

PAPER • OPEN ACCESS

Application of the information uncertainty measure when comparing planned and actual commercial losses of electricity

To cite this article: N V Dulesova *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **862** 062019

View the [article online](#) for updates and enhancements.

Application of the information uncertainty measure when comparing planned and actual commercial losses of electricity

N V Dulesova¹, A S Dulesov², D J Karandeev² and A V Malykhina²

¹ Khakas Technical Institute of Siberian Federal University, 15, Komarova ave., Abakan, 655017, Russia

² Katanov Khakass State University, 92, Lenina ave., Abakan, 655017, Russia

E-mail: den_dr_house_1991@mail.ru

Abstract. The paper considers methods for processing data on commercial losses in electric networks with subsequent analysis of the obtained results. The information processing tools included methods for determining the amount of divergence of electric power losses when comparing planned and actual data. Comparing the planned and actual values of electric power losses, a method is proposed that in the classical theory of information is called "Kullback-Leibler divergence". The rationale for its use is based on the possibility of applying a measure of information uncertainty, where information entropy is taken as a measured value. Comparing the planned and actual values of electric power losses, discrepancies between these distributions are obtained based on the application of the Kullback-Leibler model. The obtained results not only confirmed the importance of the applicability of this method of information processing, but also allowed us to draw attention to the adequacy of the planned losses to the actual ones.

1. Introduction

Commercial losses of electricity – the most important indicator of the efficiency of electric networks, a visual indicator of the system state of accounting for electricity consumption and the efficiency of power grid organizations [1].

Commercial losses cannot be calculated and measured by instruments, they are determined as the difference between the actual and planned indicators of electric power losses [2, 3]. The costs associated with the payment are not compensated by the tariff regulation, but according to the current legislation, network organizations are obliged to pay for the actual loss of electricity that occurred in their own network facilities. Network organizations are more interested in the most accurate accounting of electricity, since commercial losses are a direct financial loss of network companies.

Commercial losses are planned for regulatory reduction, but it isn't possible to eliminate them in full. Such losses are classified as excess [4]. The reason for these losses is the presence of objective processes associated with the presence of gaps between solvency of consumers and electricity tariffs, insufficient investment in the network infrastructure and the electricity metering system, the lack of full-scale automated information systems for collecting and transmitting data on the useful output of electricity, etc. [5].

Planning involves the use of regulatory documents to determine the effectiveness of measures to reduce commercial electric power losses. However, not all activities are subject to accurate calculations, which significantly reduces the level of reliable planning of the effectiveness of the entire range of



activities. Consequently, the effectiveness of some activities has to be determined empirically or by comparing planned indicators with actual ones.

2. Uncertainty of information in the task of comparing planned and actual indicators

As in the construction of the plan, and when taking into account the power consumption, you have to deal with a lot of factors of a random nature, so we will talk further about the information uncertainty. The parameter of its estimation in the organizational and economic system is entropy, considered as a measure of uncertainty. It can be used to measure the state of the system by comparing it with some "ideal conditions" where knowledge of the situation is completely deterministic.

Since we have to deal with uncertainty [6,7], it is necessary to determine the amount of entropy, without excluding the probabilistic estimates of the state of the system, which are widely used based on information theory [8,9]. In most cases, p_i is the relative number of discrete states of the system, that is, the parameter values related to each of the i -th time intervals [10]. According to the information theory, discrete states can be considered as signals that have a specific value and are distributed according to their frequency of occurrence [11]. Such definition of probability spaces allows us to use the models of Ralph Vinton Lyon Hartley and Claude Elwood Shannon to determine the entropy of the state of the system [12, 14].

The Shannon's formula allows us to determine the entropy for both the natural and statistical probability distribution of the system state, characterized by the amount of commercial losses [15]. The natural distribution is essentially "desirable", that is, related to the planned indicators, the achievement of which should be ensured by the implementation of a number of organizational measures. As for the statistical distribution, it can be built on the basis of recording actual indicators over a certain time interval. The natural and statistical distributions have a probabilistic nature of the occurrence of events, that is, the dynamics of the parameters. Therefore, it is impossible to guarantee the fact that the obtained statistical distribution completely coincides with the natural distribution of parameters due to the impact on the system of hard-to-predict factors of probabilistic nature. In this case, it is important to have information about the discrepancy between the comparable distributions in order to assess the similarity of the quantitative characteristics of the planned indicators with the actual ones.

