DOI: 10.48011/sbse.v1i1.2495

Model Based on Artificial Neural Networks for Forecasting Electricity Consumption: A Holistic Approach

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

** Center of Engineering, Modelling and Applied Social Sciences Federal University of ABC, Santo André, SP, Brazil.; E-mail's: daniel.garzon@ufabc.edu.br, calixto.junior@ufabc.edu.br, thales.sousa@ufabc.edu.br*

Abstract: Electrical demand forecasting is a key tool in making operational and strategic decisions in power companies, whose lack of accuracy can lead to high economic costs. In this sense, forecasting allows network operators to make power dispatch, maintenance program, reliability analysis and operational safety decisions. Therefore, the present work proposed the use of Artificial Neural Networks (ANN) to project the demand of the Colombian residential sector. The model presented for the forecast was based on socioeconomic variables obtained from official Colombian government data sources such as population growth, gross domestic product and residential electrical consumption. The work was developed with the aid of the MATLAB® software where a model with appreciable assertiveness margin were proposed.

Keywords: Artificial Intelligence; Artificial Neural Networks; Load Forecasting; Multimodel Forecasting; Residential Forecasting.

1. INTRODUCTION

Energy planning is extremely relevant for determining consumption in the coming years, as well as for assessing the capacity to supply human need through the use of non-renewable energy sources. To this end, studies of electric power demand are employed, which allow a broad development of electric systems, starting with the proposal of inclusion of new renewable sources, commercialization of energy and operative analysis of electric networks (Mordjaoui *et al.*, 2017).

The main objective of the forecast is to define what will be the electricity consumption of a specific place, in order to plan and make the best decisions to serve the consumers (Conde *et al.*, 2016). As mentioned by Conde *et al.* (2016) and Taylor *et al.*(2007), a forecasting model must take into account historical data that makes it possible to detect patterns that determine the behaviour of the load in the future.

Especially the forecasting of electricity consumption for the residential sector becomes very important since it is associated with the development and infrastructure of a country. This mode of forecast has peculiar characteristics as it has the most complex behaviour of studying, forecasting and analysing, although it has a strong connection with other classes of consumption such as industrial and commercial. (Conde *et al.*, 2016), (Shamsollahi *et al.*, 2001).

Different models for consumption forecasting have been developed over the last decades, including time series models, vector support machines and, neural networks. (Schachter and Mancarella, 2014), (Pessanha and Leon, 2015). The first models employed statistical methods that linearly related the different variables (Maybee and Uri, 1979). The use of this type of technique had negative consequences since the result of the predictions presented high errors and often lack of convergence caused by the modelling of variables erroneously.

Due to the lack of precision of some traditional models, models based on Artificial Intelligence (AI) have become preferably for researchers in the field. Among the various techniques employed over the last few years, Artificial Neural Networks (ANN) (Peng, Hubele and Karady, 1990). Since then, the application of this technique has been wide in several areas of knowledge, including load forecasting (Park and Mohammed, 1991).

Nowadays, Multilayer Perceptron (MLP) model has been worked by many authors due to the good quality of the results, although it requires a longer computational time (Duda, Hart and Stork, 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.

Considering the above, the present work employs MLP-type ANNs to make a short-term load forecast for Colombia's 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, a clear and objective methodology will be presented that resulted in great precision.

2. METODOLOGY

For construction and testing of the model, historical data from Colombia was used. A characteristic of this locality is that it has stratified data, it means, divided into socioeconomic sectors, which represents an advantage by allowing to analyse the behaviour 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).

2.1 Modelling Stages

The planning of the model followed five steps conventionally employed for the correct development of a short or medium term forecast (Mondragón, 2011). Fig. 1 shows the process flowchart.

