Reliability Engineering and System Safety 204 (2020) 107171

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Improvement in the computational efficiency of a technique for assessing the reliability of electric power systems based on the Monte Carlo method

Dmitry Krupenev*, Denis Boyarkin, Dmitrii Iakubovskii

Melentiev Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

ARTICLE INFO

Keywords:
Machine learning
Monte Carlo method
Power system
Random sequences
Reliability

ABSTRACT

The reliability of energy systems is assessed to control their operation and expansion. An effective method for reliability assessment is the Monte Carlo method. This process, however, is often time-consuming due to the large size of the power system. This interferes with subsequent control problems. The speed of reliability assessment and the accuracy of the result for the Monte Carlo method directly depend on the number of randomly generated states of the system, their quality and the complexity of the subproblem to be solved for each state. When solving such a subproblem for reliability assessment, random states can be defined as a shortage and shortage-free ones. To assess the reliability of power systems using the Monte Carlo method, one should analyze only the state of the system with a shortage. We suggest the use of machine learning methods to eliminate or sort the shortage and shortage-free states. The paper demonstrates the effectiveness of two methods: a support vector machine and a random forest. It also shows their performance when the Monte Carlo and quasi-Monte Carlo methods are used.

1. Introduction

Modern power systems are characterized by high complexity and a great number of problems that influence the reliability of electricity supply to consumers. Quality and continuity of electricity supply become increasingly more critical indicators for both large consumers and small consumers (householders). Requirements of modern consumers for reliable electricity supply can be satisfied by the effective measures aimed at the improvement and maintenance of the power system reliability level in system expansion planning and its operation control. However, such measures may prove to be economically inefficient and excessive, and therefore, it is necessary to timely assess power system reliability to make informed decisions. The reliability of power systems can be assessed on-line (for the current operating conditions) and off-line (for prospective expansion).

The technique based on the Monte-Carlo method is the most effective and widely used technique for power system reliability assessment [1–5]. This method is used in many software systems intended for energy systems, such as GE-MARS, GridView, PLEXOS, and DIgSILENT / PowerFactory. Unlike other widely used methods, for example, analytical ones, it can reduce the problem of high dimensionality in large systems, which are involved in practical calculations most frequently.

The reliability assessment technique based on the Monte Carlo method [6] consists of the following steps [7–10]:

1) Generation of power system random states;
2) Minimization of power shortage in power system random states;
3) Calculation of reliability indices.

Normally, for the implementation of this technique, the first two steps are integrated by a common cycle.

For the operational control of the electric power system, the reliability of the current operating condition or security is assessed, and a set of random states is formed relative to this condition, the probabilities characterizing the state of the power equipment and a load of consumers. The long-term planning of the EPS expansion involves the assessment of adequacy. In this case, the annual interval is considered, and the planned and unplanned (emergency) repairs of power equipment, a change in the load curve and its random deviations are taken into account.

High dimensionality, which increases with dimensionality and complexity of the considered systems, creates difficulty even when the Monte Carlo method is used. For example, for a small power system consisting of 50 components of generating and network equipment, the number of random states in the complete enumeration of all possible equipment failure options will be about 10^{15}. The EPS structure also affects the speed of calculation. With its complication, the process of power shortage minimization becomes more complicated, and, consequently, the time for the EPS reliability assessment is increased. The

* Corresponding author.
E-mail address: krupenev@isem.irk.ru (D. Krupenev).

https://doi.org/10.1016/j.ress.2020.107171
Received 16 October 2019; Received in revised form 9 July 2020; Accepted 8 August 2020
Available online 10 August 2020
0951-8320/ © 2020 Elsevier Ltd. All rights reserved.

speed of the EPS reliability assessment is essential for assessing both security and adequacy. After the security evaluation, a set of control actions that improve the EPS security should be proposed based on the security indices, and the faster and more accurately the reliability indices are obtained, the higher the probability of the reliable operation of the EPS. After assessing the EPS adequacy, based on the obtained adequacy indices, the generating capacity reserves, as well as the structure and transfer capability of the network, are substantiated. In this case, for all combinations of power equipment commissioning, the adequacy should be evaluated. An analysis of real electric power systems can involve thousands of options for the commissioning of power equipment. Therefore, by increasing the speed of reliability assessment for each EPS expansion option and observing the required accuracy, the validity of decisions made on the EPS expansion is increased.

At present, the machine learning methods are increasingly more used in various spheres, including the techniques based on the Monte Carlo method [11–13]. This work continues the similar studies presented in [14]. As a rule, the application of machine learning methods does not directly concern the technique, but it is connected with the modification of repeated operations in the body of the technique or setting of the parameters. In this paper, which develops our results in [15], we propose these methods for classification of power system states with and without power shortage. For this purpose, the classifiers are used to determine the shortage level of generated power system states without their calculation in the computationally expensive second stage (calculation block). The paper presents two classification methods (support vector machine [16] and random forest [17]), the assessment of their efficiency and accuracy with respect to the problem statement, and their applicability in general. Additionally, the use of machine learning methods is considered along with the use of the Monte Carlo and quasi-Monte Carlo methods. The quasi-Monte Carlo method is widely used to solve computationally difficult problems, for example, in stochastic financial mathematics or in modeling physical processes [18–21]. Its main advantage is that it has convergence close to $O(\frac{1}{N})$, while the convergence rate of the Monte Carlo method is of the order of $O(\frac{1}{\sqrt{N}})$, which is ensured by the use of quasi-random sequences (or low-discrepancy sequences (LDS)) [22] instead of random generators, which are usually pseudorandom ones. Based on the studies, a qualitative analysis of the applicability of machine learning methods for classifying design states to assess the EPS reliability using the Monte Carlo method and the benefits of using Sobol sequences is presented, which is an improvement in the methods for the EPS reliability evaluation.

The paper consists of 4 parts. The first part states the problem of adequacy assessment by the Monte Carlo method. The second part describes the classification of shortage and shortage-free EPS states based on machine learning methods. The focus of the third part is on the efficiency of the Mersenne twister and the Sobol LDS-sequence generator when used to assess the EPS reliability. The experimental part of the paper is concerned with the test of the proposed approach to increasing the computational efficiency of the technique for EPS reliability assessment using the methods of random number generation and machine learning to determine their most effective combination.

2. Power system reliability assessment based on the Monte Carlo method

As noted above, the technique for power system reliability assessment based on the Monte-Carlo method consists of three computational stages. Their more detailed description is below:

1) *Generation of power system states*. In this stage, the states of power system facilities and the value of consumer loads are modeled based on the statistics on emergency rates of energy facilities and random deviations of consumer loads. One random event K_f, $f = 1, ..., F$, with the probability p_f, $f = 1, ..., F$, is modeled when generating one random number r_f, $f = 1, ..., F$ from the equidistributed set R in the interval [0, 1]. If r_f is in the interval [0, p_f], the event is considered to occur, otherwise, not to occur, i.e.:

$$K = \begin{cases} 1, & if\ r_f \in [0,\ p_f] \\ 0, & if\ r_f \in (\ p_f,\ 1] \end{cases},\ f = 1, ..., F. \tag{1}$$

The distributed set R is a sequence of random numbers generated by the pseudo- or quasi-random numbers generators.

