НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ»
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МІЖНАРОДНИЙ НАУКОВО–ТЕХНІЧНИЙ ЖУРНАЛ

№ 3 2025

ЗАСНОВАНО У ЛИПНІ 2001 р.

АДРЕСА РЕДАКЦІЇ:
03056, м. Київ,
просп. Перемоги, 37, корп. 35,
ННК «ІПСА» КПІ ім. Ігоря Сікорського
Тел.: **204-81-44**; факс: **204-81-44**
E-mail: journal.iasa@gmail.com
http://journal.iasa.kpi.ua

У номері:

NATIONAL ACADEMY OF SCIENCES OF UKRAINE
EDUCATIONAL AND SCIENTIFIC COMPLEX
«INSTITUTE FOR APPLIED SYSTEM ANALYSIS»
OF THE NATIONAL TECHNICAL UNIVERSITY OF UKRAINE
«IGOR SIKORSKY KYIV POLYTECHNIC INSTITUTE»

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

INTERNATIONAL SCIENTIFIC AND TECHNICAL JOURNAL

№ 3 2025

IT IS FOUNDED IN JULY 2001

THE EDITION ADDRESS:

03056, Kyiv,
av. Peremogy, 37, building 35,
Institute for Applied System Analysis
at the Igor Sikorsky Kyiv Polytechnic Institute
Phone: **204-81-44**; Fax: **204-81-44**
E-mail: journal.iasa@gmail.com
http://journal.iasa.kpi.ua

In the issue:

Шановні читачі!

Навчально-науковий комплекс «Інститут прикладного системного аналізу» Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» видає міжнародний науково-технічний журнал

«СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ».

Журнал публікує праці теоретичного та прикладного характеру в широкому спектрі проблем, що стосуються системних досліджень та інформаційних технологій.

Провідні тематичні розділи журналу:

Теоретичні та прикладні проблеми і методи системного аналізу; теоретичні та прикладні проблеми інформатики; автоматизовані системи управління; прогресивні інформаційні технології, високопродуктивні комп'ютерні системи; проблеми прийняття рішень і управління в економічних, технічних, екологічних і соціальних системах; теоретичні та прикладні проблеми інтелектуальних систем підтримання прийняття рішень; проблемно і функціонально орієнтовані комп'ютерні системи та мережі; методи оптимізації, оптимальне управління і теорія ігор; математичні методи, моделі, проблеми і технології дослідження складних систем; методи аналізу та управління системами в умовах ризику і невизначеності; евристичні методи та алгоритми в системному аналізі та управлінні; нові методи в системному аналізі, інформатиці та теорії прийняття рішень; науково-методичні проблеми в освіті.

Головний редактор журналу — науковий керівник ННК «ІПСА» КПІ ім. Ігоря Сікорського, академік НАН України Михайло Захарович Згуровський.

Журнал «Системні дослідження та інформаційні технології» включено до переліку наукових фахових видань України (категорія «А»).

Журнал «Системні дослідження та інформаційні технології» входить до таких наукометричних баз даних: Scopus, EBSCO, Google Scholar, DOAJ, Index Copernicus, реферативна база даних «Україніка наукова», український реферативний журнал «Джерело», наукова періодика України.

Статті публікуються українською та англійською мовами.

Журнал рекомендовано передплатити. **Наш індекс 23918.** Якщо ви не встигли передплатити журнал, його можна придбати безпосередньо в редакції за адресою: 03056, м. Київ, просп. Перемоги, 37, корп. 35.

Завідувачка редакції **С.М. Шевченко**
Редакторка **Р.М. Шульженко**
Молодша редакторка **Л.О. Тарин**
Комп'ютерна верстка, дизайн **А.А. Патіохи**

Рішення Національної ради України з питань телебачення і радіомовлення №1794 від 21.12.2023. Ідентифікатор медіа R30-02404

Підписано до друку 29.09.2025. Формат 70×108 1/16. Папір офс. Гарнітура Times.
Спосіб друку – цифровий. Ум. друк. арк. 14,411. Обл.-вид. арк. 28,56. Наклад 153 пр. Зам. № 11/04

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Свідоцтво про державну реєстрацію: ДК № 5354 від 25.05.2017 р.
просп. Перемоги, 37, м. Київ, 03056.
ФОП Пилипенко Н.М., вул. Мічуріна, б. 2/7, м. Київ, 01014. тел. (044) 361 78 68.
Виписка з Єдиного державного реєстру № 2 070 000 0000 0214697 від 17.05.2019 р.

Dear Readers!

Educational and Scientific Complex «Institute for Applied System Analysis» of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» is published of the international scientific and technical journal

«SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES».

The Journal is printing works of a theoretical and applied character on a wide spectrum of problems, connected with system researches and information technologies.

The main thematic sections of the Journal are the following:

Theoretical and applied problems and methods of system analysis; theoretical and applied problems of computer science; automated control systems; progressive information technologies, high-efficiency computer systems; decision making and control in economic, technical, ecological and social systems; theoretical and applied problems of intellectual systems for decision making support; problem- and function-oriented computer systems and networks; methods of optimization, optimum control and theory of games; mathematical methods, models, problems and technologies for complex systems research; methods of system analysis and control in conditions of risk and uncertainty; heuristic methods and algorithms in system analysis and control; new methods in system analysis, computer science and theory of decision making; scientific and methodical problems in education.

The editor-in-chief of the Journal is scientific director of the Institute for Applied System Analysis at the Igor Sikorsky Kyiv Polytechnic Institute, academician of the NASU Michael Zaharovich Zgurovsky.

The articles to be published in the Journal in Ukrainian and English languages are accepted. Information printed in the Journal is included in the Catalogue of periodicals of Ukraine.