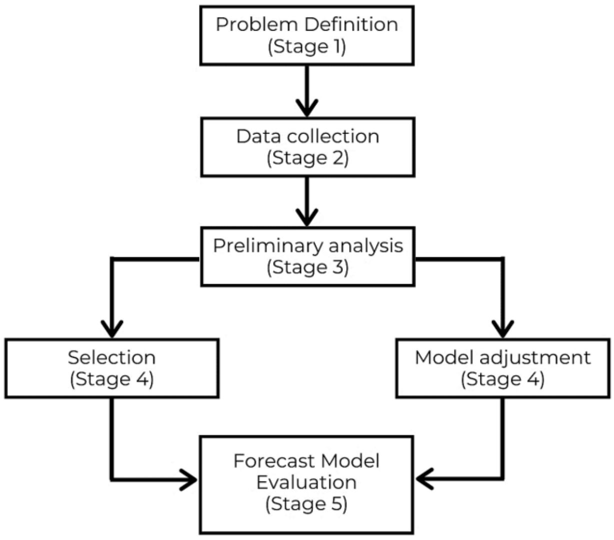

Fig. 1 Process diagram for residential load forecasting.

Looking at Fig. 1, the problem is first defined, being the most important phase in load forecasting where it is necessary to delimit the problem, as well as the time horizon to be considered. Considering the above, in the definition of the problem, it was considered the short term forecast (one year ahead) (UPME, 2016). At this stage it was considered the time series from 2006 to 2016, to serve as the basis for the 2017 forecast. The forecast year was chosen mainly because it is a real past data that allows a better view of the accuracy of the model employed.

Then, in step 2 it was seek to obtain good quality historical data because the model's effectiveness is influenced by the coherence of the obtained history, and it is also necessary to evaluate the influence that the selected variables will have on future electric consumption. The study of the literature allowed to obtain, conclusively, the most relevant variables for the model. The following were employed: Energy consumption (X_{COM}) obtained from (UPME, 2019), population growth (X_{POP}) provided by *Departamento Administrativo Nacional de Estadística* (DANE, 2019), and finally GDP (X_{GDP}) provided by *Banco Central de la Republica* (BanRep, 2019). As shown previously, the data base employed in the study has a time window of 11 years (2006 to 2016). In addition to the selection and identification of the input variables, it is also necessary to pre-treat them for better use in the model. Regarding the X_{COM} variable, it was necessary to make the decomposition in a time interval corresponding to 12, that is, monthly, $(T-n)$, where $n \in \mathbb{Z}$, $n = \{1,2 \dots 12\}$, corresponding to each month of the year, making possible the future value for the instant $(T+1)$. A similar procedure was made for (X_{POP}) and (X_{GDP}), however considering the temporal division, where the totality of samples corresponds to 11, one sample for each year. The previous steps are also called "windowing for time series forecasting".

In step 3, it was seek to perform a preliminary analysis that allows to recognize time series patterns in order to allow the choice of the most appropriate model. First, it is essential to know the behaviour of demand over the time period to be considered. Soon after, developing a prior statistical analysis of the model variables provides a better understanding of their behaviour, thus increasing their reliability (Mondragón, 2011). Thus, after knowing the demand behaviour (model dependent variable), a relationship is developed between the independent variables: population growth and GDP. Thus the variables were previously analysed using Pearson's Correlation Index (R) to ensure that they would be appropriate for use in the model created (Hernández Lalinde et al., 2018). It was possible to verify that the two independent variables contributed to the modelling and could be used in the model. The good correlation between GDP and demand was observed, with a coefficient of 94%, while population growth compared with the dependent variable obtained a correlation of 98%.

Subsequently, in step 4 it was seek to select and adjust the model based on the variables and time series acquired. Each model has its own characteristics that will determine the accuracy according to the time horizon worked. For the present work, it was considered appropriate to work with an ANN-based model due to the ability to model nonlinear relationships between the dependent variable and the independent variables (Duda, Hart and Stork, 2000).

Thus, in step 5 the validation of the models were performed. From the inputs of the variables, the result obtained was the forecast of the year 2017 (Y_t). The expression representing the result as a function of the entries is described in (1).

$$Y_t = [f(X_{CON}, X_{POP}, X_{GNP})] \tag{1}$$

In this step it is also observed the calculation of Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE) that compares the actual prediction values with those obtained in the model, as can be seen in (2) and (3) (Singh and Dwivedi, 2018). These are quality indicators widely employed for their statistical representation and simplicity.

$$MSE = \frac{\sum_{i=1}^{n}\left(Y_{m_i} - Y_{e_i}\right)^2}{n} \tag{2}$$

$$MAPE = \frac{\sum_{i=1}^{n}(Y_{e_i} - Y_{m_i)}}{Y_{ei}} x100 \tag{3}$$

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

Based on the previous steps, the ANN was modelled using the variables considered within the study as shown in Fig. 2.

