DATA ANALYSIS ON NON-RESIDENTIAL ELECTRICITY CONSUMPTION BY STATISTICAL AND MATHEMATICAL TECHNIQUES IN VIEW OF DEVISING APPROPRIATE CONSUMPTION STRATEGIES

George CĂRUȚAȘU [1]
Alexandru PÎRJAN [2]
Cristina COCULESCU [3]
Justina Lavinia STĂNICĂ [4*]
Mironela PÎRNĂU [5]

Abstract: *The aim of this paper is to analyze, process and interpret, from economic and statistical perspectives, the data regarding the quantity of electric energy, measured at the non-residential consumers' level. Our intention was to track and analyze the electric energy consumption level at hourly intervals for a real consumer in Romania. The measurements were carried out in MWh, and collected in databases, in order to facilitate the application of the calculation methods. The results and their interpretations facilitate the scientific substantiation of new policies in order to optimize the electric energy consumption. The statistical and mathematical methods employed represent viable tools in achieving an adequate data analysis on non-residential electricity consumption in view of devising appropriate consumption strategies. These will be transmitted and proposed to the real consumer as scenarios of its analyzed consumption profile. After having experimented several methods for approximating the data repartition, we have concluded that by adjusting the primary data with the estimated normal repartition one obtains the ideal model in the case of hourly electricity consumption of non-residential consumers offering valuable insights regarding the modelling of their consumption patterns.*

Keywords: *data analysis, grouping intervals, optimizing electricity consumption, intervals of variation.*

[1] Prof. PhD. habil., Faculty of Computer Science for Business Management, Romanian-American University, Bucharest, Romania, carutasu.george@profesor.rau.ro
[2] Prof. PhD. habil., Faculty of Computer Science for Business Management, Romanian-American University, Bucharest, Romania, alex@pirjan.com
[3] Assoc. prof. PhD., Faculty of Computer Science for Business Management, Romanian-American University, Bucharest, Romania, coculescu.cristina@profesor.rau.ro
[4*] corresponding author, Lecturer PhD., Faculty of Computer Science for Business Management, Romanian-American University, Bucharest, Romania, stanica.lavinia.justina@profesor.rau.ro
[5] Assoc. prof. PhD., Faculty of Informatics, Titu Maiorescu University, Faculty of Computer Science for Business Management, Romanian-American University, Bucharest, Romania, mironela.pirnau@prof.utm.ro

Data Availability Statement

Data available on request from the author: The data that support the findings of this study are available from the corresponding author upon reasonable request.

1. Introduction

Our study is based on data regarding the electric energy consumption of a non-residential consumer in Romania. The data that are part of the present study were collected during January-December 2016. The measurements were performed using specialized smart metering devices situated at the nonresidential consumers' locations and stored in databases dedicated to the analyzed field. The measurements sampling was carried out on an hourly basis over the entire period of the calendar year.

The authors' concerns regarding forecasting energy consumption and using the obtained results in reducing it, can be seen in the previous researches that they have done on residential households. In (Oprea, Pîrjan, Căruțașu, Petroșanu, Bâra, Stănică, & Coculescu, 2018) a mixed neural network approach has been used in order to provide an accurate method for forecasting the residential electricity consumption in smart homes complexes, using data recorded by sensors. The developed method was validated and further compiled, the idea being to incorporate it in the IoT cloud solution that was proposed in (Stănică, Căruțașu, Pîrjan, & Coculescu, 2018). The solution here was to optimize the electricity consumption and costs of households, based on analyzing disparate data collected from sensors and home appliances in smart homes.

In Europe, non-residential buildings represent 25% of the total building stock and are considered to be more heterogenous and more complex than residential buildings (Droutsa, Balaras, Dascalaki, Kontoyiannidis, & Argiriou, 2018). Out of these, the retail and wholesale buildings represent the leading sector, with 28% of non-residential stock floor area. However, according to the same paper, the available data and the studies that track energy performance in non-residential buildings are more limited compared to those for households.

Nevertheless, the existing reviews show that the research community is making efforts in this direction. Miller, Nagy, and Schlueter (2018) have done a review of 100 publications that used unsupervised machine learning techniques in order to analyze the performance of non-residential buildings. Most of the publications being reviewed focused on energy performance. The conclusions show that clustering algorithms (particularly k-means clustering) and visual analytics are commonly used, but other procedures and techniques are worth exploring as well.

In a similar study (Ruparathna, Hewage, & Sadiq, 2016), a number of research articles focusing on increasing energy efficiency in commercial and institutional buildings were reviewed. The study included only articles published in well-reputed journals. Three main approaches in the literature were identified, concentrating on technical, organizational, and behavioral changes. As an outcome of the comprehensive review, the authors proposed a strategy map for improving buildings energy performance, stating that their findings could set the basis for developing national and organizational strategies in this direction.

Other studies focused on identifying the most performant techniques for modelling and forecasting the energy consumption. Tso and Yau (2007) made a comparison between three different techniques for predicting energy consumption: regression analysis, decision tree, neural networks. In order to choose the best one, the authors suggested the idea of developing a platform that implements different models and therefore can assess their prediction performances.

Another article on electrical consumption forecasting methods, authored by Daut et al. (2017), is focusing on both conventional and artificial intelligence methods, comparing the performance of both of them. The article concluded that a hybrid of the two forecasting techniques could lead to better results.

Covering the same topic, Zhao and Magoules (2012) evaluated different models for energy consumption prediction, including statistical, engineering, and artificial intelligence models, and at the same time, emphasized the difficulty of making such predictions, since there are many factors that can influence them and must be taken into consideration.

Perez-Chacon, Talavera-Llames, Martinez-Alvarez, and Troncoso (2016) analyzed a big time series of data collected from the electricity consumption of two university buildings over a period of three years. For establishing patterns, the authors used the distributed version of k-means clustering algorithm for Apache Spark, for which they also tested its computational performance.

For the residential sector, the prediction of energy consumption is modelled in Fumo and Biswas (2015), which used simple and multiple linear regression analysis on hourly and daily collected data from a household. Also this paper promotes the idea of developing a user-friendly software for modelling and forecasting the energy consumption.

Another research direction aims to identify the factors that influence the electric energy consumption. In Ma et al. (2017) the authors perform a case study on a number of public non-residential buildings in China, by analyzing their energy consumption patterns and the factors influencing it. Similarly, Gutiérrez-Pedrero, Tarancon, del Rio, and Alcantara (2018) also focused on determining the main factors influencing electricity consumption of non-residential sector, their results showing that higher technological progress and higher electricity retail prices lead to a reduction of the consumption intensity.