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

3 • 2025

ЗМІСТ

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

3 • 2025

CONTENT

UDC 004.896:629.735.33
DOI: 10.20535/SRIT.2308-8893.2025.3.01

COGNITIVE AI PLATFORM FOR AUTONOMOUS NAVIGATION OF DISTRIBUTED MULTI-AGENT SYSTEMS

M.Z. ZGUROVSKY, P.O. KASYANOV, N.D. PANKRATOVA, Yu.P. ZAYCHENKO, I.O. SAVCHENKO, T.V. SHOVKOPLYAS, L.S. PALIICHUK, A.M. TYTARENKO

Abstract. This paper presents a concept for a cognitive AI platform that enables autonomous navigation of distributed multi-agent systems, exemplified by UAV swarms. The proposed architecture integrates a ground control center with cognitive services and a multi-layered onboard subsystem, supporting a continuous loop of learning, adaptation, execution, and behavioral model updates. Several core mission scenarios are introduced, such as reconnaissance, search and rescue, target neutralization, and deception, showcasing the swarm's ability to operate autonomously and in a decentralized manner, even under adversarial conditions. An example of a search and rescue mission implementation plan using a cognitive platform that includes adaptive planning, SLAM navigation, swarm coordination, and deep object recognition is presented. The results were partially supported by the National Research Foundation of Ukraine, grant No. 2025.06/0022 "AI platform with cognitive services for coordinated autonomous navigation of distributed systems consisting of a large number of objects".

Keywords: artificial intelligence, UAV swarm, autonomous navigation, cognitive platform, multi-agent systems, behavior trees, digital twin, SLAM.

INTRODUCTION

In modern conditions of increasingly complex combat environment, active electronic warfare, and loss of reliable satellite connection network, a critical need arises for creating autonomous, decentralized control framework for distributed systems, particularly swarms of unmanned aerial vehicles (UAV). In this context the development of a cognitive AI platform, capable of guaranteeing the coordinated navigation of a multitude of agents prohibited from interaction with a centralized control point or external infrastructure, becomes especially important [1–3]. This kind of environment requires not only sufficient autonomy level of individual agents (drones), but also a wholesome approach to the organization of their collective behavior implemented through cognitive self-learning, self-organization, adaptation algorithms, and resilient inter-agent information exchange. The theoretical and methodological basis for constructing this kind of platform was described in [4–10], in particular the impossibility of full consistency of agents: swarm agents cannot have a fully coordinated movement direction on spherical surfaces (as well as on large single-connected compact manifold surfaces without edges, including geoids), which compromises at least ant colony algorithms, requiring the selection of special points as regrouping zones [5, Theorem 1].

The AI platform for autonomous navigation of distributed multi-agent systems is viewed as an integral architecture that combines two closely interconnected components: the on-board component functioning directly at each of the autonomous agents, particularly the UAV, and the ground control center that provides learning, simulation, validation and strategic system control. Both components are functionally and logically interconnected, and together they form a cognitive AI platform in a broad sense – as an intellectual, self-learning architecture, capable of adaptation to the changes in environment, and self-improvement on the basis of accumulated experience.

The on-board component of the AI platform provides the completely autonomous functioning of its agents. It implements the capability for independent navigation without the GPS (Global Positioning System) satellite signals, making decisions in real time, decentralized swarm coordination, and adaptation in case of losing individual agents, or changes in the environment. Its functioning is based on the on-board AI modules, sensor systems, stygmergy algorithms, decentralized planning, reinforcement learning methods, self-learning and self-organization, SLAM (Simultaneous Localization and Mapping) methods, and other modern approaches [11–13]. This component in particular implements the cognitive behavior during missions: each drone is able to orient itself, perform the assigned tasks, and interact with other swarm agents without centralized control.

The ground control center performs the role of the strategic brain center of the system. It provides both primary, and cyclical training of the neural networks, modeling mission scenarios in the simulation environment using digital twins [14–17], testing and validation of the models, as well as the generation of the behavioral politics for on-board implementation. The ground center aggregates information from OSINT/ESINT sources, adapts the models to the operational context using analytics, supports visualization, monitoring and strategic correction. Through secure human-machine interface the operator obtains access to parametrization of missions, system state management, and updates to the AI modules software.

The interaction between the on-board and ground systems is organized as a closed cognitive loop. In the pre-missionary phase, the ground control center implements the training of models, mission modeling; creates the digital twins for drones, and uploads the updated algorithms to the on-board systems. This process involves analytical modules that aggregate OSINT (Open-Source Intelligence) for adaptation to the current context. During missions, the drones operate autonomously, performing swarm coordination, and in case the secure connection is available, transmit telemetry to the center which conducts monitoring and provides corrections if necessary. After the mission, the collected data is analyzed, log files are checked for anomalies, the models are tweaked, and the new cycle of training is started. Thus, the system is capable of continuous cognitive evolution – it learns on its own experience, gradually increasing the efficiency and resilience to new challenges of modern combat environment.

The cognitive AI platform is the only intellectual architecture system that includes ground and on-board components that jointly form the adaptive and viable complex for coordinated autonomous navigation of a UAV swarm. This complex functions within a continuous loop of adaptation and improvement, encompassing pre-mission preparation, autonomous task completion, post-mission analysis, and further additional training. This loop implements the concept of a cognitive core as a system capable of forming, updating and generalizing knowledge based on its own experience, react to the variable conditions, support collective behavior of agents, and retain efficiency in a complex, dynamic, and hostile environment.

The purpose of this research is to create architecture and principles of the system operation where each UAV behaves as an autonomous cognitive agent, capable of navigating without GPS, make decisions based on the local information, exchange signals with its neighbors using stygmergy or a mesh network, while acting within a single coordinated environment (the swarm). The construction of a new generation cognitive AI platform that combines adaptivity, resilience and scalability, is envisioned, enabling the UAV swarm to operate independently of external control, and efficiently complete the assigned tasks (missions) even under critical circumstances. This research is aimed at implementing the swarm intelligence in defense and rescue technologies, and forms the theoretical and engineering base for the next generation of double purpose autonomous systems.

