XLI CILAMCE
XLI Ibero-Latin American Congress on Computational Methods in Engineering
16-19 November, 2020 FOZ DO IGUAÇU PARANÁ – BRAZIL

FORECASTING THE SPOT PRICE BEHAVIOR IN THE BRAZILIAN ENERGY MARKET WITH STATISTICAL TOOLS

William Lagasse[1], Simone F. Venturini[1], Natália A. Weber[1], Fernando H. Scherer[1], Paulo S. Schneider[1], Amir R. De Toni[1], Walber F. Braga[2].

[1]*Dept. of Mechanical Engineering, Federal University of Rio Grande do Sul*
Rua Sarmento Leite, 425, 90046-902, Porto Alegre RS
williamlagasse@gmail.com, pss@mecanica.ufrgs.br, amir.detoni@ufrgs.br, weber.nati@gmail.com, sfventurini@gmail.com, fs.hubner@gmail.com
[2]*ILATIT, Federal University for Latin American Integration - UNILA*
Parque Tecnológico Itaipu – Complexo B, 85867-970, Foz do Iguaçu PR
walber.braga@unila.edu.br

Abstract. Electricity in the short-term market, or spot market, is traded according to the Difference Settlement Price PLD, whose calculation follows a complex and computationally demanding procedure. The present work proposes the use of statistical methods to forecast PLD values and trends in order to mitigate uncertainty in the decision making of market agents. Inputs for the proposed models are the affluent natural energy, the stored energy, the hydroelectric, thermal and wind generations, and the demand for electric energy, publicly disclosed by the system operator ONS, in addition to the historical series of the PLD itself, disclosed by Electricity Trading Chamber CCEE. One week ahead forecasts for the southern Brazilian submarket under heavy demand are proposed, based on time series and regression models for a 2015-2020 historical data. The best PLD forecast accuracy is achieved with the simple exponential smoothing time series model, with average errors of 17.91 % and 28.01 R$/MWh. The best trend forecast, of 64.77 %, is obtained by both time series exponential smoothing models.

Keywords: Spot price, Forecasting models, Regression models, Time series forecasting, PLD.

1 Introduction

The Brazilian electricity sector comprises two trading environments: Regulated Contract Environment– ACR, which includes stakeholders in power generation and distribution; and the Free Contract Environment ACL, comprising agents in power generation, distribution, traders, importers and exporters, as well as free and special consumers [1]. Trading on the ACL is affected by the pricing of the short-term market (spot market), in which differences between contracted and measured power are settled. The spot price, officially Difference Settlement Price – PLD is determined by complex, computationally demanding programs, which are operated by high skilled personnel. Market agents employ different strategies to maximize profits and sudden PLD changes may have significant effects on the economy. Therefore, an opportunity was identified to develop and compare statistical models to forecast PLD values and its trends, with a prediction horizon of one week, in the southern Brazilian submarket under heavy demand.

2 Spot price prediction

The PLD is the reference value employed to reckon the total energy traded in the National Interconnected System SIN and to settle the electricity trading in the Brazilian spot market [2]. Its value is computed by the programs NEWAVE and DECOMP, developed by the Electrical Energy Research Center – CEPEL. The Electricity Trading Chamber – CCEE weekly evaluates the PLD to report it every Friday, for each Brazilian submarket and demand level, and within a price range annually set by the National Electric Energy

Agency– ANEEL. The maximum price is due to variable costs of thermal power plants while the minimum price considers operation and maintenance costs of hydroelectric plants [3]. The present work employed two statistical methods for PLD forecasting: regression analysis and time series.

Regression analysis evaluates the relationship between a dependent variable and independent parameters. It delivers a parameter model whose coefficients are determined by the method of least squares [4]. Time series analysis uses past behavior of the studied variable to forecast its value, by means of breaking down the data into four components: trend, seasonality, cyclical variations and irregular variations [5]. These two approaches were used to forecast the PLD under heavy demand for the southern Brazilian submarket, in a methodology that comprises seven steps, as show in Fig. 1.

