We note in Figure 2 that ~~the electricity consumption~~ PEC is highly correlated with ~~the~~ it ~~consumption of the~~ value 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 prediction in this time horizon. Lasso model, for instance, have all their variables coefficients equal to zero, except for the aggregate ~~power electric consumption~~ PEC at 11 p.m. of the previous day. ~~Still,~~ Also Random Forest assigns almost no value to all 60 coefficients of electricity price set.

Power Eletricity Consumption 1 Day Forecast

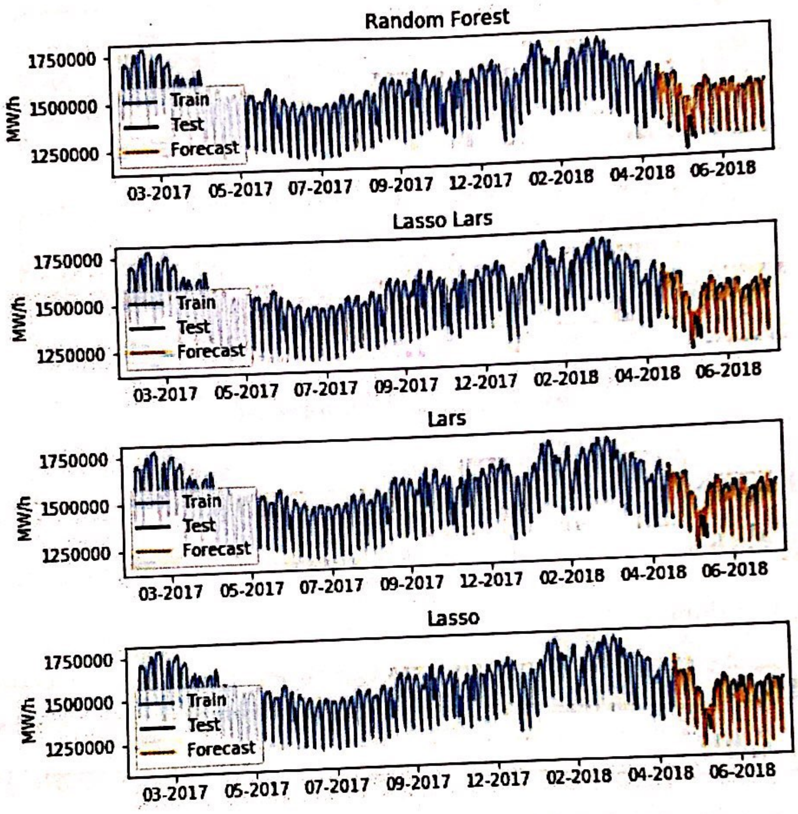


Figure 7 – 1 day forecast

Figure 7 shows that all four best forecasts are able to capture the direction and seasonality of electric power consumption with accuracy.

PEC's autocorrelation of 7 days ago is higher than the previous day autocorrelation, as we note in Figure 2. Table 3 shows that the weight attributed to lagged demand is larger for a 7-days forecast than ~~from~~ for 1-day forecast. As lagged values increase their importance in this forecast horizon, other~~s~~ sets of variables ~~relevance~~ reduce their relevance. That is, other explanatory variables are still not very important to make forecasts for the very short-term.