1) *Minimization of power shortage in generated power system states*. The steady state that is optimal in terms of minimum power shortage is calculated for each modeled state. For the power system adequacy assessment, the mathematical formulation of the problem has the following form [7,15]

for power shortage evaluation of each power system state determine:

$$\sum_{i=1}^{I} y_i \rightarrow max, \tag{2}$$

subject to the balance constraints

$$g_i - y_i + \sum_{j=1}^{J} (1 - a_{ji} z_{ji}) z_{ji} - \sum_{j=1}^{J} z_{ij} = 0,\ i = 1,\ ..., I;\ i \neq j, \tag{3}$$

and the linear constraints on variables

$$\begin{cases} y_i \le \bar{y}_i, \\ g_i \le \bar{g}_i \quad i = 1, ..., I;\ j = 1, ..., J;\ i \neq j, \\ z_{ij} \le \bar{z}_{ij}, \end{cases}$$
$$y_i \ge 0,\ g_i \ge 0,\ z_{ij} \ge 0, \quad i = 1, ..., I;\ j = 1, ..., J;\ \ i \neq j, \tag{4}$$

where g_i is the capacity used at node (reliability zone) i, MW; $\bar{g}_i$ is the available generating capacity at node i, MW; y_i is the value of firm load at node i, MW; $\bar{y}_i$ is the load at node i, MW; z_{ij} is the power flow from node i to node j, MW; $\bar{z}_{ij}$ is the transfer capability of the transmission line between nodes i and j, MW; a_{ij} are the given coefficients of specific losses of power when transmitted from node i to node j, $i = 1, ..., I;\ j = 1, ..., J;\ i \neq j$.

1) *Calculation of reliability indices*. After the calculation of all generated states, the calculated parameters of power systems are processed by statistical methods. In the end, we obtain a set of power system reliability indices. Below are some main reliability indices:
- the probability of failure-free (shortage-free) operation;
- the electricity undersupply expectation;
- the power shortage expectation;
- the coefficient of power availability;
- the probability of operating parameters deviation beyond the maximum admissible values.

3. Machine learning methods

The second stage of the electric power system adequacy assessment involves solving the problem of power shortage minimization. This process is time-consuming, therefore a reduction in the number of states to be calculated while maintaining the accuracy of the estimate will increase the computational efficiency of the whole technique.

We propose dividing a set of random states obtained during the first stage into smaller and larger data sets. The calculation of indices for the states of the first set will not differ from the initial solution, except that the states themselves, as well as the calculation result, will form a training data set, which will be used as a basis for the classifier to be built. Then, for each state of an electric power system from a larger set, the classifier predicts its shortage and, if the state does not have a

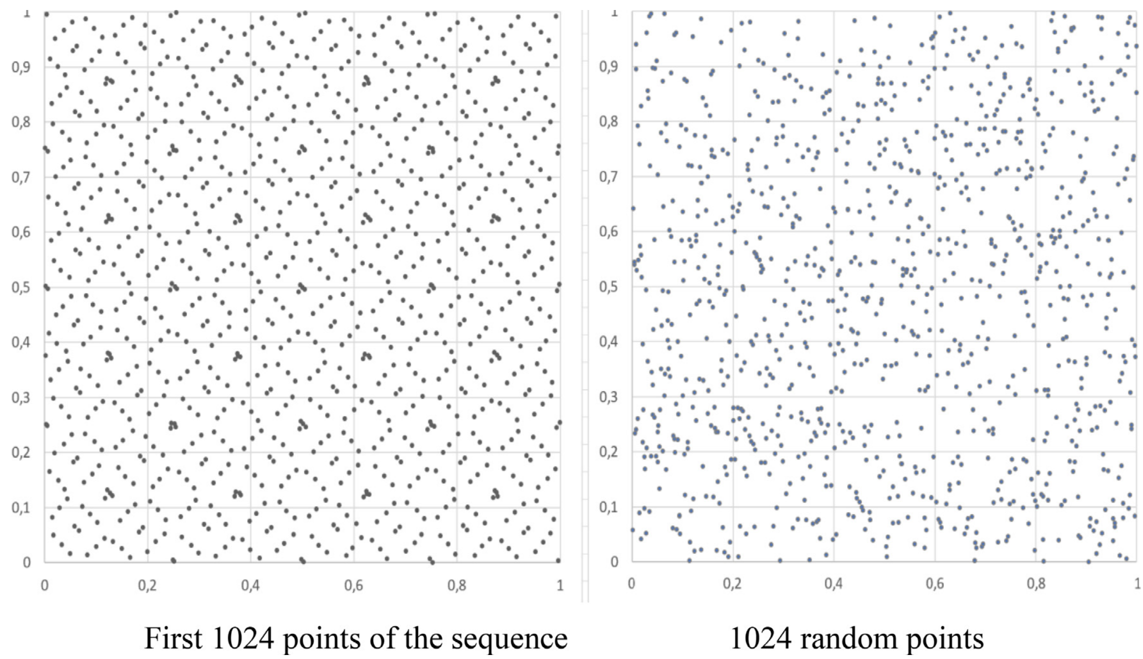

Fig. 1. Graphical representation of points. (a) First 1024 points of the sequence (b) 1024 random points.

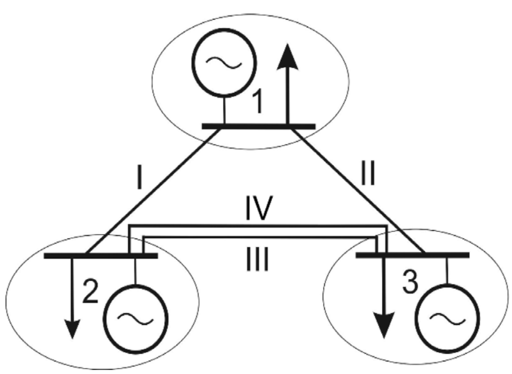

Fig. 2. Scheme of the studied 3-node power system.

shortage, this state is not analyzed for the shortage and its size. The power shortage at nodes of the electric power system is assumed to be 0. If the algorithm classifies the state as the one with shortage, then the exact value of the power shortage is determined in the second calculation block, which reduces the number of states to be considered in it.

The considered problem of state-with-shortage classification according to the configuration, and operability of electric power system equipment is a typical precedent training problem, which requires that an algorithm be constructed to approximate the objective relationship $X_A \to L$ between a set of objects X_A and a set of all answers L for the objects of training set S. In the problem of shortage classification, X_A represents a set of all EPS states, and $L \in [-1, 1]$, where 1 means a state without shortage, and 1 is a state with shortage.

Each data object $\bar{x}_a \in X_A,\ a = 1, ..., A$, represents a vector in the d-dimensional space and characterizes the electric power system state where d is a quantity of object features defining this state, A is the number of all possible states of the electric power system.

The training data S is formed from a set of objects X_U, for which the value of shortage in the electric power system state is known:

$$S = ((\bar{x}_1, l_1), (\bar{x}_2, l_2), ..., (\bar{x}_n, l_n)), \tag{5}$$

where $\bar{x}_i \in R^d$; $l_i \in L$; $i = 1, ..., n$; n is the number of objects in the training set, X_U is a set of electric power system states modeled in the first stage, $X_U \in X_A$. The number of objects n in the training set is specified by the user and normally does not exceed half of the number of modeled states N.