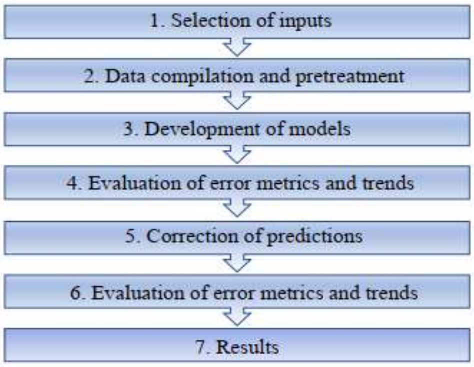

Figure 1. Methodology to forecast PLD values and trends with statistical models

The first step aims to identify and further on select the most relevant input parameters related to the PLD calculation employed by parameters NEWAVE and DECOMP. The seven chosen ones are: Affluent Natural Energy (ENA), Stored Energy (EA), Hydroelectric Generation (GH), Thermoelectric Generation (GT), Wind Power Generation (GE), Electricity Exchange between southern and southeastern submarkets (IE) and Electricity Demand (D). This data is publicly available from the National Electric System Operator – ONS < http://www.ons.org.br/paginas/resultados-da-operacao/historico-da-operacao>. The correspondent PLD values for that dataset was included from <https://www.ccee.org.br/portal/faces/pages_publico/o-que-fazemos/como_ccee_atua/precos>.

In the second step, the aforementioned parameters were obtained from the ONS and CCEE databases and preprocessed to handle missing data and to present them in a weekly basis format. The present forecasting effort employed historical data between the second operative week of 2015 until the last operative week of January 2020. Inflation's effect on electricity pricing was accounted for by use of the General Index of Market Prices (IGP-M), which was also converted to weekly basis from its original monthly basis. The next step was dedicated to the computation of the multiple regression equations and time series models. In the fourth step, a set of error metrics were computed for each forecasting model, i.e., mean absolute error (MAE), mean absolute percentage error (MAPE), and mean squared error (MSE), presented in Table 1.

Table 1. Error metrics employed to evaluate the proposed forecast models.

MAE	Mean Absolute Error	$MAE = \frac{1}{n} \times \sum_{i=1}^{n} \lvert Y_i - \hat{Y}_i \rvert$	(1)
MAPE	Mean Absolute Percentage Error	$MAPE = \left(\frac{1}{n} \times \sum_{i=1}^{n} \left\lvert \frac{Y_i - \hat{Y}_i}{Y_i} \right\rvert \right) \times 100\%$	(2)
MSE	Mean Squared Error	$MSE = \frac{1}{n} \times \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2$	(3)

with Y_i and $\hat{Y}_i$ the actual value of PLD and the predicted value at the period *i*, respectively, while *n* is the total number of periods.

Additionally, the predicted values were compared to the real data from CCEE and the accuracy of trend prediction was evaluated. The trend prediction was computed as follows: the actual PLD values for two consecutive weeks were compared and classified as increasing ($Y_{i+1} - Y_i > 0$), decreasing ($Y_{i+1} - Y_i < 0$), or stagnant ($Y_{i+1} - Y_i = 0$). Also computed is the difference between the predicted value one week ahead and the actual current value, i.e., $\hat{Y}_{i+1} - Y_i$. Therefore, a correct prediction occurs when both comparisons match.

The fifth step consists in a modified version of the third one, this time correcting predicted values in order to fit the annual price range set by ANEEL, which means that any predicted value above the maximum is corrected to be equal to the maximum, and correspondingly for the minimum. The sixth step repeats the fourth with these corrected values, while the last step consists in assessing the error metrics to determine which model is most useful.

3 Results

Regression models were obtained using 80 % of the dataset while the remaining 20 % were employed for forecast comparison. The R^2_{pred} coefficient indicates the model's capacity to predict the 20 % data and it was adopted to determine which subset of the input parameters provides the best regression models. Coefficient R^2_{adj} evaluates models with different number of input parameters and it increases as more inputs are included, while R^2 indicates the overall agreement between the equation and the input data.

Table 2 shows ten subsets of input parameters for the multiple linear regressions, in decreasing order of R^2.

