Ontology of experiment planning for obtaining a probabilistic model of single-phase electricity consumers

V A Mokhov, D V Shaikhutdinov, T N Kruglova, R G Oganyan and A N Tkachev
Platov South-Russian State Polytechnic University (NPI), 132, Prosveshcheniya, Novocherkassk, 346428, Russia

E-mail: mokhov_v@mail.ru

Abstract. The paper presents the rationale and the problem of forming a probabilistic model of single-phase electricity consumers. To solve the problem, the corresponding mathematical apparatus is stated and the ontology of experiment planning is developed. The applicability of the prepared ontology was demonstrated on a specific example of an open dataset from the Intelligent Systems Subcommittee (ISS) IEEE. At the same time, the authors identified the characteristic modes of energy consumption for a particular residential building and data sets for direct construction of a probabilistic model of single-phase electricity consumers with total data loss caused by filtering less than 10%.

1. Introduction

Often, when solving practical problems, the effectiveness of computer modeling is determined by the degree of similarity of the characteristics of the model of the investigated object with its real physical characteristics [1-2]. This is especially important in instances when a system of automatic control of a physical object is built based on the results of modeling. To solve the problem of controlling interphase switching of single-phase electricity consumers [3], a model of the consumer behavior is required. For this purpose, this work presents the developed mathematical apparatus and ontology for planning an experiment to obtain a version of such a model with specific examples of calculations.

2. Materials and methods

Previously, in the work of the authors [3], the problem of controlling interphase switching of single-phase electricity consumers was solved, the essence of which is formulated as follows:

- \( n \) non-industrial single-phase consumers (SC) of electricity are connected to the distribution network powered by the distribution transformer substations (TS);
- each of the SC is provided with dynamic switching tools in relation to the phases L1, L2, L3 within the TS;
- the implementation of the solution to the task of switching the SC is performed dynamically at regular intervals \( t \).

The input data for finding a solution is:

- original number combination of SC to TS connection phases:
  \[
  l^0 = \{ l^0_{11}, l^0_{12}, ..., l^0_{1n} \}, \quad t^0_r \in \{1,2,3\}, \quad r = \frac{1}{1,n} \quad (1)
  \]
- a set of digitized load current oscillograms corresponding to each combination of connections (1) for all SCs.
The solution to the problem is to find a new combination of numbers of SC connections to TS phases for the time $t' (t' < t)$ in real time:

$$L^* = \{i_{1}', i_{2}', ..., i_{n}'\}, \quad l_r' \in \{1, 2, 3\}, \quad r = \overline{1, n},$$

in the implementation of which, in comparison with the original combination (1), the smallest values are provided: the number of switching of the SC in phases, the values of the indicators characterizing the asymmetry of currents and their total harmonic distortions with the same set of oscillograms $I(2)$. To find the combination (3), the presented problem is reduced to a combinatorial optimization problem, and the search for its solution is performed using binary algorithms of agent-based metaheuristics. The objective function for the problem is calculated based on (2) and is presented in [3]. Obviously, to perform the calculations in the stated problem, a system for generating sets (2) is required.

The most widespread spectral decomposition of periodic functions is their representation in the form of a truncated segment of the Fourier series [4, 5]:

$$a(t) = a_0 + \sum_{k=1}^{n} a_k \sin k\omega t + \sum_{k=1}^{n} a_{n+k} \cos k\omega t$$

where $x(t)$ is a periodic function with main period $T$; $\omega = 2\pi f; f = \frac{1}{T}; a_0 = \frac{1}{T} \int_0^T x(t)dt; a_k = \frac{2}{T} \int_0^T x(t) \sin k\omega t dt; a_{n+k} = \frac{2}{T} \int_0^T x(t) \cos k\omega t dt; k = \overline{1, n}.$

In accordance with (4), the elements of the set (2) can be uniquely determined by the set of values of the Fourier coefficients $\tilde{a} = (a_0, a_1, ... a_n, a_{n+1}, ... a_{2n})$.

We assume $a_0, a_1, ... a_n, a_{n+1}, ... a_{2n}$ to be random variables. In this case, each SC or element in (2) represents a random process that can be described by the corresponding random vector $\tilde{a} = (a_0, a_1, ... a_n, a_{n+1}, ... a_{2n})$.