THE GROUND CONTROL CENTER FOR THE AI PLATFORM WITH COGNITIVE SERVICES

The ground control center for neural network training is a critical architecture element of the general AI platform for cognitive control of the autonomous drone swarm. It performs the functions of development, testing, adaptation, security check, and preparation of the behavior strategies and cognitive models that will be uploaded to each of the drones before the actual mission assignment. The structure of this center is modular, logically decentralized, but centralized by computational power. It includes the following main functional blocks (Fig. 1):

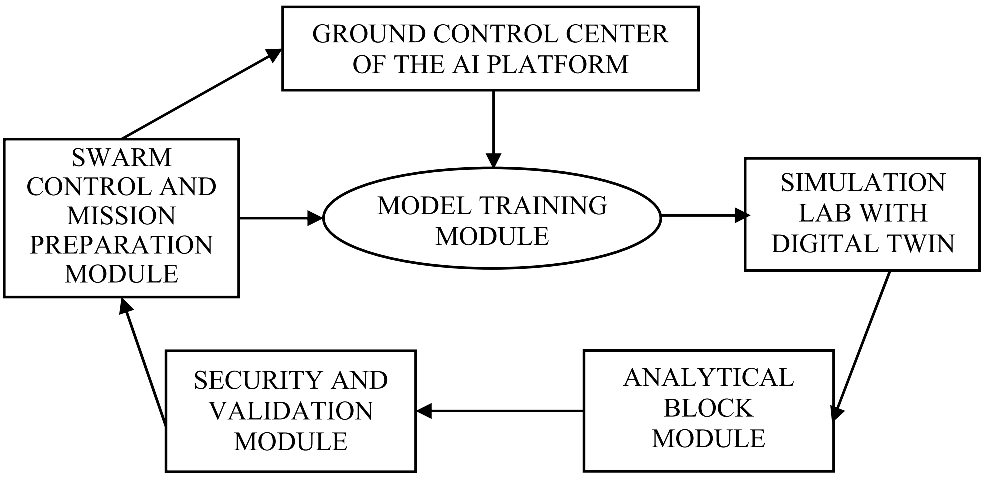


Fig. 1. The architecture of the ground center of the AI platform

Model training module. This block is responsible for the primary and recurrent training of the neural networks that will be applied in drone systems. The technologies involved include Reinforcement Learning models, self-learning models, perception models for detection and tracking of objects, as well as graph neural networks (GNN) for optimization of behavior in swarm configurations. The training is performed both on the historical data, and the data obtained during previous missions.

Simulation lab with digital twin. The digital twin of the ground center is a critical element of the general AI platform architecture that allows to test the neural network behavior in complex and variable scenarios. Here both the standard situations are simulated, and the stress scenarios, including the loss of the swarm elements, navigation under interference, electronic warfare conditions. This stage provides adaptivity and resilience of the trained behavior even before the real operation.

Analytical block. This module conducts the analysis of open-source data (OSINT). Analytical insight regarding the potential risks, typical tactics of the enemy, or features of the mission territory can be promptly integrated in the process of preparation for the real mission, increasing the relevance of the drone behavior. This may include, in particular, the location of the notable objects, relevant mission details, maps etc.

Security and validation module. Following the primary training, all models are tested to ensure they meet resilience, security, and durability requirements. In particular, this check includes a model's capability of detecting anomalies, restoration after errors, resilience to attacks at the data level, connection channels, and model integrity. Validation is the obligatory stage before the mission implementation.

Swarm control and mission preparation module. This block represents the control interface that aggregates the results from all other blocks and prepares the behavior model for uploading to the drones; forms the detailed missions; distributes the tasks among agents; plans the route networks; defines the zonal priorities. This module is used to upload the prepared cognitive software to the drones before their assignment to the real or test mission. The center also performs the functions of the swarm state monitoring, interactive control, and strategy adaptation in real time.

As the Fig. 1 shows, the interaction between the sub-systems of the ground control center is organized as a closed cognitive loop that guarantees the wholesome functioning of the drone swarm control system. In this loop the models formed in the training module are automatically transferred to the simulation lab, where they are subject to testing under the circumstances as close as possible to the real environment. The simulation results are analyzed by the validation module that makes the decision regarding the fitness of the models for combat use. The OSINT module works in parallel, generating the contextual scenarios using open-source intelligence data; these scenarios are integrated into the training processes, increasing adaptivity and relevance of the trained models.

When the neural networks complete all verification stages, the swarm control center uploads them on-board of the drones, initiating missions, and performing their accompaniment, monitoring and correction in real time. Thus, the ground center acts as a "cognitive foundry" of the system – the environment where the artificial intelligence is not only created but also evolves under the influence of the new data, combat experience, and strategical analysis. Here the intellectual potential of the swarm is formed, allowing the drones to act as intelligent autonomous agents with high adaptation abilities, mutual understanding, and collective behavior in the complex and hostile environment.

BASIC SCENARIOS (MISSIONS) FOR THE AUTONOMOUS NAVIGATION OF THE DISTRIBUTED MULTI-AGENT SYSTEMS

In modern combat and rescue conditions the scenarios for the drone swarm constitute the basis for the cognitive behavior of the autonomous agents that function within the integral AI platform. These scenarios are not just simple instructions – they represent the structured, multi-component algorithmic descriptions, preliminarily modeled in the ground control center. Due to the involvement of digital simulation environments (such as Gazebo or AirSim), analytical modeling, mission planning tools (such as QGroundControl), and machine learning methods, the scenarios achieve high adaptivity to the complex and dynamic environment. After modeling they are saved in JSON, XML, TensorRT, ONNX [18] etc. formats, and are uploaded to the computational blocks of each drone through a secure channel before the mission starts.

The content of these scenarios includes several critically important functional blocks: mission planning, autonomous navigation, recognition, decision making, and swarm coordination. The planner contains the vectorized task description, temporal parameters, action sequences, and defined objectives. The autonomous navigation modules provide route planning in real time using SLAM, localization and obstacle avoidance algorithms. The recognition components are responsible for the processing of sensor data from cameras, thermal imagers and radars, allowing them to detect objectives, obstacles and threats. The decision making is implemented through cognitive models capable of situational analysis, and producing reactions based on environment assessment. Finally, the swarm coordination provides the dynamic distribution of roles between agents, syncing of the trajectories, and coordinated behavior within the swarm [19].

The unique nature of these scenarios lies in their ability to activate the onboard drone cognitive modules that provide adaptive behavior even in case of the absent connection to the control center, external interference, or the shifting environment. In other words, the drones not only implement the previously assigned actions, but also learn from the current situation, predict risks, and react collectively. This is made possible by the integration of reinforcement learning methods, graph neural networks, and large language models that enable flexible, situational cognition at the swarm level [20].