The effectiveness and applicability of the proposed technique are determined by the increase in the calculation speed and, accordingly, by a reduction in the time expenditure. However, the complication of the technique by involving machine learning methods and the related time spent on construction of a training data set, model training, selection of hyperparameters and classification should be compensated for by reducing the number of calls to the optimization methods. Thus, a potentially successful machine learning algorithm for solving our problem should differ not only in the classification accuracy but also in the overall speed in all the above-mentioned stages of its work.

The paper considers two most common methods of machine learning. These are the linear kernel support-vector machine [16] and the random forest [17]. These methods were not chosen by chance, the application of various machine learning methods in electric power problems was analyzed [23–25]. The analysis has indicated that the linear kernel support-vector machine works faster but less accurately, while the random forest can require more time.

3.1. Support-vector machine

The support vector method is one of the most popular training methods, which is used to solve classification and regression problems. The main idea of the method is to construct a hyperplane separating the sample objects in an optimal way [16]. Optimality is understood in the sense of minimizing the upper bounds on the generalization error probability. It is assumed that the greater the distance (gap) between the separating hyperplane and the objects of the separated classes, the smaller the average error of the classifier will be.

3.2. Random forest method

The random forest is an ensemble of simple decision trees, each built based on a random sample from the original training set (bagging), and only a fraction of randomly selected features is used to split the vertices [17]. The optimal number of trees is selected to minimize the error of the classifier on the test sample. Classification of objects is carried out by voting: each committee tree assigns the classified object to one of the classes, and the winner is the class for which the largest

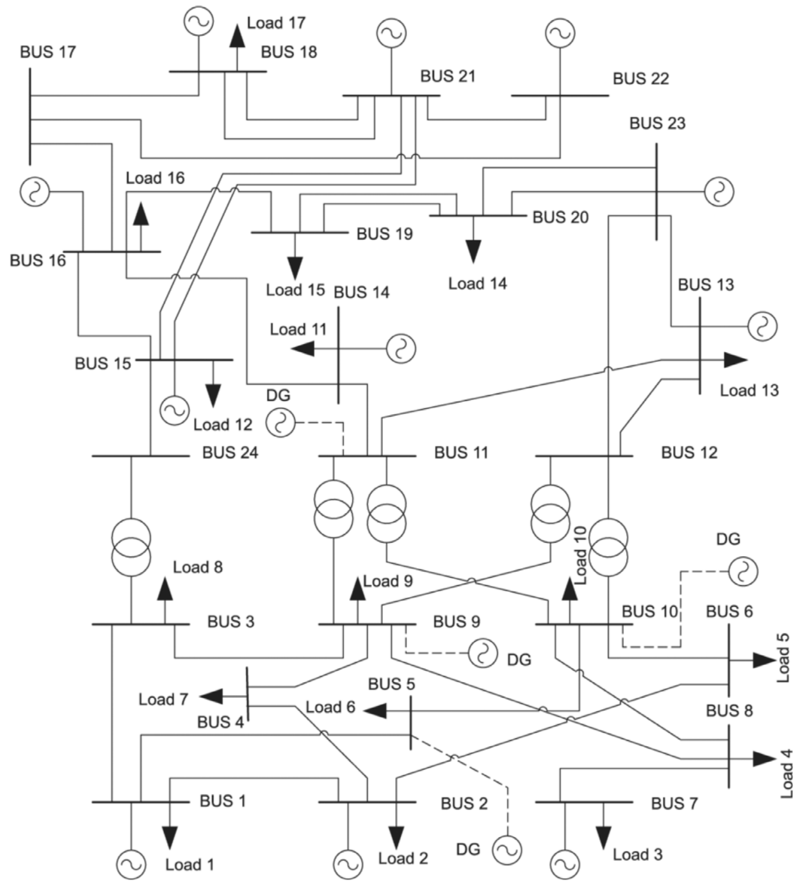

Fig. 3. Scheme of the studied 24-node (bus) IEEE RTS-96 system.

number of trees voted.

4. Method for the formation of random states of EPS

In the formation of random states of EPS, one of the important points that directly affects the computational efficiency of the whole technique is the operation of the random number generator. The requirements imposed on their work include a long cyclic period, the most accurate reproduction of a given random distribution, and the absence of correlation in the mutual arrangement of the sequence members. These properties affect the results of the simulation by the method of statistical tests.

The widespread use of the Monte Carlo method in computer modeling has predetermined the use of pseudorandom number generators as a source of entropy, which is associated with the difficulty and low efficiency of sequences obtained from the sources of real random numbers.

Another step was the use of low-discrepancy sequences [22] as a replacement for random number sequences, which gave rise to a whole subclass of quasi-Monte Carlo numerical methods. There is a relationship between the discrepancy of the sequence and the equidistribution [26,27]. Such a significant feature as a low discrepancy of these sequences provides a higher convergence rate than the classical Monte Carlo method [6].

Nevertheless, not only the algorithm used but also the implementation features affect the speed and efficiency of the generators

The paper focuses on two representatives of their subclasses of random number generators – a pseudorandom Mersenne Twister [28] and a quasi-random generator of Sobol sequences [29].

4.1. Mersenne twister

The Mersenne twister is a pseudo-random number generator based on the properties of the simple Mersenne numbers [28]. The Mersenne numbers are the numbers of the form:

$$w_l = 2^t - 1,\ \ t = 1,\ \ldots,\ T, \tag{6}$$

where: t is the natural number.

The Mersenne Twister belongs to the class of so-called generalized feedback shift register generators (TGFSR) [30]. “Twister” is the transformation that ensures equidistribution of generated pseudo-random numbers in 623 measurements. Therefore, the function of correlation between two sequences of samples in the output sequence of the Mersenne twister is negligibly small.

This algorithm quickly generates pseudo-random numbers that are of high quality by the randomness criterion, does not have many shortcomings inherent in other pseudo-random number generators, such as short period, predictability, easily revealed statistical laws.

An additional advantage is the presence of an effective algorithm implementation called SIMD-oriented Fast Mersenne Twister (SFMT) [31].

4.2. Sobol sequence

Sobol sequence is a low-discrepancy sequence introduced by I.M. Sobol [29]. It represents the points located in a certain way in a hypercube and is defined as follows: let $p_1, \ldots, p_s \in F_2(x)$ be primitive polynomials ordered by nondecreasing degrees. Then, for $1 \le i \le s$ let

$$p_i(x) = x^{e_i} + a_{1,i}x^{e_i-1} + \ldots + a_{e_i-1,i}x + 1. \tag{7}$$

Take arbitrary odd positive integers $m_{1,i}, \ldots, m_{e_i,i}$, such that $m_{k,}$

Table 1
Main characteristics of generating units of the 3-node power system and RTS-96.

