Electric Energy Consumption Forecast Based on Spatial Information

Carolina L. S. Cipriano *, Mayara G. Silva*, Weldson A. Corrêa*,
João D. S. Almeida*, Márcia I. A. Silva†, João O. B. Diniz*
*Applied Computing Group (NCA), Federal University of Maranhão (UFMA), São Luís - MA, Brazil
Email: {carol, mayara, weldson, jdallyson, joao.bandeira}@nca.ufma.br
†Equatorial Energia, Brazil
Email: marcia.alves@equatorialenergia.com.br

***Abstract*—The task of predicting the consumer's electricity consumption is currently a trend in power energy companies. This prediction becomes difficult or impractical for consumers with no history or a short history of consumption. Thus, this work deals with an alternative to the prediction of energy consumption for these consumers. The proposed method is based on the consumption of the k closest neighbors and the consumption forecast made by one of three available regression models. The regressors used, namely Autoregressive Integrated Moving Average (ARIMA), Boosting Additive Quantile Regression (BAQR) and the named Seasonal and Trend decomposition using Loess (STL), were chosen for providing the best performance. The results obtained were promising, achieved a mean of the 30.4 % in the symmetric mean absolute percentage error (sMAPE) metric in a dataset with 86,874 customers.**

Keywords–Geospatial Information; Energy Forecasting; STL; ARIMA; BAQR

I. Introduction

It is a current trend for power companies to invest in artificial intelligence and machine learning to predict the monthly behavior of their consumers' energy consumption [1][2]. Forecasting is beneficial for both energy companies and consumers. This mutual benefit comes from reducing the energy company's expenses during power distribution and increasing its revenues. By reducing financial losses caused by wrong measurements or power thefts, it can then pass on to its consumers a lower energy consumption billing rate.

The problem of predicting consumers' electricity consumption is an essential step in verifying inconsistencies in measuring monthly energy consumption. Inconsistency checking avoids both incorrect billing for a consumer and may indicate that, due to abnormal energy consumption, the consumer may be using technical arrangements to reduce his energy consumption and, therefore, not being properly registered. For this reason, power companies have invested in Pattern Recognition (PR) methods to predict their customers' energy consumption and thus improve the verification step for measuring energy consumption inconsistencies.

In practice, each company defines its criteria for checking for inconsistencies in reading data. For this verification, Equatorial Energy uses the average consumption of the last threes months as a forecast for each consumer. Forecasted consumption is used to define a minimum and maximum consumption range. This range is defined to avoid errors and anomalies in the reading performed. Readings outside the expected range are reviewed by company technicians before issuing the customer invoice.

Our motivation stems from the fact that this task of predicting the behavior of power consumption is relatively simple when its consumers' consumption history exists, but it becomes difficult or impractical for consumers with no consumption history, i.e., for new consumer installations or those who have a short history of power consumption. Consequently, the reading of these customers usually goes through technical analysis before issuing the invoice, given the impossibility of predicting consumption.

In this sense, to solve the problem of energy consumption prediction of consumers without consumption history and reduce the number of customers that are analyzed before the invoice issue, this paper proposes an energy consumption prediction method using neighborhood consumption information. Indeed, it is likely that a new customer will have a consumption similar to that of its closest neighbors, as well as its consumption range.

In the proposed work, the term spatial information is related to neighborhood identification, to compute energy consumption based on Tobler's first law of geography [3] stating that everything is related to everything else, but near things are more related than distant things.

This work is part of a Research and Development (R&D) project, contracted by Equatorial Energy under contract CELPA 962/2018 and CEMAR 30/2019, executed by the Applied Computing Group (NCA) of the Federal University of Maranhão (UFMA). This project will develop the Consumption Habit Analysis System (SisHCo). The project is organized to provide the development of methods, techniques, and tools based on computational intelligence and machine learning, to define, in an individualized and adaptive way, parameters for the critique of power consumption measurement, based on historical information.

The main contributions resulting from this work are:

1) We developed a method that solves a real problem of the energy supply companies, using simple techniques and with reasonable accuracy;
2) An alternative method was developed to forecast energy consumption in customers without a history of consumption;
3) The proposed method uses spatial data as a step in the forecast flow of energy consumption;
4) We determined and compared the most accurate prediction methods.