When comparing planned and actual indicators, first of all, they are based on the determination of the absolute difference and error. Such estimates are quite reasonable, but don't allow us to draw a specific conclusion about the connection between two patterns of distribution of the same parameter. The solution of such a problem is possible through the application of a measure of information uncertainty, that is, entropy [16, 17].

Entropy characterizes the occurrence of events with a quantitative characteristic of the parameter of commercial losses- kWh. This quantitative characteristic $w_i(t)$ is the electricity, fixed in the time interval t (hour, day, month, etc.) and related to the interval $i \in I$. The whole set of intervals I completely covers the range of the amount of electricity from w_{min} to w_{max} . The selected interval i includes two values of the considered parameter: the actual indicator, considered as the mathematical expectation of a random variable, and the planned indicator, as the expected value. It is necessary that the interval be the same for both distributions.

Considering the actual distribution of the amount of electricity in each of the intervals, we can build a histogram of the distribution of this statistical series. Based on the obtained frequencies λ_i of the histogram the probability of occurrence of events for each interval is determined:

$$p_i(\lambda) = \frac{\lambda_i}{\Lambda}, \text{ by } \sum p_i(\lambda) = 1, \quad (1)$$

where Λ is the sum of frequencies over all intervals.

Considering the planned indicators of distribution of the electric power over the intervals, by analogy with (1) the probabilities $q_i(\lambda)$ are determined, without excluding observance of the condition: $\sum q_i(\lambda) = 1$. The probability data distribution is a priori postulated, and the actual distribution $p_i(\lambda)$ –

statistical, obtained from the processing of operational data, is verifiable. Comparison of these two distributions is possible through the measure of information uncertainty, the value of which is determined by the Shannon's formula:

$$H = -\sum_{i=1}^n p_i \log_2 p_i \quad \sum_{i=1}^n p_i = 1, \quad \text{bit, by} \quad (2)$$

where n is the number of intervals.

Formula (2) is valid for determining the entropy of various distributions. The Shannon's entropy will be the greater the smaller the distribution density p_i ($p_i \rightarrow 1, H \rightarrow 0$) takes on. However, the Shannon's formula doesn't take into account the situation when there are no indicators in the considered time interval, i.e. $\lambda_i = 0$. In this case, $p_i = 0$, and the entropy is $H \rightarrow \infty$. If the system was running in a given time interval and the data were not received for processing, then λ_i can be determined by averaging: $\lambda_i = (\lambda_{i-\Delta} + \lambda_{i+\Delta})/2$, where $\lambda_{i-\Delta}$ and $\lambda_{i+\Delta}$ are the frequency of occurrence of events to the left and right of the interval t .

When comparing two distributions, it is impossible to say unequivocally that in the future the probabilities of occurrence of events p_i will coincide with the probabilities of q_i . The statistical distribution $p_i(\lambda)$ obtained from (1) is verifiable and serves as an approximation of the distribution $q_i(\lambda)$. Referring to the possibility of applying the measure of information uncertainty, we can calculate an approximation that reflects the value of the unaccounted amount of information in the transition from the expected distribution of $q_i(\lambda)$ to the statistical $p_i(\lambda)$, due to the Nature and control capabilities of the system. The measure of information considered in this case is called "cross entropy". For the two distributions $q_i(\lambda)$ and $p_i(\lambda)$, the cross entropy is determined as follows:

$$H_q(p) = -\sum_{i=1}^n q_i(\lambda) \log_2 p_i(\lambda), \quad \text{bit.} \quad (3)$$

There is a difference between the entropy (2) and the cross entropy (3) called the Kullback-Leibler divergence [18, 17]. The Kullback-Leibler divergence is the divergence of the distribution of p with respect to q :

$$D_q(p) = H_q(p) - H(q). \quad (4)$$

The value of entropy is determined by the expression (2). When substituting expressions (3) and (2) in (4) by performing simple mathematical transformations, the Kullback-Leibler divergence has the form:

$$D_q(p) = \sum_{i=1}^n q_i(\lambda) \log_2 \frac{q_i(\lambda)}{p_i(\lambda)}, \quad \text{bit.} \quad (5)$$

The Kullback-Leibler divergence is the distance between two distributions. It shows how different the distributions of random variables are. It should be borne in mind that the functional (5) is not a metric in the space of distributions, since it does not satisfy the axiom of symmetry: $D_p(q) \neq D_q(p)$. Further, on a concrete example, we will offer to consider the comparison of the planned and actual indicators of commercial losses.

