**THE JOURNAL IS INCLUDED IN THE LIST OF THE LEADING JOURNALS AND EDITIONS OF THE HIGHER ATTESTATION COMMISSION (VAK)**

*Dependability,*
*DOI:10.21683*
*ISSN 1729-2646*
*e-ISSN 2500-3909*



**THE JOURNAL PROMOTER:**
JSC «NIIAS»

*It is registered in the Russian Ministry of Press, Broadcasting and Mass Communications. Registration certificate ПИ 77-9782, September, 11, 2001.*

*Official organ of the Russian Academy of Reliability*

**Publisher of the journal**
LLC Journal "Dependability"
**Director**
Dubrovskaya A.Z.
The address: 109029, Moscow,
Str. Nizhegorodskaya, 27, Building 1, office 209
Ltd Journal "Dependability"
www.dependability.ru
Printed by OOO Otmara.net. 2/1 bldg 2,
Verkhiaya Krasnoselskaya St., floor 2, premise II,
rooms 2A, 2B, 107140, Moscow, Russia.

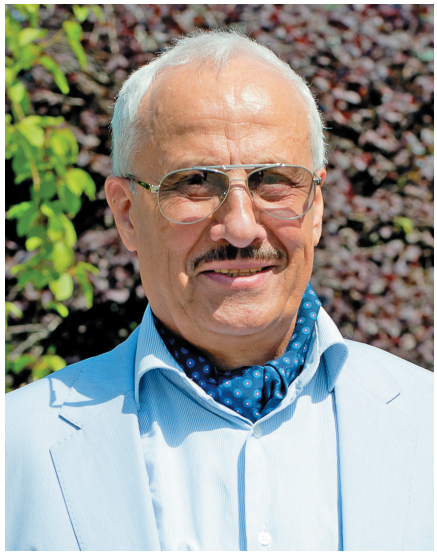
Circulation: 500 copies.
Printing order
Papers are reviewed. Signed print 18.03.2022,
Volume , Format 60x90/8, Paper gloss

The Journal is published quarterly since 2001.
The price of a single copy is 1595 Rubles, an annual subscription costs 6380 Rubles.
Phone: +7 (495) 967 77 05.
E-mail: dependability@bk.ru.

Papers are reviewed.
Papers are published in author's edition.

**THE JOURNAL IS PUBLISHED WITH THE PARTICIPATION AND SUPPORT OF THE JOINT-STOCK COMPANY «RESEARCH AND DESIGN INSTITUTE OF INFORMATISATION, AUTOMATION AND COMMUNICATION ON RAILWAY TRANSPORT» (JSC «NIIAS»)**

## CONTENTS



**Dear colleagues, authors and readers,**

The very hard year 2021 is over. The pandemic greatly affected our lives. Avoiding a coronavirus infection required self-isolation from other people. Many of us work remotely. This new phenomenon has become part of our everyday lives. Remote learning, thesis and dissertation defense, workshops and conferences are commonplace now. Today, digital human interaction is the new normal in the life of the society. Digital technology has acquired a strong momentum for development. Efficiency, safety and dependability are its key characteristics. Those matters will be at the focus of the 2022 publications of the Dependability Journal.

The Journal's subject matter emphasises technical efficiency, transportation and functional safety, information security, structural and functional dependability of technical and man-machine systems, reliability and survivability of systems, standardization and certification in terms of system dependability and safety. Risk theory and the practice of risk management is one of the primary topics of the Journal's papers. The focus will be on new findings in the area of artificial intelligence, especially with regard to efficiency, safety and dependability in transportation.

The priority is still given to the items that reflect the results of practical application of advanced technologies, methods and engineering solutions.

The Journal is open for publication of advertisement materials highlighting the latest achievements in the domain of design, application and development of technical systems and processes as regards their efficiency, dependability and safety.

In 2022, Dependability will still be published in Russian and English. At www.dependability.ru, the readers can access the digital versions of the Journal's issues. All the issues will simultaneously be published in paper form.

The Editorial Board and Editorial Council are continuing their efforts aimed at improving the scientific quality of the Journal. The publications are included in the Russian Science Citation Index, as well as the international Index Copernicus. The inclusion of the Journal's papers into the Web of Science (WoS) database and citation system is pending agreement.

The Editorial Board calls the authors and readers for active involvement with the Journal. Your observations and proposals will help improve its quality, as well as its scientific and application level.

***Best regards,***
***Editor-in-Chief, Professor***
***I.B. Shubinsky***

*Original article*
*https://doi.org/10.21683/1729-2646-2022-22-1-4-12*

# Analysing the effect of information redundancy on the dependability indicators of distributed information systems

**Vladimir V. Kulba**[1], **Sergey K. Somov**[1]*, **Alexey B. Shelkov**[1], *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russian Federation*
**ssomov2016@ipu.ru*


*Vladimir V. Kulba*


*Sergey K. Somov*


*Alexey B. Shelkov*

**Abstract. Aim**. *The paper analyses the effect of information redundancy on the functional dependability indicators of distributed automated information systems. Information redundancy in the form of hot standby and HDD archives located in the system nodes is examined.* **Methods.** *The concepts of the probability theory and Markov processes are employed.* **Results.** *Indicators of operational dependability of distributed information systems and the effect of operational and recovery redundancy of data sets on these indicators are analysed. The paper analyses the efficiency of three backup strategies in distributed systems.* **Conclusions.** *Using information redundancy significantly improves the dependability and operational efficiency of distributed systems. At the same time, this type of redundancy requires a certain increase in operating costs.*

**Keywords:** *distributed information systems, data backup and recovery, failures, faults, dependability indicators of information systems.*

**For citation:** *Kulba V.V., Somov S.K., Shelkov A.B. Analysing the effect of information redundancy on the dependability indicators of distributed information systems. Dependability 2022;1: 4-12. https://doi.org/10.21683/1729-2646-2022-22-1-4-12*

**Received on:** *06.10.2021* / **Upon revision:** *07.02.2022* / **For printing:** *18.03.2022.*

## Introduction

Today, information technology is a key element of the managerial control infrastructure that allows improving its efficiency, minimizing the cost of various resources, stimulating labour productivity and other management performance indicators.

The operational dependability of any automated information system significantly depends on the integrity of the data it uses. The matter of ensuring a high system dependability and integrity of the data it uses is of particular relevance for large, geographically distributed multi-level systems of various purpose, such as most advanced information systems operated by the Russian Railways, e.g., KASANT, the Automated System for Tracking, Supervision and Elimination of Technical Failures and Dependability Analysis [1].

Connecting hundreds and thousands of computers with communication channels into large-scale computer networks of various scales and topologies allowed creating distributed automated information systems (DAIS) that, compared with local systems, have acquired qualitatively different features and capabilities [2, 3].

Among other things, the integrity of data stored on multiple storage media may be affected by various negative factors [4]. Such factors may include: errors and malfunctions [5] of computer equipment, software errors, operator errors caused by non-observance of guidelines and regulations. Errors in the operation of data storage devices may cause distortion and even loss of data, failure of individual or several network nodes and, in severe cases, of an entire distributed system. In such cases, significant resources and time may be required to restore corrupted data.

The use of information redundancy in distributed information systems is one of the efficient methods for ensuring high data integrity and dependability of such systems. Currently, information redundancy is widely used in the form of two types of data redundancy [3, 6]:

- online redundancy that consists in creating online backup (OB) data from a certain set of copies and/or historical data arrays that are used to improve the reliability of processing of incoming inquiries by the distributed system in the event of data errors or their partial loss during inquiry processing;

- recovery redundancy that consists in creating special recovery backup (RB) data that are used only for restoring real-time data if they are affected by corruption or errors.

Second, since the examined information systems have a distributed geospatial topology, two primary methods of storing the two types of backups can be used, i.e., centralized and decentralized. In case of centralized storage, the backup is located in a single, central node of the system. In case of decentralized storage, the data backup is located in several system nodes selected in accordance with a certain backup localization algorithm [6].

Third, if the number of system nodes is high, there are many options for backup allocation, which complicates the choice of the best configuration. That causes the requirement to define and solve the problem of selecting the best backup allocation.

Fourth, when looking for the best allocation of distributed backup across network nodes, various parameters of the network itself need to be take into consideration. Those include the bandwidth of communication channels, traffic and average message latency, cost of using computers and network channels, etc.

## Strategies for online backup and redundancy

Today, three online redundancy strategies are used to ensure the integrity of data that take into account the specifics of their use in information systems [3]:

**Strategy I.** According to this strategy, a backup is created and then used that consists of a certain number of copies of the permanent (rarely modified) data array. Processing of each inquiry to the array's data starts with the main array. If the array is corrupted, the inquiry is processed using the data from the first copy, and so on.

**Strategy II.** This strategy uses a backup that includes a certain number of historical versions of an array with frequently modified data. Array history $AP_i$ is its exact copy created at time $t_i$ $\left(i = \overline{1, N}\right)$ and the change log of the array's data that occurred within the time interval $(t_i+\Delta t)$. In case the main array is corrupted, it is restored using history $AP_N$. If the history is corrupted in the course of restoration, it is restored using history $AP_{(N-1)}$.

**Strategy III.** This strategy is mixed and restores the corrupted main array first by using copies of the array (according to strategy I), and, if all copies are corrupted, by using the backup from the history (according to strategy II).

The use of OB significantly increases the dependability of the distributed system when processing inquiries, but does not completely eliminate the possibility of the OB itself becoming corrupted. The recovery backup data (RB) is used for restoring a corrupted OB. There are two main options for using the RB [6]:

1) The first option is used in case of decentralized allocation of OB in several system nodes. If an OB is corrupted in a certain node of the system, it is restored using another uncorrupted instance of the OB located in the nearest node. In this case, this OB is used as an RB.

2) The second option involves using a special RB, the magnetic media archive (MMA). The MMA is used only for processing inquiries to restore a corrupted OB. The MMA may be hosted in a single network node or multiple instances of it can be hosted in multiple nodes.

The paper examines two strategies for restoring a corrupted OB, i.e., B-1 and B-2 that help significantly improve data integrity in distributed systems [7]. According to strategy B-1, all copies of data arrays that are required for

OB restoration are created sequentially based on RB data. The second strategy, B-2, differs from the first one in that, when obtaining the next copy, data is used not only from RB, but also all previously obtained copies of the data array to be restored.

## Operational dependability indicators of a DAIS that uses online backup for restoring corrupted data

Let us consider the primary indicators of operational dependability of a DAIS that, for the purpose of improving such dependability, uses only online backup with no MMA.

In terms of dependability, the operation of a DAIS, in whose nodes an online data backup is hosted, can be represented as a process of such system's transition within the space of possible states. System transitions from one state into another occur as a result of failures of system nodes that process incoming data inquiries and/or after the restoration of previously failed nodes. Thus, the state of a DAIS at any given time can be characterized by the number of failed and that of operable nodes.

When a certain system node processes a data inquiry, the OB of such node may become corrupted. As the result, the node becomes inoperable and no longer able to process incoming inquiries. The node's transition into such state will be considered a failure of the node. Since the system under consideration does not use recovery redundancy, the failed node will remain in this state. Let us assume that after a node fails, all incoming inquiries will be evenly distributed for processing among all still operable system nodes with an OB. In case all redundant nodes fail, the system will become unable to process incoming inquiries. Such state of a DAIS we will also interpret as system failure.

Let us denote by $M$ the number of DAIS nodes hosting online backup, and by $H$ the set of all DAIS states. Set $H$ consists of the following elements: $H_0$, all system nodes are operable, $H_m$, $m$-th node failure, $H_{mn}$, failure of nodes $m$ and $n$, $H_{1,2,...,M}$, all $M$ system nodes with an OB failed, system is inoperable.

Then, set $H$ of all system states and its power $|H|$ will be equal to:

$$H=\left\{H_0,H_1,\ldots,H_M,H_{1,2},\ldots,H_{1,2,\ldots,M}\right\},\ |H|=\sum_{i=0}^{M}C_M^i=2^M.$$

At any moment in time $t$, the system may be only in one state $\xi(t)=H(t)\in H$. Let us assume that the DAIS may remain in the initial state or transition into another state at regular time intervals. At the end of each such period of time, with a certain probability, the system either transitions into another state (one or more nodes failed simultaneously) or remains in the same state (none of the nodes failed). Such transitions between possible system states are called steps of a random process. We denote by $\xi(t)$, $t\geq 0$ the random value that describes the process of a system transitioning from one state into another.

Let us assume that at the moment of time $t$ the system is in state $\xi(t)$. Let us assume that, within a single time interval, node $j$ processes $\lambda_j(t)$ inquiries, provided that the system is in state $\xi(t)$.

Let us also assume that at the initial moment of time $t_0$ the system is fully operable and has no failed nodes. Let us denote by $\xi(t_0)=H(t_0)$ the initial operable system state at the moment of time $t_0$, and by $\lambda_j^0=\lambda_j(t_0)$ the number of inquiries that node $j$ is processing at the moment of time $t_0$.

After a certain period of system operation time operable, node $j$, at the moment of time $t$, will be processing $\lambda_j(t)$inquiries:

$$\lambda_j(t)=\lambda_j^0+M_P^{-1}(t)\sum_{i\in I_O(t)}\lambda_i^0. \quad (1)$$

In formula (1), $I_o(t)$ is the set of numbers of the system nodes that failed by time $t$, while $M_P(t)=M-\left|I_O(t)\right|$ is the number of system nodes hosting a backup that are operable at time $t$.

When processing a single inquiry in node $j$, a failure may occur with probability $Q_j$. Then, for a single time interval $(t, t+1)$, probability $\tau_j(t)$ of node $j$ failing and probability $\beta_j(t)$ of no failure will be, respectively, equal to:

$$\tau_{j(t)}=1-P_j^{\lambda_j(t)};\ \beta_j(t)=1-\tau_{j(t)}=P_j^{\lambda_j(t)};\ P_j=1-Q_j. \quad (2)$$

By sequentially numbering all the elements of set $H$ we obtain set $S$ of system states that consists of the same number of elements:

$$H=S=\left\{S_0,S_1,\ldots,S_M,S_{M+1},\ldots,S_N\right\},\ N=2^M.$$

The above system transition from one state into another is a homogeneous process, as the future state of the system does not depend on its previous transitions, but only on its current state [8, 9]. Then we can state that conditional probability $P\left\{\xi(t)=S_j/\xi(u)=S_i\right\}$that the system, at moment $t$, is in state $S_j$, provided that the system, at moment $u$, was in state $S_i$, will be equal to:

$$P\left\{\xi(t)=S_j/\xi(t_1)=S_{i_1},\ldots,\xi(t_n)=S_{i_n},\xi(t_u)=S_i\right\}=$$
$$=\left\{\xi(t)=S_j/\xi(t_u)=S_i\right\}=p_{ij}(t-u).$$

At the same time:

$u>t_n>\ldots>t_1;\ t>u;\ i,j\in\left\{0,1,\ldots,N\right\}$.

That means that conditional probability $P\left\{\xi(t)=S_j/\xi(u)=S_i\right\}$ does not depend on moments of time $t$ and $u$, but depends on distance $(t–u)$ between such moments. Therefore, such conditional probability depends on the time interval from moment $u$ to moment $t$.

Let us suppose that $p_{ij}(t–u)$ is the conditional probability of an event that corresponds to the transition of the system from state $S_i$ into state $S_j$ within a time interval equal to $(t–u)$. Let us assume that the system's transitions from one state into another occur within a single time unit. Then, the difference between the moments of time $t$ and $u$ will be equal to 1 $(t–u=1)$, while the conditional probability $p_{ij}(t-u)=p_{ij}(1)=p_{ij}$ is the transition probability of the system for states $S_i$ and $S_j$.

The values $p_{ij}$ of transition probabilities of the examined process will be calculated using the formula:

$$p_{ij} = \begin{cases} 0, \text{ if } (i < j) \text{ or when } \xi(t) = S_j \neq S_i = \xi(t-1) \\ \text{and } |I_0(t)| = |I_0(t-1)|; \\ \prod_{n \in R} \tau_n(S_i) \left[ \prod_{n \in R} \beta_n(S_i) \right]^{-1} \prod_{n \in I_p(S_i)} \beta_n S_i, \text{ if other wise.} \end{cases} \quad (3)$$

Formula (3) uses the following notations:

$I_0(t)$, set of numbers of system nodes that failed by time $t$.

$I_p(S_i)$, set of numbers of operable nodes of a system that is in state $S_i$;

$\tau_n(S_i)$, probability of failure of node $n$ per unit of time when the system is in state $S_i$;

$R=[I_0(S_i) - I_0(S_j)]$, set of numbers of the nodes that failed during the transition of the system between two states;

$I_p(S_i)$, set of numbers of the nodes that are operable in system state $S_i$.

$$\beta_n(S_i) = \tau_n(S_i).$$

In the examined process, the system transitions between various states can be formally represented as an oriented graph. The system states in the graph are represented by its vertices, while oriented arcs correspond to the system's transitions between states (vertices of the graph).

Fig. 1 shows an example of an oriented graph of a random system transition process. The system consists of $M$ = 2 nodes with multiple states: $S_0=H_0$; $S_1=H_1$; $S_2=H_2$; $S_3=H_{1,2}$; $\xi(t_0)=S_0$.

Since the failed nodes are not restored in this case, the system can be considered a non-restorable item that has a finite set of operable states and one state of complete failure [8, 9]. The process of system transition between the different states is an absorbing discrete-time Markov chain [9, 10].

Let us examine the following important indicators of system dependability: $T_1$, mean time to failure; $Q(t_0)$ and $Q(t,t+t_0)$, probability of system failure within time intervals $[0,t_0]$ and $[t,t+t_0]$; $P(t_0)$ and $P(t,t+t_0)$, probability of no failure within time intervals $[0,t_0]$ and $[t,t+t_0]$;

Let us deduce and analyse the above dependability indicators for the case of a system that operates based on a homogeneous directly connected network (for the situation of a heterogeneous network, the indicators are deduced and analysed similarly using formulas (1)–(3)).

Let us denote by $S=\{S_0, S_1, \ldots, S_N\}$ the set of all states of a system, whose set of nodes with online backup is equal to $N$. Let us denote by $S_j$ such system state, in which $j$ nodes with online backup have failed. Let the initial rate of inquiries processed by each node of the system in state $S_0$ be equal to $\lambda_0$. The the rate of inquiries processed by network nodes that are operable in system state $S_j$ will be denoted as $\lambda_j$. Then, in accordance with formula (1), we obtain the following formula for calculating $\lambda_j$:

$$\lambda_j = \lambda_0 + \lambda_j (N-j)^{-1} = \lambda_0 N (N-j)^{-1}. \quad (4)$$

The probability $\tau_j$ that, within a single time interval, one of the nodes of the network in state $S_j$ fails – taking into account formula (2) – will be calculated as follows:

$$\tau_j = 1 - p^{\lambda}. \quad (5)$$

The transition probabilities for the examined network, taking into account (3), will be calculated using the following formula:

$$p_{ij} = \begin{cases} 0, \text{ if } i < j; \\ C_{N-j}^{j-i} \tau_i^{j-i} \beta_i^{N-j}, \text{ if } 0 \le i \le j \le N. \end{cases} \quad (6)$$

Since the system does not use recovery redundancy, the failed node is not restored, and the system will eventually enter state $S_N$, in which all system nodes will be inoperable. Moreover, $p_{NN}$=1, since $S_N$ is an absorbing state.

Thus, as a result, we have the matrix $P=p_{ij}$ of probability of system transition between states, initial state of the system $S_0$ is known, the system has one absorbing state $S_N$ and a set $\{S_0, S_1..., S_{N-1}\}$ of operable states. Then, it can be affirmed that there is an absorbing discrete-time Markov chain. The set $S^1=\{S_0, S_1..., S_{N-1}\}$ of non-recurrent states is defined for it. I.e., the set of operable system states, in which not all nodes have failed. As well as a single-element set of absorbing states $S^2=\{S_N\}$ (when all system nodes are inoperable).

Since a Markov chain has a single absorbing state, it will eventually transition from the initial state into such absorbing state. Let us identify the mean number $n_{ij}$ of steps, after which the chain will be in one of the non-recurrent states $S_j \in S^1$ before absorption, provided that state $S_i$ was its initial state. Each step from state to state takes the system a unit time interval. Hence, value $n_{ij}$ can be considered the mean time the system spends in state $S_j$ before absorption, provided that $S_i$ was the initial state of the system. The initial state $S_i$ itself brings to value $n_{ij}$ a contribution equal to 1 if $i=j$ and 0 if otherwise, i.e.:

$$\delta_{ij} = \begin{cases} 1, \text{ if } i = j; \\ 0, \text{ if } i \neq j. \end{cases}$$

The chain enters state $S_m$ in one step from state $S_i$ with probability $p_{im}$. If we assume that $S_m \in S^2$, then the chain will never transition into state $S_j$. If $S_m \in S^1$, then, in the course of $n_{mj}$ steps, the chain will be in state $S_j$. Hence, we can write:

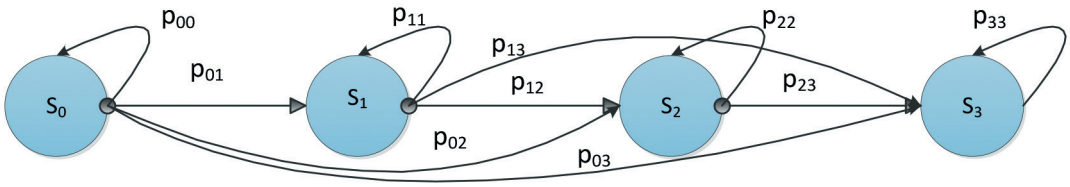



Fig. 1. Graph of the random transition process for a system of 2 nodes

$$n_{ij} = \delta_{ij} + \sum_{S_m \in S^1} p_{im} n_{mi}.$$

This equality in matrix form looks as follows:

$$\tilde{N} = I + \bar{Q}\tilde{N} \text{ or } \left(I - \bar{Q}\right)\tilde{N} = I.$$

In this formula, $I=\delta_{ij}$ is the identity matrix corresponding to the Kronecker delta of dimension $(N\times N)$; $\bar{Q}$ is a trix of dimension $(N\times N)$ that describes the chain's behaviour in the set of non-recurrent states $S^1$. It is derived from matrix $P=p_{ij}$ by removing the last column and the last row.

We will derive the fundamental matrix $\tilde{N}$ for the absorbing Markov chain by premultiplying by $\left(I - \bar{Q}\right)^{-1}$ both parts of the above equation [10]:

$$\tilde{N} = \left(I - \bar{Q}\right)^{-1}. \quad (7)$$

Let us identify the mean time $t_i$ of the chain being in the set of states $S^1$ by using matrix $\tilde{N}$, given that the initial state of the chain is state $S_i$. Obviously, $t_i=0$ if $S_i \in S^2$. Therefore:

$$t_i = \sum_{S_j \in S^1} n_{ij}; \ \ S_i \in S^1. \quad (8)$$

The validity of formula (8) follows from the fact that it is based on the premise that the time the chain remains in the set of states $S^1$ is equal to a sum of random variables. Or, in other words, it is equal to the sum of each of the individual times of the chain remaining in each of the non-recurrent states of set $S^1$. Moreover, the value of the mean sum of random variables is always equal to the sum of the mean values that make such sum [11].

As previously identified, $p_{ij}(n)$ is the probability of the system transitioning from state $S_i$ to state $S_j$ in $n$ steps. Then, taking into account the total probability formula, we will deduce that this probability is calculated using the formula:

$$p_{ij}(n) = \sum_{S_m \in S^1} p_{im} p_{mj}(n-1); \ \ p_{mj}(0) = \delta_{mj}.$$

The resulting formula in matrix form will be as follows: $P(n)=P^n$. In other words, the probability matrix of system transitions in $n$ steps is equal to the $n$-th power of the system's transition probability matrix.

Within the time interval from 0 to $t_0$, the system will complete $t_0$ steps, since, in a unit interval, the system completes one transition step. Then, given that $p_{0N}(n)=0$ if $n<N$, we deduce:

$$P(t_0) = 1 - \sum_{n=N}^{t_0} p_{0N}(n); \ Q(t_0) = \sum_{n=N}^{t_0} p_{0N}(n).$$

According to the conditional probability formula, probability of no failure $P(t,t+t_0)$ within the interval between $t$ and $(t+t_0)$ is defined as $P(t+t_0)=P(t+t_0)/P(t)$. It follows that the probability of system failure $Q(t+t_0)$ within the interval between $t$ and $(t+t_0)$ will be equal to $Q(t+t_0)=1-P(t+t_0)/P(t)$.

Using formula (8), let us identify the value of the system's mean time to failure $T_1$. Since, in our system, the initial state is $S_0$, while the absorbing state is $S_N$, the sought time $T_1$ is identified using formula:

$$T_1 = t_0 = \sum_{S_j \in S^1} n_{0j} = \sum_{j=0}^{N-1} n_{0j}.$$

If, for the examined Markov chain, matrix $\tilde{N} = n_{ij}$ is calculated using formula (7), we will deduce:

$$n_{ij} = \begin{cases} 0, \text{ if } j < i; \\ \prod_{m=i}^{j-1} p_{m,m+1} \left[ \prod_{n=i}^{j} (1 - p_{nn}) \right]^{-1}, \text{ if } i \le j. \end{cases}$$

Since, in this case, $n_{ii}=(1-p_{ii})^{-1}$, we will deduce that the system's mean time to failure $T_1$ is equal to:

$$T_1 = \prod_{j=0}^{N-1} \prod_{m=0}^{j-1} p_{m,m+1} \left[ \prod_{n=i}^{j} (1 - p_m) \right]^{-1}.$$

Let us assume that the system's parameters and the value of the unit time interval are such that the probability of an event consisting in a simultaneous failure of two or more system nodes is close to zero, i.e.:

$$p_{jj} + p_{j,j+1} \gg \sum_{n=2}^{N-j} p_{j,j+n}; \ \left(p_{jj+p_{j,j+1}} = 1\right). \quad (9)$$

Given that assumption, let us consider value $T_1$ of the system's mean time to failure. Fig. 2 shows the transition graph for the system that corresponds to the examined assumptions.

Having defined matrix $\tilde{N}$ using formula (7), we will deduce: $n_{jj}=1$ if $i>j$ and $n_{jj} = p_{j,j+1}^{-1}$ if $i \le j$. Then, the system's mean time to failure $T_1$ is equal to:

$$T_1 = \sum_{j=0}^{N-1} p_{j,j+1}^{-1} = \sum_{j=0}^{N-1} \left(1 - p_{jj}\right)^{-1}.$$

Let consider the probabilities $\rho_{jj}^{\gamma} \left(\gamma = I, II, II\right)$ of an event that consists in the fact that the system does not leave state $S_j$ within a single time interval. The system nodes use OB created in accordance with one of the three backup strategies $\left(\gamma = I, II, II\right)$. Let us prove that relation (10) is true for the examined probabilities

$$\rho_{jj}^{I} > \rho_{jj}^{III} > \rho_{jj}^{II}. \quad (10)$$

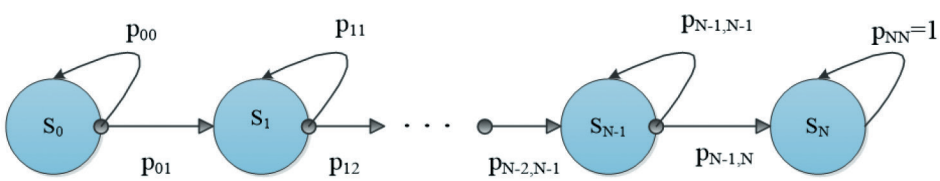



Fig. 2. Graph of a random transition process of a system in a set of states $S$

Taking into account formulas (4)–(6), we deduce:

$$\rho_{jj} = \beta_j^{(N-j)} = p^{\lambda N};\ j = 0,1,\ldots,(N-1). \qquad (11)$$

In [7], it was proved that, if OB was created using the three online backup strategies $(\gamma = I, II, II)$, for the probabilities $P^{\gamma}$ of successful processing of inquiries (if the OB of the node, in which the inquiry is processed, is not corrupted), the following formula is true:

$$P^{I} > P^{III} > P^{II}.$$

Then, both formula (10), and the expression $T_1^{\gamma}$ for the mean time to system failure is true as well:

$$T_1^{I} > T_1^{III} > T_1^{II}.$$

The above findings can be formulated as the following statement.

**Statement 1.** Using backup strategy I for creating OB in distributed systems that do not use recovery redundancy enables the longest mean time to failure compared to the other online redundancy strategies (strategies II, III).

Value $P(t_0)$ of the systems' probability of no failure within time interval $[0,t_0]$, taking into account formula (9) for the examined configuration of distributed system, i.e., with no recovery redundancy, will be equal to:

$$P(t_0) = 1 - p_{0N}(N)\sum_{n=0}^{t_0-N} B^n;$$

$$\text{где: } p_{0N}(N) = \prod_{i=0}^{N-1} p_{i,i+1} = \left(1 - p^{\lambda N}\right)^N;$$

$$B = \sum_{i=0}^{N-1} p_{ii} = Np^{\lambda N}.$$

Moreover, $P(t_0)=0$ if $t_0<N$.

## Indicators of operational dependability of a DAIS that uses magnetic media archives for restoring corrupted data

Let us examine the operational dependability indicators of a DAIS that uses recovery redundancy based on magnetic media archives.

A magnetic media archive is a special set or several sets of a certain number of copies and/or histories of data arrays. MMA is stored in one of the system's nodes (centralized archive) or in several nodes in the case of decentralized storage of several identical copies of magnetic media archives. [12]. MMA is used exclusively for restoring OB that has been corrupted in one or more nodes of a distributed system, thus improving the system's dependability.

Let us assume that, when a node with an MMA processes an inquiry for restoring a corrupted online backup, with certain probability, the node with the MMA itself may fail. Taking into account this possibility, let us analyse the operational dependability indicators of the DAIS that uses recovery redundancy in the form of magnetic media archives that themselves may be in a state of failure.

When processing a data inquiry in a node with an OB, the latter may become corrupted resulting in the failure of such node. The operability of the failed node is restored using one of two restoration strategies: B-1 or B-2 using MMA.

A failure of an entire DAIS system will be understood as such system state, whereas all system nodes with an OB and all MMA have failed.

In the state of failure, a DAIS is unable to process incoming data inquiries or restore the operability of nodes with an OB due to the failure of all MMA.

Let us assume that the following assumptions are true: 1) inquiries arriving to a failed node with an OB are not redirected to operable nodes and are not processed until the node has been restored; 2) should a node with an MMA fail, it is not restored; 3) all inquiries for restoring nodes with corrupted OB arriving to the failed node with an MMA are evenly distributed and redirected to other operable nodes with an MMA; 4) inquiries for restoring failed nodes with an OB are evenly distributed among all operable nodes with an MMA.

To describe the operation of such DAIS, we will use a discrete-time homogeneous absorbing Markov chain. Let us assume that the system parameters are such that the probability of failure of more than one node with an OB or more than one MMA over a unit time interval of system operation is close to zero. Given that assumption, let us define set $H$ of information system states $H = \{H_{m,n}\}$, $m = \overline{0,M}, n = \overline{0,N}$.

State $H_{m,n}$ corresponds to a state of the DAIS, in which $m$ magnetic media archives and $n$ nodes with an OB are in

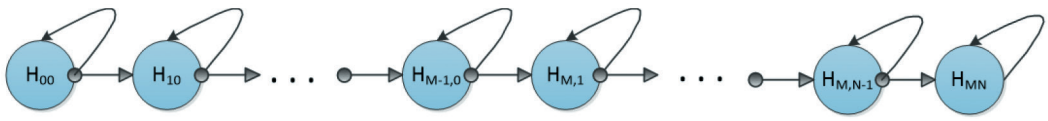



Fig. 3a. Transition graph of a distributed system in a set $H$ of possible states

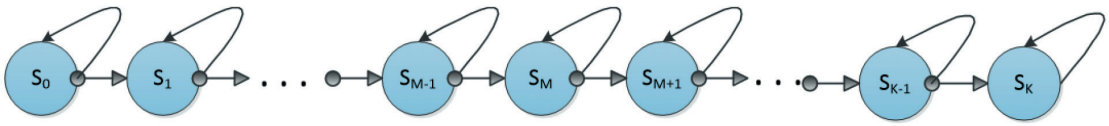



Fig. 3b. Transition graph of a distributed system in a set $S$ of possible states

a state of failure. Let us suppose that, over a unit interval of operation (one-step of the Markov process), the system may transition from state $H_{m,n}$ into state $H_{m+1,0}$ if $n = 0$. While if $m = M$, the system transitions to state $H_{M,n+1}$ or may remain in the initial state. Given the above assumptions, the graph of system transitions will be as shown in Fig. 3a.

Based on set of states $H$, let us construct set of states $S = \{S_i\}, i = \overline{1,K}$, $K=M+N-1$, where $S$: $S = \{S_i,\ (i = \overline{1,K}, k = M+N-1)\}$. Element $S_i$ of set $S$ if $0 \le i < M$ is associated with state $H_{i,0}$, while if $i=M$ state $H_{M,1}$ and if $M<i\le K$ state $H_{M,i}$. The constructed set $S$ is associated with the system transition graph shown in Fig. 3b.

Let us assume that, with probability $p_{ij}$, the system can, in one step, transition from state $S_i$ into state $S_j$. Given the above assumptions, it can be affirmed that:

$$p_{ij} \neq 0 \text{ if } i = j \text{ or}$$
$$j = i+1,\ p_{kk}=1 \text{ and } p_{ii} + p_{i,i+1} = 1 \text{ if } 0 \le i \le K-1.$$

Let us assume that, within a unit interval, λ data inquiries arrive to each of $N$ system nodes with an online backup. Each node, while a single inquiry is processed, may fail with probability $Q=1-P$ or successfully process it with probability $P$. Since we have assumed that all inquiries for restoring nodes with an OB are evenly distributed among operable nodes with an MMA, then, if the system is in state $S_0$, each node with an MMA receives $\mu_0$ restoration inquiries over a unit interval of system operation time:

$$\mu_0 = \lambda Q N M^{-1}.$$

On the other hand, in the course of OB restoration inquiry processing, the MMA itself may become corrupted with probability $Q_A=1-P_A$. Here, $P_A$ is the probability of successful processing in a node with an MMA of an OB restoration inquiry. Should a node an MMA fail, the OB restoration inquiries it received start being evenly distributed among the still operable nodes with an MMA.

