Expert Systems With Applications 187 (2022) 115917

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Machine learning models for forecasting power electricity consumption using a high dimensional dataset☆

Pedro C. Albuquerque, Daniel O. Cajueiro *, Marina D.C. Rossi

Department of Economics - Universidade de Brasília (UnB), Campus Universitário Darcy Ribeiro, Brasília-DF, CEP 70910-900, Brazil

ARTICLE INFO

Keywords:
Power electricity consumption
Machine learning
Forecast

ABSTRACT

We use regularized machine learning models to forecast Brazilian power electricity consumption for short and medium terms. We compare our models to benchmark specifications such as Random Walk and Autoregressive Integrated Moving Average. Our results show that machine learning methods, especially Random Forest and Lasso Lars, give more accurate forecasts for all horizons. Random Forest and Lasso Lars managed to keep up with the trend and the seasonality for various time horizons. The gain in predicting PEC using machine learning models relative to the benchmarks is considerably higher for the very short-term. Machine learning variable selection further shows that lagged consumption values are extremely important for very short-term forecasting due to the series high autocorrelation. Other variables such as weather and calendar variables are important for longer time horizons.

1. Introduction

Forecasting economic activity is very important in promoting economic growth. Accurate forecasts allow the government to better organize its budget and help economic agents make decisions because it aids them in setting expectations. We usually consider GDP as the main representation of economic activity, but it is released only quarterly and this data is usually adjusted afterwards. Thus, it is interesting to find alternative methods to forecast short-term economic activity. This helps surpass the problems associated with the usual measurements, such as the low frequency of the available data.

Power electricity consumption (PEC) is a good proxy for economic activity. It is released hourly and it has a strong causal relation with economic activity (Maza & Villaverde, 2007). Developed economies require a large amount of electricity to satisfy the needs of industries, households, agriculture, and government. In this sense, the literature views satellite data on lights at night as a strong and useful indicator of economic activity (Bundervoet et al., 2015; Henderson et al., 2012; NASA, 2000). The Federal Reserve Board's monthly index of industrial production (until 2005) is partially based on a survey that measures delivered electricity.

Moreover, forecasting PEC is crucial in aiding the electrical power industry in their planning and is also crucial in guaranteeing the proper operation of electrical power systems. Since this type of energy cannot be stored, accurate forecasts help ensure the balance between power electricity supply and demand. Consequently, these forecasts play an important role in future decisions on energy management and in reducing operational and maintenance costs (Almeshaiei & Soltan, 2011; Bere et al., 2018; Fan & Hyndman, 2011; Goude & Pierrot, 2011).

In this paper we accurately forecast Brazilian PEC for 1 day and up to 3 months ahead. We use traditional econometric models as well as machine learning (ML) techniques in order to get precise forecasts. Our results show that ML models outperform traditional ones. Additionally, our results indicate that the gain in forecasting with ML models relative to regular ones is larger for shorter horizons. For a 1-day forecast, our best model has a forecast error that is approximately 2 times smaller than the best benchmarks.

Our work relates to the literature that applies ML techniques to forecast variables in the energy market (G.Creamer et al., 2019). We may classify them according to the following criteria: (1) the predicted variable can be power electricity consumption, smart grid load, gas demand, transport energy demand or coal demand; (2) the data types used can be regular data,[1] multidimensional data or past variables only; (3) the frequency of the data used can be high frequency[2] or low frequency data[3]; (4) the models that are used to predict can be regularized models (Raihanian Mashhadi & Behdad, 2018), SVM (Anderson et al.,

☆ The datasets and codes used to generate all the results of the current study are available in the Zenodo repository (Albuquerque et al. 2021).

* Corresponding author.
E-mail addresses: pedrocampeloa@gmail.com (P.C. Albuquerque), danielcajueiro@gmail.com (D.O. Cajueiro), marinadcrossi@gmail.com (M.D.C. Rossi).

[1] When the number of observations is greater than or equal to the number of predictors.
[2] Hourly, daily or weekly frequency.
[3] Monthly, quarterly or yearly frequency.

https://doi.org/10.1016/j.eswa.2021.115917
Received 18 September 2020; Received in revised form 5 August 2021; Accepted 12 September 2021
Available online 24 September 2021

2011), Random Forest (Collado & Creamer, 2016), Adaboost (Huang et al., 2018), or ANN (Debnath & Mourshed, 2018). Furthermore, our work also connects to the recent literature that uses models to predict the behavior of economic variables using high dimensional data, such as in Freitas et al. (2016), Garcia et al. (2017), and Medeiros and Mendes (2016).

Our work contributes to the literature as we build a multidimensional and high frequency database to explore the PEC forecasting field using regularized models. We innovate by predicting PEC using a large set of weather, calendar, and economic variables. The use of regularized models here is essential to prevent the model from having a good forecasting capacity only in-sample and not being able to generalize to out-of-sample data. Therefore, the models we use in this work must be able to deal with a large set of variables, to select only the relevant ones, or to reduce the effect of those that are not important in the model. Our results show that these ML models can consistently outperform benchmark models.

PEC is a highly seasonal and cyclical variable. It varies, for example, with the day of the week and season of the year, and it is also highly autocorrelated. Hence, it is common to forecast short-term PEC by using only lagged variables via ARIMA models (Almeshaiei & Soltan, 2011; Calado et al., 2014; Huang & Shih, 2003). Calendar variables are also important in capturing the series' seasonality. Bere et al. (2018) and Fan and Hyndman (2011) include calendar variables in order to catch the complex nonlinear relationship between electricity demand and its driving forces.

Weather variables are relevant to predict a variety of economic variables. Dell et al. (2014) show that shocks in temperature affect many economic outcomes, such as economic growth, industrial output, and energy demand. The authors also note that the effect of weather on economic variables is more pronounced in developing countries, since they have less access to technology. In Brazil, in particular, weather variables also directly affect energy production. Hydroelectric plants are responsible for about 68% of all electricity consumed (EPE, 2017). This type of energy is highly seasonal and influenced by weather factors (Adegbehin et al., 2016), such as river flow, rain incidence, and increase in temperature. In addition, wind energy accounts for about 5.4% of Brazil's electricity production and is on a great rise, growing around 33% in 2016 (EPE, 2017). Thus, including weather variables helps to capture the highly seasonality of PEC for short-term forecasts (Bere et al., 2018; Fan & Hyndman, 2011).

Electrical energy price variables and other economic variables capture the economic environment, which is a potential determinant of PEC. This set of variables follows the trend of PEC. El-Shazly (2013) include economic variables to build a dynamic econometric model to forecast PEC for Egypt and produce reliable ex-post forecasts.

We consider an approach based on high dimensional data. We build a database with more than 1500 variables, including lagged demand, calendar variables, weather variables, electrical energy price variables, and other economic variables. We estimate our models for 6 different forecasts horizons using rolling windows.

The outline of the paper is as follows. In Section 2 we detail our dataset and analyze each set of variables that we use. In Section 3 we present the models, the specifications used for forecasting and the procedure adopted to choose the parameters. We then show our general results and examine the best models in Section 4. Lastly, in Section 5, we present a brief conclusion concerning our findings.

2. Data

We use Brazilian data from various sources. Our dataset starts on February 1, 2017, when PEC hourly data starts to be disclosed, and ends on July 31, 2018. Although PEC data are released hourly, the data for most of the other explanatory variables are at a daily or monthly frequency. Therefore, we use data at a daily frequency. Our sample consists of 546 days (one and a half year). In order to run a ML procedure, we split this data into training and test (holdout) sets.

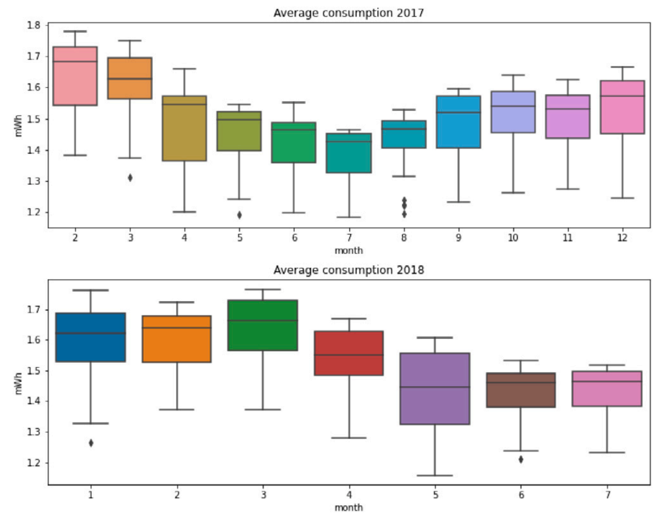

Fig. 1. Seasonal plot.

