Available online at http://www.journalijdr.com

IJDR

ISSN: 2230-9926

International Journal of Development Research
Vol. 10, Issue, 09, pp. 40905-40915, September, 2020
https://doi.org/10.37118/ijdr.20089.09.2020

RESEARCH ARTICLE **OPEN ACCESS**

MULTIMODEL NEURAL NETWORKS USING SOCIOECONOMIC VARIABLES FOR THE PREDICTION OF RESIDENTIAL ELECTRIC CONSUMPTION

Daniel Orlando Garzón Medina*, **Jose Calixto Lopes Junior and Thales Sousa**

Federal University of ABC, Santo Andre, SP, Brazil

ARTICLE INFO

Article History:
Received 18th June 2020
Received in revised form
14th July 2020
Accepted 20th August 2020
Published online 30th September 2020

Key Words:
Artificial Intelligence, Artificial Neural Networks, Load Forecasting, Multi-Model Forecasting, Socioeconomic Variables

Corresponding author:
Daniel Orlando Garzón Medina

ABSTRACT

This work intends to focus on the use of Multimodel Artificial Neural Networks (ANNs) for the projection of the residential electric demand, taking into account that this is the basis for an adequate planning of electricity distribution networks. The ANNs were developed in MATLAB®, trained according to the data recorded, and the final results of the different regions of study were compared with the official data provided by the UPME for the year 2017. Models of 2-layer ANNs capable of accurately predicting medium-term residential electrical consumption were designed, taking into account variables such as GDP per capita, population, residential electrical consumption and temperature. The ANNs were found and suggested as a model capable of incorporating the nonlinearities of the different study variables, in addition to having no complexity for the planner in their mathematical modeling. Thus, in addition to estimate the degree of precision of the forecast used, it is sought to achieve a high degree of accuracy in the decisions, taking into consideration that the increase of residential users and load are important topics for the energy supply companies in the next decade.

Citation: ***Daniel Orlando Garzón Medina, Jose Calixto Lopes Junior and Thales Sousa.*** "Multimodel Neural Networks Using Socioeconomic Variables for the Prediction of Residential Electric Consumption*", International Journal of Development Research*, 10, (09), 40905-40915.

INTRODUCTION

The significant increase in electricity consumption in urban areas, associated with major technological developments, has brought to the Electric Power System (EPS) the need for rapid modernization. Thus, in the current scenario, there have been major changes that incur uncertainties and new concerns regarding the maintenance of robustness and the ability to meet the demand for electricity, which in turn has been behaving in an increasingly dynamic manner. All of these factors make the planning and forecasting of electrical consumption of paramount importance in the electricity sector, being the focus of many studies. Derived from the Latin '*praedictĭo*', the word forecast means to see in advance, to discover. In other words, it is an expression used in the sense of predicting what is going to happen or what may happen. In the field of forecasting electricity demand, decisions are always surrounded by interconnected factors such as: investments or their postponement in infrastructure, interest of shareholders, consumer needs and the development of the country's economy.

There is currently no standardized classification for defining the planning period. In the study developed by (Hong and Fan, 2016) planning horizons were grouped into 4 groups, namely: Very Short Term Forecast (VSTLF), Short Term Forecast (STLF), Medium Term Forecast (MTLF), and finally Long Term Forecast (LTLF). The period considered for each horizon is 1 day, 2 weeks, and 3 years, respectively. In order to assist energy planning, mathematical models are proposed with the objective of predicting future consumption and, thus, assist in deliberations that reflect the period to be planned. In this sense, models based on Artificial Intelligence (AI) have been widely used where, among many techniques, the Artificial Neural Networks (ANN) stand out. ANNs have wide application in areas such as science, technology, economics and industrial processes. From the application of AI techniques in demand forecasting, great capacity for handling nonlinear functions has been achieved, which has the advantage of not requiring complex mathematical formulations or quantitative correlation between inputs and outputs (Ramirez, 2013). The first work on load forecasting used statistical methods that linearly related the different variables of a given model.

However, such methods resulted in forecasts with relatively high associated error and, often, lack of convergence when modeling variables (Zhang *et al*, 1998). The first analysis of ANN behavior was introduced in 1943 by Warren McCulloc and mathematician Walter Pitts, where the beh behaviour avior of neurons in the human central nervous system was modeled, using electrical circuits. The concept of neural learning took hold when, in 1949, Donald Hebb presented the concepts of learning, the paths and interactions between neurons, and the essence of human learning (Park *et al*., 1991). One of the great challenges of load forecasting based on mathematical models is related to precision. Accuracy is a factor of great importance, being difficult to be achieved, especially when there are variables based on external factors such as the occurrence of days with extreme weather, weekends and atypical days (of behavior outside the normal pattern analyzed initially). With the recent development of mathematical tools, data mining and analysis and AI, it has been possible to significantly improve the obtained results (Liu *et al*., 2017).

Due to the lack of precision of some traditional models, AI-based models have become a preference for researchers in the field. In (Singh and Dwivedi, 2018) a genetic algorithm was presented combining RNA and optimization techniques capable of finding the optimal network parameters that reduce errors in load predictions. The authors selected data from the interconnected system of New England, New South Wales, and Texas, in order to find an approximate error of 4.86% for the year 2015.

ANNs have been successfully applied to forecast electric consumption in the short and long term planning period (Shamsollah, 2001). Since they have important characteristics, such as the ability to learn through training, providing sufficient adaptability to produce appropriate responses to various problems in any area of science, this technique has attracted the attention of researchers (Mordjaoui *et al*., 2017).

Different authors have been working with the Feedforward Multilayer Perceptron (MLP) model for load forecasting, among which are (Shamsollah, 2001), (Shamsollah, 2001), (Duda *et al*., 2000) and (Jasiński, 2019). The main characteristic of the models employed is based on the inclusion of 1 or 2 hidden layers due to the precision in the final errors. The results obtained by the authors were considerably accurate, where associated errors were obtained, which ranged from 1 to 4%. Nowadays, MLP model has been worked by many authors due to the good quality of the results, although it requires a longer computational time (Duda *et al*., 2000). This network model was also considered by (Jasiński, 2019) and it can be observed that obtaining satisfactory results is based on the learning process of the network.

In (Miranda *et al*., 2018) a load forecast was developed based on electricity pricing using RNAs and Fuzzy Logic. The authors worked with a modified MLP in order to correlate changes in consumption and changes in energy tariffs. The best configuration of the RNA used by the authors was the composition with 21 hidden layers, employing different inputs such as temperature, population, among others. Compared with the actual data from the New England interconnected system, the results were considered to be accurate, remaining below 3.9%. In order to obtain a greater degree of assertiveness in the forecasts, the short and medium term models need to incorporate regional climatic variables, since they influence residential electricity consumption (Barman *et al*., 2018). For example, in (Kuster *et al*., 2017) a short term forecast using temperature variations was developed. The authors presented the influence of temperature on consumption in Turkey using a model based on ANN. The models presented by the authors are compared using Mean Absolute Percentage Error (MAPE), with errors of less than 5%. Currently, it is possible to find different models capable of predicting electrical consumption. Among the models most found in the literature, models based on time series are presented (Kuster *et al*., 2017), econometric models (Mohammad Zadeh and Masoumi, 2010), and models based on Artificial Neural Networks (Raza and Khosravi, 2015), (Medina *et al*., 2019). Finally, (Silva *et al*., 2019) presented a new methodology for load forecasting through Bayesian Networks and hierarchical linear models, considering different energy efficiency scenarios. To carry out the tests, a case of study was conducted in a paper industry, where it was possible to perceive the behavior of the model and the probability distribution of the forecast for the period between the years 2015 and 2050. ANNs do not require complex modeling to correlate the different variables that will be used in the prediction of electrical consumption, as explained by (Nichiforov *et al*., 2017), which is a big advantage. The different works referring to ANN agree that it is a powerful method, guaranteeing low errors although, sometimes, it requires a high computational effort, depending on the data records used for the training stage and the optimization algorithm used in the model.

