FORECASTING RESIDENTIAL ELECTRICAL CONSUMPTION FOR THE CITY OF SALTA, ARGENTINA

Franco Zanek[1,2], Judith Franco[2], Marcia Mac Gaul[1]

[1] Departamento de Informática, Facultad de Ciencias Exactas, Universidad Nacional de Salta
fzanek@di.unsa.edu.ar
[2] Instituto de Investigación en Energías No Convencionales (INENCO) – Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Resumen. This work presents a model that forecasts the residential electricity demand of the City of Salta. The consumption habits of the residents of the dwellings helped the development of the proposed model, which include the turning on and off of the artifacts, aspects related to the presence or not of people in the homes, including the day-night simulation, of temperature environment and relative humidity. Surveys and interviews were conducted to characterize the electricity consumption routines in the city's housing. The techniques proposed in the theory of probability and statistics helped the development of the functions that are integrated to form the proposed model

1 Introduction

Electricity is essential for the operation of the various activities of any country, regardless of whether it is developed or is in the process of being developed. However, the increase in global energy demand and the insufficiency of the development of new energy technologies combined with social aspects such as population growth, industrialization, and globalization have created problems to satisfy the growing demand for energy.

Also, electric power cannot be stored in large quantities, so, at each moment of time, there must be a balance in the network, so that the supply is equal or slightly higher than the demand. This planning requires coordination in the production of energy, ranging from investment decisions in a generation to the transportation of electricity. Against this background, it is essential to forecast the demand for electric power most accurately, since such prediction can determine the most efficient action of each of the parties involved in the generation and distribution of electrical energy.

During the last decades, the authors have published methods that allow estimating the temporal evolution of the energy demand, with different considerations of time and space scale. This type of studies allows us to analyze the changes in demand in the face of changes in the environment, generating tools for the preparation of economic policies or the planning of the energy sector.

The works listed in the bibliography can be classified into models based on analysis of trends and models based on artificial intelligence. Among the first ones we can find the works [1 – 3] were several models based on time series are proposed to estimate the annual electricity consumption. In all these works, they use as input variables the Historical Electric Consumption and the Variation of the gross domestic product (GDP) throughout the years of study; however, [4] proposes the use of the variation of the price of electricity, an essential factor in the development of the model, because this variable is strongly related to the costs of an organization; but, it does not take into account the evolution of GDP.

The described models are presented as excellent tools for the simulation of the electrical consumption of a particular place, since the Mean Square Error (RSME) reported by the authors, varies between 3% and 10%. However, all focus only on the model that allows simulating the electrical consumption. The study of the impact that variations in inputs can have on electrical consumption cannot be observed. Also, they do not take into account the climatic variables, temperature, and relative humidity; both linked with the electrical consumption of refrigeration and heating appliances.

Within this category of Trend Analysis, we can find jobs such as [5 - 8] proposing models based on linear and multi-linear regression. These authors include in their works the same variables that were previously considered, also incorporating the ambient temperature and in the case of [7] , to estimate the power consumption of a set of offices, take into account the thermal characteristics of the building as well as the influence of the environment.

The models that take into account the variables of the environment are more active since the RSME obtained by the authors varies between 2.5% and 4%; However, these works only seek to determine the model to simulate and do not present between their future lines the study of how the electrical demand behaves in front of the variations that the input variables of the model may suffer.

Models based on Artificial Intelligence, among which we find the works of [9 – 15], use neural networks for the prognosis of electricity consumption in function of historical, climatic data, and those related to the socio-economic growth of a country (GDP, population growth).

These models have a slightly higher efficiency, since the RSME oscillates between 2% and 4.5%; nevertheless, the models based on neural networks, require more computational time, since they determine the optimal number of hidden layers, such as the number of neurons per layer; or, they determine the weights of each one of the neurons. In the works [16 – 18] present the combined study of neural networks with different heuristics, in order to determine the optimal structure of the network, to subsequently make the corresponding predictions. These works present an advance in the field of simulations, since, by incorporating artificial intelligence, they produce better results than methods based on trends.

The objective of this article is to present a tool capable of estimating residential electricity consumption and provide a solution to the decision-making process of the electricity sector of the City of Salta, in the province of Salta, in the north of the Argentine Republic. This software incorporates data from surveys conducted on electric consumption habits and their study using statistical techniques.

It contains a first section describing the work methodology used, detailing in detail the various steps contemplated for the development of the model. The following section shows the results obtained, compared with data provided by the literature. Finally, in section 3 presents the discussion of the results, conclusions and future lines of research.