Let us provide a list of basic scenarios:

Scenario 1: *Enemy territory reconnaissance.* The drone swarm distributes the reconnaissance area (e.g., 10×10 km), with each sub-area assigned to an individual drone. The results are obtained as a shared locality map. The scenario is performed by 6–12 drones that cover up to 100 km^2 in 15–40 minutes.

Scenario 2: *Targeted strike with autonomous guidance.* Several drones attack the target from different directions, overcoming air defenses by dispersed planning. Up to 7 drones attack the target's coordinates in 3–10 minutes after its detection.

Scenario 3: *Communication relay.* The drone swarm creates a temporary mesh network, providing connection under electronic warfare. For example, 5–15 drones create a 5–10 km long linear network, providing communication for 20–60 minutes.

Scenario 4: *Search and rescue*. The swarm autonomously scouts the destruction zone, detecting people and animals by performing scanning with distribution of routes. Up to 20 drones are used, with coverage area 10–40 km^2 for one hour.

Scenario 5: *"Death ring" swarm attack.* The drones fly round the target from all directions, forming a ring, and strike it simultaneously. In this scenario 5–10 drones are used, with 100–500 m attack radius during 5–15 minutes.

Scenario 6: *Scattering false targets/misinformation.* The swarm scatters imitation objects to mislead the enemy or mask the actual swarm's goals, by performing a coordinated placement of false targets (vehicle imitations), or modeling the behavior of a real vehicle column. During 10–30 minutes 5–10 drones place signal imitators along the 20 km route, using GPS and waypoint navigation (a drone moves from one waypoint to another in a predetermined sequence).

The compiled scenario (mission) parameters are given in Table 1.

Table 1. The compiled scenario (mission) parameters for UAV swarms

Scenario	Drone quantity	Surface/length of coverage	Duration	Communication/ protocol
Scenario 1. *Enemy territory reconnaissance*	6–12	Up to 100 km²	15–40 minutes	DDS or ROS topics + sensors (LIDAR/camera)
Scenario 2. *Targeted strike with autonomous guidance*	Up to 7	Depends on target (up to 10 km)	3–10 minutes	MAVLink/mesh connection
Scenario 3. *Communication relay*	5–15	5–10 km	20–60 minutes	DDS+RTPS with real-time QoS profile
Scenario 4. *Search and rescue*	Up to 20	10–40 km²	Up to 1 hour	ROS topics + thermal imager
Scenario 5. *Death ring*	5–10	Attack radius up to 500 m	5–15 minutes	ROS2 + DDS
Scenario 6. *Scattering false targets*	5–10 depending on the route	Up to 20 km route	10–30 minutes	MAVLink with waypoint navigation

So, the scenarios for the drone swarm become the key element for the cognitive AI platform, combining high precision planning, realistic simulation, analytical adaptation, and self-learning. Their exploitation not only increases mission efficiency, but also provides resilience to the uncertainty factors, which is critical in the environment where each second and each decision is significant.

ON-BOARD COMPONENT OF THE AI PLATFORM WITH COGNITIVE SERVICES

The on-board component is a key functional environment where the autonomous intelligence of each drone in the swarm is implemented. This is the place where the integration of cognitive models, sensory perception, swarm interaction, flight control, and adaptive decision making in real time is performed. The architecture of this component (Fig. 2) is multi-layered and includes a number of modules that jointly ensure the independence of the drone from external control, its self-learning capacity, and flexible reaction to a dynamic environment.

Let us consider each module of the architecture presented in Fig. 2, reviewing its functions, and their mutual interaction.

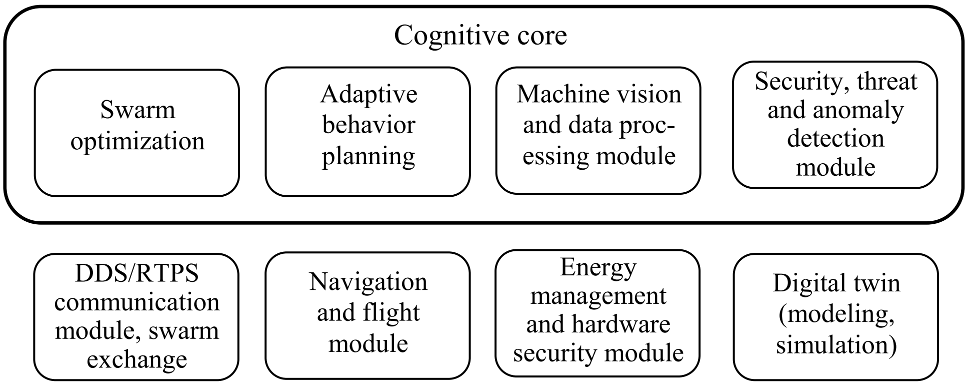


Fig. 2. The structural scheme of the on-board component of the AI platform with cognitive services

The center of the architecture is comprised of the cognitive core that acts as a drone's "brain", and is responsible for situational analysis, adaptation and decision making. Its fundament is the *Swarm coordination* module, implemented using the hybrid approach where the swarm AI methods are applied using a hybrid scheme: the Behavioral trees (BTs), and Global swarm optimization (Global Best PSO) that can reconfigure in real time depending on the changes in the environment [21–22]. This allows each agent to form the sequence of actions, independently react to the loss of communication, emergence of new threats, or changes in objectives.

Combined with the *Adaptive behavior planning* module that analyzes risks, priorities and current context, the system acquires the ability for conscious decision making even having incomplete information. It performs the incremental onboard learning (given the appropriate resources), bufferization of the field data, and the backhaul retraining loop implementation – the transmission of the collected data to the ground control center, with the subsequent updates in the models. This mechanism forms the basis of the system evolution, as it allows to take previous experience into account in the future missions. This approach allows to coordinate local trajectories, synchronize agent sub-groups, and sustain the overall mission goals at the lower autonomy level.

To enable these cognitive processes, the drone requires a constant flow of information about the environment. This task is achieved by the *Machine vision and data processing* module that aggregates the data from cameras, ultra-sonic sensors etc., forming the local space maps using SLAM algorithms [23–24]. An important feature of this layer is its capability for semantic classification of the objects (e.g., enemy units, civilians, allied units), and detection of the situational patterns that allow to construct not only a spatial, but also a behavioral model of the environment.