Considering the above, the present work employs MLP-type ANNs to make a medium-term load forecast for the residential electricity sector. Historical data on energy consumption, population growth and Gross Domestic Product (GDP) were used, using the Levenberg-Marquardt (LM) algorithm for the training stage. In this way, the followed methodology was implemented presenting a stratified-disaggregated forecasting, showing low errors in comparison to real values of the residential consumption of the year 2017. For the stages of validation and test of the model, it was considered as a study case the Colombian residential sector. The rest of the paper is organized as follows. Section 2 presents the methodology and the chosen variables for the study case. Section 3 shows the results and discussion of the 3 simulated stratums. Concluding remarks are contained in Section 4.

Perceptron Multilayer Networks: The process of learning an ANN basically consists of three stages: training, validation and testing. First, it is necessary to teach the network the standards it needs to foresee. The second step is designed to determine when the learning process should be interrupted, which is accomplished by feeding the model with a smaller set of information. It is important to supervise the network so that it is not overly trained, which would imply that the network loses the ability to learn, and starts to simply memorize, compromising the results. Finally, in the third stage, the objective is to assess the quality of the trained network. It is important to emphasize that, in order to avoid overfitting problems in the training stage, it is necessary to have an arrangement of reliable and well-distributed information for the three steps previously mentioned (Kermanshahi and Iwamiya, 2002). Figure 1 illustrates the steps previously described. In Figure 1, X_{CON}, X_{POP} e X_{GDP} represent the set of input variables used. The interactions between the hidden layers and the input layers are represented by the junctions between the nodes of the two layers, representing a set of weights for each interaction (w_m); the activation level is represented as the weighted sum of the input vector; and the activation function *f*, which represents a non-linear function applied on the activation level in order to produce an output signal ($f \varepsilon Z \rightarrow \neq 0$).

Thus, the mathematical expression that represents the set of outputs (Yt) depending on the inputs is given by (1):

$$Y_t = f[w^t.x] = f[w_1.x_1] + [w_2.x_2] + [w_3.x_3] + \cdots [w_m.x_n] \quad (1)$$

Levenberg-Marquardt Algorithm: For the present work, the Levenberg-Marquardt (LM) algorithm was chosen for training the network. It is an optimization technique that demands a high amount of memory for long historical series and multiple variables as well as relative functional complexity, which can be a disadvantage for some large applications. On the other hand, this algorithm is mathematically designed to minimize the aggregate mean square error in all training standards, being of great use for applications in different areas (Beale *et al*., 2016). When using the LM algorithm, the weights of the network are adjusted as shown in (2):

$$\Delta w_{ji}(t) = -\left[\nabla^2 E\left(w_{ji}(t)\right) + \mu I\right]^{-1} \nabla E(w_{ji}(t)) \quad (2)$$

where $\nabla^2 E\ (w_{ji}(t))$ is a Hessian matrix and $\nabla E\ (w_{ji}(t))$ is the gradient. From (2), it is worth noting that when a high value of μ is used, the algorithm becomes a descending gradient. On the other hand, if the value of μ is low, the algorithm is equivalent to the Gauss-Newton algorithm (Braga *et al*., 2007).

Quality Indicators used in Load Forecasting: Many of the authors have selected MAPE and Mean Square Error (MSE) to assess the accuracy of the models (Onoda, 1993), (González-Romera *et al*., 2008), and (Singh and Dwivedi, 2017). It is noteworthy that both MSE and MAPE are widely used in comparative analysis due to their statistical representation of relative simplicity, being preferred among other error indicators. Both can be obtained from (3) and (4), respectively:

$$MSE = \frac{\sum_{i=1}^{n}(Y_{m_i} - Y_{e_i})^2}{n} \quad (3)$$

$$MAPE = \frac{\sum_{i=1}^{n}(Y_{e_i} - Y_{m_i})}{Y_{ei}} x100 \quad (4)$$

where Y_m represents a vector of n predictions, and Y_e a vector of true values.

METHODS

The algorithm used for the development of the methodology is presented in Figure 2, where it represented the process to be repeated for each neural structure that will be used in the case of study that will be described below. For the development of the proposed methodology, an adaptation of the methodology developed by (Ahmad *et al*, 2016) for the medium term forecast was presented. The authors presented a review of the different methodologies based on computational intelligence, describing the appropriate process to treat the data before and after the ANN training and validation process focused on Heating, Ventilation and Air-Conditioning (HVAC) systems. The necessary adaptations for use in the forecast of residential electrical consumption were: the choice of the forecast type; acquisition of histories of the different input variables to be considered; selection of the forecasting technique; configuration of network parameters; and analysis of the quality indicators considered (MAPE, MSE).

Thus, the proposed methodology allows to be replicated to develop forecasts of short and medium term residential electrical consumption, being possible to identify a segmented or stratified consumption, as in the case study selected for the present work.

From the algorithm presented in Figure 2, six stages of development can be observed:

- **Step 1:** Selection of the forecast type: short- or medium-term planning;
- **Step 2:** Search and analysis of historical data with which the forecast will be worked. In this second stage, Colombian residential consumption is identified, according to the socioeconomic strata characteristic in the country (DANE, 2019), as well as the division by regions to identify temperature patterns. In addition, historical data on population and GDP per capita are consulted. The selection of the Colombian case study is justified by the availability of access to records through public portals, in addition to the possibility of correlating different socioeconomic variables to stratified electricity consumption;
- **Step 3:** Analysis of the relationship between the variables used in the selected load forecast. This criterion is important in the development of forecasts in the short- and long-term planning period, as it allows identifying the variables that contribute or affect the model. First, the relationship that residential electrical consumption has with the selected variables must be identified. For that, Pearson's correlation coefficient is evaluated for each of the considered variables. The correlation result represents the relationship of the dependent variables with respect to the independent variable (Hernández Lalinde *et al*., 2018). In the context of this step, the planner chooses the type of technique he wants to employ to develop the forecast for residential electrical consumption. For the present case of study, ANNs were used;
- **Step 4:** Adjust network parameters such as search algorithm, hidden layers and activation function. Different ANN configurations are tested and analyzed in order to find the best structure capable of predicting residential consumption. To do so, the data from Step 2 is inserted into a neural network that will be in charge of analyzing the relationships between the variables, identifying a logical medium-term pattern, which will be the result of the load forecast for the Colombian residential sector in the year 2017. Residential electrical consumption is disaggregated, which makes it possible to analyze and forecast the consumption of the different strata. For this part of the analysis, the regional division of Colombia is used, considering temperature as an independent variable. Once the network is trained and validated, the result is denormalized in order to evaluate the result obtained with the actual data for the year 2017, in GWh.