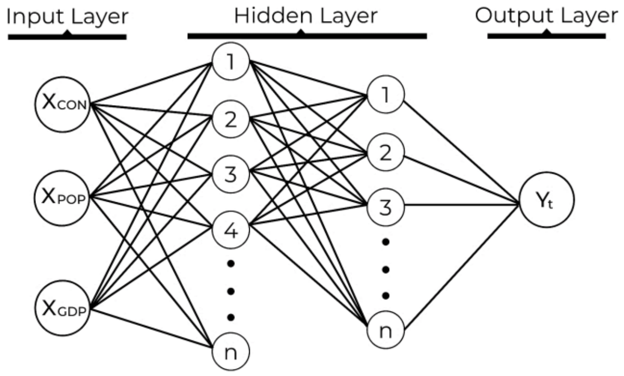

Fig. 2 ANN structure considering the variables selected for the study.

3. TESTS AND RESULTS

Finally, following the steps described in the previous section, it was possible to perform the tests to finally get the results. Different cases of studies were conceived where the first one approached the use of only one variable (Electric consumption). From then on, in the following tests the combinations of the other variables were added. The following subsections describe the procedures adopted in each case.

Regarding the division of samples from the database used, for all cases, 70% were considered for training (92 samples), 15% for validation (20 samples) and 15% for the tests (20 samples). Such parameters are at the discretion of the project's executor, and may be altered by the user, as some works in the literature have had lower error bands using the division presented here, also found in (Conde et al., 2016).

3.1 Case 1: Residential Load Forecasting using 1 variable

In the first case study, only the historical series of residential electrical consumption was considered. The appropriate ANN configuration was found by performing multiple tests to observe network performance and how the gradient descended to fit known network data (training).

The algorithm used for network training was LM since it has shorter response time and higher accuracy resulting in better convergence (Kaytez et al., 2015). The best result was obtained with a composition of two hidden layers of 10 and 3 neurons respectively and tansig activation function. In this way an appropriate response was obtained within 4 seconds and 9 iterations. The results obtained are presented in Fig. 3 and Table 1.

Percent error values were compared by MAPE and square error values by MSE (Ghalehkhondabi et al., 2017). The results showed high values for some months, such as January and May, because of the error calculation was performed from the square of the difference of the forecast and actual values.

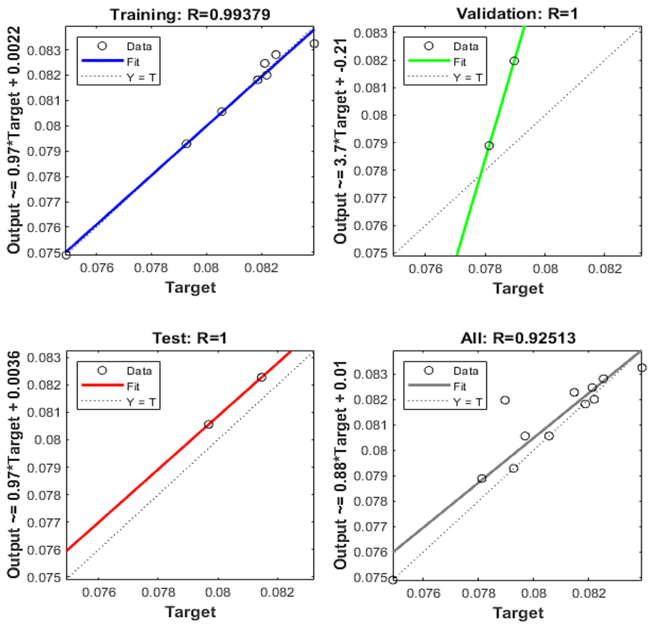

Fig. 3 ANN performance through its different stages for the first case.