By analyzing the existing body of knowledge, one can identify a necessity, a clear need for modeling the variation of non-residential electricity consumption covering various time intervals in order to identify specific consumption profiles. Therefore, the main objective of this analysis was to find the variation mode of the electric energy consumption for various time intervals as shown in Sections 2 and 3, in order to identify specific consumption profiles. Our research was aimed at identifying the main statistical sizes for modelling the collected data. For a more accurate analysis, collected data were stored in a table having the fields: month, day-number, hour-number, and energy-consumption-MWh.

The reminder of the paper is structured as follows: Section 1 presents the statistical and mathematical methods and techniques for analyzing data of electric energy consumption, Section 2 contains the processing and results, in Section 3 is presented the data analysis by grouping on intervals of variation, Section 4 contains the computer model for data analysis, followed by the Conclusions Section.

2. Statistical and mathematical methods and techniques for analysing data of electric energy consumption

Since the data in this study are linearly distributed at one-hour intervals over a calendar year, we have tracked their statistical behaviour in the case of grouping on equal intervals of variation. The statistical and mathematical methods and techniques applied in the present study allowed us to develop a specific computer model, in which we identified:

- The amplitude of variation of the general overall consumption (C) on an hourly basis during January-December 2016, using the equation (1):

$$A = C_{max} - C_{min}. \tag{1}$$

The number of groups, using Sturges' formula (Sturges, 1926; Scott, 2009)

$$k = [1 + 3.322/lg\, n], \tag{2}$$

where n, in this case, has the value of 24, i.e. the number of hours analyzed on a daily basis.

- The size of the grouping interval, denoted by h, which represents the ratio between the consumption amplitude and the identified k number of groups, was determined, the calculation formula being equation (3)

$$h = A/k. \tag{3}$$

Based on the statistical and economic support for the repartition of the value intervals samples, we used rounded intervals in order to carry out the calculations. Under these conditions, we identified the size of the grouping interval as 91 MWh.

- Starting from the minimum value of the determined sum and the size of the identified grouping interval, we constructed the vectors of the minimum and maximum limits of the grouping intervals. Based on these vectors, the grouping of the data on the electric energy consumption was made, in order to build the statistical indicators specific to the analysis of the value series on intervals of variation. The vectors of the grouping intervals limits (L) are input variables in the mathematical-computer model presented in Section 4.

$$L_{min} = [C_{min}, C_{min} + h, \dots, C_{min} + (k-1) * h),$$
$$L_{max} = [C_{min} + h, C_{min} + h, \dots, C_{min} + (k-1) * h] \,.$$

- The center of each analysed interval was identified as the simple arithmetic mean of the interval bounds, according to equation (4):

$$c_i = (c_{imin} + c_{imax})/2. \qquad (4)$$

- The absolute frequency of each group (n_i) was calculated; this is equal to the number of statistical units having the value of the characteristic greater than or equal to the lower limit of the interval and less than or equal to the upper limit.

Subsequently, based on the absolute frequencies, the ascending and descending cumulative absolute frequencies at each group level were identified. Similarly, ascending and descending cumulative relative frequencies could be determined. The absolute, relative, and cumulative frequencies represent the support that allows the identification of the overall behaviour of the distribution of values in collectivity, especially of the central tendency to normality of the frequency repartition.

Systematization of data on electric energy consumption in 11 equal intervals of variation, as well as the statistical and economic interpretation and construction of histograms (Scott, 1979) and curves of cumulative frequencies, are presented in the results section.

When applying the selection method, the most common situations are those in which the theoretical repartition law is normal N (m, σ) (Purcaru, 1997). For selections from statistical populations with normal repartitions, the probability theory states the following results:

Theorem 1. If $\{X_1, X_2, \dots, X_n\}$ is a selection of volume n in a statistical population characterized by a random variable that follows a normal distribution $N(m, \sigma)$, then the selection mean has a normal repartition of mean m and standard deviation $\frac{\sigma}{\sqrt{n}}$, i.e.:

$$\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n} \in N\left(m, \frac{\sigma}{\sqrt{n}}\right). \tag{5}$$

Theorem 2. If $X_1, X_2, \ldots, X_n$ are normally distributed random independent variables, $X_k \in N(m_k, \sigma_k)$, $k \in \overline{1,n}$, and $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$, then the random variable

$$Y = \sum_{k=1}^{n} \alpha_k X_k \in N\left(\sum_{k=1}^{n} \alpha_k m_k, \sqrt{\sum_{k=1}^{n} \alpha_k^2 \sigma_k^2}\right). \tag{6}$$

In particular, if $\alpha_1 = \alpha_2 = \cdots = \alpha_n = \frac{1}{n}$, we have:

$$Y = \frac{\sum_{k=1}^{n} X_k 1}{n} \in N\left(\frac{\sum_{k=1}^{n} m_k}{n}, \frac{\sqrt{\sum_{k=1}^{n} \sigma_k^2}}{n}\right). \tag{7}$$

From the estimation theory (Popescu, 1993), we know that the selection mean $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$ is a fixed, consistent and efficient estimator for the mean m of the general statistical population, and the dispersion of selection $S^2 = \frac{\sum_{k=1}^{n}(X_k-\bar{X})^2}{n}$ represents a sufficiently consistent estimator for the dispersion σ^2 of the general population (Popovici, 2015). In case of small volume selections, the dispersion σ^2 is evaluated with the corrected dispersion of selection, given by the formula $s^2 = \frac{\sum_{k=1}^{n}(X_k-\bar{X})^2}{n-1}$.

3. Processing and results

For reasons related to the rigor of the statistical analysis, as well as to facilitate the calculation process for limiting the field of error propagation (measurement, calculation, method), we used calculation approximations in certain data processing and analysis. When processing the data, we have used the following hardware configuration: the ASUS Rampage V Extreme motherboard, the central processing unit Intel i7-5960x with 32 GB DDR4 quad channel and the GeForce GTX 1080 TI NVIDIA graphics card. The software configuration that we have used consists in the Windows 10 Educational Version 1803 operating system. Starting from the initial data underlying the present study, and from the mathematical model in section 1, we have calculated in Table 1 statistical and mathematical indicators for data systematization.