The rest of the paper is organized as follows. In Section II, we present the main related works. Section III describes the proposed methodology of the prediction of individual power consumption of consumers. The presentation of the results is given in Section IV. Finally, a conclusion on the results obtained is drawn in Section V.

Copyright (c) IARIA, 2020. ISBN: 978-1-61208-762-7

II. Related Work

A Long Short-Term Memory (LSTM) network was used by Alonso et al. [4] for predicting the individual hourly load data of consumers. The prediction model was generated from consumption, weather and calendar data. In the construction of the model, data from 3,891 smart meters in 2013 were collected, producing 8,760 readings on each meter. Then, it was evaluated how the spatial location of the residential customers influences the load prediction. Results indicate that the proposed model mean-absolute error was 19 % better than that of the ARIMA [5] and around 24 % better than that of the seasonal naive approach [6].

Bâra and Oprea [7] evaluated the dynamic profile of energy consumption of consumers with and without power generation. Their objective was to develop a Neural Network (NN) for predicting energy demand based on smart grid consumption patterns and profiles. For this purpose, the profiles were created using a Self-Organizing Map-based pooler (SOM) and an autoregressive NN for daily prediction of energy consumption. As for the power generation properties, a feed-forward NN was used to predict the consumption. Results were obtained in a dataset consisting of 212 consumers, with approximately 1,900,000 hourly energy consumption registers from several devices. Clustering with SOM produced better results than with k-means. The consumption prediction of consumers without power generation resulted a correlation coefficient of 0.99429, Mean Squared Error (MSE) of 0.0046 and a Mean Absolute Percentage Error (MAPE) of 4.21%. Similar results were observed for consumers with power generation, in which the correlation coefficient was 0.999 and the MSE was 0.04.

Jiang et al. [8] proposed a fuzzy clustering model to categorize consumers and identify their energy consumption characteristics. The identification of such characteristics is done after each customer's consumption series are individually grouped into similar parts to detect consumption patterns. Then, the fuzzy clustering generates groups of customers with the same consumption profile, from which the features of consumption patterns are extracted. Finally, a classifier is used to categorize new consumers into one of the previously found groups. Results were obtained for a dataset of hourly energy consumption from 1,168 non-residential consumers over a year. The authors concluded that their fuzzy clustering-based method improved classification accuracy for the inclusion of new consumers.

The prediction of daily energy consumption in apartments in the Republic of South Korea appears as a problem in the work of Wahid and Kim [9]. In this work, K-Nearest Neighbors (KNN) was used as a predictor of energy consumption over hourly consumption data from 520 apartments. From the consumption history, four features were extracted: average, variance, asymmetry and kurtosis. An accuracy 95.96% was obtained as best result.

Lora et al. [10] compared the energy price time series prediction performance of two models, one based on a multilayer perceptron recurrent neural network, and the other based on a combination of KNN and Genetic Algorithm (GA). They used GA to adjust the weights for Euclidean distance. The performances of both models were compared in a small dataset of energy prices from January to August 2001, in which was obtained in the period from March to May an MSE of 0.3464, and in the period between June and August an MSE of 0.428.

Poloczek et al. [11] and Kim et al. [12] used KNN to predict the values lost in the process of sensor data acquisition, due to inactivity. Both works showed that KNN is able to produce results close to real values, using both the proximity of the data values and the sensors' spatial information. These results motivated us to use both the spatial information of energy-consuming facilities and KNN, for its simplicity in data generation.

Most of the aforementioned works use consumption readings automatically acquired on smart meters, which are less sensitive to noise caused by acquisition mistakes. Our proposed work, however, makes use of a dataset which acquisitions were made manually in electromechanical and digital meters and, as a consequence, are subject to mistakes during the readings of energy consumption. Therefore, the presence of noise makes the task of forecasting consumption more challenging.

It is worth mentioning that only a small group of clients use smart meters, which explains the small number of clients in these works. In our work, though, we use a large customer dataset with more than one year of energy consumption data. Moreover, it can be observed that in Jiang et al. [8] there is a need for initial consumption data for new customers, so that they can be inserted in a group. This is necessary to enable the use that group's model in future predictions. We highlight that this restriction is not present in our work.

Additionally, the works of Wahid and Kim [9], Lora et al. [10], Poloczek et al. [11], and Kim et al. [12] perform data prediction using only KNN and spatial information due to its excellent performance in the data regression process. Similarly, our work uses KNN, spatial information, and energy consumption data to estimate new customer consumption. However, it differs from the mentioned studies in that it uses the best prediction result among three proposed regressors, obtained in eight distinct classes of consumers.