Table 2. Input variable subsets for multiple linear regression prediction of PLD values (southern Brazilian submarket, heavy demand)

Input Parameters								# of input parameters	R^2 *pred*	R^2 *adj*	R^2
D	EA	ENA	GE	GH	GT	I_{IGPM}	IE				
X	X		X	X		X	X	6	53.0 %	54.5 %	55.5 %
X	X			X	X	X	X	6	53.0 %	54.5 %	55.5 %
X	X			X		X	X	5	53.0 %	54.3 %	55.2 %
X	X	X		X	X	X	X	7	52.8 %	54.5 %	55.7 %
X	X	X	X	X		X	X	7	52.8 %	54.5 %	55.7 %
X	X	X	X	X	X	X	X	8	52.5 %	54.4 %	55.8 %
X			X	X		X	X	5	50.3 %	51.5 %	52.4 %
X				X		X	X	4	50.3 %	51.5 %	52.3 %
	X			X	X	X		4	46.5 %	47.6 %	48.4 %
X				X			X	3	43.3 %	44.5 %	45.1 %

One can notice that three subsets of parameters presented the highest R^2_{pred}, i.e. 53 %, with two subsets comprising six input parameters and the other comprising five input parameters. Table 3 presents the multiple linear regression equations for the top three models regarding R^2_{pred}.

Table 3. Multiple linear regression equations for the best performing models

Model	Equation
i	$PLD = 619 + 0.1946 \times D + 0.0097 \times EA - 0.0527 \times GE - 0.2079 \times GH - 0.4309 \times I_{IGPM} - 0.1795 \times IE$
ii	$PLD = 598 + 0.1624 \times D + 0.01009 \times EA - 0.1738 \times GH + 0.0453 \times GT - 0.4438 \times I_{IGPM} - 0.1460 \times IE$
iii	$PLD = 652 + 0.1886 \times D + 0.00946 \times EA - 0.1999 \times GH - 0.4728 \times I_{IGPM} - 0.1708 \times IE$

Table 4 shows the error metrics for these three models, according to the comparison criteria described in the previous section.

Table 4. Error metrics and percentage of correct trend prediction for the best performing regression models

Model	Error metrics for PLD value			Correct trend prediction
	MAE	MAPE	MSE	
i	79.79	58.28 %	9705.20	56.06%
ii	79.62	57.94 %	9704.20	55.68%
iii	79.63	58.63 %	9773.06	56.44%

Model ii presents the lowest errors regarding PLD value while model iii is slightly better at trend prediction. Nevertheless, all three models are similar to each other regarding these metrics, with significant PLD errors and trend predictions close to 50 %. Therefore, as described on methodology's fifth step, a correction is implemented using the annual PLD range set by ANEEL. Table 5 presents the results after the correction.

Table 5. Error metrics and percentage of correct trend prediction for the best performing regression models including annual PLD limits.

Model	Error metrics for PLD value			Correct trend prediction
	MAE	MAPE	MSE	
i	75.20	56.58 %	9277.70	60.98 %
ii	75.06	56.28 %	9278.82	60.61 %
iii	75.53	56.98 %	9411.94	61.36 %

After the correction, model ii again presents the lowest errors (MAE and MAPE) while model iii correctly predicts the trend in 61,36 % of the cases. Yet, all three models remain similar, with the correction improving trend prediction from 56 % to 61 % while percentage errors decreased from 58 % to 56 %.

Regarding the use of time series, three moving mean models were applied: simple moving mean of size 4 (MMS4), in which the prediction is equal to the simple mean of the last four values; and, accordingly, simple moving mean models of size 3 and 2 (MMS3 and MMS2). Table 6 presents the error metrics and trend predictions of these time series models. The lowest errors were obtained with model vi (MMS2) while the trending prediction was close to 50 % for the three time series models.

Table 6. Error metrics and percentage of correct trend prediction for simple moving mean time series models

Model	Method	Error metrics for PLD value			Correct trend prediction
		MAE	MAPE	MSE	
iv	MMS4	47.88	33.34 %	5943.29	49.81 %
v	MMS3	41.39	28.85 %	4732.63	50.00 %
vi	MMS2	34.60	23.95 %	3570.35	50.19 %

An additional time series model developed uses simple exponential smoothing (SES), which smooths the data via exponentially weighted averages. This method employs a smoothing parameter α, which determines the weight of more recent data, and its value is computed in an iterative way, in order to minimize the quadratic errors. The optimized value for α was 1,12211, for the historical PLD data of the southern Brazilian submarket under heavy demand.