Let us suppose that the system is in state $S_i \left(i = \overline{1, M-1}\right)$, then, within a unit interval, each node with an MMA receives $\mu_i$ inquiries for restoring corrupted online backups in the network nodes:

$$\mu_i = \mu_0 + i\mu_0 \left(M-i\right)^{-1} = \mu_0 M \left(M-i\right)^{-1}.$$

Then, within a unit interval, an operable MMA in state $S_i$, may fail with probability $\varphi_i$. Probability $\varphi_i$ is equal to $\varphi_i = 1 - P_A^{\mu_i}$. On the other hand, a node with an MMA can successfully process an inquiry to restore the OB with probability $\Psi_i = 1 - \varphi_i$. The transition of the system in one step from state $S_i$ into state $S_{i+1} \left(i = \overline{1 = M-1}\right)$ occurs with probability $p_{i,i+1}$ that is calculated using the following formula:

$$p_{i,i+1} = 1 - \Psi_i^{M-i} = 1 - P_A^{\mu_0 M} = 1 - P_A^{\lambda Q N};$$
$$p_{ii} = 1 - p_{i,i+1} = P_A^{\mu_0 M}. \qquad (12)$$

Taking into account the findings of the above paragraph, we obtain that the value of probability $p_{i,i+1}$, if $i = M, \left(K-1\right)$, is equal to:

$$p_{i,i+1} = 1 - p^{\lambda N};\ \ p_{ii} = p^{\lambda N}. \qquad (13)$$

Let us assume that, should all nodes with an MMA be corrupted ($i \ge M$), the failed node with an OB is not restored. All data inquiries received by such node are evenly distributed among the still operable nodes with an online backup.

For the examined distributed system configuration, in accordance with formula (7) and taking into account formulas (12) and (13), we obtain the fundamental matrix $\tilde{N} = n_{ij}$ of a Markov chain, in which its element is equal to:

$$n_{ij} = \begin{cases} 0, \text{ if } i > j; \\ p_{j,j+1}^{-1}, \text{ if } i \le j. \end{cases}$$

Next, let us identify the operational dependability indicators of the examined system that uses recovery redundancy in the form of undependable MMA. Nodes with MMA may fail when processing inquiries for restoring a corrupted OB in system nodes. Such distributed system may be considered as a non-restorable item.

Let us assume that the system is in initial state $S_0$. Then, it can be asserted that the distributed system's mean time to failure $T_1$ is equal to the mean time $T_1$ the system will spend in the set of non-recurrent states. The formula for calculating time $T_1$ is set forth below:

$$T_1 = \sum_{j=0}^{K-1} n_{0j} = \sum_{j=0}^{K-1} p_{j,j+1}^{(K-1)}$$

Taking into account formulas (12) and (13), the formula for calculating $T_1$ is transformed as follows:

$$T_1 = \sum_{j=0}^{M-1} \left[1 - P_A^{\lambda Q N}\right]^{-1} + \sum_{j=0}^{K-1} \left(1 - P^{\lambda N}\right)^{-1}$$

In [7], for the two restoration strategies B-1 and B-2, inequality $P_A^{B-1} < P_A^{B-2}$ was proved, out of which follows that the following similar inequality for the mean time to failure is true:

$$T_1^{B-1} < T_1^{B-2}.$$

Out of that inequality follows that the following statement is true.

**Statement 2.** Recovery strategy B-2 in distributed systems that use MMA ensures a mean time to failure greater than recovery strategy B-1.

In [7], it was proved that if OB is created using three backup strategies $\left(\gamma = I, II, II\right)$, for probabilities $P^{\gamma}$ of successful inquiry processing (if the OB of the node, in which the inquiry is processed, is not corrupted) the below formula is true.

$$P^{I} > P^{III} > P^{II}.$$

If OB in the nodes of a distributed system is created using one of the three backup strategies with redundancy parameters $(x_i > 1, 0 < q_i < 1/2)$, then, in accordance with the findings of [7], for probability ($P$=1–$Q$) of successful processing of data inquiry in node $i$, the following formula is true:

$$P_i^I(x_i) > P_i^{III}(x_i) > P_i^{II}(x_i).$$

Taking this formula into account, the validity of the following inequality is proven:

$$T_1^I > T_1^{III} > T_1^{II}.$$

Let us formulate the findings in the form of the following statement.

**Statement 3.** Applying strategy I of online backup in distributed information systems enables a mean time to failure greater than that ensured by strategies II and III of online backup.

Let us consider the probability $P(t_0)$ of no failure and the probability $Q(t_0)$ of failure of a distributed system within the time interval $[0,t_0]$.

Based on the earlier findings, we deduce:

$$P(t_0) = 1 - \sum_{n=K}^{t_0} p_{0,K}(n) = 1 - p_{0,K}(K) \sum_{m=0}^{t_0-K} B^m,$$

where:

$$p_{0,K}(K) = \prod_{i=0}^{K-1} p_{i,i+1} = \left[1 - P_A^{\mu_0 M}\right]^M \left[1 - P^{\lambda N}\right]^{(N-1)};$$

$$B = \sum_{i=0}^{K-1} p_{ii} = MP_A^{\mu_0 M} + (N-1)P^{\lambda N}.$$

Within the time interval $[0,t_0]$, the system will fail with probability $Q(t_0)$=1–$P(t_0)$. For time interval $[t,t+t_0]$, the values of probability $P$ of no failure and probability $Q$ of system failure will be calculated using the following formulas:

$$P(t, t+t_0) = P(t+t_0)/P(t);$$

$$Q(t, t+t_0) = 1 - P(t, t+t_0).$$

## Conclusion

The paper examines methods for improving the operational dependability of distributed automated information systems by means of information redundancy. An analysis is made of the efficiency of the online backup strategies in nodes of a distributed system and strategies of restoring a corrupted OB. The paper analyses the effect of online and recovery redundancy strategies on such indicators of DAIS operational dependability as the mean time to failure, probability of system failure and probability of no failure within a given time interval. A number of statements regarding the efficiency of the examined strategies in terms of the time of DAIS time to failure were substantiated.

The findings referred to in this paper can be used at the stages of design, development and operation of DAIS of various classes and purposes. These findings may be of particular relevance for such large-scale geographically distributed multi-level automated systems as railway ACS-class systems. For such systems, the problems of ensuring the operational dependability and data integrity become of particular importance and relevance.

## References

1. Shestiukov O.S. [Automated system for tracking, elimination supervision of technical failures, dependability analysis]. [Global Trends in Science, Education, Technology: Proceedings of the International Research and Practice Conference]. Belgorod (Russia): Agetstvo perspektivnykh nauchnykh issledovaniy; 2021. P. 42-46. [accessed: 05.02.2022]. Available at: https://apni.ru/article/2543-avtomatizirovannaya-sistema-uchyota-kontrolya. (in Russ.)

2. Shubinsky I.B. [Dependable failsafe information systems. Synthesis methods]. Moscow: Dependability Journal; 2016. (in Russ.)

3. Mikrin E.A., Kulba V.V., editors. [Information support of managerial control systems (theoretical foundations). In 3 parts]. Moscow: Izdatelstvo fiziko-matematicheskoy literatury; 2012. (in Russ.)

4. [Data recovery]. [accessed: 10.09.2021]. Available at: http://www.datarecovery.ru/datarecovery.htm. (in Russ.).

5. Shubinsky I.B, Schäbe H. Errors, faults and failures. *Dependability* 2021;2:24-27.

6. Somov S.K. [Information integrity in distributed data processing systems]. Moscow: ICS RAS; 2009. (in Russ.)

7. Kulba V.V., Somov S.K., Shelkov A.B. [Data redundancy in computer networks]. Kazan: Kazan University Publishing; 1987. (in Russ.)

8. Tikhonov V.I., Shakhtarin B.I., Sizykh V.V. [Random processes. Examples and problems. Volume 1, Random values and processes. Study guide for higher education]. Moscow: Goriachaya liniya – Telekom; 2014. (in Russ.)

9. Rozanov Yu.A. [Random processes]. Moscow: Nauka; 1979. (in Russ.)

10. Kemeny J., Snell J. Finite Markov chains. Moscow: Mir; 1970.

11. Kovalenko I.N., Filippova A.A. [Probability theory and mathematical statistics]. Moscow: Vysshaya shkola; 1978. (in Russ.)

12. Mikrin E.A., Somov S.K. [Analysis of the efficiency of strategies of information recovery in distributed data processing systems]. *Informatsionnye tekhnologii i vychislitelnye sistemy* 2016;3:5-19. (in Russ.)

## About the authors

**Vladimir V. Kulba**, Honoured Science Worker of the Russian Federation, Doctor of Engineering, Professor, Head Researcher, V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences. Address: 65 Profsoyuznaya St., Moscow, 117997, Russian Federation, e-mail: kulba@ipu.ru.

**Sergey K. Somov**, Candidate of Engineering, Senior Researcher, V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences. Address: 65 Profsoyuznaya St., Moscow, 117997, Russian Federation, e-mail: ssomov2016@ipu.ru.

**Alexey B. Shelkov**, Candidate of Engineering, Lead Researcher, V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences. Address: 65 Profsoyuznaya St., Moscow, 117997, Russian Federation, e-mail: shelkov@ipu.ru.

## The authors' contribution

**Kulba V.V.** Authored the idea of using information redundancy for ensuring data integrity and operational dependability of information systems. Overall leadership and participation in the preparation of the paper.

**Somov S.K.** Problem definition. Developed the formal model of an information system that uses online redundancy and magnetic media archives. Analysed the model, proved the assertions set forth in the paper.

**Shelkov A.B.** Participated in the preparation of the Introduction and the formal system model. Advisory on the specifics of the information systems employed by JSC RZD.

## Conflict of interests

The authors declare the absence of a conflict of interests.

*Original article*
*https://doi.org/10.21683/1729-2646-2022-22-1-13-19*

# Optimizing the timeframe of scheduled repairs using valuation techniques

**Sergey A. Smolyak**, *Central Economics and Mathematics Institute, Russian Academy of Sciences, Moscow, Russian Federation,*
*smolyak1@yandex.ru*

*Sergey A. Smolyak*

**Abstract. Aim.** *The paper examines technical systems (machinery and equipment), whose condition deteriorates in the course of operation, yet can be improved through repairs (overhaul). The items are subject to random failures. After another failure, an item can be repaired or disposed of. A new or repaired item is to be assigned the date of the next scheduled repairs. Regarding a failed item, the decision is to be taken as to unscheduled repairs or disposal. We are solving the problem of optimization of such repair policy. At the same time, it proves to be important to take into consideration the effect of repairs, first, on the choice of appropriate indicators of item condition that define its primary operational characteristics, and second, on a sufficiently adequate description of the dynamics of items' performance indicators.* **Methods.** *Assigning the timeframe of scheduled repairs normally involves the construction of economic and mathematical optimization models that are the subject matter of a vast number of publications. They use various optimality criteria, i.e., probability of no failure over a given period of time, average repair costs per service life or per unit of time, etc. However, criteria of this kind do not take into account the performance dynamics of degrading items and do not fully meet the business interests of the item owners. The criterion of maximum expected total discounted benefits is more adequate in such cases. It is adopted in the theory of investment projects efficiency estimation and the cost estimation theory and is, ultimately, focused on maximizing a company's value. The model's formulas associate the item's benefit stream with its primary characteristics (hazard of failure, operating costs, performance), which, in turn, depend on the item's condition. The condition of non-repairable items is usually characterized by their age (operating time). Yet the characteristics of repairable items change significantly after repairs, and, in recent years, their dynamics have been described by various models using Kijima's virtual age indicator (a similar indicator of effective age has long been used in the valuation of buildings, machinery and equipment). That allows associating the characteristics of items in the first and subsequent inter-repair cycles. However, analysis shows that this indicator does not allow taking into consideration the incurable physical deterioration of repaired items. The paper suggests a different approach to describing the condition of such objects that does not have the above shortcoming.* **Conclusions.** *The author constructed and analysed an economic and mathematical model for repair policy optimisation that is focused on maximizing the market value of the company that owns the item. It is suggested describing the condition of an item with two indicators, i.e., the age at the beginning of the current inter-repair cycle and time of operation within the current cycle. It proves to be possible to simplify the dependence of an item's characteristics on its condition by using the general idea of Kijima models, but more adequately taking into consideration the incurable physical deterioration of such item. The author conducted experimental calculations that show a reduction of the duration of planned repairs as machinery ages at the beginning of an inter-repair cycle. Some well-known repair policies were critically evaluated.*

**Keywords:** *service life of a technical system, repair policy, optimization criterion, degradation, Kijima models, cost estimation, revenue approach.*

**For citation:** *Smolyak S.A. Optimizing the timeframe of scheduled repairs using valuation techniques. Dependability 2022;1: 13-19. https://doi.org/10.21683/1729-2646-2022-22-1-13-19*

**Received on:** *14.10.2021 /* **Upon revision:** *10.02.2022 /* **For printing:** *18.03.2022.*

## Basic definitions. Problem definition

The paper examines mass-produced and marketed technical systems (TS, i.e., machines and equipment) that are operated by companies, can be repaired and are subject to failures. We will call their state at the moment of release *new*. All TS that are identical in the new state are combined into a single *brand*. The paper examines the process of operation of single-brand TS. TS is useful for *market players* and therefore, according to valuation standards [1], has a certain *market value* (MV) that depends on the state of the TS. In the course of operation, the state of TS deteriorates due to physical wear (degradation). Possible failure of TS causes losses for the company. In case of failure, a TS is to be either disposed of, or submitted to emergency repairs. The efficiency of TS operation can be improved by assigning scheduled repairs (overhauls), as well as modifying its service life (i.e., the time of disposal). The MV of the disposed TS is called *salvage value* and is usually defined as the value of the elements (components, parts, scrap metal) suitable for further use less the cost of dismantling the TS and transporting its elements. Usually, this value is not high and we (for the purpose of simplification) will deem it equal to zero. The work performed by the TS is also useful for the market players and has a certain market value.

The *benefits* from the use of TS within a certain period are defined as the MV of the deliverables less the costs incurred within such period. Accordingly, the benefits of TS operation are equal to the MV of the performed work less the operating costs (that, among other things, include the cost of maintenance and scheduled repairs), while the benefits of TS disposal are zero.

Repairs improve the state of TS by eliminating some of the effects of physical wear. That is a case of *curable* deterioration. However, other effects accumulate and, eventually, may cause TS failure. Such deterioration is called *incurable* [1, 2]. The service life of TS is divided by repairs into inter-repair cycles (IRC). The *assigned duration* of IRC is the period, at the end of which the TS that has not failed earlier within this cycle is to be disposed of or repaired.

A *repair policy* is understood as the rule for assigning IRC durations and the rule for choosing a solution regarding a TS at the end of an IRC (that failed or reached the end of a designated period). We are solving the problem associated with the development of an optimal repair policy for single-brand TS. For that purpose, an optimality criterion is to be defined and the variation of TS characteristics in the course of operation is to be described. At the same time, almost until the end of the paper, we will assume the absence of inflation.

## Optimality criterion

The numerous works on the dependability theory used various optimality criteria, e.g., the life-average number or cost of repairs, total discounted costs [3] or equivalent annuity [4, 5], ratio of life-average costs to the average service life [6]. However, as it is correctly noted in [7], the optimality criteria were normally chosen without proper substantiation, out of real business context.

If we consider the acquisition and use of TS as an investment project, the optimal repair policy is to comply with its best version, the one that provides the highest expected net present value (NPV) [8; 9]. A similar criterion is also used in property valuation. Here, the *market value* (MV) is considered the primary type of value. This concept is defined and commented in the valuation standards [1], and we will not repeat that here. Let us just note that the MV of a valuation item at a certain date (valuation date) reflects both the price of the item in the transaction made on the valuation date between independent and economically rational market players under certain conditions (specified in the valuation standards) and the contribution of the item into the MV of the company that owns it. Three approaches are used for determining an item's MV.

In the comparative (market) approach, an item's MV is estimated based on the prices of the deals concluded on the valuation date with similar items.

In the cost-based approach, an item's MV is estimated based on the costs required for its creation. This approach is primarily used for evaluating buildings and structures, but its applicability is limited by evaluation standards and, in general, appears to be controversial, especially as regards machinery and equipment [10, 11].

The income-based approach is based on the principle of expected benefits that is mentioned, but not detailed in [1]. We will use the following definition [1, 2].

**The MV of the assessed item at the date of valuation is equal to the expected amount of discounted benefits from its use within the projection period and the item's MV at the end of the period, if the item is used most efficiently, and not less than the above sum if otherwise. The end of the projection period can be chosen arbitrarily, as the item's MV does not depend on it.**

A number of important comments should be made on this definition.

1. The term "expected", in the context of probabilistic uncertainty, is understood as the expectation (in [1], "weighted by probability"). In the following formulas, it is denoted as **E**.

2. Adding the item's MV at the end of the period can also be interpreted as benefits from the (virtual) sale of the item at the current MV. According to this interpretation, the most efficient use of an item may also include its sale at the MV at some point in time.

3. According to [1], the benefits are to be discounted at the after- or pre-tax rate, depending on whether the income tax was included in the cost. We assume the second option and discount the benefits at the pre-tax rate $r$.

As it can be seen, the item's MV reflects the maximum amount of the expected total discounted benefits (ETDB) that corresponds to its most efficient use. In this context, the ETDB from the use of TS is to be the criterion of optimal repair policy, which contributes to the growth of the company value. It is difficult to directly apply the principle of

expected benefits to the optimization of the repair policy, as the MV of the work performed by the TS (denoted as *B*) is usually unknown. The following considerations help solve the problem. Let us consider a TS at the beginning of its use (moment in time 0). Its MV *K* is a known value that reflects the costs of acquisition, delivery and installation. Let P be a certain repair policy. With this policy, at time *t*, TS performance *Q*(t) and the rate of operating costs *C*(t) will be random functions of time. The moments $s_1, s_2,\ldots$ of repairs (and, in general, the cost of such repairs $R_1, R_2,\ldots$) will also be random.

The expected amount of discounted benefits of policy P, in this case, will be

$$B_{\Sigma}(\Pi) = \mathrm{E}\left\{\int_0^{\infty}\left[BQ(t) - C(t)\right]e^{-rt}dt + \sum_i R_i e^{-rs_i}\right\}.$$

However, due to the principle of expected benefits, this value is not greater than the TS market value *K* and is identical to *K* in case of an optimal policy. Out of that easily follows that

$$B \geq \frac{K + \mathrm{E}\left\{\int_0^{\infty} C(t)e^{-rt}dt + \sum_i R_i e^{-rs_i}\right\}}{\mathrm{E}\left\{\int_0^{\infty} Q(t)e^{-rt}dt\right\}},$$

with equality when the most efficient policy P is used.

It follows that an optimal repair policy is to ensure minimal expected unit costs (EUC), i.e., the ratio of the expected discounted costs for the purchase, operation and repair of the TS to the expected discounted scope of work that it performs. For the deterministic case, a similar criterion was proposed in [1-3] and practically used in the development of depreciation rates of construction machines. However, since in our case the repair policy cannot be defined with a finite number of scalar parameters, it proves to be difficult to optimize it according to the EUC criterion. Further on, we will suggest a more convenient solution of this problem that is based on the same idea.

## Characterization of the state of TS under repair

Kijima has suggested [14] characterizing the state of repaired items by *virtual age* (VA) that increases synchronously with the chronological age, but after repairs, rapidly decreases proportionally to the item's VA before the repairs (model I) or the duration of the previous IRC (model II). If proportionality coefficient $\beta = 1$ or 0, the item's state after the repairs becomes either new, or same as before the repairs. We consider both of these cases unrealistic, and assume that $0 < \beta < 1$.

Meanwhile, the basic idea of describing the state of a TS with a single indicator was proposed much earlier. Thus, in [15], it was stated that some valuators use the *effective age* (EA) for appraising the value of buildings. This indicator reflects the age of a typically used similar building that is in the same state as the one being evaluated. Since the 1950s, the concept of EA has been used, first in the US, then in other countries for the purpose of valuating buildings, machinery and equipment. Initially, the valuators assessed the EV of items using expert methods. Later, more substantiated methods and tables were developed, of which neither Kijima, nor his followers were apparently aware. In this context, Kijima models can be considered an application of the EV concept in dependability. These models have been studied by many authors (e.g., in [16]) and used for solving practical problems.

However, the virtual age and similar indicators cannot adequately describe the state of a repaired TS. Indeed, otherwise, after the first repairs, when the virtual age of TS decreases, it will be in the same state as it was at some point in time within the first IRC. But then it is to further be used in the same way, i.e., work until failure or until the assigned time of the first repairs, etc. In this case, its service life will prove to be infinite, which is impossible for the TS exposed to *incurable* wear. At the same time, the state of regular TS can be adequately described by *two* indicators, i.e., age *s* at the beginning of the current IRC and the time of operation within this cycle *t* [17, 18]. Then, the dependence of TS characteristics on its state will have to be described by functions of two variables. It turns out that they can be simplified using Kijima's idea.

Let us take a certain TS operational characteristic (e.g., performance). We will denote its value for the TS in state *(s,t)* as *Z(s,t)* and assume that *z(t)=Z(0,t)*. As with the Kijima model I, we will assume that the characteristic of a TS that underwent the first repairs at age *s* becomes the same as that of a TS of a smaller age of $\beta s$: $Z(s,0) = Z(0,\beta s) = z(\beta s)$. But the incurable wear of the first TS is greater than that of the second one, therefore, further on, its characteristic will deteriorate faster, and the faster the greater the age difference. Generally speaking, in its regard, the time will as if "accelerate" by a certain rate $k(s) > 1$ times, i.e., after time *t*, it will be $z(\beta s + k(s)t)$. If the first TS, after repairs at age *s*, undergoes second repairs after time *s'*, then, in the next IRC, for it, time should "accelerate" by another *k(s')* times, i.e., by *k(s)k(s')* times compared to the second TS. It is logical to assume that the result of the second repairs will be the same as that of the second TS that underwent repairs at the same age $s + s'$, which means an "acceleration" of $k(s + s') > 1$ times. But then $k(s + s') = k(s)k(s')$, and that is only possible if $k(s) = \gamma^s$, where $\gamma > 1$ is the "degradation acceleration" coefficient. In such case, the characteristic of a TS that underwent repairs at age *s* will be $Z(s,0) = z(\beta s)$ after the repairs, while after more time *t* it will be $Z(s,t) = z(\beta s + t\gamma^s)$. Such model that is applicable to any characteristics of a TS can be called a modified Kijima model. It appears that it more adequately describes the dynamics of the characteristics of repaired TS. A similar model of geometrical process, in which "degradation acceleration" is associated with the *ordinal number of the IRC*, was proposed in [19] and subsequently examined by many authors.

## Optimization model

Let us first find out how the value of a TS varies within a single IRC. We will characterize each IRC not by its ordinal number, as it is usually done, but by the age of the TS at the beginning of the cycle. Let us introduce the following designations: $M_s$ is the IRC, at the beginning of which the TS has the age of $s$, $T_s$ is its designated duration, $B$ is the MV of the work performed by an operable TS within a small unit of time, $R$ is the cost of TS repairs (we deem it to be identical for scheduled and emergency repairs), $L$ is the company's losses caused by TS failure, $Q(s,t)$ is the TS performance in state $(s,t)$, $C(s,t)$ is the rate of its operational costs, $\lambda(s,t)$ is the hazard of TS failure, $\Lambda(s,t)=\int_0^t \lambda(s,x)\,dx$ is the mean number of failures over time $t$ within cycle $M_s$.

We will assume that function $Q(s,t)$ is non-increasing, while functions $\lambda(s,t)$ and $C(s,t)$ are non-decreasing with respect to their arguments, while at least one of them grows indefinitely if $s\to\infty$ and $t\to\infty$. The time of disposal and repair is considered negligible.

Let us denote the cost of a TS in state $(s,0)$ as $f(s)=V(s,0)$. Let us evaluate function $f(s)$ from above. In order to do that, let us note that the TS in state $(s,t)$ brings benefits with the rate of $BQ(s,t)$ - $C(s,t)$. In particular, a TS at the beginning of cycle $M_s$ brings benefits with the rate of $B_0=BQ(s,0)$ - $C(s,0)$ and has the hazard of failure $\lambda(s,0)$. Then, those characteristics deteriorate until the TS enters the next IRC or is disposed of. It can be seen that the cost of the TS is not greater than the MV $W(s)$ of a virtual item that always provides benefits with the rate $B_0$, whose failures occur with a constant rate $\lambda(s,0)$ and do not cause losses. But such item, within the short time $dt$, fails with the probability $\lambda(s,0)\,dt$, therefore, requiring expected repair costs $\lambda(s,0)Rdt$ and providing expected benefits $[B_0 - \lambda(s,0)R]dt$. That is why the ETDB from its use over an infinite service life is equal to $[B_0 - \lambda(s,0)R]/r$. If this value is positive, it is identical to the MV of the virtual item $W(s)$, otherwise using such item is inefficient and it has $W(s)=0$. Noting that the TS at the beginning of the cycle $M_s$ has a MV $f(s)$ that does not exceed $W(s)$, we obtain: $f(s)\le W(s)=\max\{[BQ(s,0)-C(s,0)-\lambda(s,0)R]/r;\,0\}$. But if $s\to\infty$, at least one of the functions $C(s,0)$ and $\lambda(s,0)$ increases indefinitely, therefore if $s$ is sufficiently large $f(s)=0$.

Let $g(x)$ be the cost of a TS with the age of $x$, that needs to be disposed of or repaired. It corresponds to the total benefits from the best possible further use of such TS. But disposing of the TS provides zero benefits, while repairs require costs $R$ and put the TS at the beginning of the next cycle, i.e., state $(x,0)$, where it will have a MV $f(x)$. Therefore,

$$g(x)=\max\left[f(x)-R;0\right]. \tag{4}$$

Cycles $M_s$, in which $f(s)>0=g(s+T_s)$, will be called *terminal*. In them, using a TS for its intended purpose is efficient, but at the end of the cycle it should be disposed of.

In this situation, the repair policy consists in assigning for each cycle $M_s$ a duration $T_s$ and specifying, which of them are terminal.

Let us assume that for cycle $M_s$ duration $T$ was assigned. Let us take a TS at the beginning of this cycle and find the expected sum $G(s,T)$ of discounted (at the beginning of the cycle) benefits from its use in cycle $M_s$ (including the cost of the TS at the end of the cycle).

Note that the duration of cycle $M_s$ is random. With probability $e^{-\Lambda(s,T)}$, it is equal to $T$, while with probability $\lambda(s,x)e^{-\Lambda(s,x)}dx$, it lies within the interval $(x, x+dx)$ if $x<T$.

In the first case, there will be no loss from failure, and at the end of the cycle, the TS will have an age of $s+T$ and PC $g(s+T)$. In the second case, the TS fails having operated for time $x$, i.e., at the age of $s+x$. Its MV will be $g(s+x)$ and there will be failure-related losses L.

The benefits of using the TS for its intended purpose at time $x$ after the start of the cycle, i.e., in state $(s,x)$, for the small period $dx$ are $[BQ(s,x)-C(s,x)]dx$. However, the TS will provide them only if it does not fail during time $x$ of its operation within the cycle, i.e., with probability $e^{-\Lambda(s,x)}$.

Now, taking into account the amount of possible benefits and their probabilities, we find:

$$\begin{aligned}G(s,T)&=e^{-rT}\cdot e^{-\Lambda(s,T)}g(s+T)+\\&+\int_0^T e^{-rx}\cdot\left[g(s+x)-L\right]\lambda(s,x)e^{-\Lambda(s,x)}dx+\\&+\int_0^T e^{-rx}\cdot e^{-\Lambda(s,x)}\left[BQ(s,x)-C(s,x)\right]dx=\\&=e^{-N(s,T)}g(s+T)+\int_0^T e^{-N(s,x)}H(s,x)\,dx,\end{aligned} \tag{5}$$

where

$$\begin{aligned}N(s,x)&=rx+\Lambda(s,x);\\H(s,x)&=BQ(s,x)-C(s,x)+\lambda(s,x)\left[g(s+x)-L\right].\end{aligned} \tag{6}$$

Let us note that the cost of TS $f(s)$ at the beginning of cycle $M_s$ is equal to the maximum value of $G(s,T)$, out of which and by virtue of (5) we obtain:

$$\begin{aligned}f(s)&=\max_T G(s,T)=\\&=\max_T\left\{e^{-N(s,T)}g(s+T)+\int_0^T e^{-N(s,x)}H(s,x)\,dx\right\}.\end{aligned} \tag{7}$$

The optimal $T_s$ will be the value of $T$, under which $G(s,T)$ is maximal. But, perhaps, such $T$ are more than one, or $T=\infty$. Let us consider both of the options.

1. Let us assume that the maximum $G(s,T)$ is reached both if $T=T'$, and if $T=T''>T'$. But a TS operating within cycle $M_s$, before it reaches state $(s,T'')$, must first be in state $(s,T')$, where it will be decided upon its repairs or disposal. Therefore, if used rationally, it simply will not "live until" state $(s,T'')$. That means that $T_s$ must be the smallest of the values of $T$ that maximize $Q(T)$.

2. The case of $T = \infty$ is impossible, as $G(s,T)$ decreases if $T$ are large. Indeed, as it was shown above, if $s$ are sufficiently large, $f(s) = 0$. Out of that and (4) follows that if $T$ is sufficiently large, $g(s + T) = 0$. But then by virtue of (5) and (6) $G(s,T) = \int_0^T e^{-N(s,x)} H(s,x)\,dx$ and $G'_T(s,T) = e^{-N(s,T)}\left[BQ(s,T) - C(s,T) - \lambda(s,T)L\right]$. Since at least one of the functions $C(s,t)$ and $\lambda(s,t)$ increases without limit if $t \to \infty$, $G'_T(s,T)$ becomes negative if $T$ is sufficiently large, while $G(s,T)$ will decrease, which was to be proven.

Let us note that function $G(s,T)$ is continuous, but not monotone in terms of $T$, and therefore can have several local maxima, out of which only one will be chosen as $T_s$, i.e. the global one. However, if $s$ changes, some of the local maxima may simply disappear, while the global maximum may "jump" from one local maximum to another. As a result, the dependence of $T_s$ on $s$ may be discontinuous, and, in the points of discontinuity, the maximum value of $G(s,T)$ will be achieved at two points at once. A similar situations may arise if $s$ is fixed if the initial data, e.g., loss value $L$ or dependencies $\lambda(s,t)$ and $C(s,t)$, are modified.

## Algorithm of model solution

To solve the problem, let us substitute (4) into formula (7) and represent it as follows:

$$f(s) = \mathrm{G}\left(f(s)\right), \tag{8}$$

where **G** is an operator that translates the function of one variable $\varphi(s)$ into another function $\phi(s)$ as follows:

$$\phi(s) = G(\varphi(s)) \triangleq \max_{T \geq 0} \left\{ \max\left[\varphi(s+T) - R; 0\right] e^{-N(s,T)} + \right.$$
$$\left. + \int_0^T e^{-N(s,x)} \begin{bmatrix} B - C(s,x) + \lambda(s,x) \cdot \\ \cdot \max\left[\varphi(s+x) - R; 0\right] - \lambda(s,x) L \end{bmatrix} dx \right\}. \tag{9}$$

It is easy to see that if function $\varphi(s)$ is continuous, non-negative and confined for $s \geq 0$, then function $\phi = \mathbf{G}(\varphi)$ will be identical. It is also easy to see that operator **G** is monotone: if $\varphi_1(s) \geq \varphi_2(s)$, then $\phi_1(s) \geq \phi_2(s)$, too. That allows solving equation (8) using the iterative method. For example, for the first approximation we can take $f_1(s) = 0$, and the subsequent ones we can find using formula $f_{n+1} = \mathbf{G}(f_n)$. Then, sequence $\{f_n(s)\}$ will be monotone and bounded, and, therefore, will have a limit (the "fixed point" of operator G), the required $f(s)$ equal to zero for sufficiently large $s$. In case of a numerical solution, the values of $f(s)$ were identified at the points of a uniformly spaced small-stepped grid, while the integrals were calculated using the Simpson formula.

In the course of the solution, for each cycle $M_s$, its assigned duration $T_s$ is also defined as the least $T$ that enables the maximum of (5). Naturally, the durations $T_s$ of different IRC will be different, which was revealed back when the deterministic situation was considered, e.g., in [17]. Further, all the terminal cycles can be identified (they will have $f(s) > 0 = g(s + T_s)$ along with the maximum service life of TS $T_{max}$ (it corresponds to the least $s$, under which $f(s) = 0$). Knowing $f(s)$, the cost of TS in any other states $(s,t)$ can be calculated as well. The corresponding formulas are derived in the same way as formulas (5) and (7), but we do not need them.

In the above procedure, the MV of the work performed by a serviceable TS per unit of time $B$ was considered known, although the owners of machinery and equipment are not usually aware of it, and valuators almost never estimate the cost of work (except, probably, that of construction, installation and repair activities). To solve this problem, let us note that the above procedure can be performed for different values of $B$, and all the costs of $f(s)$ will be non-decreasing functions of $B$. That is also true for the cost of the TS at the beginning of its use $f(0)$. But this cost is known and is equal to $K$. Therefore, the desired value $B$ must be the root of the equation $f(0) = K$. This method of estimating the cost of work strictly corresponds to the cost-based approach to valuation, although, in this form, it has not yet been used by valuators.

The model assumed there was no inflation. However, the model is also applicable under conditions of inflation, if, according to valuation standards [1], cost indicators are measured in prices at the valuation date, and the real (rather than nominal) pre-tax discount rate is used. Such procedure can also be substantiated by the method described in [12, section 4.3].

## Experimental calculations

According to model (7), experimental calculations were performed with the following input data:

- TS performance at the beginning of operation is adopted as 1: - $Q(0,0) = 1$, rate of operating costs $C_0 = C(0,0) = 40$, market value $K = 100$;
- repair costs $R = 25$;
- in the first IRC, as the TS ages, its performance deteriorates exponentially at a rate of $\alpha = 0.02$ 1/year, the rate of operating costs increases linearly at the rate of $i = 0.03$, while the failures have a Rayleigh distribution with the parameter $\omega$ (mean time to failure is equal to $\omega\sqrt{\pi/2}$).
- in other IRCs, the TS characteristics were described by a modified Kijima model:

$$Q(s,t) = e^{-\alpha\left(\beta s + t\gamma^s\right)};\ C(s,t) = C_0\left[1 + i\left(\beta s + t\gamma^s\right)\right];$$
$$\lambda(s,t) = \left(\beta s + t\gamma^s\right)/\omega^2,$$

where $\beta = 0.4$, $\gamma = 1.2$.

Failure-related losses $L$ and the failure distribution parameter $\omega$ varied. The MV of work performed by the TS per a unit of time, $B$, was determined from the condition $f(0) = K$.

We examined the effect of parameters $L$ and $\omega$ on the assigned times of preventive repairs and the maximum service life of the TS $T_{max}$. Let us set forth only some of the findings (in their entirety, they would take up too much space).

Value $L$ ranged from 100 to 1000 (from one to ten times the MV of the TS cost). Its effect, if ω = 4 and 8 years, is shown in Fig. 1 to 3. Figure 1 shows the dependence of the assigned time of the first repairs ($T_0$, years) on $L$. Disturbances on the graph occur when $L$ is small. They correspond to the above situations, whereas the maximum value of $Q$(s,T) is reached at two points at once. The dependence of the maximum service life of the TS ($T_{max}$, years) on $L$ is shown in Fig. 2.