III. Materials and method

This section describes the materials and the proposed method for consumption estimation of new installations. The steps of the proposed method are presented in the sequence they are applied, as illustrated in Figure 1. First, the process of data acquisition is described. Second, the dataset goes through a preprocessing step. Third, the neighborhood of each customer is determined. Fourth, consumption of the customer's neighbors is predicted. Fifth, consumption of customers with short series is predicted. And, finally, results are validated in step six.

A. Data acquisition

The dataset consists of power consumption data from 2,316,760 active customers from the state of Maranhão, Brazil. The data was collected monthly from January 2017 to April 2019, and form the series consumption history for each customer.

Customers are organized into classes and subclasses, according to ANEEL Normative Resolution no. 414/2010 [13], repealed by ANEEL Normative Resolution no. 800/2017 [14]. The consumption classes applicable to consumers are:

- **Residential**: this category includes consumer units with residential purposes;

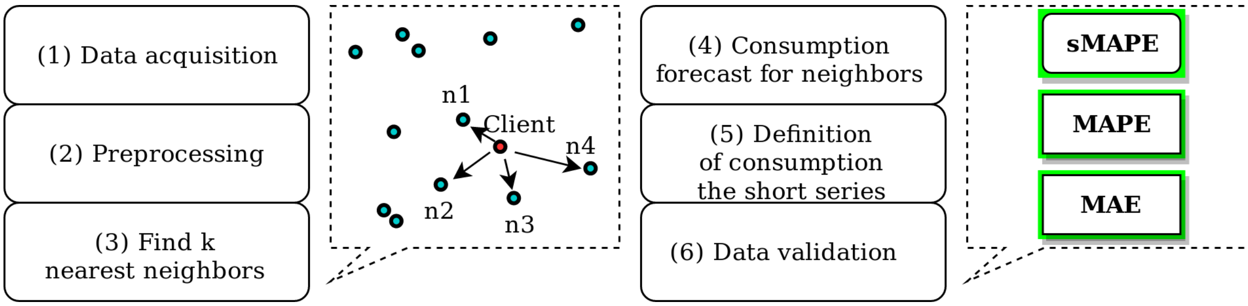

Figure 1. Steps of the proposed method.

- **Industrial**: are the consumer units in which industrial activity is developed;
- **Commercial, services and other activities**: this includes the consumer units where the service rendering activities are developed and others not provided for in the other classes;
- **Public service**: consumer units are intended exclusively for the supply of engines, machinery and cargo essential for the operation of public water, sewage, sanitation and urban or railway traction services, operated directly by the Government or through concession or authorization;
- **Self-Consumption**: the consumer units owned by the distributors are included;
- **Rural**: consumer units that develop activities of agriculture, livestock or aquaculture;
- **Government**: consumer units that are consumers of a legal entity governed by public law are independent of the activity developed, including illumination on roads and traffic lights, radars and traffic monitoring cameras, except for those classified as public irrigation services, schools, agrotechnics, street lighting and public service;
- **Public Lighting**: public service whose sole purpose is to provide clarity to public places on a periodic, continuous or occasional basis.

B. Preprocessing

Before the short-series consumption forecast step, the dataset is subjected to a preprocessing step. In the dataset there are customers with short series, ranging from zero to four months of registered consumption, totaling 95,052. For this reason, this data is separated and used in the tests of the proposed method. In addition, we ignore consumer series that: (1) do not have at least two neighbors to be considered in their estimation, i.e., series categorized in classes that have less than threes installations; (2) cases where clients do not have valid coordinates, which makes it impossible to identify their location and distance to their neighbors; and finally, (3) cases in which the series do not have consumption registered in the reference month of this study, making it impossible to validate the estimated consumption.

C. Finding K-Nearest Neighbor

The k-nearest neighbors are defined based on the customer's geographic coordinates. Fig. 2 shows the information available on each neighbor, which is the prediction of consumption for the reference month of June 2018, the prediction interval with minimum and maximum (after III-D); the coordinates with latitude and longitude; and the Reading Unit (RU). The reading unit represents a set of installations that are read by a particular reader on a reading day. RU information is used to narrow the search scope of K-neighbors. Instead of searching for K-Neighbors throughout the municipality, only the neighbors belonging to the same RU of the analyzed series are searched.