2 Work Methodology

The quantitative and qualitative logics provide a working guide, given the complexity of the object of the study. Consultations with experts in the area of electrical energy of the University and professionals of the electricity distribution company of the province of Salta (EDESa), together with the surveys and interviews carried out with residents, provided the necessary tools to establish the modeling decisions. With these instruments, quantitative data matrices were elaborated, which were treated statistically, which allowed having an approach to the reality investigated and to validate the model.

2.1 Data survey

The users of the different geographical areas of the city answered interviews and surveys with the objective of knowing the consumption habits of residential homes. An aspect related to the energy consumption of each appliance present in the home were consulted, discriminating them according to the following categories:

- Illumination
- Food refrigeration
- Heating and cooling of the home
- Entertainment
- Kitchen appliances
- Clothes washing and drying

The users answered about the number of electrical devices present, the average time of ignition or the frequency of weekly use. Also, they answered on aspects related to:

1. The number of members of the family group, discriminated by age
2. Periods of time in which there are no people at home
3. Periods of time in which there is a minimal activity that requires electrical energy in the home, assuming that people are resting

These last two aspects are closely related to the electrical behavior of a home, because, the ignition of electrical appliances or their connection to the electrical network depends exclusively on human action.

The optimal size of the sample must be determined, in such a way that the results obtained have statistical validity and the error committed is the least possible. In this line, the works [19 – 20] develop the necessary methods to determine the sample size. The following expression was used to calculate the sample size:

$$n = \frac{Z_\alpha^2 N \sigma^2}{e^2(N-1) + Z_\alpha^2 \sigma^2} \qquad (1)$$

Where:

- n: sample size

- N: size of the population
- e: sample error that can be accepted
- σ^2: standard deviation of the population. The literature establishes that this value should be 0.5, in case of not knowing its value.
- Z_α^2: constant that depends on the level of confidence that we assign. The values most used in the literature are:

Table. 1. Most used values for Z_α^2 [19]

Z_α^2	1.15	1.28	1.44	1.65	1.96	2	2.58
Confidence level	75%	80%	85%	90%	95%	95,5%	99%

The sample must integrate the most extensive variety of housing types in the city, so it was necessary to consider the classification of the homes that EDESA proposes, according to its rate schedule, for conducting surveys and interviews. The total number of homes provided by EDESA was considered taking into account a 95% confidence.

Table. 2. Categorization of housing (EDESA) and number of homes to survey

Category	Maximum monthly consumption (kWh)	Total Households	Sub-Total to be surveyed
R1	Until 192	160570	391
R2	Between 192 and 500	116964	391
R3	Between 500 and 700	13382	381
R4	Between 700 and 1400	7765	373
R5	More than 1400	1319	302
		Total to be surveyed:	1838

2.2 Climate data simulation

The variables related to climate play a fundamental role when forecasting energy consumption since these variables define the thermal comfort zone, which determines whether or not a person wants to turn on and turn off a heating or cooling appliances.

For this reason, it was necessary to introduce to the model the necessary procedures to be able to estimate both temperature and relative humidity.

In [21] is proposed a procedure to simulate the relative humidity of a city, considering the same in time periods. For this, the minimum and average temperature are taken into account until the period before the one to be simulated. The following equation describes the procedure:

$$HR = 100\frac{e_s(T_n)}{e_s(T_m)} \quad \textbf{(2)}$$

Where $e_s(T_n)$ corresponds to the saturation pressure (hPa), determined by the minimum temperature of the day. While $e_s(T_m)$ it also corresponds to the saturation pressure but calculated from the hourly temperature. The expression that allows calculating these pressures is:

$$e_s(T) = 6{,}11 * e^{25{,}22(1-\frac{213{,}16}{T})} * \left(\frac{213{,}16}{T}\right)^{5{,}31} \quad \textbf{(3)}$$

This procedure is valid with the data provided by the INTA, obtaining a value of the mean square error (RMSE) of 2.66. Therefore, the procedure is considered valid from the statistical point of view, to simulate the relative humidity of the city of Salta.

