

# Prediction of the power shortage in the electric power system by means of regression analysis by machine learning methods

D.A. Boyarkin<sup>1,2\*</sup>, D.S. Krupenev<sup>1,2</sup>, and D.V. Iakubovsky<sup>1,2</sup>

<sup>1</sup>Melentiev Energy Systems Institute of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russian Federation;

<sup>2</sup>Irkutsk National Research Technical University, Irkutsk, Russia

**Abstract.** Modern electricity consumers place increasingly high demands on the level of reliability of power supply and, correspondingly, the reliability of electric power systems (EPS). These requirements should be directly addressed in the EPS development planning tasks. To this end, the evaluation of the level of EPS reliability is performed by employing software and computer systems that have various methods of reliability evaluation implemented therein. Among the variety of methods for identifying reliability indicators to evaluate resource adequacy the most appropriate one is the Monte Carlo method (the method of statistical tests): it enables to perform calculations within a reasonable time with the required accuracy, while the calculation of complex EPS-like systems by means of analytical methods proves impossible because of the large dimensionality of the problem. However, even when using the Monte Carlo method, the difficulties arise as well, namely the problem of the need to reproduce a large number of random states of the simulated EPS and the calculation of the operating mode of each of them. There are several ways to reduce the overall calculation time by either efficient random number generators and optimizers or alternative methods of the calculation of operating modes. The article deals with the issue of bringing up to date the method behind reliability calculation that makes use of the Monte Carlo method. We propose to use regression analysis methods when calculating operating modes of random states of the EPS. To this end, we adopt the support-vector machine and the random forest method. Experimental studies covered in the article attest to the efficiency of the new method, for the 24-node system IEEE RTS-96 the calculation speed has been increased by almost a factor of 4 while maintaining accuracy.

## 1 Introduction

When planning the development of electric power systems (EPS) and justifying the construction of new power plants and networks, one has to evaluate the economic consequences of the measures to be taken. For this purpose, many different factors are considered, including the reliability of power supply to consumers, which in turn depends on the reliability of the EPS. We study resource adequacy as applied to the problems of EPS development planning.

Adequacy assessment is carried out using the mathematical models of EPSs that simulate their operation [1-3]. Among the variety of methods for identifying reliability indicators for adequacy assessment, the most appropriate one is the Monte Carlo method (the method of statistical tests): it enables to perform the reliability calculation within a reasonable time with the required accuracy, while the calculation of complex EPS-like systems by means of analytical

methods proves impossible because of the large dimensionality of the problem [4].

The methodology behind the evaluation of the resource adequacy of the EPS that is based on the Monte Carlo method is made up of the three computational steps:

1. The stage of defining the set of states of the EPS.

At this stage, a set of random states of the considered EPS is defined. By a random state we understand such configuration of the equipment that is part of the EPS that have the performance of each of its components as determined randomly based on the statistical data on their failure rate.

2. The stage of calculating the operating modes for the defined states.

At this stage, for each defined random state of the EPS, we calculate the established operating mode that is optimal with respect to minimizing the power shortage.

3. The stage of calculating reliability indicators.

At this stage, the reliability indicators are calculated by statistical processing of the results obtained at the

---

boyarkin\_denis@mail.ru

second stage.

When solving the problem of adequacy assessment, the most challenging is its dimensionality [5]. To achieve acceptable accuracy of the result one has to define and calculate a large number of states, the number of which depends on the complexity of a given EPS. Due to the computational complexity of the second stage, the overall calculation for the problem of adequacy assessment of the EPS may take an unacceptably long time, which reduces the efficiency of the EPS design.

This paper proposes a modification of the method of adequacy assessment by the Monte Carlo method with the aid of machine learning methods [6]. The essence of the proposed modification is to divide the total number of random states obtained at the first stage of adequacy assessment of the EPS into two sub-samples. After obtaining the power shortage values at the second stage for the first sub-sample, a training sample is defined. It is a set of key-value correspondences for each random state, where the key is a description of the performance of the EPS equipment and the loads at the nodes, while the value is the calculated value of the power shortage of the system. The remaining states from the second sub-sample are no longer processed at the computationally intensive second calculation stage, the power shortage value for the state is determined by the method of machine learning calibrated by the training sample. The problem of this kind is referred to as the regression analysis problem [7].

## 2 The algorithm of the application of machine learning to the problem of adequacy assessment of the EPS.

### *The regression problem.*

We propose to divide the set of random states obtained during the first stage into smaller (subject to training) and larger (subject to analysis) samples. The calculation algorithm for each sample will be different:

1. The evaluation of the power shortage for the states of the first sample is similar to the initial one (the first and second stages of the calculation) except that the description of the states themselves, as well as the result of the calculation, will define a training sample, on the basis of which the regression model will be built.