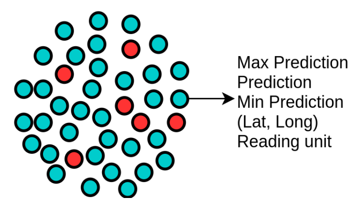

Figure 2. New customers in red and their neighborhood in green.

On the other hand, in the case of a larger number than K-neighbors are available with the same coordinate, only the k-neighbors with the greatest consumption history will be selected.

D. Consumption forecast for neighbors

Consumption forecast and neighbors minimum and maximum prediction interval were performed for the reference month of June 2018. Consumption was estimated using statistical methods (STL [15] and ARIMA [5]) and methods based on machine learning (BAQR) [16]. These methods were empirically chosen based on tests because they outperform most classes over other methods, such as LSTM [17], SGD [18]. STL was better for the Industrial and Self-consumption class, ARIMA in the Public Lighting and Public Power class and BAQR was superior in the Residential, Rural, Public Service and Commercial classes.

1) Boosting Additive Quantile Regression (BAQR): It is a quantile regression model that uses additive models to relax the assumption of linearity [16][19]. It is based on the following equation:

$$\hat{q_\alpha}\left(x_1^{(i)},...,x_d^{(i)},z_1^{(i)},...,z_J^{(i)}\right) = \beta_0 + \sum_{j=1}^{d} \beta_j x_j^{(i)} + \sum_{j=1}^{J} g_j(z_j^{(i)}) \tag{1}$$

Copyright (c) IARIA, 2020. ISBN: 978-1-61208-762-7

where g_j is a variable smoothing function $z_j^{(i)}$. The method uses the boosting technique to estimate the model by minimizing a loss function using the descending gradient method.

2) STL: According to Cleveland et al. [15], STL is a filtering procedure that decomposes the series into Trend (T), Seasonality (S) and Error (E) components. Thus, the original series can be formed by the sum of these components. So, seasonality determines the existence of a cyclic pattern in a time series and the trend is characterized by the behavior of growth or decrease in the long-term amplitude of the time series. Therefore, the power consumption series was filtered using STL and the result was used to choose two regressors: *Simple Exponential Smoothing* (SES) and *Holt's Linear Smoothing*(Holt).

The first regressor, SES [20], shows better results when the dataset has no trend or seasonality, while the second, Holt [21], when there is a trend but no seasonality. Therefore, because these regression methods were input using the result of STL filtering, this method was named STL.

3) Autoregressive Integrated Moving Average - ARIMA: It is defined as a generalization of the Autoregressive Moving Average (ARMA) model. These models are generally applied to non-stationary data because, through the differentiation step, this data is transformed into stationary data [5].

E. Definition of consumption of the short series

This section presents the method for estimating the k-nearest neighbors of the new customer. Consumption estimation for new, short-series, non-consumer customers depends on the discovery of k-nearest neighbors, as well as the consumption estimation and prediction interval of the k-nearest neighbor which is described in the Section III-D.

In the study, the developed methodology considers four scenarios for consumption estimation and prediction interval. The first scenario is made for customers without consumption history and the other scenarios are made for customers with short series. June 2018 is used for validation of results only, not entering each client's series size count. Therefore, customers who have only the June 2018 validation reference month, for example, have zero-size series.

The first scenario considers only customers who have no history of consumption. The second scenario considers customers who have a single month of consumption. The third scenario considers customers who have two months in their consumption history. And finally, the fourth scenario considers customers who have three or four months in their consumption history.

For the first scenario, in which the customer has no history, the consumption estimation and prediction interval are generated from its vicinity, as shown in Fig. 3. In this figure, in (1) are the nearest neighbors of the new customer, all belonging to the same RU, which is defined by the company. In (2), the median of consumption predictions and prediction intervals of the k-nearest neighbor are calculated.

For the second scenario, the approach repeats the previous month's consumption for the prediction of consumption. For the third scenario, the approach calculates the median consumption of previous months to predict consumption. However, the consumption interval estimation of these scenarios continues to be made by estimating the neighborhood consumption interval.