Then, to build the model (2), it becomes necessary to preliminarily estimate the statistical parameters of the distribution $a_0, a_1, ... a_n, a_{n+1}, ... a_{2n}$. The solution to the assessment problem assumes the presence of $\gamma$ corresponding statistical observations, for which, by means of (4), a statistical sample of the form must be formed:

$$\tilde{a}_1 \leftrightarrow \{a_{01}, a_{11}, ..., a_{n1}, a_{n+11}, ..., a_{2n1}\}$$

$$\vdots$$

$$\tilde{a}_\gamma \leftrightarrow \{a_{0\gamma}, a_{1\gamma}, ..., a_{n\gamma}, a_{n+1\gamma}, ..., a_{2n\gamma}\}. $$

The statistical estimates obtained on the basis of (5) will make it possible to determine the parameters of the distribution functions

$$\eta_k = \mathcal{F}_k(\zeta_k), \quad \zeta_k \in [0; 1]$$

each of $2n$ values as well as to provide the opportunity to generate random vectors

$$\tilde{A} \leftrightarrow \{a_0, a_1, ... a_n, a_{n+1}, ... a_{2n}\},$$

for a given number of $p$ individual consumers in solving the problem of search (3).

As a statistical sample for the formation (5), we will use the information on the amplitude spectrum of the current consumption from the open dataset Private Home 1 [6], provided by the Intelligent Systems Subcommittee (ISS) IEEE at [7].
3. Results

Generally, a means of constructing mathematical models of various processes in order to increase the efficiency of experimental research is the planning of an experiment [8]. It allows to increase the reliability of the research results [9] and to reduce the time and funds for the experiment. The main stages of the planned experiment [10] can be described as follows.

1) The goal of planning an experiment in this case is to obtain parameters for distribution functions (6) at a given level of confidence.

2) With regard to the conditions of the experiment, information is available that the tests were carried out on the basis of installation single-phase 5.75 kVA in the period from 3rd to 18th June 2011 for a single family housing, with a sampling period of 5 minutes by means of the analyzer quality of electrical energy Chauvin Arnoux 8335 (France). Single family housing has the following characteristics:

   a) number of people per house: 3 adult people;
   b) number of rooms in the house: 3 bedrooms, 3 bathrooms, 1 living room, 1 kitchen, 1 laundry room, 1 hallway and corridor;
   c) main electrical equipment used: 1 washer / dryer, 1 dishwasher, 1 electric stove, 1 electric oven, 1 refrigerator, 1 microwave, 1 range hood, 1 vacuum cleaner, 1 hair dryer, 3 TVs, 1 LCD display, 2 laptops, 1 router, 2 electric heaters.

3) Input data is time counts with a period of 5 minutes (registered parameter) and corresponding random combinations of equipment active/switched on during this period (non-registered parameter). The output data is the data of the amplitude spectrum of the current consumption.

4) The output data is presented as a set of 4320 measurements for each harmonic, their values are measured in amperes with an accuracy of representation to the fifth decimal place.

5) Obviously, when using the data from the open ISS IEEE sets indicated in the previous paragraph, the experiment can be considered passive, based on the registration of input and output parameters characterizing the object of research, without interfering with the experiment in the course of its conduct. In this case, experimental data is processed after the end of the experiment.

6) Statistical processing of the experimental results to construct a mathematical model of the behavior of the characteristics under study was performed as follows.

   a) In the general dataset, all measurements related to the inactive period of residents were excluded: from 00:00 hours to 09:00 hours. This period of time is characterized by the fact that during its course the elements of electrical equipment with low energy consumption operate mainly in automatic mode, which does not determine the main trend of daily energy consumption. As a result, the amount of data for research was reduced to 2700 measurements for each harmonic.

   b) Based on RMS statistics for current consumption $I_{rms}$, calculated through the effective values of harmonics of various orders $I_h$ according to the formula (8)

$$I_{rms} = \sqrt{\sum_{n=1}^{\infty} I_n^2},$$  \hspace{1cm} (8)

all measurements were ranked.

   c) For the sequence of data obtained in 6.b, subsamples were identified that define the typical energy consumption patterns for a residential building. For this:

   i. using the method of a sliding window of variable length [11, 12], estimates of the significance level were calculated at which the set of measurements within the bounds obeyed one of the known distribution laws;

   ii. for each of the laws involved, possible bound configurations were recorded (numbers of boundary elements and estimates of the level of significance);
iii. from the set of all possible configurations, those were selected that, with their minimum number and a given level of significance $\alpha$ ensured maximum coverage of calculated values $I_{rms}$ with minimal intersection;

iv. for uncovered bounds (with a size of 100 or more elements and an insufficient level of significance), based on binary optimization algorithms, their elements were filtered to obtain a configuration corresponding to one of the known distribution laws with a significance level $\alpha$ while maintaining the maximum possible number of elements.