2. Training of the regression model.

3. In order to estimate the power shortage for each state of the analyzed sample at the second stage of the calculation, a regression model that is more efficient with respect to its time-intensity is used instead of optimization methods.

In the given problem of estimating the state-specific shortage based on the configuration of the EPS equipment performance, it is required to construct a regression model approximating the target dependence of  $X_A \rightarrow L$  between a set of  $X_A$  entities and a set of all  $L$  responses for the entities of the training sample  $S$ . In regression analysis,  $X_A$  represents a set of all states of the EPS, and  $L$  represents a set of all shortages specific to such states.

Each data entity  $\bar{x}_a \in X_A$ ,  $a = 1, \dots, A$ , represents a

vector in the  $d$ -dimensional space and is descriptive of the state of the EPS, where  $d$  is the number of features of the entity that describe this state,  $A$  is the number of all possible states of the EPS.

Training sample  $S$  is formed from a set of entities  $X_U$ , for which the value of the shortage level of the state of the EES is known:

$$S = ((\bar{x}_1, l_1), (\bar{x}_2, l_2), \dots, (\bar{x}_I, l_I)), \quad (1)$$

where  $\bar{x}_i \in R^d$ ;  $l_i \in L$ ;  $i = 1, \dots, I$ ;  $I$  is the number of objects in the training sample,  $X_U$  is the set of states of the EPS handled at the first stage,  $X_U \in X_A$ . The number of entities  $I$  in the training sample and, accordingly, their relation to the number of entities of all states, is set by the user as per the principle of achieving the best trade-off between accuracy and speed.

### *The classification problem.*

The problem of classification differs from the problem of regression analysis in the way of processing power shortages of random states. The algorithm behind its operation also consists of three steps, but instead of training the regression model, it is the classification algorithm that is trained. This requires the preparation of data, as the  $L$  responses in the training sample now represent a certain group, i.e. a class. In the problem of adequacy assessment  $L \in [-1, 1]$ , where -1 means a shortage-free state and 1 means a shortage state. For this reason, the third step of the calculations is also modified. Here, for each state of the EPS from a larger sample, the classifier predicts its level of shortage and, if it is shortage-free, this state is not analyzed for the availability and magnitude of the shortage, while the power shortage in the reliability zones and the EPS as a whole is assumed to be equal to 0. If the algorithm classifies the state as a shortage, the exact power shortage value is determined at the second calculation stage. This reduces the number of states handled therein.

Modification of the Monte Carlo method by means of classification has been discussed in detail in previous studies [8]. This article assesses the advantages the regression model has over them.

Both problems are based on regression [9]: dependence of the expected value of a random variable on one or several other random variables (free variables):

$$E(y|x) = f(x), \quad (2)$$

Regression analysis is the search for such function  $f$  that describes this dependence. For this purpose, we set a regression model that in its generic form is as follows:

$$f: (W * X) \rightarrow Y, \quad (3)$$

where  $W$  is the set of all parameters,  $x \in X$  is the space of free variables,  $Y$  is the space of dependent variables. Since the regression analysis assumes the search for the dependence of the expected value of a random variable on free variables (2), the function (3) will include random variable  $e$ :

$$y = f(w, x) + e, \quad (4)$$

One has to find such parameters  $w$ , which would provide the least error between the actual values of the dependent variable and the reconstructed ones when substituting them into function  $f(w, x)$ .

The classification problem differs from the regression problem in that the dependent variable  $y$  takes

one of the two values 0 or 1, and the dependence function is the probability of the event  $y=1$ :

$$P\{y = 1|x\} = f(z), \tag{5}$$

where  $f$  is the logistic function (sigmoid function):

$$f(z) = \frac{1}{1 + e^{-z}}, \tag{6}$$

and where  $z = (W * X)$ .

The sigmoid function will change within (0, 1) and thus the classification problem gets solved, entity  $x$  can be classified as belonging to the class  $y=1$ , if the probability predicted by the model  $P\{y = 1|x\} > 0.5$ , and can be classified as belonging to the class  $y=0$  otherwise.

Just like in the regression problem, one has to find such parameters  $w$ , which would provide the least error between the actual values of the dependent variable and the reconstructed ones when substituting them into the function  $f(w, x)$ .

The efficiency and applicability of the regression analysis are determined by the advantage in the speed of the calculation and, consequently, by the reduction of the time spent. However, increasing the complexity of the method by introducing machine learning methods and the associated time spent on defining training samples, model training, selection of hyperparameters, and prediction should be compensated by time savings due to reducing the number of calls to optimization methods. Hence, it follows that a potentially successful algorithm of machine learning for solving our problem should differ not only in its accuracy but also in its overall execution speed at all the above-mentioned stages of its operation.