In scenarios two and three, a problem was encountered regarding the generated consumption interval. In some cases, the prediction based on previous customer consumption may be outside the range generated by neighbors. This problem was mitigated as follows: the ratio between the upper limit and the estimated customer prediction found from the neighborhood method is obtained and this factor is multiplied by the customer's predicted consumption, due to its short history of consumption.

F. Validation of Results

The proposed method was evaluated using the MAPE, Symmetric Mean Absolute Percentage Error (sMAPE) and Mean Absolute Error (MAE). These metrics are commonly used to evaluate value estimation techniques. The lower the error metrics, the better.

The MAPE is a percentage relative error, as in (2), that expresses how much the absolute error between the real value (y_i) and the predicted value ($\hat{y}_i$) is greater than the real value, for a point i in the time series. According to Yorucu [22], a prediction with MAPE percentage below 10 % is interpreted as highly accurate; forecast greater than 10 % and less than 20 % is interpreted as good; forecast greater than 20 % and less than 50 % is reasonable; and prediction greater than 50 % is considered inaccurate.

$$MAPE = \frac{1}{n}\sum_{i}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2}$$

The sMAPE is a percentage measure of prediction errors according to (3) and indicates how much the observed error is greater than the sum of the modules of the real value (y_i) and the predicted value ($\hat{y}_i$), for the N available points. Therefore, the closer to zero, the better the prediction.

$$sMAPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|\hat{y}_t - y_t|}{|\hat{y}_t| + |y_t|} \tag{3}$$

MAE is simply the average of the absolute values of errors, i.e., the differences between the real value (y_i) and the predicted value ($\hat{y}_i$), according to (4).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{4}$$

The proposed validation metrics are commonly used in prediction tasks. Thus, although the MAE is useful to verify the magnitude of the errors found along the predictions, the visualization together with MAPE and sMAPE is very important to understand the performance of the proposed method.

IV. Results and Discussion

This section presents the results of the first experiments needed for parameter definition, case studies and the final result of the proposed method.

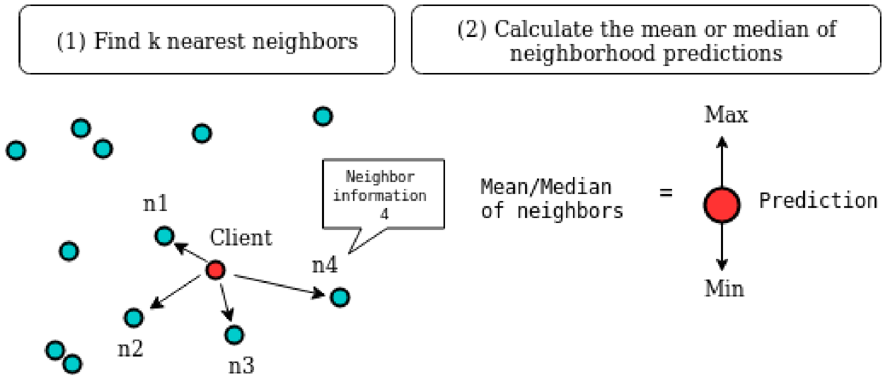

Figure 3. First scenario: a new customer with no history.

A. First experiments

The first experiments were performed in the data set of the Residential class of a municipality X of Brazil. This dataset contains 107,738 customers. The objective of these experiments was to verify the influence of the calculation of distances on the results for the first scenario, where both consumption and interval predictions are generated only by the nearest neighbors. Thus, the Euclidean and Manhattan distances were tested. First, 10 % of these customers were randomly separated. We then simulated short series using only the customer's first four months and applied the consumption and interval estimation method described for the first scenario (III-E, using the predicted consumption median of the 10 nearest customers. Next, the result found for the Manhattan distance was an sMAPE of 35.4 % and the Euclidean distance obtained an sMAPE of 35.42 %. Despite the slight difference, the distance from Manhattan was chosen to be applied in real cases.

1) Study of case: To exemplify tests performed in the first experiments during the development of the methodology, specific cases to be analyzed were removed.

Fig. 4 presents a case considered highly accurate, according to Yorucu [22], with a 4.28 % sMAPE, an 8.20 % MAPE and a 6.14 MAE. In this case, the customer has, in the reference month analyzed, real consumption of 75 kWh and had a predicted consumption of 68.85 kWh. The neighborhood consumption range relative to the customer is close, resulting in a closer prediction of the real.