Table 1. Statistical and mathematical indicators for data systematization

Intervals of variation of electric energy consumption	Value of class 1	Value of class 2	Number of hours frequency	Percentage	Center of interval	Ascending cumulative absolute frequencies	Descending cumulative absolute frequencies
887.02 – 978.02	887.02	978.02	5	0.21	932.52	5	24
978.02 – 1069.02	978.02	1069.02	2	0.08	1023.52	7	19
1069.02 – 1160.02	1069.02	1160.02	0	0	1114.52	7	17
1160.02 – 1251.02	1160.02	1251.02	1	0.04	1205.52	8	17
1251.02 – 1342.02	1251.02	1342.02	1	0.04	1296.52	9	16
1342.02 – 1433.02	1342.02	1433.02	1	0.04	1387.52	10	15
1433.02 – 1524.02	1433.02	1524.02	1	0.04	1478.52	11	14
1524.02 – 1615.02	1524.02	1615.02	0	0	1569.52	11	13
1615.02 – 1706.02	1615.02	1706.02	2	0.08	1660.52	13	13
1706.02 – 1797.02	1706.02	1797.02	4	0.17	1751.52	17	11
1797.02 – 1888.02	1797.02	1888.02	7	0.29	1842.52	24	7

Table 1 contains the main numerical characteristics that allow the statistical and mathematical systematization of the recorded values for the intervals of variation of electric energy consumption, number of hours frequency, ascending cumulative absolute frequencies, descending cumulative absolute frequencies etc.

The statistical results led to the histograms represented by the Figures 1 and 2.

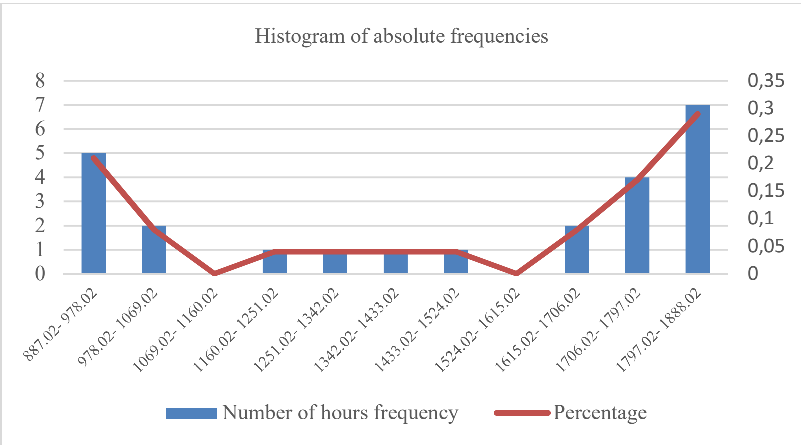

Figure 1. Histogram of the absolute frequencies calculated based on data in Table 1

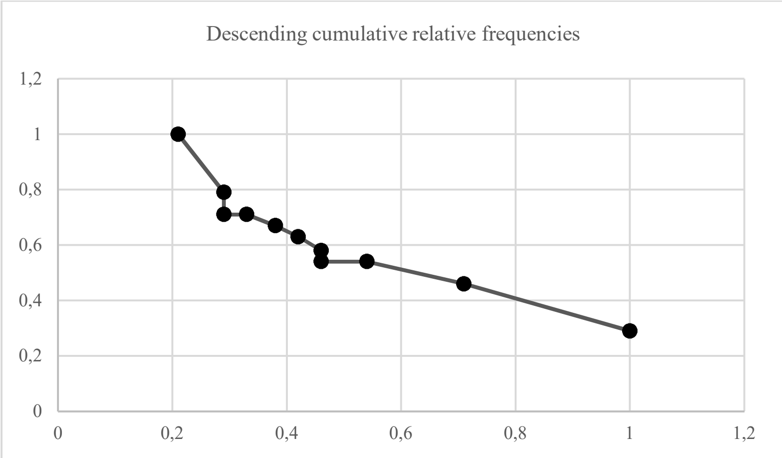

Figure 2. Descending cumulative relative frequencies

Figure 1 highlights the fact that since there are two grouping intervals with null absolute frequency, then it is necessary to remake the systematization.

Based on the experience gained from the analysis of previous studies in the electric energy field, we have reduced the number of grouping intervals to avoid the excessive fragmentation of the processed statistical collectivity. Thus, by using 6 grouping intervals corresponding to an amplitude of h = 180, Table 2 resulted.

Table 2. Statistical and mathematical indicators for amplitude h=180

Intervals of variation of electric energy consumption	Absolute frequencies (number of hours)	Relative frequencies (percentage)	Ascending cumulative absolute frequencies	Descending cumulative absolute frequencies	Ascending cumulative relative frequencies
887.02 – 1067.02	6	0.25	6	24	0.25
1067.02 – 1247.02	2	0.08	8	18	0.33
1247.02 – 1427.02	2	0.08	10	16	0.42
1427.02 – 1607.02	1	0.04	11	14	0.46
1607.02 – 1787.02	6	0.25	17	13	0.71
1787.02 – 1967.02	7	0.29	24	7	1

By analyzing the results in Table 2, one can observe that the possibilities of the occurrence of null absolute frequencies were eliminated.

Corresponding to the values calculated in Table 2, histograms for absolute frequencies, relative frequencies, and ascending cumulative relative frequencies are shown in Figures 3 and 4.

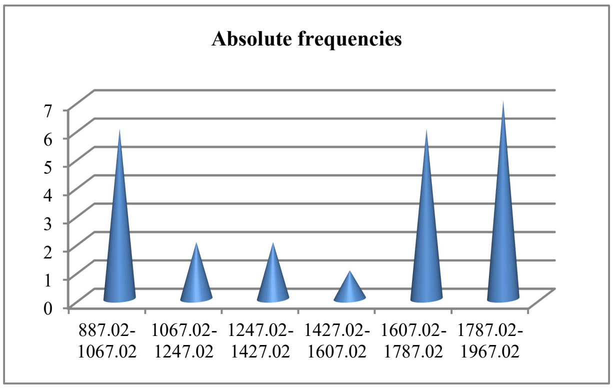

Figure 3. Absolute frequencies (number of hours)

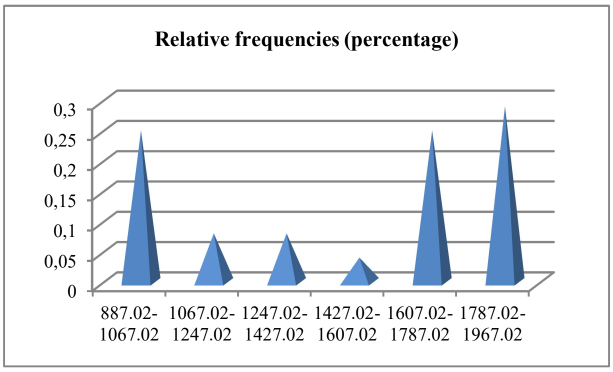

Figure 4. a – Relative frequencies

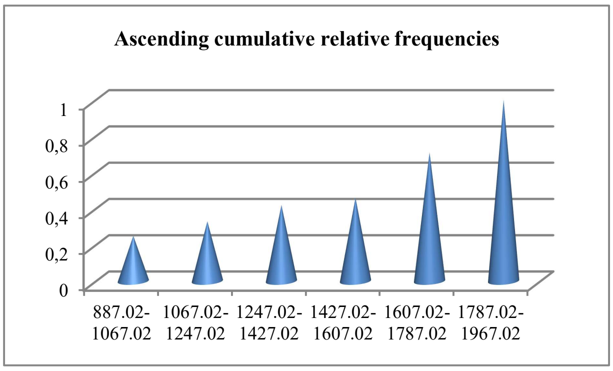

Figure 4. b – Ascending cumulative relative frequencies

The charts of histograms (Feedman & Diaconis, 1981) and cumulative frequencies indicate that the distribution of hourly electric energy consumption within a full 24-hour horizon has a normal tendency. Our research aimed, for the argumentation of the normality hypothesis of theoretical repartition, to apply a concordance test, by which we verified the possibility of concordance between the data provided on the experience and the hypothesis made on the form of the theoretical repartition law.