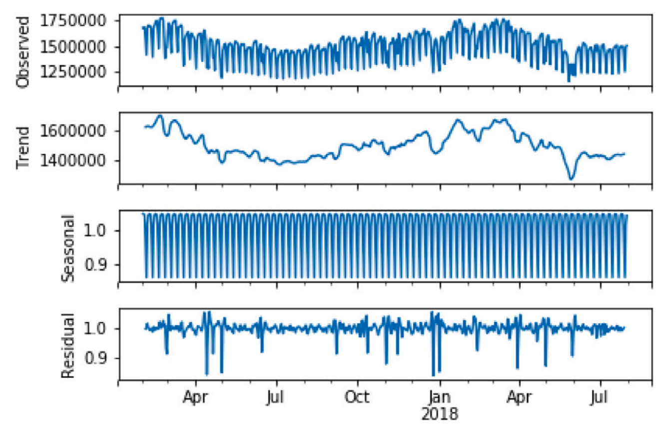

Fig. 2. Seasonal plot.

Besides the data of PEC and its lagged values, as we mention in Section 1, we also use calendar variables, weather variables, a price of energy dataset, and several economic variables. Since we use a large number of variables, we present the summary statistics and the correlation matrix of this set of variables in our Zenodo repository (Albuquerque et al., 2021).

We normalize all the variables in order to allow for the proper comparison of the importance of these variables in forecasting PEC.

2.1. Power electricity consumption

We use hourly energy load by subsystem (mWh) for Brazil, which is released by the National Electricity System Operator (ONS), as the measure of PEC. Fig. 1, using boxplots, shows the average, concentration, and dispersion of PEC over the analyzed period, while Fig. 2 shows the trend, seasonality, and residual of this variable in Brazil. We can see from these figures that PEC is a highly seasonal variable.

From Fig. 2 we can also see the effect of the truckers strike on PEC. At the end of May 2018, truck drivers blocked roads across Brazil demanding a reduction in the price of diesel, which had risen by more than 50% in the previous 12 months. With trucks stalled, partially blocking roads, fuels were no longer delivered to several gas stations. Other activities that required raw materials and essential products, such

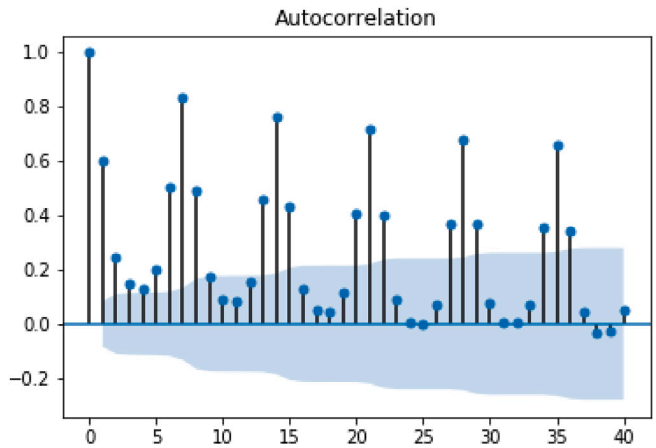

Fig. 3. Autocorrelation plot.

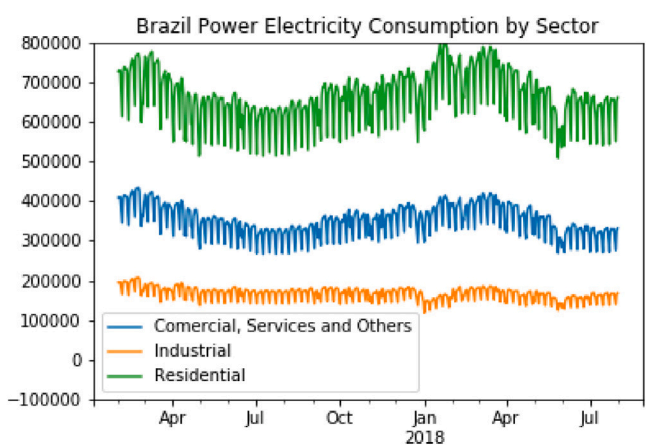

Fig. 4. PEC by sector.

as food, also ran out of supplies (BBC, 2018; Times, 2018). We note that the level of PEC fell by about 30% during the truckers strike in Brazil. This episode further exemplifies the strong causal relationship between economic activity and electrical energy.

Fig. 3 shows how PEC is correlated with its past values. There is a strong correlation with its value from the previous day. However, the correlation is even stronger with the value seven days before, corroborating the cyclical behavior of PEC present in Fig. 2. This figure indicates the importance of including lagged PEC values in the forecasting models.

We obtain the hourly energy load by subsystem (mWh) for Brazil and by Region from the ONS website.[4] We also get daily PEC data separated by consumption class for each region from the Brazilian Electricity Regulatory Agency (ANEEL) website.[5] Consumption classes includes PEC variables divided into groups such as industrial, commercial and services, own consumption, residential, rural, street lighting, public service, and others.[6]

Separating by consumption class is important in understanding how the consumption of electricity is distributed across various economic activities or classes. In our sample, most of the electricity consumption is concentrated in commercial and services (23%), industrial (12%), and residential consumption (44%). The first two are directly linked

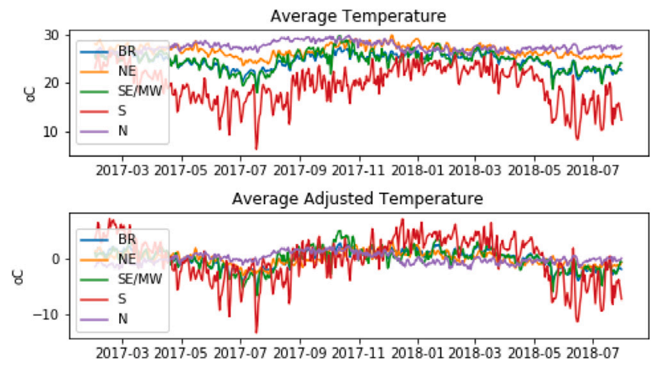

Fig. 5. Average temperature.

to economic activity, while the third one is indirectly related. Fig. 4 plots the series for these variables and we can see that together they represent around 80% of all PEC in Brazil.

We have a total of 185 PEC variables after taking into account the classification by consumption class and region.

2.2. Calendar variables

We use calendar information as a set of variables. This set includes day of the week, day of the month, month, season of the year, year, and a dummy for holidays, totaling 6 calendar variables.

2.3. Weather variables

We obtain this set of variables from the Brazilian Institute of Meteorology (INMET) website,[7] which releases the historical series issued by stations scattered throughout Brazil. This set of variables includes air nebulosity, atmospheric pressure (mbar), dry bulb temperature (°C), humidity bulb temperature (°C), relative humidity (%), wind direction, and wind speed (m/s) disclosed for 9 a.m., 3 p.m., and 9 p.m. for each station.

The most important weather variable for forecasting PEC is the dry bulb temperature, which is also a highly seasonal variable with a large variance. Inspired by Boldin and Wright (2015), we adjust the temperature by subtracting the average of the entire training set, if the day belongs to the training set; and over the entire test set, if the day belongs to the test sample. Adjusted temperature controls both for the seasonal effect and for the variance of temperature. This deviation from temperature's average is also able to capture information that is no longer in lagged values. This methodology is aligned with the World Meteorological Organization guidelines for the calculation of climate norms (ORGANIZATION, 2017).

Figs. 5 and 6 show how adjusting the temperature partly controls for the large variance and seasonality. When comparing Figs. 2 and 6 we observe that PEC and temperature have positive correlation. This is a further indication that weather variables are potentially important in predicting PEC.

We have a total of 180 weather variables after aggregating the data by region and including the adjusted temperature.

2.4. Price of electrical energy

We retrieve the price dataset from Electrical Energy National Agency (ANEEL) website[8] and divide by consumption class for each region. However, price of electricity is a monthly variable and as our variable

[4] The ONS website is http://sdro.ons.org.br/SDRO/DIARIO/index.htm.
[5] The ANEEL website is http://www.aneel.gov.br.
[6] We list all consumption classes in Appendix A.

[7] The INMET website is http://www.inmet.gov.br.
[8] The ANEEL website is http://www.aneel.gov.br.

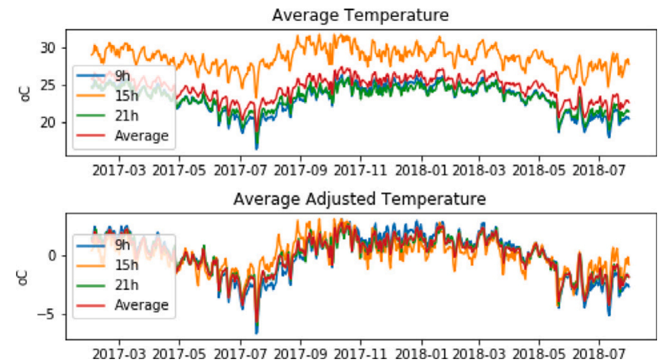

Fig. 6. Adjusted average temperature.