In [22] several methods are proposed for the simulation of the hourly temperature of a city. For this, it is necessary to have the minimum and maximum temperature of the previous day and the apparent solar time of the period to be simulated.

$$T_{ab} = T_1 + T_2\cos\left((14 - AST)\frac{\pi}{12}\right) \quad \textbf{(4)}$$

Where

- T_{ab} corresponds to the ambient temperature

- AST corresponds to the apparent solar time
- T_1 y T_2 correspond to the output of the following equations:

$$T_1 = \left(\frac{T_{max} + T_{min}}{2}\right) \quad \textbf{(5)}$$

$$T_2 = \left(\frac{T_{max} - T_{min}}{2}\right) \quad \textbf{(6)}$$

This procedure allows to calculate only for a particular day; then it was modified to simulate the hourly temperatures of a whole year.

$$T_{max}(i+1) = T_{max}(i) + value1 \quad \textbf{(7)}$$

$$T_{min}(i+1) = T_{min}(i) + value2 \quad (8)$$

Where, at the maximum and minimum temperatures to be considered for the calculation of the next day, they add value1 and value2, which are obtained from the analysis of the distributions that follow the monthly variations of the temperatures. This study is carried out based on the data provided by INTA.

The value of 4.23 obtained for the RSME allows us to ensure that this model can be considered valid from the statistical point of view, to simulate the hourly temperature during a whole year.

2.3 Sunrise and Sunset

Another essential aspect incorporated into the model, are the procedures that calculate the sunrise and sunset since these values are closely related to the time interval of turning on or off of appliances, the following formulas were used [23]

$$h_{ss} = 12 - \frac{\omega_s}{15} \quad (9)$$

$$h_{ps} = 12 + \frac{\omega_s}{15} \quad (10)$$

Where ω_s corresponds to the hour angle to the sunrise and sunrise and is calculated as:

$$\omega_s = \operatorname{acos}(-\tan\delta * \tan\phi) \quad (11)$$

From this last equation is defined

- Φ as the latitude of the place
- δ as the declination obtained from the following formula, which has as input the day of the year, expressed in Julian time.

$$\delta = 23{,}45\, sen\left(360\frac{284+n}{365}\right) \quad (12)$$

The results obtained by equations 10 and 11 are expressed regarding the Real Solar Time (HSR), but what we need to work is the Apparent Solar Time (HSA), and for this, the following equation is used:

$$HSA = HSR - 4\ (L_s - L_E) - E_t \quad (13)$$

Where,

- L_s corresponds to the reference meridian
- corresponds to the meridian of the place
- corresponds to the equation of time that is obtained from the formula:

$$E_t = 229{,}2 * (0{,}000075 + 0{,}001868 \cos B - 0{,}032077\ sen\ B - 0{,}014615 \cos 2B - 0{,}04089\ sen\ 2B) \tag{14}$$

Finally, the value of B corresponds to the daily angle obtained from:

$$B = (n-1)\frac{360}{365} \tag{15}$$

In the last equation, n corresponds to the day of the year expressed in Julian time.

2.4 On/off switching of appliances

The next step was to simulate the on/off switching of appliances, to calculate the energy consumed. The appliances were divided into the following categories; to determine the period an appliance is on:

Appliances that are connected to the electrical network throughout the day, such as wireless phones and modems. For them, the ignition time is 24 hours a day.

Appliances whose operation depends on the amount of natural light that exists at a certain time of day. They are artifacts that are characterized because their operation has a close relationship between the current time and the hours of sunrise and sunset. In periods of low natural light, the appliance tends to be lit longer, than in periods of natural light abundance. The results of equations 9, 10 and 13 of section 2.3, allow simulating the on/off switching of appliances. Within this category are the artifacts related to the lighting of the house.

Appliances related to refrigeration of the dwellings. It is necessary to simulate the hourly temperature, and relative humidity, with the equations, developed in section 2.2, to determine the on/off switching of appliances. These values together with the psychometric charts - in particular with the region called comfort zone1 - made it possible to define the procedure for this category. Given a value for the temperature and a

[1] Thermal comfort is defined in ISO 7730 Norm as “Condition of the ambient in which satisfaction with the thermal ambient is expressed.” relative temperature and humidity are the most important variables in this definition.

value for the relative humidity, if these were within the comfort zone2, the appliance was expected to be off, or if it were on, the procedure would stimulate its off/switching. Otherwise, the appliance must be turned on according to the relation between the variables and the comfort zone.

Appliances related to washing and drying clothes. The amount of use per day and the most used washing cycle was used to determine the operating time of the appliance. For artifacts related to drying, the algorithm has an extra control condition, which considers ignition only if the use of a clothes washer is previously registered.

The following category covers all the devices that have an internal battery giving them autonomy once they have been disconnected from the electrical network. For these devices, switching on depends solely on the battery running out or being about to run out. The procedure determines whether the appliance is not within its autonomy period, to be connected again to the electrical network. For example, cell phones are included in this category

Finally, for the remaining appliances, the operating time is randomly generated, regardless of the cases described above.