- **Step 5:** Calculation of the corresponding forecast errors in relation to the real values of the year 2017 for the different study cases through MAPE and MSE.
- **Step 6:** Analysis of the results.

Load Curve Neural Model: Architecture: The definition of the architecture of the neural model of the load curve aims to present the general functioning of an MLP ANN, illustrating the advantages of this type of network both in approximating functions and in its ability to predict future behaviors of a time series (Eljazzar and Hemayed, 2016). The determination of the number of neurons in the hidden layers is developed through a heuristic process, where different configurations are tested, looking for the least possible error (Ahmad *et al*., 2014). In this sense, the main objective is to obtain an adequate configuration of layers and neurons that allows ANN to learn the peculiar characteristics of each input variable. On the other hand, a large number of layers can generate an over-adjustment, resulting in memorization instead of learning, implying in inappropriate results for different data sets of employees in training. Otherwise, if the number of layers is insufficient, under-tuning may occur, where the network would be unable to learn mutual relations from the input variables. Therefore, it is essential to test different structures in order to obtain the smallest error in the final result (Beaule *et al*., 2018).

CASE STUDY

To test and validate the model, Colombia was chosen to develop the case study. A characteristic of this locality is that it has stratified data, it means, divided into socioeconomic sectors, which represents an advantage by allowing to analyze the behavior of the variables more independently and less generalized. In addition, the Colombian residential sector has become of great importance to planners, especially due to the significant increase in electricity consumption in recent years. According to official data from the Unidad de Planeación Minero Energética (UPME), there was a 25.72% increase in residential electricity consumption in the period from 2006 to 2016 (UPME, 2019). The forecast year selected was 2017, since this year's information is known, allowing a better view of the accuracy of the model employed.

Colombian Regional Division and Temperature Standards: The Colombian political-administrative division was considered which makes it possible to group together the states that have similar temperature patterns and electrical consumption. This grouping of states in Colombia are called Regions and, according to (Morales, 2019), are characterized by the following weather patterns:

- Caribbean Region: Tropical region with average temperatures of approximately 30 degrees. It is mainly characterized by the presence of coastal plateaus and mountainous groups, such as the Sierra Nevada de Santa Marta;
- Pacific Region: Known as one of the most humid regions on the planet, it is characterized by having high temperature humid zones throughout the year;
- Andean Region: It is the most important region, concentrating 75% of the country's population. It is characterized by a wide variety of climates, mainly due to the division of the Andes into 3 sub-ranges: Cordillera Occidental, Cordillera Central, and finally, Cordillera Oriental;
- Orinoquia Region: Also known as "Los llanos Orientales", it is a region characterized by a hot and dry climate, providing savanna vegetation and natural pastures, in addition to a rich and varied fauna;
- Amazon Region: Region characterized by hot weather and equatorial rains, which allow the birth of a tropical jungle, and predominant fauna.

Figure 3 presents the grouping of states in the five geographic regions of Colombia (Sun *et al*., 2016). The temperature of the regions has a pattern of influence on residential electrical consumption, representing greater complexity when being statistically modeled (Sun *et al*., 2016); (Wang *et al*., 2016), which is why it was important to include it among the input variables of the ANN-based models. On the other hand, it is important to note that the non-linearities of electrical consumption in relation to different temperatures are mainly due to the increase in electrical consumption when the temperature fluctuates (mainly attributed to the use of air conditioning). The study regions are characterized by not having a homogeneous climate, ranging from 12 to 40 degrees.

Based on this limitation, a temperature indicator was obtained taking into account:

- Analysis and determination of homogeneous climatic zones: Based on the average values recorded by the different observatories of the Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) located throughout the national territory, it was possible to identify 5 geographical zones that have temperature patterns similar (UPME, 2016);
- Grouping of 27 of the 32 states within the 5 selected regions: The 27 states are grouped into regions that have a similar temperature pattern, as previously explained. Thus, it is possible to average the temperatures of similar states in order to obtain a unique temperature indicator by region, characterized by having a low error and little discrepancy, such as the procedure adopted by (Moral-Carcedo and J. Vicéns-Otero, 2005).

Another important feature is the absence of seasons such as autumn and spring throughout the year, this behavior being justified by its location on the equator. On the other hand, the summer and winter seasons are intense in a large part of the territory, generating similar behaviors between the Andean-Pacifica and Amazonas-Orinoquia regions.

Stratification of the Colombian Residential Sector and Possible Applications: In many countries, the concept of social stratification is adopted, which means that society is segmented by levels. It is a sociological concept normally used to classify individuals and groups based on their socioeconomic conditions. It is commonly used as a basis for understanding the behavior of society in the characterization of social inequalities. In many countries this concept is applied which makes it possible to group data in a sectorized way, bringing the advantage of making future studies more decentralized. The Colombian socio-economic stratification was born in the 1980s with the aim of distributing state aid equally. The concept of "social stratification" includes the division of society into classes, but this division is established based on criterias such as type of housing and incomes (DANE, 2019). In other words, there are 6 social strata that reflect the country's income levels, with number 1 corresponding to the grouping of the lowest income sectors and, on the contrary, strata 6 the highest income sectors. One strategy of the Colombian government is to take advantage of social stratification to maintain several indicators separated by social class. Thus, many public data can be found divided by stratum. Whether due to the strict need for energy planning or the need to study contracted demand, energy consumption must contain stratified and diversified knowledge in order to obtain a better understanding of the electrical system as a whole. However, not only the data related to electrical consumption is stratified. Virtually all historical data are maintained in the same way, for example, historical consumption of public services (water and gas), GDP and many others. Considering the above, the present methodology can also be applied to several countries with characteristics similar to Colombia. For example in Brazil, although residential electric consumption is not stratified, as in the Colombian case, the data recorded by the Energy Research Company (EPE, of the acronym in Portuguese) allows to use the same methodology to develop a forecast for the 5 regions that make up the country (Navarrete, 2015). Other compatible cases that can be cited are the countries that make up the Central American Integration System (SICA, of the acronym in Spanish), since the electricity consumption data are thoroughly analyzed by region and users, being publicly disclosed through statistical reports (Navarrete, 2015).

Selected Variables: For carrying out a load forecast, it is essential to consider the historical data of electrical consumption, which makes it possible to know the consumption profile in a future time window, as presented by (Conde *et al.*, 2016) and (Taylor and McSharry, 2007). In addition to the consumption history, for better robustness of the model, other variables were associated that correlate with the characteristics of demand, as suggested by (Eljazzar and Hemayed, 2016). Therefore, 5 sets of data were used to feed the ANN:

- Historical data of residential electrical consumption obtained from *Unidad de Planeación Minero Energética* (UPME , 2019);
- Population growth historical data obtained from *Departamento Administrativo Nacional de Estadística* (DANE, 2019);
- GDP per capita historical data obtained from *Banco de la Republica* (BanRep, 2019);
- Temperature records of the different study regions, obtained from IDEAM (UPME, 2016).