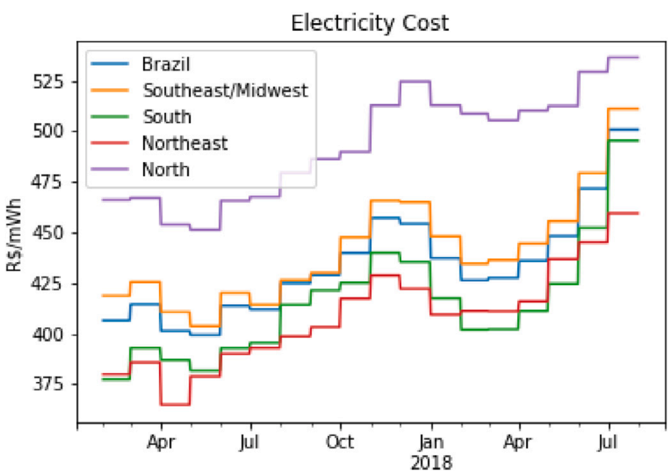

Fig. 7. Average price by region.

of interest has daily frequency, we set the same price for all days within a month. Fig. 7 shows price of power electricity for Brazil and its regions.

After classifying the price by region and consumption class, we have a total of 60 price variables.

2.5. Economic variables

We use a database from the Brazilian Central Bank (BCB) website[9] that includes several price indexes, traffic of vehicles, unemployment, wage, industrial production variables, monetary variables, consultations with credit bureaus, Brazilian Bank for Economic and Social Development (BNDES) disbursements, and several others. We include a wide list of economic variables, some at a daily and others at a monthly frequency. For those at a monthly frequency, we also set the same value for all days within a month.

We have a total of 79 economic variables. After merging all sets of explanatory variables listed in this section, we have a total of 510 variables.[10]

3. Methods

In this section, we present the methods used to forecast PEC. This section is divided in two parts. We first present our set of models and then describe the procedure we use to choose the hyper-parameters and parameters and to valuate these models.

3.1. Models

Our benchmark models, ARIMA and Random Walk, use only lagged PEC. We build a high dimensional dataset with all candidate variables. We compare these benchmark models with the regularized "industrial models"[11] in terms of the ability to forecast the PEC dynamics.[12] Since we work with a database with a huge number of variables, we need to use models that are not very computationally expensive.[13] We expect these models to perform better than traditional ones when working with big data. We detail all the models in the following subsections.

3.1.1. ARIMA

Autoregressive integrated moving average (ARIMA) uses just lagged values of PEC. The ARIMA (p,i,q) equation can be represented by:

$$y_t^* = \gamma + \Phi_0 + \Phi_1 y_{t-1}^* + \cdots + \Phi_p y_{t-p}^* + u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q},$$

in which y_t is the PEC at time t, $y_t^* = y_t - y_{t-1}$, and u_t is an error term. We use Augmented Dickey–Fuller test and find that our series has an unit root (i.e., $i = 1$) and therefore we use its first difference in the ARIMA.

3.1.2. Random walk

Random Walk (without drift) is a process in which the current value of the variable is the sum of its past value and an error term. That is, the best forecast for k periods ahead is the value available today.

Random Walk without drift can outperform the Random Walk with drift in terms of the RMSE (Burns & Moosa, 2016). The model can be represented by:

$$y_t = y_{t-k} + u_t,$$

in which y_t is PEC at time t and u_t is an error term. This model is good in capturing the autocorrelation in a series, since the last available value is probably correlated with the forecast. Thus, Random Walk is commonly used as a benchmark in forecasting.

3.1.3. Lasso

Although high-dimensional databases are available in many empirical settings, in most cases only a small number of variables actually affect the dependent variable. In these cases, a sparse representation is more adequate and shrinkage models become particularly useful since they vanish coefficients associated with the variables that have little or no effect on the dependent variable (Horowitz, 2015).

The least absolute shrinkage and selection operator (Lasso) minimizes the mean squared error (MSE), just as OLS. However, Lasso also has a shrinkage coefficient, λ, that forces irrelevant variables to

[9] The Brazilian Central Bank's website is https://www.bcb.gov.br/estatisticas/indicadoresconsolidados.

[10] We list all variables in Appendix A.

[11] We use the term "industrial models" to refer to models that we can choose the hyperparameters based on a simple cross validation mechanism, are well known, and may be found in different libraries of different programming languages.

[12] Another interesting class of machine learning models are the so-called Deep Learning models, such as the Recurrent Neural Networks. These models stand out when the dataset is a long time series, since we need a large number of examples to train them (Baraniuk et al., 2020; Sejnowski, 2020), which is not our case. Furthermore, these models present a much higher computational cost when compared to the models we use. Therefore, we do not consider these models in our paper.

[13] It is worth mentioning that a complete simulation including all horizons takes less than a second for the Random Walk, which is the fastest model and it does not require any kind of estimation or validation, and about 8 h for the Random Forest, which is the slowest model. A table with the time required to run each model for each time horizon is available in the Zenodo repository (Albuquerque et al., 2021). All the simulations were developed in Python and we use a SSD-equipped PC with a 3.8 GHz Octa-Core and with 16G of RAM.

zero (Santosa & Symes, 1986; Tibshirani, 1996). The parameters are determined by:

$$\hat{\beta} = \underset{\hat{\beta}}{argmin} \left[\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1 \right].$$

Lasso can be seen as a generic OLS model because if $\lambda = 0$, then $\hat{\beta}$ and $\hat{\beta}_{OLS}$ are the same. On the other hand, if the shrinkage coefficient is equal to some $\lambda_c > 0$ large enough, then all the variables are considered irrelevant and all β_i's are equal to zero. Generically we have, $\lambda_c \geq \lambda \geq \lambda_{OLS} = 0$.

3.1.4. Lars

Least angle regression (Lars) is also a good model to use when working with high dimensional data because it provides a method to evaluate which variables to include. The main idea of this algorithm is to start with all values of β_i equal to zero and to increase all the coefficients associated with the x_i that is most correlated with y. In other words, the algorithm increases the values of the parameters in an equiangular direction to each one's correlation with the residual.

Lars algorithm and its variations work gracefully for the case in which there are many more variables than observations. Lars is easy to modify to produce efficient algorithms, like Lasso Lars, and is useful in cross-validation or similar attempts (Efron et al., 2004).

3.1.5. Lasso Lars

A simple modification in the Lars algorithm can result in the simulation of all Lasso models for all possible λ values (Efron et al., 2004). Thus, in the traditional Lars method, if a coefficient changes the signal, the direction remains the same. However, in the Lasso model, when a coefficient reaches the value 0, it is discarded from the active variable set. For that reason, Lasso Lars algorithm makes a simple modification in the original Lars. That is, if any coefficient becomes zero this variable is discarded and the model recalculates the search direction.

3.1.6. Ridge

Ridge Regression, like Lasso, is a generic version of OLS that also has a shrinkage coefficient λ (Hoerl & Kennard, 1970). However, the difference is that in Ridge the penalization term is squared instead of linear, as in Lasso. Its parameters are determined by:

$$\hat{\beta} = \underset{\hat{\beta}}{argmin} \left[\|Y - X\beta\|_2^2 + \lambda\|\beta\|_2^2 \right].$$

Although this model also reduces overfitting presented in traditional models it differs from Lasso because the penalization factor uses the Euclidean norm. Therefore, it is harder to make the coefficients vanish in the Ridge algorithm and, consequently, it cannot completely eliminate some irrelevant variables. However, one particular characteristic of this model is that it treats correlated variables in a close way.

3.1.7. Elastic Net

Elastic Net is also a regularized regression model that combines the restrictions in Lasso and Ridge Regression (Zou & Hastie, 2005). The parameters are given by:

$$\hat{\beta} = \underset{\hat{\beta}}{argmin} \left[\|Y - X\beta\|_2^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2 \right].$$

Elastic Net becomes a general case of Lasso and Ridge Regression. Therefore, as in the Ridge model, the Elastic Net method makes the loss function strictly convex, forcing it to have a unique minimum.

3.1.8. Random Forest

Random Forest is a machine learning method that combines decision tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees (Breiman, 2001). Briefly, this model grows a forest of trees and lets them vote for the most popular class (if in a classification problem) or averaging the forecasts (if in a regression problem).

Although models based on conventional decision trees reduce the possibility of overfitting by choosing parameters that control their depth and number of leaves, Random Forest also seeks to reduce the overfitting of traditional trees by combining different trees.

3.2. Parameters choice and model valuation

Since PEC is a highly seasonal variable, we consider that its forecast h periods ahead is a function of three blocks of predictors. Each block of predictors consists of all our explanatory variables in a different period of time, which depends on our forecast horizon h. The first block uses the data from the last available date, the second one uses data from the past h days, and the third block uses data from the past $2h$ days. Once we consider all blocks of predictors, we have a total of 1530 candidate variables for 546 observations.