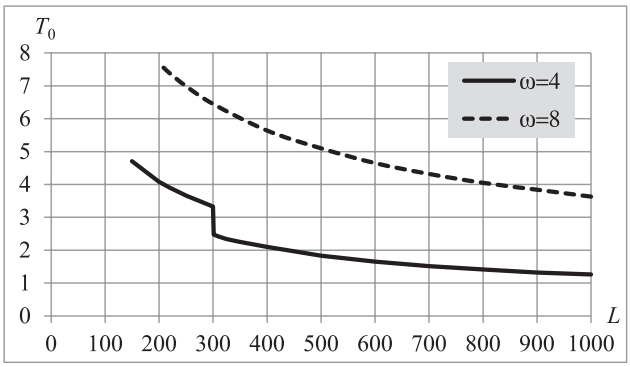



Fig. 1. Dependences of the assigned time of the first repairs ($T_0$, years) on $L$ if ω = 4 and 8 years

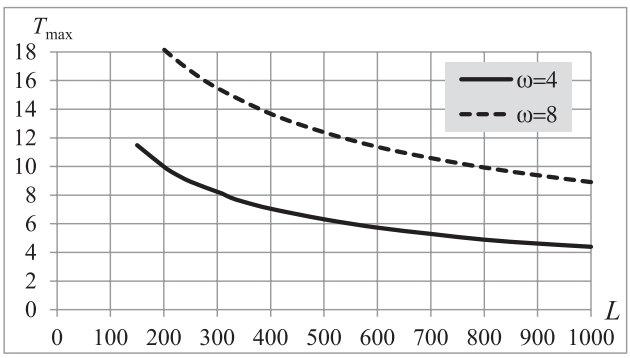



Fig. 2. Dependences of the maximum TS service life ($T_{max}$, years) on $L$ if ω = 4 and 8 years

Figures 3 and 4 show the dependences of the cycle duration ($T_s$, years) on the age of the TS at the beginning of the cycle ($s$, years) for various combinations of $L$ and ω. It can be seen that the optimal policy is significantly different from the common one, whereas the time of scheduled repairs is assigned identical or according to the serial number of the repairs regardless of the failure-related damage. Note that TS of a "sufficient" age is to be assigned very short times of scheduled repairs, which is technically inconvenient and provides a small economic effect. Therefore, such TS should not be assigned a time of the next preventive repairs at all. They should be disposed of only upon the next failure.

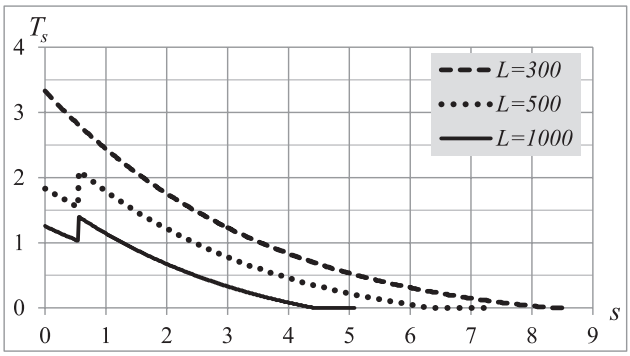



Fig. 3. Dependences of the assigned time of preventive maintenance ($T_s$, years) on the age of the TS at the beginning of the cycle ($s$, years) if ω = 4 years and varied $L$

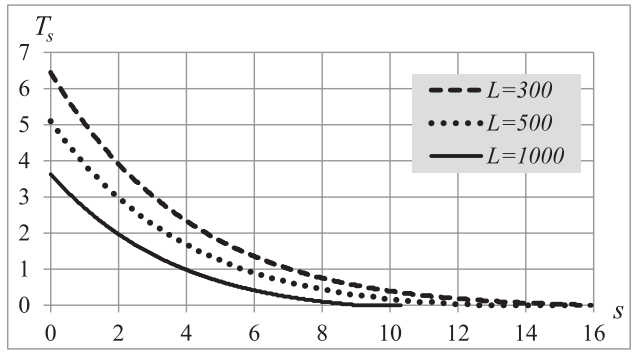



Fig. 4. Dependences of the assigned time of preventive maintenance ($T_s$, years) on the age of the TS at the beginning of the cycle ($s$, years) if ω = 8 years and varied $L$

## Conclusion

The cost criteria used in the dependability theory for optimising TS repair policy do not fully meet the business interests of companies. An economically substantiated solution of such problems is ensured by methods and criteria used in the valuation theory. They allow estimating the cost of work (services) performed by the TS, and, in particular cases, result in the criterion of the minimum expected discounted unit costs that can be rarely found in dependability-related literature.

In it, changes in the performance characteristics of TS after repairs are described by Kijima's virtual (effective) age models. Half a century before Kijima, a similar indicator was proposed for the valuation of assets and is still practically used by appraisers today. However, we show the inadequacy of describing the condition of repairable TS by any one such indicator. It appears to be more appropriate to characterize their condition by two indicators, i.e., the operating time at the beginning of the IRC and in the course of such cycle.

The above provisions allow constructing models for optimising a repair policy that meets the economic interests of market players. It is shown that, in each IRC, the time of the next scheduled repairs is to be assigned depending on the damage caused by a failure and the age of the TS at the beginning of the cycle, not on the serial number of repairs.

## References

1. International Valuation Standards (IVS) (2019). Effective 31 January 2020. International Valuation Standards Council.

2. Fedotova M.A., Koroliov I.V., Kovaliov A.P., et al. Fedotova M.A, editor. [Evaluation of machines and equipment:

Textbook. Second edition, updated and revised]. Moscow: INFRA-M; 2018. (in Russ.)

3. Aven T. Optimal replacement under a minimal repair strategy. *Advances in Applied Probability* 1983;15(1):198-211. DOI: 10.2307/1426990.

4. Bergman B. Optimal Replacement under a general failure model. Advances in Applied Probability 2;10(2):431-451.

5. Christer A.H., Waller M.W. Tax-Adjusted Replacement Models. *Journal of the Operational Research Society* 1987; 38(11):993-1006. DOI: 10.1057/jors.1987.170.

6. Jiang R. Performance evaluation of seven optimization models of age replacement policy. *Reliability Engineering & System Safety* 2018;180(C):302-311. DOI: 10.1016/j.ress.2018.07.030.

7. Van Horenbeek A., Pintelon L., Muchiri P. Maintenance optimization models and criteria. *International Journal of System Assurance Engineering and Management* 2010;1(3):189-200. DOI: 10.1007/s13198-011-0045-//.

8. [Guidelines for efficiency assessment of investment projects. Second edition. Approved by the Ministry of Economy of the RF, Ministry of Finance of the RF, Gosstroy of the RF, 21.06.1999, no. VK477]. Moscow: Ekonomika; 2000. (in Russ.)

9. Vilensky P.L., Livshits V.N., Smolyak S.A. [Efficiency assessment of investment projects: theory and practice. Study guide. Fifth edition]. Moscow: PoliPrintServis; 2015. (in Russ.)

10. Mikerin G.I., Smolyak S.A. [Efficiency assessment of investment projects and property valuation: convergence opportunities]. Moscow: CEMI RAS Publishing; 2010. (in Russ.)

11. Smolyak S.A. [Preventive repairs of machines in a Kijima-type model]. *Vestnik CEMI RAS* 2018;1(2). Available at: https://cemi.jes.su/s111111110000091-3-1/. DOI: 10.33276/S0000091-3-1. (in Russ.)

12. Smolyak S.A. [Valuation of machines and equipment (secrets of discounted cash flow)]. Moscow: Optsion; 2016. (in Russ.)

13. Livshits V.N., Smolyak S.A. [Service life of fixed assets in an optimal plan]. In: Proceedings of the First Conference for Optimal Economy Planning and Management. Section 1, Issue 1. Moscow: CEMI RAS; 1971. P. 352-357. (in Russ.)

14. Kijima M. Some results for repairable systems with general repair. *Journal of Applied Probability* 1989;26:89-102.

15. Welch R.B. Depreciation of buildings for assessment purposes. Chicago: International Association of Assessing Officers; 1943.

16. Chumakov I.A., Chepurko V.A., Antonov A.V. [On some properties of Kijima incomplete recovery models]. Dependability 2015;3(54):10-15.

17. Smolyak S.A. Overhaul policy optimization and equipment valuation concerning its reliability. *Journal of the New Economic Association* 2014;2(22):102-131. (in Russ.)

18. Smolyak S.A. Optimization of the Number and Frequency of Repairs. *Economics of Contemporary Russia* 2019;2:84-103. DOI: 10.33293/1609-1442-2019-2(85)-84-103. (In Russ.)

19. Lin Ye (Lam Yeh). Geometric processes and replacement problem. *Acta Mathematicae Applicatae Sinica* 1988,4:366-377. DOI: 10.1007/BF02007241.

## About the author

**Sergey A. Smolyak**, Doctor of Economics, Chief Researcher, Central Economics and Mathematics Institute, Russian Academy of Sciences. Address: 27 Krasnobogatyrskaya St., app. 129, 107564, Moscow, Russian Federation, e-mail: smolyak1@yandex.ru.

## The author's contribution

The author analysed the existing methods of accounting for the effects of repairs on the operational characteristics of technical systems, suggested an alternative method. A model was constructed for the purpose of optimizing the timeframe of scheduled repairs of technical systems that is based on the valuation theory and ensures maximized market value of companies that own such systems. Calculations using the model show that an optimal timeframe of scheduled repairs of a technical system significantly depends not only on the failure rate, but the system's age at the start of the inter-repair cycle and losses in production caused by its failures.

## Conflict of interests

The author declares the absence of a conflict of interests.

# Design engineering approach to ensuring specified dependability. Case study of unique, highly critical systems with short operation life

**Yuri P. Pokhabov**, *Joint Stock Company NPO PM – Maloe Konstruktorskoye Buro (AO NPO PM MKB), Zheleznogorsk, Krasnoyarsk Krai, Russian Federation*
*pokhabov_yury@mail.ru*

*Yuri P. Pokhabov*

**Abstract. Aim.** *To examine the design engineering approach to ensuring specified dependability on the basis of engineering disciplines and design engineering methods of quality and dependability assurance using the case of unique, highly critical products with short operation life. Such approach, unlike the statistical procedures of modern dependability, allows associating the dependability indicator calculations with the calculated operability parameters and established design criteria that are to be met in order to confirm the specified dependability indicators for products with an indefinite number of critical elements, each of which operates according to a functional principle that is different in its nature.* **Methods.** *The paper examined the prerequisites for the implementation of the design engineering approach to dependability, such as the distinctive features of ensuring the dependability of unique, highly critical products with short operation life, the applicability of design engineering approach to dependability, the effect of the genesis on the assurance of design engineering dependability, behavioural models of technical products in terms of dependability and specifics of highly critical product calculation. It was identified that, for items with high specified probability of no failure exceeding three-sigma random value variation interval, dependability is to be calculated not by identifying the dependability function, but rather by proving that undependability function is below the acceptable value, which ultimately ensures the specified dependability. Such approach enables the development of methods of early failure prevention using procedures of design engineering analysis of dependability for the purpose of achieving the required parameters of functionality, operability and dependability of products on the basis of a generalised parametric functional model.* **Results.** *The design engineering analysis of dependability allows substantiating the criteria for error-free design (selection of sound principles of operability and validation of engineering solutions for achieving the required dependability indicators). The effect of the error-free engineering criteria combined with the criteria for defect-free engineering (observance of the generally accepted principles, rules, requirements, norms and standards of drawing generation) and defect-free manufacture (strict adherence to the requirements of drawings with no deviation permits) enables a designer to achieve the specified dependability values without using the statistical methods of the modern dependability theory.* **Conclusion.** *Dependability as a comprehensive property is characterised by a probability that, on the one hand, determines the rate of possible failures, and, on the other hand, indicates the number of errors that were made by engineers during the design, manufacture and operation of products and can lead to failures. Additionally, the failure rate is determined by the engineers' efforts to eliminate or mitigate the consequences of possible failures at each life cycle stage. The greater and earlier are such efforts adopted, the higher the product's dependability will be. Ultimately, dependability is determined by consistent and rigorous implementation of error-free design, defect-free design and defect-free manufacture procedures whose efficiency is in no way associated with the number of manufactured products. Their efficiency and effectiveness are determined by specific decisions and actions by the engineers who make sure that the product performs the required functions with the specified dependability in the established modes and conditions of operation. Ensuring that only takes using engineering disciplines, as well as design engineering methods for quality and dependability assurance.*

**Keywords:** *dependability calculation, error-free design, defect-free design, defect-free manufacture, spacecraft, design engineering analysis of dependability (DEAD).*

**For citation:** *Pokhabov Yu.P. Design engineering approach to ensuring specified dependability. Case study of unique, highly critical systems with short operation life. Dependability 2022;1: 20-29. https://doi.org/10.21683/1729-2646-2022-22-1-20-29*

**Received on:** *18.12.2021 /* **Revised on:** *21.01.2022 /* **For printing:** *18.03.2022.*

## 1. Introduction

Ensuring faultless operation of single-use mechanical devices of spacecraft normally involves using methods of analytical and experimental verification that, in practice, have little in common with the stochastic methods of modern dependability [1, Chapters 16, 22]. Additionally, the methodological gap is so great that designers simply do not understand and are not aware of the relationship between the decisions they make and the specified dependability indicators, while the results of practical activities and calculations of dependability indicators do not correspond to each other so much that the founding fathers of the Russian aerospace industry generated a meme: "Dependability is calculated by people who cannot achieve it". It is still true today.

This practice is primarily due to the fact that calculations of dependability indicators are in no way associated with the types and tasks of parameter calculations (kinematic, electrical, thermal, hydraulic, pneumatic, etc.) that confirm the operability of products and serve as a ground for design solutions. The only exception is the dependability problems, in which the reliability indicators depend only on the strength parameters. In this case, it is normally considered that the *probability of no failure is identical to the probability that, within a specified time interval, the value of the loading parameter will never exceed the value that the strength parameter takes* ([2][1], see Reference Annex, comment on term "Reliability measure"), while strength calculations are performed taking into account the specified safety factors and strength margins that ensure the required probability of random parameters of loads and strength being within the acceptable range of values [3]. However, in most cases, the dependability objectives go beyond the matters of strength. The strength-specific dependability is regarded only as a conditional probability of failure that is based on the assumption that all other factors that can affect dependability are not critical.

It is still being debated in the research and engineering community as to what calculations of operability parameters and design criteria (apart from strength) are to be performed in order to confirm the specified dependability indicators for products with an uncertain number of critical elements, each of which operates according to principles that are different in their nature [4]. Ultimately, this problem is one of the causes of the widespread use of the statistical methods of modern dependability, as such methods do not require engineering analysis of the operability parameters of critical elements with different nature of operation. However, after the emergence of unique, highly critical products[2], the use of statistical methods eventually not only aggravated the problem due to the requirement to ensure the dependability of almost failsafe products, but also caused a complete misunderstanding of how to verify dependability in the situation of unavailable or insufficient failure statistics. Within the scope of the generally accepted approaches to dependability, there is still no scientifically substantiated solution for the problem of dependability of unique, highly critical products, which is confirmed by such regulatory documents as GOST RO 1410-001-2009, GOST 27.301-95, RD 50-476-84, etc. Additionally, the national standard GOST R 27.013-2019 clearly states that "*the probability of no failure is an indicator that cannot be evaluated using data for a single item.*" Regulatory documents do not clarify how to proceed, if the manufacturing procedure requires assessing the dependability of a product that was manufactured in a single instance and has no comparable items.

The paper presents and substantiates an approach to ensuring specified dependability based on engineering disciplines and design engineering methods of quality and dependability assurance using the case of unique, highly critical products with short operation time [5–7]. If required, the laws of the presented approaches allow extending them to any other technical products as forks [8].

## 2. Prerequisites for implementing the approach to design engineering dependability

Literally all engineering practices are based on the confirmation of physical principles of item operation and the application of design engineering methods of quality and dependability assurance, while no one doubts that this is the only way to achieve the required level of dependability. Nevertheless, designers continue using such approaches to dependability that do not allow understanding what dependability indicators can be achieved, given certain engineering practices, without involving the statistical methods of modern dependability. However, the statistical rules of dependability are only a consequence of an engineering practice in the form of quantitative interpretation of adverse events that allow judging upon product dependability based on data regarding failures that have already affected similar items. Additionally, if statistics are not available, such rules are of no practical importance, while if dependability requirements are high (with the probability above the three-sigma range of random value variation), obtaining the required reliable statistics may prove to be impossible. In particular,

[1] GOST 27.002-89 that is referred to herein is historical and has been replaced by GOST 27.002-2015 with removal of the reference annex, whose contents are of interest as regards the matters considered in this paper.

[2] Unique highly vital products are understood as virtually failsafe products that are unique (rare) in terms of their design, manufactured not more than in small series and operating in unique environmental conditions.

in case of unique, highly critical systems, it is virtually impossible to identify the dependability indicators using statistical methods due to financial/economic and/or physical/technical considerations (e.g., due to the large number of required test items and/or the need to conduct the tests in conditions drastically different from those on Earth, for example, in zero gravity and/or in increased radiation) [4].

At the same time, it is known that dependability as a property remains relevant for single or mass-produced items, with a long or short operation time, regardless of the availability of failure statistics [9]. Everything is defined by the ability of items to retain their properties over time under given modes and conditions of operation. The difference is that in the case of failures associated with long periods of operation, we are dealing with unacceptable (fatal) deterioration of functional properties of items over time, and in the case of short periods of operation the matter consists in various errors, i.e., actions or inaction of people (designers, fabricators, operators) that cause unintended results and eventually failures. Any unacceptable deterioration of item properties under the specified (known at the beginning of development) modes and conditions of operation are also errors, only associated with insufficient knowledge regarding the item operation, both in terms of its design (internal structure of elements and their interaction) and environmental conditions of application. Accordingly, statistical dependability may well be used to characterise cumulative errors that unintentionally occur in the course of design, development, manufacture and operation of products.

It is obvious that the case of unique, highly critical systems with short operation time simplified the identification of the effect of design engineering factors on dependability, as for such systems the reliability depends on a single performance of the required functions, rather than on the duration of operation exposed to the effect of modes and conditions of operation, which in itself is a complex scientific and technical problem (that distracts from the assessment of the criticality of human errors). Failures of products with short operation time are defined by the substantiated quality of the engineers' decisions and most often manifest as professional errors unlike failures caused by long deterioration of product characteristics over time that leads to gradual (implicit) decline in performance. However, in both cases, failures can be represented by a universally applicable diagram that describes the performance values of critical components going outside the admissible domain. The only difference is that the process may be sudden (instantaneous) or gradual (monotonous), which is determined by the physical processes that accompany the performance of the required functions by items. That is the context, in which are examined the various aspects of design engineering analysis of dependability of unique, highly critical systems that go beyond the statistical approaches of modern dependability [5].

## 3. Specificity of ensuring the dependability of unique, highly critical systems with short operation times

Virtually each spacecraft (being, in terms of its design, a rare and valuable product) in the orbit needs to deploy its folded structures (solar panels, antennas, reflectors, rods, etc.) into the operational position and only then is able to become fully functional for its intended purpose, e.g., as a repeater satellite [10–12]. The reliability requirements for such mechanical devices are so high that without using the engineering methods of identifying the potential critical failure hazards there is no point in creating spacecraft at all, which is evidenced by the fatal outcomes of the SinoSat 2 (2006), Kanopus-ST (2015), Mayak (2017), Zuma (2018), ChinaSat-18 (2019) missions, as well as the launches of many other artificial satellites and space devices [13–17].

The folded structures can deploy and assume their operational position in orbit only after the completion of a number of successive stages of spacecraft operation:

• ground transport and storage during and after exposure to transport loads and ground climatic conditions;

• final check of the mechanisms' operation in the technical area, where the "last draw" takes effect (possible unintentional disruption of mechanisms' operation before the flight as the result of personnel's action);

• flight as part of the launch vehicle during and after exposure to quasi-static, acoustic and vibration loads;

• separation from the last stage of the launch vehicle during and after impact loads;

• orbital flight in a folded position, when space factors that are sharply different from the atmospheric conditions on Earth (abnormally low or high temperature, temperature gradients, thermocycling, vacuum, microgravity, etc.) begin manifesting themselves;

• automatic deployment of mechanisms in presence of unstationary thermal processes of outer space and possible changes of dynamic dimensions of adjacent structures caused by microgravity (creating conditions for entanglement of moving parts);

• locking in the operating position with exposure to dynamic loads at the moment of end position lock operation.

The above sequence of events and states of mechanical devices of spacecraft is associated with integrated effects of modes and conditions of operation, which requires ensuring necessary and sufficient redundancy of product design to enable the specified dependability and is a complex engineering problem. It should also be taken into consideration that the products are manufactured in single instances, which is associated with a predominant share of manual labour in the assembly of unique systems (may result in anthropogenic risks of defects), the effect of technological heredity on the operation of mechanisms (in the form of assembly stresses, errors in the settings and adjustments of mechanisms, errors in the assembly operations, etc.), the practical impossibility of ensuring

redundancy of functional elements due to the high cost of the launched payload and strict weight and dimension restrictions on the satellite design, as well as the non-availability of reliable statistics on the operation of functional units in outer space. All of the above features apply to space structures that deploy immediately after launching into the orbit, unlike, e.g., delayed deployment after a long stay in outer space [18–19] or deployment of mechanical devices of landing modules on destination planets exposed to climatic, atmospheric and gravitational effects of poorly studied environments that require taking into account additional environmental effects affecting the reliability [20].

The special methods of calculating the dependability of deploying spacecraft structures (that, along with strength, take into account the requirement of mechanical unit mobility) were developed in the late 1970s [21–22], but largely lost their relevance due to the increased dependability requirements, which is evidenced by the mechanical failure statistics of space launches over the last few years [23–25]. The existing dependability requirements (about 0.999÷0.9999 and higher) create an objective need to take into account the design engineering factors of dependability assurance that guarantee maximum reliability of highly critical products manufactured virtually in a single instance with no critical element redundancy [23]. Additionally, when it is required due to practical considerations, it is important not to reject the statistical theory of dependability (at least, as one of the starting points), as mechanisms may include components and elements that obey statistical rules of modern dependability, e.g., pyrotechnic devices or electrical and electronic components [5]. The legitimacy of using statistical approaches to dependability is thoroughly substantiated in the reference annex to GOST 27.002-89 [2]. However, the difficulty of applying the statistical rules of modern dependability to deploying space structures consists in the fact that such rules are at the foundation of the series of standards 27, R 27, RV 27 and many other standards that do not imply other approaches even if no failure statistics are available. At the same time, the demand for complex, unique, highly critical systems complying with the specified dependability indicators for the military, nuclear and space industries is on a constant rise [26].

## 4. On the applicability of the design engineering approach to dependability

In practice, failures of unique, highly critical systems show that dependability problems exist not only for systems with long operating lives, but also for those with single operation [23]. Moreover, in the first case, the failures are primarily caused by various factors of damage to the structure of materials and joint assemblies, i.e., ageing, degradation, fatigue, wear, etc., while in the second case, those are mainly due to erroneous design solutions adopted on the basis of the distinctive features of the manufacturing process (design engineering solutions) [27–29]. Assuming that the causes of failures in both cases are inferior design or process engineering solutions [27], it is always possible to identify and apply those out of them that allow eliminating failures or reduce their probability. Accordingly, since the designer proceeds from the knowledge available to him/her under the process-specific constraints of production, the product's dependability will be fully defined by the designer's decisions. Moreover, design engineering methods allow handling failures of any nature (physical, stochastic, design engineering), which enables the migration from failure simulation using stochastic methods of modern dependability to managing failures at the physical level by choosing the required product parameters.

The fact that the modern research and engineering literature on dependability, with rare exceptions [30–31], does not discuss design factors (i.e., those associated with the designer) of dependability assurance can be easily explained. The designer's work in any field of technical activity is, by its nature, difficult to understand by those who are not directly involved in it. Moreover, the further from the drawing board, the more, at best, is visible only the tangible result of the designer's work – the drawings – yet the process of their conception, i.e., the origin and substantiation of the design concept that most often defines the causes of future failures, is completely incomprehensible (and indifferent). The design concept is the cumulative result of the use of a person's natural abilities and individual knowledge that he/she accumulates, preserves and applies to the creation of technical items throughout the professional life. It has nothing to do with the computerisation of business that aims to reduce the share of routine operations, therefore substituting the designer's knowledge and skills with computer capabilities cannot improve the dependability of developed technology [32–34].

No educational and academic institution or industrial agency has or is not involved with the development of scientific and methodological foundations of dependability assurance at the stage of design. In Soviet times, it was believed that fundamental engineering education was sufficient for designers to be able to develop quality and dependable equipment. Nevertheless, every major company created specialised engineering schools continuously enriched by the experience and knowledge of many generations of engineers that, for various reasons, was not properly formalized, but passed on by word of mouth from generation to generation [23]. At the same time, all research in the field of dependability assumed that the operability of products by the beginning of operation is ensured by default (due to the high qualification of designers), i.e., virtually out of the context of the genesis of dependability. In the modern world, the hopes are set on the computer-assisted design in belief that computers do not make mistakes and, therefore, design ensures dependability automatically [32]. However, the fact is

ignored that this dramatically increases the computational potential of technology and (through a misunderstanding) the educational level of engineers is unjustifiably reduced. In the author's opinion, that is a thoughtless mistake that needs to be addressed as soon as possible, but without developing and applying research and methodological approaches to dependability based on design engineering methods that would be almost impossible to do [33–34].

## 5. Genesis of the foundations of design engineering dependability

Philosophically speaking, all technical items that man creates are, in a sense, "prosthetics", devices that replace unobtainable functions or compensate for those that are not characteristic and difficult to achieve for a human being, e.g., to move in space (technical devices for transporting people and goods), communicate at a distance (means of telecommunication), live in comfortable climatic and other conditions (housing), etc. "Prosthetics", in the broad meaning of this word, that are commonly called technical items, are not the creation of nature existing by its laws, but something people artificially create owing to an understanding of the laws of nature (sometimes incorrectly or incompletely comprehended). Technical items make human life convenient, complete and comfortable, but are totally alien to the world around us and even ultimately harmful to humans when it comes to their disposal, and if so, then technical items are required and are created solely to satisfy the human needs[1]. Only man is able to conceive and impart to them a certain (required for him) functionality as a *set of properties defined by the presence and specific features of a set of functions capable of meeting given or implied needs* (GOST 28806-90). Moreover, such functionality of technical items must from the beginning (before their creation) be known and clear to man, otherwise significant safety risks may arise, if control is lost. The same principle applies to assembly drawings. All, even the smallest parts (e.g., bolts, nuts and washers) must be specified, each fulfilling a strictly defined function, for which they are all used. Each such function does not just (and only) exist, but can be formalised by a third person who is not directly associated with the design concept for the purpose of independent substantiation of its performance.

The understanding of functionality as the presence of a set of required functions ultimately underpins dependability that can only be achieved by focused and consistent human actions. Accordingly, without formalising what the required functionality is, it is virtually impossible to achieve dependability close to one.

## 6. Behaviour models of technical products in terms of dependability

In principle, any manufacture of products is organised in such a way that there are two ways of producing something. The first is "jury-rigging" according to the principle "good as done". The second one involves following a pre-designed plan, for which are used drawings of products with clear and known functionality, primarily as regards durability [35]. Drawings are important due to the fact that prior to the commencement of production, the information contained within them can be used to conduct the required engineering calculations, thus reducing the risk of errors, and to plan the production to improve its efficiency. The purpose of drawings is that they contain complete information on the performance by the product of its required functions, as well as the obligatory and sufficient requirements for its manufacture and operation. The absence or insufficiency of such information in the drawings inevitably reduces the product's dependability (the whole matter consists in the extent of such reduction). There are also two models of product behaviour in terms of dependability that are associated with drawings.

When no drawings for a product are available (they are not provided to the operator or they simply do not exist, e.g., they have been lost), the model of its behaviour in terms of dependability can only be identified by observing its operation (or through statistical tests). Such behaviour can be described using failure statistics, processing which using mathematical methods various dependability indicators can be obtained. For the purpose of implementing such approach, the methods of modern dependability were created, when it is not relevant which of a product's components causes failure or why the failure even occurs. Here, a person is only an observer who studies and generalises the laws of technical items' behaviour based on the results of their operation.

The fact that statistical methods of dependability are a special case of the physical understanding of various processes and phenomena was repeatedly pointed out by Soviet scientists A.I. Berg [36], V.V. Bolotin [9, 37], A.S. Pronikov [38–39], A.M. Polovko [4], I.A. Ushakov [40] and many others, but no fundamental changes have taken place yet. Various predictions of a product's future behaviour are usually based on the data on technical items that came to the end of their useful lives [39], while no effective methods of failure management at the earliest possible stages of newly created items' life cycle, primarily in mechanical engineering, have yet appeared [41]. There are only general guidelines for the design and development of products that have been worked-out on the basis of a long practical activity of engineers, following which high performance and dependability can be ensured [42–46]. However, such guidelines have

[1] By the way, the proverbial artificial intelligence does not need the human "prosthetics" either. And why would an artificial intelligence create technical objects that humans need (the "prosthetics") if it does not need those, and why would it know better than humans what humans need (the same is the case for any digital technology, primarily, in the area of design).

nothing to do with providing evidence of the achieved/not achieved product dependability indicators based on specific decisions made by the designer in the course of product development, i.e., they do not answer the question: "How much the designer's mistakes may weigh in terms of dependability indicator reduction" [7]. Consequently, various assumptions and restrictions inevitably arise that are associated with the concepts of early failure prevention models. For instance, it is assumed that, at the initial moment of operation, an automatic spacecraft is operable (GOST R 56526-2015), it is impossible to describe the first hump of the *U*-shaped dependability curve by mathematical formulae suitable for engineering calculations [47], the dependability of power structures of spacecraft is close to one, if their strength has the required safety coefficients [48], system dependability is the higher the less functional elements it contains [4], etc.

The second model of technical system dependability-specific behaviour is based on the fact that the drawing contains all the obligatory and sufficient requirements for manufacturing and operating the product that, within the specified operation time, in the given modes and conditions of application, will work without failure. Virtually, the point is that the design of such products is based on the assumption of unacceptable failures, or acceptable risks of failures, in the worst-case scenario. The premises of that approach are described in the foundations of dependability-specific design, when it is required to observe the principle of redundancy in order to eliminate (or reduce) the uncertainty between the "required" product structure and the "randomness" of environmental factors, whereas the degree of redundancy defines the acceptable ratio between the specified dependability and the possible undependability [49]. That should mean that if no errors were made in the process of design and development, the manufacture was done without damage or defects, while, in operation, the requirements of the operational documentation were not violated, then failures simply cannot occur. Should deviations occur at any of the life cycle stages, a risk of failure appears. Therefore, the primary problem of any development is to prevent design and development errors and to take measures to prevent defects in the manufacture and operation of products. The solution of the problem is examined in detail using the case of deployable spacecraft structures in papers that can serve as guidelines for engineers for using design engineering approaches to dependability assurance suitable for practical application (implementation) [5, 23]. In this case, it can be considered that technical documentation (design and process engineering) is a textual model of the product that contains all the required and sufficient information for the performance of the required functions. In particular, the geometric parameters correspond to the specified dimensions and tolerances, the choice of materials is made based on scientifically substantiated physical and mechanical characteristics and established safety margins, the structural depths and wall thickness of structural elements are selected subject to the specified safety coefficients, etc., therefore, the output parameters of any actual implementation of the product in the course of manufacture will meet the requirements of the design documentation, and the product itself, accordingly, will operate as the designer intended it to. A logical result of this model of dependability-specific product behaviour are the well-known methods of defect-free design (compliance with the generally accepted principles, rules, requirements, norms and standards of drawing development) and defect-free manufacture (work in strict compliance with drawing requirements without deviation permission cards) [50–51].

If the second, technical documentation-based model of dependability-specific product behaviour is used, three problems arise [23]:

1) identifying its dependability using hard-copy (design and process engineering documentation) and electronic documents (e.g., an annotation 3D model);

2) defining the obligatory and sufficient requirements for the manufacture in the design and process engineering documentation to ensure its specified dependability;

3) conducting the required technical inspection of the defined requirements.

In a certain sense, such statement is a trivial engineering problem, the solution of which can be appropriately organized and directed, e.g., using the methods of early failure prevention. For example, using the procedures of design engineering analysis of dependability to achieve the required indicators of functionality, operability and dependability of products based on a generalized parametric model of operation [5–7, 23, 33–34]. Moreover, if economically and financially feasible, quantitative dependability indicators can be ensured as per the standards, based on the statistical approaches of modern dependability [52].

## 7. Specificity of highly critical product calculation

When it comes to ensuring reliability above three nines (i.e., 0.997, which corresponds to the three-sigma rule), any stochastics-based calculations become meaningless [4, Chapter 14]. All possible failures in this case will fall within the category of rare events that do not match statistical patterns due to the fact that any set will always be smaller than the required entire assembly. In fact, proper engineering analysis shows that such failures have perfectly rational causal relationships. The purpose of such analysis may be to prove that system undependability $Q(t)$ will be below a certain value

$$Q(t) \leq 1 - P(t).$$

The analysis should result in the planning and execution of calculations and tests aiming not so much to identify the dependability – as it is usually done in modern

dependability – but to confirm the required undependability using the method of negative judgements (antitheses). In this case, if it is proven that the undependability is less than, e.g., 0.0001, then the dependability would indeed be greater than 0.9999 [23].

Are "black swans" possible in this case? Certainly, they are (no one is safe from errors), but their number will obviously be much lower if left unaddressed in the belief that it is impossible to avoid errors anyway or by neglecting the development of the methodological framework for such analyses. It is only a matter of choice, i.e., to manage the risks of possible rare failures, or to reasonably reject this opportunity [7]. For example, if the specified dependability is not higher than 0.99, the use of the methods of modern dependability may well be justified, but if it is 0.999, those will prove to be absolutely insufficient and additional methods of early failure prevention will have to be employed enabling the designer to make timely and substantiated technical decisions for the purpose of failure prevention based on engineering disciplines and design engineering methods of quality and dependability assurance.