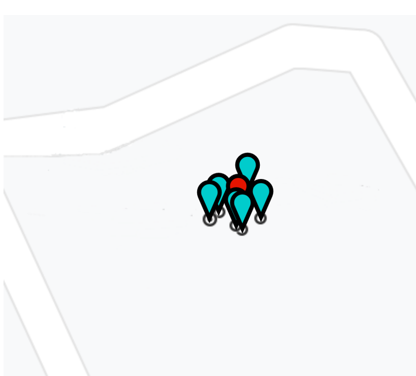

Figure 4. First case study: high accuracy of consumption prediction.

Fig. 5 presents a bad case, where the customer's consumption range is much lower than its neighborhood consumption range. In this case, the real customer consumption is 7 kWh, but the predicted was 358.21 kWh, with an sMAPE of 96.17 %, a MAPE of 351.21 % and an MAE of 5,017.29.

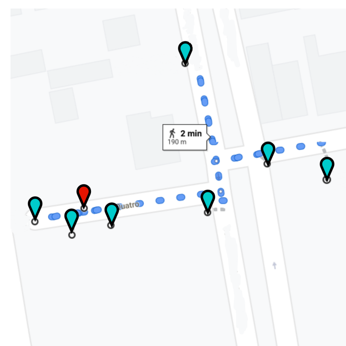

Figure 5. Second case study: low accuracy of consumption prediction.

B. Results of the proposed method

This section describes the results of the experiment conducted with a short series of clients. In total, 86,874 customers remained after the preprocessing. In this experiment, the distance from Manhattan was used as a metric of the proximity of the points and ten neighbors closest to the new client were experimentally defined.

Table I shows the results by class of the methods chosen to predict neighborhood consumption, from the stage prior to the consumption forecast for new customers. The metric used to choose the methods was MAPE and each line represents a class with their respective values in each method used. From this same table, it is possible to verify that BAQR was the most used method among classes.

Table II presents the Quantity (QTY) of customers separated by class, which were processed by the proposed method and the results obtained in the consumption estimation for the short series. The residential class, with 79,871, contains the largest number of customers. While the lighting public class contains only 12. The best results were obtained with the sMAPE metric, as it is a symmetric metric, i.e., it limits the extreme effects, as well as avoiding null results. However,

Copyright (c) IARIA, 2020.

TABLE I. RESULTS BY CLASS OF THE THREE BEST METHODS CHOSEN TO PREDICT NEIGHBORHOOD CONSUMPTION.

	Methods		
Classes	BAQR	STL	ARIMA
Residential	**36.20%**	47.00%	57.10%
Industrial	41.80%	**32.90%**	62.10%
Commercial	**32.60%**	61.90%	157.00%
Rural	**38.90%**	48.70%	203.70%
Government	129.40%	77.50%	**57.90%**
Lighting public	29.50%	32.90%	**25.10%**
Public service	**31.30%**	337.80%	229.70%
Self-Consumption	14.40%	**10.00%**	12.70%

when analyzing the prediction results in terms of MAPE and MAE, these were not considered ideal but promising given the diversity of scenarios and consumption classes found in the dataset.

MAPE had the highest values where consumption patterns are highest. For example, in the class Self-consumption, with minimum consumption of 0, the average is 15,596 and a maximum of 82,740 kWh.

TABLE II. RESULTS OF VALIDATION METRICS BY CONSUMPTION CLASS.

CLASSES	QTY.	sMAPE (%)	MAPE (%)	MAE	Within range (%)
Residential	79,871	30.92	1,136.56	490.05	96.14%
Industrial	76	26.57	133.29	379.28	94.74%
Commercial	4,032	36.42	301.43	1,292.05	94.10%
Rural	1,475	33.17	614.16	3,472.73	93.42%
Government	577	25.15	91.11	443.54	93.41%
Lighting public	12	22.71	45.12	667.67	91.67%
Public service	98	41.57	200.91	1,548.66	83.67%
Self-Consumption	733	22.69	447.65	86,477.52	89.50%

Table III presents the percentage distribution of customers to sMAPE and MAPE percentage interval. The largest number of customers (37.39 %) were found with up to 10 % of the sMAPE metric. Analyzing sMAPE according to Yorucu's classification [22], 37 % of customers presented a highly accurate forecast, 16 % good, 22 % reasonable and only 23 % of customers presented an inaccurate forecast.