The lag structure means that a forecast for $t+h$ periods ahead uses variables from t, $t-h$ and $t-2h$. That is, in a forecast for 7 days ahead, we use data from today and from the past 7th and 14th days. We test several other lags structures and this one presents the smallest forecast error. We see from Fig. 3 that PEC has a large autocorrelation with the same day of the week for past weeks and this structure builds a database that captures information from this past data. With a nearby lag structure,[14] our database is unable to capture some of the information that explains PEC's seasonality. The lag structure is the first hyper-parameter that we define. Thus, our estimated equation is given by:

$$y_{t+h} = \alpha_0 + \sum_{i=0}^{2} \gamma_i y_{t-(h*i)} + \sum_{i=0}^{2} \beta_i X_{t-(h*i)} + u_{t+h},$$

in which y_t is PEC (mWh) in Brazil at time t, α_0 is a constant term, X_t is a matrix containing all candidate variables, $\boldsymbol{B} = [\gamma, \beta]$ is a vector with all the linear parameters, and u_t is an error term.

We take the initial training and test (holdout) sets roughly equal to 70% and 30% of the sample. After that we use a rolling sample approach to determine the next training and test sets, removing the first observation of the training set and including the next observation in it and removing the first observation of the test set. Since we are working with time series, we use a walk forward cross validation approach to select the hyper-parameters of the model. In this approach, since we cannot use future data to validate past data, each test set is a single observation with corresponding training set formed by the observations that occurred prior to the observation that forms the test set (Hyndman & Athanasopoulos, 2018). Thus, we start with a folder of 60 days and use the observation of the 61th as the test set. In the next step, we use a folder of 61 days and use the observation of the 62th as the test set and so on. We proceed like this until we use the entire training set. We run several hyper-parameters values for each model and we present a summary of these hyper-parameters in Appendix B. We select the hyper-parameters that deliver the lowest root mean square error (RMSE) for each forecast horizon. We use the test set to evaluate the forecast models.

After we select the hyper-parameters, we run the models for the entire training set and obtain each model's parameters ($\boldsymbol{B} = [\gamma, \beta]$). Next, we use these parameters in our test set and obtain the forecasts for

[14] Example of a nearby lag structure: forecast $t+h$ use variables from t, $t-1$ and $t-2$. In this case, a 7 day ahead forecast use data from today, yesterday and the day before yesterday.

Table 1
Models results.

Model rank	Model	MCS	RMSE	MAE	MAPE	R^2
1 day						
1st	Random forest	1	44943	32046	2.17	0.9
2nd	Lasso lars	0	53893	39672	2.66	0.85
3rd	ARIMA	0	55894	38484	2.62	0.84
4th	Elastic net	0	74583	50734	3.42	0.72
7 days						
1st	Random forest	1	77703	52865	3.64	0.67
2nd	Lars	0	78279	56364	3.94	0.66
3rd	Lasso lars	0	78279	56364	3.94	0.66
4th	Random Walk	0	78570	54148	3.7	0.66
15 days						
1st	Random forest	1	90316	62137	4.42	0.47
2nd	Arima	0	97862	70948	4.99	0.38
3rd	Lars	0	123014	89834	6.51	0.02
4th	Lasso lars	0	126624	93079	6.71	−0.04
30 days						
1st	Random forest	1	125554	99320	7.08	−0.33
2nd	Lasso lars	0	141189	118063	8.47	−0.69
3rd	Arima	0	147249	119464	8.62	−0.84
4th	Lars	0	161549	121461	8.88	−1.21
60 days						
1st	Random forest	1	75700	63314	4.55	0.28
2nd	Lasso lars	0	91556	68535	4.81	−0.06
3rd	Lars	0	123430	100117	7.25	−0.92
4th	Arima	0	169528	133701	9.67	−2.62
90 days						
1st	Random forest	1	102613	92413	6.52	−0.31
2nd	Lasso lars	0	101068	77089	5.66	−0.27
3rd	Lars	0	101370	77421	5.68	−0.28
4th	Random Walk	0	138982	115909	8.37	−1.41

the different horizons. We use several metrics to evaluate our models. The main one is the Model Confidence Set (MCS) developed by Hansen et al. (2011). This technique is similar to a confidence interval for a parameter. It is a set of models that is build such that it will contain the best model given a confidence level. We set the confidence level to 5%. This method only identifies the best one. Thus, in order to select the second best model and onward, we use three errors metrics (RMSE, MAE, and MAPE) and one accuracy metric (R^2).

Table 2
Models rank.

Rank	Model		
	Very short-term	Short-term	Medium-term
1st	Random forest	Random forest	Random forest
2nd	Lasso lars	Lasso lars	Lasso lars
3rd	Lars	Arima	Lars
4th	Elastic net	Lars	Arima

This table shows the four best models for each forecast horizon group.

4. Results

We make predictions for 6 different forecast horizons: 1, 7, 15, 30, 60, and 90 days ahead. We divide these horizons into three groups: (1) very short-term forecast group (VSTFG) that includes 1 and 7 days, (2) short-term forecast group (STFG) containing 15 and 30 days and (3) medium-term forecast group (MTFG) including 60 and 90 days.

Table 1 exhibits the four models that present better forecast based on MCS, RMSE, MAPE, MAE, and R^2 for each forecast horizon. These results corroborate what we conjectured, namely, that machine learning models present a much better performance relative to traditional ones.

In Table 2, we classify the five best models according to the results present in Table 1. For all groups, Random Forest is the model that presents the lowest metric errors. Right after, and even surpassing for a few periods and metrics, are the regularized linear models, such as Lasso Lars, Ridge, and Elastic Net.

In order to analyze the gain in predicting PEC using ML models instead of our benchmark models, we normalize the error by dividing the RMSE of the benchmarks by Random Forest's RMSE for each time horizon. We use Random Forest because even though it is not the model with the lowest RMSE for all forecast horizons, it remains consistently at the top of all evaluation metrics across all horizons. Although the ability to forecast PEC depends on the forecast horizons, Random Forest has an error, on average, about 2 times smaller than our benchmark models' errors. In Fig. 8, we present the comparison of the RMSE for the random forest and our benchmark models.

Even when using almost the same variables as benchmark models to predict short-term PEC,[15] ML models have a much better performance.

[15] Benchmark models only use lagged variables to predict and these variables are the main ones for predicting short-term PEC with ML models.

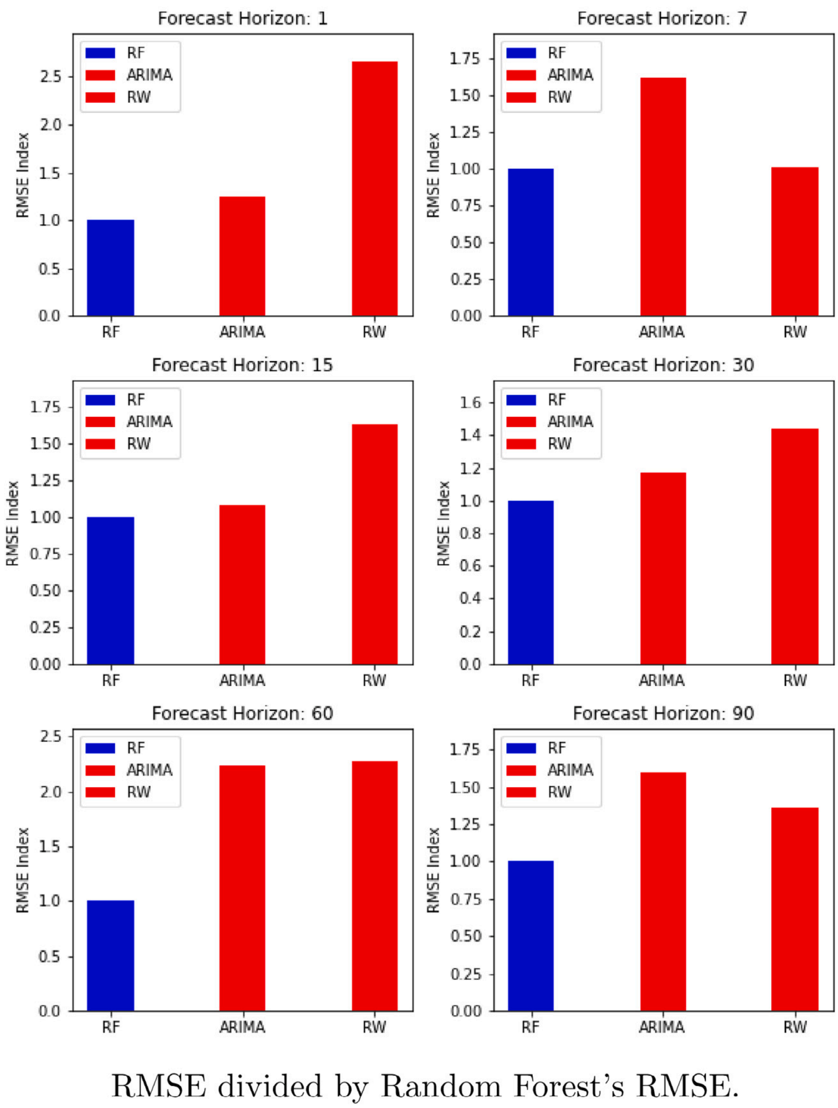

Fig. 8. Relative gain of the Random Forest model.