## 8. On the requirement to apply the methods of design engineering analysis of dependability

As it is known, dependability is the property of an item to retain in time the ability to perform the required functions in the specified modes and conditions of operation, maintenance, storage and transportation [52]. If a product does not yet exist, but the design documentation has already been developed, its dependability is objectively determined by the technical requirements of the design documentation for the manufacture and operation that define the ability of the product to display the specified dependability. This ability does not appear out of nowhere. It is defined by the designer in the course of development as a result of heuristic thinking, knowledge of the process and conditions of operation, engineering logic, calculated decisions, engineering calculations and development tests. In the process of manufacture, this ability can be reduced due to manufacturing defects and damage, or retained at the level of the design concept, if the conditions of defect-free manufacture are fulfilled [23]. Deviations from the requirements of operational documentation in operation have a similar effect. That is why it is believed that *it is impossible to improve equipment dependability in the course of operation. It can only be ensured and maintained at the required level* [4]. In this context, the design and engineering solutions directly determine the ability of a product to achieve a specified dependability. It is those solutions that define the product's dependability at the beginning of operation (at the stage of running-in) that, in turn, corresponds to the first "hump" on the *U*-shaped dependability curve. If the dependability genesis factors are taken into consideration, there is no business secret about the causes of the first "hump", as it is mentioned in [47], as well as about the possibility to describe the first "hump" of the curve by "simple mathematical formulas suitable for engineering calculations". Everything depends on the efficiency of the early failure prevention methods that the designer does or does not use.

The concept of dependability as a property and the ability to manifest such property does not contradict the definition of the term "probability" in GOST R 50779.10, where probability is considered as a real number between 0 and 1 associated with a random event that may reflect the relative frequency in a series of observations or the degree of confidence that a certain event will take place. The performance of the required functions by a product is conventionally characterized with the probability of no failure, i.e., the frequency probability that no failure will occur within a given operation time. However, there are no reasons not to characterize the operation of future products – in the course of design documentation development – with the conditional probability that the logical or subjective probability of its operation – should it be manufactured in accordance with the design and manufacturing documentation – is ensured, if the conditions of defect-free and manufacture were fulfilled (i.e., with no deterioration of the product's ability to manifest dependability the way that the designer has intended) [50-51].

The duality of the concept of "probability" leads to two ways of designing and manufacturing products. In the first case (frequency probability), whatever happens during the product's design and manufacture, with or without the application of quality management standards, such as the ISO 9000 series, its reliability can be characterised by a frequency probability that, within a certain (economically substantiated) range, can be monitored using statistical testing.

In the second case (conditional probability), product dependability can be based on the designer's confidence that all the technical requirements that he/she established in the design documentation are sufficiently substantiated and allow an actual product manufactured defect-free performing the specified functions regardless of the number of manufactured products. Additionally, the validity of the technical requirements means that any of the hypothetical (i.e., possible, yet for some reason not implemented in manufacture) or actual (as the result of actual manufacture) states and successions of product-related events would allow (or will allow) performing the required functions if the conditions of defect-free manufacture are fulfilled. A formalized description of such states and successions of events in the form of a set of parameters that characterize the ability to perform the required functions and the allowable limits of parameter value variation is identical to the concept of the digital twin, i.e., *"a single model that reliably describes all characteristics, processes and relationships both for an individual item and for the entire business process"*

[53]. In practice, the above confidence is supported by a check list of evidence of, e.g., the selection of materials and non-acceptability of substitution, specified physical dimensions, tolerances and their unconditional observance, specification of functional characteristics and their confirmation in the design, coordination of design requirements and manufacturing capabilities and limitations, compliance of the technological heredity factors with the requirements specified in the design documentation, acceptance testing of acquired products for compliance with the specified requirements, etc. This approach enables an ultimate dependability of a product manufactured even in a single instance without recurring to critical element redundancy. However, in this case, a method is required that would enable error-free design, i.e., choosing substantiated principles of operability and confirming engineering solutions for the purpose of achieving the specified dependability indicators.

The meaning of error-free design can be shown by the example used by the English naturalist T.H. Huxley to describe the essence of mathematics. Defect-free design (as in Uniform System for Design Documentation) and defect-free manufacture (as in ISO 9000) are millstones. If we fill them with wheat grains (error-free design), we will produce flour. If we mix wheat grain with litter (faulty design solutions), will not produce flour. The millstones (defect-free design and defect-free manufacture) will obediently grind litter (faulty solutions), producing the same litter (products with uncontrollable dependability).

Defect-free design is enabled by unbiased substantiation of critical solutions based on the assessment of the risks associated with the performance of each required product function for strict execution of the documentation (as is). The model involves that the designer predefines the performance of the required functions by means of the conditions that he/she examines based on the design and process constraints and specifies them in the form of drawing specifications that must be fulfilled and supervised in production. In this case, the dependability assessment at the stage of documentation preparation and manufacture is done by means of dependability calculation based on the probabilities of performance of the required functions by components and elements using the method of structural dependability [7]. The above method of dependability calculation can only be used along with the method of design engineering analysis of dependability, which allows obtaining a complete list of critical parameters and calculation criteria that affect dependability. That allows defining the tasks for engineering calculation and perfection of critical parameters of product operation subject to the established design margins [5].

## 9. Conclusion

Dependability as a comprehensive property is characterised by a probability that, on the one hand, determines the rate of possible failures, and, on the other hand, indicates the number of errors that were made by engineers during the design, manufacture and operation of products and can lead to failures. Additionally, the failure rate is determined by the engineers' efforts to eliminate or mitigate the consequences of possible failures at each life cycle stage. The greater and earlier are such efforts adopted, the higher the product's dependability will be.

Ultimately, dependability is determined by consistent and rigorous implementation of error-free design, defect-free design and defect-free manufacture procedures whose efficiency is in no way associated with the number of manufactured products. Their efficiency and effectiveness are determined by specific decisions and actions by the engineers who make sure that the product performs the required functions with the specified dependability in the established modes and conditions of operation.

Procedures for error-free design, defect-free design and defect-free manufacture are based on the results of design and process dependability analysis designed to achieve the required functionality, operability and dependability of products based on a generalised parametric model of operation. The methodology of such analysis uses the required engineering disciplines and design engineering methods for quality and dependability assurance, and is not bound to statistical rules of modern dependability.

## References

1. Conley P.L., editor. Space Vehicle Mechanisms – Elements of Successful Design. NJ: John Wiley & Sons; 1998.

2. GOST 27.002-89. Industrial product dependability. General principles. Terms and definitions. Moscow: Izdatelstvo Standartov; 1990. (in Russ.)

3. Biriukov G.P., Kukushkin Yu.F., Torpachev A.V. [Fundamentals of dependability and safety of launch facilities]. Moscow: MAI Publishing; 2002. (in Russ.)

4. Polovko A.M., Gurov S.V. [Foundations of the dependability theory]. Saint Petersburg: BHV-Peterburg; 2006. (in Russ.)

5. Pokhabov Yu.P. [Design engineering analysis of dependability. Guidelines. Case study of spacecraft separation system]. Zheleznogorsk: AO NPO PM MKB; 2020. [Issued certificate of copyright registration no. 3644 of 27.05.2020 registered by OOO Sibkopirait, Novosibirsk]. [accessed 20.10.2021]. Available at: https://gnedenko.net. (in Russ.)

6. Pokhabov Yu.P., Ushakov I.A. [On the fail-safety of unique highly critical systems]. *Metody menedzhmenta kachestva* 2014;11:50-56. (in Russ.)

7. Pokhabov Yu.P. On the dependability of highly critical non-recoverable space entities with short operation life. Case study of single-use mechanical devices. *Dependability* 2021;21(3):3-12.

8. Artyushenko A.G., Pokhabov Yu.P. Design and technology reliability analysis: fork. *IOP Conference Series: Materials Science and Engineering* 2020;862(2):022001(1–6). doi:10.1088/1757-899X/862/2/022001.

9. Bolotin V.V. [Application of probability theory and dependability theory methods in structural analysis]. Moscow: Izdatelstvo literatury po stroitelstvu; 1971. (in Russ.)

10. Always P. Rockets of the world. Saturn Press; 1999.

11. Fortescue P., Stark J., Swinerd G. Spacecraft Systems Engineering. NJ: John Wiley & Sons; 2003.

12. Testoyedov N.A., Kosenko V.E., Vygonsky Yu.G. et al. [Space relay systems]. Moscow: Radiotekhnika; 2017. (in Russ.)

13. Fusaro R.L. NASA Space Mechanisms Handbook – Lessons Learned Documented. *Research & Technology 1998. NASA/TM* 1999:138–140.

14. Shapiro W. et al. Space Mechanisms Lessons Learned Study, Volume I – Summary. NASA/TM-107046; 1995.

15. Shapiro W. et al. Space Mechanisms Lessons Learned Study, Volume II – Literature Review. NASA/TM-107047; 1995.

16. Gore B.W. Critical Clearances in Space Vehicles. The Aerospace Corporation ATR-2009(9369)-1; 2008.

17. Harland D.M., Lorenz R.D. Space systems failures: disasters and rescues of satellites, rockets and space probes. Berlin: Springer; 2005.

18. Shtokal A.O., Rykov E.V., Dobrosovestnov K.B. et al. Ways of dependability enhancement of spacecraft deployment units with suspended actuation operating. Vestnik NPO im. S.A. Lavochkina 2017;4:60-67. (in Russ.)

19. Merstallinger A., Sales M., Semerad E. et al. Assessment of Cold Welding between Separable Contact Surfaces due to Impact and Fretting under Vacuum. ESA STM-279. Nordwijk; 2009.

20. Pokhabov Yu.P., Makarov V.P., Kolobov A.Yu. et al. [Aspects of ensuring the operational dependability of the mechanical devices for deployment and locking of landing module structures]. *Aktualnye voprosy proektirovaniya kosmicheskikh sistem i kompleksov. Sbornik nauchnykh trudov* 2019;20:151-166. (in Russ.)

21. Kuznetsov A.A. [Structural dependability of ballistic missiles]. Moscow: Mashinostroenie; 1978. (in Russ.)

22. Kuznetsov A.A., Zolotov A.A., Komyagin V.A. et al. [Dependability of mechanical parts of aircraft design]. Moscow: Mashinostroenie; 1979. (in Russ.)

23. Pokhabov Yu.P. [Theory and practice of ensuring the dependability of single-use mechanical devices]. Krasnoyarsk: SFU; 2018. (in Russ.)

24. Saleh J.H., Caster J.-F. Reliability and multi-state failures: a statistical approach. First Edition. NJ: John Wiley & Sons; 2011.

25. [Failures of rocket and space technology]. [Launch vehicles, satellites, planes, devices: website]. [accessed 20.10.2021]. Available at: http://ecoruspace.me. (in Russ.)

26. Levenchuk A. [Systems engineering thinking in life cycle management]. [accessed 20.10.2021]. Available at: https://ailev.livejournal.com/1121478.html. (in Russ.)

27. Hecht H., Hecht M. Reliability prediction for spacecraft, Report prepared for Rome Air Development Center: no. RADC-TR-85-229, Dec. Rome Air Development Center; 1985.

28. Tumanov A.V., Zelentsov V.V., Shcheglov G.A. [Fundamentals of spacecraft on-board equipment layout design]. Moscow: Bauman MSTU Publishing; 2010. (in Russ.)

29. Sevastianov N.N., Andreev A.I. [Fundamentals of dependability management of spacecraft with long service life]. Tomsk: TSU Publishing; 2015. (in Russ.)

30. Van-Jelen V. [Physical theory of dependability]. Simferopol: Krym; 1998 [Russian].

31. Kurylenko A.M., Ledovsky A.D. [Quality of ship dynamic control systems]. Saint Petersburg: Sudostroyenie; 1994. (in Russ.)

32. Kuleshov A.P. To overcome the resistance of materials: February 2, 2018 interview]. *Stimul: zhurnal ob innovatsiyakh v Rossii*. [accesed 20.10.2021]. https://stimul.online/articles/interview/preodolet-soprotivlenie-materialov/?sphrase_id=1295. (in Russ.)

33. Pokhabov Yu.P. Dependability in digital technology. *Dependability* 2020;2:3-11.

34. Pokhabov Yu.P. Dependability from a designer's standpoint. *Dependability* 2020;4:13-20.

35. Haeder H. Konstruieren und Rechnen für Praxis und Schule. Saint Petersburg: Izdatelstvo K. Rikkera; 1904.

36. Berg A.I. [Selected works]. Energia; 1964. (in Russ.)

37. Bolotin V.V. [Theory of dependability of mechanical systems with a finite number of degrees of freedom]. *Izvestiya AN SSSR. Mechanics of solids* 1969;5:74-81. (in Russ.)

38. Pronikov A.S. [Parametric dependability of machines]. Moscow: Bauman MSTU Publishing; 2002. (in Russ.)

39. Pronikov A.S. [Dependability of machines]. Moscow: Mashinostroenie; 1978. (in Russ.)

40. Ushakov I.A. [Dependability: past, present, future: keynote speech of the opening of Mathematical Methods in Reliability (MMR–2000) conference, Bordeau, France, 2000]. *Reliability: Theory & Applications* 2016;1(1):17-27. (accessed 20.10.2021). Available at: http://www.gnedenko.net/Journal/2006/RTA_1_2006.pdf. (in Russ.)

41. Plahotnikova E.V., Safonov A.S., Ushakov M.V. The design of products with requirements of reliability parameters. *Izvestiya TulGU: Teknicheskie nauki* 2015;7(1):134-139. (in Russ.)

42. Yendogur A.I. [Aeronautical structure design. Structural design of parts and units]. Moscow: Izdatelstvo MAI-PRINT; 2009. (in Russ.)

43. Orlov P.I. Uchaev P.N., editor. Introduction into design in 2 volumes. Volume 1]. Moscow: Mashinostoenie; 1988. (in Russ.)

44. Khoroshev A.N. [Introduction into the design management of mechanical systems]. Belgorod; 1999. (in Russ.)

45. Lelikov O.P. [Fundamentals of calculation and design of machine parts and assemblies]. Moscow: Mashinostroenie; 2007. (in Russ.)

46. Bushuev V.V. [Practice of machine design]. Moscow: Mashonostrienie; 2006. (in Russ.)

47. Timoshenkov S.P., Simonov B.M., Goroshko V.N. [Fundamentals of the dependability theory]. Moscow: Yurait; 2015. (in Russ.)

48. Patraev V.E., Khalimanovich V.I. [Dependability of support spacecraft]. Krasnoyarsk: SibGAU; 2016. (in Russ.)

49. Venikov G.V. [Dependability and design]. Moscow: Znanie; 1971. (in Russ.)

50. Gorokhova V.V. [Application of the Saratov system in research and design]. Moscow: Izdatelstvo standartov; 1969. (in Russ.)

51. Dubovikov B.A. [Fundamentals of scientific quality management (practical experience and theoretical substantiation of the system for defect-free work organization). Moscow: Ekonomika; 1966. (in Russ.)

52. GOST 27.002-2015. Dependability in technics. General principles. Terms and definitions. Moscow: Standartinform; 2016. (in Russ.)

53. Borovkov A.I., Riabov Yu.A., Kukushkin K.V. et al. [Digital twins and the digital transformation of defense industry companies]. *Oboronnaya tekhnika* 2018;1:6-33. (in Russ.)

## About the author

Yuri Pokhabov, Candidate of Engineering, Joint Stock Company NPO PM – Maloe konstruktorskoye buro (OAO NPO PM MKB), Head of Research and Development Center, Zheleznogorsk, Krasnoyarsk Krai, Russian Federation, e-mail: pokhabov_yury@mail.ru

## The author's contribution

The paper continues the author's 2015–2021 series of publications in the Dependability Journal dedicated to the dependability of unique, highly critical systems using design engineering analysis of dependability based on a generalized parametric model of operation.

## Conflict of interests

The author declares the absence of a conflict of interests.

*Original article*
*https://doi.org/10.21683/1729-2646-2022-22-1-30-37*

# Efficiency criterion of biased estimates. A new take on old problems

**Viktor S. Mikhailov**, *D.I. Mendeleev Central Research and Design Institute of Chemistry and Mechanics, Moscow, Russian Federation*
*mvs1956@list.ru*


*Viktor S. Mikhailov*

**Abstract.** *The perfect case estimation scenario involves unbiased estimation with minimal variance, if such estimate exists. Currently, there are no means of obtaining unbiased estimates (if they do exist!). For instance, a maximum likelihood estimate (NBT test plan) of a mean time to failure $T_{mn}$ = (total operation time)/(number of failures) is highly biased. Those involved in solving applied problems are not satisfied with the situation. Efficient unbiased estimates are used whenever such are available. If it is impossible to find an efficient unbiased estimate in terms of standard deviation, then biased estimate comparison is to be mastered. The vast majority of problems is associated with biased estimates. Within the class of biased estimates, estimates with minimal bias are to be sought, and, among the latter, those with minimal bias. Such estimates in the class of biased estimates should be called bias-efficient or simply efficient, which does not contradict the conventional definition, but only extends it. Such search process guarantees that the obtained estimates are highly accurate. However, with this definition of a bias-efficient estimate, there will always be a pair of compared estimates, in which the total bias of one estimate is slightly higher than that of the other, the same being the case with the total variances of such estimates, but in a different order. In this setting, a formal selection of a bias-efficient estimate becomes impossible and is arbitrary, i.e., the test engineer selects a bias-efficient estimate intuitively. In this case, the test engineer's choice may prove to be incorrect. Thus arises the problem of constructing a criterion of efficiency that would enable a formal selection of a bias-efficient estimate.* **The Aim of the paper.** *The paper aims to build an efficiency criterion, using which the choice of a bias-efficient estimate is unambiguously defined through computation.* **Methods of research.** *To find the bias-efficient estimate, we used integral numerical characteristics of the accuracy of the estimate, namely, the total square of the offset of the expected implementation of a certain variant estimate from the examined parameters of the distribution laws, etc.* **Conclusions.** *1) For the binomial plan and the test plan with recovery and limited test time, performance criteria were constructed that allow unambiguously identifying the bias-efficient estimate out of the submitted estimates. 2) Based on the constructed performance criteria for various test plans, bias-efficient estimates were selected out of the submitted ones.*

**Keywords:** *estimation, efficient estimation, criterion of efficiency, test plan, biased estimates.*

**For citation:** *Mikhailov V.S. Efficiency criterion of biased estimates. A new take on old problems. Dependability 2022;1: 30-37. https://doi.org/10.21683/1729-2646-2022-22-1-30-37*

**Received on:** *10.11.2021 /* **Revised on:** *24.01.2022 /* **For printing:** *18.03.2022.*

## Introduction

An efficient estimate is defined as [1]: "An estimate of a parameter that has the lowest expected squared deviation from the estimated parameter for any parameter value is called efficient." The classical theory of mathematical statistics [1] notes that within the class of all possible parameter estimates, there is no efficient estimate. Therefore, the author of [1] further writes: "It is required to impose certain restrictions on the set of estimates, within which we are seeking the best efficient estimate. A natural restriction of the class of estimates is the class of so-called unbiased parameter estimates." In this case, the efficient estimate for the scalar parameter is an unbiased estimate with minimal variance. In some cases, Cramér-Rao inequalities help find the best unbiased estimate [1]: if an estimate is efficient, then, in the above sense, it also is the best, as it has the lowest possible variance.

In estimation, the perfect case scenario involves the use of unbiased estimates with minimal variance, if such estimate exists. For that purpose, in order to identify an efficient estimate, within the class of unbiased estimates, it should be analytically proven that the Cramér-Rao inequality is fulfilled for such estimate. It should be noted that Cramér-Rao inequalities are to be satisfied for all values of the estimated parameters. However, even for exponential families of distributions, for which only efficient estimates exist, an efficient estimation using a Cramér-Rao inequality is only possible for a single function of a parameter. The question is even more relevant as regards families of distributions that are not exponential. If it is difficult to obtain such proof analytically, the total variance should be calculated for all values of the estimated parameter. For an efficient unbiased estimate, the total variance should be minimal.

Currently, there are no means of obtaining unbiased estimates (should such exist!). For instance, a maximum likelihood estimate (NBT test plan) of the mean time to failure $T_{mn}$ = (total operation time)/(number of failures) is highly biased. Those involved in solving applied problems are not satisfied with the situation. Efficient unbiased estimates are used whenever such are available. If it is impossible to find an efficient unbiased estimate in terms of mean square variance, then biased estimate comparison is to be mastered. The vast majority of problems is associated with biased estimates.

Within the class of biased estimates, estimates with minimal bias are to be sought, and, among the latter, those with minimal variance [2]. Such estimates in the class of biased estimates should be called bias-efficient or simply efficient, which does not contradict the conventional definition, but only extends it. Such search process guarantees that the obtained estimates are highly accurate. Note that the experience of constructing efficient estimates shows that the resulting unbiased efficient estimate will not always have a minimum variance [2]. Rather, on the contrary, there will always be an estimate that has minimal variance compared to the unbiased estimate. In all cases where there is an efficient (unbiased) estimate, there is a biased estimate that is more accurate than the efficient one, i.e., with a smaller squared error [3, p. 284]. That fact favours bias as the primary factor in constructing the evaluation efficiency criterion. In order to determine the bias-efficient estimate, the total biases and variances are to be calculated for all values of the estimated parameter. For an efficient biased estimate, each sum must be minimal. Such definition of an efficient estimate within a particular class of biased estimates does not contradict the definition of an efficient estimate within a class of unbiased estimates. On the contrary, defining an efficient estimate within a class of unbiased estimates is a frequent case of defining an efficient estimate within a certain individual class of biased estimates that includes a subclass of unbiased estimates.

Why the integral approach? When comparing using the classical method, whereas the variance should be minimal for all parameter values at once, we deduce that one of the compared biased estimates will have a lower variance in one part of the parameter values, while the other will have a lower variance in the remaining part, with a comparable bias. Comparing them is what the summation of all variances (biases) is required for. The sums of biases and variances define the efficiency criterion.

However, with this definition of a bias-efficient estimate, there will always be a pair of compared estimates, in which the total bias of one estimate is slightly higher than that of the other, the same being the case with the total variances of such estimates, but in a different order. In this setting, a formal selection of a bias-efficient estimate becomes impossible and is arbitrary, i.e., the test engineer selects a bias-efficient estimate intuitively. In this case, the test engineer's choice may prove to be incorrect. Thus arises the problem of constructing an efficiency criterion that would enable a formal selection of a bias-efficient estimate.

## The Aim of the paper

The paper aims to build an efficiency criterion, using which the choice of a bias-efficient estimate is unambiguously defined through computation.

## Methods of research

The bias-efficient estimate was found using integral numerical characteristics of the accuracy of estimate, i.e., the sum square of the bias of the expected realization of an estimate from the considered parameters of the distribution laws, etc. [2].

## Constructing the estimate efficiency criterion

Let us denote by $A(\theta)$ the total bias of estimate $\theta$ from estimated parameter $t$, and by $B(\theta)$ the total variance of estimate $\theta$ from estimated parameter $t$. Note that summation is done within the operating range both for all values

of estimated parameter $t$, and all values of the test plan and other parameters (e.g., time it takes to estimate the probability of no failure (PNF).

For the purpose of constructing an efficiency criterion of biased estimates we will characterize arbitrary statistical estimate θ by bias and variance. Let us denote by $b = E(\theta) - t$ the bias of estimate θ from parameter $t$, where $E$ is the mathematical expectation, and by $D$ the variance of estimate θ. Then the variance (in the mean square sense) of a certain estimate θ from the estimated parameter $t$ is expressed by the following formula [1, 4, 5]:

$$B(\theta) = E(\theta - t)^2 = D + b^2. \quad (1)$$

Note that, when dispersion changes, the variance as an efficiency characteristic also changes by the same value (see formula (1)). That is, it changes regardless of the dependence on the specific value of estimate bias. Let us try to associate the dispersion and bias square in such a way as to make the variance change adjusted to bias whenever dispersion variates. We will take into consideration the fact that bias is the primary factor in choosing an efficient estimate. The newly built characteristic $C(\theta)$ must be such as, when the dispersion changes by the value of $\delta D$, for small biases $b \approx 0 + \delta$, the adjustment for the effect of the bias on the characteristic was insignificant, and vice versa, for large biases $b >> 0$, the adjustment for the effect of bias on characteristic $C(\theta)$ was significant. We will require that the variation of characteristic $C(\theta)$ was linear with respect to characteristics $D$ and $b^2$. The product of characteristics $D$ and $b^2$ fulfils this requirement to the fullest:

$$C(\theta) = D \cdot b^2. \quad (2)$$

Out of formula (2) follows that, as dispersion changes by value $\delta D$, characteristic $C(\theta) = (D + \delta D) \cdot b^2 = D \cdot b^2 + \delta D \cdot b^2$ changes by a value that takes into account the squared bias linearly. The opposite is also true, i.e., when the squared bias changes by a certain value, characteristic $C(\theta)$ changes by a value that takes into account the dispersion value linearly. Figuratively speaking, characteristic $C(\theta)$ reflects on the Cartesian axes $D$ and $b^2$ as a rectangle with the area of $D \cdot b^2$. Any slight change to characteristics $D$ and $b^2$ modifies the area or configuration of the rectangle. Thus, in case of slightly different characteristics $D$ and $b^2$, the estimate with the minimum characteristic $C(\theta)$ (area) should be chosen as the bias-efficient. If characteristics $C(\theta)$ (areas) are equal, the estimate with the lowest bias should be chosen as the bias-efficient. Let it be reminded that the criterion was constructed only for biased estimates. In the case of unbiased estimates, variance $B(\theta)$ (see formula (1)) is such characteristic (criterion). Note that, for unbiased estimates, their realizations are grouped around the true quantitative value of the estimated parameter from different sides. When defining the efficiency criterion, similar properties are to be required from biased estimates.

Let us define the requirements for the process of selecting bias-efficient estimates:

– the proposed estimates must be strictly monotonous in all their parameters;

– estimates with a minimum bias of $A(\theta) = b^2$ or close to such are selected.

If, in the process of selection out of a number submitted estimates, there is a single unbiased estimate, then the latter is the bias-efficient one. For this estimate to be efficient in the class of unbiased estimates, it is required to prove the Cramér-Rao for such estimate:

– estimates, for which inequality $A = b^2 > D$ is fulfilled, i.e., the bias prevails over the value scatter of such estimate, are excluded;

– estimates are selected, for which the inequality $D/A > 4$ is fulfilled, i.e., the estimates, for which the realizations are grouped around the true quantitative value of the estimated parameter from different sides;

– out of the remaining estimates, the estimate with the minimum bias $A(\theta) = b^2$ or close to such (+5 ... +20%) is selected. In the case a single estimate with minimum bias $A$ was selected, such estimate is considered bias-efficient;

– in case $A$ are equal, the estimate with minimal variance is chosen as the bias-efficient one.

The majority of manipulations is replaced by the proposed criterion $C(\theta) = D \cdot b^2$.

Let us consider examples of constructing a criterion for bias-efficient estimate selection.

## Binomial test plan. Probability of no failure

Here and further, we will use the findings of [2]. Let us denote by θ a certain abstract estimate of the probability of failure in the course of testing of $n$ products. We will limit the scope of testing to $0 < n \le 10$, which is the cost limit for highly dependable and complex products. Then, the total bias formula will be as follows:

$$A\left(\theta\left(n;R\right)\right) = \frac{1}{10}\sum_{n=1}^{10}\int_0^1 \left(E\theta\left(n;R\right) - p\right)^2 dp.$$

The formula for the total variance is as follows:

$$D\left(\theta\left(n;R\right)\right) = \frac{1}{10}\sum_{n=1}^{10}\int_0^1 E\left(\theta\left(n;R\right) - E\theta\left(n;R\right)\right)^2 dp.$$

Let us note that the probability function of the binomial test plan $P_\Sigma$ steadily decreases as $p$ grows [5], therefore, equations $P_\Sigma\left(R = r\right) = \sum_{k=0}^{r} P_n\left(k,w\right) = 0,5 + x$ and $P_\Sigma\left(R = r\right) = \sum_{k=0}^{r} P_n\left(k,v\right) = 0,5$ have a single solution, where $P_n(k,p) = C_n^r p^r (1-p)^{n-r}$.

Calculations show that probability $\gamma = 0.5 + x = 0.8181$ corresponds to estimate $w$ that minimizes functional $A(\theta(n;R))$. Table 1 shows the results of substituting into functionals $A(\theta(n;R))$, $D(\theta(n;R))$ of the following failure

**Table 1. Results of substituting the proposed failure probability estimates into functionals $A(\theta(n;R))$, $D(\theta(n;R))$ for the binomial test plan**

| Type of functional | $v$ $\gamma = 0.5$ | $w$ $\gamma = 0.81$ | $p_1$ $\gamma = 0.5$ | $p_2$ $\gamma = 0.81$ | $p_3$ $\gamma = 0.81$ | $u = (R + 1)/(n + 2)$ | $p_0 = R / n$ |
|---|---|---|---|---|---|---|---|
| $A$ | 0.0176 | 0.0037 | 0.0113 | 0.0015 | 0.0070 | 0.0104 | $6 \cdot 10^{-33}$ |
| $D$ | 0.0270 | 0.0402 | 0.0288 | 0.0401 | 0.0226 | 0.0162 | 0.0488 |
| $D / A$ | 1.53 | 10.86 | 2.54 | 26.73 | 3.22 | 1.55 | $\infty$ |
| $C = D \cdot A \cdot 10^4$ | 4.752 | 1.4874 | 3.2544 | 0.6015 | 1.595 | 1.6848 | $10^{-30}$ |

**Table 2. Results of substitution of failure probability estimates $\hat{}v$, $\hat{}w$, $p_{10}$, $p_{20}$ into functionals $A(\theta(n;R))$, $D(\theta(n;R))$ for the binomial test plan**

| Type of functional | $\hat{}v$ $\gamma = 0.5$ | $\hat{}w$ $\gamma = 0.81$ | $p_{10}$ $\gamma = 0.5$ | $p_{20}$ $\gamma = 0.81$ |
|---|---|---|---|---|
| $A$ | 0.0034 | 0.0030 | 0.000680 | 0.000355 |
| $D$ | 0.0356 | 0.0427 | 0.0425 | 0.0443 |
| $D / A$ | 10.47 | 14.23 | 62.5 | 124.7 |
| $C = D \cdot A \cdot 10^4$ | 1.210 | 1.28 | 0.289 | 0.157 |

probability estimates: $v$, $w$, $p_0 = R / n$, $p_1$, $p_2$, $p_3$ [5] and $u = (R + 1)/(n + 2)$, where

$p_1 = v(0.5;n)$, $R = 0$ and $p_1 = R / n$, $R > 0$;

$p_2 = w(0.81;n)$, $R = 0$ and $p_2 = R / n$, $R > 0$;

$p_3 = w(0.81;n)$, $R = 0$ and $p_3 = u$, $R > 0$.

Functionals $A(\theta(n;R))$ and $D(\theta(n;R))$ were calculated with the step of $\partial p = 10^{-3}$. Implicit estimates $w$ and $v$ were calculated with the accuracy of $10^{-4}$.

Here and further, for the purpose of table construction, as part of calculation of characteristic $C = D \cdot A$, functionals $A$ and $D$ were calculated for each value of parameters $n$ and $p$ with subsequent individual summation, and based on the obtained total values of $A$ and $D$, characteristic $C = D \cdot A$ was calculated.

Note that calculating characteristic $C$ directly as a functional

$$C\left(\theta\left(n;R\right)\right) =$$

$$= \frac{1}{10}\sum_{n=1}^{10}\int_0^1 E\left\{\theta\left(n;R\right) - E\theta\left(n;R\right)\right\}^2 \cdot \left\{E\theta\left(n;R\right) - p\right\}^2 dp$$

is associated with great computational difficulties due to the limited word length in the computer system, which, in the course of computation, causes clearing of significant summable values. That affects the final result.

Unbiased estimate $p_0 = R / n$ that was given for comparison is excluded from consideration as a bias-efficient one despite the fact that it is efficient.

Out of Table 1 follows that estimates $v$, $p_1$, $p_3$, $u$ are to be excluded from consideration, as inequality $D / A > 4$ does not apply to them. Then, out of Table 1 also follows that estimates $w$ and $p_2$ have minimal and comparable biases. Their values do not differ by more than $(0.0037 - 0.0015) \cdot 100/0.0037 = 59\%$. In accordance with the proposed efficiency criterion of biased estimates, estimate $p_2$ is to be definitely considered efficient. Out of the construction follows that the criterion constructed based on characteristic $C = D \cdot A$ unambiguously determines the bias-efficient estimate without recurring to most of the above reasonings in this paragraph.

The proposed estimates $v$, $w$, $p_1$, $p_2$ for the binomial test plan have a bias that can be reduced, which slightly modifies the estimates as follows:

$\hat{}v = v(0.5;n,R) - 0.4 / ((R + 1)n)$;

$\hat{}w = w(0.81;n,R) - 0.1/((R+1)n)$;

$p_{10} = \hat{}v(0.5;n)$, $R = 0$ and $p_{10} = R / n$, $R > 0$;

$p_{20} = \hat{}w(0.81;n)$, $R = 0$ and $p_{20} = R / n$, $R > 0$.

Table 2 shows the results of substituting into functionals $A(\theta(n;R))$, $D(\theta(n;R))$ of the following probability of failure estimates: $\hat{}v$, $\hat{}w$, $p_{10}$, $p_{20}$.

Out of Table 2 follows that, for all available estimates, inequality $D / A > 4$ is correct. In accordance with the proposed efficiency criterion of biased estimates, estimate $p_{20}$ is to be definitely considered efficient.

## Binomial test plan. Mean time to failure

Let us assume that the products' time to failure follows the exponential law of probability distribution (d.l.) with parameter $T_0$, where the latter is identical to the mean time to failure. Let us denote the test time of each of the $N$ products as $\tau$.

As the criterion of efficient MTF estimate, a functional is constructed that is based on summing the squared relative biases of expected estimates $\theta(R,n)$ from the parameter $t$ of the exponential d.l. (MTF) for all possible values of $N$, $\tau$, $T_0 = t$ [2]

$$A\left(\theta\left(n;R\right)\right) = \frac{1}{3}\sum_{\tau=10^3}^{10^5}\frac{1}{10}\sum_{n=1}^{10}\int_0^\infty \frac{1}{t^2}\left\{E\theta\left(n;R,\tau\right) - t\right\}^2 dt.$$

Integration is done for all possible values of parameter (MTF) $t$ out of $[0;\infty]$.