TABLE III. COSTUMERS DISTRIBUTION ACCORDING TO SMAPE AND MAPE METRIC VALUE RANGE.

	Percentage of customers by metric	
Range	sMAPE (%)	MAPE (%)
[0 - 10%[	37.39	28.70
[10% - 20%[	16.57	15.16
[20% - 50%[	22.89	23.43
[50% - inf[	23.15	32.70

During the results analysis, it was observed that close customers may have characteristics in common, such as purchasing power, consequently the same consumption pattern. However, there are areas where customer consumption does not have a common range, e.g., a new customer A, with a consumption range of 10 to 50 kWh, while its vicinity has a range of 100 to 300 kWh. Thus, it was found that new customers usually have low consumption in the first months or zero consumption, which was often incompatible with the consumption of their neighbors, who already have a stable consumption pattern. Therefore, this situation contributes to the increase of the prediction error.

Likewise, customers whose neighborhood may be located in a border region between two neighborhoods with different consumption characteristics, may experience an increase in the prediction error, since only the distance is used to determine that neighborhood.

The consumption prediction interval for new customers was not previously found by the company. Thus, the proposed method appears as an important tool for theses cases, due to the reasonable accuracy obtained. As a result, at least 83 % of the real cases were within the generated interval, in the worst case, as shown in the Table II, in Public service class.

This result is significantly relevant since this method, for this example, would prevent around 92 % (79,900) of these customers from going into the billing sector, avoiding the need for manual analysis of the recorded consumption for these customers. In the current practice of the company, all customers with short series end up going to this sector.

V. Conclusion

In the present work, a method of prediction of power consumption for consumers with short or no consumption history was proposed. The method made use of machine learning techniques such as k-nearest neighbors, different distance measurements and various regressors such as SGD, STL, ARIMA, BAQR, LSTM, which are used to predict the energy consumption of consumers had their installation recently connected.

The proposed methodology used several approaches, such as Euclidean distance and Manhattan distance to find k-nearest neighbors; different regressors; and the median predicted neighborhood consumption. From the evaluated approaches, the distance from Manhattan showed a small advantage over Euclidean and among the regressors, the best estimates were made with STL, ARIMA, and BAQR.

From the above, it can be concluded that the neighborhood-based estimation method for power consumption is a promising method for new consumers, with no consumption history yet. In addition, it was possible to observe from the two selected case studies, that the proximity to neighbors results in a good result of prediction of consumption and its corresponding consumption interval, according to the first case study. The opposite, that is, when these neighbors are distant from each other, resulted in a low accuracy of prediction and their consumption range, as shown in Fig. 5, of the second case study. Therefore, the results presented contribute to reduce the volume of customers that need to be analyzed by the company.

Although promising, the method can be improved by utilizing other network-based serial data estimation techniques such as TCN [23] and N-BEATS [24].

Acknowledgments

The authors would like to thank Equatorial Energy for the financial support provided through the National Electric Energy Agency (ANEEL) Research and Development Program (R&D), PD-00037-0036/2019.

References

[1] A. Mosavi, M. Salimi, S. ardabili, T. Rabczuk, S. Shamshirband, and A. Varkonyi-Koczy, "State of the art of machine learning models in energy systems, a systematic review," Energies, vol. 12, 04 2019.

[2] J. Schneider, M. Dziubany, A. Schmeink, G. Dartmann, K.-U. Gollmer, and S. Naumann, "Chapter 8 - predicting energy consumption using machine learning," in Big Data Analytics for Cyber-Physical Systems, G. Dartmann, H. Song, and A. Schmeink, Eds. Elsevier, 2019, pp. 167 – 186, [accessed: 2019-10-10]. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128166376000087

[3] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," Economic Geography, vol. 46, no. sup1, 1970, pp. 234–240, [accessed: 2020-01-30]. [Online]. Available: https://www.tandfonline.com/doi/abs/10.2307/143141

[4] A. M. Alonso, F. J. Nogales, and C. Ruiz, "A single scalable lstm model for short-term forecasting of disaggregated electricity loads," 2019.

[5] K. M. Vu, The ARIMA and VARIMA time series: their modelings, Analyses and Applications. AuLac Technologies Inc., 2007.

[6] P. Goodwin, "Using naïve forecasts to assess limits to forecast accuracy and the quality of fit of forecasts to time series data (working paper)," 11 2014.