It is noteworthy that all the required records are public and free of charge. The data used are for the period from 2006 to 2016 (10 years) and can be extracted on a monthly basis with the exception of GDP and population growth, which are made available on an annual basis.

RESULTS AND DISCUSSION

Following the steps described in the previous sections, it was possible to perform the tests to finally get the results. Different cases of studies were conceived getting interested results in real and possible scenarios of the Colombian residential sector. For the development of the case study, 3 of the 6 socioeconomic strata were considered which are the most important, since they cover the majority of the population and have greater demand for electrical consumption. Thus, the strata addressed were 1, 3 and 6. The purpose of the case study was to make the forecast for the base year of 2017. The choice for a known year is justified by the fact that it makes it possible to compare the consumption already consolidated with the result of the forecast, analyzing the associated errors and checking the model effectiveness. As seen in Subsection 3.3, the database considered for each variable comprises the period from 2006 to 2016. Thus, the data were used to feed the ANN of each case through which the forecast for the base year was made. For each of the simulated scenarios, an adjustment of 2 ºC (blue line of the simulations) was considered in order to observe the influence and trend of temperature on residential electrical consumption, thus highlighting the advantage of the method employed. The procedure for net training was developed taking into account the following configuration of ANNs: two hidden layers of 6 and 4 neurons, respectively, LM training algorithm, tansig activation function, and a sample split 70%, 15% and 15% samples for the training, validation and testing steps. The previous simulated parameters showed the best results, as shown in section 3. For each of the 3 socioeconomic strata considered, RNA with 68 entries for the Caribbean region, 113 entries for the Andean region, 47 entries for the Pacific region, 36 entries for the Orinoquia region and finally 36 entries for the Amazon region were created and tested. For the input configurations, two hidden layers of 3 and 5 neurons were used, respectively, and a single output. The following subsections describe the procedures adopted in each case.

Socioeconomic Stratum 1: Stratum 1 is characterized by allocating the low-income population, with a stratum benefiting from subsidies from the state in public household services (water, electricity, gas). Thus, through multi-layer Perceptron ANN models it was possible to associate different variables in order to capture characteristic patterns that directly influence residential electrical consumption. The results are presented in Figure 4 where the graphs of the load forecast for stratum 1 and Table 1 are presented, where the quality indicators are obtained and the real values of residential electrical consumption are presented. It is observed that for stratum 1 tolerable results were obtained for 4 of the 5 study regions, with the exception of the Amazonas region, which presented high MAPE values for the months of February, April, May and June. On the other hand, it is noticed that the neural structures considered have adapted adequately to the 5 simulated regions, thus obtaining mean MAPE values less than 5%. Figure 4 allows observing the behavior of electrical consumption over the months of the year 2017. The black line represents the real consumption of stratum 1. The load forecast using ANN is shown in red. Finally, in blue, the load forecast from ANN is presented, considering a 2 ºC increase in temperature. Thus, we can observe the effectiveness of ANN in learning and predicting the behavior of residential electrical consumption in a satisfactory way. It is observed that for the Amazon region, the neural structure has adapted appropriately to the associated patterns, which influenced electricity consumption throughout the year, as presented in the months of February, March, May and June. On the other hand, the structure was unable to adapt to the data presented in September, generating a MAPE of 8.46%.

Table 1. Quality Indicators of the Different Analyzed Regions: Stratum 1

Stratum 1										
	MAPE (%)					MSE				
Month \ Region	Caribe	Andina	Pacifico	Orinoquia	Amazonas	Caribe	Andina	Pacifico	Orinoquia	Amazonas
January	4,30	6,99	3,61	4,31	2,36	3,285	0,753	0,502	0,062	0,006
February	3,40	1,71	0,36	5,86	9,43	1,799	0,040	0,005	0,101	0,081
March	0,68	0,81	7,91	2,38	6,08	0,087	0,010	2,231	0,019	0,029
April	0,88	1,23	2,24	0,01	22,77	0,146	0,022	0,188	0,000	0,366
May	0,15	1,39	2,77	2,50	10,11	0,004	0,029	0,270	0,021	0,088
June	1,24	0,96	3,18	4,71	12,03	0,307	0,014	0,366	0,070	0,130
July	1,19	5,88	3,88	3,88	0,43	0,295	0,527	0,497	0,059	0,000
August	1,75	2,88	0,22	5,59	7,79	0,629	0,129	0,002	0,095	0,064
September	5,80	3,46	2,78	2,34	8,52	8,153	0,194	0,301	0,017	0,081
October	1,01	3,12	0,95	6,47	0,68	0,227	0,149	0,036	0,170	0,001
November	3,04	0,88	4,24	4,97	4,48	2,093	0,012	0,652	0,080	0,025
December	1,22	0,94	3,11	4,83	0,65	0,309	0,014	0,363	0,093	0,000

Table 2. Quality Indicators of the Different Analyzed Regions: Stratum 3

Stratum 3										
	MAPE (%)					MSE				
Month \ Region	Caribe	Andina	Pacifico	Orinoquia	Amazonas	Caribe	Andina	Pacifico	Orinoquia	Amazonas
January	1,30	2,88	1,81	0,71	2,22	0,017	0,244	0,052	0,001	87,125
February	7,83	2,94	6,03	6,41	2,26	0,505	0,238	0,542	0,072	90,609
March	1,81	1,58	2,94	2,73	10,52	0,031	0,072	0,133	0,016	1016,080
April	6,10	1,85	1,83	1,69	8,38	0,398	0,097	0,051	0,006	992,566
May	2,15	1,78	0,83	8,98	3,78	0,054	0,093	0,010	0,168	217,203
June	9,93	3,08	3,68	0,88	0,67	1,182	0,287	0,216	0,001	6,827
July	8,57	1,11	1,63	2,44	0,73	0,917	0,037	0,041	0,011	7,605
August	4,71	0,42	6,40	2,52	5,12	0,275	0,005	0,683	0,011	388,281
September	8,46	1,69	7,37	0,88	0,65	0,913	0,089	0,920	0,002	6,887
October	2,12	0,24	6,14	5,93	2,58	0,055	0,002	0,625	0,074	111,114
November	3,02	1,16	0,78	10,80	4,28	0,111	0,040	0,009	0,229	331,434
December	3,87	0,84	1,94	5,21	0,86	0,159	0,021	0,059	0,061	12,318

Table 3. Quality Indicators of the Different Analyzed Regions: Stratum 6

Stratum 6										
	MAPE (%)					MSE				
Month \ Region	Caribe	Andina	Pacifico	Orinoquia	Amazonas	Caribe	Andina	Pacifico	Orinoquia	Amazonas
January	1,76	5,69	2,41	1,56		0,004	0,015	0,003	36,297	
February	21,50	1,68	9,69	5,95		0,441	0,001	0,047	407,718	
March	7,23	5,82	2,77	1,59		0,055	0,016	0,004	38,180	
April	10,14	0,34	9,04	1,26		0,116	0,000	0,040	20,902	
May	9,56	9,80	0,62	0,38		0,124	0,040	0,000	3,036	
June	15,59	1,94	2,93	3,66		0,365	0,002	0,005	244,351	
July	10,28	7,79	3,10	1,53		0,170	0,028	0,005	44,629	
August	10,18	6,01	7,05	3,60		0,152	0,016	0,031	235,102	
September	14,86	2,43	3,79	10,13		0,343	0,003	0,009	2302,141	
October	10,31	4,46	0,02	0,17		0,145	0,009	0,000	0,618	
November	27,01	6,62	0,10	14,54		1,080	0,020	0,000	4549,855	
December	7,84	5,56	0,47	1,56		0,077	0,013	0,000	60,963	

The same behavior can be observed for the month of July in the Pacific region and March in the Orinoquia region. This behavior presented in the different simulations is associated with atypical variations in electrical consumption, which, as recorded in the electrical consumption historic, interferes with the ANN learning process, influencing the final result of the model. Such consumption variations and their influence on load forecasts are a subject of study today (Lusis *et al*., 2017).