No of node	Rated capacity of generating unit, MW	The number of generating units, pcs.	Emergency rate of generating unit, p.u.	Load, MW
3-node system				
1	50	5	0.05	450
	100	2	0.05	
2	50	5	0.05	400
	100	1	0.05	
3	50	5	0.05	490
	100	1	0.05	
	200	1	0.05	
24-node IEEE RTS-96 system				
1	20	2	0.10	108
	76	2	0.02	
2	20	2	0.10	97
	76	2	0.02	
3	-	-	-	180
4	-	-	-	74
5	-	-	-	71
6	-	-	-	136
7	100	3	0.04	125
8	-	-	-	171
9	-	-	-	175
10	-	-	-	195
11	-	-	-	0
12	-	-	-	0
13	197	3	0.05	265
14	-	-	-	194
15	12	5	0.02	317
	155	1	0.04	
16	155	1	0.04	100
17	-	-	-	0
18	400	1	0.12	333
19	-	-	-	181
20	-	-	-	128
21	400	1	0.12	0
22	50	6	0.01	0
23	155	2	0.04	0
	350	1	0.08	
24	-	-	-	0

$_{i} < 2^k$ for $1 \le k \le e_i$. For all$k > e_i$, the numbers $m_{k,\ i}$ are determined recursively using the bitwise operator XOR (exclusive or $\oplus$):

$$m_{k,i} = 2a_{1,i}m_{k-1,i} \ \oplus \ 2^2a_{2,i}m_{k-2,i} \ \oplus \ \ldots \oplus 2^{e_i-1}a_{e_i-1,i}m_{k-e_i-1,i} \ \oplus \ 2^{e_i}a_{e_i,i} m_{k-e_i,i}\ . \tag{8}$$

Next, the direction numbers $\vartheta_{k,\ i}$ are determined as

$$\vartheta_{k,i} = \frac{m_{k,i}}{2^k}. \tag{9}$$

As a result, for an arbitrary point in the sequence n$\in$ N_0 having a binary decomposition, which is expressed as $n = n_0 + 2n_1 + \ldots + 2^{(r-1)} n_{r-1}$, the ith coordinate has the form:

$$x_{n,i} = n_0\vartheta_{1,i} \oplus n_1\vartheta_{2,i} \oplus \ldots \oplus n_{r-1}\vartheta_{r,i}. \tag{10}$$

Thus, the Sobol sequence is defined as the collection $(x_0, x_1, \ldots)$, where $x_n = (x_{(n,1)}, \ldots, x_{n,s})$.

There is also an advantage of the Sobol sequence over other quasi-random sequences, which is the fact that for this sequence we know the effective sequential algorithm of Antonov I.A. and Saleev V.M. [32] based on Gray codes.

5. Experimental studies

The first stage of the experimental studies involved a statistical analysis of the sequences of numbers generated by the Mersenne twister and Sobol sequence. The obtained two-dimensional points are presented in Fig. 1 as a point diagram.

Table 2
Main characteristics of transmission lines of the 3-node power system and RTS-96.

No of transmission line	Transmission line vector	Transfer capability, MW	Emergency rate per 100 km, p.u.	Length, km
3-node power system				
1	1-2	150	0,001	400
2	1-3	150	0,001	400
3	2-3	150	0,001	400
4	2-3	150	0,001	400
24-node IEEE RTS-96 system				
1	1-2	175	0.009079396	4.828
2	1-3	175	0.00065774	88.5139
3	1-5	175	0.001063992	35.4056
4	2-4	175	0.000838297	53.1084
5	2-6	175	0.000680955	80.4672
6	3-9	175	0.000869499	49.8897
7	3-24	500	0.0	0.0
8	4-9	175	0.00094577	43.4523
9	5-10	175	0.001048572	37.0149
10	6-10	175	0.00512046	25.7495
11	7-8	175	0.00132999	25.7495
12	8-9	175	0.000725824	69.2018
13	8-10	175	0.000725824	69.2018
14	9-11	500	0.0	0.0
15	9-12	500	0.0	0.0
16	10-11	500	0.0	0.0
17	10-12	500	0.0	0.0
18	11-13	500	0.00094577	53.1084
19	11-14	500	0.001049316	46.671
20	12-13	500	0.00094577	53.1084
21	12-23	500	0.000605575	107.826
22	13-23	500	0.000637213	96.5606
23	14-16	500	0.001098145	43.4529
24	15-16	500	0.002145717	19.3121
25	15-21	500	0.000940903	54.7177
26	15-21	500	0.000940903	54.7177
27	15-24	500	0.00088863	57.9364
28	16-17	500	0.001517173	28.9682
29	16-19	500	0.001658054	25.7495
30	17-18	500	0.002496834	16.0934
31	17-22	500	0.000577179	117.4821
32	18-21	500	0.001517173	28.9682
33	18-21	500	0.001517173	28.9682
34	19-20	500	0.001078178	44.257
35	19-20	500	0.001078178	44.257
36	20-23	500	0.001768591	24.1402
37	20-23	500	0.001768591	24.1402
38	21-22	500	0.000747058	75.6392

Fig. 1 shows that the points of the Sobol sequence form a grid of points filling the space uniformly. The points generated by the Mersenne twister algorithm form areas of concentrations and rarefactions, which are inherent in all random number generators to a varying extent.

The more objective evaluation is obtained using the Kolmogorov-Smirnov criterion [33]. This criterion is intended to test simple hypotheses that the analyzed sequences belong to some completely known distribution law, in our case to equidistribution. The closer the criterion value to 0.87, the higher the randomness degree of equidistributed random numbers. The analysis was based on 1024 one-dimensional points generated for each algorithm, the criterion values equaled 0.84 for the Sobol sequence and 0.97 for the Mersenne twister, which indicates higher randomness of the first algorithm.

The proposed methods were evaluated in the second stage of experimental studies by the practical analysis. The reliability indices were calculated for two test schemes – the 3-node scheme presented in Fig. 2

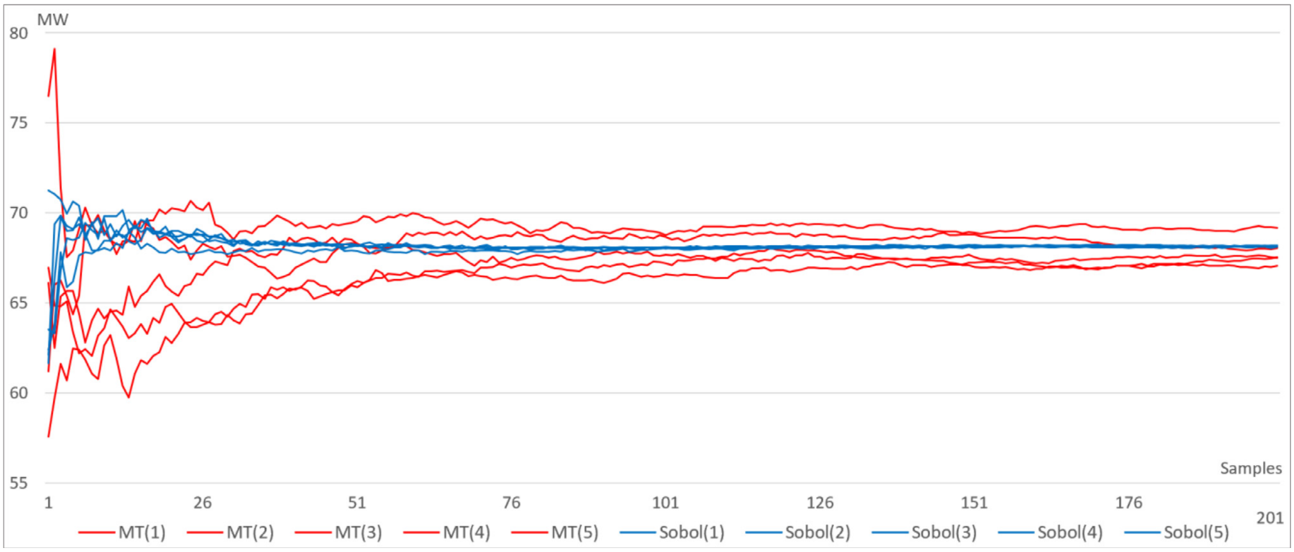

Fig. 4. Power shortage expectation in the test 3-node system.