The last PLD forecasting model employs a double exponential smoothing (SED), which is recommended for data with noticeable trends but essentially no seasonality [8]. As the name implies, this model works with two smoothing parameters, the level constant α, similar to the smoothing parameter of SES, and the trend constant β, a moving difference between consecutive observations. Once again, an iterative calculation was employed to optimize these constants, thus obtaining α=1.20731 and β=0.01109. Table 7 presents the error metrics and trend prediction of these exponential smoothing models.

Table 7. Error metrics and percentage of correct trend prediction for exponential smoothing time series models

Model	Method	Error metrics for PLD value			Correct trend prediction
		MAE	MAPE	MSE	
vii	SES	28.30	18.30 %	2524.03	54.54 %
viii	SED	29.85	19.44 %	2574.97	41.67 %

One can notice that both exponential smoothing models displayed similar error metrics, with model vii performing slightly better. However, the trend prediction showed a significant difference of 12.87 % in favor of model vii. Following the preceding methods, these models were also subjected to a correction of predicted values according to the ANEEL's annual PLD range. Table 8 shows the results obtained with this correction.

Table 8. Error metrics and percentage of correct trend prediction for exponential smoothing time series models including annual PLD limits.

Model	Method	Error metrics for PLD value			Correct trend prediction
		MAE	MAPE	MSE	
vii	SES	28.01	17.91 %	2510.15	59.85 %
viii	SED	29.01	18.13 %	2574.74	50.38 %

That the correction slightly improved the error metrics for both models, while the trend predictions shows a more pronounced improvement, i.e., model vii was enhanced by 5.31 % and model viii improved 8.71 %.

4 Discussion

Given the preceding comparisons, the corrected vii model provided the lowest error metrics, with a MAE of 28.01 R$/MWh, MAPE of 17.91 % and MSE of 2510.15. Figure 2 presents the comparison between actual PLD values and predictions from the corrected vii model for the entire range of historical data.

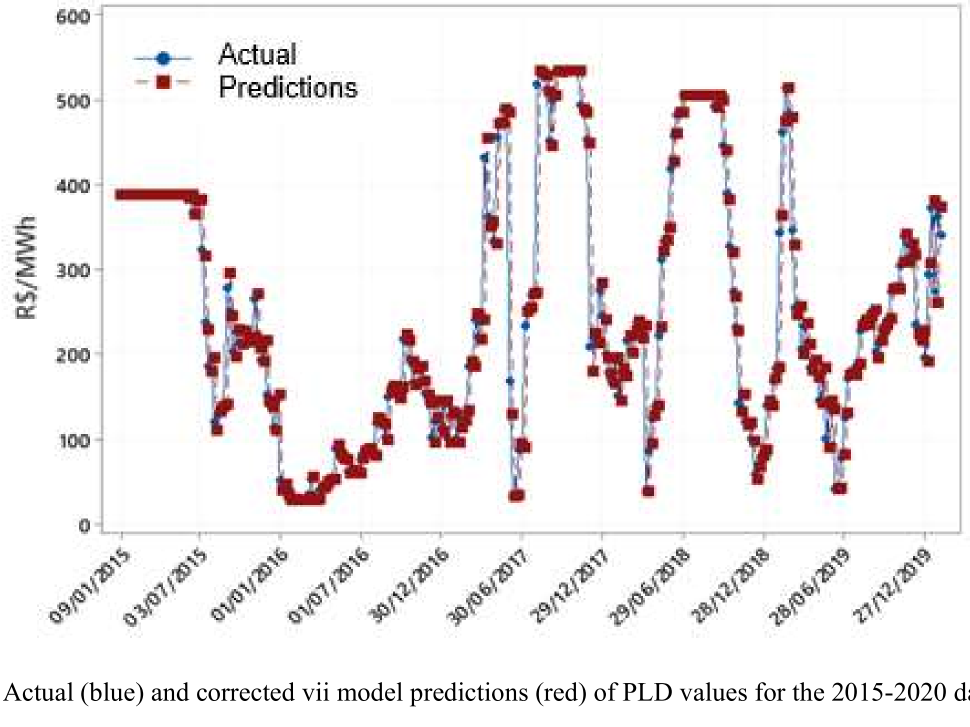

Figure 2. Actual (blue) and corrected vii model predictions (red) of PLD values for the 2015-2020 data period.