Socioeconomic Stratum 3: Stratum 3 is characterized by allocating the middle-income population, being a stratum that

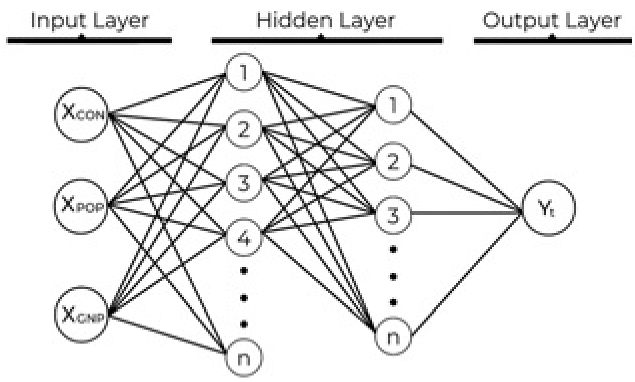

Figure 1. Basic Structure of an ANN-PMN

Figure 2. Load Forecasting Developed Algorithm

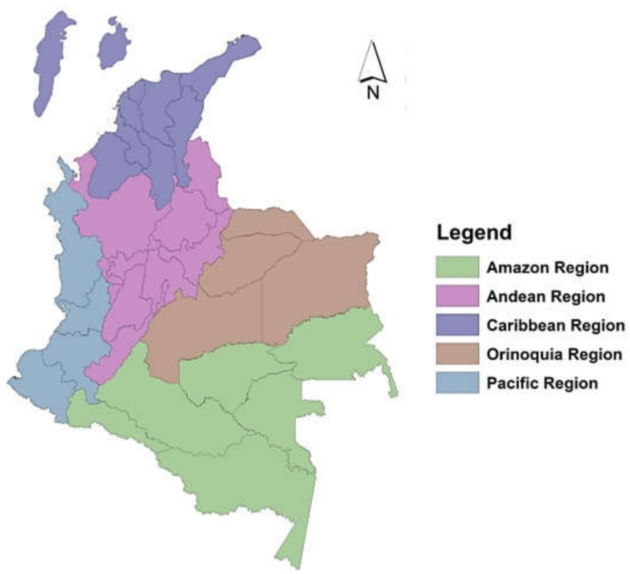

Figure 3. Classification of Study Regions: Grouping of states with similar temperature patterns and residential electrical consumption

has no subsidies or costs for public services (water, electricity, gas). Also, as in the case addressed in the previous section, for this stratum, the historical data discussed in the introduction to this section were used.

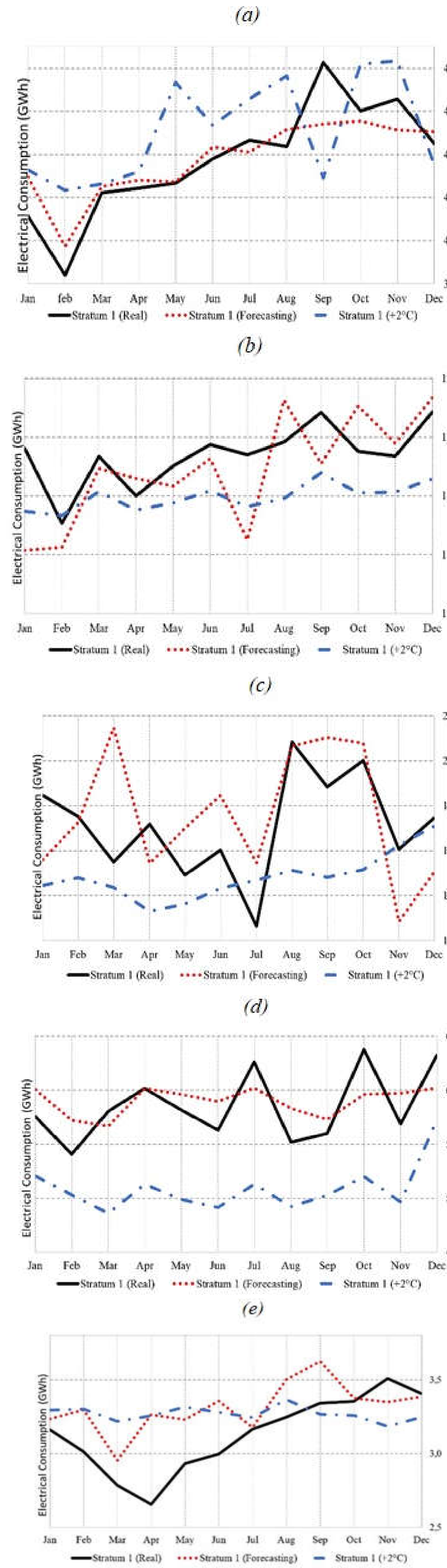

Figure 4. Load Forecast-Stratum 1. Caribbean (a), Andean (b), Pacific (c), Orinoquia (d) and Amazon (e) regions