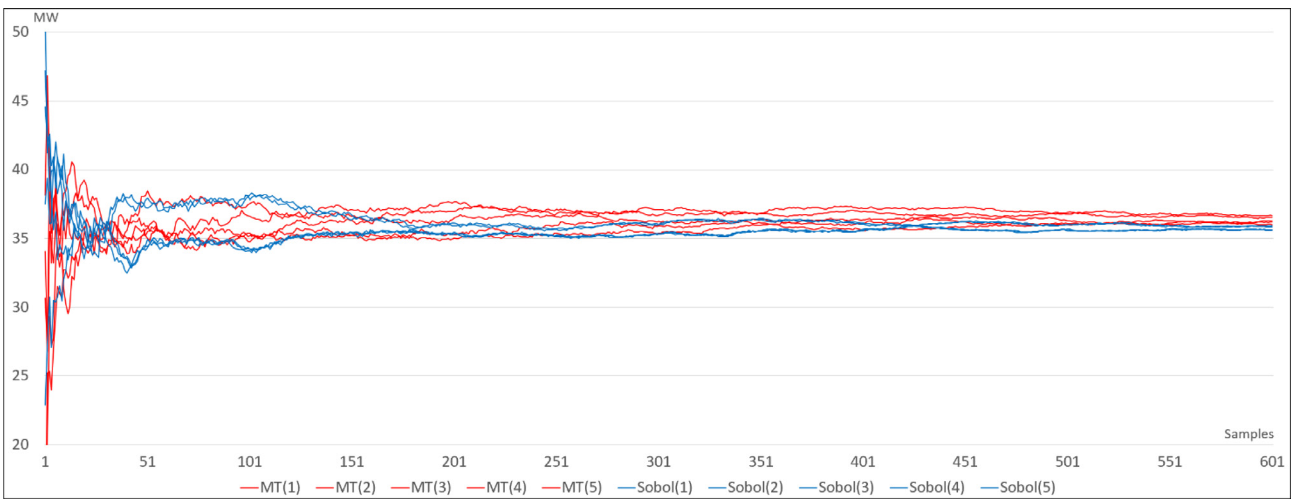

Fig. 5. Power shortage expectation in the RTS-96 system.

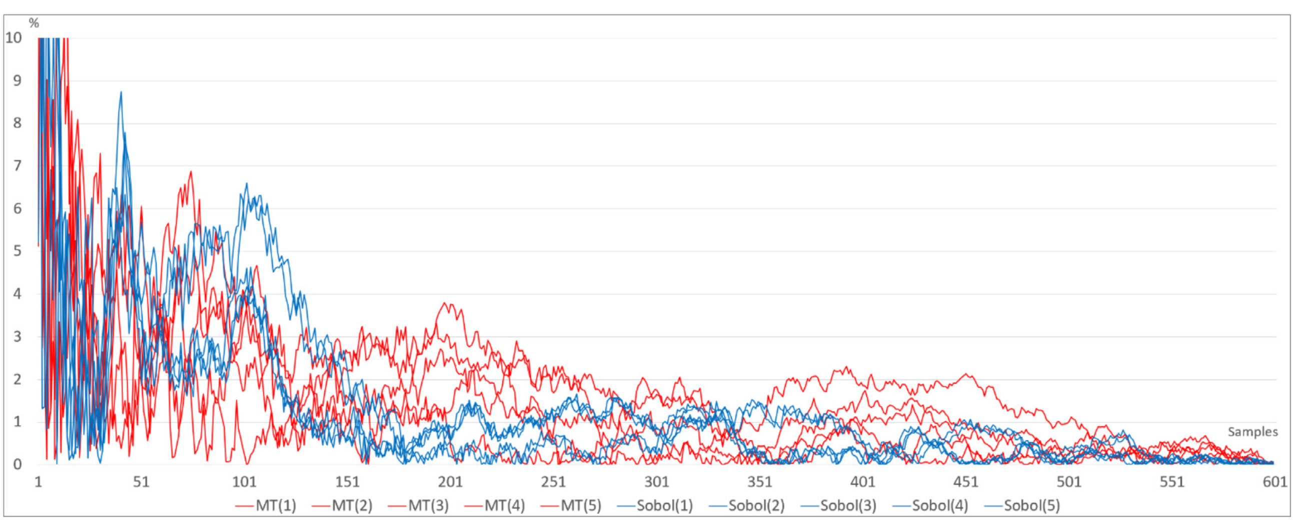

Fig. 6. Diagram of a decrease in the relative error versus the number of tests.

Table 3
The system used for estimation.

Python version	3.6
Operating system	Windows 10 × 64
Processor	Intel i7 8700K
RAM	16 GB
Video card	No

and the 24-node (bus) IEEE RTS-96 system [34] presented in Fig. 3. Table 1 demonstrates the main characteristics of power system nodes. Table 2 presents the main characteristics of transmission lines (the characteristics of the transmission lines for RTS-96 are converted into metric units).

The computational efficiency of the described generators was compared based on a set of experiments on reliability assessment that involved the determination of power shortage expectation for 10000 states of one operating condition for the 3-node system and 30000 states for RTS-96 system. We calculated the operating condition at the maximum shortage hour of the year. Five experiments were carried out for each generator. The results of power shortage stabilization for the 3-node system are presented in Fig. 4, for RTS-96 system – in Fig. 5. Stabilization for the Mersenne twister is shown in red and that for Sobol sequence- in blue.

Fig. 4 shows that the power shortage expectation for the Sobol sequences converges to the “true” value faster than for the pseudorandom numbers generated by the Mersenne Twister. Moreover, uniformity in the variance decrease allows the automatic termination of calculations, when the required accuracy of the result is achieved. This advantage of Sobol sequences was noted in other works, for example in [18]. However, the result obtained for a larger scheme (for the RTS-96 system) does not allow similar firm conclusions to be drawn. Hence, to assess the efficiency of the algorithms for larger schemes, it is appropriate to apply statistical analysis methods, for example, the relative error estimate:

$$\delta = \frac{|I_u - I_{true}|}{I_{true.}} \cdot 100\%. \tag{11}$$

Since the true value is unknown, the average of the last 250 values of power shortage expectation was taken as the true value in each calculation. The diagram of a decrease in the relative error is presented in Fig. 6.