The formula for total variance $D$ is

$$D\left(\theta\left(n;R\right)\right) = \frac{1}{3}\sum_{\tau=10^3}^{10^5}\frac{1}{10}\sum_{n=1}^{10}\int_0^\infty \frac{1}{t^2} E\left\{\theta\left(n;R,\tau\right) - E\theta\left(n;R,\tau\right)\right\}^2 dt.$$

**Table 3. Results of substitution of suggested MTF estimates into functionals $A(\theta(n;R))$, $D(\theta(n;R))$ for the binomial test plan**

| Type of functional | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ |
|---|---|---|---|---|---|---|
| $A$ | 1513 | 11.27 | 11.26 | 11.09 | 11.01 | 10.59 |
| $D$ | 1.962 | 3.679 | 7.402 | 7.534 | 4.983 | 9.157 |
| $D / A$ | ≈0.01 | 0.32 | 0.65 | 0.67 | 0.45 | 0.86 |
| $C = D{\cdot}A$ | 2968 | 41.4 | 83.3 | 83.6 | 54.8 | 96.9 |

**Table 4. Results of substitution of suggested MTF estimates into functionals $A(\theta(n;R))$, $D(\theta(n;R))$ for the binomial test plan**

| Type of functional | $T_{10}$ | $T_{20}$ | $T_{30}$ | $T_{40}$ | $T_{50}$ | $T_{60}$ |
|---|---|---|---|---|---|---|
| $A$ | 5.67 | 4.62 | 5.34 | 5.27 | 5.03 | 4.85 |
| $D$ | 9.65 | 7.06 | 3.62 | 3.69 | 4.98 | 5.47 |
| $D / A$ | 1.70 | 1.52 | 0.67 | 0.70 | 0.99 | 1.12 |
| $C = D{\cdot}A$ | 54 | 32.61 | 19.33 | 19.44 | 25.04 | 26.52 |

Table 3 shows the results of substitution of the following MTF estimates into functionals *A(θ(n;R)), D(θ(n;R))*:

$$T_1 = ((n - R)\cdot\tau + R\cdot\tau / 2) / (R + 1);$$
$$T_2 = -\tau / \mathrm{Ln}(1 - (R + 1) / (n + 1));$$
$$T_3 = -\tau / \mathrm{Ln}(1 - p_1);$$
$$T_4 = -\tau/\mathrm{Ln}(1 - p_4), \text{ where } p_4 = u = (R + 1) / (n + 2), R = 0$$
$$\text{and } p_4 = p_0 = R / n, R > 0;$$
$$T_5 = -\tau / \mathrm{Ln}(1 - v(R,n,\gamma = 0.5);$$
$$T_6 = -\tau/\mathrm{Ln}(1 - v(R,n,\gamma = 0.62).$$

Out of Table 3 follows that, in accordance with the constructed criterion, all estimates are to be excluded from consideration, as the critical condition $D / A > 4$ is not fulfilled for them. However, due to the need to make a choice, estimate $T_6 = -\tau/\mathrm{Ln}(1 - v(R,n,\gamma = 0.62))$ with a minimum bias and maximum characteristic $D / A = 0.86$ should be considered conditionally bias-efficient.

The proposed MTF estimates for the binomial test plan are strongly biased, yet this bias can be reduced. The type of estimates will change slightly as follows:

$$T_{10} = 400 + 0.015\cdot\tau + \tau\cdot(n - R + R\cdot 0.02) / (R + 0.5));$$
$$T_{20} = 400 + 0.015\cdot\tau + (-\tau\cdot 0.7/\mathrm{Ln}(1 - (R + 0.4)/(n + 0.4)));$$
$$T_{30} = 400 + 0.015\cdot\tau + (-\tau\cdot 0.7 - \tau/\mathrm{Ln}(1 - p_1));$$
$$T_{40} = 400 + 0.015\cdot\tau + (-\tau\cdot 0.7/\mathrm{Ln}(1 - p_4)),$$
$$\text{where } p_4 = u=(R + 1)/(n + 2), R = 0$$
$$\text{and } p_4 = p_0 = R / n, R > 0;$$
$$T_{50} = 400 + 0.015\cdot\tau + (-\tau/\mathrm{Ln}(1 - v(R,n,\gamma = 0.5));$$
$$T_{60} = 400 + 0.015\cdot\tau + (-\tau\cdot 0.75/\mathrm{Ln}(1 - v(R,n,\gamma = 0.62)).$$

Variants of the suggested estimates with smaller biases are shown in Table 4.

Out of Table 4 follows that, in accordance with the constructed criterion, all estimates are to be excluded from consideration, as critical condition $D / A < 4$ is fulfilled. However, as a choice has to be made, the minimum bias estimate $T_{20} = 400 + 0.015\cdot\tau - \tau\cdot 0.7 / \mathrm{Ln}(1 - (R + 0.4) / / (n + 0.4))$ should be regarded as conditionally bias-efficient.

Further reducing the bias on the selected class of estimates would be quite challenging. In this case, the problem of bias reduction is solved by searching a wider class of estimates that includes a class of unbiased or similar estimates. Note that the closer an estimate is to unbiased (characteristic $A$ tends to zero), if it exists, its variance increases (see Table 1), below tending to the variance of an unbiased estimate, or decreases, above tending to the variance of an unbiased estimate, which forces their realizations to cluster around the true quantitative value of the estimated parameter from different sides similarly to the realizations of unbiased estimates. This fact follows directly from the Cramér-Rao inequality for biased estimates [5, f. 2.14.14]. Therefore, for estimates with a near-zero bias, condition $D / A > 4$ will always be fulfilled. It is important to note that the estimates of the selected class intended for finding bias-efficient estimates are to be strictly monotone with respect to all parameters ($R$, $\tau$, $n$).

## *NB* test plan. MTF

In what follows, the designations of the test plan are according to [6, 7]. For the $NB\tau$ plan, the number of observed failures ($r$) is a sufficient statistic [6, 7]. Let us denote a random number of failures as $R$, then, for a $NB\tau$ test plan, the random value $R$ (hereinafter referred to as r.v.) has a Poisson distribution $L(r;\Delta)$ with the parameter $\Delta = n\tau / T_0$, $n = N$ [4–7]. Then, by definition, $r$ is the realization of r.v. $R$. On the other hand, $R$ is the sum of r.v. $X_i$, each of which is a random number of failures of one of the $N$ tested products ($1 < i < n$). R.v. $X_i$ have a Poisson distribution with parameter $\Delta / n$

$$L(r;\Delta) = \sum_{k=0}^{X_1+\ldots+X_n=r} \exp\{-\Delta\}\cdot\frac{\Delta^k}{k!}. \qquad (3)$$

Let us use formula (3) and examine the properties of the parameter estimate $\Delta$ obtained from the equation

$$L(r;\Delta) = \sum_{k=0}^{r} \exp\{-\Delta\}\cdot\frac{\Delta^k}{k!} = 0,5 \text{ or}$$

$$\varepsilon(\Delta) = \ln(2) + \ln\left(\sum_{k=0}^{r}\frac{\Delta^k}{k!}\right) - \Delta. \qquad (4)$$

Minimizing the absolute value $\varepsilon(\Delta)$ in formula (4), with the required accuracy, we obtain the sought point estimate of

**Table 5. Results of substituting the suggested PNF estimates into functionals $A(\theta(n;R))$, $D(\theta(n;R))$ for the $NB\tau$ test plan.**

| Type of functional | $A$ | $D$ | $D/A$ | $C=D\cdot A$ |
|---|---|---|---|---|
| $T_{11} = 2.2n\tau$ if $R = 0$ and $T_{11} = n\tau / (R + 1 + 1 / R)$ if $R > 0$ | 0.214 | 3.93 | 18.36 | 0.841 |
| $T_{10} = 2.1n\tau$ if $R = 0$ and $T_{10} = n\tau / (R + 1.2)$ if $R > 0$ | 0.234 | 3.89 | 16.62 | 0.910 |
| $T_6 = 1.5n\tau / \Lambda$ if $R = 0$ and $T_6 = n\tau / (\Lambda + 0.5)$ if $R > 0$ | 0.234 | 3.98 | 17.00 | 0.931 |
| $T_1 = 2n\tau$ if $R = 0$ and $T_1 = n\tau / (R + 1)$ if $R > 0$ | 0.25 | 4.12 | 16.48 | 1.03 |
| $T_8 = n\tau / (R + 1) + n\tau 10^{-(R+0.5)} / (R + 0.5)$ | 0.28 | 4.00 | 14.28 | 1.134 |
| $T_7 = n\tau / (R + 1) + n\tau e^{-(R+1)} / (R + 1)$ [8] | 0.34 | 4.1 | 12.05 | 1.394 |
| $T_9 = n\tau / (R+0.7)$ | 0.364 | 4.43 | 12.17 | 1.61 |
| $T_5 = n\tau / \Lambda$ | 0.37 | 4.51 | 12.18 | 1.66 |
| $T_3 = n\tau / (R + 1)$ | 0.500 | 3.72 | 7.44 | 2.30 |
| $T_2 = 2n\tau$ if $R = 0$ and $T_2 = n\tau / R$ if $R > 0$ | 1.437 | 7.94 | 5.52 | 11.40 |
| $T_4 = 6n\tau$ if $R = 0$ and $T_4 = n\tau / (R + 0.5)$ if $R > 0$ | 5.36 | 10.21 | 1.90 | 54.72 |

**Table 6. Results of substituting the proposed PNF estimates into functionals $A$ $(\theta(m,g;R))$, $D(\theta(m,g;R))$ for the $NB\tau$ test plan**

| Type of functional | $e^{-g/T_1}$ | $e^{-g/T_2}$ | $e^{-g/T_3}$ | $e^{-g/T_4}$ | $e^{-g/T_5}$ | $e^{-g/T_9}$ | $e^{-g/T_7}$ |
|---|---|---|---|---|---|---|---|
| $A$ | 0.0346 | 0.0300 | 0.0641 | 0.0156 | 0.0410 | 0.0157 | 0.0458 |
| $D$ | 0.0987 | 0.1066 | 0.0740 | 0.1501 | 0.0876 | 0.1486 | 0.0851 |
| $D / A$ | 2.85 | 3.55 | 1.15 | 9.62 | 2.13 | 9.46 | 1.85 |
| $C = D\cdot A\cdot 10^3$ | 3.415 | 3.198 | 47.43 | 2.341 | 35.91 | 2.333 | 3.914 |

the Poisson parameter $\Lambda = \Lambda(R)$. Having estimate $\Lambda(R)$, we easily obtain the MTF estimate $T_5 = n\tau / \Lambda$. Let us examine the following MTF estimates:

– implicit estimate $T_5 = n\tau / \Lambda$;
– $T_1 = 2n\tau$ if $R = 0$ and $T_1 = n\tau / (R + 1)$ if $R > 0$;
– $T_2 = 2n\tau$ if $R = 0$ and $T_2 = n\tau / R$ if $R > 0$;
– $T_3 = n\tau / (R + 1)$;
– $T_4 = 6n\tau$ if $R = 0$ and $T_4 = n\tau / (R + 0.5)$ if $R > 0$;
– $T_6 = 1.5n\tau / \Lambda$ if $R = 0$ and $T_6 = n\tau / (\Lambda + 0.5)$ if $R > 0$;
– $T_7 = n\tau / (R + 1) + n\tau e^{-(R+1)} / (R + 1)$ [8];
– $T_8 = n\tau / (R + 1) + n\tau 10^{-(R+0.5)} / (R + 0.5)$;
– $T_9 = n\tau / (R + \beta(R))$ if $\beta = 0.7$;
– $T_{10} = 2.1n\tau$ if $R = 0$ and $T_{10} = n\tau / (R + 1.2)$ if $R > 0$;
– $T_{11} = 2.2n\tau$ if $R = 0$ and $T_{11} = n\tau / (R + 1 + 1 / R)$ if $R > 0$.

These bias estimates are based on functional ($T_0 = t$) [2]

$$A\left(\theta\left(n;R\right)\right) = \int_0^\infty \frac{1}{t^2}\left\{E\theta\left(n;R\right) - t\right\}^2 d\Delta.$$

The formula for the normalized variance $D$ is

$$D\left(\theta\left(n;R\right)\right) = \int_0^\infty \frac{1}{t^2} E\left\{\theta\left(n;R\right) - E\theta\left(n;R\right)\right\}^2 d\Delta.$$

Table 5 shows the results of substituting the suggested PNF estimates into functionals $A(\theta(n;R))$, $D(\theta(n;R))$ for the $NB\tau$ test plan.

Out of Table 5 follows that estimates $T_1$, $T_6$, $T_8$, $T_{10}$ and $T$ have approximately the same biases. The greatest difference between their values is $(0.28 - 0.214)\cdot 100 / 0.28 = 23$ %. In accordance with the suggested efficiency criterion of biased estimates, estimate $T_{11}$ with the minimum value of characteristic $C = 0.841$ must certainly be regarded as the most efficient.

Note that [2] provides the evidence of the fact that, in the class of estimates $T_R = n\tau / (R + 1) + n\tau f(R)$, estimate $T_1 = 2n\tau$ if $R = 0$ and $T_1 = n\tau / (R + 1)$ if $R > 0$ affords a minimum to functional $A = 0.25$. Let us prove that estimate $T_9 = n\tau / (R + \beta(R))$ does not belong to the class of estimates $T_R$, for which it suffices to represent estimate $T_9$ as $T_9 = n\tau(R + 2) / (R + 1)(R + \beta(R)) - n\tau / (R + 1)(R + \beta(R))$, hence the statement. The only estimate out of class $T_9$ that belongs to the class of estimates $T_R$ is the estimate of type

$T_9 = n\tau / (R + \beta(R)) = n\tau(R + 2) / (R + 1)(R + \beta(R)) - n\tau / (R + 1)(R + \beta(R)) = = n\tau(R + 2) / (R + 1)(R + 2) - n\tau / (R + 1)(R + 2) = n\tau / (R + 1) - n\tau / (R + 1)(R + 2)$

if $\beta(R) = 2$ (or if $\beta(R) = 0$, i.e., $T_2 = 2n\tau$ if $R = 0$ and $T_2 = n\tau / R = n\tau / (R + 1) + n\tau / R(R + 1)$ if $R > 0$). Where it is easy to see that $n\tau f(R) = -n\tau / (R + 1)(R + 2)$. Therefore, the occurrence of the values of the functional $A(T_{10}) = 0.234 < 0.25$ on estimate $T_{10}$ and $A(T_{11}) = 0.214 < 0.25$ on estimate $T_{11}$ is quite justified.

## *NB* test plan. Probability of no failure

Let us denote $m = n\tau$. Let us examine the PNF estimates for the time interval $g$ of the form $\theta(m,g;R) = \exp\{-g / T_i\}$, where $T_i$ is a certain MTF estimate (see Table 5). Instead of estimate $T_6$, let us examine estimate $T_9 = 4n\tau / \Lambda$ if $R = 0$ and $T_9 = n\tau / \Lambda$ if $R > 0$.

The comparison the PNF estimates in terms of the total bias value is based on a functional of the form [2]

$$A\left(\theta\right) = \frac{1}{3}\sum_{m=10^3}^{10^5}\frac{1}{10}\sum_{g=10^3}^{10^5}\int_0^\infty \frac{1}{t^2}\left\{E\theta\left(n;R,m,g\right) - \exp\left(-g\Delta/m\right)\right\}^2 d\Delta.$$

The formula for the normalized variance $D$ is

**Table 7. Results of calculating the PNF of Example 1 ($\tau = g$, R = 0)**

| $N = n$ | $p_{20} = \hat{}w(0.81;n)$, $R = 0$ and $p_{20} = R / n$, $R > 0$; $P_{20} = 1 - p_{20}(R = 0) = 1 - \hat{}w(\gamma = 0.81, R = 0)$ Binomial plan | $P_{NB\tau}(T_9) = \exp\{-g\Lambda / 4n\tau\}$, $g = \tau$, $R = 0$, $\Lambda(R) = 0.693148$ $NB\tau$ plan |
|---|---|---|
| 1 | 0.91 | 0.841 |
| 2 | 0.95 | 0.917 |
| 3 | 0.965 | 0.944 |
| 4 | 0.973 | 0.958 |
| 5 | 0.978 | 0.966 |
| 6 | 0.982 | 0.972 |
| 7 | 0.984 | 0.976 |
| 8 | 0.986 | 0.979 |
| 9 | 0.988 | 0.981 |
| 10 | 0.989 | 0.983 |

**Table 8. Results of MTF calculation for Example 2 ($\tau$ = 1000, $R$ = 0)**

| $N = n$ | $T_{20} = 400 + 0.015\cdot\tau +$ $+ (-\tau\cdot 0.7 / \mathrm{Ln}(1 - (R + 0.4)/(n + 0.4)))$ Binomial plan | $T_{11} = 2.2n\tau$, if $R = 0$ and $T_{11} = n\tau / (R + 1 + 1 / R)$, if $R > 0$ $NB\tau$ plan |
|---|---|---|
| 1 | 2495 | 2200 |
| 2 | 4254 | 4400 |
| 3 | 6008 | 6600 |
| 4 | 7759 | 8800 |
| 5 | 9511 | 11000 |
| 6 | 11261 | 13200 |
| 7 | 13012 | 15400 |
| 8 | 14762 | 17600 |
| 9 | 16512 | 19800 |
| 10 | 18263 | 22000 |

$$D(\theta) = \frac{1}{3}\sum_{m=10^3}^{10^5}\frac{1}{10}\sum_{g=10^3}^{10^5}\int_0^{\infty}\frac{1}{t^2}E\left\{\begin{matrix}\theta(n;R,m,g)-\\ -E\theta(n;R,m,g)\end{matrix}\right\}^2 d\Delta.$$

Table 6 shows the results of substituting the proposed PNF estimates into functionals $A$ (θ($m$,$g$;$R$)), $D$(θ($m$,$g$;$R$)) for the $NB\tau$ test plan.

Out of Table 6 follows that estimates $e^{-g/T_4}$ and $e^{-g/T_9}$ have approximately the same biases. Their values differ by (0.0157 – 0.0156) · 100/0.0157 = 0.63%. According to the proposed efficiency criterion of biased estimates, estimate $e^{-g/T_9}$ with the minimum value of characteristic C = 2.333 is to be regarded as the most efficient.

**Example 1.** In the course of dependability testing of a set of 1, 2, ..., 10 products, no failures occurred. It is required to estimate the PNF of the inspected batch of products using bias-efficient estimates for the binomial test plan and the test plan with recovery and limited test time. The calculation results are given in Table 7.

Out of Example 1 follows that for the binomial plan and the test plan with recovery and limited test time, in the setting of Example 1, the bias-efficient estimates differ (in case of $R = 0$). It is up to the test engineer to choose which estimates to use in this case.

**Example 2.** In the course of 1000-hour dependability tests of a set of 1, 2,..., 10 products, no failures occurred. It is required to estimate the MTF of the inspected batch of products using efficient estimates for the binomial test plan and the test plan with recovery and limited test time. The calculation results are given in Table 8.

Out of examples 1 and 2 follows that for the binomial plan and the test plan with recovery and limited test time, the outputs of bias-efficient estimates differ (case of $R = 0$). It is up to the test engineer to choose which estimates to use in this case.

## Afterword

A general approach is defined to constructing an efficiency criterion of biased estimates. For various test plans, performance criteria were constructed that allow unambiguously identifying the bias-efficient estimate out of those submitted. However, the problem of constructing (obtaining) efficient estimates (biased and not) with good statistical properties remains at the focus of the dependability theory and awaiting a solution.

## Conclusions

1) For the binomial plan and the test plan with recovery and limited test time, performance criteria were constructed that allow unambiguously identifying the bias-efficient estimate out of the submitted estimates.

2) Based on the constructed performance criteria for various test plans, bias-efficient estimates were selected out of the submitted ones.

## References

1. Yasnogorodsky R.M. [Probability theory and mathematical statistics. Textbook]. Saint Petersburg: Naukoyomkiye tekhnologii; 2019. (in Russ.)

2. Mikhailov V.S., Yurkov N.K. [Integral estimation in the dependability theory. Introduction and main findings]. TEKHNOSFERA; 2020. (in Russ.)

3. Pugachiov V.S. [Probability theory and mathematical statistics: textbook: 2-nd edition, corrected and extended]. Moscow: FIZMATLIT; 2002. (in Russ.)

4. Borovkov A.A. [Probability theory]. Moscow: Editorial URSS; 1999. (in Russ.)

5. Shulenin V.P. [Mathematical statistics. Part 1. Parametric statistics]. Tomsk: Izdatelstvo NTL; 2012. (in Russ.)

6. Barzilovich E.Yu., Beliaev Yu.K., Kashtanov V.A. et al. Gnedenko B.V., editor. [Matters of mathematical dependability theory]. Moscow: Radio i sviaz; 1983. (in Russ.)

7. Gnedenko B.V., Beliaev Yu.K., Soloviev A.D. [Mathematical methods in the dependability theory. Primary dependability characteristics and their statistical analysis: Second edition, corrected and extended]. Moscow: Knizhny dom LIBROKOM; 2013. (in Russ.)

8. Mikhailov V.S. [Efficient estimation of mean time to failure]. *Nadiozhnost i kontrol kachestva* 1988;9:6-11. (in Russ.)

## About the authors

**Viktor S. Mikhailov**, Lead Engineer, D.I. Mendeleev Central Research and Design Institute of Chemistry and Mechanics (FGUP CNIIHM). Address: 16a Nagatinskaya St., Moscow, 115487, Russian Federation, e-mail: Mvs1956@list.ru.

## The author's contribution

The author proposed a new criterion of bias-specific efficiency and used it to obtain bias-efficient estimates of various test plans.

## Conflict of interests

The author declares the absence of a conflict of interests.

# Correlations between states and events in the simulation of dependability using Markov processes

**Boris P. Zelentsov,** *Siberian State University of Telecommunications and Informatics, Novosibirsk, Russian Federation*
*zelentsovb@mail.ru*

*Boris P. Zelentsov*

**Abstract.** *The paper examines the correlations between states and events that are used in the construction of process diagrams that describe the dependability of items. Based on the constructed state and event diagram, input data is generated and the mathematical method is selected that is implemented in accordance with the problem at hand. The distinctive features and advantages of the matrix method are presented.* **Aim.** *To improve the simulation methods by clarifying the correlation between states and events and using matrix methods of calculation.* **Methods.** *The examined causal relationships between states and events allowed establishing correlations between them, i.e., an event can be the cause of a state change, then a state change is a consequence; a state can be the cause of an event, then an event is a consequence of a state. Under this approach, an event can cause a state change, while at the same time an event is a consequence of a state. The situation with states is similar. A state can be the cause of an event, while at the same time a state is the consequence of an event. It is also noted that a single state may cause a number of events, while an event can also cause a number of states. Examples of such correlations are given. It is noted that the duration of a state can be constant, random or zero. The examined correlations between states and events enable a substantiated construction of a diagram of states and transitions. A substantiated construction of a diagram of states and transitions results from a conceptual model, in which all states and events are given a physical and technical interpretation that transforms into a formal state-transition diagram. A special attention is given to the matrix methods that have a number of advantages, i.e., compactness and simplicity of converting the input characteristics into output characteristics, availability of standard software, use of verification procedures, feasibility of implementation using standard computer-based tools. The input data is also generated in matrix form. The paper indicates the characteristics of a state-transition diagram that can be calculated from the input data. Note is made of the use of methods based on semi-Markov processes. The author points out that, while using matrix methods, cycles should be generated. A relevant matter associated with the large number of states and the consequent problem of aggregation of states is touched upon. Two approaches to the aggregation of states are set forth that allow keeping the system's output characteristics unchanged.* **Results.** *A proposal is formulated for the construction of a dependability model involving a number of stages, i.e., definition of the goal of simulation with the indication of the used dependability indicators, description of the conceptual model, construction of a substantiated state-transition diagram, selection of the mathematical method, calculations, discussing the findings, conclusions and suggestions based on the performed simulation.* **Discussion and conclusions.** *A dependability model should take into consideration the causal relationships between states and events that are established based on the physical, as well as the engineering and technical features of the item. Taking these relationships into account, a state diagram is generated that enables initial data compilation. The matrix method is efficient and has a number of useful features. The above considerations are methodological in their nature. They can be helpful for generating dependability models of technical systems and studying the dependability theory in educational institutions.*

**Keywords:** *item dependability, state transition diagram, matrix methods for Markov process simulation.*

**For citation:** *Zelentsov B.P. Correlations between states and events in the simulation of dependability using Markov processes. Dependability 2022;1: 38-43. https://doi.org/10.21683/1729-2646-2022-22-1-38-43*

**Received on:** *26.11.2021 /* **Revised on:** *21.01.2022 /* **For printing:** *18.03.2022.*

## Introduction

One of the problems of complex systems research consists in constructing such models of actual systems that are suitable for theoretical and experimental study of their properties. At the same time, item dependability models are to adequately represent actual processes in existing systems. Mathematical simulation is the most common and promising method for studying complex systems that allows conducting research at the design stage, solving analysis and synthesis problems, predicting the quality and efficiency of system operation, substantiating the required or optimal structure when designing new and improving existing systems and correctly interpreting statistical data.

Normally, dependability models are constructed on the basis of a discrete set of states, transitions between which occur in continuous time. Such processes are graphically represented as a state-transition diagram. The paper examines the causal correlations between states and events (transitions between states) that form a diagram. Further use of the diagram is associated with the selection of a mathematical method and calculations in accordance with the defined dependability process simulation objective.

The considerations presented in the paper are methodological in their nature and reflect the author's individual opinion as regards technical system dependability simulation.

## Source overview

State standard [1] establishes guidelines for the application of Markov methods for simulating the dependability of systems with discrete states in continuous time. Markov methods can be used for dependability simulation of various technical systems. When applying Markov analysis, a state-transition diagram is used, which is a graphical representation of the conceptual model and simulates the behaviour of the system over time. The rules for constructing state and transition diagrams are described and examples of applying these rules are given. Accordingly, the state space analysis is used. State space analysis is used in the study of the dependability of various system architectures, i.e., redundant systems, systems with complex maintenance strategies, etc. It is stressed that the key problem solved by Markov analysis is the correct construction of the state space diagram. Additionally, a homogeneous Markov process is completely characterized by a transfer rate matrix.

Standard [1] further notes the advantages of Markov analysis, i.e., the ability to simulate various maintenance strategies. The assumption of constant recovery rate is to be substantiated, if the mean recovery time is not negligible compared to the corresponding mean time to failure. It is also noted that the use of Markov analysis requires special precautions associated with the increasing number of system states. In case of a large number of states and transitions, the probability of errors and distortions grows. Additionally, the computational methods also become more complex and may require the use of special software. For practical reasons, it is allowed excluding states with very low probabilities from the model of system operation.

There are numerous publications in the Russian and foreign literature dedicated to the study of the properties of Markov processes in discrete and continuous time, as well as their application to simulating probabilistic systems of various purpose, e.g., [2 – 4]. Of particular note is the widespread use of the Markov process theory involving state transition diagrams employed for solving dependability-related problems. Thus, in [5], using continuous-time Markov processes, functional models of recoverable and non-recoverable systems were developed, methods for calculating dependability indicators (availability coefficient, mean time between failures, etc.) were given for various conditions associated with equipment specificity. In particular, [5] sets forth dependability models of systems tested at random periods. In [6], methods are examined for calculating dependability based on Markov processes taking into account the completeness of testing. The above works examine Markov models, in which the future state of a system does not depend on the evolution of states up to the current one.

When Markov methods are used, mathematical models of dependability clearly show the process of state transition of the item (element, system). This process reflects the actual processes within technical systems. First of all, let us define the term "state of item".

The technical state of an item (technical state, state of item, state) is a set of the item's properties that are subject to change during its manufacture, operation, transportation and storage that are characterized by documented parameter values and/or qualitative characteristics [7]. Out of the above definition follows that a state is characterized by the time elapsed from the beginning of the state to its end, while the beginning and the end of the state are events. It should be noted that [1] uses the term "state transition", which is synonymous with the term "event". The term "event" (more precisely, "random event") is a basic term of probability theory. Further, we will use both terms that have the same meaning.

In [8], recovery is considered as a process and an event that consists in an item's transition from a non-operable to an operable state. Out of this definition, as well as the definition of recovery rate, follows that what is meant here is an event associated with the completion of recovery. This understanding involves that recovery is associated with two events, i.e., the beginning of recovery and the completion of recovery. Mathematical models often use the "recovery rate" parameter that characterises the completion of recovery, provided that the "beginning of recovery" event has taken place. Note that a "beginning of recovery" event may occur under various conditions, i.e., immediately after a failure, with a delay due to limited recovery capabilities, operation of the item in nonoperable state after a hidden failure, etc.

## Correlation between the terms «state» and «event»

For a common understanding of the process of state transition, let us note the causal relationships between the terms "state" and "event". Mathematical models generally assume that events (state transitions) occur instantaneously.

An event can cause a state transition, whereas the changed state is a consequence of the event. Examples:

– failure causes the operable state to change into inoperable state, i.e., the inoperable state is a consequence of the failure;

– completion of recovery causes transition from recovery to operable state, i.e., the operable state is a consequence of the completion of recovery.

– failure detection can cause the start of recovery or blocking of an item, i.e., the start of recovery or blocking of the item are consequences of failure detection.

A state can be the cause of an event. In this case, an event is a consequence of a state. Examples:

– using an item for its intended purpose is the cause of the failure, i.e., the failure is a consequence of the item being used;

– repair (restoration) of an item is the cause of the "start of operation" or "start of storage" events, i.e., "start of operation" or "start of storage" are the consequences of repair (restoration);

– an operation of incorrect technical condition inspection can cause such events as a type I inspection error and type II inspection error, i.e., type I and II inspection errors are the consequences of an incorrect inspection operation.

Thus, an event can be the cause of a state change, while at the same time being a consequence of a state. The situation is similar with a state. It may cause an event, while at the same time being a consequence of an event.

It should be noted that the same event is normally the end of one state and the beginning of another. Therefore, events may have different names depending on what state they are assigned to when the model is developed.

In the examples given in standard [1], the beginning of recovery coincides with the item's failure. In actual systems, different situations may take place. Transition into the recovery state may occur with the following events:

– recovery after waiting in queue (restricted recovery);

– detection of a hidden failure during item diagnostics;

– inspection error causes false recovery of operable item.

Cases may be noted, whereas a failure does not cause recovery at the moment of failure:

– a hidden failure occurs;

– an explicit failure occurs and the item is queued for recovery (restricted recovery);

– upon verification of the technical state, no failure was detected.

It should be kept in mind that, within a single state, several events may occur, e.g., when an item is used for its intended purpose, hidden failures, explicit failures, pre-failures, damage may occur. An event can also cause a number of states: a technical state inspection operation may be valid or may cause type I and II inspection errors.

Thus, states and events are temporally associated with causal relationships. The new state of an item (an element or an entire system) is a consequence of a certain event, while any event is a consequence of the preceding state. Each state corresponds to two events, the beginning of the state and its completion.

The duration of a state may be of three types: constant (fixed, regular, deterministic), random or zero. If the state duration is zero, the beginning and completion of the state coincide. Such state can be called both a state, and an event.

An example of a constant-duration state is diagnostics of an item with a constant diagnostic time. An example of a random-duration state is the random time of item recovery (repair). An example of a zero-duration state: item diagnostics operation is performed within a time that is significantly shorter than the duration of other states, therefore, in models, the duration of the diagnostic operation is assumed to be zero.

When building dependability models, the system features are taken into account that cannot be covered in a single paper. However, the following factors can be noted:

– presence of hidden failures, explicit failures, pre-failures;

– application of a technical state monitoring system;

– use of maintenance system;

– maintenance with periodic or continuous inspection;

– possibility of type I and II inspection errors

and much more.

Hence, the state transition $s_i \rightarrow s_j$ is an event that is a consequence of state $s_i$ and the cause of state $s_j$, i.e., state $s_j$ is a consequence of this event. It should be noted that, in [1], state transitions (events) are often given with not due explanation. The expressions "state transition", "transition from one state to another" and "return from one state to another" are used.

In order to define a substantiated diagram, a complete description of the states and transitions (events) should be done. For each state, the following is to be specified:

1) name;

2) transition into a state as a consequence of an event;

3) termination of a state as a consequence of another event.

For each event (transition), the following should be specified:

1) name;

2) the state that causes the event;

3) the state that is a consequence of the event.

Mathematical methods based on Markov chains and processes use various types of states associated with diagrams, i.e., neighbouring states, reachable states, communicating states, isolated state, absorbing state, non-

essential and essential state, recurrent and non-recurrent state. The specificity of diagrams is expressed in the use of different types of sets and subsets of states, i.e., associated set, isolated and non-isolated set, transitive subset, subset of essential and non-essential states, ergodic set. These terms are extensively covered in academic and research literature.

When formulating the specific features of state-transition diagrams, "terminological perfection" is to be ensured. It comes down to the non-ambiguity of the terms, consistency within themselves and with state standards [9].

## Matrix methods for Markov process simulation

Standard [1] notes that a homogeneous continuous-time Markov process is fully characterized by a rate matrix that is used for constructing and solving a matrix differential equation that allows finding the probabilities of states or events as function of time. It also refers to a method based on algebraic equations for calculating the limit probabilities of states. Hybrid models are mentioned, i.e., fault tree analysis, dependability structure diagram, Petri nets.

The matrix method is one of the most efficient mathematical methods for simulating Markov processes. The initial data of the matrix method for a continuous-time process are in the rate matrix. The use of constant rates for the time of occurrence of events or the duration of states is to be substantiated.

There are numerous Russian and foreign publications associated with studying the properties of discrete-time and continuous-time Markov processes and their application for simulating various probabilistic systems, e.g., for dependability-related purposes. In [10], matrix methods are presented for simulating discrete-time and continuous-time Markov processes that allow calculating probabilistic, temporal and frequency characteristics of states and subsets of states associated with the specificity of the examined system. Those characteristics are easily converted into dependability indicators, such as probability of no failure, mean time to failure, failure rate, availability and unavailability coefficients, etc.

The interest in the matrix methods is due to their advantages, i.e., compactness and simplicity of converting the input characteristics into output characteristics, availability of standard software, feasibility of implementation using state-of-the-art computer-based tools. It should be noted that matrix methods are classified as numerical analytical methods, i.e., applicable for both numerical calculations, and analytical studies.

Let us briefly represent the matrix method for a continuous-time process described in [10]. A transfer rate matrix is compiled based on the state transition diagram. The rate matrix can be used to calculate the following:

– state probabilities as functions of time under any initial state and specified initial distribution by defining and solving a matrix differential equation;

– limiting state probabilities for an ergodic process according to two analytical formulas, i.e., using matrix inversion and determinants;

– probabilities of being in a subset of states;

– mean time spent in a subset of states by inverting the rate matrix;

– variance of the time spent by the system in a subset of states using an inverted rate matrix and the matrix made on the basis of the initial distribution of state probabilities.

Depending on the chosen dependability model, the process of state transition may be either discrete-time, or continuous-time. In dependability, mathematical models are most often continuous-time. Similar procedures and characteristics for the discrete-time process are described in [10].

For the purpose of simulating processes that describe the dependability of systems, semi-Markov process-based methods can be used. The difference between a semi-Markov process and a discrete-time and continuous-time Markov process is that transitions are considered not at discrete moments of time and not in continuous time, but at moments of exiting states (or moments of state transition).

The process of state transition in a semi-Markov process is defined by the so-called probabilities of passing. A passing probability matrix can be defined based on a transition probability matrix for a discrete-time process and based on a rate matrix for a continuous-time process. The probabilities of passing a semi-Markov process can also be calculated for cases with a constant or random duration with an unknown distribution.

Probabilities of passing do not contain information on the duration of states. If such characteristics are required for simulating the system, those are defined together with the probabilities of passing as initial data. Such input data may include, e.g., the mean times in states after entering.