Note that each of these procedures is carried out as long as the appliance to be turned on exists in the household and that the people are at home and they are awake.

2.5 Generation of the households

One of the most important processes is that which is responsible for the generation of the houses, that is, the one responsible for determining which artifacts a specific housing unit will contain, as well as the quantity. For this process, the results of the surveys and interviews, described in section 2.1, were analyzed and the density functions corresponding to each of the random variables were determined, defined as the number of Appliances of a specific type.

[2] For the city of Salta, the comfort zone is defined between the temperature values of 20 and 35 degrees Celsius and the relative humidity between 30% and 60% (Psychometric charts of the Province of Salta)

Once the generation process is finished, the results obtained are verified. Because it is a random process, it is necessary to impose restrictions at the time of generation. This restriction is that the monthly consumption of the house generated is within limits established for the category to which it corresponds, according to the EDESA rate chart. For this purpose, the house was simulated for June and January, where the consumption maximums related to the winter and summer periods, are produced. This situation is justified by the fact that these months are the coldest and warmest months, and are the months where people tend to use heating and cooling appliances, which consume the most energy, in a higher proportion. In case the generated housing exceeds the maximum allowed consumption for the assigned category, the algorithm discards the generated housing and starts this process again; repeating until the condition mentioned above is met.

2.6 Description of the model

Our proposal is characterized as a discrete, dynamic and stochastic model. It is discreet because it studies the system at a fixed time interval of one minute. It is dynamic because it simulates the course of time regarding minutes, hours, months and years. It is stochastic because randomness lies in the generation of entities, based on the probability distributions derived from the data survey of the real system.

The pseudo-code of the model is shown below:

```
// Generation of climate data for all homes for the entire year
      Repeat for the 365 days of the year
            Generate temperature and hourly humidity (Section 2.2)

// Generation of the households
     Repeat for the 5 Categories
            // In proportion to the distribution of the city
            Repeat for each house in the Category
                   Generate and verify Housing (Section 2.5)

// Simulation of city housing for one year and calculation of electricity consumption
     Repeat for the 365 days of the year
          Repeat for each of the house generated
            Repeat for every minute of the day
               Determine on/off switching of each appliance (Sección 2.4)
           Calculate the daily consumption
            Increase day
            Accumulate monthly consumption
            If (the month changes)
               Show monthly consumption
```

3. Results

To validate the software, we compared the results with the data provided by EDESA (Data Consumption). The data correspond to residential consumption, of all housing in the city of Salta for three years. The results obtained were as follows:

Table. 3. Results

	Consumption 2014 (GWh)	Consumption 2015 (GWh)	Consumption 2016 (GWh)	SimulatedConsumption (GWh)
January	74.89	76.01	84.57	83.04
February	59.61	67.00	75.44	77.29
March	61.08	70.47	73.02	73.18
April	58.11	63.88	69.16	74.45
May	62.61	67.47	80.99	81.77
June	68.25	71.88	88.15	84.58
July	72.18	78.95	83.10	81.97
August	65.78	71.21	72.67	74.90
September	61.36	65.25	70.03	73.30
October	68.37	70.08	72.45	75.42
November	68.24	72.57	72.30	77.10
December	72.11	81.73	83.32	81.10

The value of RMSE between the simulated data and the data for the year 2014 is 12.77. The RMSE value between the simulated data and the data for the year 2015 is 8.05. Finally, the value of the RMSE between the simulated data and the data for the year 2016 is 2.9. The average value of the RMSE for the three years is 7.9.