The obtained results indicate that the application of the Sobol sequences to assess power system reliability by the Monte-Carlo method provides more qualitative calculation results and their achievement based on a fewer number of random states.

The efficiency of the technique for improving the current method of adequacy assessment was tested at the second stage of experimental studies by the machine learning methods. For this purpose, we made calculations similar to the first stage and estimated the program run speed for different combinations of machine learning methods (random forest and support vector machine) and generation of random numbers (Mersenne twister and Sobol sequences).

The model parameters were selected by monitoring the AUC ROC metric value [35] for the fivefold cross-validation. The penalty coefficient was varied in the method of support vector machine [16]. The depth of trees and the number of considered random features during division were selected for the random forest method. The class weight imbalance was revealed for both methods. For prediction, preference is given to a shortage state. This is explained by the fact that in practical problems, the training set is usually unbalanced, the vast majority of generated states of real power systems are shortage-free. It is also worth noting that the statement of our problem differs from that of classical machine learning problems that normally have one studied object, to which the model is adjusted in the best possible way. In our case, each calculation power system scheme is an independent object. Therefore, it is impossible to create and adjust the only one model, because these operations are performed automatically at each new scheme calculation using cross-validation. As a result, the optimal parameters can differ even for two identical calculations due to the existing dependence on the randomly formed training set.

The test schemes were calculated by the Python-based program for the conditions indicated in Table 3. Parallel calculations were not used.

Fig. 7 presents the calculation results for the 3-node scheme, and Table 4 shows the time spent on the corresponding calculation and the number of states indicated by the classifier as shortage-free (hereafter – rejected states). Here, the total time of calculations by the techniques based on the machine learning methods includes the time spent on the formation of the training set, model learning, and selection of parameters.

Fig. 7 shows that the methods and the combinations of methods

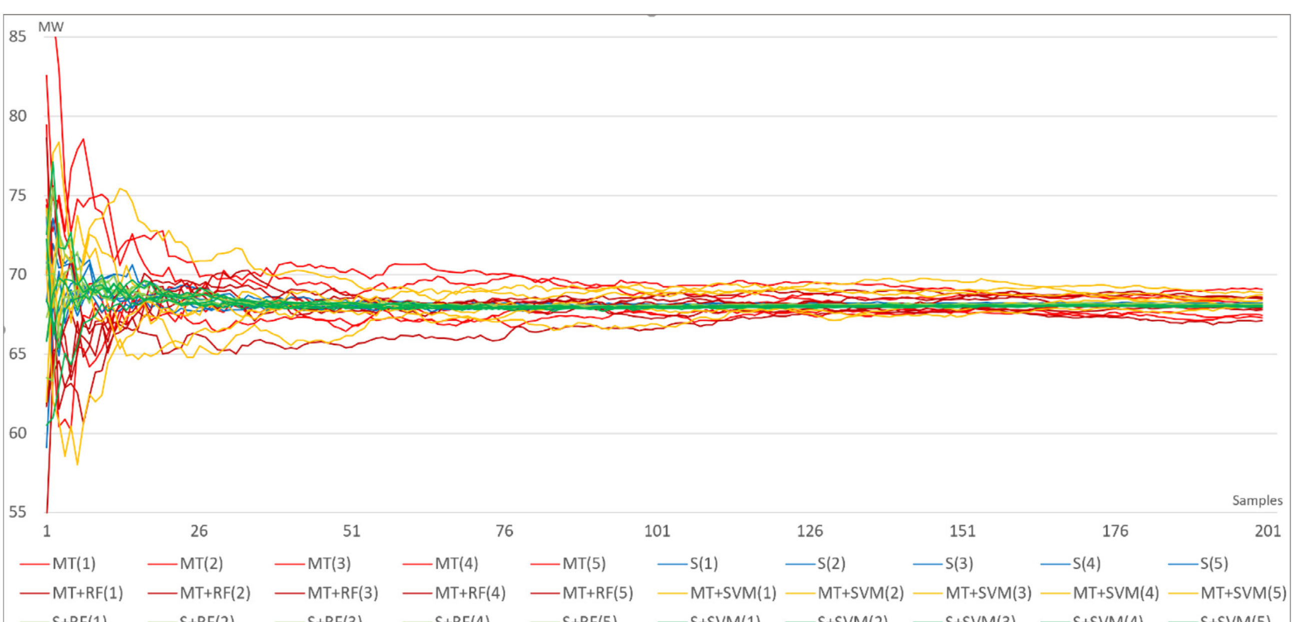

Fig. 7. Power shortage expectation in the test 3-node system.

Table 4
Operation time of combinations of the methods and the number of rejected states in the calculation of the test 3-node system.

Combination of methods	Average time of work (sec.)	The average number of rejected states
Mersenne twister (MT)	28.33520942	-
Sobol sequence (S)	29.45201483	-
Mersenne twister + Random forest (MT + RF)	30.80900955	1302
Mersenne twister + Support vector machine (MT + SVM)	24.99896854	1584
Sobol sequence + Random forest (S + RF)	31.61228132	1400
Sobol sequence + Support vector machine (S + SVM)	25.94910803	1778

Table 5
The values of model quality metrics based on the holdout data set.

Methods	Accuracy	F1-metric
Support vector machine	0.8945	0.9130
Random forest	0.9482	0.9650

based on the Sobol sequences are the best from the standpoint of convergence speed and quality (blue and green color shades). Classification accuracy is better for the random forest method, which is confirmed by the analysis of metrics in Table 4. However, this method requires more time for calculations, which contradicts the task stated in the study. The support vector machine is worse in accuracy, but this deviation is negligible, and the technique speed with its use is much higher, as is seen from the graph.

Fig. 8 presents the calculation results for the RTS-96 system, and Table 6 shows the time spent on the corresponding calculation and the number of states indicated by the classifier as shortage-free. The Table also includes the information on the range of values, as it is of concern in the analysis of the results obtained.

Similar to the first stage of experimental studies, the result presented by the graph is ambiguous. However, the information of the Table can be used to carry out a preliminary analysis. As in the case of the 3-node scheme calculation, the best stabilization results are achieved by the methods based on the Sobol sequences. It is more interesting to compare the methods of random forest and the support vector machine. As seen from Table 6, the range of variation in the number of rejected states by the random forest method is quite wide, which indicates strong dependence of the classification quality on the initial data set, which is formed randomly. In turn, the total calculation speed is directly proportional to the number of rejected states, and the best time corresponds to the least number of rejected states. However, this is not so important, because the random forest does not demonstrate an increase in the calculation speed even with optimal parameters, and in combination with the Sobol sequences retards it. The support vector machine is learned fast and provides sensibly accurate results, the speed of power shortage calculation increases on the average threefold without loss of the result accuracy.

We analyze the convergence speed of power shortage expectation using the relative error formula (11). The results obtained by the random forest method are excluded from the graph as contradicting the problem statement. Fig. 9 presents the diagram of a change in the relative error for the calculations performed.

The trend of the green-blue color in Fig. 9 shows that from the standpoint of relative error reduction, the combinations with the Sobol sequences are the most effective for power system reliability assessment by the Monte-Carlo method.