3. Practical implementation of the distribution matching problem

Based on the statistical data of monthly electricity consumption for 6 years (from 2013 to 2018), we present: analysis of the dynamics of electricity sales with its division into electricity supply to the network and consumers; comparison of the dynamics of actual losses and regulatory losses; distribution comparison results.

Based on the statistical data of monthly electricity consumption for 6 years (from 2013 to 2018), we present: analysis of the dynamics of electricity sales with its division into electricity supply to the network and consumers; comparison of the dynamics of actual losses and regulatory losses; distribution comparison results.

To determine the volume of electric power losses for the period under review, according to [20] the balance of electricity supplied through the electric network to consumers is compiled (Table 1).

Table 1. Technical and economic indicators of the enterprise for the period 2013-2018.

Year	Supply of electric power in the networks, MWh			Electric power losses in the networks, MWh			Electric power losses in the networks, %		
	Planned	Actual	Deviation	Planned	Actual	Deviation	Planned	Actual	Deviation
2013	127760	126044	-1716,5	32767	41607,8	8840,8	25,6	33,01	7,41
2014	181155	197354	16198,6	43755	66846,8	23091,8	24,15	33,87	9,71
2015	179580	189808	10228,2	43640	59034	15394	24,3	31,1	6,8
2016	179628	190539	10910,5	43508	57447,4	13939,4	24,22	30,15	5,93
2017	179585	184107	4522,35	43298	49366,4	6068,4	24,11	27,81	3,7
2018	188662	206604	17942,3	43849	54343,9	10494,9	23,24	26,3	3,06

The results presented in table 1 are displayed graphically (figure 1), where the annual excess of actual electric power losses over the planned ones is clearly visible.

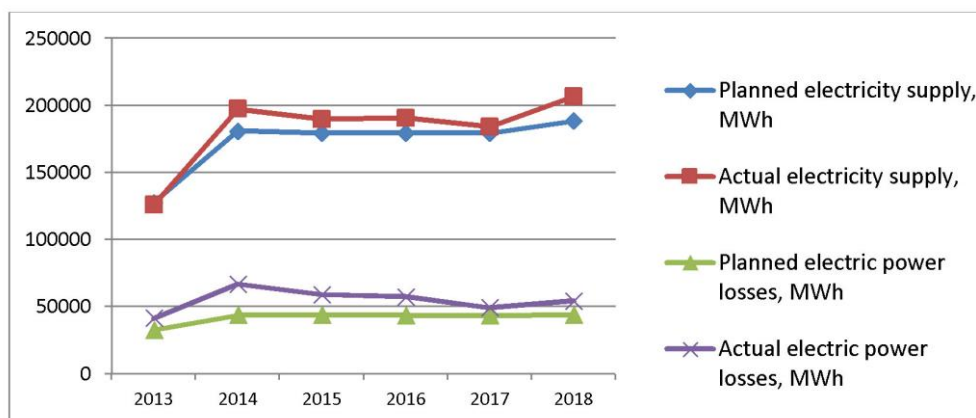


Figure 1. Performance indicators of the electric power company for 2013-2018.

Analysis of the monthly divergence between actual and planned losses indicates a high divergence, especially in the winter period. Significant factors of this divergence include the increase in electricity consumption in the winter period and the theft of electricity mainly by household consumers.

Next, we propose to consider the results of applying the formula:

$$D_q(p) = \sum_{i=1}^n q_i(y) \log_2 \frac{q_i(y)}{p_i(x)}, \text{ bit}, \quad (6)$$

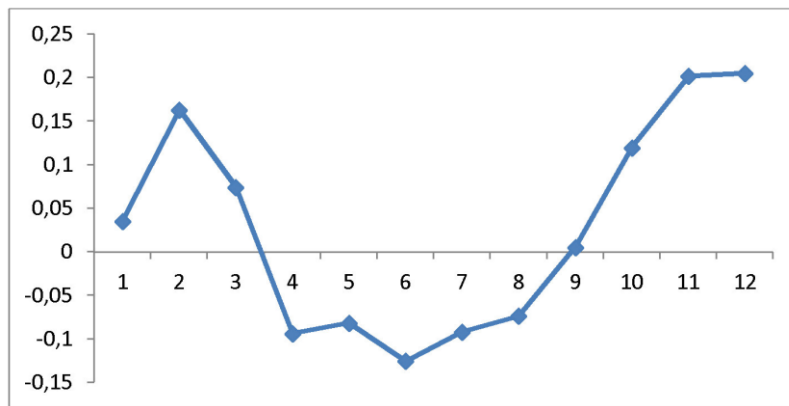
where the probabilities are determined by formula: $p_i(x) = \frac{x_i}{\sum x_i}$; $q_i(y) = \frac{y_i}{\sum y_i}$.