Table1. Results of Simulation

Month	Out RNA (GWh)	Real Value (GWh)	MAPE (%)	MSE
January	1967.286	1895.217	3.802	5193.89
February	1797.604	1797.705	0.005	0.01
March	1893.368	1875.096	0.974	333.86
April	1902.959	1902.708	0.013	0.06
May	1933.443	1912.655	1.086	432.14
June	1968.034	1972.823	0.242	22.93
July	1933.475	1933.499	0.001	0.00
August	1974.669	1955.368	0.987	372.54
September	1997.872	2014.284	0.814	269.35
October	1987.470	1980.788	0.337	44.65
November	1963.601	1964.942	0.068	1.80
December	1979.152	1971.065	0.410	65.40

3.2 Case 2: Residential Load Forecasting using 2 variables

It was considered a new configuration whose inputs were electric consumption and population growth. As in the first case, the appropriate ANN configuration was found by performing multiple tests. Through a hidden two-layer configuration of 3 and 6 neurons, respectively, LM training algorithm, and tansig activation function, an appropriate response was achieved within 4 seconds and 10 iterations. Results are shown in Fig. 4 and Table 2.

It is noteworthy that the population growth registered by DANE is annual, but a constant value that did not vary during the year was simulated. Thus, 11 different values were used, representing the population value for the period from 2006 to 2016.

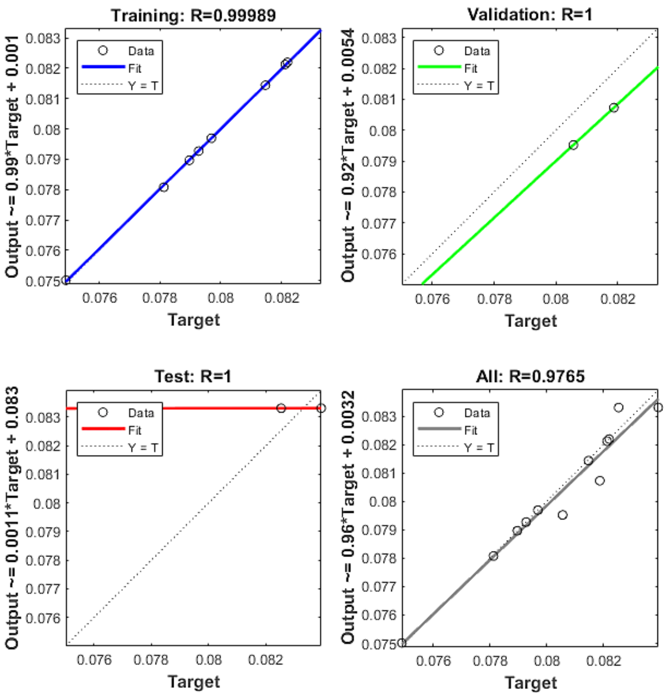

Fig. 4 ANN performance through its different stages for the second case.

Table2. Results of Simulation

Month	Out RNA (GWh)	Real Value (GWh)	MAPE (%)	MSE
January	1895.226	1895.217	0.0004	0.00
February	1800.384	1797.705	0.1490	7.18
March	1873.938	1875.096	0.0617	1.34
April	1902.564	1902.708	0.0075	0.02
May	1912.657	1912.655	0.0001	0.00
June	1972.740	1972.823	0.0042	0.01
July	1908.559	1933.499	1.2899	622.04
August	1954.570	1955.368	0.0408	0.64
September	1999.428	2014.284	0.7375	220.70
October	1999.392	1980.788	0.9392	346.11
November	1937.486	1964.942	1.3973	753.87
December	1970.981	1971.065	0.0042	0.01

The analysis of the results concluded that there is a great influence of the use of a second variable in the model. For the second case, it was possible to observe the improvement in the obtained results, since the quality indicators presented lower values compared to the first case.

3.3 Case 3: Residential Load Forecasting using 3 variables

In the third case, it was applied as input: the historical series of residential electric consumption, the population, and the GDP. The procedure for net training was identical to the previous cases but, in this case, the appropriate configuration of ANN was two hidden layers of 6 and 4 neurons, respectively, LM training algorithm and tansig activation function. The results are shown in Fig. 5 and Table 3.

Considering that the population growth and per capita GDP data were obtained annually from official public portals, it was considered a constant value over the 12 months, similar to the process implemented in Case 2. Thus, 11 different values were used. GDP and population, which represent the values of the years 2006-2016.

For this case it is of paramount importance to note that the addition of a third variable adversely impacted the results when compared to the previous case which obtained a relatively minor associated error.