6. Conclusion

The computational efficiency of the techniques and software for power system reliability assessment is one of the criteria of their effective application to practical problems.

In the assessment of the EPS reliability using the Monte Carlo method, one of the factors affecting computational efficiency is the number of random states of EPSs that are necessary to achieve the required accuracy of the power shortage value. The reduction in the

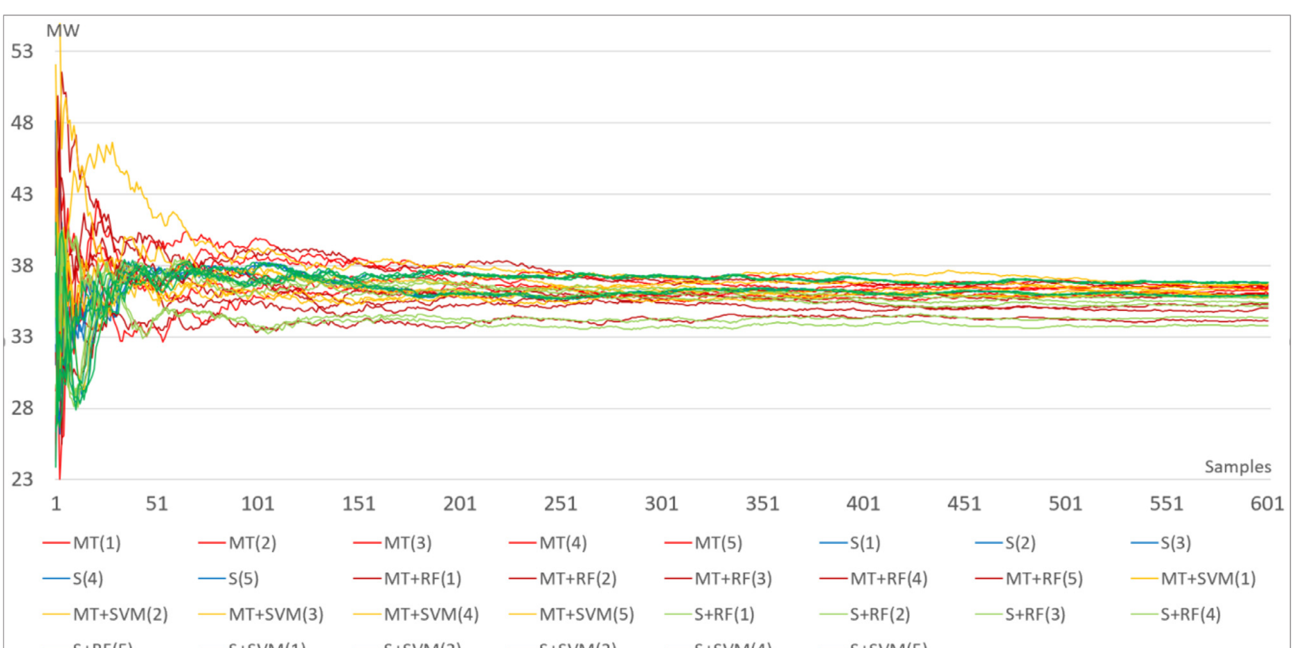

Fig. 8. Power shortage expectation in the RTS-96 system.

Table 6
Operation time of combinations of methods and the number of rejected shortage-free states in the calculation of the RTS-96 system for one condition.

Combination of methods	Average time of work and range of values (sec.)		The average number of rejected states and a range of values	
Mersenne twister (MT)	1311.073	Min:1290.22 Max: 1333.88	-	-
Sobol sequence (S)	1327.08	Min: 1318.46 Max: 1340.27	-	-
Mersenne twister + Random forest (MT + RF)	1068.324	Min: 834.279 Max: 1246.326	12536	Min: 8932 Max: 17611
Mersenne twister + Support vector machine (MT + SVM)	474.786	Min: 454.96 Max: 498.749	20460	Min: 20409 Max: 20517
Sobol sequence + Random forest (S + RF)	1751.371	Min: 1546.144 Max: 2117.412	7396	Min:1854 Max: 11326
Sobol sequence + Support vector machine (S + SVM)	478.78	Min: 473.1639 Max: 481.423	20458	Min: 20439 Max: 20472

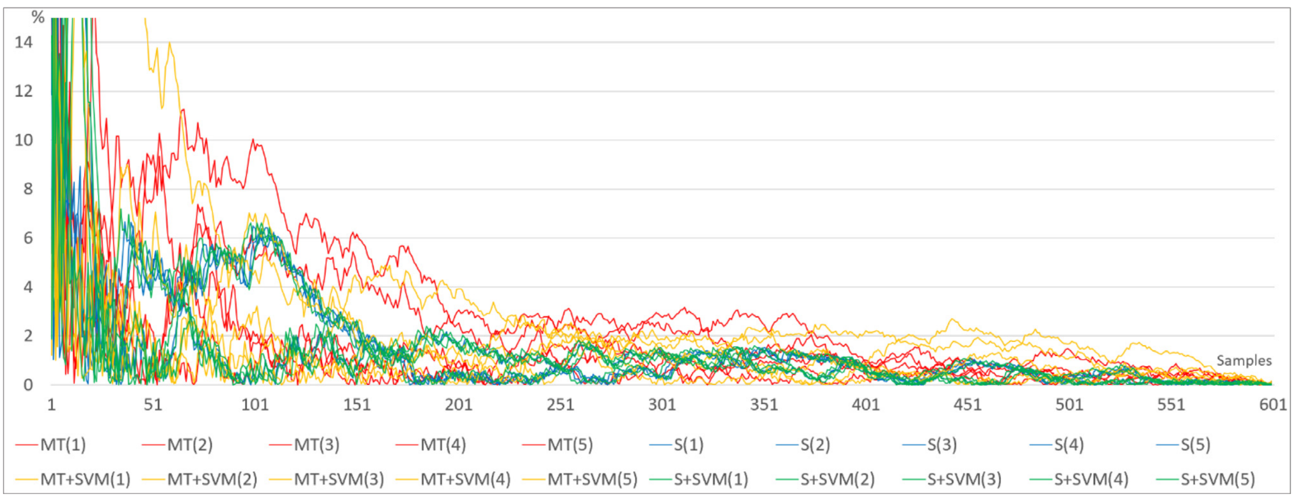

Fig. 9. Diagram of relative error reduction versus the number of tests.

number of such system states will increase the speed and accuracy of calculations overall.

The first stage of the study involved an analysis of two random number generators for the formation of random EPS states: the Mersenne Twister and Sobol sequences without using machine learning methods. Calculations were performed for two test power systems using the data from these generators. Expectedly, the results of the calculations indicated that the quasi-random Sobol sequence proved to be the most effective generator of random states for assessing the EPS reliability.

The second stage of the studies employed the machine learning methods (support vector machine and random forest) to reduce the number of analyzed power system states. The efficiency of both methods and the technique overall were analyzed.

