/gj	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18
g1	0.07	0.07	0.29	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.78	1.00	1.00	0.92	0.72	0.08
g2	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.00	0.56	0.35	0.20	0.37	0.00	0.13	0.25	1.00	0.24	0.00
g3	1.00	0.40	1.00	0.58	0.00	0.60	0.58	0.58	1.00	0.58	0.42	0.42	0.00	0.00	0.00	0.00	0.42	0.60
g4	0.75	0.50	1.00	0.75	0.25	0.50	1.00	1.00	1.00	0.25	0.00	0.25	0.00	0.00	0.00	0.00	0.25	0.50
g5	0.00	0.33	0.00	0.67	0.00	0.33	0.00	0.67	1.00	0.33	0.00	0.33	0.00	0.00	0.00	0.00	0.33	0.33

After that, the concordance matrix was initially calculated (Table 13). Each index that composes this matrix was determined as shown by the calculation of alternative one in relation to two, considering all criteria (Equation (14)).

$$C_{1,2} = g_1(if\ a_1 \geq a_2) + g_2(if\ a_1 \geq a_2) + g_3(if\ a_1 \geq a_2) + g_4(if\ a_1 \geq a_2) + g_5(if\ a_1 \geq a_2)/\sum g \tag{14}$$

where the variables were defined in Tables 2 and 3.

The concordance index $C_{1,2}$ represents the weights of the criteria that are added when the alternative 'a1' exceeds 'a2'. The calculation is performed in the pairwise comparison of all alternatives until the matrix was established. In this example, we have: $C_{1,2} = 0.24(a1 = a2) + 0.52(a1 = a2) + 0.1(a1 > a2) + 0.05(a1 > a2) + 0(a1 < a2)/\Sigma g = 0.90$.

For the discordance indexes, the scales of each criterion were first calculated, corresponding to the value of the difference of the highest value assigned to the alternative minus the lowest value. Thus, the scale of all criteria resulted in 1.0. Therefore, the discordance matrix (Table 14) could be elaborated, comparing the alternative 'a1' in relation to 'a2'. Equation (15) shows this process:

$$D_{1,2} = Max\left[g_1\frac{(a2-a1)}{1}; g_2\frac{(a2-a1)}{1}; g_3\frac{(a2-a1)}{1}; g_4\frac{(a2-a1)}{1}; g_5\frac{(a2-a1)}{1}\right] \quad (15)$$

Finally, for the alternatives to be classified, a selection matrix containing only the values 0 and 1 was created (Table 15).

Table 13. Concordance matrix generated by pairwise comparison of all alternatives.

C	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18
a1		0.90	0.71	0.15	1.00	0.90	0.71	0.85	0.10	0.15	0.25	0.15	0.76	0.25	0.25	0.25	0.15	0.66
a2	0.85		0.61	0.00	1.00	0.90	0.61	0.75	0.00	0.15	0.15	0.15	0.76	0.25	0.25	0.25	0.15	0.66
a3	1.00	0.90		0.15	1.00	0.90	0.76	0.90	0.15	0.15	0.25	0.15	0.76	0.25	0.25	0.25	0.15	0.90
a4	0.90	1.00	0.85		1.00	0.90	0.95	0.95	0.24	1.00	1.00	1.00	1.00	1.00	1.00	0.49	1.00	0.90
a5	0.61	0.51	0.61	0.00		0.75	0.61	0.75	0.00	0.05	0.15	0.05	0.76	0.25	0.25	0.25	0.05	0.51
a6	0.61	0.76	0.61	0.10	1.00		0.71	0.85	0.00	0.25	0.25	0.25	0.76	0.25	0.25	0.25	0.25	0.76
a7	0.90	0.90	0.90	0.39	1.00	0.80		0.90	0.29	0.39	0.49	0.39	1.00	0.49	0.49	0.49	0.39	0.80
a8	0.66	0.76	0.66	0.25	1.00	0.90	0.76		0.05	0.25	0.25	0.25	0.76	0.25	0.25	0.25	0.25	0.66
a9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	1.00	1.00	0.49	1.00	1.00
a10	0.85	0.95	0.85	0.34	1.00	0.85	0.95	0.85	0.24		1.00	0.49	1.00	1.00	1.00	0.49	1.00	0.85
a11	0.85	0.85	0.85	0.24	0.95	0.75	0.85	0.75	0.24	0.24		0.34	1.00	1.00	0.49	0.49	0.34	0.75
a12	0.85	0.95	0.85	0.24	1.00	0.85	0.85	0.75	0.24	0.90	1.00		1.00	1.00	1.00	0.49	1.00	0.85
a13	0.85	0.75	0.85	0.00	0.95	0.75	0.61	0.75	0.00	0.00	0.15	0.00		0.25	0.25	0.25	0.24	0.75
a14	0.85	0.75	0.85	0.24	0.95	0.75	0.85	0.75	0.24	0.24	0.39	0.24	1.00		0.49	0.49	0.24	0.75
a15	0.85	0.75	0.85	0.24	0.95	0.75	0.85	0.75	0.24	0.24	0.90	0.24	1.00	1.00		0.49	0.75	0.75
a16	0.85	0.75	0.85	0.51	0.95	0.75	0.61	0.75	0.51	0.51	0.66	0.51	1.00	0.76	0.76		0.75	0.75
a17	0.85	0.95	0.85	0.00	1.00	0.85	0.61	0.75	0.00	0.15	0.76	0.25	0.76	0.76	0.25	0.25		0.85
a18	0.85	1.00	0.61	0.10	1.00	1.00	0.71	0.85	0.00	0.25	0.25	0.25	0.76	0.25	0.25	0.25	0.25	

Table 14. Discordance matrix generated by pairwise comparison of all alternatives.