The set of configurations obtained in this way determined the size, boundaries, data sets and distribution laws that determine the characteristic modes of energy consumption.

d) In order to obtain configurations for the effective values of harmonics (within the limits of the power consumption modes defined in clause 6.c), the measurements were also filtered for compliance with one of the known distribution laws with a given level of significance $\alpha$ using the maximum number of elements (similar to n. 6.c.iv).

The resulting set of configurations determined the numerical composition, boundaries, data sets and distribution laws that determine the characteristic modes of energy consumption at the level of effective harmonic values.

7) An explanation of the results obtained and the formulation of recommendations for their use assumes that, based on the results of the implementation of the stages 6.a÷6.d obtained data sets are associated with statistical samples (5), on the basis of which estimates are calculated to determine the parameters of the distribution functions (6). In this case, the generation of random vectors (7) within each characteristic mode of energy consumption for a residential building with level of significance $\alpha$ becomes trivial.

Graph illustrating the curve of change of 2700 ranked values of the calculated value $I_{rms}$, is shown in the figure 1.

![Figure 1. Ranged Values Curve $I_{rms}$](image)

Figure 2 is a graph showing an initial set of window configurations selected for subsequent filtering and extraction of typical power consumption modes.
When constructing the graph (figure 2), bounds with the following configuration were displayed: window width - not less than 205 elements, confidence level - not less than 0.7. The distribution laws that sub-samples within bounds obey are indicated by colors: magenta for Rayleigh, blue for Uniform, red for Lognormal, black for Weibull, green for Exponential.

Subsequent $I_{rms}$ data processing based on the graph (figure 2) made it possible to identify 6 typical power consumption modes, the configurations of which are presented in table 1.

Table 1. Configurations of typical power consumption modes of the object under study.

| Mode | Boundary indices of elements | Number of elements, units | Range of values $I_{rms}$, A | Share in the total range of values $I_{rms}$, % | Law of distribution | Confidence probability $\alpha$ | Filtering loss, units | Filtering loss, % |
|------|-----------------------------|---------------------------|-----------------------------|-----------------------------------------------|---------------------|-------------------------------|---------------------|------------------|
| 1    | 1-190                       | 190                       | 0.323-0.3788                | 0.36%                                         | Weibull             | 0.91343                       | 65                  | 34.21%           |
| 2    | 191-449                     | 259                       | 0.3791-0.4030              | 0.15%                                         | Weibull             | 0.90267                       | 10                  | 3.86%            |
| 3    | 450-610                     | 161                       | 0.4032-0.4218              | 0.12%                                         | Lognormal           | 0.90312                       | 5                   | 3.11%            |
| 4    | 611-1210                    | 600                       | 0.4224-0.7009              | 1.79%                                         | Uniform             | 0.90252                       | 77                  | 12.83%           |
| 5    | 1211-2560                   | 1350                      | 0.7012-3.7011             | 19.27%                                        | Rayleigh            | 0.90722                       | 79                  | 5.85%            |
| 6    | 2561-2700                   | 140                       | 3.7064-15.8924             | 78.27%                                        | Exponential         | 0.90157                       | 3                   | 2.14%            |
| Overall | 1-2700                    | 2700                      | 0.323-15.8924             | 100%                                          | --                 | --                            | 239                 | 8.85%            |

The first mode corresponds to the first (from left to right in figure 2) uncovered bound, which, according to the filtering results, best matches the Weibull distribution law. The second mode is also defined on the basis of the Weibull distribution law (in figure 2 - the first covered, black bound from left to right. The third mode is determined on the basis of the Lognormal distribution law (in figure 2 - the first red box on the left). The fourth mode is determined on the basis of the Uniform distribution law (figure 2 shows a frame between the red and magenta bounds, partially covered with many blue bounds). The fifth mode is defined on the basis of the Rayleigh distribution law (in figure 2, the larger magenta frame). The sixth mode is determined on the basis of the Exponential distribution law (in figure 2 - the rightmost one, starting from the left border of the magenta frame and partially covered with a green frame). The complete final configurations of power consumption modes are presented in the corresponding rows of table 1.
Total data filtering loss for a collection of 2700 values $I_{\text{rms}}$ was less than 9%. In this case, the highest percentage of relative losses was obtained for the first mode, the share of value range of which in the total value range is only 0.36%.