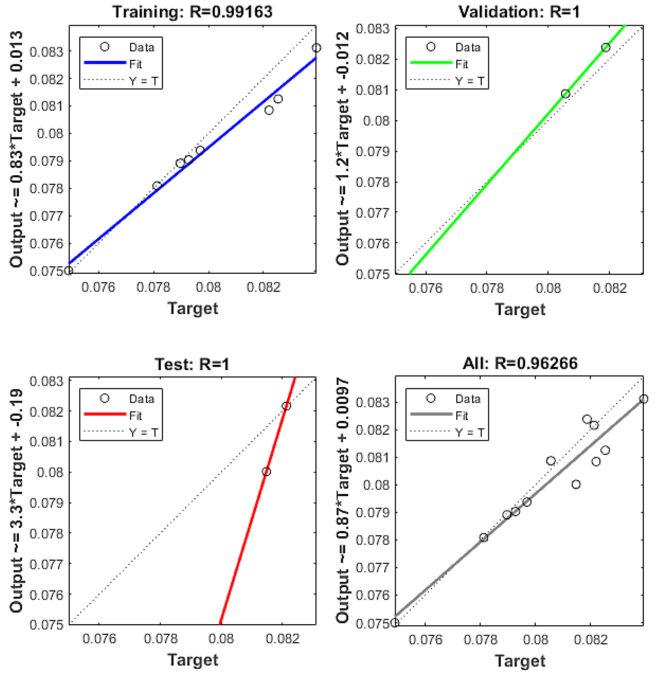

Fig. 5 ANN performance through its different stages for the third case.

Table3. Results of Simulation

Month	Out RNA (GWh)	Real Value (GWh)	MAPE (%)	MSE
January	1893.807	1895.217	0.0743	1.99
February	1799.924	1797.705	0.1234	4.92
March	1874.086	1875.096	0.0538	1.02
April	1896.715	1902.708	0.3149	35.92
May	1904.984	1912.655	0.4010	58.85
June	1940.157	1972.823	1.6557	1067.03
July	1940.827	1933.499	0.3790	53.70
August	1920.330	1955.368	1.7918	1227.65
September	1994.698	2014.284	0.9723	383.61
October	1950.059	1980.788	1.5513	944.25
November	1977.050	1964.942	0.6161	146.59
December	1971.778	1971.065	0.0361	0.51

3.4 Discussions and Analysis

Figures 3, 4 and 5 show the network outputs in relation to the targets for the training, validation and testing steps. For a perfect fit, data should fall along a 45 degree line where network outputs are equal to targets (Mathworks, 2018). This could be seen in the results.

Model data fit was considered reasonably good for all data sets, with R values equal to or greater than 0.96 (Hernández Lalinde et al., 2018). Another point that denotes good performance of the results are the quality indicators employed to validate the model (MSE and MAPE). It can be seen from the tables presented in each case that all of the errors have remained quite below the models normally employed in the literature, which demonstrates the model's reliability. Additionally, it is possible to retrain the net in order to modify the initial weights, thus producing improvements in the results.

For better comparative effect between cases, Fig. 6 illustrates the load curve of 2017, comparing the results of the three cases analysed, where it is possible to see the deviation of the simulated cases to the real consumption. It can be observed that the closest result to the real one is the case with 2 variables (case 2).

It is important to note that until may the models have a similar behaviour to the real consumption. This is explained by the influence of the variables in each case and the way the neural structures of the three cases were parameterized. Note that the curves deviated slightly from May, approaching different forms of the real consumption curve (in red).

4. CONCLUSIONS

In the present work it was proposed a short-term demand forecasting model for the Colombian residential sector. In this sense, the behaviour of residential electricity consumption was simulated including socioeconomic variables such as population growth and GDP associated with the history of energy consumption in order to predict residential electrical consumption in 2017.

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.

Each time an ANN is trained, different solutions can be obtained as a result of the initial weights, as well as a function of the different divisions of the training, validation and testing sets. As a result, different ANNs trained on the same problem may give different solutions for the same set of entries. To ensure that RNA has found the right answer, it needs to be trained several times, thus analysing the stopping conditions and the simulation time that the user deems appropriate.