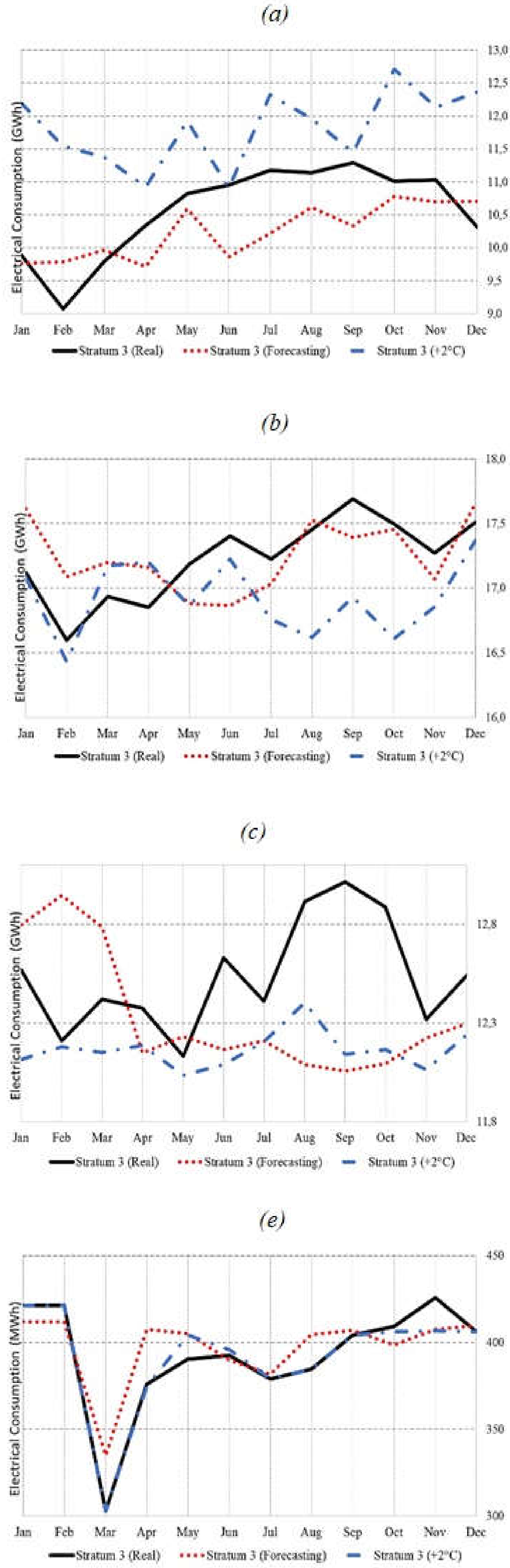

Figure 5. Load Forecast-Stratum 3. Caribbean (a), Andean (b), Pacific (c), Orinoquia (d) and Amazon (e) regions

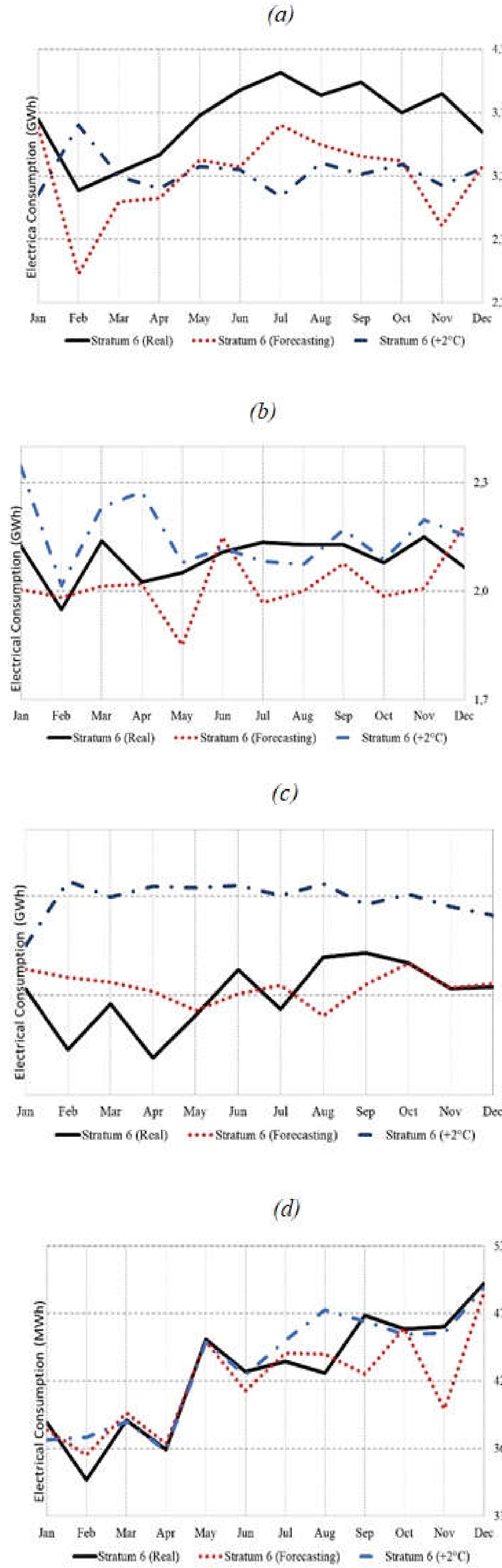

Figure 6. Load Forecast-Stratum 6. Caribbean (a), Andean (b), Pacific (c) and Orinoquia (d) regions

Two hidden layers of 4 and 6 neurons were considered, respectively. The output of the ANN was represented by a single vector of 12 positions, which will represent residential electricity consumption for the 12 months of the year 2017. Table 2 presents the quality indicators obtained between the load forecast obtained through the ANN and the real values of residential electrical consumption of stratum 3 registered by the UPME. As seen in Table 2, the forecasting models worked satisfactorily for most regions, especially the Andean, Pacific and Orinoquia regions. However, errors were obtained outside the standard behavior, for example, the month of November in the Orinoquia region and the month of March in the Amazon region. One reason for this behavior is the low correlation of variables within the model. Another reason that deserves to be considered is the occurrence of atypical events, which ends up influencing energy consumption, mainly due to exogenous variables (external to the model) such as political decisions or rationing that were not considered in this paper. In Figure 5 the results of the stratum 3 load forecast are presented. As noted, the configuration of hidden layers employed showed a level of accuracy within the tolerable range, proving capable of adapting to variations in residential electrical consumption. For example, it was observed that in the months of February and November in the Andean region, there was a relative decrease in real consumption (black line). Analyzing the forecast together (red line), it can be noted that the ANN behaved following the trend line. For this case, the model with the highest accuracy was the one used in the Andean region, where the critical month had a MAPE of 3.08%. The second most accurate model was the one used for the Amazon region, whose critical MAPE is associated with the month of March (10.52%), suggesting an atypical consumption.

Socioeconomic Stratum 6: Stratum 6 is characterized by allocating the high-income population, being a stratum that is not benefited by any type of subsidies, therefore the portion of the population that pays the total for the costs related to public services (water, electricity, gas). For this stratum a configuration of 4 and 7 neurons was used. The Amazon region, made up of the states Putumayo, Caquetá and Guaviare, does not have this socioeconomic stratum, so it was disregarded in this case. Table 3 shows the quality indicators obtained between the load forecast developed through the ANNs and the real values of residential electrical consumption of stratum 6 registered by the UPME. Also, in this case, it was possible to observe good efficacy of neural structures for the Andean and Pacific regions, where the biggest errors were obtained for the months of February (9.69%) and April (9.04%). On the other hand, the Caribbean region had the biggest error among all the strata and regions analyzed, with MAPE quite critical for most of the months of the year, with November (27.01%) being the most distant from the real. The error behavior found in this stratum is based on the fact that the electrical consumption of the states that make up the region is not homogeneous. The lack of homogeneity of the data refers to the significant variations between electrical consumption from one state to another, mainly due to population and territorial differences. This happens mainly in the 6 states analyzed, with Atlántico and Bolívar being the states with the highest electricity consumption, while Sucre and Córdoba are those with the lowest residential electricity consumption. Figure 6 shows the results of the load forecast for stratum 6. The results showed the good behavior of the model when compared to the real values for the Caribbean region, except for the forecast value obtained for the month of November. It is observed that the adopted neural structure follows a behavior similar to the real consumption of the year 2017, replicating characteristic consumption behaviors, such as those presented in the months of February and July. On the other hand, the consumption of the month of November was not accurately replicated by the ANN, obtaining a MAPE of 27.01%, being the biggest error reached in the different scenarios considered. For the Andean region, the ANN configured with an increase in temperature of 2 ºC (Blue line) performed better and was even more accurate than the base forecast model (red). Finally, the model that best adapted to the two considered scenarios was the one employed in the Orinoquia region, whose results were similar to the real data for the year 2017, except for the data recorded for the month of November, whose MAPE was 14.54%. It is worth pointing out that stratum 6 represents a social minority whose electric consumption habits are totally different from strata 1 and 3. Thus, the data recorded in the history of electric consumption allowed us to observe a different consumption for the months of November, February and July, replicating different behaviors in the result of the developed electricity consumption forecasts.