D	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18
a1		0.33	0.25	0.93	0.00	0.33	0.93	0.67	1.00	0.93	0.93	0.93	0.71	0.93	0.93	1.00	0.65	0.33
a2	0.60		0.60	0.93	0.00	0.20	0.93	0.50	0.93	0.93	0.93	0.93	0.71	0.93	0.93	1.00	0.65	0.20
a3	0.00	0.33		0.71	0.00	0.33	0.71	0.67	1.00	0.71	0.71	0.71	0.49	0.71	0.71	1.00	0.43	0.33
a4	0.42	−0.18	0.42		−0.48	0.02	0.25	0.25	0.42	0.00	0.00	0.00	−0.22	0.00	0.00	0.52	−0.16	0.02
a5	1.00	0.40	1.00	1.00		0.60	1.00	0.75	1.00	1.00	1.00	1.00	0.78	1.00	1.00	1.00	0.72	0.60
a6	0.40	0.07	0.50	1.00	0.00		1.00	0.50	1.00	1.00	1.00	1.00	0.78	1.00	1.00	1.00	0.72	0.08
a7	0.42	0.33	0.42	0.67	0.00	0.33		0.67	1.00	0.35	0.20	0.37	0.00	0.13	0.25	1.00	0.33	0.33
a8	0.42	0.07	0.42	1.00	0.00	0.02	1.00		1.00	1.00	1.00	1.00	0.78	1.00	1.00	1.00	0.72	0.08
a9	0.00	−0.50	0.00	0.00	−0.56	−0.40	0.00	0.00		0.00	0.00	0.00	−0.22	0.00	0.00	0.44	−0.28	−0.40
a10	0.50	0.25	0.75	0.50	0.00	0.25	0.75	0.75	0.75		0.00	0.02	−0.22	0.00	0.00	0.65	0.00	0.25
a11	0.75	0.50	1.00	0.75	0.25	0.50	1.00	1.00	1.00	0.33		0.33	0.00	0.00	0.05	0.80	0.33	0.50
a12	0.58	0.25	0.75	0.50	0.00	0.25	0.75	0.75	0.75	0.16	0.00		−0.22	0.00	0.00	0.63	0.00	0.25
a13	1.00	0.50	1.00	0.75	0.25	0.60	1.00	1.00	1.00	0.58	0.42	0.42		0.22	0.25	1.00	0.42	0.60
a14	1.00	0.50	1.00	0.75	0.25	0.60	1.00	1.00	1.00	0.58	0.42	0.42	0.00		0.12	0.87	0.42	0.60
a15	1.00	0.50	1.00	0.75	0.25	0.60	1.00	1.00	1.00	0.58	0.42	0.42	0.00	0.00		0.75	0.42	0.60
a16	1.00	0.50	1.00	0.75	0.25	0.60	1.00	1.00	1.00	0.58	0.42	0.42	0.00	0.08	0.08		0.42	0.60
a17	0.58	0.25	0.75	0.50	0.00	0.25	0.75	0.75	0.75	0.28	0.28	0.28	0.06	0.28	0.28	0.76		0.25
a18	0.40	0.00	0.50	0.92	0.00	0.00	0.92	0.50	0.92	0.92	0.92	0.92	0.70	0.92	0.92	1.00	0.64	

Table 15. Selection matrix generated by the condition defined by the concordance and discordance and thresholds.

D	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18
a1		0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
a2	0		0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
a3	1	0		0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
a4	0	1	0		1	1	1	1	0	1	1	1	1	1	1	0	1	1
a5	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0
a6	0	1	0	0	1		0	0	0	0	0	0	0	0	0	0	0	1
a7	0	0	0	0	1	0		0	0	0	0	0	1	0	0	0	0	0
a8	0	1	0	0	1	1	0		0	0	0	0	0	0	0	0	0	0
a9	1	1	1	1	1	1	1	1		1	1	1	1	1	1	0	1	1
a10	0	1	0	0	1	1	0	0	0		1	0	1	1	1	0	1	1
a11	0	0	0	0	1	0	0	0	0	0		0	1	1	0	0	0	0
a12	0	1	0	0	1	1	0	0	0	1	1		1	1	1	0	1	1
a13	0	0	0	0	1	0	0	0	0	0	0	0		0	0	0	0	0
a14	0	0	0	0	1	0	0	0	0	0	0	0	1		0	0	0	0
a15	0	0	0	0	1	0	0	0	0	0	0	0	1	1		0	0	0
a16	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1		0	0
a17	0	1	0	0	1	1	0	0	0	0	1	0	1	1	0	0		1
a18	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	

According to Saaty [34], it is assigned the value one only if alternative "a" over classify "b", for example. For the calculation of this matrix, the decision-makers defined the thresholds of concordance (0.7) and discordance (0.3).

This outranking relationship occurs if the indexes of the concordance matrix comply with the condition of equal to the concordance threshold and if the indexes of the discordance matrix comply with the condition of equal to the threshold of discordance. For cases that do not obey the condition, the assigned value is 0 [35].

As a result, according to the established criteria, the alternatives that were selected as independent variables were a9 and a16, which correspond to temperature and evaporation (according to Table 1).