This approach involves that an event is a dependent event if it is caused by a certain state. In this case, events should be characterized with conditional probabilities. In this context, the events reflected in dependability models may be deterministic or random. A deterministic event is the only event that is a consequence of a state. Its conditional probability is 1. Random events include those whose conditional probabilities are below 1.

An example of a random event is a failure of an item that has been used for its intended purpose for some time with no failure during that period, i.e., opposite random events. An example of a deterministic event is commencement of the use of an operable item after recovery.

Based on the semi-Markov process, the following can be calculated:

– expected number of times in the states of a subset called mean relative state rates and represented in matrix form;

– mean time in a subset of states based on the mean relative state rates.

Computation and verification should be implemented using computer mathematics. In particular, ready-made formulas are used to calculate states as functions of time, limit probabilities of states, mean time and variance of time spent in a subset of states and other characteristics. The use of computer mathematics allows reducing the relevance of many states.

Computational procedures based on computer mathematics can be performed both numerically and analytically.

## Relevant problems solved using the matrix method

Let us briefly mention two problems solved using the matrix method, i.e., aggregation of states and cyclic system operation.

In standard [1], it is noted that if the number of states is large, difficulties associated with possible errors and distortions may arise. At the same time, it is allowed to exclude from the system operation model the very-low-probability states. Research literature notes that reducing the number of states by discarding unlikely ones may cause significant errors in the system's output characteristics. Therefore, such approaches are to include error calculation, i.e., with no error calculating such approach is irrelevant.

When using the matrix methods, the difficulties associated with the large number of states are quite easily addresses by means of verification procedures for both initial data generation, and calculation. Verification procedures allow quickly finding data input and computation errors; therefore, verification procedures help improve the efficiency of mathematical methods.

In [10], two approaches are described to the problem of aggregation of states, i.e., aggregation using truncation of matrix characteristics and frequency-based. These approaches use the matrix method and allow keeping the system's output characteristics unchanged.

The operation of long-term use systems is associated with repeating cycles. That fact was one of the reasons that determined the development of the model of cyclic system operation [10]. The model of cyclic operation describes transitions between subsets of states by means of manipulations with matrices. This approach allowed calculating system characteristics in transient and stationary modes. Formulas are given for calculating the mean times in subsets of states in the transient and stationary modes, as well as for limit probabilities of subsets.

## Results

Out of the above reasoning follows that dependability simulation should be carried out in the following stages.

1. Definition of the simulation objective specifying the employed dependability indicators.

2. Presentation of the conceptual model that contains the initial representation of the item. It sets out the physical and operational features of the facility and provides an engineering description of the processes in terms of dependability.

3. Construction of the state transition diagram based on the conceptual model.

4. Selection of the mathematical method. There should be a clear understanding of the source data and the output characteristics obtained using the method.

5. Calculations.

6. Findings, suggestions and conclusions based on the conducted simulation.

## Conclusions

1. A proposal was defined for dependability model preparation that contains the goal, conceptual model, state transition diagram, mathematical method, calculations and conclusions.

2. When constructing state-transition diagrams, the causality relationships between states and events should be taken into consideration. The establishment of these relationships is based on the physical and engineering features of the examined systems.

3. One of the efficient methods of Markov process simulation is the matrix method that has a number of useful features, i.e., compactness and simplicity of transformations, availability of standard software.

4. The above matrix method allows constructing analytical and algorithmic models of equipment operation as part of various technical systems.

5. The matrix method provides verification procedures for every stage of the simulation for the purpose of eliminating errors and distortions in the input data generation and computational procedure implementation.

6. Matrix manipulations should be performed using such modern software tools as Mathcad and Matlab.

The above materials can be used as guidelines for efficient construction of dependability models of technical systems and in studying the dependability theory in educational institutions.

## References

1. GOST R IEC 61165-2019. Dependability in technics. Application of Markov techniques. Moscow: Standartinform; 2019. (in Russ.)

2. Mataltsky M.A., Khatskevich G.A. [Probability theory and mathematical statistics]. Minsk: Vyshaishaya shkola; 2017. (in Russ.)

3. Birolini A. Reliability Engineering. Theory and Practice, 8th ed. Springer; 2017.

4. Knill O. Probability and Stochastic Processes with Applications. Overseas Press, India Private Limited; 2009.

5. Ushakov I.A. [Course of systems dependability theory]. Moscow: Mir; 2008.

6. Viktorova V.S., Stepaniants A.S. [Models and methods of technical system dependability calculation]. Moscow: Lenand; 2016. (in Russ.)

7. GOST 18322-2016. Maintenance and repair system of engineering. Terms and definitions. Moscow: Standartinform; 2017. (in Russ.)

8. GOST 27.002-2015. Dependability in technics. Terms and definitions. Moscow: Standartinform; 2016. (in Russ.)

9. Zelentsov B.P. Comments on the contents of the dependability terminology standard. *Dependability* 2021;1: 34-37.

10. Zelentsov B.P. [Matrix methods of simulating homogeneous Markov processes]. Academic Publishing; 2017. (in Russ.)

## About the author

**Boris P. Zelentsov**, Doctor of Engineering, Professor, Department of Further Mathematics, Siberian State University of Telecommunications and Information Sciences, Novosibirsk, Russian Federation, e-mail: zelentsovb@mail.ru

## The author's contribution

The author examined the possible causal relationships between states and events and provided examples of a significant dependence of the future on the past in dependability process simulation.

## Conflict of interests

The author declares the absence of a conflict of interests.

*Original article*
*https://doi.org/10.21683/1729-2646-2022-22-1-44-51*

# Analysis of the functional dependability of underground gas storage compressor stations in cases when actual performance indicators deviate from the design values

**Vorontsov M.A.[1,2*], Grachiov A.S.[1**], Grachiova A.O.[1,2], Kirkin M.A.[3], Melnikova A.V.[1]**
*[1]Gazprom VNIIGAZ, Razvilka, Russian Federation, [2]Gubkin University, Moscow , Russian Federation, [3]Gazprom, Saint Petersburg, Russian Federation*
**m_vorontsov@vniigaz.gazprom.ru*
***grachev.anatoliy@yandex.ru*


*Vorontsov M.A.*


*Grachiov A.S.*


*Grachiova A.O.*


*Kirkin M.A.*


*Melnikova A.V.*

**Abstract. Aim.** *The paper examined the matter of assessment of the functional dependability of compressor stations (CS) of underground gas storage (UGS) facilities. A definition of CS functional dependability and guidelines for its assessment were proposed.* **Methods.** *Design calculation of compressor stations, scenario analysis.* **Results.** *The paper presents: a) a definition, indicators of CS functional dependability and guidelines for its assessment; b) an example of the guidelines application for UGS CS; c) a comparative analysis of UGS CS functional dependability in a number of various versions: use of single-unit and two-unit centrifugal compressors as part of gas turbine gas pumping units for two-stage compression with intercooling.* **Conclusion.** *The paper shows the requirement to analyse the functional dependability of various versions of UGS CS for the purpose of identifying the most rational option that ensures unconditional performance of the key UGS CS function under uncertain initial design data.*

**Keywords:** *compressor station, UGS compressor station, primary target process functions, gas compressor unit, centrifugal compressor, two-unit centrifugal compressor, functional dependability.*

**For citation:** *Vorontsov M.A., Grachiov A.S., Grachiova A.O., Kirkin M.A., Melnikova A.V. Analysis of the functional dependability of underground gas storage compressor stations in cases when actual performance indicators deviate from the design values. Dependability 2022;1: 44-51. https://doi.org/10.21683/1729-2646-2022-22-1-44-51*

**Received on***: 23.10.2021 /* **Upon revision:** *13.02.2022 /* **For printing:** *18.03.2022.*

## 1. Introduction

In order to ensure peak consumption volumes, as well as flexibility and dependability of gas supply, UGS are created for the purpose of collection and storage of natural gas and its subsequent prompt delivery to consumers as required [1]. There is experience in creating UGS facilities based on depleted hydrocarbon deposits, water-bearing formations or salt caverns. Regardless of the type of underground reservoir, each UGS facility uses compressor stations required for ensuring the temperature and pressure conditions for gas withdrawal and injection into the formation.

CS is a critical element of the UGS process system, which defines the high requirements for its dependability and efficiency. That is largely due to the fact that a CS is an "active" process facility, as it ensures increased gas pressure, whereas the other UGS facilities are primarily "clients" of the target value of pressure. For example, the operation of gas purification systems and injection process require the design value of pressure, while gas collection systems cause losses of pressure that need to be amended, etc. Therefore, the CS UGS is the only facility, modifying the operating modes of which the changes in the operating conditions can be compensated for the purpose of enabling the planned performance indicators.

In general, as regards ensuring the dependability of technical systems, two classes of tasks are to be distinguished. The first class includes problems of structural dependability. Those are solved using the methods of the traditional dependability theory that studies the processes of item failure and restoration (of an entire technical system and its elements). The second class includes problems related to the analysis of functional dependability (FD) of technical systems that characterizes the reliability of a system's target functions when actual operating conditions deviate from the design values [2]. The analysis of structural dependability is based on classical methods of statistical analysis, which significantly limits its applicability to complex systems, while the FD is analysed using modern methods of computer simulation, queuing theory, machine learning, etc.

The methods of FD analysis of complex technical systems are used (and are being actively developed) for identifying the most efficient technical and process-specific solutions in the power industry, including nuclear [3], development of information management and operating systems [4], security system [5], etc. Problems, whose solution involves evaluating the FD of gas industry process systems, are examined in a number of papers, including those by Gazprom VNIIGAZ [6, 7].

This paper examines the solution of the problem of ensuring reliable performance of primary UGS CS target functions subject to uncertain operational indicators (OI) its design is based upon (temperature and pressure, consumption parameters, etc.). This problem belongs to the second class of problems of technical system dependability. To solve it, a methodological approach has been developed that consists in identifying the UGS CS FD and indicators for its quantification. The paper presents the methodological approach and an example of its application for comparing the FD of equipment options of UGS CS with a gas turbine-driven gas compressor unit (GCU) and various types of centrifugal compressors (CS), single-section single-unit and single-unit two-section.

## 2. Key provisions of the methodological approach, concepts and terms

UGS CS FD research includes the following main activities:

– definition of the list of functional failures;

– identification of UGS CS FD factors (threats) (evaluation of the probability of their future occurrence);

– development of a system of UGS CS FD indicators;

– development of methods for FD indicator calculation;

– definition of UGS CS FD requirements.

A functional failure (FF) should be identified as nonperformance (entirely or in part) of a system's primary functions, while FD factors should be defined as the causes and events that entail an FF, i.e. FD threats. Therefore, identifying an FF and UGS CS FD threats (factors) requires defining the meaning and formalizing the concept of UGS CS FD.

## 3. Functional dependability and primary process-specific functions of compressor stations of underground gas storage facilities

In the academic community, there is still no generally accepted definition of the term "functional reliability/dependability", and, consequently, there is still no single understanding of the subject and goals of FD analysis of technical systems.

In this paper, by analogy with [2, 8 – 11], it is assumed that UGS CS FD is the ability of CS to ensures the performance of its primary target process functions (PTPF) when the primary OI (pressure at the output and input of the compressor station, volume and/or composition of the compressed natural gas, its temperature at the input of the compressor station, etc.) deviate from the design values.

The PTPF of UGS CS is ensuring that the pressure of a given amount of gas increases to the values required for: a) injection of the required amount of gas into UGS; b) withdrawal of the required amount of gas in autumn and winter for delivery to the main gas line (MG).

## 4. Functional failures and factors of functional dependability of compressor stations of underground gas storage facilities

As it was mentioned above, FF is the impossibility of performing a system's primary process functions. Taking into account the above definition of PTPF, the FF of UGS CS are:

– impossibility to increase the pressure for the required amounts of gas for withdrawal or injection to the design values;

– maintaining PTPF in the course of long CS operation in suboptimal operating modes, e.g., in case of the low polytropic efficiency of the compression process, deviation from the nominal value by more than 20 % (rel.) or if backup GPUs need to be put in operation with violation of regulatory requirements for redundancy.

It is important to note that, in terms of FD evaluation, only those cases are of interest when the above FFs occur with fully operable equipment, i.e., not due to accidents or GPU failures, etc., but due to changes in the operating conditions. Accordingly, for UGS CS, those are deviations of consumption and temperature and pressure indicators from the design values that may be due to the following causes:

– decreased pressure in the UGS due to a reduced amount of stored gas, e.g., when the amount of withdrawn gas is higher than planned;

– increased pressure in the UGS due to an increased amount of stored gas, e.g., when the amount of injected gas is higher than planned;

– decreased or increased pressure in the UGS due to changes or better definition of the structure and properties of the UGS reservoir;

– changed temperature of gas at the input to the UGS CS;

– changing MG operating mode;

– increased rate of gas withdrawal, etc.

The above events are among UGS CS FD factors, i.e., threats that may cause FFs, and their elimination may require additional costs, e.g., associated with the reconstruction of the CS (deployment of additional GPUs, re-wheeling, etc.) Therefore, when choosing the UGS CS equipment option, the results of the FD analysis should be taken into consideration, since that allows determining the facility alternatives that ensure the PTPF performance within a wide range of consumption and pressure parameters with no additional material costs. Quantitative FD analysis requires a system of UGS CS FD indicators.

## 5. UGS CS functional dependability indicators

As primary UGS CS FD indicators, parameters are adopted that allow quantifying the consequences of changes in the CS operating mode as an FF occurs:

– required deployment of backup GCUs (with no violation of redundancy requirements);

– required installation and commissioning of additional GCUs;

– CS performance margins when operating with the predefined number of GCUs and in compliance with the GCU redundancy requirements;

– variation of required fuel gas.

The requirement to deploy backup GCU with no violation of redundancy requirements characterizes the CS's ability to maintain the PTPF subject to consumption, temperature and pressure indicators deviating from the design values without additional costs for the installation and commissioning of equipment. The required installation and commissioning of additional GCUs indicates the requirement to overhaul the CS, which means additional capital investment.

The CS performance margin characterizes the difference between the design performance and the maximum possible performance if pressure deviates from the design values (characterizes the existence of a performance margin). In other words, it indicates the feasibility of intensifying the scope of useful UGS performance.

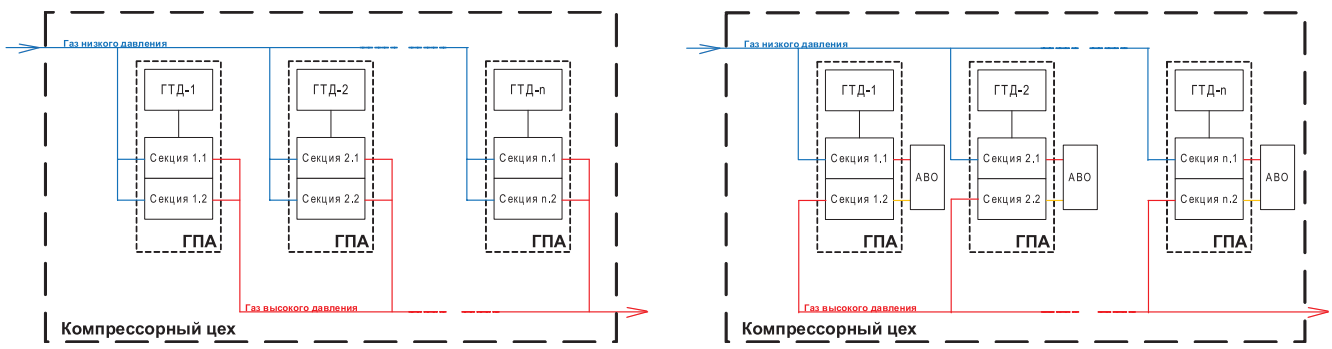
The variation in the fuel gas demand allows comparing the UGS CS equipment options based on the variation of the energy efficiency indicators of the gas compression process in changing operating conditions. That means that it characterizes the increase in operating costs.

The presented system of indicators is the foundation of the developed methodological approach to assessing the UGS CS FD. Each of the system's indicators characterizes both the process-specific, and economic aspects of UGS CS operation.

## 6. Methodological approach and an example of its implementation

Quantifying the FD of various design and engineering solutions in terms of UGS CS equipment requires performing the following calculations and analytical studies for various station equipment options:

1. Quantitative assessment of possible deviation of the consumption and temperature and pressure parameters of




a) b)

Fig. 1. UGS CS option with a GCU with a two-unit CC: a, single-stage compression (parallel operation of units); b, two-stage compression (sequential operation of units); where ACU is air cooler unit

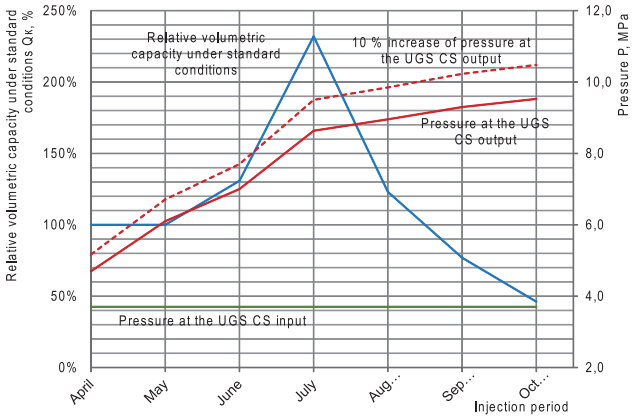



Fig. 2. Graph of performance dynamics (relative values) and gas pressure at the CS inlet and outlet during UGS injection

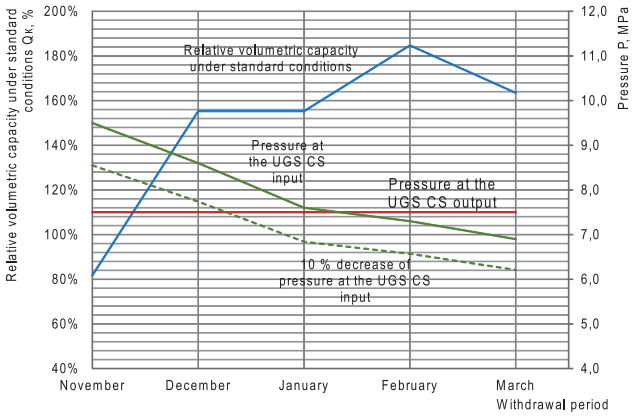



Fig. 3. Graph of productivity dynamics (relative values) and gas pressure at the CS inlet and outlet during UGS withdrawal

the actual operating modes from the design values. This stage is to be carried out in cooperation with geology experts involved with the examined UGS, gas transport experts, etc.

2. Calculations of UGS CS operation modes for various equipment options:

– calculation of design operation modes;

– calculation of operating modes in cases of OI deviation from the design values;

– calculation of quantitative FD indicators through comparative analysis of the UGS CS operating mode calculations for cases when design OI values are and are not fulfilled.

3. Comparative analysis of UGS CS FD indicators for various equipment options for the purpose of developing guidelines for primary station design solutions.

Below is an example of implementation of the examined methodological approach for UGS CS with gas-turbine GCUs with single-unit and two-unit CCs.

In case single-unit CCs are used, the UGS CS includes one compressor shop, whose GCUs operate in parallel.

In case two-unit CCs are used, the UGS CS includes one compressor shop. Additionally, special process piping of the CC units is foreseen for the purpose of ensuring their sequential or parallel operation (see Fig. 1).

The initial data used in the calculations are shown in Fig. 2 and 3 and in Tables 1 to 4. It is accepted that in the course of the injection period, gas comes from the main gas line with a constant pressure of 3.7 MPa, while in case of gas withdrawal from UGS it is required to supply gas to a pipeline with the working pressure of 7.5 MPa.

Scenarios of UGS CS operation under the design operating conditions and FD factors (threats) caused by geological risks were considered. It is accepted that operation in off-design conditions causes the requirement for a 10% increase of the UGS CS output pressure when injecting gas and a 10% reduction of the UGS CS input pressure when withdrawing gas.

The calculation data are shown in Fig. 4 to 6 and Table 5.

**Table 1. Temperature and pressure parameters of the gas injected into UGS**

| Month | Gas pressure at the station input $P_1$, MPa | Gas temperature at the station input $T_1$, K | Gas pressure at the station output $P_2$, MPa (per design / deviates from design) |
|---|---|---|---|
| April | 3.70. | 283.15 | 4.70 / 5.17 |
| May | | | 6.10 / 6.71 |
| June | | | 7.00 / 7.70 |
| July | | | 8.64 / 9.50 |
| August | | | 8.95 / 9.85 |
| September | | | 9.30 / 10.23 |
| October | | | 9.53 / 10.48 |

**Table 2. Temperature and pressure parameters of the gas withdrawn from UGS**

| Month | Gas pressure at the station input $P_1$, MPa (per design / deviates from design) | Gas temperature at the station input $T_1$, K | Gas pressure at the station output $P_2$, MPa |
|---|---|---|---|
| November | 9.50 / 8.55 | 283.15 | 7.50 |
| December | 8.60 / 7.74 | | |
| January | 7.60 / 6.84 | | |
| February | 7.30 / 6.57 | | |
| March | 6.90 / 6.21 | | |

**Table 3. Compressible gas composition**

| No. | Name | Composition | Molar concentration, % |
|---|---|---|---|
| 1 | Methane | $CH_4$ | 98.511 |
| 2 | Ethane | $C_2H_6$ | 0.360 |
| 3 | Propane | $C_3H_8$ | 0.066 |
| 4 | n-Butane | n-$C_4H_{12}$ | 0.013 |
| 5 | n-Pentane | n-$C_5H_{12}$ | 0.028 |
| 6 | Nitrogen | $N_2$ | 0.782 |
| 7 | Carbon dioxide | $CO_2$ | 0.280 |

**Table 4. Primary technical characteristics of GCU with single and two-unit CCU**

| Primary characteristics | GCU with single-unit CC | GCU with two-unit CC | |
|---|---|---|---|
| Unit power GCU $N_{e0}$, MW | 8.0 | 8.0 | 10.0 |
| Mechanical efficiency $\eta_{MEX}$, % | 98 | 96 | 96 |
| Pressure loss between compression stages $\Delta\rho_{cm}$, MPa | –* | 0.50 | 0.50 |
| Nominal pressure ratio of one unit | 3.0 | 1.7 | 1.7 |

*Note: * – operation involves one compression stage*

**Table 5. Calculated indicators of UGS CS FD**

| Quantitative indicators of FD | Gas injection into UGS ($P_{out}$ increased by 10%) | | | Withdrawal from UGS ($P_{in}$ reduced by 10%) | | |
|---|---|---|---|---|---|---|
| | Single-unit | Two-unit | | Single-unit | Two-unit | |
| Unit power of GCU, MW | 8 | 8 | 10 | 8 | 8 | 10 |
| Required backup GCUs, pcs | 1 | – | – | – | – | – |
| Required installation and commissioning of additional GCUs, pcs | – | – | – | – | – | – |
| Margin of cubic capacity under standard conditions $q_{\Delta}$, mln $m^3$/day | -0.39*...27.42 | 0.11...34.51 | 6.94...57.96 | 21.27...23.56 | 70.65...72.95 | 91.83...94.13 |
| Increased fuel gas consumption as compared to the reference case $\Delta q_{fg}$, mln $m^3$ | 4.07 (13.9%) | 2.75 (10.7%) | 2.87 (10.9%) | 3.37 (32.4%) | 1.50 (32.5%) | 1.64 (32.6%) |

*Note: * The minus symbol indicates that the station will not be able to provide the specified gas flow rate with the designed number of operating GCUs. Implementing this mode will require increasing the number of operating units compared to the design values.*

The analysis of the calculation data per the UGS CS FD assessment system shows that:

1. Deployment of backup GCUs with violation of redundancy requirements is only necessary if the pressure at the CS outlet increases by 10 % if the CS is equipped with a UGS GCU with a single-unit CC (see Fig. 4a, Table 5).

2. Installation and commissioning of additional GCUs is not required in the above situations.

3. The minimal margin of CS volume efficiency when operating with the predefined number of GCUs is:

– minus 2.2%[1] when UGS CS is equipped with a GCU with a single-unit CC with the unit capacity of 8.0 MW (see Fig. 6a, Table 5);

– 0.6% when UGS CS is equipped with a GCU with a two-unit CC with the unit capacity of 8.0 MW (see Fig. 6a, Table 5);

– 38.4% when UGS CS is equipped with a GCU with a two-unit CC with the unit capacity of 10.0 MW (see Fig. 6a, Table 5);

4. If pressure deviates 10 % from the design values, fuel gas consumption grows:

– by 13.9% in the course of injection and by 32.4% in the course of withdrawal if a GCU with a single-unit CC with a unit capacity of 8.0 MW is used (see Fig. 4b and Table 5);

– by 10.7% in the course of injection and by 32.5% in the course of withdrawal if a GCU with a two-unit CC with a unit capacity of 8.0 MW is used (see Fig. 4d and Table 5);

[1] * The minus symbol indicates that the station will not be able to provide the specified gas flow rate with the designed number of operating GCUs and required redundancy. Implementing this mode will require a larger number of operating units.

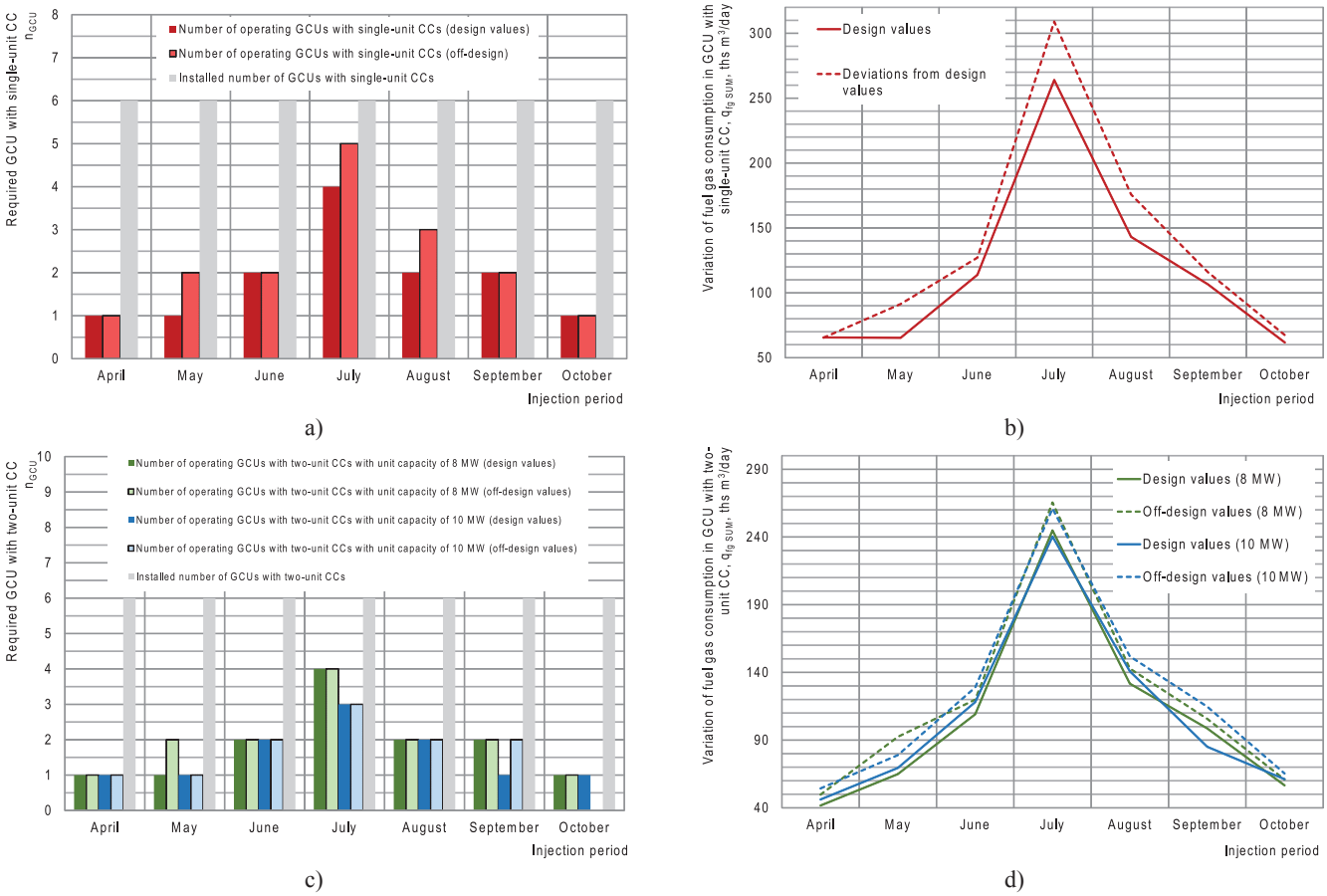



Fig. 4. Calculation data for the UGS operation in the course of gas injection in the reference conditions and when affected by FD factors (threats): a, required GCU with single-unit CC; b, variation of fuel gas consumption in GCUs with single-unit CCs; c, requirement for GCUs with two-unit CCs of different unit capacity; d, variation of fuel gas consumption in GCUs with two-unit CCs of various unit capacities

– by 10.9% in the course of injection and by 32.6% in the course of withdrawal if a GCU with a two-unit CC with a unit capacity of 10.0 MW is used (see Fig. 4d and Table 5).

The analysis of the FD assessment results showed that UGS CSs using GCUs with two-unit CCs have higher FD indicators as compared with GCUs with single-unit CCs as part of the UGS CS.

Thus, the analysis of the UGS CS FD allowed:

– identifying the optimal UGS CS options enabling MTF performance when the OI deviate from the design values, while ensuring the GCU redundancy standards with insignificantly decreased energy efficiency of the compression process;

– reducing the number of considered options for a detailed trade-off study.

## 7. Conclusion

A methodological approach to assessing the UGS CS FD was developed that consists in quantifying the negative consequences that may be caused by deviations of the actual operational parameters from the design values.

The methodological approach to FD assessment includes the following primary stages:

– quantification of FD factors (threats);

– computation of UGS CS operation modes for various versions operating per design and under FD factors (threats);

– comparative analysis of FD indicators for various UGS CS versions.

It was proposed quantifying FD with a number of indicators that affect the capital and operating costs:

– required backup GCUs;

– required installation and commissioning of additional GCUs;

– CS performance margins with a predefined number of GCUs and in compliance with the GCU redundancy requirements;

– changing requirements for fuel gas and installed GCUs.

The practical use of the methodological approach is shown by comparing the UGS CS GCU variants with single-unit and two-unit CCs. Based on the results of FD assessment, it was determined that UGS CSs equipped with a GCUs with two-unit CCs have higher functional dependability indicators as compared with single-unit CCs as part of the GCU.

The developed methodological approach to FD assessment has the potential for further improvement for the purpose of establishing a common approach for CS of various technical designation.

FD assessment allows comparing various CS versions in terms of the feasibility of ensuring the target design indica-

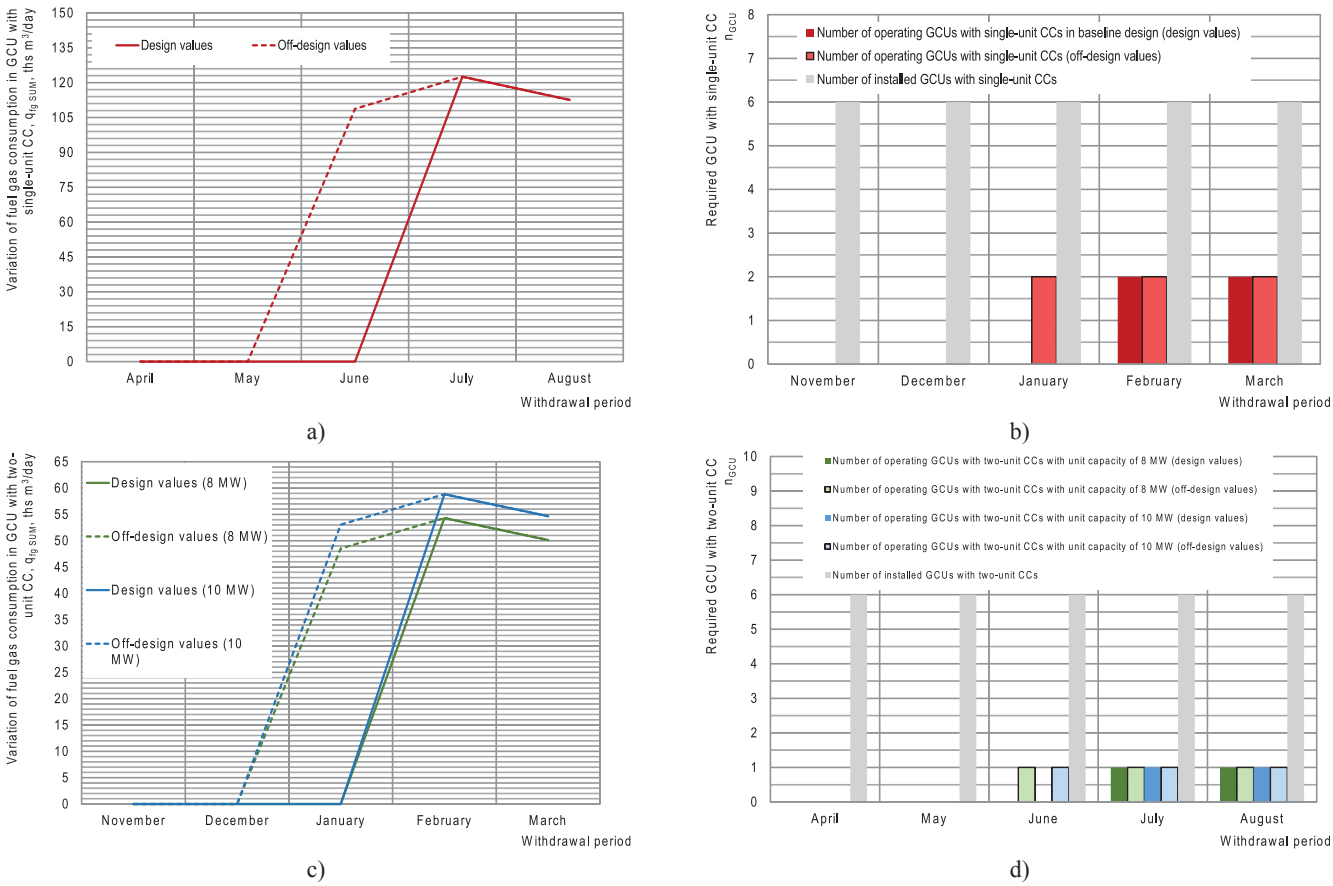



Fig. 5. Calculation data for the UGS operation in the course of gas withdrawal in the reference conditions and when affected by FD factors (threats): a, required GCU with single-unit CC; b, variation of fuel gas consumption in GCUs with single-unit CCs; c, requirement for GCUs with two-unit CCs of different unit capacity; d, variation of fuel gas consumption in GCUs with two-unit CCs of various unit capacities

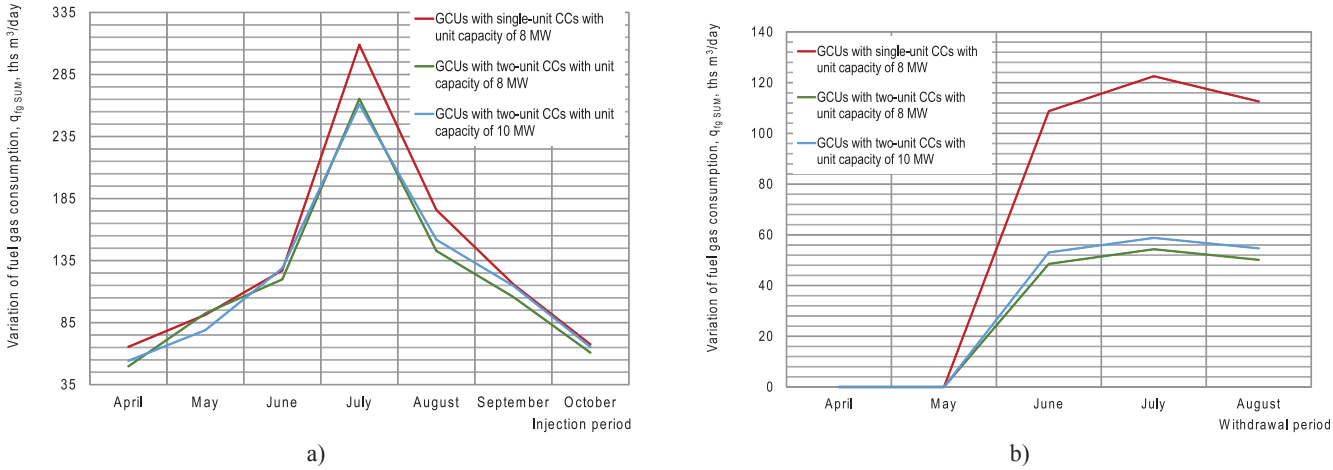



Fig. 6. Variation of fuel gas consumption for GCUs with single-unit and two-unit CCs (of various unit capacities) when affected by FD factors (threats): a, when injecting into UGS; b, when withdrawing from UGS

tors and CS performance variation when the actual operating conditions deviate from the design. It is recommended defining requirements for the UGS CS process system in terms of meeting the designed volume of gas extraction and injection, including in cases when the consumption, as well as the temperature and pressure performance, deviate from the design values, and assessing FD as early as at the CS design stage in order to substantiate the primary engineering solutions.