For coordinated interaction among the swarm elements, the platform contains the *communication* module, based on low latency DDS/RTPS protocols. It provides the interchange of statuses between agents by behavioral subtree broadcast, and allows to maintain the swarm coordination without the centralized control [25–27]. Even in case of losses or disruption in network channels the module remains operational due to the QoS control network that allows to duplicate critical data, and adapt priorities.

The physical implementation of the cognitive core is done by the *Navigation and flight* module that is the interface to the autopilots like PX4 or ArduPilot. It performs maneuvers, passing route points and avoiding obstacles, while relying on the visual odometry and SLAM data to ensure collision safety.

At the same time, the *Security, threat and anomaly detection* module is responsible for self-observation: temperature monitoring, CPU/GPU load, system degradation detection, and activates fail-safe scenarios, or dynamically reschedules the swarm tasks in case of losses of individual agents. Detecting anomalies in time series of sensory indicators allows the system to automatically react to potential threats, detect compromised swarm participants, analyzing the irregular patterns in input data. This approach is more flexible than the traditional heuristic rules in robotized systems [4].

A strategically important link is the *Digital twin* module – a limited representation of a fully functional digital twin deployed in the ground control center. On-board this module is responsible for maintaining the relevant strategies, simulation of the partial actions, and asynchronous renewal of the behavioral models [14–17]. It guarantees the autonomous behavior even in case of a complete connection loss, synchronizing data later.

Finally, the stability and security of the system is sustained by the *Energy management and hardware security* module that includes communication encryption, agent authentication, multi-layered service backup, and power management. This module allows the system to adapt to power supply limitations, lowering the sensor operation intensity, or switching to the energy-saving mode in critical moments. The whole multi-layered system provides the autonomous, adaptive and resilient UAV swarm operation even in hostile or unpredictable environment, implementing the modern approaches to the on-board cognitive management.

SCENARIO 4 (SEARCH AND RESCUE) IMPLEMENTATION PLAN EXAMPLE

The operational situation: after a large-scale earthquake in some region several settlements were ruined. There is a risk of further collapses, and the access for the ground rescue groups is limited. An autonomous scanning of territory with a total area of nearly 30 km^2 is required to find the victims, designate safe evacuation zones, and transmit the coordinates to the ground forces.

The employment of the AI platform. To implement the scenario, a swarm system of 16 autonomous quadcopters will be deployed. The drones will be equipped with thermal imagers, RGB cameras, and laser rangefinders (LiDARs). The computational platform of each drone allows local image processing, map charting, and decision making. SLAM navigation, along with visual odometry and obstacle avoidance module, will be used to form local maps, and dynamically plan routes in real time. The behavioral coordination in the swarm will be implemented on the base of combined Behavior Trees and Graph Neural Networks that will allow adaptively distribute the tasks between agents, avoid duplication of the search zones, and optimize the area coverage.

The platform will ensure:

- distribution of the swarm into sub-groups of 4 drones with partial (~10%) overlap of the areas for increased probability of object detection;
- detection of heat anomalies using a pre-trained neural network;
- suppressing background noise (e.g., heat from transport or infrastructure);
- exchanging scanned area tags, and analysis results between participants.

For synchronization of the swarm behavior the implementation of the sub-tree broadcast protocol is planned that will transmit the minimal context every few seconds. Communication between agents is planned to be achieved through the ROS Topics + DDS with QoS parameters stack, providing reliable data exchange.

The expected data to be utilized includes:

- previous mission simulation models, formed on the base of satellite image data, topographical data and OSINT;
- fallback behavior scenarios for cases of connection loss or situation change.

The transmission to the ground center is conducted through relay drones that hover at up to 120 m height and form the mesh network. They transmit:

- local maps;
- visual confirmations;
- coordinates of detected objects and safe areas;
- GPS/SLAM log files.

The expected results include:

- detection of the potentially alive targets using thermal signatures;
- coverage map charting, and marking the risk areas;
- designation of safe routes for evacuation;
- transmission of the structured coordinates and statuses to the operational headquarters.

CONCLUSIONS

1. The developed AI platform for the autonomous navigation of UAV swarms presents a fundamentally new approach to handling the distributed multi-agent systems under conditions of a complex, dynamic, and hostile environment. Its architecture combines the ground control center, and the autonomous on-board subsystem, providing a continuous loop of adaptation, learning and evolution for artificial intelligence during each of the mission stages, from pre-mission modeling, to post-mission analysis. The ground control center performs the functions of simulation, training, validation and strategic coordination, while each drone, due to its cognitive core, sensory stack and communication modules, implements autonomous navigation, recognition, and decision making without centralized control.

2. A number of basic scenarios (missions) is formed that cover a broad spectrum of combat and humanitarian tasks. These scenarios include both classic objectives (reconnaissance, targeted strikes, communication relay), and specialized missions (search and rescue, misinformation, "death ring" strike), proving the platform's scalable and universal nature in dynamic environments. Formalization and typification of such scenarios allow to not just quickly adapt the swarm to new conditions but also form a repository for behavioral patterns that will be improved using the principles of cognitive learning over time.

3. An on-board component of the AI platform with cognitive services was developed by combining a cognitive core, a sensor and analytical layer, navigation, communication, and security modules. Each drone in the system can act independently, adapt to the changes in environment, make critical decisions in real time, and interact with other agents without centralized control. The hybrid application of the AI swarm intelligence methods "Behavior Trees" and "Global swarm optimization", and SLAM methods provides situational prediction and flexible reaction. The availability of power management, self-observation, and local knowledge updates additionally fortifies the system's survivability, and the digital twin module provides the asynchronous swarm evolution even after connection loss. All these functional capabilities prove that the on-board component is not just a computational node, but an accomplished cognitive agent, able to conduct missions within the decentralized new generation architecture.

4. A model search and rescue scenario of people after a catastrophe is proposed, where a drone swarm autonomously scans the investigated area, detects heat anomalies, identifies casualties, and transmits the coordinates for evacuation.

In preparing this manuscript, we used ChatGPT 4.0 to improve the style.