Note that the correlation coefficients between the target variable (power demand) and a9 and a16 are 0.1568 and −0.2782, respectively.

4.2. Forecasting with Exogenous Variables

After the selection of the exogenous variables, the data were collected to identify the regularity. To this purpose, the data of evaporation and temperature were plotted in Figure 7, from which it is possible to observe the presence of seasonality in both series and the absence of a trend.

Aiming to confirm or not the absence of trend and presence of seasonality, the Cox-Stuart and the Friedman non-parametric statistical tests were applied. The samples preprocessed by deseasonalization of Equation (2) of both series are shown in Figure 8.

The results listed in Table 16 show that the *p*-value for the trend of both analyzed series was greater than 5%, indicating that there is no trend in the series, while the *p*-value for the seasonality was less than 5%, statistically proving the presence of seasonality. After removing the seasonality, the series became stationary, and the Friedman test *p*-values for evaporation and temperature were 0.997947317 and 0.999969254, respectively.

It is noteworthy that there was no trend in the evaporation series. Additionally, the missing data from September to December 2015 and April to December 2016 were replaced by the respective mean values corresponding to the months analyzed.

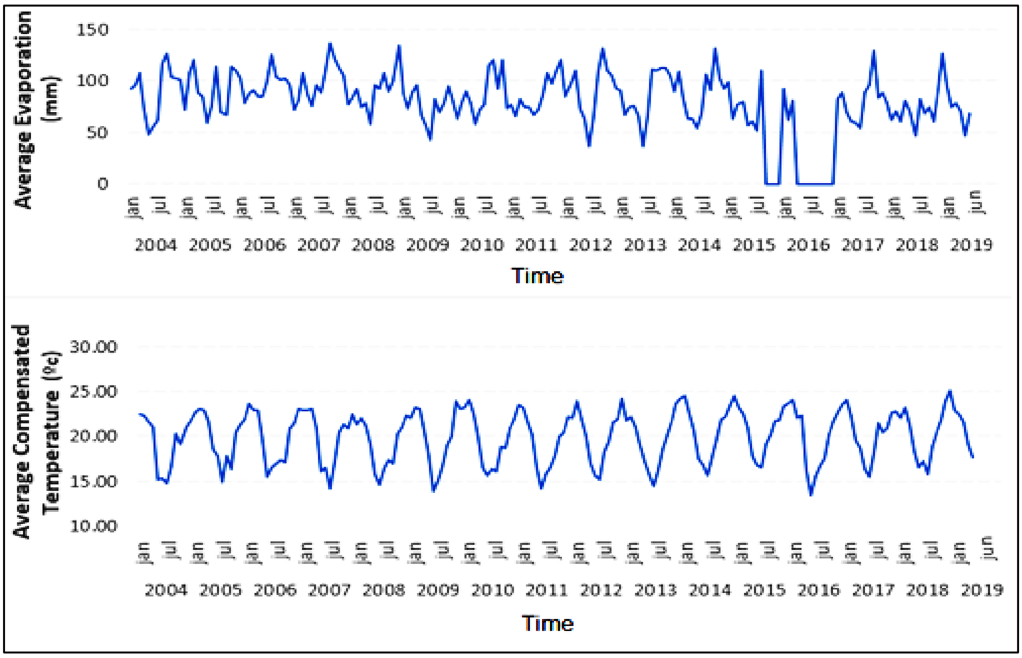

Figure 7. Evaporation and temperature time series for the period from January 2004 to June 2019.

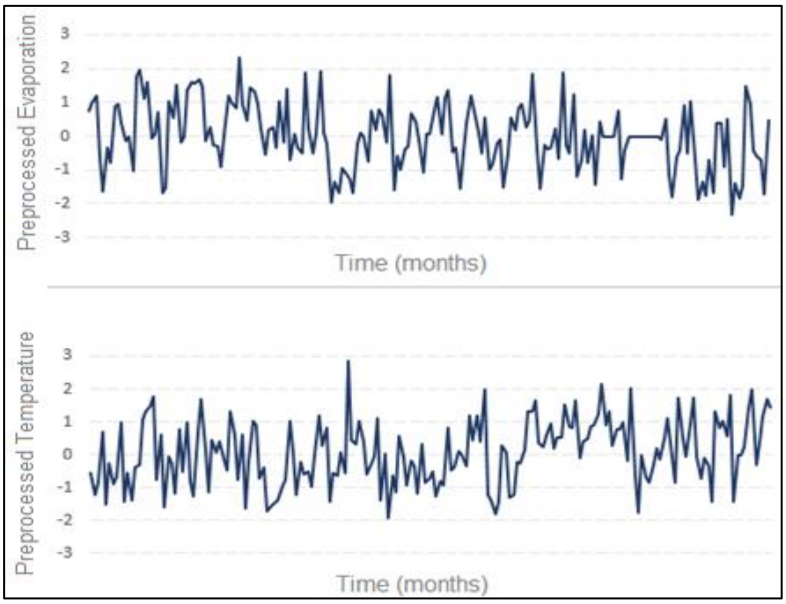

Figure 8. Evaporation and average temperature data for the State of Paraná for the period from January 2004 to June 2019, after passing through the deseasonalization process.

Table 16. Non-parametric tests for trend and seasonality analysis.