Time series models offered PLD predictions with the lower error metrics than the regression ones, but this does not necessarily mean the time series models provided correct trend prediction. Among the time series, modified model vii presented the best trend prediction, 59.85 %, even though it did not include a measurement of overall trend like model viii. Among the regression models, error metrics were higher, but trend prediction was marginally better than their time series counterparts. A likely explanation for this better performance is the causal relationship between the input parameters of the regression model and the desired output. The parameters Electricity Demand, Hydroelectric Generation, and Electricity Exchange were all included on the better performing regression models, with the best trend prediction being provided by the modified/limited model iii, with 61.36 %. Figure 3 shows the comparison between actual PLD values and corrected model iii for the entire range of historical data.

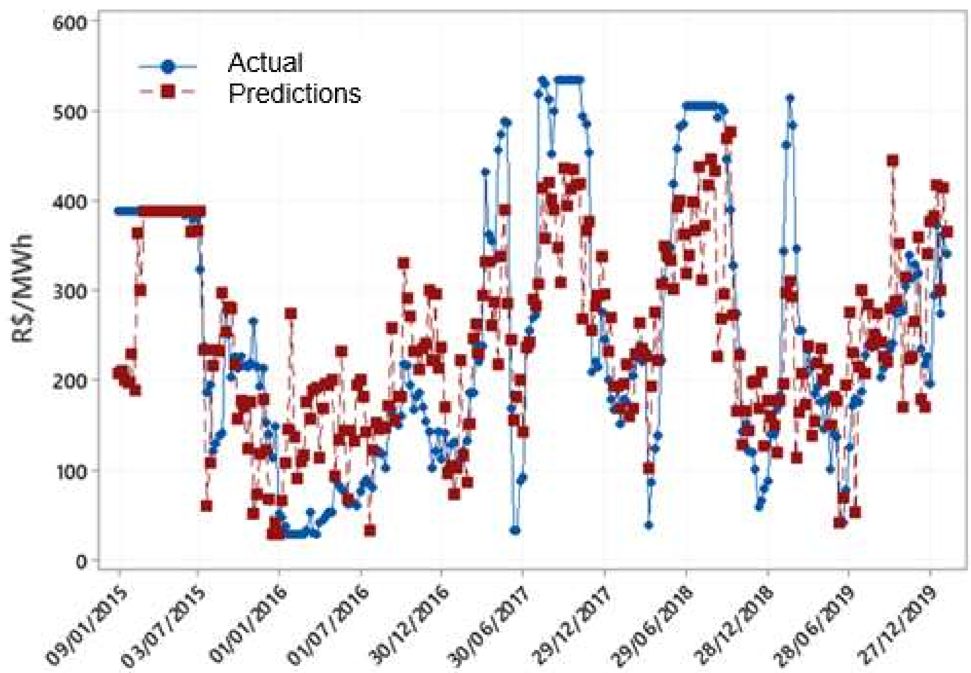

Figure 3. Actual (blue) and corrected iii model predictions (red) of PLD values for the 2015-2020 data period.

A closer look at the historical data shows that PLD values are unlikely to show outliers, i.e., a three-week period of low-high-low or high-low-high prices. The mean absolute difference between consecutive weeks was 28.29 R$/MWh with standard deviation of 48.21 R$/MWh. For comparison, the error metrics of a moving mean of size are MAE of 28.18 R$/MWh, MAPE of 18.21 % and MSE of 2565.06. These error metrics are similar to the ones provided by the exponential smoothing models.