REFERENCES

1. S.-J. Chung, A.A. Paranjape, P. Dames, S. Shen, V. Kumar, "A survey on aerial swarm robotics," *IEEE Transactions on Robotics*, 34(4), pp. 837–855, 2018. doi: https://doi.org/10.1109/TRO.2018.2857475
2. M. Brambilla, E. Ferrante, M. Birattari, M. Dorigo, "Swarm robotics: A review from the swarm engineering perspective," *Swarm Intelligence*, 7(1), pp. 1–41, 2013. doi: https://doi.org/10.1007/s11721-012-0075-2
3. Z. Xia, J. Du, C. Jiang, J. Wang, Y. Ren, G. Li, "Multi-UAV cooperative target tracking based on swarm intelligence," in *ICC 2021 – IEEE International Conference on Communications*, pp. 1–6. doi: https://doi.org/10.1109/ICC42927.2021.9500771
4. A. Tytarenko, "Detecting unsafe behavior in neural network imitation policies for caregiving robotics," *System Research and Information Technologies*, no. 4, pp. 86–96, 2024. doi: https://doi.org/10.20535/SRIT.2308-8893.2024.4.07
5. P.O. Kasyanov, L.S. Paliichuk, "Acute angle lemma for noncompact image sets," *Journal of Fixed Point Theory and Applications*, 27(3), Article 67, 2025. doi: https://doi.org/10.1007/s11784-025-01220-4
6. M.Z. Zgurovsky, P.O. Kasyanov, L.B. Levenchuk, "Formalization of Methods for the Development of Autonomous Artificial Intelligence Systems," *Cybern. Syst. Anal.*, 59, pp. 763–771, 2023. doi: https://doi.org/10.1007/s10559-023-00612-z
7. M.Z. Zgurovsky, Y.P. Zaychenko, *Big Data: Conceptual Analysis and Applications*. Springer, 2020. doi: https://doi.org/10.1007/978-3-030-14298-8
8. M.Z. Zgurovsky, Y.P. Zaychenko, *The Fundamentals of Computational Intelligence: System Approach* (Studies in Computational Intelligence, Vol. 652). Springer, 2017. doi: https://doi.org/10.1007/978-3-319-35162-9
9. N.D. Pankratova, K.D. Grishyn, V.E. Barilko, "Digital twins: Stages of concept development, areas of use, prospects," *System Research and Information Technologies*, no. 2, pp. 7–21, 2023. doi: https://doi.org/10.20535/SRIT.2308-8893.2023.2.01
10. N.D. Pankratova, I.M. Golinko, "Development of digital twins to support the functioning of cyber-physical systems," *Computer Science Journal of Moldova,* 31(3(93)), pp. 299–320, 2023. doi: https://doi.org/10.56415/csjm.v31.15
11. Y. Alqudsi, M. Makaracı, "UAV swarms: Research, challenges, and future directions," *Journal of Engineering and Applied Science*, 72, Article 12, 2025. doi: https://doi.org/10.1186/s44147-025-00582-3

12. R. Arranz, D. Carramiñana, G. de Miguel, J.A. Besada, A.M. Bernardos, "Application of deep reinforcement learning to UAV swarming for ground surveillance," *Sensors*, 23(21), 8766, 2025. doi: https://doi.org/10.3390/s23218766
13. L. Tan et al., "Multi-UAV-enabled collaborative edge computing: Deployment, offloading and resource optimization," *IEEE Transactions on Intelligent Transportation Systems*, 25(11), pp. 13741–13754, 2022. doi: https://doi.org/10.1109/TITS.2024.3432818
14. Z. Li, L. Lei, G. Shen, X. Liu, X. Liu, "Digital twin-enabled deep reinforcement learning for safety-guaranteed flocking motion of UAV swarm," *Transactions on Emerging Telecommunications Technologies*, 35(11), e70011, 2024. doi: https://doi.org/10.1002/ett.70011
15. L.R. Salinas, G. Tzoumas, L. Pitonakova, S. Hauert, "Digital twin technology for wildfire monitoring using UAV swarms," in *2023 International Conference on Unmanned Aircraft Systems (ICUAS)*, IEEE, 2023, pp. 586–593. doi: https://doi.org/10.1109/ICUAS57906.2023.10155819
16. B.R. Barricelli, E. Casiraghi, D. Fogli, "A survey on digital twin: Definitions, characteristics, applications, and design implications," *IEEE Access*, 7, pp. 167653–167671, 2019. doi: https://doi.org/10.1109/ACCESS.2019.2953499
17. T. Li et al., "Digital twin-based task-driven resource management in intelligent UAV swarms," *IEEE Transactions on Intelligent Transportation Systems*, 26(4), pp. 3905–3918, 2025. doi: https://doi.org/10.1109/TITS.2025.3531120
18. T. Bray, J. Paoli, C.M. Sperberg-McQueen, E. Maler, F. Yergeau, "Extensible Markup Language (XML) 1.0," *W3C*, 2008. Avaialble: https://www.w3.org/TR/xml/
19. H.R. Ahmed, J.I. Glasgow, *Swarm intelligence: Concepts, models and applications; Technical Report 2012-585*. Queen's University, School of Computing Technical Reports, 2012. doi: https://doi.org/10.13140/2.1.1320.2568
20. D. Vernon, "Cognitive system," in K. Ikeuchi (Ed.), *Computer Vision*. Springer, Cham, 2021. doi: https://doi.org/10.1007/978-3-030-63416-2_82
21. H. Huang, J. Qiu, K. Riedl, "On the global convergence of particle swarm optimization methods," *Applied Mathematics & Optimization*, 88, 30, 2023. doi: https://doi.org/10.1007/s00245-023-09983-3
22. M. Iovino, E. Scukins, J. Styrud, P. Ögren, C. Smith, "A survey of behavior trees in robotics and AI," *Robotics and Autonomous Systems*, 154, 104096, 2022. doi: https://doi.org/10.1016/j.robot.2022.104096
23. C. Theodorou, V. Velisavljevic, V. Dyo, F. Nonyelu, "Visual SLAM algorithms and their application for AR, mapping, localization and wayfinding," *Array*, 15, 100222, 2022. doi: https://doi.org/10.1016/j.array.2022.100222
24. P. Su, S. Luo, X. Huang, "Real-time dynamic SLAM algorithm based on deep learning," *IEEE Access*, 10, pp. 87264–87276, 2022. doi: https://doi.org/10.1109/ACCESS.2022.3199350
25. "Data Distribution Service (DDS) Specification Version 1.4," *Object Management Group*, 2015. doi: https://www.omg.org/spec/DDS/
26. B. Tekinerdogan, Ö. Köksal, T. Çelik, "Data distribution service-based architecture design for the Internet of Things systems," in Z. Mahmood (Ed.), *Connected Environments for the Internet of Things*, pp. 201–220. Springer, Cham, 2017. doi: https://doi.org/10.1007/978-3-319-70102-8_13
27. V. Bode, C. Trinitis, M. Schulz, D. Buettner, T. Preclik, "DDS implementations as real-time middleware – A systematic evaluation," in *2023 IEEE 29th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, pp. 39–46. doi: https://doi.org/10.1109/RTCSA58653.2023.00030