### 3 Experimental calculations

To evaluate the proposed methodology, a series of experiments was conducted to determine the power shortage of the EPS. The 24-node IEEE RTS-96 was used as the system under study [10], the diagram of which is presented in Figure 1. The main characteristics of nodes and links are given in [10]. This system is characterized by high reliability and being shortage-free, so as the single operation mode for all calculations we chose the most intense one, with the highest load on December 17, 18:00.

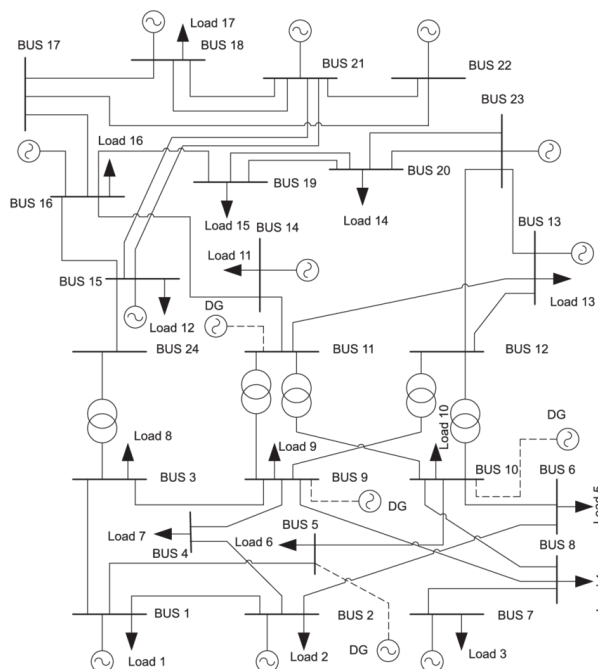


Fig. 1. IEEE RTS-96 24-node system diagram.

In total, three groups of calculations of 5 computational experiments were carried out within the scope of the experiment:

- 1) Determination of the power shortage of the EPS without applying machine learning methods. The obtained result was later used as a benchmark to assess the accuracy of the proposed methods and their computational efficiency.
- 2) Determination of the power shortage of the EPS with the use of machine training. The problem of machine learning is stated as a classification problem.
- 3) Determination of the power shortage of the EPS with the use of machine training. The problem of machine learning is stated as a problem of regression analysis.

The results of the first stage are shown in Figure 2. Values of the power shortage of the EPS are obtained in the conventional way that makes no use of machine training methods. In each calculation, 30,000 random system states were considered. Table 1 shows the time taken to perform each calculation.

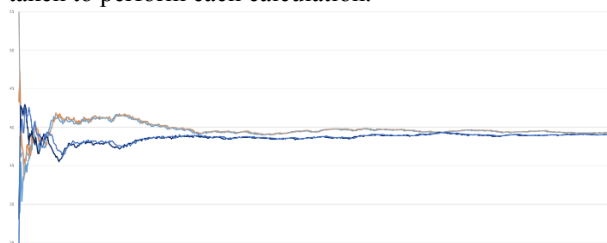


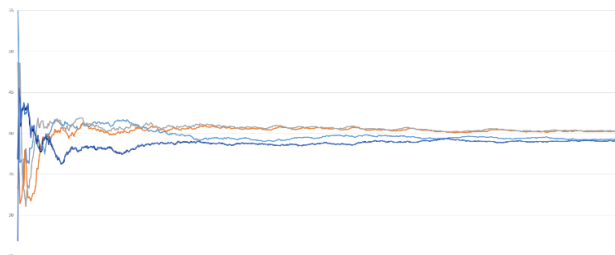
Fig. 2. The expected value of the power shortage as per the conventional calculation.

Table 1 Time of execution of each calculation experiment and the average over them.

Calculation	1	2	3	4	5
Execution time (sec)	236.94	237.012	235.958	238.579	238.595
Average execution time (sec.)	237.4168				

The chart shows that the average value of the established expected value of the power shortage for all five experiments is equal to 39 MW. The deviation of the calculation time is caused by a unique set of calculated states in each calculation experiment.

At the second stage, the support-vector machine (SVM) was used, that is the classification method that showed the best results according to the results of the comparative analysis of machine learning methods in [8]. The training sample was 5% of the total number of random system states. Figure 3 shows the results of a series of calculations, and Table 2 shows the time taken to perform each of them. Table 2 also shows the number of states marked as shortage-free by the model.