Benchmark models end up being overfitted and present poor results for nearly all time horizons analyzed. Thus, our four best prediction models are practically always the machine learning ones that better select the relevant variables.

A relevant contribution of our work is to understand the relative importance of the variables in each model for each time horizon. Since our dataset has thousands of variables and we estimate the coefficients of the models for each time window and each time horizon, we provide an aggregate analysis for each group of variables. Additionally, in the Zenodo repository (Albuquerque et al., 2021) we provide a table with a list of averaged values of coefficients of the variables (features) for each time horizon.

We sum the absolute values of the coefficients of all time windows for each set of (normalized) variables described in Section 2 to better analyze what are the most important variables for each model. Table 3 shows this sum for the four best models for each forecast horizon. Comparing the sum among them does not give us much since each model behaves differently in their valuation of the parameters. Also, each set of variables has a different size and the comparison between them also makes no sense. Therefore, we compare the sum of the parameters for the same model and same variable set across different forecast horizons. This comparison is only for the main models for each forecast horizon group. In other words, we analyze how the sum of the coefficients varies for each variable set across each time horizon and, thus, infer the importance of a particular set of variables for each horizon.

Furthermore, Fig. 9 shows for each forecast horizon an index of relative importance of the absolute sum of the coefficients for each class of variable and for the top four models. In order to create this index, we first identify separately the relative importance of each sum of coefficients group for each forecast horizon, that is, calculating:

$$ind_{\sum |var_i|,t} = \frac{\sum |var_i|_t - \min_k \sum |var_i|_k}{\max_k \sum |var_i|_k - \min_k \sum |var_i|_k}.$$

Then, we normalize all groups of variables in order to make the sum of the groups equals to one for each forecast horizon, obtaining:

$$\overline{ind}_{\sum |var_i|,t} = \frac{ind_{\sum |var_i|,t}}{\sum_j ind_{\sum |var_j|,t}},$$

in which $var_i \in \{\text{PEC, calendar, weather, price, economic}\}$ is the group of variables and $t \in \{1, 7, 15, 30, 60, 90\}$ is the forecasting horizon. If we analyze Random Forest, for instance, the relationship between the predictors and the target is clear. PEC variables are extremely important in the VSTFG, due to the large serial correlation of the series. This significance diminishes as the forecast horizon increases because the autocorrelation decreases. The importance of Calendar and Weather Variables rise over the forecast horizon.

Table 3
Parameters relevance.

VSTFG								
Parameters	1 day				7 days			
	Random forest	Lasso lars	Elastic net	Lars	Random forest	Lars	Lasso lars	Lasso
∑ |PEC variables|	0.77	572.99	9666.98	308.66	0.9	246.9	246.9	15842.85
∑ |Calendar variables|	0.17	71053.52	59484.61	22081.42	0	1342.7	1342.7	157329.87
∑ |Weather variables|	0.06	90482.3	536238.51	37669.68	0.08	2040.3	2040.3	1757849.02
∑ |Price variables|	0	351.78	9689.83	278.39	0	48.35	48.35	25404.76
∑ |Economic variables|	0.01	33000.7	162791.92	20622.72	0.01	3325.35	3325.35	752702.22
STFG								
Parameters	15 days				30 days			
	Random forest	Lars	Lasso lars	Lasso	Random forest	Lasso lars	Lars	Lasso
∑ |PEC variables|	0.46	146.98	274.98	20928.32	0.33	444.35	542.68	14176
∑ |Calendar variables|	0.27	7367.29	37117.11	211425.78	0.4	70639.35	36503.57	209475
∑ |Weather variables|	0.17	15419.55	72530.44	2393313.2	0.17	128534.03	67804.31	1478187
∑ |Price variables|	0.02	444.44	555.2	29228.74	0.02	654.75	9686.96	16040
∑ |Economic variables|	0.08	25084.22	50702.26	1061607.17	0.08	120282.78	128119.6	379589
MTFG								
Parameters	60 days				90 days			
	Random forest	Lasso lars	Larsd	Lasso	Random forest	Lasso lars	Lars	Lasso
∑ |PEC variables|	0.2	478.2	335.02	16711.57	0.24	387.77	350.97	10057.65
∑ |Calendar variables|	0.49	63160.85	24734.3	265304.37	0.51	3703.82	1358.13	117510.03
∑ |Weather variables|	0.29	221453.57	74408.94	1621278.69	0.19	24829.05	13402.54	1651818.16
∑ |Price variables|	0	449.12	880.26	23926.31	0	290.65	275.62	23675.84
∑ |Economic variables|	0.02	206157.1	116752.21	274965.1	0.06	43749.44	35739.98	520342.06

Sum of the absolute values of the coefficients for each set of variables and forecast horizon.

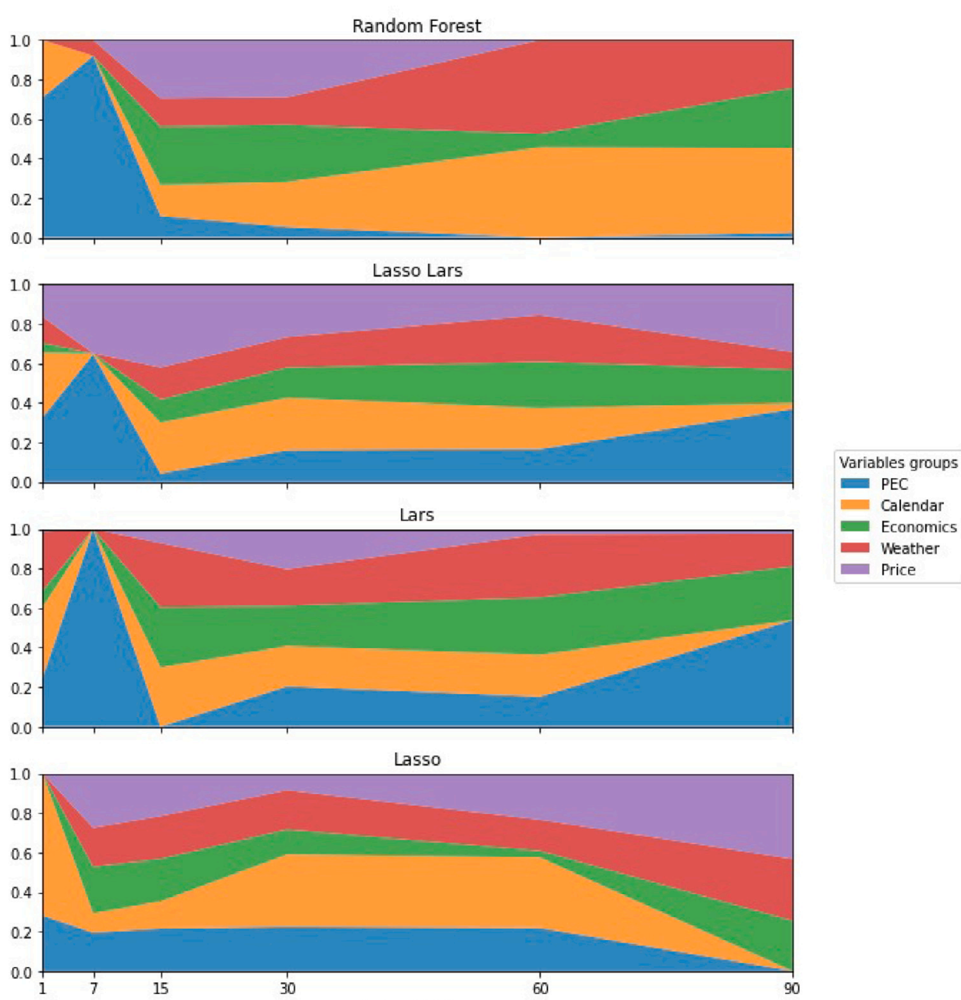

Fig. 9. Relative importance of the models coefficients sum for each time horizon.