[7] A. Bâra and S. V. Oprea, "Electricity consumption and generation forecasting with artificial neural networks," in Advanced Applications for Artificial Neural Networks, A. El-Shahat, Ed. Rijeka: IntechOpen, 2018, ch. 7, [accessed: 2019-10-16]. [Online]. Available: https://doi.org/10.5772/intechopen.71239

[8] Z. Jiang, R. Lin, and F. Yang, "A hybrid machine learning model for electricity consumer categorization using smart meter data," Energies, vol. 11, no. 9, 2018, [accessed: 2019-10-15]. [Online]. Available: https://www.mdpi.com/1996-1073/11/9/2235

[9] F. Wahid and D. Kim, "A prediction approach for demand analysis of energy consumption using k-nearest neighbor in residential buildings," International Journal of Smart Home, vol. 10, 02 2016, pp. 97–108.

[10] A. T. Lora, J. R. Santos, J. R. Santos, J. L. M. Ramos, and A. G. Exposito, "Electricity market price forecasting: Neural networks versus weighted-distance k nearest neighbours," in Database and Expert Systems Applications, A. Hameurlain, R. Cicchetti, and R. Traunmüller, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 321–330.

[11] J. Poloczek, N. A. Treiber, and O. Kramer, "Knn regression as geo-imputation method for spatio-temporal wind data," in International Joint Conference SOCO'14-CISIS'14-ICEUTE'14, J. G. de la Puerta, I. G. Ferreira, P. G. Bringas, F. Klett, A. Abraham, A. C. de Carvalho, Á. Herrero, B. Baruque, H. Quintián, and E. Corchado, Eds. Cham: Springer International Publishing, 2014, pp. 185–193.

[12] M. Kim, S. Park, J. Lee, Y. Joo, and J. K. Choi, "Learning-based adaptive imputation methodwith knn algorithm for missing power data," Energies, vol. 10, no. 10, 2017, [accessed: 2019-10-16]. [Online]. Available: https://www.mdpi.com/1996-1073/10/10/1668

[13] N. E. E. Agency. Normative resolution no. 414/2010. [Online]. Available: www.aneel.gov.br/cedoc/ren2010414.pdf [retrieved: oct, 2019]

[14] ——. Normative resolution no. 800/2017. [Online]. Available: http://www2.aneel.gov.br/cedoc/ren2017800.pdf [retrieved: oct, 2019]

[15] R. B. Cleveland, "Stl : A seasonal-trend decomposition procedure based on loess," 1990.

[16] S. B. Taieb, R. Huser, R. J. Hyndman, and M. G. Genton, "Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression," IEEE Transactions on Smart Grid, vol. 7, no. 5, 2016, pp. 2448–2455.

[17] D. M. Nelson, A. C. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with lstm neural networks," in 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, 2017, pp. 1419–1426.

[18] S. Ruder, "An overview of gradient descent optimization algorithms," arXiv preprint arXiv:1609.04747, 2016.

[19] D. Kraus and C. Czado, "D-vine copula based quantile regression," Computational Statistics Data Analysis, vol. 110, 2017, pp. 1 – 18, [accessed: 2019-10-15]. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167947316303073

[20] R. J. Hyndman, G. Athanasopoulos, and OTexts.com, Forecasting : principles and practice / Rob J Hyndman and George Athanasopoulos, print edition. ed. OTexts.com [Heathmont?, Victoria], 2014 2014.

[21] B. Etienne, "Time series in python - exponential smoothing and arima processes," Mar 2019, [accessed: 2019-10-11]. [Online]. Available: https://towardsdatascience.com/time-series-in-python-exponential-smoothing-and-arima-processes-2c67f2a52788

[22] V. Yorucu, "The analysis of forecasting performance by using time series data for two mediterranean islands," Review of Social, Economic & Business Studies, vol. 2, 2003, pp. 175–196.

[23] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "Lstm fully convolutional networks for time series classification," IEEE Access, vol. 6, 2017, pp. 1662–1669.

[24] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-beats: Neural basis expansion analysis for interpretable time series forecasting," arXiv preprint arXiv:1905.10437, 2019.

[25] M. Laurinec, P. & Lucká, "Clustering-based forecasting method for individual consumers electricity load using time series representations," Open Computer Science, no. 8, 2018, pp. 38–50.