CONCLUSION

The present work presented the development of a multivariable model based on MLP ANN for the medium-term forecast of residential electrical consumption. Variables of great adherence to the sector were incorporated and correlated in order to increase the model's reliability. For testing and validating the forecast from the ANN, information from Colombia was chosen as a case study, being one of the many countries to which the present model can be applied, with historical data and electrical consumption separated by regions and layers / social strata. The results obtained from the computational structures presented and tested showed a clear influence of the different variables in the forecast of Colombian residential electrical consumption. This shows the importance of correlating different variables in order to enable a better performance of the electrical network, reducing costs and increasing the effectiveness of the models. As can be seen, it is possible to achieve relatively low margins of error with the use of the employed technique. This shows the great advantage of using ANNs compared to other methodologies commonly found in the literature. From the developed work, a set of parameters was observed that can be modified in order to find the best configuration, always aiming at reliable and accurate results. Among the suggested alteration parameters, the user can increase the number of hidden layers, increase the number of samples in the training stage, increase the number of inputs (if possible) or even consider different training algorithms. The residential electrical sector represents an important focus of study due to its increase over time, influenced mainly by the socioeconomic and climatic variations of the study region. However, the literature review points out the need to advance new models in this follow-up in view of the major changes that the world has gone through with regard to the electricity sector. In fact, also considering the advances in technology, it is possible to obtain increasingly reliable forecasting models that can contribute to better planning of residential energy demand.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support provided by the National Council for Scientific and Technological Development (CNPq) and the Coordination of Improvement of Higher Education Personnel (CAPES).

AUTHOR´S CONTRIBUTIONS

The main contributions of the present work were:

- Development of an improved model compared to the approaches in the literature for the forecast of medium-term load considering the use of AI, more specifically ANNs, which allows the reduction of computational effort;
- Use of multi-model architecture, that is, subdivision of the forecast model into sub-models that allow to address the problem considering the peculiarity of each case, since the approach to obtaining stratified data was addressed, aiming at greater precision;
- Multivariable approach, considering not only electrical consumption as an input parameter, but also socioeconomic and climatological variables that have an influence on the behavior of energy consumption;
- Stratified implementation by means of variables that are grouped according to the social strata, facilitating the development of planning and operation tools that punctually meet the needs of each of the country's socioeconomic strata.

REFERENCES

Ahmad A. S. *et al*. 2014. “A review on applications of ANN and SVM for building electrical energy consumption forecasting,” Renew. Sustain. Energy Rev., vol. 33, pp. 102–109, doi: 10.1016/j.rser.2014.01.069.

Ahmad, M. W. M. Mourshed, B. Yuce, and Y. Rezgui 2016. “Computational intelligence techniques for HVAC systems: A review,” Build. Simul., vol. 9, no. 4, pp. 359–398, Aug., doi: 10.1007/s12273-016-0285-4.

BanRep, “Producto Interno Bruto 2019.” Banco de la Republica de Colombia. https://www.banrep.gov.co/es/ estadisticas/producto-interno-bruto-pib accessed Nov. 06, 2019.

Barman, M. N. B. Dev Choudhury, and S. Sutradhar 2018. “A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India,” *Energy*, vol. 145, pp. 710–720, Feb., doi: 10.1016/j.energy.2017.12.156.

Beale, M. H. M. T. Hagan, and H. B. Demuth 2016. “Neural Network Toolbox TM Reference How to Contact MathWorks,” Natick.

Conde, G. A. B. Á. L. de Santana, C. R. L. Francês, C. A. Rocha, L. Rego, and V. Gato 2016. “Estratégias De Previsão De Carga E De Consumo De Energia Elétrica Baseadas Em Modelos Estatísticos E Redes Neurais Artificiais: Um Estudo De Caso Nas Concessionárias De Energia Do Estado Do Pará,” in Anais do 8. Congresso Brasileiro de Redes Neurais, Apr., no. April, pp. 1–6, doi: 10.21528/CBRN2007-088.

da Silva, F. L. C. F. L. Cyrino Oliveira, and R. C. Souza 2019. “A Bottom-up Bayesian Extension for Long Term Electricity Consumption Forecasting,” *Energy*, vol. 167, pp. 198–210, doi: 10.1016/j.energy.2018.10.201.

DANE 2019. “Demografia y Población,” Departamento Administrativo Nacional de Estadistica. https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion accessed Nov. 06, 2019.

de A. P. Braga, A. P. de L. F. de Carvalho, and T. B. Ludermir 2007. Redes Neurais Artificiais: Teoria e Pratica, Grupo Edit., no. 2 Edição. Rio de Janeiro: LTC.

Duda, R. O. P. E. Hart, and D. G. Stork 2000. Pattern Classification, 2nd ed. New York, NY, USA: John Wiley & Sons, Inc..

Eljazzar M. M. and E. E. Hemayed 2016. “Feature selection and optimization of artificial neural network for short term load forecasting,” in 2016 Eighteenth International Middle East Power Systems Conference MEPCON, Dec., pp. 827–831, doi: 10.1109/MEPCON.2016.7836990.

Empresa de Pequisa Energética 2018. “Consumo Mensal de Energia Elétrica por Classe regiões e subsistemas,” 2018, Jan.. https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/Consumo-mensal-de-energia-eletrica-por-classe-regioes-e-subsistemas accessed Jun. 29, 2020.

Garzon Medina, D. O. R. Caneloi dos Santos, T. Sousa, and J. C. Lopes 2019. “Comparative Analysis of Artificial Neural Networks and Statistical Models Applied to Demand Forecasting,” in *2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America ISGT Latin America*, Sep., pp. 1–6, doi: 10.1109/ISGT-LA.2019.8895277.

González-Romera, E. M. A. Jaramillo-Morán, and D. Carmona-Fernández 2008. “Monthly electric energy demand forecasting with neural networks and Fourier series,” Energy Convers. Manag., vol. 49, no. 11, pp. 3135–3142, Nov., doi: 10.1016/j.enconman.2008.06.004.

Hernández Lalinde J. *et al*. 2018. “Sobre el uso Adecuado del Coeficiente de Correlación de Pearson: Definición, Propiedades y Suposiciones,” Arch. Venez. Farmacol. y Ter., vol. 37, no. 5, p. 8, Accessed: Nov. 06, 2019. [Online]. Available: http://190.169.30.98/ojs/index.php/rev_aavft/article/view/16165.