For the application of the concordance tests, the selection repartition function is determined in advance, based on the observed data, grouped by intervals and expressed using the relative frequencies and the cumulative relative frequencies. Subsequently, the selection repartition function is compared with the hypothetical theoretical repartition of the general population (Poisson, binomial, exponential, normal repartition). The literature mentions several methodologies (Sivilevičius, Vislavičius & Bražiūnas, 2017; Teodorescu, 2015; Ahmad, Ahmed, Vveinhardt & Streimikiene, 2016) for the implementation of these studies: Pearson's χ^2 test, Kolmogorov-Smirnov's test.

In the case of normal repartition, Kolmogorov is one of the most used tests of concordance. According to this test, the selection repartition function of the observed data noted as $F_n^*(x)$ is compared to the hypothetical theoretical repartition of the general population noted as $F_0(x)$:

- if $max|F_0(x) - F_n^*(x)| < \frac{\lambda_\alpha}{\sqrt{n}}$, then there is concordance between $F_n^*(x)$ and $F_0(x)$ and the hypothesis H_0: $F(x) = F_0(x)$ is accepted;
- if $max|F_0(x) - F_n^*(x)| \geq \frac{\lambda_\alpha}{\sqrt{n}}$, then there is no concordance between $F_n^*(x)$ and $F_0(x)$ and the hypothesis H_0 is rejected,

where, to the given significance threshold α it corresponds, by the formula $K(\lambda_\alpha) = 1 - \alpha$, a value of λ_α, such that, for a given *n* volume of the selection, we identify the value λ_α (Popescu, 1993).

Starting from the observations regarding the annual electric energy consumption on an hourly basis, grouped on intervals of variation and expressed by means of the relative frequencies and ascending relative cumulative frequencies, we checked the normality hypothesis of the repartition of the observed values.

The concordance hypothesis was created with the following formula:

$$H_0:\ F(x) = F_0(x, m, \sigma^2),$$

where F_0 is the normal repartition function of parameters m and σ^2 , which are unknown, but estimated by:

- the selection mean $\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$,

respectively

- the dispersion of selection $s^2 = \frac{\sum_{k=1}^{n}(X_k - \bar{X})^2}{n-1}$.

We calculated the differences $F_0(x) - F_n^*(x)$ in Table 3 where: X successively takes the values of the right bounds of the intervals of variation.

Table 3. Calculation of the differences $F_0(x) - F_n^*(x)$

Intervals of variation of the electric energy consumption	Interval right limit (x)	Number hours frequency (n_k)	Relative frequencies $\left(\frac{n_k}{n}\right)$	Ascending cumulative relative frequency (F_n^*)	Reduced standardised values $\left(\hat{z} = \frac{x - \bar{x}}{s}\right)$	Laplace values $\Phi(\hat{z})$	Reduced normal repartition function $F_0(x) = \frac{1}{2} + \Phi(\hat{z})$	$F_0(x) - F_n^*(x)$
1	2	3	4	5	6	7	8	9
887.02 – 1067.02	1067.02	6	0.25	0.25	-1.05	-0.35314	0.14686	-0.10314
1067.02 – 1247.02	1247.02	2	0.08	0.33	-0.58	-0.21904	0.28096	-0.04904
1247.02 – 1427.02	1427.02	2	0.08	0.42	-0.12	-0.04776	0.45224	0.03224
1427.02 – 1607.02	1607.02	1	0.04	0.46	0.35	0.13683	0.63683	0.17683
1607.02 – 1787.02	1787.02	6	0.25	0.71	0.82	0.29389	0.79389	0.08389
1787.02 – 1967.02	1967.02	7	0.29	1	1.28	0.39973	0.89973	-0.10027
Total		24						

As can be seen in Table 3, in column 4 we calculated the relative frequencies corresponding to each interval, and in column 5 the cumulative relative frequencies, i.e. the values of the repartition function of the selection $F_n^*(x)$. For calculating the values of the theoretical repartition function $F_0(x)$ in column 8, we calculated the reduced standardised selection values (column 6) and the corresponding values of the Laplace function (column 7).

To test the H_0 concordance hypothesis, in column 9 we calculated the differences $F_0(x) - F_n^*(x)$ from which we obtained $max|F_0(x) - F_n^*(x)| = 0.17683$.

Considering the significance threshold α = 0.005, we correspondingly found $\lambda_\alpha = 1.358$, resulting that $\frac{\lambda_\alpha}{\sqrt{n}} = 0.2772$.

Since $max|F_0(x) - F_n^*(x)| = 0.17683 < 0.2772$, then the repartition normality hypothesis in Table 3 is accepted.

Therefore, we can assume that the evolution of the annual electric energy consumption has a normal repartition, with the parameters m = 1471.942625 and σ = 385.7714135. This allowed us to use the theoretical normal repartition constructed beforehand, in order to evaluate the probability of the electric energy consumption, for any real value of it between the minimum and maximum limits of the possible field of variation.

The adjustment of the observation data based on this repartition has led to the results in Table 4 and the histogram in Figure 5.