The findings can be used for creating risk-oriented approaches to the design of compression systems, i.e., those based on the risk and uncertainty assessment and management [7], as well as for assessing risks as part of investment decision support.

## References

1. Lyugay D.V, Dolgov S.I., Rakitina G.S. A role of underground gas storages in provision of stable functioning of the Unified System of Gas Supply of Russia. *Povyshenie nadezhnosti i bezopasnosti obektov gazovoy promyshlennosti* 2018;2(34):101-108. (in Russ.)

2. Shubinsky I.B., Schäbe H. On the definition of functional dependability. *Reliability: Theory & Applications* 2012:7;4(27):8-18.

3. Wakankar A., Kabra A., Bhattacharjee A.K., Karmakar G. Architectural model driven dependability analysis of computer based safety system in nuclear power plant. *Nuclear Engineering and Technology* 2019;51(2):463-478.

4. Kleiman L., Freyman V. Improving the functioning dependability of the information management system elements, using built-in diagnostic tools. *Radio Electronics, Computer Science, Control* 2021;1:158-171.

5. RD 25.03.001-2002. Guard and security systems of objects. Terms and definitions. Moscow: Izdatelstvo standartov; 2002. (in Russ.)

6. Vorontsov M.A., Grachiov A.S., Kirkin M.A., Bogatyriova E.V., Nuriev M.F., Drozdov A.V., Novikov A.I., Snezhko D.N. [Technological risks of the operation of underwater compressor units]. In: *[Proceedings of the 2018 International contest of research, engineering and innovative solutions aimed at the development of the Arctic and the continental shelf]* 2018. [accessed 11.02.2022]. Available at: https://in.minenergo.gov.ru/upload/tek/analitika/Arctica-2018-web.pdf . (in Russ.)

7. Vasiliev Yu.N., Gimadeyeva R.N., Ilnitskaya V.G. Uncertainties and risks of gas field development design and management. *Vesti gazovoy nauki: Problems of development of gas, gas condensate and oil/gas/condensate fields* 2014;4:16-22. (in Russ.)

8. GOST R 56205-2014 IEC/TS 62443-1-1:2009 Industrial communication networks. Network and system security. Part 1-1. Terminology, concepts and models. Moscow: Standartinform; 2014. (in Russ.)

9. GOST R 57329-2016/EN 13306:2010 Automation systems and integration. Maintenance and repair systems. Terms and definitions. Moscow: Standartinform; 2020. (in Russ.)

10. GOST R ISO/IEC 15026-1-2016 Systems and software engineering. Systems and software assurance. Part 1. Concepts and vocabulary. Moscow: Standartinform; 2016 (in Russ.)

11. Burgazzi L. Reliability Evaluation of Passive Systems Through Functional Reliability Assessment. *Nuclear Technology* 2003;144:145-151.

## About the authors

**Mikhail A. Vorontsov**, Candidate of Engineering, Senior Lecturer, Department of Thermodynamics and Thermal Engines, Gubkin University, Head of Laboratory for Oilfield Compressor and Turborefrigeration Systems, Gazprom VNIIGAZ, 15, 1 Proektiruemy proyezd no. 5537, Leninsky urban district, Razvilka, Moscow Region, Russian Federation, 142717, e-mail: m_vorontsov@list.ru.

**Anatoly S. Grachiov**, Researcher, Laboratory for Oilfield Compressor and Turborefrigeration Systems, Gazprom VNIIGAZ, 15, 1 Proektiruemy proyezd no. 5537, Leninsky urban district, Razvilka, Moscow, Region, Russian Federation, 142717, e-mail: grachev.anatoliy@yandex.ru.

**Alina O. Grachiova**, Second Year Postgraduate Student, Gubkin University, Engineer, Laboratory for Oilfield Compressor and Turborefrigeration Systems, Gazprom VNIIGAZ, 15, 1 Proektiruemy proyezd no. 5537, Leninsky urban district, Razvilka, Moscow Region, Russian Federation, 142717, e-mail: doki03@mail.ru.

**Maksim A. Kirkin**, Chief Engineer of Department (V.I. Dontsov), Gazprom, 156A Moskovsky pr-t, Saint Petersburg, Russian Federation, 196105, e-mail: M.Kirkin@adm.gazprom.ru.

**Anna V. Melnikova**, Candidate of Engineering, Chief Specialist, Laboratory for Predictive Simulation of Damage to Continuous and Area UGSS Facilities, Gazprom VNIIGAZ, 15, 1 Proektiruemy proyezd no. 5537, Leninsky urban district, Razvilka, Moscow, Region, Russian Federation, 142717, e-mail: A_Melnikova@vniigaz.gazprom.ru.

## The authors' contribution

**Vorontsov M.A.** Developed the method for assessing functional dependability, organized the research to test the method.

**Grachiov A.S.** Developed the method for assessing functional dependability, collected the input data for the calculation study.

**Grachiova A.O.** Conducted the calculation study, made the search and analysed sources, prepared an overview of research findings regarding the subject matter.

**Kirkin M.A.** Developed the method for assessing functional dependability, organized the research to test the method, prepared an overview of research findings regarding the subject matter.

**Melnikova A.V.** Participated in the discussion of the research methods and analysis of the findings.

## Conflict of interests

The authors declare the absence of a conflict of interests.

*Original article*
*https://doi.org/10.21683/1729-2646-2022-22-1-52-55*

# Synthesis of new, more powerful statistical tests through multiplicative clustering of classical Frozini and Murota-Takeuchi tests with the Hurst test for the purpose of testing small samples for normality

**Alexander I. Ivanov[1], Evgeny N. Kupriyanov[2]***
[1] *Penza Research and Design Electrical Engineering Institute, Penza, Russian Federation,*
[2] *Penza State University, Penza, Russian Federation*
**tsib@pnzgu.ru*


*Alexander I. Ivanov*


*Evgeny N. Kupriyanov*

**Summary. Aim.** *The paper examines the problem of small sample analysis by means of synthesizing new statistical tests generated by the clustering of the Hurst statistical test with the Frozini test, as well as with the Murota-Takeuchi test. The problem of normal distribution hypothesis testing on samples of 16 to 25 experiments is solved. Such significant limitations of the sample size arise in subject areas that include biometrics, biology, medical science and economics. In this case, the problem can be solved by applying not one, but a number of statistical tests to the analysis of the same small sample.* **Methods.** *It is suggested multiplying the Hurst test outputs by the Frozini test and/or the Murota-Takeuchi test outputs. A multiplicative clustering was performed for pairs of examined tests and their combination. It was shown that for each known statistical test, an equivalent artificial neuron can be constructed. A neural network integration of about 21 classical statistical tests constructed in the last century becomes possible. It is expected that the addition of new statistical tests in the form of artificial neurons will improve the quality of multi-criteria analysis solutions. Formally, the products of non-recurrent pairs of 21 original classical statistical tests should produce 210 new statistical tests. That is significantly more than the total number of statistical tests developed in the last century for the purpose of normality testing.* **Results.** *Pairwise product of the examined tests allows reducing the probability of errors of the first and second kind by more than 1.55 times as compared to the basic Hurst test. In case of triple product of the tests, the probabilities of error decrease relative to the basic Hurst test and to the associated second test. It is noted that there is no steady improvement in the quality of the decisions made by multiplicative mathematical constructions. The probabilities of error of the new test obtained by multiplying three of the examined tests are approximately 1.5% worse as compared to those of the tests obtained by multiplying pairs of the original tests.* **Conclusions.** *By analogy with the examined tests, the proposed data processing methods can also be applied to other known statistical tests. In theory, it becomes possible to significantly increase the number of new statistical tests by multiplying their final values. Unfortunately, as the number of clustered statistical tests grows, mutual correlations between the newly synthesized tests grow as well. The latter fact limits the capabilities of the method proposed in the paper. Further research is required in order to identify the most efficient combinations of pairs, triples or large groups for known statistical tests.*

**Keywords:** *statistical analysis of small samples, normality testing, Hurst test, Frozini test, Murota-Takeuchi test.*

**For citation:** *Ivanov A.I., Kupriyanov E.N. Synthesis of new, more powerful statistical tests through multiplicative clustering of classical Frozini and Murota-Takeuchi tests with the Hurst test for the purpose of testing small samples for normality. Dependability 2022;1: 52-55. https://doi.org/10.21683/1729-2646-2022-22-1-52-55*

**Received on:** *08.10.2021* / **Revised on:** *28.01.2022* / **For printing:** *18.03.2022*

## Introduction

Neural networks are trained to convert biometrics into an authentication code per GOST R 52633.5 [1] using 16 examples of a "Friend" image. "Good" biometrics data have a normal distribution, while "bad" data with gross errors have a near-uniform distribution. Eventually, when evaluating the quality of small learning samples, the hypothesis of normal distribution of a small sample of 16 examples needs to be tested.

One of the methods of testing the normality hypothesis involves using the Hurst test (the ratio between the scope of data and the standard deviation of the sample that is commonly used in economics [2]). Unfortunately, this statistical test does not perform well with small samples. The distribution of data is shown in Fig. 1.

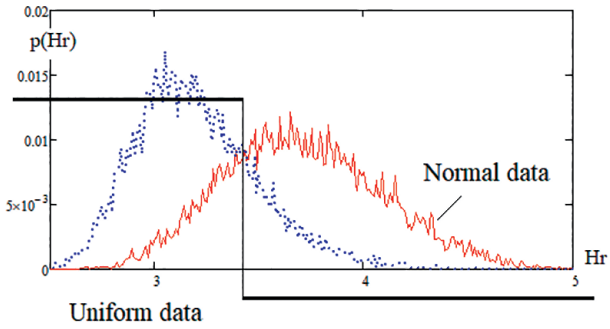



Fig. 1. Example of poor linear separability by an artificial neuron of the output states of the classical Hurst test for small samples of 16 experiments

It is obvious that, for small samples, the probabilities of errors of the first and second kind are high: $P_1 = P_2 = P_{EE} \approx 0.228$. In this context, according to the standard recommended [3, 4] acceptable values of the confidence probability, the classical tests are to be applied to samples of 200 or more experiments. This condition cannot be fulfilled for neural network biometrics.

A similar situation occurs when we try using another statistical test. Fig. 2 shows the density functions of the output states of the Frozini test.

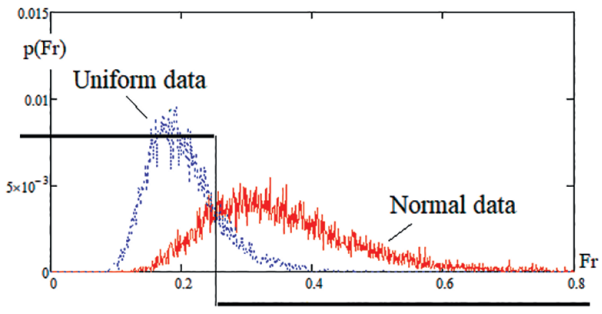



Fig. 2. Output states of the classical Frozini test for small samples of 16 experiments

Comparing Fig. 1 and Fig. 2 clearly shows that the linear separability of normal and uniform small sample data of the Frozini test is significantly better $P_1 = P_2 = P_{EE} \approx 0.172$ as compared to the Hurst test. We observe a 1.33-fold decrease in the probability of errors of the first and second kind. The effect of linear separability of data is even higher for the Murota-Takeuchi test, Fig. 3.

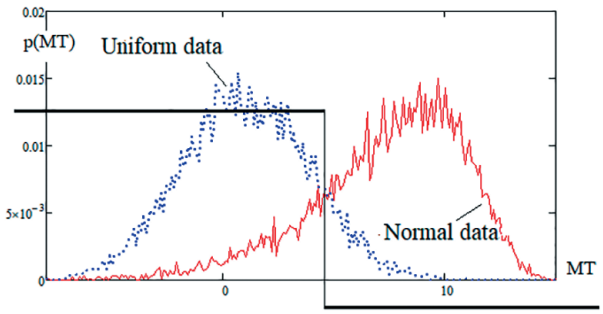



Fig. 3. Distribution of the output data of the Murota-Takeuchi test

For the Murota-Takeuchi test, the probability of errors of the first and second kind can be reduced to $P_1 = P_2 = P_{EE} \approx 0.152$. That test is the most powerful out of the three examined.

It is also obvious that for each statistical test [5, 6], an equivalent artificial neuron can be constructed that quantizes data in the point of equally probable errors of the first and second kind $P_1 = P_2 = P_{EE}$, if the quantizer outputs "0" for normal data. In this case, the three artificial neurons under consideration will likely output code "000" with triple redundancy if the input is data with a normal-like distribution.

Moreover, the reference book [7] describes 21 statistical tests for normality testing. In other words, we can obtain 21 artificial neurons that solve the same problem simultaneously. At the same time, formally, we will obtain output codes with a 21-fold redundancy. This redundancy can be contracted using codes that enable detection and correction of errors [8].

Unfortunately, most of the statistical tests created in the last century involve a strong correlation between the output states. Taking into account the effect of the correlations [9] causes the situation whereas the statistical tests created in the last century are not sufficient for a confidence probability of 0.99. About 40 new statistical tests need to be synthesised in the near future.

## Method of increasing the number of statistical tests through pairwise multiplication of their final results

It should be noted that multiplying the outputs of the formula of a certain statistical test should cause increased linear separability of small samples with a normal and uniform distribution. This fact is easily verified through a numerical experiment. The Hurst, Frozini and Murota-Takeuchi tests, as well as their multiplicative clusterings are calculated using software written in MathCAD and shown Fig. 4.

$$sx := \left| \begin{array}{l} x \leftarrow sort(morm(16,0,1)) \\ m \leftarrow mean(x) \\ \sigma \leftarrow stdev(x) \\ Hr \leftarrow \dfrac{x_{15} - x_0}{\sigma} \\ Fr \leftarrow \sum_{i=0}^{15} \left[ \left| \left( pnorm(x_i, m, \sigma) - \dfrac{i - 0.5}{16} \right) \right| \cdot dnorm(x_i, m, \sigma) \right] \\ MT \leftarrow \sum_{i=0}^{15} \left( \cos\left( \dfrac{x_i - x_{15}}{1.8} \right) \right) \\ (Hr \;\; Fr \;\; MT \;\; Hr{\cdot}Fr \;\; Hr{\cdot}MT \;\; Hr{\cdot}Fr{\cdot}MT)^T \end{array} \right.$$

$$sxr := \left| \begin{array}{l} x \leftarrow sort(runif(16, -3, 3)) \\ m \leftarrow mean(x) \\ \sigma \leftarrow stdev(x) \\ Hr \leftarrow \dfrac{x_{15} - x_0}{\sigma} \\ Fr \leftarrow \sum_{i=0}^{15} \left[ \left| \left( pnorm(x_i, m, \sigma) - \dfrac{i - 0.5}{16} \right) \right| \cdot dnorm(x_i, m, \sigma) \right] \\ MT \leftarrow \sum_{i=0}^{15} \left( \cos\left( \dfrac{x_i - x_{15}}{1.8} \right) \right) \\ (Hr \;\; Fr \;\; MT \;\; Hr{\cdot}Fr \;\; Hr{\cdot}MT \;\; Hr{\cdot}Fr{\cdot}MT)^T \end{array} \right.$$

Fig. 4. Software for numerical simulation of three examined statistical tests and their multiplicative combinations

Among other things, the above software allows calculating the Hurst-Frozini multiplicative test. Data on the density function of the test's output states are shown in Fig. 5.

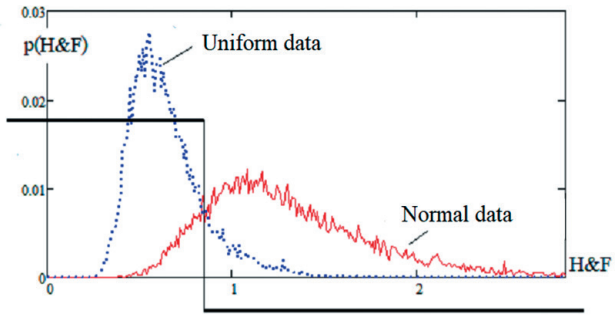



Fig. 5. Density functions of the output states of the Hurst and Frozini multiplicative test

The multiplicative Hurst-Frozini hybrid has a 28% lower probability of errors of the first and second kind $P_1 = P_2 = P_{EE} \approx 0.134$ as compared with the Frozini test, the most powerful one out of them. Another version of the new Hurst (and Murota-Takeuchi) multiplicative statistical test also has a significant, 14% reduction in the probability of errors of the first and second kind as compared with the Murota-Takeuchi test, the most powerful one out of them. The distribution of the response probability functions of the synthesized multiplicative test are shown in Fig. 6.

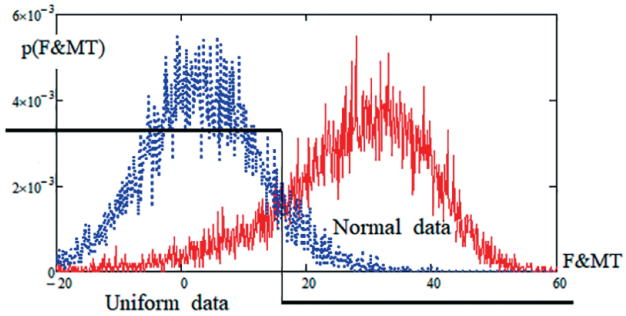



Fig. 6. Density functions of the output states of the Hurst (and Murota-Takeuchi) multiplicative test with probability of error $P_1 = P_2 = P_{EE} \approx 0.133$

## Synthesising another test by multiplying the output states of all three examined statistical tests

As the above transformations show, multiplying the responses of two statistical tests significantly reduces the probability of errors of the first and second kind. In theory, the effect should grow if more than two statistical test outputs are multiplied. Fig. 7 shows data on the probability of errors of the first and second kind obtained by multiplying all three examined statistical tests.

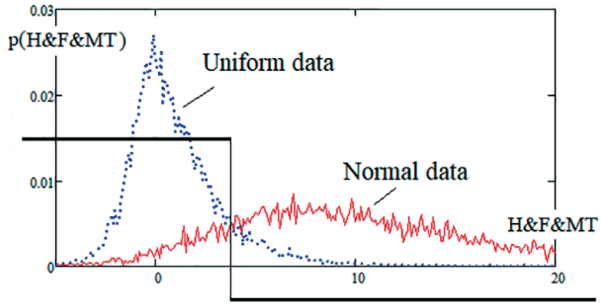



Fig. 7. Density functions of the output states of the Hurst-Frozini-Murota-Takeuchi multiplicative test obtained by multiplying the three examined tests

Multiplying a group of three statistical tests produces the probability of errors of the first and second kind $P_1 = P_2 = P_{EE} \approx 0.136$, which is 1.5% worse as compared with the paired multiplicative tests. It appears that the probability of errors hardly always occurs when the number of multiplied partial parameters grows. Unfortunately, as the number of multiplied tests (the multiplicativity) grows, the correlation of their responses increases as well. That appears to be the exact factor that limits the decrease of the probabilities when it is attempted to increase the number of multiplicatively clustered initial tests.

## Conclusion

All known statistical tests can be divided into two classes. In the case under consideration, all three tests belong to the same class. They are similar with respect to the point of equal

probability of error in the shared data $P_1 = P_2 = P_{EE}$. The distribution of normal small sample data for the Hurst, Frazini and Murota-Takeuchi tests is always to the left of the point $P_1 = P_2 = P_{EE}$ (see continuous graphs in Fig. 1–3, 5, 6). The distribution of uniform data for such tests is always to the right of the point $P_1 = P_2 = P_{EE}$ (see the dotted graphs in Fig. 1–3, 5, 6). That allows clustering the tests multiplicatively.

On the one hand, multiplicative clustering of known statistical tests allows synthesizing quite a number of new tests. However, such attempts cause a growing rate of correlation of new data, which is a negative phenomenon. In general, synthesizing new statistical tests through multiplicative clustering of existing tests is impossible without taking into account the growing correlations between the new tests.

## References

1. GOST R 52633.5-2011. Information protection. Information protection technology. The neural net biometry-code converter automatic training. Moscow: Standartinform; 2018. (in Russ.)

2. Mandelbrot B., Hudson R.L. The (Mis)Behaviour of Markets: A Fractal View of Risk, Ruin and Reward. Moscow, Saint Petersburg, Kiev: Izdatelstvo Viliams; 2006.

3. R 50.1.037-2002. [Applied statistics. Rules for compliance verification of experimental distribution with the theoretical one. Part I. Chi-square-type criteria]. Moscow: Gosstandart of Russia; 2001. (in Russ.)

4. R 50.1.037-2002. [Applied statistics. Rules for compliance verification of experimental distribution with the theoretical one. Part II. Non-parametric tests]. Moscow: Gosstandart of Russia; 2002. (in Russ.)

5. Ivanov A.I., Kupriyanov E.N., Tureev S.V. Neural network integration of classical statistical tests for processing small samples of biometrics data. *Dependability* 2019;2:22-27. DOI: 10.21683/1729-2646-2019-19-2-22-27.

6. Ivanov A.I., Bannykh A.G., Bezyaev A.V. Artificial molecules assembled from artificial neurons that reproduce the work of classical statistical criteria. *Vestnik permskogo unoversiteta. Seria: Matematika. Mekhanika. Informatika* 2020;1(48):26-32. (in Russ.)

7. Kobzar A.I. [Applied mathematical statistics. For engineers and researchers]. Moscow: FIZMATLIT; 2006. (in Russ.)

8. Morelos-Zaragoza R. The R, Art of Error Correcting Coding. Moscow: Tekhnosfera; 2007.

9. Ivanov A.I., Bannykh A.G., Serikova Yu.I. Accounting for the effect of correlations by modulo averaging as part of neural network integration of statistical tests for small samples. *Dependability* 2020; 20(2):28-34. DOI: 10.21683/1729-2646-2020-20-2-28-34.

## About the authors

**Alexander I. Ivanov**, Doctor of Engineering, Associate Professor, Lead Researcher, Laboratory of Biometric and Neural Network Technologies, Penza Research and Design Electrical Engineering Institute, Address: 9 Sovetskaya St., Penza, Russian Federation, phone: (841-2) 59 33 10, e-mail: ivan@pniei.penza.ru.

**Evgeny N. Kupriyanov,** post-graduate student, Department of Information Security Technology, Penza State University, Address: 40 Krasnaya St., Penza, Russian Federation, e-mail: tsib@pnzgu.ru.

## The authors' contribution

**Ivanov A.I., Kupriyanov E.N.** jointly suggested multiplying outputs of computations based on known statistical tests assuming a steadily increasing quality of the solutions made as the number of multiplied components (tests) grows.

**Kupriyanov E.N.** conducted a numerical experiment that identified the absence of a steady decrease of the probability of errors of the first and second kind as the number components multiplied as part of synthesis grows.

## Conflict of interests

The authors declare the absence of a conflict of interests.

*Original article*
*https://doi.org/10.21683/1729-2646-2022-22-1-56-62*

# Developing a method for recovering data on storage media

**Vasily S. Zamolotskikh[1]*, Valentina G. Sidorenko[1]**
*[1]Russian University of Transport, Moscow, Russian Federation*
**netaonh11@gmail.com*


*Vasily S. Zamolotskikh*


*Valentina G. Sidorenko*

**Abstract. Aim.** *The development of digital technology brings about the need to digitize data with their subsequent storage on digital media. Regardless of how information is stored, it is of value and its loss may cause harm. There are a number of preventive measures (hardware and logical redundancy of various types) to prevent such loss. Should the preventive measures – due to certain reasons and circumstances – fail to protect the data and access to the latter was lost, it must be recovered in a complete and timely manner. In this context, a need arises for a data recovery algorithm that would take into account the hardware features of today's storage media, their logical structure, as well as the specificity of the stored data.* **Methods.** *There are two approaches to information recovery, i.e., all-purpose and personal. The all-purpose approach involves using a minimal number of programs and tools that work with all items. The personal approach implies a large number of programs and tools that address specific issues associated with the loss of access to information. That enables a faster, higher-quality recovery as compared to the all-purpose approach. Additionally, personal programs are normally cheaper than all-purpose software. All-purpose information recovery tools do not provide quality results when applied to large numbers of failure scenarios. A single utility may not be enough for resolving all issues caused by an incident. A readily available template for obtaining an acceptable result does not exist either. Aside from personal software, there are other alternatives to the all-purpose approach, i.e., the manual data recovery programs and hardware and software systems. In cases of minor logical faults (master boot record corruption) manual data recovery software is used. If a drive is affected by critical hardware issues, hardware and software systems are used.* **Results.** *A method of recovering data on storage media of various types was created. It includes an all-purpose and a personal approaches to information recovery, use of software for manual data recovery, as well as hardware and software systems. The method allows recovering data of popular extensions from common file systems and storage media. Compatibility with RAID arrays of all levels is provided. Programs were selected out of eight sets using the analytic hierarchy process with the priority given to the performance criterion. The method was submitted to a number of tests. Testing involved emulation of incidents associated with the loss of access to data. The cost of eliminating various incidents using the developed methodology is estimated.* **Conclusions.** *Based on the obtained test results, conclusions are set forth regarding the efficiency of the personal approach to information recovery.*

**Keywords:** *storage medium, data organization, analytic hierarchy process.*

**For citation:** *Zamolotskikh V.S., Sidorenko V.S. Developing a method for recovering data on storage media. Dependability 2022;1: 56-62. https://doi.org/10.21683/1729-2646-2022-22-1-56-62*

**Received on:** *21.01.2022 /* **Upon revision:** *03.02.2022 /* **For printing:** *18.03.2022.*

## Introduction

Data is of paramount importance, and in some cases, it is worth more than the media it is stored on. The cost of recovery may also exceed the cost of the storage media. It is therefore very important to properly evaluate the information and determine if it needs to be recovered and whether it is cost-effective to do so. This evaluation, as well as the cause of the loss of access to information, will help choose the method for recovering information on storage media that is described in this paper.

There are several definitions of data recovery. Data recovery is a sequence of actions aimed at retrieving information from a storage medium when such information cannot be read in the conventional way [1].

Data recovery is a sequence of actions, in which damaged or unreadable information is accessed and transferred to another device [2]. The key difference between the two is that the second definition directly refers to the transferring of the recovered data to another (i.e., operable) device.

## 1. Data organisation on storage media

Developing a data recovery method involves analysing the causes of data loss. Classifying the latter requires taking into consideration the way the data is organised on a medium. Data organisation is examined from two aspects, i.e., hardware-specific and logical.

The hardware-specific data organisation depends on the type of storage medium, its physical component. In hard disk drives (HDDs), it is stored on magnetic plates within the hermetic block [3]. In solid state drives, the data is stored on floating gates that are used in transistors of the same name. Solid state drives include an external interface (in case of peripheral connection to a computer), a controller and memory cells. [1]. In optical storage media, the reading beam passes through the bottom polycarbonate layer, hits the information layer and is reflected. [4, 5].

The hardware-specific data organisation has an effect on the causes of data loss.

### 1.2. Logical organisation

The logical organisation is common and does not depend on the type of storage medium. It is a hierarchy: disk, partition, directory, file [4]. The first sector of a disk contains the master boot record (MBR) that is required for operating system booting. The partition layout depends on the size of the storage medium. Media larger than 2 TB use global unique identifiers (GUID). The file structure has a main header that stores the file signature. File recovery software uses this record to detect files on a drive.

## 2. Causes of information loss

The causes of information loss are shown in Fig. 1 [6]. Based on data organisation information, three groups of causes of information loss can be distinguished, i.e., hardware-specific, logical and the human factor (Fig. 2).

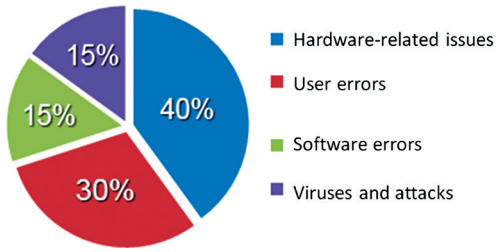



Fig. 1. Distribution of the causes of information loss

## 3. Algorithm for recovering information on storage media

Based on an analysis of the physical and logical organisation of data on storage media, as well as the causes of data loss, the authors have created an algorithm for data recovery on storage media (Fig. 3). It is based on the use of hardware and logical redundancy in the data organisation, as well as equivalent storage media. Hardware redundancy is implemented by copying service data, backing up and virtualisation, while logical redundancy is implemented by adding a signature to the file header. It is convenient to use operable equivalent storage media for manual data recovery.

The choice of recovery direction depends on the degree of damage to the service information that includes data of the file tables that contain file addresses. If they are severely damaged, the emphasis is placed on the bit sequence analysis, i.e., on signature search. Note that this method is the most popular and is used in most software applications. They, however, also use service information for data recovery. Manual data recovery is done with hex and disk editors. It is convenient to use the disk editor for service data recovery,

**Table 1. Baseline data and AHP outputs**

| Item | SW type | Number of examined SW applications | Winner |
|---|---|---|---|
| 1 | Storage media testing | 5 | R.tester |
| 2 | Copying media byte-to-byte | 5 | Active Disk Image |
| 3 | Manual data recovery | 7 | Acronis Disk Director |
| 4 | All-purpose | 13 | Hetman Partition Recovery |
| 5 | RAID array recovery | 3 | RS RAID Retrieve |
| 6 | Removed file recovery | 4 | Active Undelete |
| 7 | Corrupted file recovery | 12 | RS File Repair, Recovery Toolbox |
| 8 | Personal | 25 | Recuva, EasyRecovery |

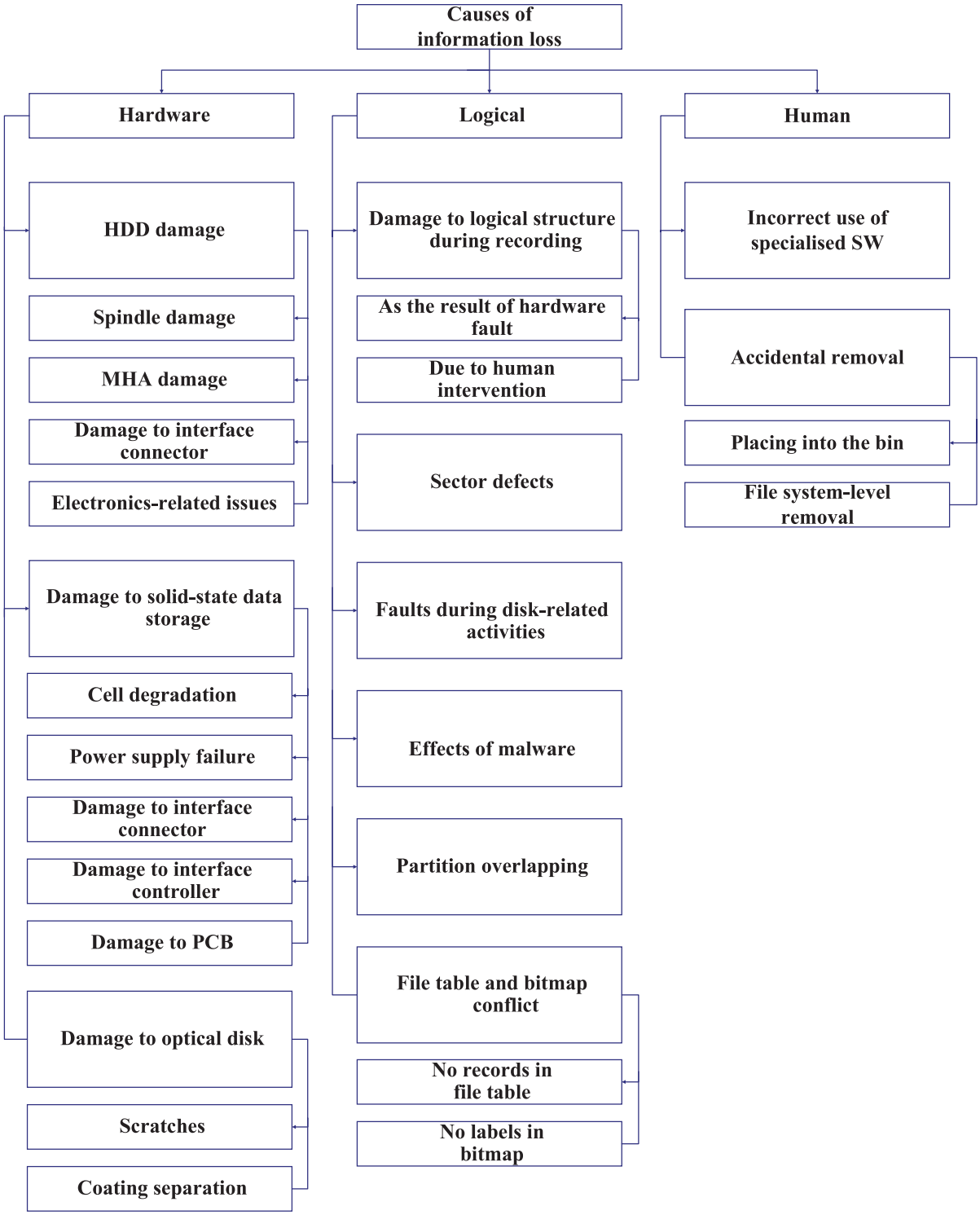



Fig. 2. Causes of information loss

as it represents the storage medium as a partitioned image. Signature search can be done by entering file signatures into the search window [4].