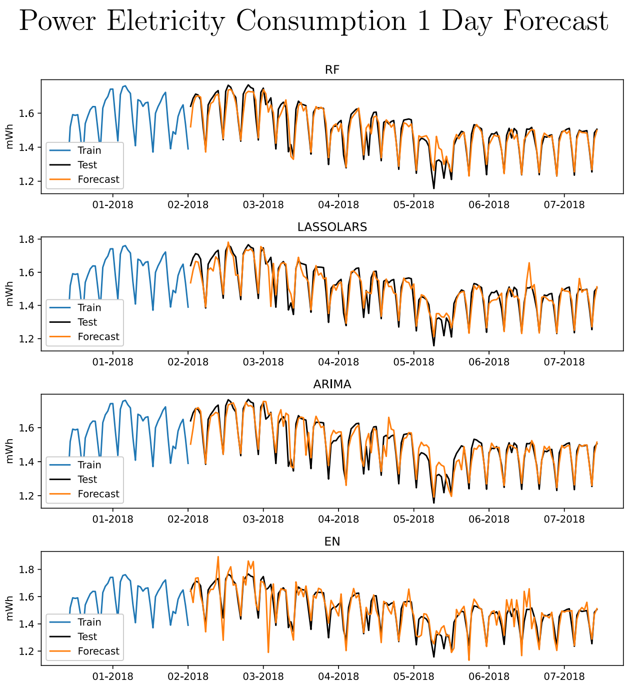

Fig. 10. The dynamics of PEC and 1 day forecast.

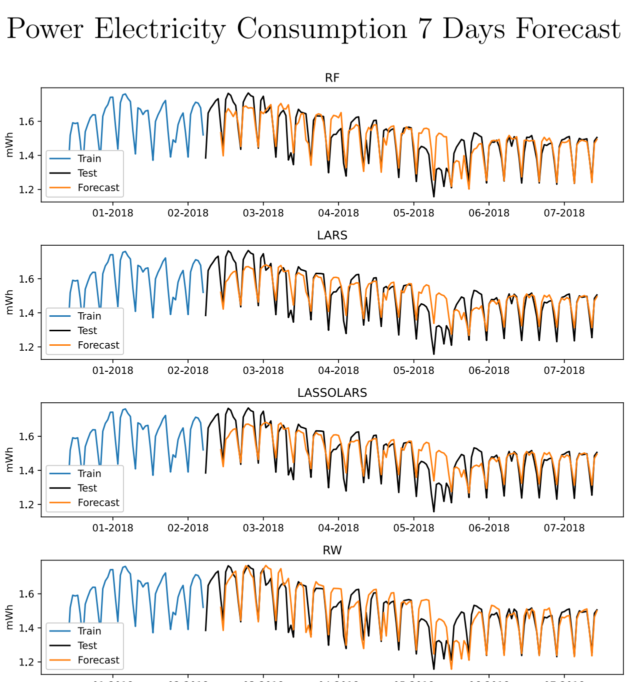

Fig. 11. The dynamics of PEC and 7 days forecast.

From Fig. 3 we see that PEC is highly correlated with its value for the previous day. Consequently, all major models attach great importance to lagged values in a 1-day forecast. Other variables have almost no relevance for the forecast for this time horizon. Random Forest assigns almost no value to all 60 coefficients associated with the variables in the electricity price set.

As the forecasting horizon increases, the series depends less on its past values and more on seasonal factors, which are captured by the calendar and weather Variables. Finally, the importance of price and economic variables rises for the STFG because it controls for the macroeconomic environment associated with the current and previous month of the forecast.

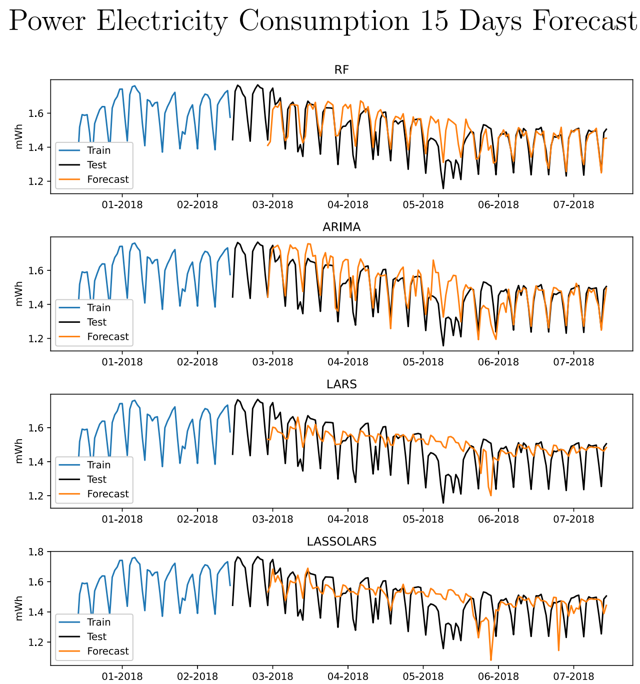

Fig. 12. The dynamics of PEC and 15 days forecast.

Table A.1
List of consumption classes.

Index	Class
CSO	Commercial, services and others
OC	Own consumption
SL	Street lighting
IND	Industrial
PP	Public power
RES	Residential
RR	Rural
ARR	Aquicultor rural
IRR	Irrigating rural
PS1	Public service (water, sewage and sanitation)
PS2	Public service (electrical traction)
TOT	Total
AVG	Average

Table A.2
List of regions.

Index	Region
SE	Southeast
MW	Midwest
S	South
NE	Northeast
N	North
BR	Brazil

Table A.3
Power electricity consumption variables (185 lagged variables).

Index	Variable
PEC_R_i	Hourly PEC of the hour i and region R (mWh) - (i = 0, ⋯, 24,)
PEC_R_TOT	Sum of daily PEC of region R (mWh)
PEC_R_AVG	Average daily PEC of region R (mWh)
PEC_R_C	Daily PEC of the region R for consumption class C (mWh)

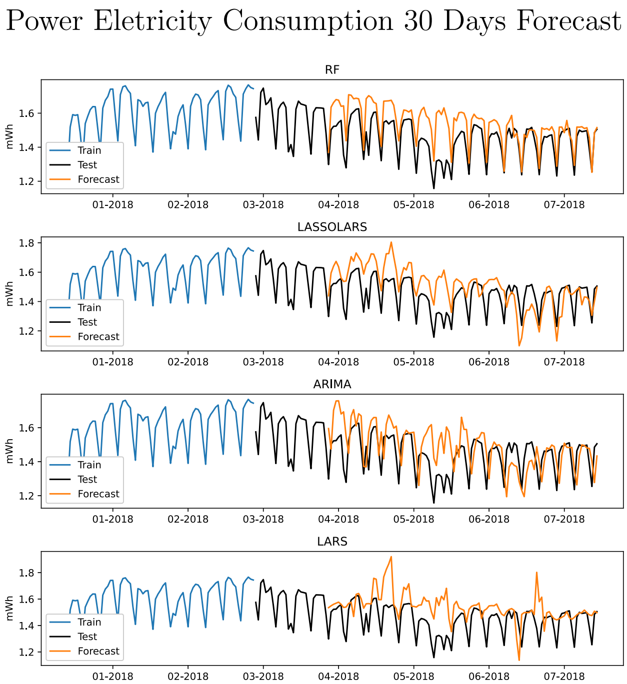

Fig. 13. The dynamics of PEC and 30 days forecast.

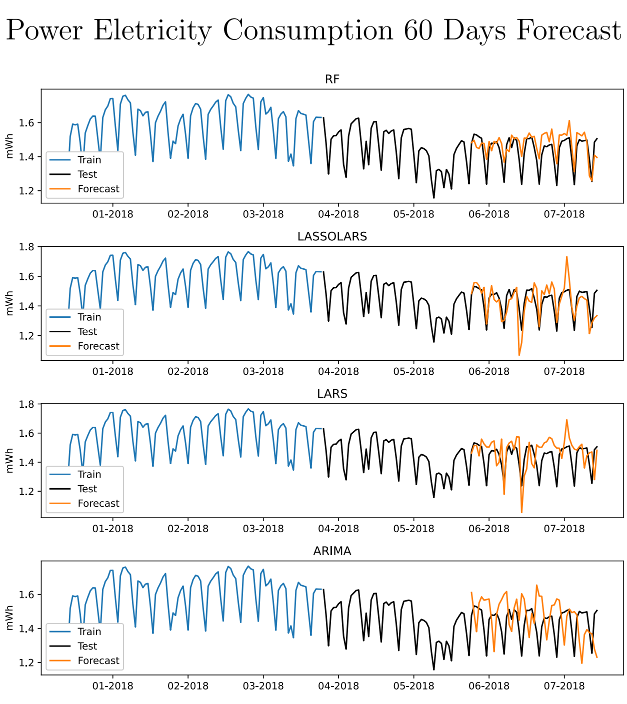

Fig. 14. The dynamics of PEC and 60 days forecast.

Figs. 10 through 15 show the dynamics of PEC and its forecasts for different horizons. Fig. 10 shows that all four best forecasts are able to capture the direction and seasonality of PEC with accuracy. Lagged values are very relevant in this forecast horizon. Fig. 11 shows that the predictions of the main models are able to follow PEC's trend and cyclicity. Additionally, even with a high autocorrelation, our main models perform better than the forecasts produced by the Random Walk.