For the first analysed study case, considerable errors were obtained starting from the relation of electric consumption over 10 years of study. For the second case, a smaller error was found compared to the first. The population growth variable, together with an excellent ANN structure, allowed us to find better results, where the highest MAPE found was 1.397% relative to November. Finally, for the third case it was observed that the performance obtained was worse than in the previous case in relation to errors. Previous ANN configurations found allowed reliable results, thus achieving minimal errors in short-term residential load forecasting.

From the work developed it was possible to realize that in order to find reliable and accurate results, the planner has a set of parameters that can modify in order to find the best setting that guarantees the best results. Among the suggested parameters, the user can increase the number of hidden layers; increase the number of samples in the training stage; increase the number of entries (if possible) as well as consider different training algorithms.

5. ACKNOWLEDMENTS

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).

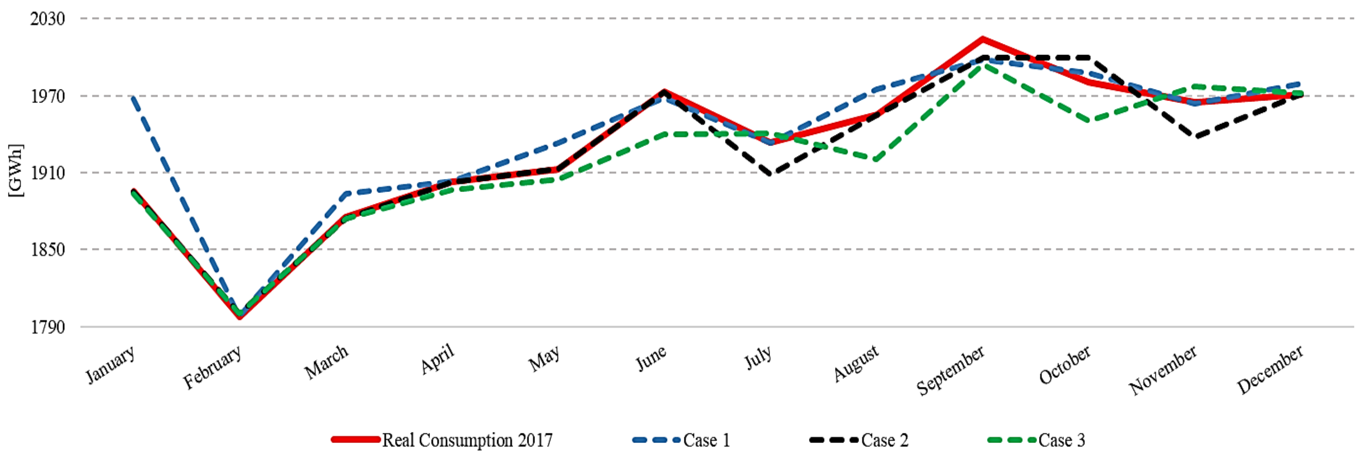

Fig. 6 Demand curve: behaviour of the 3 cases with the consolidated result of the year 2017.

6. REFERENCES

BanRep (2019) Producto Interno Bruto (PIB), Banco de la Republica de Colombia. Available at: https://www.banrep.gov.co/es/estadisticas/producto-interno-bruto-pib (Accessed: 6 November 2019).

Conde, G. A. B. et al. (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 VIII Congresso Brasileiro de Redes Neurais. Florianópolis, SC, Brazil: ABRICOM, pp. 1–6. doi: 10.21528/CBRN2007-088.

DANE (2019) Demografia y Población, Departamento Administrativo Nacional de Estadistica. Available at: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion (Accessed: 6 November 2019).

Duda, R. O., Hart, P. E. and Stork, D. G. (2000) Pattern Classification. 2nd edn. New York, NY, USA: John Wiley & Sons, Inc. Available at: https://cds.cern.ch/record/683166/files/0471056693_TOC.pdf (Accessed: 6 July 2019).

Ghalehkhondabi, I. et al. (2017) 'An Overview of Energy Demand Forecasting Methods Published in 2005–2015', Energy Systems, 8(2), pp. 411–447. doi: 10.1007/s12667-016-0203-y.