Power Electricity Consumption 7 Days Forecast

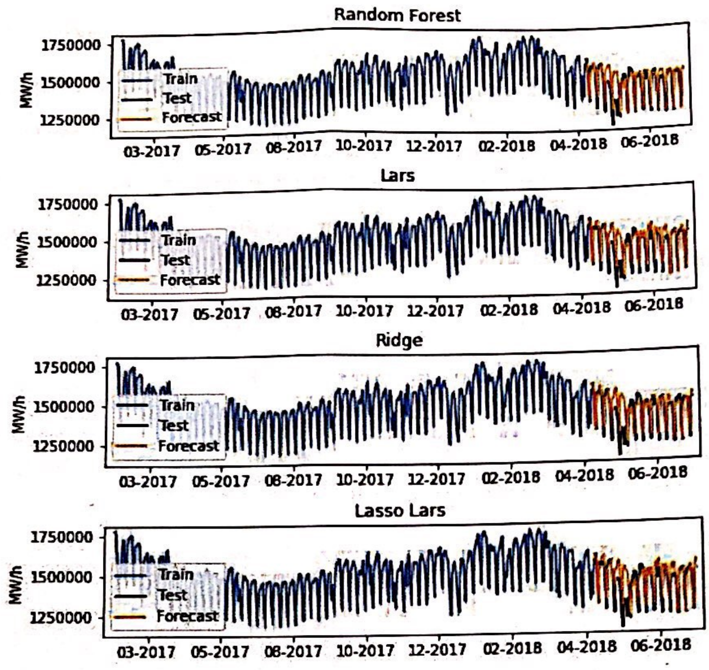


Figure 8 – 7 days forecast

Figure 8 shows that the predictions of the main models are able to follow PEC trend and cyclicity. Additionally, even with a high autocorrelation, our main models perform better than the Random Walk forecasts.

The set containing lagged values of PEC loses its relevance for short and medium term forecasts.

~~For short and medium term forecasts, lagged demand variable set loses its relevance~~ for the following dates as its autocorrelation ~~decreases~~. ~~In a 15-day forecast~~ for a 15-days forecast, the correlation with the past demand values are still present, however ~~increasingly~~ weak. When comparing with the 7-day forecast, we note that the sum of the parameters associated with lagged demand ~~lagged demand parameters sum~~ decreases in our main models. In Random Forest, this set declines faster than by more than a half.

Consequently, for the following horizons all other variables significantly expand their importance in the major models. The results are intuitive since PEC is highly autocorrelated with its initial lags. As the forecast horizon increases, other variables gain more importance, ~~capturing~~ better capturing the series seasonality and the whole macroeconomic environment.

Power Electricity Consumption 15 Days Forecast

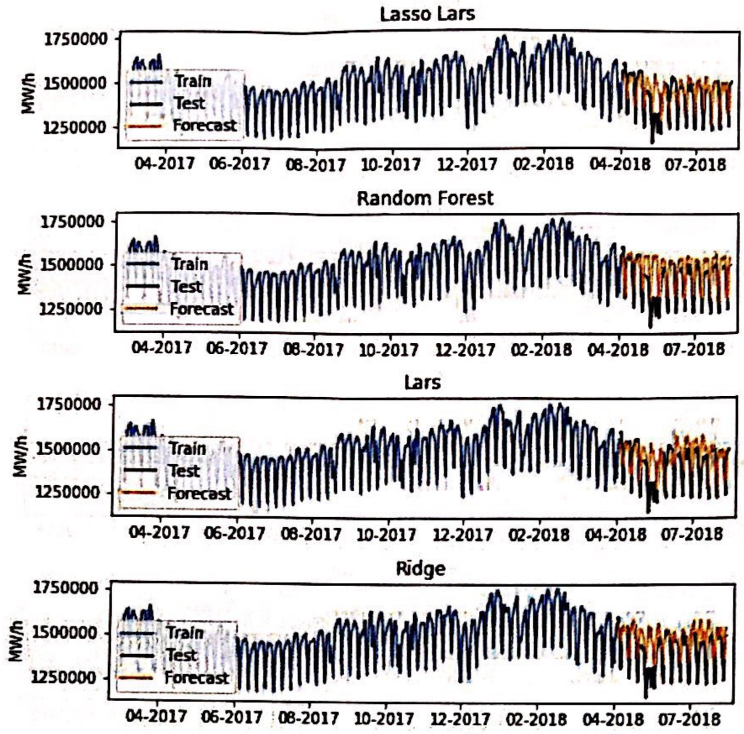


Figure 9 – 15 days forecast

(a)

Figure 9 again shows ~~shows again~~ that the best forecasts are able to ~~can~~ follow the series' trend and cyclicity, with a small difficulty to find the local maximums and minimums~~ of the series~~.

Moreover, we still have accurate results in a 30-day~~s ahead~~ forecast. Figure 10 shows that the best forecasts can follow the series trend and cyclicity. We note that the predictions made using Lasso model is ~~are~~ able to follow the series seasonality even when forecasting a month ahead. ~~in month-ahead forecast~~.