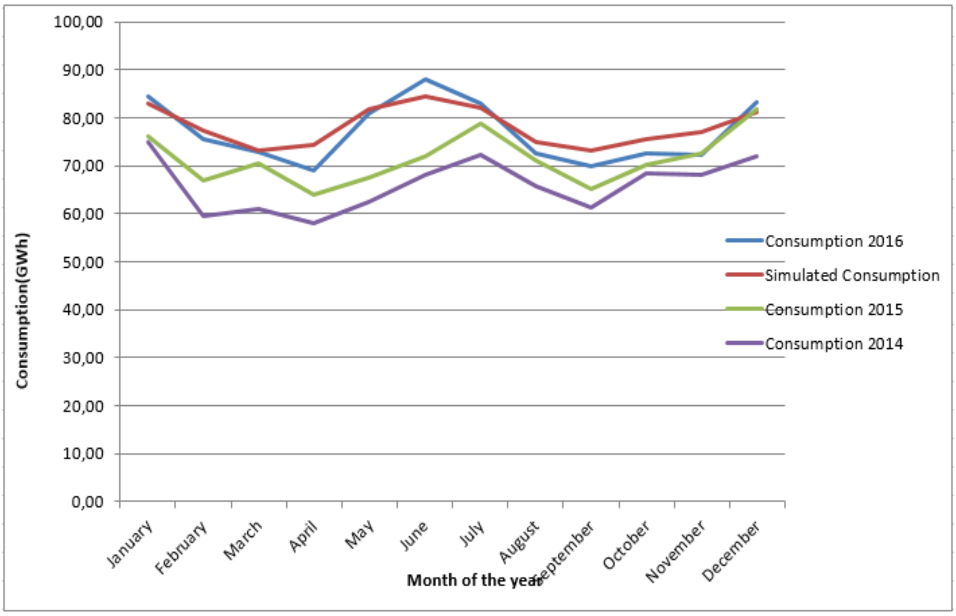

Fig. 1. Validations of resutls

Figure 1 shows the comparison between the real values of the years 2014 to 2016 and the simulated values. The execution was carried out on an 8 GB RAM computer with Intel Core 7 processor. Regarding the computation times, the following table shows the times, discriminated by the average generation time of the housing type and the average time of simulation.

Table. 4. Generation and Simulation times

Category	Generation Time (seg)	Simulation Time (seg)
R1	499.03	49.25
R2	151.41	36.24
R3	177.22	36.29
R4	154.60	36.15
R5	508.23	36.45

4. Conclusions and future work

A stochastic model, which is dynamic and discrete, has been proposed for simulating the electricity consumption of the households in the city of Salta, where the codification of the households and the on/off switching of appliances is done at random. By adopting a randomized generation of households, the implementation of an additional procedure is required to validate the households within each category according to the current rate chart.

From the values shown by the RMSE, it can be concluded that the results obtained are statistically valid and therefore the proposed model reflects the reality of the consumption habits of the city of Salta; However, it can be seen that the execution of the model takes much time in the generation of housing. The program requires several runs to adjust it to the corresponding category, because when each house is generated randomly. On the other hand, the simulation of the on/off switching, to determine the monthly consumption, is relatively fast in computational time, compared with the generation times. In the future, it is expected to complement the program with tools that allow graphing the geo-positioning, to analyze the relationship between the geographical position of a house within the city and energy consumption. Also, the development of more efficient algorithms regarding computational times, but above all looking for efficiency in the results, which allows its subsequent application in the city. Within these future studies, we propose the development of a model based on neural networks to make predictions.

The model will also be complemented so that it makes the estimations of the whole city including all sectors of the social actors, to be able to have a complete vision of the consumption habits of the city. In turn, complement this study with the realization of experiments that allow analyzing the behavior of the total electrical demand, depending on the variation of the input parameters.

References

1. Kumar, U., & Jain, V. K. (2010). Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. Energy, 35(4), 1709–1716. https://doi.org/10.1016/j.energy.2009.12.021.
2. Al-Shobaki, S., & Mohsen, M. (2008). Modeling and forecasting of electrical power demands for capacity planning. Energy Conversion and Management, 49(11), 3367–3375. https://doi.org/10.1016/j.enconman.2008.05.005.

3. Magazzino, C. (2014). Electricity demand, GDP and employment: Evidence from Italy. Frontiers in Energy, 8(1), 31–40. https://doi.org/10.1007/s11708-014-0296-8.

4. DilaverZafer, Z., Hunt, L. C. (2011). Industrial electricity demand for Turkey: A structural time series analysis. Energy Economics, 33(3), 426–436. https://doi.org/10.1016/j.eneco.2010.10.001.

5. Mohamed, Z., &Bodger, P. (2005). Forecasting electricity consumption in New Zealand using economic and demographic variables. Energy, 30(10), 1833–1843. https://doi.org/10.1016/j.energy.2004.08.012.

6. Braun, M. R., Altan, H., & Beck, S. B. M. (2014). Using regression analysis to predict the future energy consumption of a supermarket in the UK. Applied Energy, 130, 305–313. https://doi.org/10.1016/j.apenergy.2014.05.062.

7. Vu, D. H., Muttaqi, K. M., &Agalgaonkar, A. P. (2015). A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. Applied Energy, 140, 385–394. https://doi.org/10.1016/j.apenergy.2014.12.011.