Tests	Cox-Stuart p-Value > 5%	Friedman p-Value < 5%
Evaporation	0.146177029	7.58419×10^{-8}
Temperature	1	1.37641×10^{-26}

The next step was to define the parameters for predicting the ANNs with the exogenous variables. The parameters of the ANNs that ensured the best prediction were as follows:

- MLP with 35 neurons in the intermediate layer, using the first energy lag, sixth temperature lag, and evaporation room as inputs;
- RBF with 50 neurons and lag 1 for energy, lag 6 for temperature, and lag 4 for evaporation;
- ELM with 45 neurons and lag 1 and 6 for energy, lag 3 for evaporation, and lag 6 for temperature.

After the parametrization of the parameters of the models, the electricity demand was forecast for one step ahead. Table 17 gathers the results obtained for the test set and the performance ranking, with errors referring to the best of 30 simulations performed for each model. It can be seen that the RBF ensured the best performance, followed by ELM for all error measures, while the MLP was the worst one. Figure 9 shows the predictions made by the models and the actual data in the test set.

Table 17. Forecast results for one step ahead, compared by the error measures MSE, MAE, MAPE, and their respective performance rankings.

One Step Ahead						
Model	**Model's Parameters**	**MSE**	**MSE (d)**	**MAE**	**MAE (d)**	**MAPE**
MLP	neurons = 35, lags = 1, 6, 4	5854808140.90	0.2363	62140.20	0.4015	2.3501
RBF	neurons = 50, lags = 2, 6, 6	3900550265.86	0.2142	50749.56	0.3643	1.9354
ELM	neurons = 45, lags = 1, 6, 3, 6	4154631401.10	0.2180	51866.49	0.3673	1.9694
Performance ranking						
Model	**MSE**	**MSE (d)**	**MAE**	**MAE (d)**	**MAPE**	
MLP	3	3	3	3	3	
RBF	1	1	1	1	1	
ELM	2	2	2	2	2	

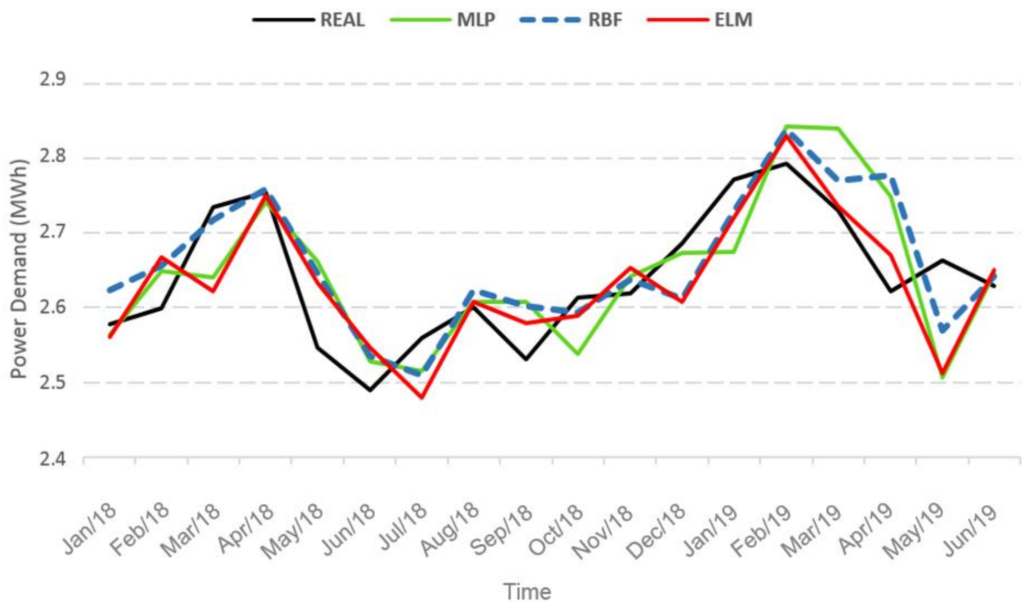

Figure 9. Graph comparing actual data with data predicted for one step ahead, using MLP, RBF, and ELM.

Table 18 shows the predictions for the same series considering three steps ahead and the performance ranking.

Table 18. Forecast results for three steps ahead, compared by the error measures MSE, MAE, MAPE, and their respective performance rankings.

Three Steps Ahead						
Model	**Model's Parameters**	**MSE**	**MSE (d)**	**MAE**	**MAE (d)**	**MAPE**
MLP	neurons = 35, lags = 1, 6, 4	6248871037.89	0.2823	66536.95	0.4368	2.5146
RBF	neurons = 50, lags = 2, 6, 6	6205419571.17	0.2761	67111.50	0.4544	2.5330
ELM	neurons = 45, lags = 1, 6, 3, 6	5536080184.82	0.2594	58071.25	0.4042	2.2150

Performance ranking					
Model	**MSE**	**MSE (d)**	**MAE**	**MAE (d)**	**MAPE**
MLP	3	3	2	2	2
RBF	2	2	3	3	3
ELM	1	1	1	1	1

Comparing MSE, ELM was the best among the tested models, followed by the RBF network. However, the positioning of RBF and MLP depended on the other error measures. Figure 10 shows predictions made by the models and the actual data separated for testing.