When forecasting 60-day ahead, we see an increase in the sum of the coefficients associated w/ price variables. ~~note a increse in price variables coefficients sum. Which is intuitive~~ This is intuitive, since economic agents tend to be more sensitive to price changes for longer horizons, as they have more time to adjust behavior.

~~For smaller horizons~~, Our predictions for short horizons are very close to the original values of PEC. ~~converge near to the original consumption values~~. At this stage, our forecasts are able to catch the direction and cyclicity of the original series ~~direction and cyclicity~~, as we can note in Figure 11. That is, as the forecast horizon increases ~~with an increase in the forecast horizon~~ our predictions are less successful in reaching the ~~obtain a greater difficult to reach~~ local maximums and minimums of the series.

When forecasting PEC 90-days ahead we have a smaller accuracy but we still have good

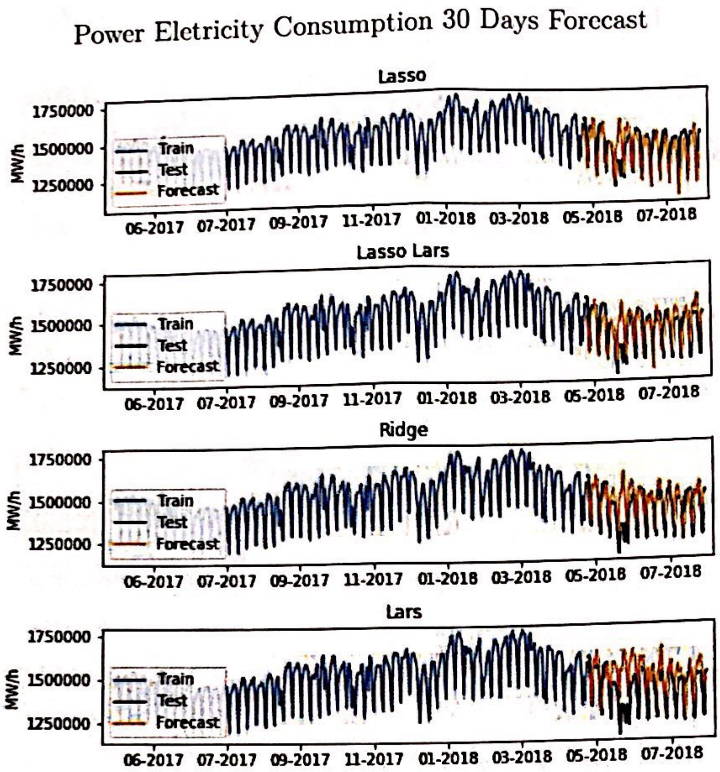


Figure 10 – 30 days forecast

results. As we note in Figure 12, the ~~main~~ best models have a greater difficulty to follow the series cyclicity, but still, they are able to follow its trend. Random Forest does not perform better due the unpredictable Brazilian truckers strike, when the model predict an increase in electricity consumption. Random Forest, although does not have the smallest RMSE, it manages to capture series direction and seasonality better than the other models, with the exception of the negative shock.

NOT SURE THAT YOU SHOULD SAY THIS.

In this way, if we are predicting 90 days ahead, Random Forest remains the model that best follows the series trend and the cyclicity. Even though, other models can give the direction of electricity consumption for the next 3 months, which is satisfactory to measure economic activity level.

stick to ya chosen criteria! RMSE

We also analyze the ~~gain to predict PEC with~~ gain in predicting PEC using machine learning models ~~compared~~ instead of ~~with traditional ones~~ our benchmark models. We normalize the error by dividing the RMSE of the best models ~~and the benchmarks~~ and of the random walk by ARIMA's RMSE for each time horizon. Figure 13 shows this gain.