The analysis showed that both methods allowed reducing the number of analyzed states necessary to determine the value of the mathematical expectation of a system power shortage. In some cases, however, this does not lead to a reduction in time expenditure. When applying the random forest method in combination with the Sobol sequence, the time spent on calculations was higher than in the option with the Sobol sequence alone, which is due to the low speed of the random forest method (training and prediction). With the increase in the size of the electric power system, however, this disadvantage is reduced. Already in the calculation of the RTS-96 system, which consists of 24 nodes, both machine learning methods showed a positive impact on the calculation efficiency, i.e., reduced the number of analyzed random states of the EPS and calculation time. For this reason, the proposed technique for the improvement of the speed and accuracy of calculations can be evaluated as positive for power system adequacy assessment.

Based on the results obtained, it can be concluded that the best combination of methods for calculating the power shortage of EPS is the use of the Sobol sequence combined with the support vector machine. The obtained result improves the practice of EPS adequacy assessment and allows us to move on to solving the problems of synthesizing the reliability of EPSs, substantiating the level of reservation of generating capacities, as well as the structure and bandwidth of the electric network at a new higher level.

CRediT authorship contribution statement

Dmitry Krupenev: Conceptualization, Methodology, Supervision, Writing - original draft, Writing - review & editing. **Denis Boyarkin:** Methodology, Investigation, Software, Resources, Writing - review & editing. **Dmitrii Iakubovskii:** Methodology, Investigation, Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was carried out as part of the project of the Melentiev Energy Systems Institute of Siberian Branch of the Russian Academy of

Sciences "Methodological framework considering the reliability factor in managing the development of intelligent energy systems" (No. AAAA-A17-117030310450-3).

References

[1] All Continental Europe and Nordic TSOs' proposal for assumptions and a Cost Benefit Analysis methodology in accordance with Article 156(11) of the Commission Regulation (EU) 2017/1485 of 2 August 2017 establishing a guideline on electricity transmission system operation. European Network of Transmission System Operators for Electricity, March 2018.

[2] European Resource Adequacy Assessment - Methodology Proposal in accordance with Article 23 of the Electricity Regulation of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (recast). European Network of Transmission System Operators for Electricity, December 2019.

[3] Fishman GS. A comparison of four Monte Carlo methods for estimating the probability of s − t connectedness. IEEE Trans Reliability 1986;35(2):145–55.

[4] Karger DR. A randomized fully polynomial time approximation scheme for the all terminal reliability problem. SIAM Rev 2001;43(3):499–522.

[5] Nel LD, Colbourn CJ. Combining Monte Carlo estimates and bounds for network reliability. Networks 1990;20:277–98.

[6] Sobol IM. Numerical Monte Carlo methods. / The main edition of the Physical and Mathematical Literature of the Nauka Publishing House; 1973. (in Russian).

[7] Kovalev GF, Lebedeva LM. Reliability of power systems. Springer; 2019. p. 237. 1st ed.

[8] Li W. Probabilistic transmission system planning. Wiley-IEEE Press; 2011. p. 376.

[9] Billinton R, Allan R. Reliability evaluation of power systems. Plenum Press; 1996. (2nd ed.).

[10] Ghajar R, Billinton R. A Monte Carlo simulation model for the adequacy evaluation of generating systems. Reliab Eng Syst Saf 1988;20(3):173–86.

[11] Sterna RE, Song J, Work DB. Accelerated Monte Carlo system reliability analysis through machine-learning-based surrogate models of network connectivity. Reliab Eng Syst Saf 2017;164:1–9. August.

[12] Tracey B, Wolpert D, Alonso JJ. Using supervised learning to improve Monte Carlo integral estimation. AIAA J. 2011;51(8).

[13] Montemanni R, D'Ignazio F, Chou X, Gambardella LM. Machine learning and Monte Carlo sampling for the probabilistic orienteering problem. Joint 10th international conference on soft computing and intelligent systems and 19th international symposium on advanced intelligent systems. 2018.

[14] Rocco CM, Moreno JA. Fast Monte Carlo reliability evaluation using support vector machine. Reliab Eng Syst Saf 2002;76(3):237–43.

[15] Iakubovskii DV, Krupenev DS, Boyarkin DA. An analysis of shortage minimization models to assess power system adequacy. Energy Syst Res 2018;1(3):25–32.

[16] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97.

[17] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.

[18] Paskov S, Traub J. Faster valuation of financial derivatives. J Portf Manage 1995;22:113–20. Fall.

[19] Galanti S, Jung A. Low-discrepancy sequences: Monte-Carlo simulation of option prices. J Deriv New York 1997;5:63–83.

[20] Alban A, Darjia H, Imamurac A, Nakayama MK. Efficient Monte Carlo methods for estimating failure probabilities. Reliab Eng Syst Saf 2017;165:376–94. September.

[21] Hou T, Nuyens D, Roelsa S, Janssen H. Quasi-Monte Carlo based uncertainty analysis: sampling efficiency and error estimation in engineering applications. Reliab Eng Syst Saf 2019;191:1–14. November.

[22] Niederreiter H. Low-discrepancy and low-dispersion sequences. J Number Theory 1988;30(1):51–70. September.

[23] Negnevitsky, M., Tomin, N., Kurbatsky, V., Panasetsky, D., Zhukov, A., Rehtanz, C. A random forest-based approach for voltage security monitoring in a power system (2015) 2015 IEEE Eindhoven PowerTech, PowerTech 2015, art. no. 7232460. doi: 10.1109/PTC.2015.7232460.

[24] (...) Diao R, Sun K, Vittal V, O'Keefe RJ, Richardson MR, Bhatt N, Stradford D, Sarawgi SK. Decision tree-based online voltage security assessment using PMU measurements. IEEE Trans Power Syst 2009;24(2):832–9. https://doi.org/10.1109/TPWRS.2009.2016528.

[25] Kalyani S, Swarup KS. Classification of static security status using multi-class support vector machines. J Eng Res 2012;9(1):21–30. https://doi.org/10.24200/tjer.vol9iss1pp21-30.

[26] Niederreiter H. Random number generation and quasi-Monte Carlo methods. Dover: Dover Publications; 2006.

[27] Sobol IM. Multidimensional quadrature formulas and Haar functions. M., Nauka; 1969. p. 288.

[28] Matsumoto M, Nishimura T. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. ACM Trans Model Comput Simul 1998;8(1):3–30.

[29] Sobol IM. Distribution of points in a cube and approximate evaluation of integrals. U.S.S.R Comput Maths Math Phys 1967;7:86–112.

[30] Matsumoto M, Kurita Y. Twisted GFSR generators. ACM Trans Model Comput Simul 1992;2(3):179–94. July.

[31] Saito M, Matsumoto M. SIMD-oriented Fast Mersenne Twister: a 128-bit pseudorandom number generator. Monte Carlo and quasi-Monte Carlo methods 2006. Springer; 2008. p. 607–22.

[32] Antonov IA, Saleev VM. An economical method for calculating sequences. Zh. Vychisl. Mat. Mat. Fiz (J Comput Math Math Phys) 1979;19(1):243–5.

[33] Kolmogorov A. Sulla determinazione empirica di una legge di distribuzione. Guornale dell' Inst Ital Degli Attuari 1933;4(1):83–91. V.

[34] Grigg C, Wong P. The IEEE reliability test system - 1996. IEEE Trans Power Syst 1999;14(3):1010–20.

[35] Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn 1997;30(7):1145–59.