Table 4. Adjusted values

Hour	Hourly annual consumption *x*	Hourly annual standardised consumption *(x-m)/σ*	Normal standardised distribution of the consumption *N(0,1)*
1	981.487	-1.27136332	0.101799713
2	942.031	-1.373641505	0.084776502
3	920.86	-1.428521155	0.076570955
4	899.077	-1.484987236	0.068773603
5	887.02	-1.516241496	0.064729149
6	955.821	-1.337894948	0.090465342
7	1286.612	-0.480415652	0.315465934
8	1418.438	-0.138695152	0.444845524
9	1671.148	0.516381899	0.697206147
10	1741.403	0.698497518	0.757566945
11	1779.658	0.797662461	0.787466803
12	1830.843	0.930344661	0.82390367
13	1859.225	1.003916728	0.842290623
14	1876.665	1.049124847	0.852939669
15	1883.546	1.066961834	0.857005465
16	1863.549	1.015125438	0.844976981
17	1837.069	0.946483752	0.828049047
18	1818.669	0.898787113	0.815616967
19	1764.055	0.757216229	0.775539836
20	1711.728	0.621573726	0.732888899
21	1636.821	0.427399152	0.665455688
22	1475.068	0.008101624	0.503232045
23	1218.669	-0.656538085	0.255738986
24	1067.161	-1.049278435	0.147024994

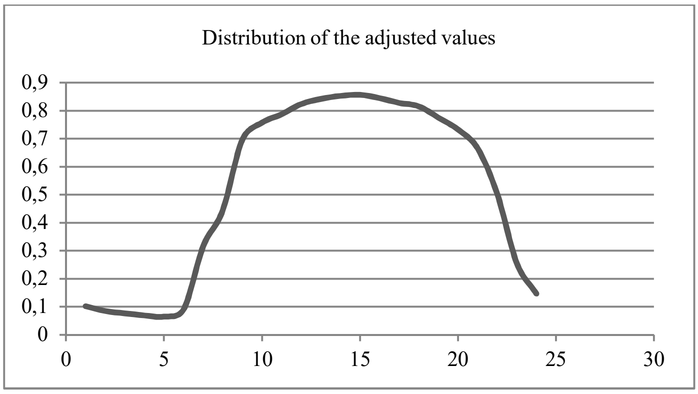

Figure 5. Distribution of the adjusted values

4. Data analysis by grouping on intervals of variation

The results of the analysis and the grouping of data on intervals of variation are presented in Table 5 and were based on the estimation of the parameters (mean and dispersion) of the theoretical normal repartitions that approximate the selection repartitions. Within the intervals of variation ($I_1..I_6$) obtained by the data analysis in Table 2, we calculated the selection mean and the dispersion for each set of selection data.

Table 5. Intervals of variation of energy consumption I_1 = [887,02-1067,02]

Interval	hour	X selection mean	stdev
I_1 = [887.02 – 1067.02]	1	81.79058333	8.806464
	2	78.50258333	8.862347
	3	76.73833333	7.876533
	4	74.92308333	7.823663
	5	73.91833333	7.611099
	6	79.65175	7.912416
I_2 = [1067.02 – 1247.02]	23	101.5558	17.4152642
	24	88.93008	12.5056917
I_3 = [1247.02 – 1427.02]	7	107.2176667	9.460543232
	8	118.2031667	11.19081416
I_4 = [1427.02 – 1607.02]	22	122.9223	18.3328563
I_5 = [1607.02 – 1787.02]	9	139.2623	15.09059277
	10	145.1169	17.40209719
	11	148.3048	19.02851054

	19	147.0046	23.7693493
	20	142.644	23.10819802
	21	136.4018	21.54638284
I_6 = [1787.02 – 1967.02]	12	152.5703	21.15601997
	13	154.9354	22.50710768
	14	156.3888	24.16318601
	15	156.9622	24.87144687
	16	155.2958	25.09644122
	17	153.0891	25.07614269
	18	151.5558	25.21181051

As a result of researching various methods for approximation of data repartition, we identified that the adjustment of primary data by estimated normal repartition provides the ideal model applied to the hourly electric energy consumption for the January-December time series, as it can be seen in Table 6 for the interval of variation I_1 = [887.02–1067.02], Table 7 for intervals I_2 = [1067.02–1247.02], I_3 = [1247.02–1427.02], and I_4 = [1427.02–1607. 02]; Table 8 for I_5 = [1607.02–1787.02] and Table 9 for the interval of variation I_6 = [1787.02–1967.02].

The histogram of data grouped over the 6 determined intervals is shown in Figure 6.

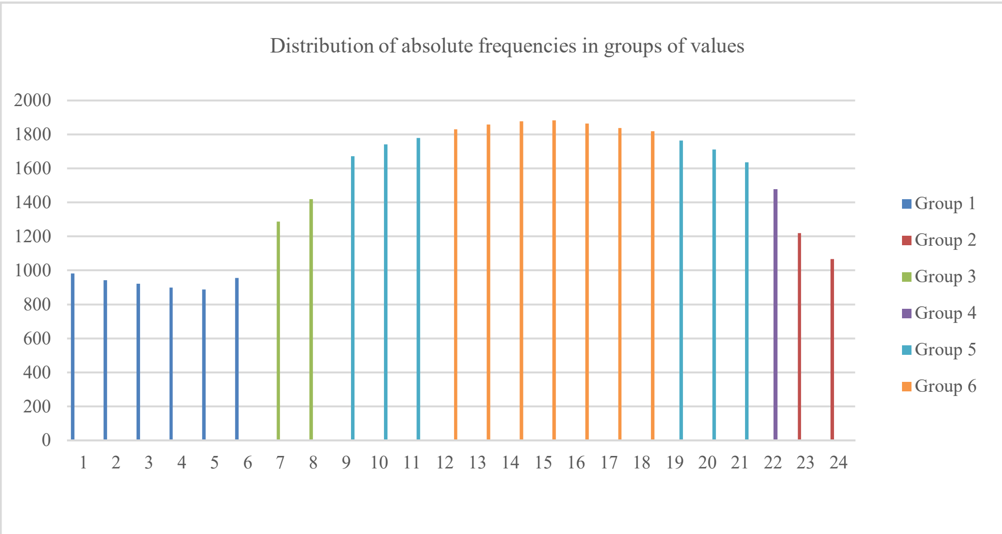

Figure 6. Distribution of absolute frequencies in groups of values

Table 6. Intervals of variation of the electric energy consumption I_1= [887.02-1067.02]

I_1= [887.02 – 1067.02]						
	Hour					
Month	1	2	3	4	5	6
Jan	0.355	0.394	0.437	0.491	0.438	0.412
Feb	0.138	0.155	0.142	0.139	0.122	0.131
March	0.161	0.174	0.187	0.129	0.171	0.159
April	0.184	0.168	0.132	0.135	0.117	0.124
May	0.265	0.274	0.281	0.266	0.264	0.255
June	0.908	0.897	0.885	0.862	0.88	0.849
July	0.968	0.968	0.966	0.962	0.955	0.949
Aug	0.926	0.935	0.934	0.934	0.941	0.948
Sept	0.549	0.542	0.528	0.502	0.518	0.544
Oct	0.428	0.435	0.5	0.6	0.53	0.567
Nov	0.394	0.378	0.431	0.422	0.481	0.595
Dec	0.375	0.313	0.27	0.318	0.337	0.255

Table 7. Intervals of variation of the electric energy consumption I_2, I_3 and I_4.