In order to better aid market agents, a modification was introduced to improve the information provided by the proposed models. This modification consists in a new classification of increasing, decreasing or stagnant trend, which is an enhancement of the fourth step described in previous section. The new classification rule considers an increasing trend if $Y_{i+1} - Y_i > 25$, a decreasing trend if $Y_{i+1} - Y_i < -25$, and stagnant value if $25 > Y_{i+1} - Y_i > -25$, with the same criteria applied to predicted values, $\hat{Y}_{i+1} - Y_i$. Table 9 presents the correct trend predictions for all developed models using this new classification.

Table 9. Correct trend prediction of PLD values considering a stagnant trend if week-to-week variation is inferior to R$ 25.00 R$/MWh

Model	Method	Correct trend prediction
i	Regression	50.76 %
ii	Regression	52.27 %
iii	Regression	50.38 %
iv	Time series - Simple moving mean of size 4	50.57 %
v	Time series - Simple moving mean of size 3	52.67 %
vi	Time series - Simple moving mean of size 2	59.70 %
vii	Time series - Simple exponential smoothing	64.77 %
viii	Time series - Double exponential smoothing	64.77 %

One can see that both time series models vii and viii show correct trend predictions of 64.77 %, thus being the best performing model developed in the present work. On the other hand, the new criteria negatively impacted the trend predictions of the regression models, which is expected since the expanded stagnant category does not match well with the overprediction and underpredictions of the regression model.

5 Conclusions

The present work reported the development of linear regression and time series models to forecast electricity spot market price (PLD) and its trend for a horizon of one week. Regression models employed eight input parameters: Affluent Natural Energy (ENA), Stored Energy (EA), Hydroelectric Generation (GH), Thermoelectric Generation (GT), Wind Power Generation (GE), Electricity Exchange between southern and southeastern submarkets (IE), Electricity Demand (D) and General Index of Market Prices (IGP-M). The proposed models were evaluated based on mean absolute error (MAE), mean absolute percentage error (MAPE), and mean squared error (MSE). The time series model vii, with simple exponential smoothing, presents the lowest error metrics, with MAE of 28,01 R$/MWh and MAPE of 17.91 %. Regarding trend prediction, two different criteria were considered: when a stagnant trend was considered only when two consecutive PLD values were identical, the regression model iii correctly predicted the trend 61.36 % of the time. When the stagnant trend criteria was broadened to ± 25.00 R$/MWh, the time series exponential smoothing models showed 64.77 % of correct trend prediction.

Acknowledgements. S.F. Venturini acknowledges the financial support from Capes for her MSc scholarship; P.S. Schneider acknowledges CNPq for his research grant (CNPq-PQ 301619/2019-0).

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property and authorship of the authors, or has the permission of the owners to be included here.

References

[1] CCEE, 2020. Câmara de Comercialização de Energia Elétrica. Ambiente livre e ambiente regulado. 01 abr 2020 <https://www.ccee.org.br/>
[2] Brasil, 2004. Lei Nº 10.848, de 15 de março de 2004. Dispõe sobre a comercialização de energia elétrica. 10 abr 2020 <http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Lei/L10.848.htm>
[3] Brasil, 2004. Decreto Nº 5.163 de 30 de julho de 2004. Regulamenta a comercialização de energia elétrica, o processo de outorga de concessões e de autorizações de geração de energia elétrica, e dá outras providências. 10 abr 2020 <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/decreto/D5163.htm>
[4] B. H. Archer. Forecasting demand: Quantitative and intuitive techniques. *International Journal of Tourism Management*, pp. 5–12.
[5] P. A., Morettin and C. M. C. Toloi. *Análise de séries temporais*. Edgard Blucher, 2ª edição, 2006.
[6] ONS, 2020. Operador Nacional do Sistema. Histórico da operação. 15 abr 2020 <http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/>
[7] CCEE, 2020. Câmara de Comercialização de Energia Elétrica. Preços. 15 abr 2020 <https://www.ccee.org.br/>
[8] Nau, R. *Statistical forecasting: notes on regression and time series analysis*. Duke University. 27 mai 2020 <http://people.duke.edu/~rnau/411home.htm>.