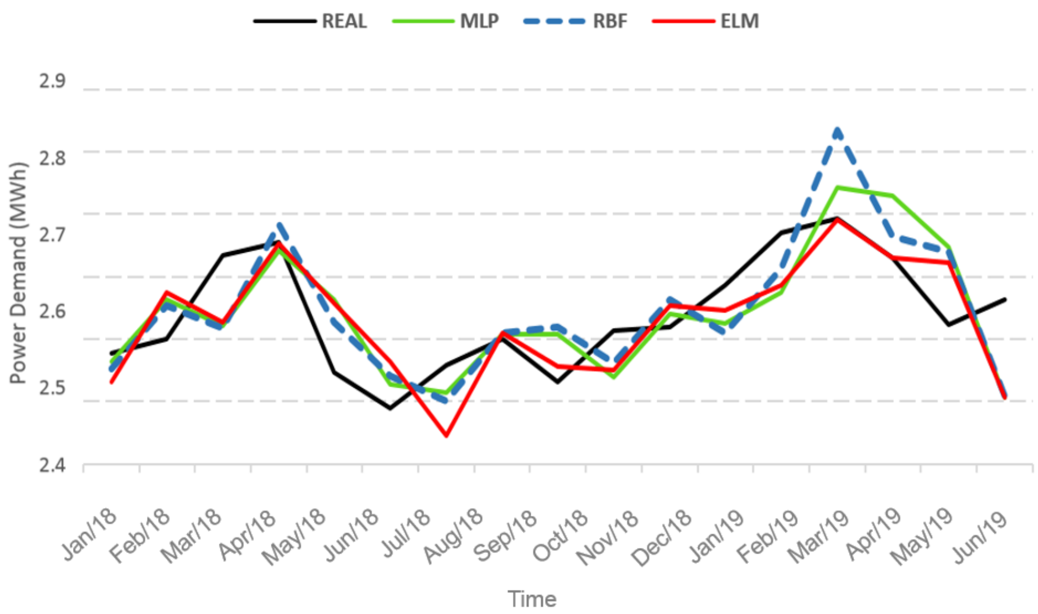

Figure 10. Graph comparing actual data with data predicted for three steps ahead, using MLP, RBF, and ELM.

Finally, predictions were also made considering the six steps ahead, whose results are presented in Table 19 by the error metrics, as well as the performance ranking.

On the other hand, Figure 11 depicts the predictions made in the models and the data used in the test set. The ELM showed the best performance for the MSE, while the MLP for the MAE and MAPE. It is noticed RBF was the second best in the ranking for all forecasting horizons. As mentioned, MAE and MAPE do not penalize large errors. MSE percentage reduction by the ELM in comparison with RBF and MLP was 5.11% and 6.36%, respectively.

Table 19. Forecast results for six steps ahead, compared by the error measures MSE, MAE, MAPE, and their respective performance rankings.

Six Steps Ahead						
Model	**Model's Parameters**	**MSE**	**MSE (d)**	**MAE**	**MAE (d)**	**MAPE**
MLP	neurons = 35, lags = 1, 6, 4	6433456508.91	0.3195	68395.90	0.4847	2.5895
RBF	neurons = 50, lags = 2, 6, 6	6358076170.74	0.3180	68783.06	0.4873	2.6132
ELM	neurons = 45, lags = 1, 6, 3, 6	6049034166.40	0.3165	70008.28	0.4886	2.6619

Performance ranking					
Model	**MSE**	**MSE (d)**	**MAE**	**MAE (d)**	**MAPE**
MLP	3	3	1	1	1
RBF	2	2	2	2	2
ELM	1	1	3	3	3

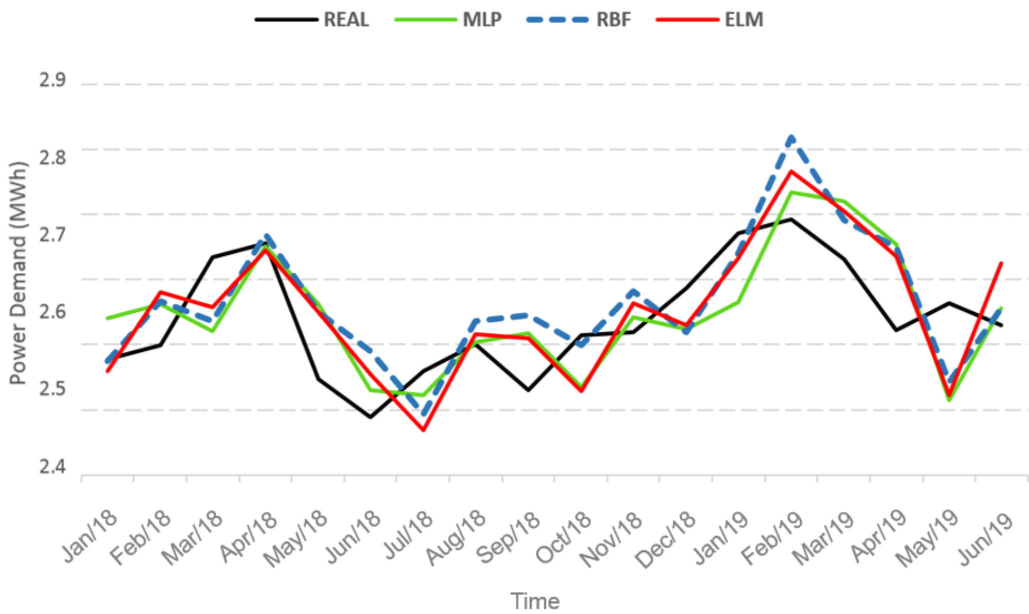

Figure 11. Graph comparing actual data with data predicted for six steps ahead, using MLP, RBF, and ELM.

Comparing the best predictor for each forecasting horizon, considering the univariate forecasting (ELM), and considering the best model's proposed method, we have the following percentage gains: 17.93% for one step ahead, 15.01% for three steps ahead, and 5.96% for six steps ahead.