fol sempre

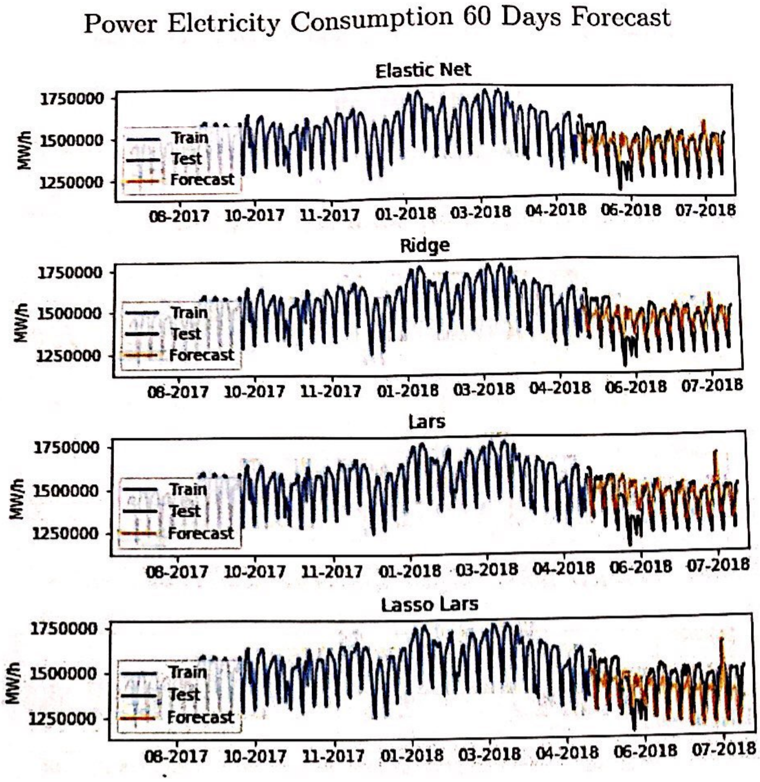


Figure 11 – 60 days forecast

~~We have a much higher~~ The relative gain in predicting PEC ~~with~~ using machine learning models is larger for short-term predictions ~~in the initial periods~~. The best model has an error that is about 40 times lower than the error of the ARIMA for 1-day forecast and that is about 20 times smaller for the ~~following~~ subsequent forecasts horizons. ~~Comparing~~ When compared to ~~with~~ the Random Walk, our best model has a forecast error that is about 5 times lower for the 1-day forecast and about 3 times lower for the ~~following~~ other forecast horizons.

Even when using almost the same variables to predict short-term PEC [1] as traditional models, machine learning models have a much better performance. This result occurs due to the high precision that these models have in short-term forecasts.

↳ isso ~~é medida~~ não diz nada - Apenas states o que já concluimos

↳ não diz ~~muito~~ nada sobre o mecanismo.

[1] Traditional models only use lagged variables to predict and these variables are the main ones for predicting short-term PEC with machine learning models.