The set of configurations for the effective values of harmonics (within the bounds determined by the configurations of the modes from Table 1) was obtained in a similar way by filtering the corresponding data sets based on modifications of the binary algorithms of agent-based metaheuristics [13-15] with total data losses within 10%.

4. Discussion
The applicability of the described experiment planning ontology for obtaining a probabilistic model of single-phase electricity consumers is demonstrated with the example of one of the open ISS IEEE sets. Its applicability in relation to other datasets, datasets of a harmonic nature from other subject areas, as well as the study of the possibility of constructing appropriate probabilistic models with high specified values of significance levels is the subject of further research.

5. Conclusion
The obvious results of the work presented by the authors are:

1) ontology of experiment planning to obtain a probabilistic model of single-phase electricity consumers;
2) a demonstration of its application in the example of one of the open ISS IEEE sets;
3) a list of typical energy consumption modes for a particular residential building, obtained on the basis of processing statistics of its energy consumption;
4) datasets for direct construction of a probabilistic model of single-phase electricity consumers, obtained on the basis of filtering measurement results by means of modifications of binary algorithms of agent-based metaheuristics with total losses less than 10% of the total volume of measurements.

Acknowledgments
The reported study was funded by RFBR according to the research project № 18-38-20188 «Development and research of tools for intellectual modeling and synthesis of electric power net-works technological processes trajectories» using equipment of shared facility «Diagnosis and energy-effi cient electrical equipment» (NPI).

References
[1] Weisberg M 2012 *Simulation and Similarity: Using Models to Understand the World* (Oxford: Oxford University Press) 288
[2] Diedrichsen J and Kriegeskorte N 2017 Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS computational biology* 13(4) e1005508
[3] Mokhov V, Tkachev A and Shaykhutdinov D 2019 Optimal Control of Phase-to-Phase Switching of Single-Phase Power Consumers Based on Binary Algorithms of Agent Metaheuristics. 2019 *International Russian Automation Conference (RusAutoCon)* Sochi: IEEE 1-6
[4] Elbi M D and Kizilkaya A 2020 Multicomponent signal analysis: Interwoven Fourier decomposition method. *Digital Signal Processing* 104 102771
[5] Singh P, Joshi S D, Patney R K and Saha K 2017. The Fourier Decomposition Method for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 473(2199) 20160871
[6] Canizes B et al 2015 Resource Scheduling in Residential Microgrids Considering Energy Selling to External Players. *Clemson University Power Systems Conference 2015 (PSC 2015 Clemson: USA)* 10-13
[7] *Open Data Sets* Retrieved from: http://www.gecad.isep.ipp.pt/ies2/Consumption/PrivateHome_1/2011_June_03_to_17.xls
[8] Wu C J and Hamada M S 2011 Experiments: planning, analysis, and optimization (New York: John Wiley & Sons) 552
[9] Baird D C 1962 Experimentation: an introduction to measurement theory and experiment design (USA: Prentice Hall) 212
[10] Antony J 2014 Design of experiments for engineers and scientists (Elsevier) 672
[11] Sedyakin I, Ulyanov M V 2018 Time-efficient algorithm for the generalized entropy calculating of two-dimensional words by the method sliding window. Modern information technologies and IT education 14(3) 560-566
[12] Du L, Song Q and Jia X 2014 Detecting concept drift: an information entropy based method using an adaptive sliding window. Intelligent Data Analysis 18(3) 337-364
[13] Mirjalili S, Mirjalili S M and Yang X S 2014 Binary bat algorithm. Neural Computing and Applications 25(3-4) 663-681
[14] Mirjalili S and Lewis A 2013 S-shaped versus V-shaped transfer functions for binary particle swarm optimization. Swarm and Evolutionary Computation 9 1-14
[15] Mafarja M, Aljarah I, Heidari A A, Faris H, Fournier-Viger P, Li X and Mirjalili S 2018 Binary dragonfly optimization for feature selection using time-varying transfer functions. Knowledge-Based Systems 161 185-204