	I_2 = [1067.02 – 1247.02]		I_3 = [1247.02 – 1427.02]		I_4 = [1427.02 – 1607.02]
	Hour		Hour		Hour
Month	23	24	7	8	22
Jan	0.301	0.304	0.343	0.388	0.347835
Feb	0.136	0.106	0.104	0.112	0.226021
March	0.175	0.122	0.196	0.175	0.2908
April	0.203	0.21	0.146	0.129	0.287687
May	0.259	0.244	0.228	0.214	0.358772
June	0.887	0.874	0.827	0.831	0.18391
July	0.962	0.953	0.953	0.959	0.072961
Aug	0.941	0.909	0.926	0.931	0.115001
Sept	0.643	0.577	0.579	0.478	0.393751
Oct	0.36	0.405	0.606	0.59	0.383884
Nov	0.325	0.341	0.718	0.686	0.392983
Dec	0.47	0.765	0.216	0.316	0.308018

Table 8. Intervals of variation of the electric energy consumption I_5

I_5 = [1607.02 – 1787.02]						
	Hour					
Month	9	10	11	19	20	21
Jan	0.343	0.338	0.327	0.327	0.327	0.31

Feb	0.119	0.126	0.126	0.158	0.151	0.14
March	0.222	0.238	0.231	0.214	0.222	0.22
April	0.157	0.174	0.179	0.192	0.185	0.207
May	0.283	0.263	0.267	0.271	0.279	0.301
June	0.856	0.849	0.854	0.873	0.871	0.875
July	0.973	0.977	0.979	0.971	0.971	0.971
Aug	0.94	0.938	0.929	0.944	0.942	0.942
Sept	0.446	0.473	0.521	0.598	0.596	0.604
Oct	0.491	0.445	0.435	0.365	0.392	0.394
Nov	0.526	0.505	0.469	0.432	0.435	0.417
Dec	0.304	0.299	0.307	0.264	0.246	0.238

Table 9. Intervals of variation of the electric energy consumption I_6

I_6 = [1787.02 – 1967.02]							
	Hour						
Month	12	13	14	15	16	17	18
Jan	0.301	0.3	0.294	0.3	0.299	0.302	0.309
Feb	0.132	0.142	0.146	0.145	0.15	0.142	0.143
March	0.227	0.225	0.222	0.212	0.213	0.209	0.202
April	0.199	0.206	0.206	0.208	0.212	0.215	0.208
May	0.286	0.289	0.294	0.308	0.309	0.311	0.295
June	0.874	0.874	0.876	0.879	0.882	0.876	0.872
July	0.976	0.976	0.974	0.972	0.971	0.97	0.971
Aug	0.931	0.934	0.935	0.937	0.94	0.942	0.942
Sept	0.558	0.577	0.607	0.618	0.624	0.622	0.612
Oct	0.408	0.383	0.383	0.384	0.364	0.363	0.369
Nov	0.45	0.424	0.415	0.405	0.383	0.389	0.419
Dec	0.274	0.271	0.259	0.25	0.259	0.269	0.275

The normal distribution (Kosareva & Krylovas, 2011) of the hourly electric energy consumption values is confirmed by the graph of the repartition of the adjusted values for Figure 7.

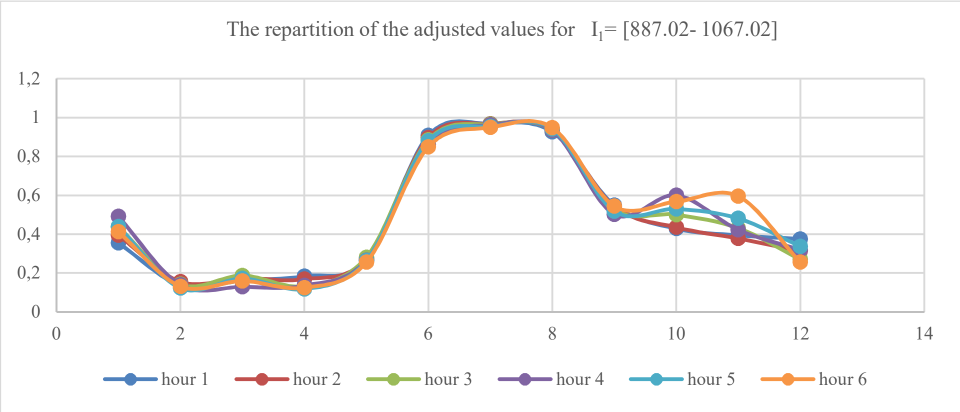

Figure 7. a – the repartition of the adjusted values for I1.

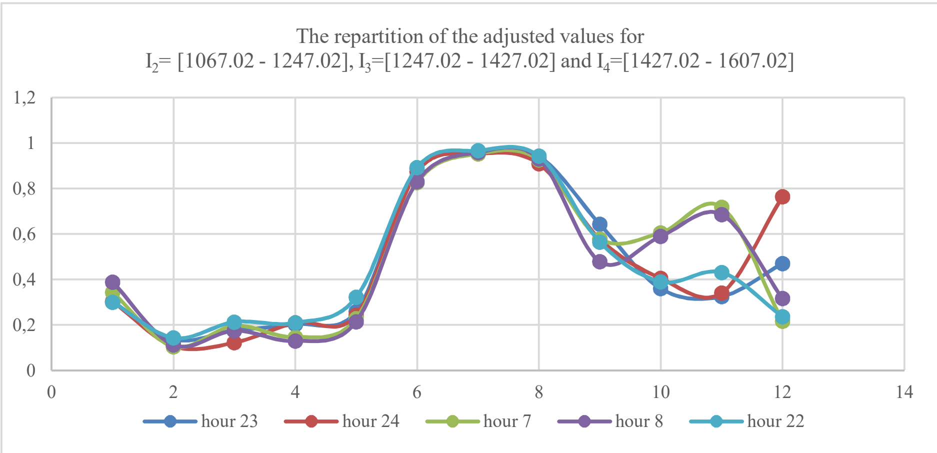

Figure 7. b – the repartition of the adjusted values for I_2, I_3, I_4.