8. Mottahedi, M., Mohammadpour, A., Amiri, S. S., Riley, D., &Asadi, S. (2015). Multi-linear Regression Models to Predict the Annual Energy Consumption of an Office Building with Different Shapes. In Procedia Engineering (Vol. 118, pp. 622–629). https://doi.org/10.1016/j.proeng.2015.08.495.

9. Al-Hamadi, H.M., &Soliman, S.A. (2005). Long-term/mid-term electric load forecasting based on short-term correlation and annual growth. Electric Power Systems Research 74(3), 353-361.

10. Rodrigues, F., Cardeira, C., &Calado, J. M. F. (2014). The daily and hourly energy consumption and load forecasting using artificial neural network method: A case study using a set of 93 households in Portugal. In Energy Procedia (Vol. 62). https://doi.org/10.1016/j.egypro.2014.12.383.

11. Yokoyama, R., Wakui, T., &Satake, R. (2009). Prediction of energy demands using neural network with model identification by global optimization. Energy Conversion and Management, 50(2), 319–327. https://doi.org/10.1016/j.enconman.2008.09.017.

12. Kankal, M., Akpınar, A., Kömürcü, M. İ., &Özşahin, T. Ş. (2011). Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. Applied Energy, 88(5), 1927–1939. https://doi.org/10.1016/j.apenergy.2010.12.005.

13. Sözen, A. (2009). Future projection of the energy dependency of Turkey using artificial neural network. Energy Policy, 37(11), 4827–4833. https://doi.org/10.1016/j.enpol.2009.06.040.

14. Martellotta, F., Ayr, U., Stefanizzi, P., Sacchetti, A., &Riganti, G. (2017). On the use of artificial neural networks to model household energy consumptions. In En-

ergy Procedia (Vol. 126, pp. 250–257). https://doi.org/10.1016/j.egypro.2017.08.149.

15. González, P. A., & Zamarreño, J. M. (2005). Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. Energy and Buildings, 37(6), 595–601. https://doi.org/10.1016/j.enbuild.2004.09.006.
16. Mishra, S. K., Panda, G., &Meher, S. (2009). Multi-objective particle swarm optimization approach to portfolio optimization. In 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC) (pp. 1612–1615). https://doi.org/10.1109/NABIC.2009.5393659.
17. Shangdong, Y., & Xiang, L. (2007). A new ANN optimized by improved PSO algorithm combined with chaos and its application in short-term load forecasting. In 2006 International Conference on Computational Intelligence and Security, ICCIAS 2006 (Vol. 2, pp. 945–948). https://doi.org/10.1109/ICCIAS.2006.295400.
18. Bashir, Z. A., & El-Hawary, M. E. (2007). Short-Term Load ForecastingUsing Artificial Neural Network BasedonParticleSwarmOptimizationAlgorithm. In 2007 Canadian ConferenceonElectrical and ComputerEngineering (pp. 272–275). https://doi.org/10.1109/CCECE.2007.74
19. García-García, J. A., Reding-Bernal, A., & López-Alvarenga, J. C. (2013). Cálculo del tamaño de la muestraeninvestigacióneneducaciónmédica. InvestigaciónEnEducaciónMédica, 2(8), 217–224. https://doi.org/10.1016/S2007-5057(13)72715-7.
20. Valdivieso, C., Valdivieso, R., &Valdivieso, O. (2011). Determinación del tamaño muestral mediante el uso de árboles de decisión. UPB- Investigación&Desarrollo, 11(2000), 148–176. https://doi.org/10.1007/s13398-014-0173-7.2.
21. Meira, G. R., Padaratz, I. J., Alonso, C., & Andrade, C. (2003). Efecto de la distancia al mar en la agresividadporclorurosenestructuras de hormigónen la costa brasileña. Materiales de Construcción, 53(271–272), 179–188. Retrieved from http://materconstrucc.revistas.csic.es/index.php/materconstrucc/article/view/302/347.
22. Chabane, F., Moummi, N., Brima, A., &Moummi, A. (2016). Prediction of the theoretical and semi-empirical model of ambient temperature. Frontiers in Energy, 10(3), 268–276. https://doi.org/10.1007/s11708-016-0413-y.
23. J.A. Duffie, W.A. Beckman, Solar Engineering of thermal processes, Editorial Wiley & Sons, INC.
24. Psychometric chart of the City of Salta.
25. [dataset] Data Consumption. Power distribution plant of Salta province (EDESA).