5. Discussion

In an initial analysis considering only the univariate electricity demand forecasting, it was possible to notice that all neural models with the exception of the MLP network ensured the best one-step-ahead forecasting performances. However, even though the MLP did not perform well compared to the other models, with increasing steps ahead, this behavior was reversed. Its performance for six steps ahead, for example, was better than those of AR, ARIMA, and HW.

The ELM performed better in all scenarios and horizons than the other models. It is important to note that the errors for MLP, ELM, RBF, and HW increased when the forecasting horizon grew. Interestingly, for the AR and ARIMA models, the error for three steps ahead was smaller than for six steps ahead.

As expected, linear models provided worse results compared to ANN due to the generalization capability of the latter to map the inputs in the desired output. As a result, ANNs predictions are well-adjusted to the actual values. Another point to comment on is that the ELM also requests a smaller computational effort, spending less processing time than other networks [27,36–42]. This occurs because its hidden layer remains untuned, and the output layer is adjusted based on a deterministic method during the training phase.

Despite the linear models being simpler and faster to be implemented and executed, they did not present good results in relation to neural approaches. Just in a few cases, they may exhibit equivalent performances.

In addition, it is perceived that when the explanatory variables are included, all ANNs increase their predictive capability for all the steps ahead. In this case, the RBF and ELM occupied the first and second positions in almost all cases, respectively. This behavior indicates that the use of ELECTRE I for variable selection was effective in improving the overall performance of the models [43,44]. Although we addressed up to eighteen exogenous variables, only two were selected by ELECTRE I when considering the five criteria and their degrees of importance.

Finally, although positive results have been presented for the Paraná state, these same variables can be used in similar contexts and may result in more accurate predictions. In addition to the characteristics of the data, each application site has other characteristics that experts who understand the electricity demand and know the application's place would need to consider.

6. Conclusions

The forecasting process for electricity demand is essential to plan the energy distribution and to avoid waste or scarcity of this resource. In the literature, different authors have proposed solutions with linear models, also adding other independent predicting variables intuitively. The independent variables are often chosen based on correlation analysis or on the authors' knowledge, which does not guarantee that all explaining variables are selected. Different criteria are related in this process, and correlation does not necessarily recognize the impact of these variables.

It is crucial to include decision criteria involved in order to do the predictive process more efficient. Insufficient data, correlation with energy, and consumption patterns by sector, calendar, and region are the most important criteria to help the selection of independent variables since these also can be validated by the decision-makers.

This investigation involved eighteen exogenous variables listed in the literature and five criteria with different degrees of importance. As a relatively large problem, the multi-criteria approach has advantages and helps the conducting for more assertive decision-making, using the ELECTRE I method to select variables that influence electricity demand. There were no indications in the literature review of multi-criteria methods especially used for this purpose. Because of this, this work contributes to forecasts of electricity demand modeling, initially in the selection of alternative inputs, such as temperature and evaporation, which influence the demand. Subsequently, the consideration of univariate monthly time series of electricity consumption to perform forecasts with linear models HW, AR, and ARIMA; and neural models MLP, RBF, and ELM. After selecting the independent variables, the forecasts were made considering the period between January 2018 and June 2019 with one, three, and six horizons ahead.

Additionally, the comparison regarding forecasting electricity demand models, as performed in this work, contributes to enlarging the theory, mainly showing what the potential integrations between models can be made, allied with the multi-criteria analysis. In general, the ANN presented the lowest errors for all horizons, except for MLP one step ahead which showed the worst performance. However, increasing the horizon, MLP behaved better than the linear models for three and six steps ahead. The higher performance of the ANN compared to linear models was already expected because they are nonlinear mappers with a generalization capability. In addition, RBF and ELM request a lower computational cost compared to MLP.

After performing the predictions considering only the univariate time series, we added as input for the predictive model's data of the variables selected by ELECTRE I. With the addition of these inputs, there was an improvement in the performance of all ANNs considering all horizons. Once again, the RBF and ELM networks occupied the first and second positions, considering the overall performances.

Future investigations should be developed using other neural approaches or combination models, such as ensembles. Moreover, other databases should be applied. In addition, the selected input variables can be investigated considering their influence on the forecasting horizon separately.