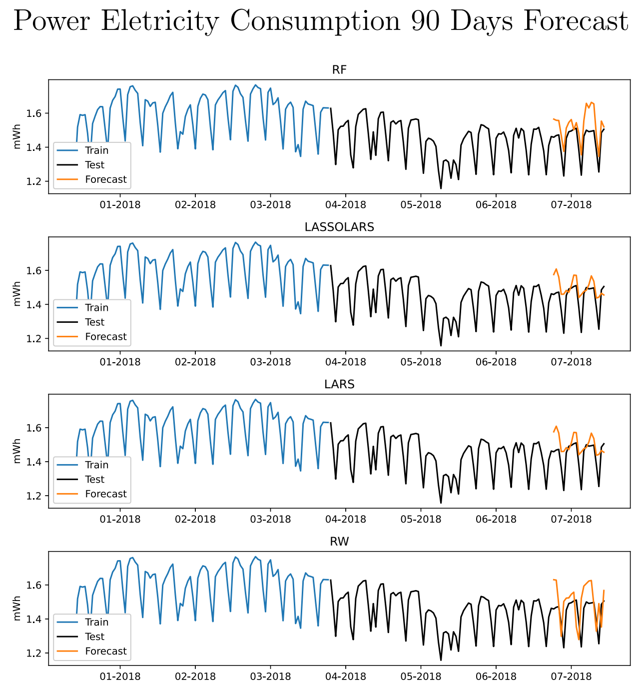

Fig. 15. The dynamics of PEC and 90 days forecast.

Table A.4
Calendar variables (6 variables).

Calendar variables
Day of the week (1–7)
Day of the month (1–30)
Month (1–12)
Year (2017, 2018)
Dummy for holiday
Seasons (1–4)

Table A.5
Weather variables (180 variables).

Index	Variable
CLD_R_i	Cloudiness of region R and hour i (i = 9,15,24, AVG)
AP_R_i	Atmospheric pressure (mbar) of region R and hour i (i = 9,15,24, AVG)
DBT_R_i	Dry bulb temperature (°C) of region R and hour i (i = 9,15,24, AVG)
HBT_R_i	Humid bulb temperature (°C) of region R and hour i (i = 9,15,24, AVG)
RH_R_i	Relative humidity (%) of region R and hour i (i = 9,15,24, AVG)
WD_R_i	Wind direction of region R and hour i (i = 9,15,24, AVG)
WV_R_i	Wind velocity (m/s) of region R and hour i (i = 9,15,24, AVG)

Table A.6
Electrical energy price variables (60 variables).

Index	Variable
PECR_R_C	PEC price of region R and consumption class C (R$)
PECR_R_AVG	Average PEC price of region R (R$)

The set containing lagged values of PEC loses (on average) its relevance for short and medium-term forecasts. Consequently, for the subsequent horizons the importance of all other variables significantly increase in the best ranked models. The results are intuitive since PEC is highly autocorrelated with its initial lags. As the forecast horizon increases, other variables gain importance, better capturing the series' seasonality and the economic environment. Fig. 12 shows that the best forecasts are able to follow both the trend and cyclicity of the series, with a small difficulty to find the local maximums and minimums. Moreover, we still have accurate results in a 30-days forecast. Fig. 13 shows that the best forecasts can follow the series' trend and cyclicity.

Table A.7
Economic variables (79 variables).

Variables
Selic target imposed by BCB
SELIC effective rate
CDI
Dollar for purchase
Dollar for purchase variance
Dollar for sale
Dollar for sale variance
Euro for purchase
Euro for purchase variance
Euro for sale
Euro for sale variance
IBOVESPA quotation
IBOVESPA daily minimum value
IBOVESPA daily maximum value
IBOVESPA daily absolute variance
IBOVESPA daily percentage variance
IBOVESPA daily volatility
INPC monthly variance
INPC cumulative
IPCA monthly variance
IPCA accumulated
IPA-M monthly variance
IPA-M accumulated
IPA-DI monthly variance
IPA-DI accumulated
IGP-M monthly variance
IGP-M accumulated
IGP-DI monthly variance
IGP-DI accumulated
Crude steel production (observed)
Crude steel production (seasonally adjusted)
Heavy vehicles traffic on toll roads (observed)
Heavy vehicles traffic on toll roads (seasonally adjusted)
Construction production inputs (observed)
Construction production inputs (seasonally adjusted)
CPS and Usecheque consultations (observed)
CPS and Usecheque consultations (seasonally adjusted)
Serasa consultations(observed)
Serasa consultations (seasonally adjusted)
Installed capacity utilization in the manufacturing industry - FGV (observed)
Installed capacity utilization in the manufacturing industry - FGV (seasonally adjusted)
Installed capacity utilization in the processing industry - CNI (observed)
Installed capacity utilization in the manufacturing industry - CNI (seasonally adjusted)
Real industrial sales (observed)
Real industrial sales (seasonally adjusted)
Production hours worked in the manufacturing industry (observed)
Production hours worked in the processing industry (seasonally adjusted)
Real wage in manufacturing industry (observed)
Real wage in the manufacturing industry (seasonally adjusted)
Oil and gross oil production (monthly production in m3)
Natural gas production (monthly production in m3)
General industrial production (observed)
General industrial production (seasonally adjusted)

(*continued on next page*)

The predictions made using Lasso model are able to follow the series seasonality even when forecasting a month ahead.

As we can note in Fig. 14, as the forecast horizon increases our predictions are less successful in reaching the local maximums and minimums of the series. As expected, when forecasting PEC 90-days ahead we have a smaller accuracy, but we still have good results. As we note in Fig. 15, the best models have a greater difficulty to follow the series cyclicity, but are still able to follow its trend.

5. Conclusion

This work uses high dimensional data to forecast PEC. We show that ML models outperform traditional ones when forecasting using this kind of database. We forecast PEC for 6 different horizons, which we divide into three groups: very short-term, short-term, and medium-term. The variables in our database can be classified into five sets: lagged demand, calendar variables, weather variables, electrical energy price variables, and economic variables. We make predictions using different models for each forecast horizon, among which are ARIMA, Random Walk, and several ML methods.

Table A.7 (*continued*).

Variables
Industrial production - capital goods (observed)
Industrial production - capital goods (seasonally adjusted)
Industrial production - intermediate goods (observed)
Industrial production - intermediate goods (seasonally adjusted)
Industrial production - consumer goods (observed)
Industrial production - consumer goods (seasonally adjusted)
Automotive industry production (observed)
Automotive industry production (seasonally adjusted)
Consumer confidence index
National consumer expectations index
Industrial entrepreneur confidence index
BNDES disbursements - accumulated amounts
Total employment index (observed)
Total employment index (seasonally adjusted)
Total employment index- manufacturing industry (observed)
Total employment index - manufacturing industry (seasonally adjusted)
Total employment index- commercial (observed)
Total employment index - commercial (seasonally adjusted)
Total employment index - services (observed)
Total employment index - services (seasonally adjusted)
Total employment index - civil construction (observed)
Total employment index - civil construction (seasonally adjusted)
Employed people (formally)
Unemployment rate (brazil)
Monetary base
Currency paper issued

The results show that the machine learning models with regularized coefficients using high dimensional data, especially Random Forest, consistently outperform the benchmark models. We have a larger relative gain in the medium term when predicting with ML models. On average, the prediction error of ML models is about 2 times smaller than the error of benchmark models. Furthermore, we conclude that lagged demands are the most relevant variables for very short-term forecast, but this does not necessarily happen for short and medium terms.

Although the focus of this paper is to forecast a specific variable (PEC) using a high dimensional dataset, we understand that we may successfully apply this approach to other high dimensional datasets with a time series structure. As in our paper, the regularized models will be able to select the most relevant variables to explain the dependent variable.

Finally, despite the fact that a high dimensional dataset is one of the main merits of our work, it is also one of the main constraints. In order to deal directly with such a high dimensional dataset, we have had to choose a class of regularized, easy-to-tune models and with low computational cost of estimation. An alternative approach is to separate the step of variables choice from the step of parameters choice. With that approach, we can explore other models that we have not considered here because they are not regularized, not easy to tune in or because they take too much time to converge. We can explore that approach in a future work.

CRediT authorship contribution statement

Pedro C. Albuquerque: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Data curation, Software. **Daniel O. Cajueiro:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision. **Marina D.C. Rossi:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table B.1
Hyperparameters set.

Model	Hyperparameter
Lasso	$\alpha \in [1e-20, 100]$
Lars	$\epsilon \in [1e-20, 100]$
Lasso lars	$\epsilon \in [1e-20, 100]$
Ridge	$\alpha \in [1e-20, 100]$
Elastic net	$\alpha \in [1e-20, 100]$, $ratio \in [0, 1]$
Random forest	$trees \in [10, 1000]$, $depth \in [10, 100]$, $samples\_leafs \in [1, 5]$, $samples\_splits \in [2, 10]$

Acknowledgments

The first and the second author are respectively indebted to CAPES, Brazil and CNPQ, Brazil (grant 302629/2019-0) for partial financial support.