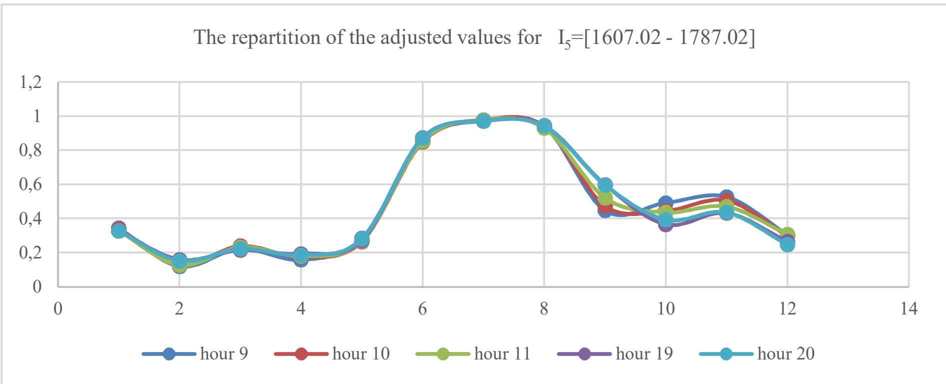

Figure 7. c – the repartition of the adjusted values for I_5.

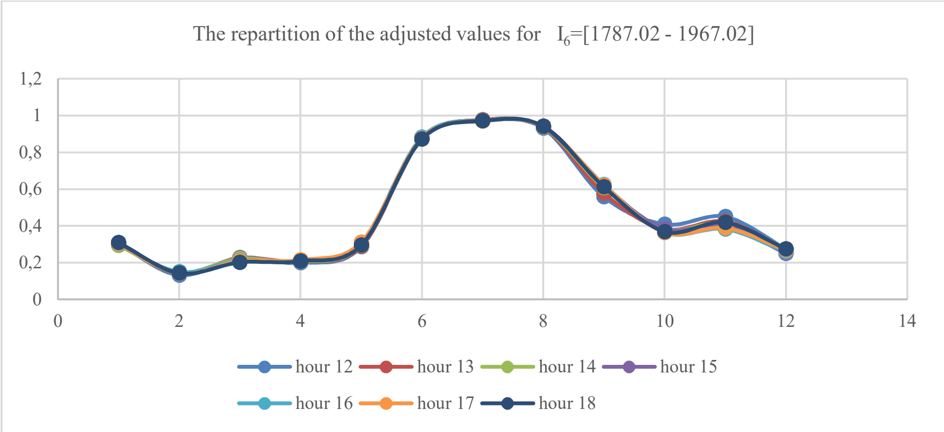

Figure 7. d – the repartition of the adjusted values for I_6.

Figures 7 a-d show that the processing of initial data led to obtaining adjusted values whose distribution corresponds to the normal distribution. This demonstrates the possibility of forecasting the electricity consumption based on the estimated normal distribution.

5. Computer model for analysis

In order to make this study more efficient, we propose the construction of a software system that corresponds to the mathematical support presented in section 1. The functional diagram of the proposed system is presented in Figure 8.

For the modelling and validation module, the analysis and processing techniques are specific to each methodology. The principle of their application is common, and it seeks to specify the form of the theoretical repartition function both in the case when its parameters are known, but also when the parameters are estimated based on the research data.

The decisional situation is characterized by the degree of certainty of the consequences of each formulated alternative. For the choice of decisions, the Electre method was used in situations where there are several possible variants Vi (i=1,m) to reach a goal, the evaluation is based on Cj (j=1,n) criteria, based on which the possible variants are compared two by two.

Various software for data analysis exist, but the data included in the present study required specific processing, which led to the need to develop our own software application for implementing the mathematical model used in the analysis.

The software application has been developed using a modular approach. Therefore, the “Data collection” module offers the possibility to collect, store, process, and archive data in a database. The "Modelling" module provides functionalities to model data, obtain decisions based on the modeled data, and achieve forecasts of the electricity consumption for non-residential consumers. The "Statistical indicators" module provides the possibility to compute the statistical indicators, to build and process grouping intervals. The "Mathematics of data" module implements the statistical tests and methods for verifying and validating the statistical repartitions, used for approximating the repartitions of the experimental data. It offers the possibility to use mathematical techniques in order to model data, to test them based on the Kolmogorov test, and to build assignment functions.

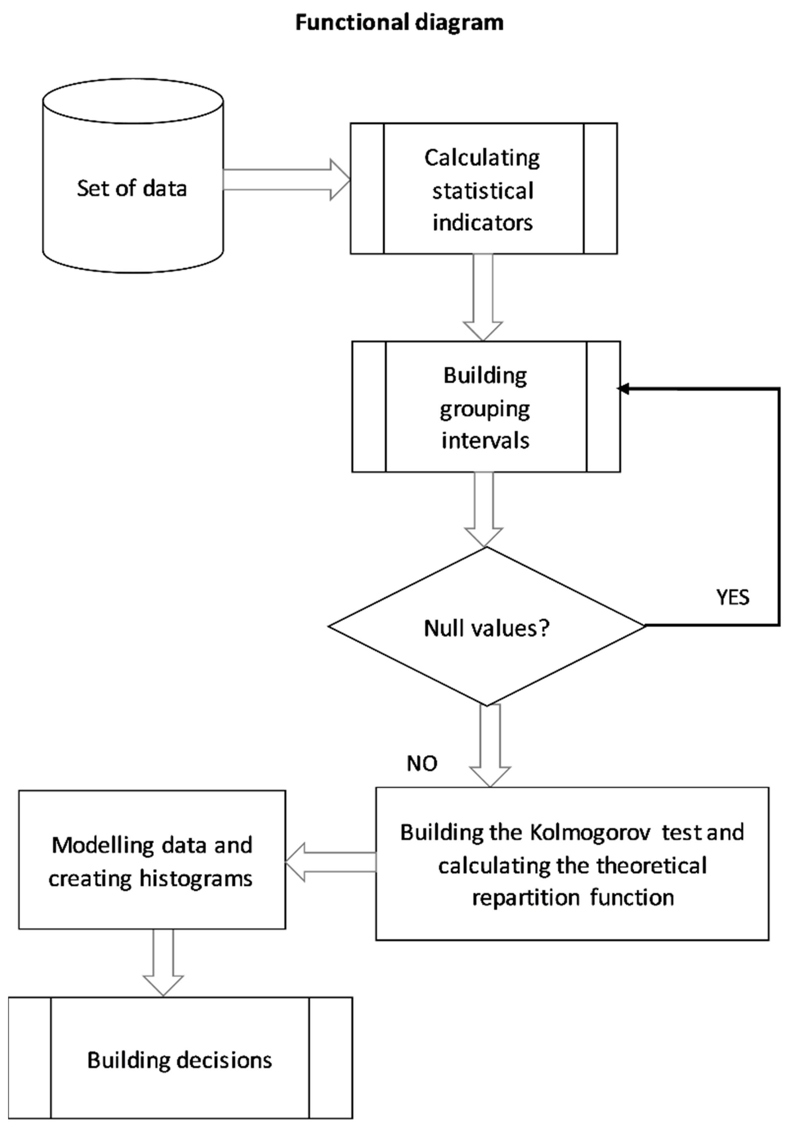

Figure 8. Functional diagram

6. Conclusions

Assuming that the analyzed phenomenon keeps its trend of evolution, the estimated normal repartition can be used to forecast the electric energy consumption. The data sampling allowed a detailed analysis that reflects as accurately as possible the actual process studied for the analyzed consumer data. As a result of the research of various methods for approximating the data repartition, we have identified that the adjustment of the primary data with the estimated normal repartition provides the ideal model for hourly electric energy consumption for the January-December time series. Using the normal theoretical repartition obtained, we can assess the likelihood that the electric energy consumption varies continuously in the analyzed intervals. Furthermore, in order to model the data, it is necessary to use a dedicated computer system that contains specific analysis functions, which continuously adapt for new input data as well.