Hong T. and S. Fan 2016. “Probabilistic electric load forecasting: A tutorial review,” *Int. J. Forecast.*, vol. 32, no. 3, pp. 914–938, Jul., doi: 10.1016/j.ijforecast.2015.11.011.

Hudson Beaule, M. M. T. Hagan, and H. B. Demuth 2018. “Neural Network Toolbox TM Getting Started Guide,” Natick, MA, USA. [Online]. Available: https://la.mathworks.com/help/simulink/index.html.

Instituto Geografico Agustin Codazzi IGAC 2019. “Geoportal de Mapas Interactivos y de Relieve Colombiano,” Acceso virtual. https://www.igac. gov.co/es/ contenido/geoportal accessed Jul. 04, 2019.

Jasiński T. 2019. “Modeling electricity consumption using nighttime light images and artificial neural networks,” *Energy*, vol. 179, pp. 831–842, Jul., doi: 10.1016/j.energy.2019.04.221.

Kermanshahi B. and H. Iwamiya 2002., “Up to year 2020 load forecasting using neural nets,” *Int. J. Electr. Power Energy Syst.*, vol. 24, no. 9, pp. 789–797, Nov., doi: 10.1016/S0142-06150100086-2.

Kuster, C. Y. Rezgui, and M. Mourshed 2017. “Electrical load forecasting models: A critical systematic review,” *Sustain. Cities Soc.*, vol. 35, no. August, pp. 257–270, Nov., doi: 10.1016/j.scs.2017.08.009.

Liu, C. Z. Jin, J. Gu, and C. Qiu 2017. "Short-term load forecasting using a long short-term memory network," in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe ISGT-Europe*, Sep., vol. 2018-Janua, pp. 1–6, doi: 10.1109/ISGTEurope.2017.8260110.

Lusis, P. K. R. Khalilpour, L. Andrew, and A. Liebman 2017. "Short-term residential load forecasting: Impact of calendar effects and forecast granularity," Appl. Energy, vol. 205, pp. 654–669, Nov., doi: 10.1016/j.apenergy.2017.07.114.

Makalesi, A. D. Aydın, and H. Toros 2018. "Prediction of Short-Term Electricity Consumption by Artificial Neural Networks Using Temperature Variables," *Eur. J. Sci. Technol.*, no. 14, pp. 393–398, doi: 10.31590.

Mendoza Morales A. 2006. "Colombia: Estado Regional," Bogotá. Accessed: Jun. 26, 2019. [Online]. Available: http://www.fuac.edu.co/recursos_web/observatorio/02/AL BERTO MENDOZA. COLOMBIA ESTADO REGIONAL.pdf.

MohammadZadeh S. and A. A. Masoumi 2010. "Modeling residential electricity demand using neural network and econometrics approaches," in *The 40th International Conference on Computers & Indutrial Engineering*, Jul., pp. 1–6, doi: 10.1109/ICCIE.2010.5668322.

Moral-Carcedo J. and J. Vicéns-Otero 2005. "Modelling the non-linear response of Spanish electricity demand to temperature variations," Energy Econ., vol. 27, no. 3, pp. 477–494, May, doi: 10.1016/j.eneco.2005.01.003.

Mordjaoui, M. S. Haddad, A. Medoued, and A. Laouafi 2017. "Electric load forecasting by using dynamic neural network," *Int. J. Hydrogen Energy*, vol. 42, no. 28, pp. 17655–17663, Jul., doi: 10.1016/j.ijhydene.2017.03.101.

Nichiforov, C. I. Stamatescu, I. Fagarasan, and G. Stamatescu 2017. "Energy Consumption Forecasting Using ARIMA and Neural Network Models," in *5th International Symposium on Electrical and Electronics Engineering ISEEE*, pp. 1–4, doi: 10.1109/ISEEE.2017.8170657.

Onoda T. 1993. "Next day's peak load forecasting using an artificial neural network," in Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems, pp. 284–289, doi: 10.1109/ANN.1993.264333.

Park, D. C. M. A. El-Sharkawi, R. J. Marks, L. E. Atlas, and M. J. Damborg 1991. "Electric load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 442–449, May, doi: 10.1109/59.76685.

Ramirez Adriana Marcela A. 2013. "Metodos Utilizados para el Pronostico de Demanda de Energía Eléctrica en Sistemas de Distribución," Universidad Tecnologica de Pereira.

Raza M. Q. and A. Khosravi 2015. "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 1352–1372, Jun., doi: 10.1016/j.rser.2015.04.065.

Rojas Navarrete M. E. 2020. "Estadisticas del subsector eéctrico de los paises del Sistema de la Integración Centroamericana SICA," Ciudad de Mexico, Jan. 2015. Accessed: Jun.. [Online]. Available: https://repositorio.cepal.org/bitstream/handle/11362/40910/1/S1700038_es.pdf.

Shamsollahi, P. K. W. Cheung, Quan Chen, and E. H. Germain 2001. "A neural network based very short term load forecaster for the interim ISO New England electricity market system," in Innovative Computing for Power - Electric Energy Meets the Market. *22nd International Conference on Power Industry Computer Applications* Cat. No.01CH37195, pp. 217–222, doi: 10.1109/PICA.2001.932351.

Singh P. and P. Dwivedi 2018. "Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem," *Appl. Energy*, vol. 217, no. October 2017, pp. 537–549, May, doi: 10.1016/j.apenergy.2018.02.131.

Sun X. *et al*. 2016. "An Efficient Approach to Short-Term Load Forecasting at the Distribution Level," IEEE Trans. Power Syst., vol. 31, no. 4, pp. 2526–2537, Jul., doi: 10.1109/TPWRS.2015.2489679.

Taylor J. W. and P. E. McSharry 2007. "Short-Term Load Forecasting Methods: An Evaluation Based on European Data," IEEE Trans. Power Syst., vol. 22, no. 4, pp. 2213–2219, Nov., doi: 10.1109/TPWRS.2007.907583.

Tondolo de Miranda, S. A. Abaide, M. Sperandio, M. M. Santos, and E. Zanghi 2018. "Application of artificial neural networks and fuzzy logic to long-term load forecast considering the price elasticity of electricity demand," *Int. Trans. Electr. Energy Syst.*, vol. 28, no. 10, p. 17, Oct., doi: 10.1002/etep.2606.

UPME 2016. "Proyección de la Demanda de Energía Eléctrica y Potencia Máxima en Colombia," Medellin. Accessed: Nov. 06, 2019. [Online]. Available: www.upme.gov.co.

UPME 2019. "Demanda y Eficiencia Energetica," Unidad de Planeación Minero Energetica. https://www1.upme.gov.co/Paginas/Demanda-y-Eficiencia-Energetica.aspx accessed Nov. 06, 2019.

Wang, P. B. Liu, and T. Hong 2016. "Electric load forecasting with recency effect: A big data approach," Int. J. Forecast., vol. 32, no. 3, pp. 585–597, Jul., doi: 10.1016/j.ijforecast.2015.09.006.

Zhang, G. B. Eddy Patuwo, and M. Y. Hu 1998. "Forecasting with artificial neural networks:," *Int. J. Forecast.*, vol. 14, no. 1, pp. 35–62, Mar., doi: 10.1016/S0169-20709700044-7.

****