The results of the divergence between the actual probability distribution and the planned one (for the period from 2013 to 2018) obtained by expression (6) are presented in table 2.

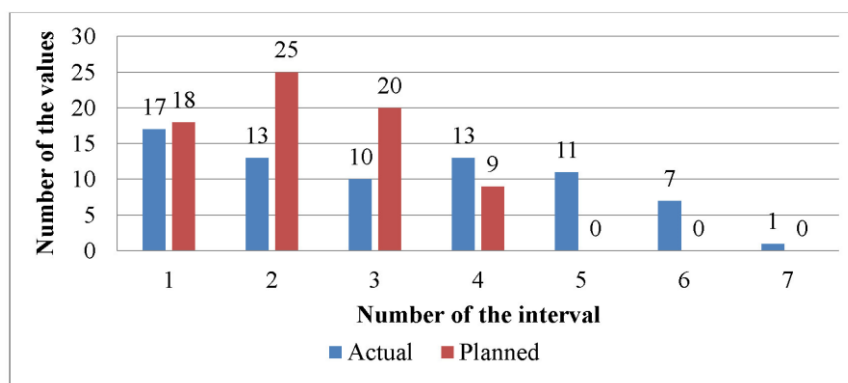
Table 2. Discrepancy between planned and actual distributions.

Month	2013	2014	2015	2016	2017	2018
January	0,002	0,006	-0,019	0,036	-0,016	0,027
February	0,071	0,046	0,012	0,019	0,027	-0,013
March	0,034	0,008	0,018	0,002	0,024	-0,011
April	0,006	-0,016	-0,004	-0,021	-0,028	-0,031
May	-0,021	-0,020	0,003	-0,005	-0,024	-0,015
June	-0,021	-0,021	-0,024	-0,023	-0,019	-0,018
July	-0,009	-0,018	-0,019	-0,020	-0,007	-0,018
August	-0,021	-0,003	-0,018	-0,012	-0,012	-0,007
September	-0,012	0,001	0,022	0,001	0,006	-0,006
October	-0,001	-0,006	0,015	0,013	0,051	0,047
November	0,022	0,016	0,025	0,029	0,059	0,051
December	0,001	0,061	0,037	0,043	-0,015	0,079

The divergence over the entire 6-year period is shown in figure 2.

**Figure 2.** Divergence for the period from 2013 to 2018.

The following can be seen from the table and figure: in February and December, the discrepancy between the planned and actual indicators is significant. Obviously, there are additional factors of a natural and social nature that are difficult to take into account. In turn, the size of the divergence shows, on the one hand: if the divergence is large, then there is enough insignificant information to identify factors that affect the deviation of the actual indicators from the planned ones; on the other – for small divergence, more information is required. Thus, measures to reduce commercial losses and, consequently, reduce financial risks are in demand.

**Figure 3.** Number of the values fall within the selected intervals.

In order to compare the planned losses with the distribution pattern of actual losses, the entire quantitative range from 117 to 4655 MWh (megawatt) was divided into 7 intervals (for each of the considered distributions). The obtained frequencies λ (indicate that the values fall within the selected intervals (figure 3) allowed to determine $q_i(\lambda)$, $p_i(\lambda)$ and use the equation (5) to determine the divergence, the values of which are given in table 3.

Table 3. Results of the divergence calculation.

№ interval	1	2	3	4	5	6	7
$D_q(p)$	0,015	0,336	0,280	-0,061	0.0	0.0	0.0

Summing up the results, we take into account the following. Having two probability distributions of a random loss value, one of which is assumed to be true. The hypothesis of truth should be confirmed by the actual data obtained during the operation of the technical system. If the two distributions completely coincide, the divergence is zero.

By looking at the results of divergence, you can control the confidence in the comparison of distributions. If the differences are large, then the development of measures aimed at reducing losses will not take much time, since the influence of factors on the change in the state of the system is clearly visible. However, if their differences are insignificant, that is, when the state of the system doesn't significantly deviate from the predicted one, then you will have to spend much more time looking at a lot of data and searching for insignificant factors.

Analysis of the results presented in table 2 showed the following: there is a significant divergence between the distributions (the total divergence is equal to 0.57 bits); the divergence in the first interval of 115÷765 thousand kWh can be considered satisfactory, and in the 2nd and 3rd intervals - unsatisfactory; in the intervals from the 5th to the 7th inclusive, the divergence values indicate the absence of the planned loss values, while the actual losses fell in the interval 2715÷4655 MWh.