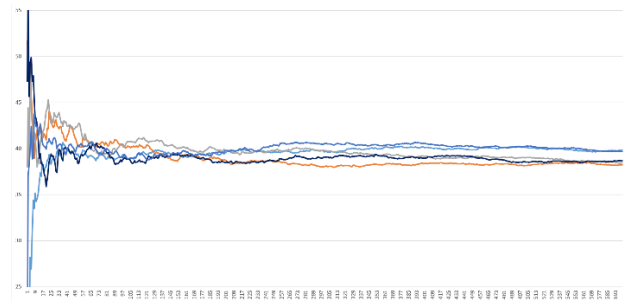
**Fig. 3.** The expected values of the power shortage in the calculation using the SVM method for the classification of the power shortage level of operating modes.

**Table 2.** Runtime and the number of rejected shortage-free states of each calculation experiment and the average over them.

Calculation	1	2	3	4	5
Execution time (sec)	139.962	142.447	142.256	140.854	141.929
Average execution time (sec)	141.4897				
The number of rejected shortage-free states	19,400	19,231	19,236	19,347	19,351
The average number of rejected shortage-free states	19,313				

As can be seen from the table, the average speed of the calculation increased and amounted to 141 seconds instead of 237 seconds for a similar calculation. As can be seen from the figure, the accuracy is preserved, with the spread in values not exceeding 4% of the reference value. The value of the established expected value falls within [38.2, 40.8].

At the third stage, the problem was transformed into a simpler problem of regression analysis. The random forest method was used for the calculation. The training sample was 5% of the total number of random system states. Figure 4 shows the results of a series of calculations, and Table 3 shows the time taken to complete each of them.



**Fig. 4.** The expected values of the power shortage in the calculation using the RF method to determine the power shortage level of operating modes.

**Table 3** Time of execution of each calculation experiment and the average over them.

Calculation	1	2	3	4	5
Execution time (sec)	84.0492	84.2189	85.9105	85.1065	85.2673
Average execution time (sec)	84.910488				

The result yielded by the experiment shows that the calculation time was reduced to 85 seconds and amounted to only ~25% of the initial time. As can be seen from the figure, the accuracy is preserved, with the spread in values not exceeding 3% of the reference value. The value of the established expected value falls within [38.8, 40.1].

## 4 Conclusion

The computational efficiency of the means of evaluating the reliability of the EPS is one of the criteria that is to be met for their efficient application to real-world problems. Reduction of time of the calculation of the system will enable to reduce the requirements for computational resources and to expand the scope of problems for which the method can be applied.

Within the scope of this study, we proposed a method of modification of the conventional method of the power shortage calculation by means of the Monte Carlo method. Its application makes it possible to significantly increase the speed of calculation of the EPS reliability indicators.

Experimental studies reported in the article confirm the efficiency of the new technique with the speed increase in the case of the 24-node system IEEE RTS-96 amounting to 153 seconds, which is to say that the speed of calculation increased by nearly a factor of 4.

The efficiency of the proposed methodology can therefore be concluded.

The research was financially supported by the Russian Foundation for Basic Research within the framework of scientific project No. 18-37-00234.

## References

1. Kovalev G. F., Lebedeva L. M. Nadeshnost' sistem elektroenergetiki. Novosibirsk: Nauka, 2015. (In Russian)
2. Wenyuan Li Probabilistic Transmission System Planning // Wiley-IEEE Press, 2011, 376 p.
3. R. Billinton, R. Allan Reliability Evaluation of Power Systems. (2nd ed.). Plenum Press. 1996.
4. Kovalev G.F., Krupenev D.S., Lebedeva L.M. Models and Methods for Estimation and Optimization of Electric Power System Reliability // Reliability: Theory & Applications. Vol.12. No.2. 2011.
5. Boyarkin D.A., Krupenev D.S., Iakubovskiy D.V. Formirovanie sluchajnykh sostojanij elektroenergeticheskikh sistem pri otsenke ikh nadezhnosti metodom statisticheskikh ispytanij / «Nadezhnost i bezopasnost energetiki». – 2017. – No. 1. – P. 33-41, Moscow. (In Russian)
6. Vapnik V. N., Chervonenkis A. Ya. Teoriya raspoznavanija obrazov. — Moscow.: Nauka, 1974. (In Russian)
7. Draper N., Smith G. Prikladnoj regressionnyj analiz. Moscow: Izdatelskij dom «Viljams», 2007. (In Russian)
8. Boyarkin D. A., Klassifikatsiya sostojanij elektroenergeticheskoy sistemy pri otsenke nadyozhnosti metodom Monte-Karlo. / Sistemnye issledovaniya v energetike, Trudy molodykh uchenykh ISEM SO RAN. 2018. Issue 48, P. 98-106. (In Russian)
9. Brandt Z. Analiz dannyh. Moscow: Mir. 2003. (In Russian)
10. C. Grigg, P. Wong, etc. The IEEE Reliability Test System - 1996 / IEEE Transactions on Power Systems, Vol. 14, NO. 3, P. 1010 – 1020, 1999.