Author Contributions: Conceptualization, C.D., H.V.S., F.T.; methodology, C.D., H.V.S., F.T.; software, C.D., M.L., J.L.F.d.S.; validation L.H.B., C.D., H.V.S., F.T.; formal analysis, A.C., H.V.S., F.T.; investigation, C.D., M.L., J.L.F.d.S., L.H.B.; resources: C.D.; writing original draft preparation, C.D., H.V.S., F.T.; writing, review and editing, J.L.F.d.S., A.C.; visualization, A.C.; supervision, A.C. and H.V.S.; project administration, F.T; funding acquisition, M.L., H.V.S., A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the support of Brazilian agencies Coordination for the Improvement of Higher Education Personnel (CAPES)-Financing Code 001, Brazilian National Council for Scientific and Technological Development (CNPq), process number 315298/2020-0, and Araucaria Foundation, process number 51497.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu, L.; Gao, X.; Xiao, Y.; Yang, Y.; Chen, X. Using a novel multi variable grey model to forecast the electricity consumption of Shandong Province in China. *Energy* **2018**, *157*, 327–335. [CrossRef]
2. Pao, H.T. Comparing linear and nonlinear forecasts for Taiwan's electricity consumption. *Energy* **2006**, *31*, 2129–2141. [CrossRef]
3. Sadownik, R.; Barbosa, E.P. Short-term forecasting of industrial electricity consumption in Brazil. *J. Forecast.* **1999**, *18*, 215–224. [CrossRef]
4. Darbellay, G.A.; Slama, M. Forecasting the short-term demand for electricity: Do neural networks stand a better chance? *Int. J. Forecast.* **2000**, *16*, 71–83. [CrossRef]
5. Cui, H.; Peng, X. Short-term city electric load forecasting with considering temperature effects: An improved ARIMAX model. *Math. Probl. Eng.* **2015**, *2015*, 589374. [CrossRef]
6. Torrini, F.C.; Souza, R.C.; Oliveira, F.L.C.; Pessanha, J.F.M. Long term electricity consumption forecast in Brazil: A fuzzy logic approach. *Socio-Econom. Plan. Sci.* **2016**, *54*, 18–27. [CrossRef]
7. Wang, J.; Zhang, J.; Nie, J. An improved artificial colony algorithm model for forecasting Chinese electricity consumption and analyzing effect mechanism. *Math. Probl. Eng.* **2016**, *2016*, 8496971. [CrossRef]
8. MME, Ministry of Mines and Energy. Statistics of the Brazilian Energy Sector. Available online: http://www.mme.gov.br/ (accessed on 20 February 2019).
9. CIA, Central Intelligence Agency. Electricity, Global Consumption Map. Available online: https://www.cia.gov/library/publications/the-world-factbook/ (accessed on 20 February 2019).
10. IEA, International Energy Agency. Electricity Statistics. Available online: https://www.iea.org/statistics/electricity/ (accessed on 20 February 2019).
11. Chujai, P.; Kerdprasop, N.; Kerdprasop, K. Time series analysis of household electric consumption with ARIMA and ARMA models. In Proceedings of the International MultiConference of Engineers and Computer Scientists 2013, Hong Kong, 13–15 March 2013.
12. Tutun, S.; Chou, C.A.; Canylmaz, E. A new forecasting framework for volatile behavior in net electricity consumption: A case study in Turkey. *Energy* **2015**, *93*, 2406–2422. [CrossRef]
13. Adam, N.R.B.; Elahee, M.K.; Dauhoo, M.Z. Forecasting of peak electricity demand in Mauritius using the non-homogeneous Gompertz diffusion process. *Energy* **2011**, *36*, 6763–6769. [CrossRef]
14. Hernández, L.; Baladrón, C.; Aguiar, J.M.; Calavia, L.; Carro, B.; Esguevillas, A.S.; García, P.; Lloret, J. Experimental analysis of the input variables' relevance to forecast next day's aggregated electric demand using neural networks. *Energies* **2013**, *6*, 2927–2948. [CrossRef]
15. Vu, D.H.; Muttaqi, K.M.; Agalgaonkar, A.P.E. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Appl. Energy* **2015**, *140*, 385–394. [CrossRef]