4. Conclusion

Among the mathematical models based on statistical data processing, we can distinguish methods that allow us to determine the amplitude of fluctuations of the parameters, the percentage of deviation, the mathematical expectation, the variance and the standard deviation of a random variable. However, when the number of random factors is large and they are closely related, information theory is in demand. It opens up perspectives on information processing that have the goal of formally and concretely describing many things: such as comparing the probabilities of the desired and statistical distributions.

Taking into account the methods of measuring uncertainty and two sets of distributions, a number of questions can be answered:

- How are two different probability distributions related to each other?
- What is the distance between the probability distributions?
- How much does the desired state of the system differ from the natural one?
- How closely are the two variables dependent?

The application of the presented models from the information theory is clear, simple and endowed with good properties and fundamental origin.

Acknowledgments

The research was carried out with the financial support Program of the Russian Foundation for Assistance to Small Innovative Enterprises in Science and Technology by the grant "UMNIK" №13138GU/2018, code № 0040353.

References

- [1] Mokhov S L 2013 Commercial power losses and their reduction Retrieved from

- <http://energoser18.ru/energoseberezhenie/propaganda/publikaczii/kommercheskie-poteri-elektroenergii-i-ix-snizhenie.html>
- [2] Savina N V and Voropay N I 2008 System analysis of power losses in electric distribution networks *Novosibirsk: Nauka* p 228
- [3] Zhelezko Yu S, Artemyev A V and Savchenko O V 2007 Calculation of technological losses of electricity in electric networks *Energetik* **2** 29–30
- [4] Mogilenko A 2006 Reduction of power losses. Approach to planning and evaluating events *Electrical engineering news* **4**(40)
- [5] Vorotnitsky V E 2003 Standard of electric power losses in electric networks. How do I define and execute it? *Electrical engineering news* **6**(24) Retrieved from <http://www.news.elteh.ru/arh/2003/24/11.php>
- [6] Dulesov A S, Ereemeeva O S, Karandeev D J and Dulesova N V 2018 Approaches to Information Measurement of the Structure State of Technical Systems *IEEE FarEastCon* pp 1-6 Retrieved from 10.1109/FarEastCon.2018.8602799
- [7] Dulesov A S, Karandeev D J and Dulesova N V 2018 Improving the operation quality of technical systems using information theory models *MATEC Web Conf* **224** 1-6 DOI: <https://doi.org/10.1051/mateconf/201822404006>
- [8] Cover T M and Thomas J A 2006 *Elements of Information Theory* (New Jersey: Wiley and Sons) p 792
- [9] Kolmogorov A N 1965 Three approaches to the quantitative definition of information *International Journal of Computer Mathematics* **1** 3-11
- [10] Ross S M 2009 *Introduction to probability and statistics for engineers and scientists* (Associated Press) p 267
- [11] Lawless J F and Fredette M 2005 Frequentist predictions intervals and predictive distributions *Biometrika* **92** 529–42
- [12] Hartley R V L 1928 Transmission of Information *Bell System Tech. J* **7** 535–63
- [13] Shannon C E 1948 Mathematical Theory of Communication *Bell System Tech. J* **27** 379-423
- [14] Shannon C E 1949 Communication Theory of Secrecy Systems *Bell System Tech. J* **28** 656-715
- [15] Dulesov A S, Karandeev D J and Krasnova T G 2017 The evaluation of the correlation between entropy and negentropy in the structure of a technical system *MATEC Web Conf* **129** 1-4 DOI: <https://doi.org/10.1051/mateconf/201712903005>
- [16] Dulesov A S, Karandeev D J and Dulesova N V 2017 Reliability analysis of distribution network of mining enterprises electrical power supply based on measure of information uncertainty *IOP Conference Series: Earth and Environmental Science (EES)* **87** 032008
- [17] Dulesov A S, Karandeev D J, Khrustalev V I, Dulesova N V and Krasnova T G 2019 Determining the number of redundant elements of the distribution network in compliance with the specified amount of information entropy *Journal of Physics: Conference Series* **1399** 055030
- [18] Kullback S 1959 *Information Theory and Statistics* (John Wiley & Sons) p 395
- [19] Kullback S and Leibler R A 1951 On information and sufficiency *The Annals of Mathematical Statistics* **22** 79-86
- [20] Vorotnitsky V E, Kalinkina M A and Apriatin V N 2009 Measures to reduce power losses in the electric networks of power supply organizations *Energy Saving* **3** 53-6