Power Eletricity Consumption 90 Days Forecast

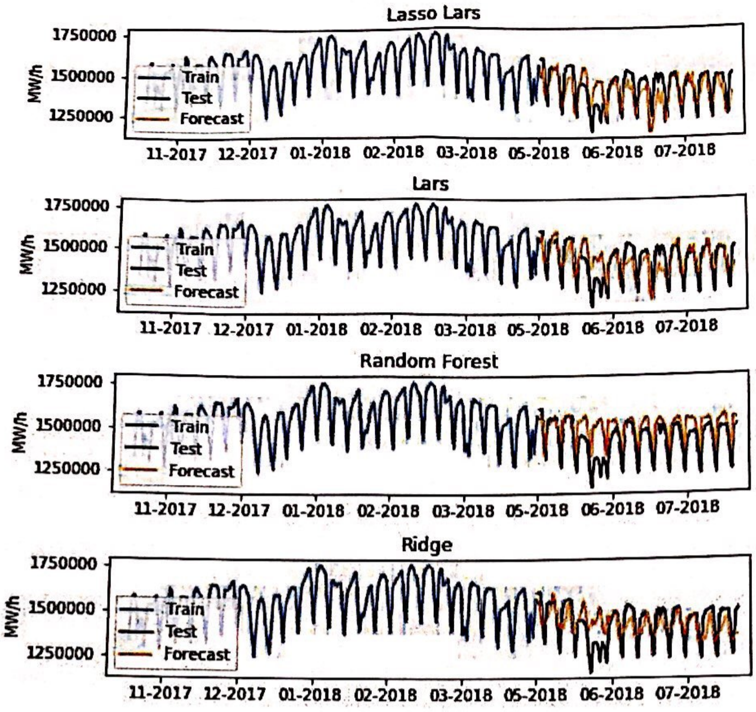
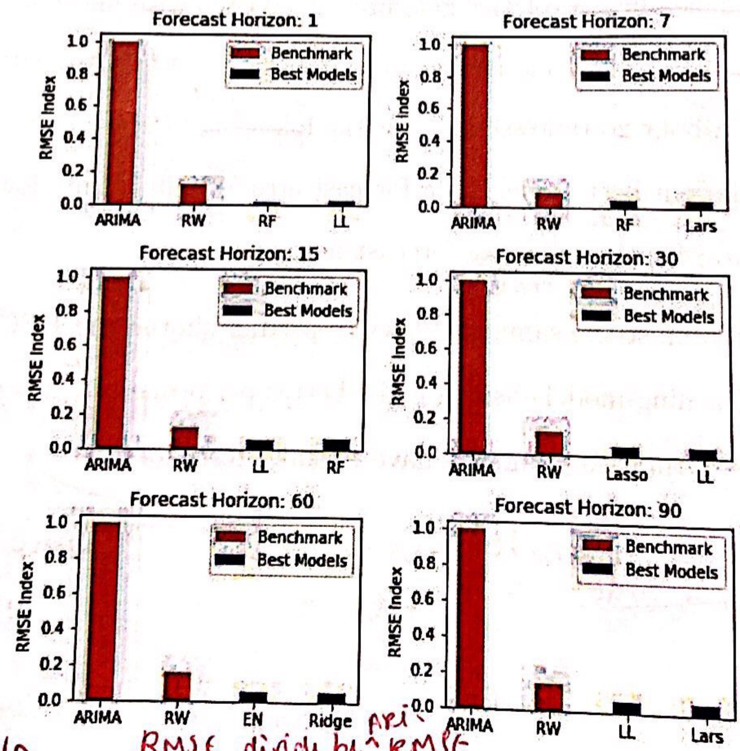


Figure 12 – 90 days forecast



No RMSE divid by ARI RMSE

We normalize ~~the error by dividing all RMSE analyzed by ARIMA's RMSE.~~

Figure 13 – Bar Plot Relative gain?

5 Conclusion

This work ~~proposes the use of~~ uses high dimensional data to forecast PEC. We show that ~~these~~ machine learning models perform better than traditional ones when forecasting ~~with~~ using this kind of database. We ~~build a structure to predict~~ forecast PEC for 6 different ~~forecasts~~ horizons, ~~divided between~~ which we divide into three groups: very short term, short term, and medium term. We shape a high dimensional dat~~abase with 5 main sets of variables~~ The variables in our database can be classified into five sets: lagged demand, calendar variables, weather variables, electric energy price variables, and economic variables. We make predictions ~~for~~ using more than ~~10~~ ten different models for ~~all~~ each forecast horizon~~s~~, among which are ARIMA, Random Walk, and ~~some~~ several machine learning methods.

The results show that ~~the best predictions are exclusively~~ the high dimensional models with regularized coefficients, especially Random Forest and Lasso Lars, consistently outperform the benchmark models. We have a bigger relative gain in the very short term when predicting with machine learning models. In a 1-day forecast, the prediction error of machine learning models is about 40 times smaller than the error of traditional models, while for the ~~following~~ subsequent forecast horizons ~~is 20 on average.~~ this error is, on average, 20 times smaller. Futhermore, we conclude that lagged demands are the most relevant variables for very short-term forecast and as we ~~advance in the time horizon~~ consider longer horizon, ~~all other variables set gain more importance.~~ the other sets of variables become more important.

Our paper shows how machine learning models predict PEC better than traditional ones. Future works may attempt to predict power electricity using deep learning models. We also may facilitate the variable selection for future works that intend to predict PEC.

?

melhorar esse parágrafo.

falar mais de future work or limitations.

Scanned with CamScanner