16. Suganthi, L.; Samuel, A.A. Modelling and forecasting energy consumption in India: Influence of socioeconomic variables. *Energy Sources B Econ. Plan. Policy* **2016**, *11*, 404–411. [CrossRef]
17. Mohammed, N.A. Modelling of unsuppressed electrical demand forecasting in Iraq for long term. *Energy* **2018**, *162*, 354–363. [CrossRef]
18. Sahay, K.B.; Sahu, S.; Singh, P. Short-term load forecasting of Toronto Canada by using different ANN algorithms. In Proceedings of the IEEE 6th International Conference on Power Systems (ICPS), New Delhi, India, 4–6 March 2016; pp. 1–6.
19. Ahmad, T.; Chen, H. Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment. *Energy* **2018**, *160*, 1008–1020. [CrossRef]
20. Kim, M.; Choi, W.; Jeon, Y.; Liu, L. A hybrid neural network model for power demand forecasting. *Energies* **2019**, *12*, 931. [CrossRef]
21. Albuquerque, P.C.; Cajueiro, D.O.; Rossi, M.D. Machine learning models for forecasting power electricity consumption using a high dimensional dataset. *Expert Syst. Appl.* **2022**, *187*, 115917. [CrossRef]
22. Mehmood, F.; Ghani, M.U.; Ghafoor, H.; Shahzadi, R.; Asim, M.N.; Mahmood, W. EGD-SNet: A computational search engine for predicting an end-to-end machine learning pipeline for Energy Generation & Demand Forecasting. *Appl. Energy* **2022**, *324*, 119754.
23. Raza, M.A.; Khatri, K.L.; Israr, A.; Haque, M.I.U.; Ahmed, M.; Rafique, K.; Saand, A.S. Energy demand and production forecasting in Pakistan. *Energy Strategy Rev.* **2022**, *39*, 100788. [CrossRef]
24. Rick, R.; Berton, L. Energy forecasting model based on CNN-LSTM-AE for many time series with unequal lengths. *Eng. Appl. Artif. Intell.* **2022**, *113*, 104998. [CrossRef]
25. Rao, C.; Zhang, Y.; Wen, J.; Xiao, X.; Oh, M. Energy demand forecasting in China: A support vector regression-compositional data second exponential smoothing model. *Energy* **2023**, *263*, 125955. [CrossRef]
26. Figueira, J.; Mousseau, V.; Roy, B. ELECTRE methods. In *International Series in Operations Research & Management Science book series (ISOR, Volume 78)*; Springer: Berlin/Heidelberg, Germany, 2004; Chapter 1. [CrossRef]
27. Belotti, J.; Siqueira, H.; Araujo, L.; Stevan, S.L., Jr.; de Mattos Neto, P.S.; Marinho, M.H.; de Oliveira, J.F.L.; Usberti, F.; de Almeida Leone Filho, M.; Converti, A.; et al. Neural-based ensembles and unorganized machines to predict streamflow series from hydroelectric plants. *Energies* **2020**, *13*, 4769. [CrossRef]
28. Belotti, J.T.; Mendes Junior, J.J.A.; Luna, I.; Asano, P.T.L.; de Souza Tadano, Y.; Alves, T.A.; Martins, M.S.R.; Usberti, F.L.; Stevan, S.L., Jr.; Trojan, F.; et al. Linear Models Applied to Monthly Seasonal Streamflow Series Prediction. *Learn. Nonlinear Model.* **2022**, *20*, 47–61. [CrossRef]
29. Box, G.; Jenkins, G.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis, Forecasting and Control*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2015.
30. Haykin, S. *Neural Networks and Learning Machines*, 3rd ed.; Pearson Education: Bengaluru, India, 2009.
31. Montgomery, D.; Johnson, L.; Gardiner, J. *Forecasting and Time Series Analysis*; McGraw-Hill: New York, NY, USA, 1990.
32. EPE, Energy Research Company. Open Publications/Data. Ministry of Mines and Energy, 2019, Rio de Janeiro, Brazil. Available online: http://www.epe.gov.br/pt/publicacoes-dados-abertos (accessed on 20 February 2019).
33. Judge, G.G.; .Hill, R.C.; Griffiths, E.W.; Lutkepohl, H.; Lee, T. *Introduction to the Theory and Practice of Econometrics*, 2nd ed.; John Wiley& Sons: New York, NY, USA, 1988.
34. Saaty, T.L. *The Analytic Hierarchy Process*; McGraw-Hill International: New York, NY, USA, 1980.
35. Figueira, J.; Mousseau, V.; Roy, B. Electre Methods. In *Multiple Criteria Decision Analysis: State of the Art Surveys*; Figueira, J., Greco, S., Ehrgott, M., Eds.; Springer: New York, NY, USA, 2005; pp. 133–162.
36. Silva, N.; Tay, I.; Okida, S.; Stevan, S.L., Jr.; Siqueira, H. Neural networks for predicting prices of sugarcane derivatives. *Sugar Tech.* **2018**, *21*, 514–523. [CrossRef]
37. Kachba, Y.; Chiroli, D.M.D.G.; Belotti, J.T.; Antonini Alves, T.; de Souza Tadano, Y.; Siqueira, H. Artificial neural networks to estimate the influence of vehicular emission variables on morbidity and mortality in the largest metropolis in South America. *Sustainability* **2020**, *12*, 2621. [CrossRef]
38. Siqueira, H.; Boccato, L.; Attux, R. Echo state networks for seasonal streamflow series forecasting. In *International Conference on Intelligent Data Engineering and Automated Learning, Lecture Notes in Computer Science*; Yin, H., Costa, J.A.F., Barreto, G., Eds.; Springer: Berlin, Heidelberg, Germany, 2012; Volume 7435, pp. 226–236. [CrossRef]
39. de Souza Tadano, Y.; Siqueira, H.V.; Alves, T.A. Unorganized machines to predict hospital admissions for respiratory diseases. In Proceedings of the IEEE Latin American Conference on Computational Intelligence (LA-CCI), Cartagena, Colombia, 27 March 2017. [CrossRef]
40. Tadano, Y.S.; Potgieter-Vermaak, S.; Kachba, Y.R.; Chiroli, D.M.; Casacio, L.; Santos-Silva, J.C.; Moreira, C.A.B.; Machado, V.; Alves, T.A.; Siqueira, H.; et al. Dynamic model to predict the association between air quality, COVID-19 cases, and level of lockdown. *Environ. Pollut.* **2021**, *268*, 115920. [CrossRef] [PubMed]
41. Neto, P.S.D.M.; Firmino, P.R.A.; Siqueira, H.; Tadano, Y.D.S.; Alves, T.A.; de Oliveira, J.F.L.; da Nobrega Marinho, M.H.; Madeiro, F. Neural-based ensembles for particulate matter forecasting. *IEEE Access* **2021**, *9*, 14470–14490. [CrossRef]
42. Siqueira, H.; Macedo, M.; Tadano, Y.D.S.; Alves, T.A.; Stevan, S.L., Jr.; Oliveira, D.S., Jr.; Marinho, M.H.; de Mattos Neto, P.S.; de Oliveira, J.F.L.; Luna, I.; et al. Selection of temporal lags for predicting riverflow series from hydroelectric plants using variable selection methods. *Energies* **2020**, *13*, 4236. [CrossRef]

43. Lizot, M.; Trojan, F.; Afonso, P. Combining total cost of ownership and multi-criteria decision analysis to improve cost management in family farming. *Agriculture* **2021**, *11*, 139. [CrossRef]
44. Lizot, M.; Goffi, A.S.; Thesari, S.S.; Trojan, F.; Afonso, P.S.L.P.; Ferreira, P.F.V. Multi-criteria methodology for selection of wastewater treatment systems with economic, social, technical and environmental aspects. *Environ. Dev. Sustain.* **2021**, *23*, 9827–9851. [CrossRef]