Hernández Lalinde, J. et al. (2018) 'Sobre el uso Adecuado del Coeficiente de Correlación de Pearson: Definición, Propiedades y Suposiciones', Archivos Venezolanos de Farmacología y Terapéutica (AVFT), 37(5), p. 8. Available at: http://190.169.30.98/ojs/index.php/rev_aavft/article/view/16165 (Accessed: 6 November 2019).

Jasiński, T. (2019) 'Modeling Electricity Consumption Using Nighttime Light Images and Artificial Neural Networks', Energy. Elsevier, 179, pp. 831–842. doi: 10.1016/j.energy.2019.04.221.

Kaytez, F. et al. (2015) 'Forecasting Electricity Consumption: A Comparison of Regression Analysis, Neural Networks and Least Squares Support Vector Machines', International Journal of Electrical Power & Energy Systems, 67, pp. 431–438. doi: 10.1016/j.ijepes.2014.12.036.

Mathworks (2018) 'Neural Network Toolbox TM Getting Started Guide'. Natick, MA, USA, p. 136. Available at: https://la.mathworks.com/help/simulink/index.html.

Maybee, S. and Uri, N. D. (1979) 'Time Series Forecasting of Utility Load Duration Curves', Canadian Electrical Engineering Journal. The Canadian Society for Electrical Engineering, 4(1), pp. 4–8. doi: 10.1109/CEEJ.1979.6592401.

Mondragón, P. O. (2011) Pronóstico De La Demanda De Energía Eléctrica a Corto Plazo Utilizando Una Red Neuronal Óptima. Instituto Politecnico Nacional.

Mordjaoui, M. et al. (2017) 'Electric Load Forecasting by Using Dynamic Neural Network', International Journal of Hydrogen Energy. Elsevier, 42(28), pp. 17655–17663. doi: 10.1016/j.ijhydene.2017.03.101.

Park, D. C. and Mohammed, O. (1991) 'Artificial Neural Network Based Electric Peak Load Forecasting', in IEEE Proceedings of the SOUTHEASTCON '91. Williamsburg, VA, USA,: IEEE, pp. 225–228. doi: 10.1109/SECON.1991.147742.

Peng, T. M., Hubele, N. F. and Karady, G. G. (1990) 'Conceptual Approach to the Application of Neural Network for Short-Term Load Forecasting', in IEEE International Symposium on Circuits and Systems. New Orleans, LA, USA: IEEE, pp. 2942–2945. doi: 10.1109/ISCAS.1990.112627.

Pessanha, J. F. M. and Leon, N. (2015) 'Forecasting Long-term Electricity Demand in the Residential Sector', Procedia Computer Science. Elsevier, 55(ITQM), pp. 529–538. doi: 10.1016/j.procs.2015.07.032.

Schachter, J. and Mancarella, P. (2014) 'A Short-Term Load Forecasting Model for Demand Response Applications', in 11th International Conference on the European Energy Market (EEM14). Krakow, Poland: IEEE, pp. 1–5. doi: 10.1109/EEM.2014.6861220.

Shamsollahi, P. et al. (2001) 'A Neural Network Based Very Short Term Load Forecaster for the Interim ISO New England Electricity Market System', in PICA 2001. Innovative Computing for Power - Electric Energy Meets the Market. 22nd IEEE Power Engineering Society. International Conference on Power Industry Computer Applications (Cat. No.01CH37195). Sydney, NSW, Australia: IEEE, pp. 217–222. doi: 10.1109/PICA.2001.932351.

Singh, P. and Dwivedi, P. (2018) 'Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem', Applied Energy. Elsevier, 217(October 2017), pp. 537–549. doi: 10.1016/j.apenergy.2018.02.131.

Taylor, J. W. and McSharry, P. E. (2007) 'Short-Term Load Forecasting Methods: An Evaluation Based on European Data', IEEE Transactions on Power Systems, 22(4), pp. 2213–2219. doi: 10.1109/TPWRS.2007.907583.

UPME (2016) Proyección de la Demanda de Energía Eléctrica y Potencia Máxima en Colombia, Ministerio de Minas y Energias. Medellin. Available at: www.upme.gov.co (Accessed: 6 November 2019).

UPME (2019) Demanda y Eficiencia Energetica, Unidad de Planeación Minero Energetica. Available at: https://www1.upme.gov.co/Paginas/Demanda-y-Eficiencia-Energetica.aspx (Accessed: 6 November 2019).