## 4. Selecting software by means of analytic hierarchy process

For the algorithm to be usable, it must be equipped with software (SW). The SW that is available on the market and can be used for implementing the stages of the developed algorithm was divided into 8 groups. In each group, SW was selected through the analytic hierarchy process (AHP) [7] with priority given to the performance criterion (Table 1).

AHP enables pairwise comparison of items, as well as decision-making based on the item-specific criteria. The choice of this method is due to the following advantages [8]: the possibility of pairwise comparison of criteria, which

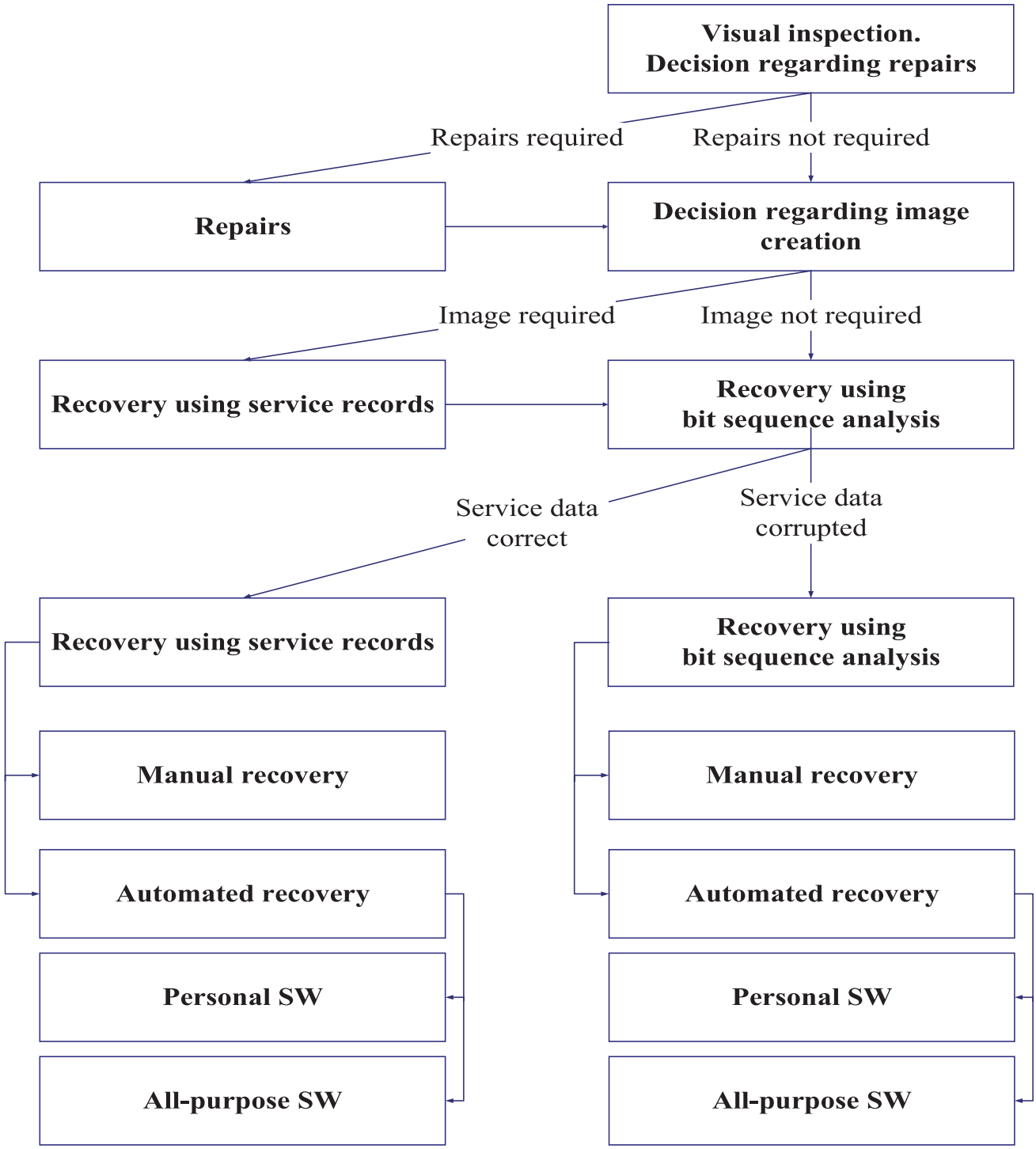



Fig. 3. Information recovery algorithm

facilitates calculations, as there is no need to "keep in sight" all the criteria that can be numerous; the existence of a verbal-numerical scale, which simplifies evaluation, as it compares numbers with easy-to-understand logical comparative constructions; the expert quality assessment (consistency ratio) built into the method.

Formulas (1) to (3) allow estimating items in terms of characteristic criteria.

$$\text{MNGEOM}_x = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n}, \quad (1)$$

where $\text{MNGEOM}_x$ is the geometric mean of judgements $x$ (rowwise);

$x$ is a positive number;

$n$ is the number of numbers $x$.

$$\text{SUM}_a = \sum_{i=1}^{n} a_i, \quad (2)$$

where $\text{SUM}_a$ is the result of the sum of the numbers $a$;

$a_i$ is a number with an ordinal number $i$;

$n$ is the number of numbers $a$.

$$\mathrm{VECTOR}_x = \frac{\mathrm{MNGEOM}_x}{\mathrm{SUM}_{\mathrm{MNGEOM}_x}}, \qquad (3)$$

where $\mathrm{VECTOR}_x$ is the value of the priority vector of criterion/SW $x$;

$\mathrm{MNGEOM}_x$ is the geometric mean of the numbers $x$;

$\mathrm{SUM}_{\mathrm{MNGEOM}_x}$ is the sum of the geometric mean values of $x$.

The final estimate will be obtained by using formula 4:

$$\mathrm{GLOBVECTROR}_x = \sum_{i=1}^{n} \mathrm{VECTOR}_{\mathrm{criterion}_i} \times \mathrm{VECTOR}_{\mathrm{SW}_i}, \qquad (4)$$

where $\mathrm{GLOBVECTROR}_x$ is the value of the global priority vector of SW $x$;

$n$ is the number of criteria.

$\mathrm{VECTOR}_{\mathrm{criterion}_i}$ is the value of the priority vector of the criterion with serial number $i$;

$\mathrm{VECTOR}_{\mathrm{SW}_i}$ is the value of the priority vector of SW with serial number $i$.

The group of formulas from (5) to (7) is required for calculating the conformity relation (CR) that determines the correctness of the judgements. According to the recommendations, it is not supposed to exceed 10%. The recommendation has been fulfilled.

$$\mathrm{NORM}_x = \mathrm{SUM}_x \times \mathrm{VECTOR}_x, \qquad (5)$$

where $\mathrm{NORM}_x$ is the normalised judgement vector $x$;

$\mathrm{SUM}_x$ is the sum of the judgements values $x$ (columnwise);

where $\mathrm{VECTOR}_x$ is the value of the priority vector of criterion/SW $x$;

$$\mathrm{CI} = \frac{\mathrm{SUM}_{\mathrm{NORM}} - n}{n-1}, \qquad (6)$$

where CI is the conformity index of the criterion/SW estimates;

$\mathrm{SUM}_{\mathrm{NORM}}$ is the sum of the normalised judgement vectors;

$n$ is the number of criteria or SW.

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RNCOH}} \times 100\%, \qquad (7)$$

where CR is the conformity relation;

CI is the conformity index;

RNCOH is the measure of random coherence.

**Table 2. Assessment of AHP criteria**

| Criterion | Performance | Flexibility in terms of media types | Price | Additional capabilities | Priority vector | Geometric mean | CI | AHP CR, % |
|---|---|---|---|---|---|---|---|---|
| Performance | 1.0 | 3.0 | 4.0 | 4.0 | 0.5285 | 2.6 | 0.03 | 3.52 |
| Flexibility in terms of media types | 0.3 | 1.0 | 2.0 | 3.0 | 0.2388 | 1.2 | | |
| Price | 0.3 | 0.5 | 1.0 | 2.0 | 0.1420 | 0.7 | | |
| Additional capabilities | 0.3 | 0.3 | 0.5 | 1.0 | 0.0907 | 0.5 | | |
| Sum | 1.8 | 4.8 | 7.5 | 10.0 | 1 | 5.0 | | |

**Table 3. Evaluation of SW in terms of performance**

| | Everest Ultimate Edition/AIDA 64 | Victoria | MHDD | R.tester | HDDScan | Priority vector | Geometric mean | CI | AHP CR, % |
|---|---|---|---|---|---|---|---|---|---|
| Everest Ultimate Edition/AIDA 64 | 1.0 | 2.0 | 2.0 | 1.0 | 2.0 | 0.2857 | 1.5 | 0.00 | 0.00 |
| Victoria | 0.5 | 1.0 | 1.0 | 0.5 | 1.0 | 0.1429 | 0.8 | | |
| MHDD | 0.5 | 1.0 | 1.0 | 0.5 | 1.0 | 0.1429 | 0.8 | | |
| R.tester | 1.0 | 2.0 | 2.0 | 1.0 | 2.0 | 0.2857 | 1.5 | | |
| HDDScan | 0.5 | 1.0 | 1.0 | 0.5 | 1.0 | 0.1429 | 0.8 | | |
| Sum | 3.5 | 7.0 | 7.0 | 3.5 | 7.0 | 1.0000 | 5.3 | | |

**Table 4. Incident models**

| No. of incident | Storage medium | File system | Cause of incident | Incident manifestation |
|---|---|---|---|---|
| 1 | HDD | *NTFS* | HDD impact | MHA damage |
| 2 | *SSD* | *NTFS* | Electrical breakdown | Controller damage |
| 3 | Flash disc | *NTFS* | *MBR* removal (values replaced with 0) | Flash disc not detected |
| 4 | Flash disc | *NTFS* | File deletion at the FS level | Files unavailable |
| 5 | Flash disc | *NTFS* | Values of main file headers replaced with 0 | Files unreadable |
| 6 | HDD | *NTFS* | *MFT* not matching the bitmap (values corresponding to file addresses in *MFT* replaced with 0) | Files unavailable |
| 7 | *SSD* | *APFS* | Values replaced with 0: container boot block, disc headers, as well as file removal at the FS level | Drive recognized incorrectly |

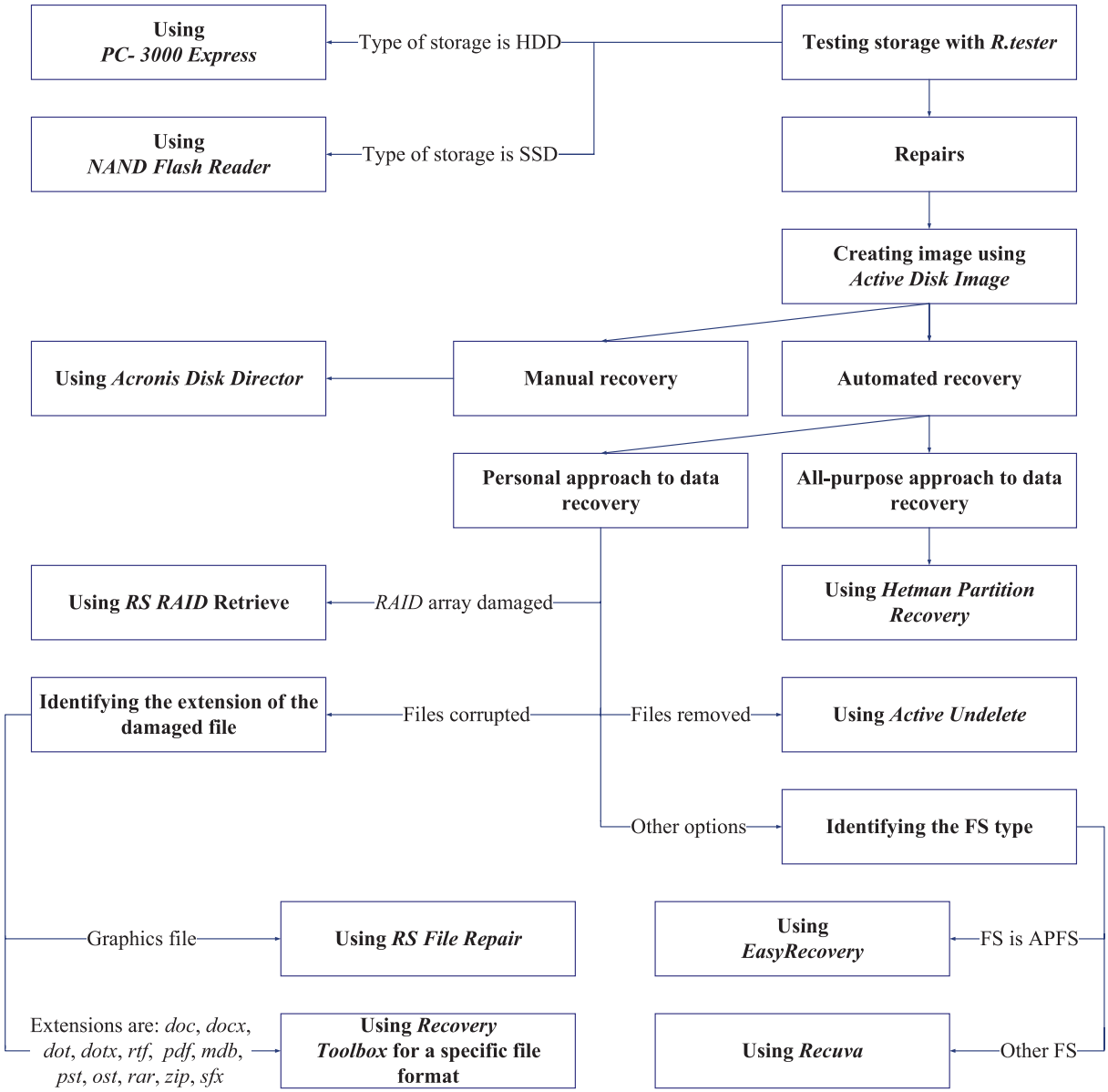



Fig. 4. Method for information recovery

Let us examine the essence of AHP using the case of SW selection for media testing. First, the criteria specific to the test software are evaluated (Table 2). Then, the programs are evaluated in terms of each of the criteria, e.g., the performance (Table 3).

## 5. Data recovery method

The result is the following method for data recovery on storage media (Fig. 4).

A separate group includes hardware and software systems that should be used if the media are affected by major hardware issues. Manual data recovery software is to be used in case of minor logical faults. An all-purpose approach to data recovery is implemented using *Hetman Partition Recovery*, while a personal approach is implemented using a combination of software at the bottom of Fig. 4.

## 6. Testing the method

The developed method was tested by solving data access loss incidents (Table 4). The first and second incidents involve the use of hardware and software systems, the third incident involves the use of software for manual information recovery. In the fourth to the sixth incidents, personal information recovery software was used, while the seventh incident involved the use of general-purpose software.

All incidents were successfully resolved. Personal software provided the quickest results, which gives reasons to recommend it for data recovery in most cases.

Table 5 shows the economic models that reflect the options for applying the data recovery method to the incidents, on which the developed method was tested. The cost of technical equipment includes the cost of equipment and software. The cost of operation depends on the consumed time and electrical power. The rapid operation of personal

**Table 5. Economic models of the applications of the developed data recovery method**

| No. of economic model | No. of incident | Used tools | Cost of equipment, RUB | Cost of operation, RUB |
|---|---|---|---|---|
| 1 | 1 | R.tester, PC-3000 Express | 191 900 | 4841.58 |
| 2 | 2 | R.tester, NAND Flash Reader | 238 243.18 | 4024.94 |
| 3 | 3 | R.tester, Active Disc Image, Acronis Disk Director | 54 709.82 | 8494.19 |
| 4 | - | R.tester, Active Disc Image, RS RAID Retrieve | 59 318.13 | 3298.1 |
| 5 | 4 | R.tester, Active Disc Image, Active Undelete | 53 264.11 | 4068.11 |
| 6 | 5 | R.tester, Active Disc Image, RS File Repair | 52 938.13 | 2869.55 |
| 7 | 5 | R.tester, Active Disc Image, Recovery Toolbox | 53 770.84 | 2869.55 |
| 8 | - | R.tester, Active Disc Image, EasyRecovery | 57 115.37 | 2896.34 |
| 9 | 6 | R.tester, Active Disc Image, Recuva | 52 619.13 | 3298.1 |
| 10 | 7 | R.tester, Active Disc Image, Hetman Partition Recovery | 54 818.13 | 4101.62 |

software (economic models 4 to 9) ensures monetary savings in operation.

## Conclusion

Based on the conducted activities and research, a method for recovering information on storage media was developed (see Fig. 4).

According to test results and economic model calculations, the authors recommend using personal information recovery software.

## References

1. [Dictionaries and encyclopaedia on academic.ru. Data recovery]. [accessed: 02.09.2021]. Available at: https://dic.academic.ru/dic.nsf/ruwiki/1369930. (in Russ.)

2. [Storelab. What is data recovery?]. [accessed: 02.09.2021]. Available at: https://storelab-rc.ru/vosstanovlenie-dannih.html. (in Russ.)

3. [How a hard drive works. What is inside an HDD and SSHD]. [accessed: 13.09.2021]. Available at: ttps://www.youtube.com/watch?v=77Ugghti1b0. (in Russ.).

4. Senkevich G.E. [The art of data recovery]. Saint Petersburg: BHV-Petersburg; 2011. (in Russ.)

5. [L-pro. DVD]. [accessed: 07.09.2021]. Available at: http://l-pro.com/uslugi/tirazhirovanie-cd-dvd-diskov/disk-formata-dvd. (in Russ.).

6. [Loss of information leads to loss of business]. [accessed: 26.11.2021]. Available at: https://www.cnews.ru/reviews/itinfrastruktura_predpriyatiya_2013/interviews/aleksandr_frolikov. (in Russ.).

7. Saaty T. Decision making – the Analytic Hierarchy and Network Processes. Moscow: Radio i sviaz; 1989.

8. Korobov V.B., Tutygin A.G. [Advantages and disadvantages of the analytic hierarchy process]. *Izvestia: Herzen University Journal of Humanities & Sciences* 2010;122:108-115. [accessed: 20.01.2022]. Available at: https://cyberleninka.ru/article/n/preimuschestva-i-nedostatki-metoda-analiza-ierarhiy. (in Russ.).

## About the authors

**Vasily S. Zamolotskikh**, Student, Russian University of Transport, Moscow, Russian Federation, e-mail: netaonh11@gmail.com.

**Valentina G. Sidorenko**, Doctor of Engineering, Professor, Chair Professor, Department of Management and Protection of Information, Russian University of Transport, Moscow, Russian Federation, e-mail: valenfalk@mail.ru.

## The authors' contribution

**Zamolotskikh V.S.** Development of the method for recovering information from storage media, preparation of test models of incidents, testing and analysis of findings, conclusions.

**Sidorenko V.G.** Analysis of methods and principles of information recovery, overview of the methods for recovering information on storage media.

## Conflict of interests

The authors declare the absence of a conflict of interests.

# GNEDENKO FORUM

INTERNATIONAL GROUP ON RELIABILITY


The Gnedenko Forum was founded in 2004 by an unofficial international group of experts in the dependability theory for the purpose of professional support of researches from all over the world who are interested in studying and developing the scientific, technical and other aspects of the dependability theory, risk analysis and safety in the theoretical and practical domains.

The Forum exists on the Internet as a non-for-profit organization. It aims to involve into joint discussion and communication technical experts interested in developing the dependability theory, safety and risk analysis regardless of their home country and membership in whichever organization.

The Forum acts as an impartial and neutral entity that delivers scientific information to the press and public as regards the matters of safety, risk analysis and dependability of complex technical systems. It publishes reviews, technical documents, technical reports and research essays for the purpose of dissemination of knowledge and information.

The Forum is named after Boris V. Gnedenko, an outstanding Soviet mathematician, expert in the probability theory and its applications, member of the Ukrainian Academy of Sciences. The Forum is the platform for distribution of information on educational grants, academic and professional positions related to dependability, safety and risk analysis all over the world.

Currently, the Forum has 500 members from 47 countries.

Since January 2006, the Forum has been publishing its quarterly journal, Reliability: Theory & Applications (www.gnedenko.net/RTA). The Journal is registered in the Library of Congress (ISSN 1932-2321) and publishes articles, reviews, memories, information and literature references regarding the theory and application of dependability, survivability, maintenance, risk analysis and management methods.

Since 2017, the Journal is indexed in Scopus.

Membership in the Gnedenko Forum does not imply any obligations. It is only required to send your photograph and a brief professional biography (resume) to a.bochkov@gmail.com.
Templates can be found at http://www.gnedenko.net/personalities.htm.

www.gnedenko.net

# DEPENDABILITY JOURNAL ARTICLE SUBMISSION GUIDELINES

***Article formatting requirements***

Articles must be submitted to the editorial office in electronic form as a Microsoft Office Word file (*.doc or *.docx extension). The text must be in black, on a A4 sheet with the following margins: 2 cm for the left, top and bottom margins; 1.5 or 2 cm for the right margin. An article cannot be shorter than 5 pages and longer than 12 pages (can be extended upon agreement with the editorial office). The article is to include the structural elements described below.

***Structure of the article***

The following structural elements must be separated with an *empty line*. Examples of how they must look in the text are shown *in blue*.

## 1) Title of the article

The title of the article is given in the English language.

*Presentation:* The title must be in 12-point Times New Roman, with 1.5 line spacing, fully justified, with no indentation on the left. The font face must be bold. The title is not followed by a full stop.

An example:

**Improving the dependability of electronic components**

## 2) Author(s)' name.

This structural element for each author includes:

In English: second name and first name as "First name, Second name" (John Johnson).

*Presentation:* The authors' names must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with no indentation on the left. The font face must be bold. The authors' names are separated with a comma. The line is not followed by a full stop.

*An example:*

**John Johnson[1], Karen Smith[2*]**

## 3) The author(s)' place of employment

The authors' place of employment is given in English. Before the place of employment, the superscripted number of the respective reference to the author's name is written.

*Presentation:* The reference to the place of employment must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with no indentation on the left. The font face must be normal. Each place of employment is written in a new line. The lines are not followed by a full stop.

*An example:*

[1] Moscow State University, Russian Federation, Moscow

[2] Saint Petersburg Institute of Heat Power Engineering, Russian Federation, Saint Petersburg

## 4) The e-mail address of the author responsible for maintaining correspondence with the editorial office

*Presentation:* The address must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with no indentation on the left. The font face must be normal, all symbols must be lower-case. Before the address reference, symbol * is written. The title is not followed by a full stop.

*An example:*

*johnson_j@aaa.net

## 5) Abstract of the article

This structural element includes a structured summary of the article with the minimal size of 350 words and maximum size of 400 words. The abstract is given in the English language. The abstract must include (preferably explicitly) the following sections: Aim; Methods; Results/Findings; Conclusions. The abstract of the article should not include newly introduced terms, abbreviations (unless universally accepted), references to literature.

*Presentation:* The abstract must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with no indentation on the left. The font face must be normal, except "**Abstract**", "**Aim**", "**Methods**", "**Conclusions**", that (along with the full stop) must be in bold. The text of the abstract must not be paragraphed (written in a single paragraph).

*An example:*

**Abstract.** Aim.Proposing an approach … taking into consideration the current methods. **Methods.** The paper uses methods of mathematical analysis,…, probability theory. **Results.** The following findings were obtained using the proposed method … **Conclusion**. The approach proposed in the paper allows…

## 6) Keywords

5 to 7 words associated with the paper's subject matter must be listed. It is advisable that the keywords complimented the abstract and title of the article. The keywords are written in English.

*Presentation:* The text must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with no indentation on the left. The font face must be normal, except "**Keywords:**" that (along with the colon) must be in bold. The text must not be paragraphed (written in a single paragraph). The text must be followed by a full stop.

*An example:*

**Keywords:** dependability, functional safety, technical systems, risk management, operational efficiency.

## 7) Text of the article

It is recommended to structure the text of the article in the following sections: Introduction, Overview of the sources, Methods, Results, Discussion, Conclusions. Figures and tables are included in the text of the article (the figures must be "In line with text", not "behind text" or "in front of text"; not "With Text Wrapping").

*Presentation:*

The titles of the sections must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with no indentation on the left. The font face must be bold. The titles of the sections (except the Introduction and Conclusions) may be numbered in Arabic figures with a full stop after the number of a section. The number with a full stop must be separated from the title with a no-break space (Ctrl+Shift+Spacebar).

The text of the sections must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with a 1.25-cm indent. The font face must be normal. The text of the sections must be paragraphed. There must be no indent in the paragraph that follows a formula and contain notes to such formula, e.g.:

where *n* is the number of products.

*An example:*

**1. State of the art of improving the dependability of electronic components**

An analysis of Russian and foreign literature on the topic of this study has shown that …

Figures (photographs, screenshots) must be of good quality, suitable for printing. The resolution must be at least 300 dpi. If a figure is a diagram, drawing, etc. it should be inserted into the text in editable form (Microsoft Visio). All figures must be captioned. Figures are numbered in Arabic figures in the order of their appearance in the text. If a text has one figure, it is not numbered. References to figures must be written as follows: "Fig. 3. shows that …" or "It is shown that … (see. Fig. 3.)." The abbreviation "Fig." and number of the figure (if any) are always separated with a no-break space (Ctrl+Shift+Spacebar). The caption must include the counting number of the figure and its title. It must be placed a line below the figure and center justified:

Fig. 2. Description of vital process

Captions are not followed by a full stop. *With center justification there must be no indent!* All designations shown in figures must be explained in the main text or the captions. The designations in the text and the figure must be identical (including the differences between the upright and oblique fonts). *In case of difficulties with in-text figure formatting, the authors must – at the editorial office's request – provide such figures in a graphics format (files with the* *.tiff, *.png, *.gif, *.jpg, *.eps extensions).

The tables must be of good quality, suitable for printing. The tables must be editable (not scanned or in image format). All tables must be titled. Tables are numbered in Arabic figures in the order of their appearance in the text. If a text has one table, it is not numbered. References to tables must be written as follows: "Tab. 3. shows that …" or "It is shown that … (see. tab. 3.)." The abbreviation "tab." and number of the table (if any) must be always separated with a no-break space (Ctrl+Shift+Spacebar). The title of a table must include the counting number and its title. It is placed a line above the table with center justification:

Table 2. Description of vital process

The title of a table is not followed by a full stop. *With center justification there must be no indent!* All designations featured in tables must be explained in the main text. The designations in the text and tables must be identical (including the differences between the upright and oblique fonts).

Mathematical notations in the text must be written in capital and lower-case letters of the Latin and Greek alphabets. Latin symbols must always be oblique, except function designators, such as sin, cos, max, min, etc., that must be written in an upright font. Greek symbols must always be written in an upright font. The font size of the main text and mathematical notations (including formulas) must be identical; in Microsoft Word upper and lower indices are scaled automatically.

Formulas may de added directly into the text, for instance:

Let $y = a \cdot x + b$, then…,

or written in a separate line with center justification, e.g.:

$$y = a \cdot x + b.$$

In formulas both in the text, and in separate lines, the punctuation must be according to the normal rules, i.e. if a formula concludes a sentence, it is followed by a full stop; if the sentence continues after a formula, it is followed by a comma (or no punctuation mark). In order to separate formulas from the text, it is recommended to set the spacing for the formula line 6 points before and 6 points after). If a formula is referenced in the text of an article, such formula must be written in a separate line with the number of the formula written by the right edge in round brackets, for instance:

$$y = a \cdot x + b. \quad (1)$$

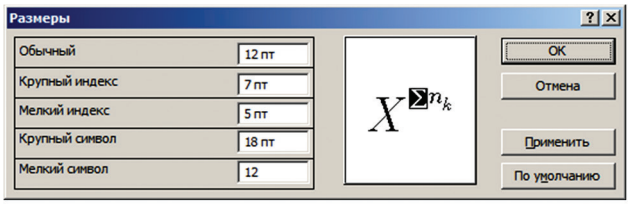
If a formula is written in a separate line and has a number, such line must be right justified, and the formula and its number must be tab-separated; tab position (in cm) is to be chosen in such a way as to place the formula roughly at the center. Formulas that are referenced in the text must be numbered in Arabic figures in the order of their appearance in the text.

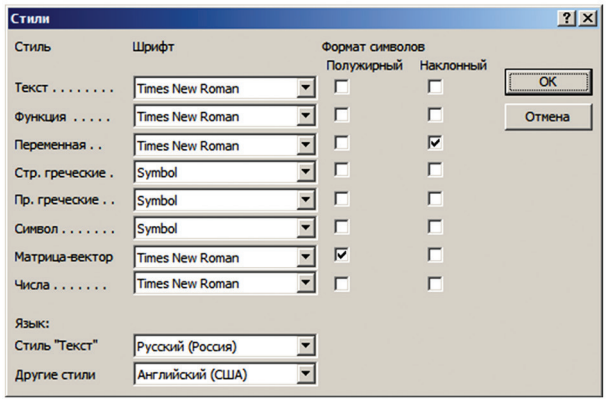
Simple formulas should be written without using formula editors (in MS Word, Latin should be used, as well as the "Insert" menu + "Special Characters", if Greek letters and mathematical operators are required), while observing the required slope for Latin symbols, for example:

$$\Omega = a + b \cdot \theta.$$

If a formula is written without using a formula editor, letters and +, –, = signs must be separated with no-break spaces (Ctrl+Shift+Spacebar).

Complex formulas must be written using a formula editor. In order to avoid problems when editing and formatting formulas it is highly recommended to use Microsoft Equation 3.0 or MathType 6.x. In order to ensure correct formula input (symbol size, slope, etc.), below are given the recommended editor settings.




When writing formulas in an editor, if brackets are required, those from the formula editor should be used and not typed on the keyboard (to ensure correct bracket height depending on the formula contents), for example (Equation 3.0):

$$Z = \frac{a \cdot \left( \sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_i \right)}{n+m}. \quad (2)$$

Footnotes in the text are numbered with Arabic figures, placed page by page. Footnotes may include: references to anonymous sources on the Internet, textbooks, study guides, standards, information from websites, statistic reports, publications in newspapers, magazines, autoabstracts, dissertations (if the articles published as the result of thesis research cannot be quoted), the author's comments.

References to bibliographic sources are written in the text in square brackets, and the sources are listed in the order of citation (end references). The page number is given within the brackets, separated with a comma and a space, after the source number: [6, p. 8].

## 8) Acknowledgements

This section contains the mentions of all sources of funds for the study, as well as acknowledgements to people who took part in the article preparation, but are not among the authors. Participation in the article preparation implies: recommendations regarding improvements to the study, provision of premises for research, institutional supervision, financial support, individual analytical operations, provision of reagents/patients/animals/other materials for the study.

*Presentation:*

The information must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with no indentation on the left. The font face must be normal.

## 9) References

The References must include only peer-reviewed sources (articles from academic journals and monographs) mentioned in the text of the article. It is not advised to references autoabstracts, dissertations, textbooks, study guides, standards, information from websites, statistic reports, publications in newspapers, websites and social media. If such information must be referred to, the source should be quoted in a footnote.

The description of a source should include its DOI, if it can be found (for foreign sources, that is possible in 95% of cases).

References to articles that have been accepted, but not yet published must be marked "in press"; the authors must obtain a written permission in order to reference such documents and confirmation that they have been accepted for publication. Information from unpublished sources must be marked "unpublished data/documents"; the authors also must obtain a written permission to use such materials.

References to journal articles must contain the year of publication, volume and issue, page numbers.

The description of each source must mention all of its authors.

The references, imprint must be verified according to the journals' or publishers' official websites.

*Presentation:*

References must be written in accordance with the Vancouver system.

The references must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with a 1.25-cm indent on the left. The font face must be normal. Each entry must be numbered in Arabic figures with a full stop after the number. The number with a full stop must be separated from the entry with a no-break space (Ctrl+Shift+Spacebar).

## 10) About the authors

Full second name, first name (in English); complete mailing address (including the postal code, city and country); complete name of the place of employment, position; academic degree, academic title, honorary degrees; membership in public associations, organizations, unions, etc.; official name of the organization in English; e-mail address; list and numbers of journals with the author's previous publications; the authors' photographs for publication in the journal.

*Presentation:*

The information must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with no indentation on the left. The font face must be normal.

## 11) The authors' contribution

Detailed information as to each author's contribution to the article. For example: Author A analyzed literature on the topic of the paper, author B has developed a model of real-life facility operation, performed example calculation, etc. Even if the article has only one author, his/her contribution must be specified.

*Presentation:*

The information must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with no indentation on the left. The font face must be normal.

## 12) Conflict of interests

A conflict of interests is a situation when people have conflicting and competing interests that may affect editorial decisions. Conflicts of interests may be potential or conscious, as well as actually existing. The objectivity may be affected by personal, political, financial, scientific or religious factors.

The author must notify the editorial office on an existing or a potential conflict of interests by including the corresponding information into the article.

If there is no conflict of interests, the author must also make it known. An example of wording: "The author declares the absence of a conflict of interests".

*Presentation:*

The text must be in 12-point Times New Roman, with a 1.5-line spacing, fully justified, with no indentation on the left. The font face must be normal.

**THE JOURNAL IS PUBLISHED WITH PARTICIPATION AND SUPPORT**
OF JOINT-STOCK COMPANY RESEARCH & DESIGN INSTITUTE FOR INFORMATION TECHNOLOGY, SIGNALLING AND TELECOMMUNICATIONS ON RAILWAY TRANSPORT
(JSC NIIAS)

**JSC NIIAS** is RZD's leading company in the field of development of train control and safety systems, traffic management systems, GIS support technology, railway fleet and infrastructure monitoring systems

**Mission:**

transportation

- ❑ efficiency,
- ❑ safety,
- ❑ reliability

## Key areas of activity

- Intellectual control and management systems
- Transportation management systems and transport service technology
- Signalling and remote control systems
- Automated transportation management centers
- Railway transport information systems
- Geoinformation systems and satellite technology
- Transport safety systems
- Infrastructure management systems
- Power consumption and energy management systems
- Testing, certification and expert assessment
- Information security
- Regulatory support

**www.vniias.ru**