Appendix A. Variables

Tables A.1 to A.7 detail the variables used in this work.

Appendix B. Hyperparameters sets

In this section, we present the hyperparameters sets that we test for our models. Table B.1 exhibits all sets of hyperparameters that we use in each machine learning model.

References

Adegbehin, A. B., Anumonye, E. N., Iguisi, E. O., Yusuf, Y., & Zubairu, I. (2016). Effect of weather parameters on hydroelectric power generation in kainji dam niger state, Nigeria. In *57th annual conference of the association of nigerian geographers* (pp. 1–11).

Albuquerque, P. C., Cajueiro, D. O., & Rossi, M. D. C. (2021). High dimensional models for forecasting power electricity consumption. URL: http://doi.org/10.5281/zenodo.4703186.

Almeshaiei, E., & Soltan, H. (2011). A methodology for electric power load forecasting. *Alexandria Engineering Journal*, *50*(2), 137–144. http://dx.doi.org/10.1016/j.aej.2011.01.015, URL: http://www.sciencedirect.com/science/article/pii/S1110016811000330.

Anderson, R., Boulanger, A., Chow, M., Dutta, H., Gross, P., Huang, B., Ierome, S., Isaac, D., Kressner, A., Passonneau, R., Radeva, A., Rudin, C., Salleb-Aouissi, A., Waltz, D., & Wu, L. (2011). Machine learning for the new york city power grid. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*, 328–345. http://dx.doi.org/10.1109/TPAMI.2011.108.

Baraniuk, R., Donoho, D., & Gavish, M. (2020). The science of deep learning. *Proceedings of the National Academy of Sciences*, *117*(48), 30029–30032.

BBC (2018). Greve dos caminhoneiros: a cronologia dos 10 dias que pararam o Brasil. [Online;] URL: https://www.bbc.com/portuguese/brasil-44302137. (Accessed 30 June 2019).

Bere, A., Boylan, J. E., Fildes, R., Lebotsa, M. E., & Sigauke, C. (2018). Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem. *Applied Energy*, *222*, 104–118. http://dx.doi.org/10.1016/j.apenergy.2018.03.155, URL: http://www.sciencedirect.com/science/article/pii/S030626191830504X.

Boldin, M., & Wright, J. H. (2015). Weather-adjusting economic data. *Brookings Papers on Economic Activity*, *46*(2), 227–278, URL: https://EconPapers.repec.org/RePEc:bin:bpeajo:v:46:y:2015:i:2015-02:p:227-278.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Bundervoet, T., Maiyo, L., & Sanghi, A. (2015). Bright Lights, Big Cities: measuring national and subnational economic growth in Africa from outer space, with an application to Kenya and Rwanda. Policy Research Working Paper 1 (7461. World Bank, Washington, DC.). URL: https://openknowledge.worldbank.org/handle/10986/22883.

Burns, K., & Moosa, I. (2016). The random walk as a forecasting benchmark: Drift or no drift? *Applied Economics*, *48*(43), 4131–4142. http://dx.doi.org/10.1080/00036846.2016.1153788, arXiv:https://doi.org/10.1080/00036846.2016.115378.

Calado, J., Cardeira, C., & Rodrigues, F. (2014). The daily and hourly energy consumption and load forecasting using artificial neural network method: A case study using a set of 93 households in portugal. *Energy Procedia*, *62*, 220–229. http://dx.doi.org/10.1016/j.egypro.2014.12.383, 6th International Conference on Sustainability in Energy and Buildings, SEB-14. URL: http://www.sciencedirect.com/science/article/pii/S1876610214034146.

Collado, R., & Creamer, G. (2016). Time series forecasting with a learning algorithm: An approximate dynamic programming approach. In *22nd international conference on computational statistics* (pp. 111–122).

Debnath, K., & Mourshed, M. (2018). Forecasting methods in energy planning models. *Renewable and Sustainable Energy Reviews*, *88*, 297–325. http://dx.doi.org/10.1016/j.rser.2018.02.002.

Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, *52*(3), 740–798. http://dx.doi.org/10.1257/jel.52.3.740, URL: http://www.aeaweb.org/articles?id=10.1257/jel.52.3.740.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*, 407–499.

El-Shazly, A. (2013). Electricity demand analysis and forecasting: A panel cointegration approach. *Energy Economics*, *40*, 251–258. http://dx.doi.org/10.1016/j.eneco.2013.07.003, URL: http://www.sciencedirect.com/science/article/pii/S0140988313001485.

EPE (2017). Balanço Energético Nacional.

Fan, S., & Hyndman (2011). Short-term load forecasting using semi-parametric additive models. In *2011 IEEE PES general meeting* (pp. 1–7). United States: IEEE, Institute of Electrical and Electronics Engineers, http://dx.doi.org/10.1109/PES.2011.6039207.

Freitas, E., Medeiros, M. C., & Vasconcelos, G. (2016). Forecasting brazilian inflation with high-dimensional models. *Brazilian Review of Econometrics*, *36*, 1980–2447. http://dx.doi.org/10.12660/bre.v99n992016.52273.

Garcia, M. G., Medeiros, M. C., & Vasconcelos, G. F. (2017). Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting*, *33*(3), 679–693. http://dx.doi.org/10.1016/j.ijforecast.2017.02.002, URL: http://www.sciencedirect.com/science/article/pii/S0169207017300262.

G.Creamer, G., Ghoddusi, H., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, *81*, 709–727, URL: https://doi.org/10.1016/j.eneco.2019.05.006.

Goude, Y., & Pierrot, A. (2011). Short-term electricity load forecasting with generalized additive models. In *ISAP power* (pp. 593–600).

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497, URL: http://www.jstor.org/stable/41057463.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, *102*(2), 994–1028. http://dx.doi.org/10.1257/aer.102.2.994, URL: http://www.aeaweb.org/articles?id=10.1257/aer.102.2.994.

Hoerl, A., & Kennard, R. (1970). Ridge regression: Biasedestimation for nonorthogonal. *Technometrics*, *12*, 55–67, URL: https://www.jstor.org/stable/1271436?seq=1#page_scan_tab_contents.

Horowitz, J. L. (2015). Variable selection and estimation in high-dimensional models. *Canadian Journal of Economics*, *48*, 389–407, URL: https://doi.org/10.1111/caje.12130.

Huang, J., Li, Y., Liu, D., Xiao, J., & Xie, L. (2018). A hybrid model based on selective ensemble for energy consumption forecasting in China. *Energy*, *159*(C), 534–546. http://dx.doi.org/10.1016/j.energy.2018.06.

Huang, S.-J., & Shih, K. R. (2003). Short-term load forecasting via ARMA model identification including non-Gaussian process considerations. *IEEE Transactions on Power Systems*, *18*(2), 673–679. http://dx.doi.org/10.1109/TPWRS.2003.811010.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. Otexts.

Maza, A., & Villaverde, J. (2007). A state-space approach to the analysis of economic shocks in Spain. *Journal of Policy Modeling*, *29*(1), 55–63, URL: http://www.sciencedirect.com/science/article/B6V82-4N1CRSM-7/1/cc6d22223592fe7fbcba46a024018163.

Medeiros, M. C., & Mendes, E. F. (2016). ℓ_1-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *Journal of Econometrics*, *191*(1), 255–271. http://dx.doi.org/10.1016/j.jeconom.2015.10.011, URL: http://www.sciencedirect.com/science/article/pii/S0304407615002638.

NASA (2000). Bright lights, big city. [Online;] URL: https://earthobservatory.nasa.gov/features/Lights. (Accessed 10 April 2019).

ORGANIZATION, W. W. M. (2017). *Guidelines on the calculation of climete normals*. Secretariat of the World Meteorological Organization.

Raihanian Mashhadi, A., & Behdad, S. (2018). Discriminant effects of consumer electronics use-phase attributes on household energy prediction. *Energy Policy*, *118*(C), 346–355. http://dx.doi.org/10.1016/j.enpol.2018.03.0, URL: https://ideas.repec.org/a/eee/enepol/v118y2018icp346-355.html.

Santosa, F., & Symes, W. W. (1986). Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, *7*, 1307–1330, URL: https://epubs.siam.org/doi/abs/10.1137/0907087.

Sejnowski, T. J. (2020). The unreasonable effectiveness of deep learning in artificial intelligence. *Proceedings of the National Academy of Sciences*, *117*(48), 30033–30038.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*, 267–288, URL: https://www.jstor.org/stable/2346178?seq=1#page_scan_tab_contents.

Times, N. (2018). Truckers' strike paralyzes brazil as president courts investors. URL: https://www.nytimes.com/2018/05/28/world/americas/brazil-truckers-strike-economy.html.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B.*, *67*, 301–320, URL: https://www.jstor.org/stable/3647580?seq=1#page_scan_tab_contents.