Received 14.08.2025

INFORMATION ON THE ARTICLE

Michael Z. Zgurovsky, ORCID: 0000-0001-5896-7466, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: zgurovsm@hotmail.com

Pavlo O. Kasyanov, ORCID: 0000-0002-6662-0160, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: p.o.kasyanov@gmail.com

Nataliya D. Pankratova, ORCID: 0000-0002-6372-5813, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: natalidmp@gmail.com

Yuriy P. Zaychenko, ORCID: 0000-0001-9662-3269, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: zaychenkoyuri@ukr.net

Illia O. Savchenko, ORCID: 0000-0002-0921-5425, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: i.savchenko@kpi.ua

Tetyana V. Shovkoplyas, ORCID: 0009-0004-8991-0285, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: from_tatyana@ukr.net

Liliia S. Paliichuk, ORCID: 0000-0003-1654-4371, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: lili262808@gmail.com

Andrii M. Tytarenko, ORCID: 0000-0002-8265-642X, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: titarenkoan@gmail.com

КОГНІТИВНА AI-ПЛАТФОРМА ДЛЯ АВТОНОМНОЇ НАВІГАЦІЇ РОЗПОДІЛЕНИХ БАГАТОАГЕНТНИХ СИСТЕМ / М.З. Згуровський, П.О. Касьянов, Н.Д. Панкратова, Ю.П. Зайченко, І.О. Савченко, Т.В. Шовкопляс, Л.С. Палійчук, А.М. Титаренко

Анотація. Подано концепцію когнітивної AI-платформи для автономної навігації розподілених багатоагентних систем на прикладі рою безпілотних літальних апаратів. Запропоновано архітектуру, яка поєднує наземний центр із когнітивними сервісами та багаторівневу бортову підсистему, що забезпечують безперервний цикл навчання, адаптації, виконання та оновлення поведінкових моделей. Сформульовано базові сценарії місій, зокрема розвідка, пошук і рятування, ураження цілей, дезінформація, які демонструють можливості рою до автономної, децентралізованої взаємодії навіть у ворожому середовищі. Представлено приклад плану реалізації місії пошуку і рятування із використанням когнітивної платформи, що включає адаптивне планування, SLAM-навігацію, ройову координацію та глибоке розпізнавання об'єктів. Результати частково підтримано Національним фондом досліджень України, грант № 2025.06/0022 «AI-платформа з когнітивними сервісами для координованої автономної навігації розподілених систем, що складаються з великої кількості об'єктів».

Ключові слова: штучний інтелект, рій дронів, автономна навігація, когнітивна платформа, мультиагентні системи, поведінкові дерева, цифровий двійник, SLAM.

UDC 004.8:681.518:629.7.052
DOI: 10.20535/SRIT.2308-8893.2025.3.02

DIGITAL TWINS IN AI-CONTROLLED NAVIGATION TASKS FOR AUTONOMOUS UAV SWARM

M.Z. ZGUROVSKY, N.D. PANKRATOVA, I.M. GOLINKO, K.D. GRISHYN

Abstract. The article presents the concept and architecture of digital twins (DT) in the tasks of autonomous swarm navigation for unmanned aerial vehicles (UAVs) controlled by artificial intelligence. Study demonstrated that the effective operation of a drone swarm under conditions of disrupted or absent communication with the ground center is enabled by the functional distribution of DT components between the ground center and onboard levels of AI agents. Mathematical models of ground center's DT provide strategic modeling, training, mission simulation, and post-mission analysis, while onboard AI agents focus on local adaptation, diagnostics, environmental reconstruction, and cognitive behavior control. Special attention is paid to the interface module of the DT, which provides asynchronous interaction with the ground infrastructure. A functional division on the swarm-level, environment, mission, telemetry, and agent-level DTs is proposed. The effectiveness of the "Learn–Simulate–Deploy–Adapt" cycle for continuous improvement of swarm systems in the context of electronic warfare (EW) and dynamic operational environments was justified. The results were partially supported by the National Research Foundation of Ukraine, grant No. 2025.06/0022 "AI platform with cognitive services for coordinated autonomous navigation of distributed systems consisting of a large number of objects".

Keywords: digital twin, swarm intelligence, autonomous navigation, unmanned aerial vehicles (UAVs), cognitive artificial intelligence platform, decentralized control, simulation modeling, simultaneous localization and mapping (SLAM), behavior trees (BT), electronic warfare.

INTRODUCTION

In today's rapidly evolving world, the application of DT has gained significant momentum, becoming a key success factor across various industries. Virtual replicas of physical objects, systems, or processes open up opportunities for real-time analysis, modeling, and optimization. The DT enables companies to reduce costs, predict malfunctions, improve the management of production processes, and develop new products with minimal risks. This technology becomes particularly crucial as industry, healthcare, transportation, and urban planning undergo digital transformation [1].

For instance, in [2], the author conducted a study on the implementation of DT in manufacturing using reinforcement learning models. Compared to traditional management methods, production efficiency was improved by 18%, energy consumption was reduced by 12%, and system downtime was decreased by 15%.

DT toolkit enables significant advancements in intelligent management across various fields of activity. Although the concept of DT existed for over two decades, scientific discussions regarding its precise definition are still ongoing. A comprehensive definition of a DT is presented in [3], which incorporates

Grieves' definition [4] and distinguishes DTs from digital models and digital shadows based on the presence of information flows between the physical system and its digital counterpart. If there is no automated data flow between the physical system and its digital representation, such object is considered a digital model, an example could be a CAD model of a technical system (e.g., an aircraft). A digital shadow refers to a digital object (model) that receives data from the physical one. Its primary function is to automatically track certain changes in the physical system in order to represent its properties. If data flows from the physical system to the digital object and vice versa, then digital object is considered a DT, as changes in the digital representation affect the physical system.