Acknowledgements

This work was funded by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS (National Research Council) / CCCDI (Advisory Council for Research, Development and Innovation) – UEFISCDI (Executive Agency for Higher Education, Research, Development and Innovation Funding), project number PN-III-P2-2.1-BG-2016-0286 “Informatics solutions for electricity consumption analysis and optimization in smart grids” and contract no. 77BG/2016, within the National Plan for Research, Development and Innovation for the period 2015-2020 (PNCDI III).

References

[1] Ahmad, N., Raheem Ahmed, R., Vveinhardt, J., & Streimikiene, D. (2016). Empirical analysis of stock returns and volatility: evidence from Asian stock markets. Technological and Economic Development of Economy, 22(6), 808–829. https://doi.org/10.3846/20294913.2016.1213204

[2] Daut, M. A. M., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., & Hussin, F. (2017). Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. Renewable and Sustainable Energy Reviews, 70, 1108-1118.

[3] Droutsa, K. G., Balaras, C. A., Dascalaki, E. G., Kontoyiannidis, S., & Argiriou, A. A. (2018). Energy Use Intensities for Asset Rating of Hellenic Non-Residential Buildings. Global Journal of Energy Technology Research Updates, 5, 19-36. https://www.researchgate.net/publication/327667818_Energy_Use_Intensities_for_Asset_Rating_of_Hellenic_Non-Residential_Buildings

[4] Freedman, D., & Diaconis, P. (1981), On the histogram as a density estimator: L2 theory, Probability Theory and Related Fields, 57(4), 453–476

[5] Fumo, N., & Biswas, M. R. (2015). Regression analysis for prediction of residential energy consumption. Renewable and Sustainable Energy Reviews, 47, 332-343. https://doi.org/10.1016/j.rser.2015.03.035

[6] Gutiérrez-Pedrero, M. J., Tarancón, M. Á., del Río, P., & Alcántara, V. (2018). Analysing the drivers of the intensity of electricity consumption of non-residential sectors in Europe. Applied Energy, 211, 743-754. https://doi.org/10.1016/j.apenergy.2017.10.115

[7] Ma, H., Du, N., Yu, S., Lu, W., Zhang, Z., Deng, N., & Li, C. (2017). Analysis of typical public building energy consumption in northern China. Energy and Buildings, (136), 139-150. https://doi.org/10.1016/j.enbuild.2016.11.037

[8] Miller, C., Nagy, Z., & Schlueter, A. (2018). A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings. Renewable and Sustainable Energy Reviews, 81(P1), 1365-1377. https://doi.org/10.1016/j.rser.2017.05.124

[9] Kosareva, N., & Krylovas. A. (2011). A numerical experiment on mathematical model of forecasting the results of knowledge testing. Technological and Economic Development of Economy, 17(1), 42–61

[10] Oprea S.V., Pîrjan A., Căruțașu G., Petroșanu D.M., Bâra A., Stănică J.L., Coculescu C. (2018), Developing a Mixed Neural Network Approach to Forecast the Residential Electricity Consumption Based on Sensor Recorded Data, MDPI Sensors Journal, Volume 18, Issue 5: 1443, May 2018, 1424-8220, https://doi.org/10.3390/s18051443

[11] Pérez-Chacón, R., Talavera-Llames, R. L., Martinez-Alvarez, F., & Troncoso, A. (2016). Finding electric energy consumption patterns in big time series data. In Distributed Computing and Artificial Intelligence, 13th International Conference (Vol. 474, p. 231-238). Springer, Cham. https://pdfs.semanticscholar.org/86d4/7d46c13ed334c5d51b4797caf412b8b75764.pdf

[12] Popescu, O. (coord.) (1993). Applications of Mathematics in Economics. Vol. I, II. Didactic and Pedagogical Publishing House, Bucharest (in Romanian).

[13] Popovici, A. (2015). Probabilities, statistics and econometrics, assisted by Excel software, Niculescu Publishing House, Bucharest (in Romanian).

[14] Purcaru, I. (1997). Matematici generale & elemente de optimizare. Teorie și aplicații, Economica Publishing House, Bucharest, pp. 622-635, Bucharest (in Romanian).

[15] Ruparathna, R., Hewage, K., & Sadiq, R. (2016). Improving the energy efficiency of the existing building stock: A critical review of commercial and institutional buildings. Renewable and sustainable energy reviews, 53, 1032-1045. https://doi.org/10.1016/j.rser.2015.09.084

[16] Scott, D.W. (1979). On optimal and data-based histograms. Biometrika, 66(3), 605–610.

[17] Scott, D.W. (2009), Sturges' rule, Wiley Interdiscipl. Rev.: Comput. Statist., 1, pp. 303-306

[18] Sivilevičius, H., Vislavičius, K., & Bražiūnas, J. (2017) Technological and economic design of asphalt mixture composition based on optimization methods. Technological and Economic Development of Economy, 23(4), 627–648. doi:10.3846/13923730.2016.1210223

[19] Stănică J.L., Căruțașu G., Pîrjan A., Coculescu C., (2018), IoT Cloud Solution for Efficient Electricity Consumption, Journal of Information Systems and Operations Management (JISOM), Vol. 12 No. 1 / May 2018, 2018, pp. 45-57, ISSN 1843-4711;

[20] Sturges, H. A. (1926), The choice of a class interval. Journal of the American Statistical Association, 21(153), 65–66

[21] Teodorescu, H.-N. (2015), On the Regularities and Randomness of the Dynamics of Simple and Composed CAs with Applications, Romanian Journal of Information Science and Technology, Romanian Academy, vol. 18, no. 2, pp. 166-181.

[22] Tso, G. K., & Yau, K. K. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy, 32(9), 1761-1768. https://doi.org/10.1016/j.energy.2006.11.010

[23] Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 16(6), 3586-3592. https://doi.org/10.1016/j.rser.2012.02.049

[24] Wand, M. P., Jones, M. C., (1994) Kernel Smoothing (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). Chapman and Hall/CRC