These examples of interaction between a physical system and a digital copy enable to distinguish at a qualitative level the categories of concepts of a digital model, a digital shadow and a DT, but they do not specify details regarding the DT components. One of the foundational studies on the standardization of DT is the Industrial Internet Reference Architecture (IIRA), proposed by the Industrial Internet Consortium (IIC) [5]. This document provides guidelines for the development of systems, solutions, and applications that incorporate DTs in industrial and infrastructure domains. This architecture contains general definitions for interested parties, the order of system decomposition, design patterns and a list of terms. The IIRA model defines at least four types of interested parties: business; use; operation; implementation. Each area focuses on the implementation of the corresponding functional model of the DT, structure, interfaces, internal component interactions, as well as on the system of DT models interaction with physical object's external elements. According to the IIRA model, information about the DT includes (but is not limited to) a combination of the following categories: physical model and data; analytical model and data; archives of time variables; transaction data; master data; visual models and calculations. Thus, the concept of DT has a multifaceted architecture and, therefore, complex mathematical support for implementation.

A promising area for the application of DT toolkit is UAV control. An additional challenge in this domain is the cognitive coordination of UAV swarms during flight. The autonomous navigation of UAV swarms is based on the integration of two key system components:

- a ground center with module DT designed for UAV training, validation, and control;
- an onboard AI-platform for UAV with cognitive services.

The ground center functions as a strategic control center, where training, testing, and validation of neural networks, used on board of the drones, are carried out [6]–[8]. This center hosts an infrastructure for simulating combat missions in a virtual environment using onboard platform of UAVs [3], [5]. This approach ensures a high degree of realism, allowing to test the system's behavior under load, estimate mission losses and adapt swarm architecture to changing conditions.

The onboard segment of the UAVs is responsible for executing the mission (scenario) defined by the ground center, either with continuous data exchange with the ground center or in full autonomy mode. Also, it enables the swarm to independently navigate, make real-time decisions, avoid obstacles, stabilize flight, and coordinate swarm members without the need for constant communication with the ground center.

The interaction between the onboard and ground centers constitutes a continuous cycle of adaptation, learning, and improvement. During the pre-mission phase, the ground center trains the models, simulates the mission execution, compiles UAV operational algorithms, and uploads updated algorithms to the onboard systems. During the mission, the drones operate autonomously but send telemetry data to the center whenever communication is available. The center monitors the mission and, if necessary, sends corrective commands. After mission, the collected data are analyzed, checked for anomalies, models are refined, and a new training cycle is being initiated.

The evolution of the cognitive component of the UAV swarm AI system is realized through iterative model training using feedback obtained after mission. The models are based on a combination of reinforcement learning, local decision-making via BT, and neural network-based anomaly detection. This architecture enables the system to self-learn and improve strategies without compromising autonomy. A distinctive feature of the AI-system is the integration of memory and logging mechanisms that accumulate data from mission to mission, forming the foundation for the swarm's cognitive adaptation. Thus, the system acquires the capability for cognitive evolution — learning from its own experience to enhance efficiency and resilience in the dynamic challenges of the modern battlefield.

ARCHITECTURE AND INTERACTION LOGIC OF DT FOR SWARM-ORIENTED AUTONOMOUS UAV NAVIGATION

The DT is deployed at the ground center, while a limited interface module is implemented onboard the UAV, which provides (Fig. 1):

- data and telemetry buffering;
- local scenarios adaptation in case of communication loss.

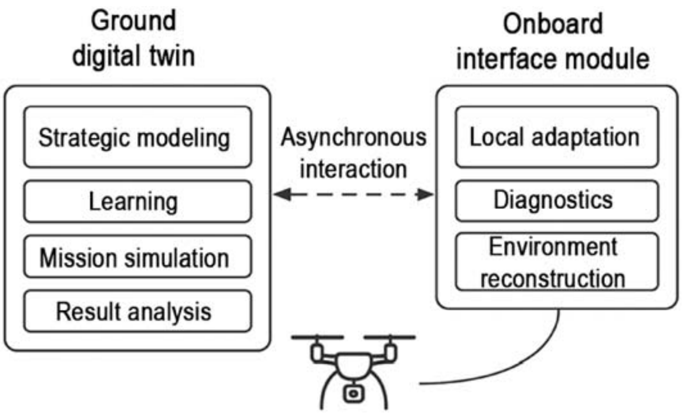


Fig. 1. Systemic interaction between the ground DT and the UAV's onboard subsystem

Fig. 1 schematically illustrates the fundamental architecture of interaction between the ground-based DT and the onboard interface module of the drone, which are the key components of the autonomous cognitive artificial intelligence platform for swarm control. The ground DT performs strategic-level functions — modeling swarm behavior, training neural networks, simulating combat missions in a virtual environment, and conducting in-depth post-mission analysis. This environment acts as a virtual test bed where adaptive strategies are developed and verified before their implementation on real platforms.

On the other hand, the onboard interface module, integrated into each drone, ensures autonomous system operation under conditions of partial or complete communication loss with the ground center. It implements local adaptation to environmental changes, performs internal diagnostics of the system's technical state, and reconstructs the surroundings using onboard sensors and odometry [9], or simultaneous localization and mapping (SLAM) algorithms [10], [14]. This approach enables each agent to make real-time decisions independently, ensuring decentralized, flexible, and fail-safe swarm behavior.

Asynchronous interaction between the ground-based DT and the onboard module emphasizes the key concept of drone independence during flight. Communication between the two levels is not continuous and may occur only at specific moments, when external environmental conditions allow it. This design enables the maintenance of autonomous navigation even in hostile environments, including when electronic warfare (EW) systems are active. At the same time, data accumulated during the mission is buffered and transmitted to the ground center when communication becomes available for analysis and further model retraining, starting the next cycle of cognitive improvement. Therefore, the structure illustrated in Fig. 1 represents a dynamic and distributed system, where the ground and onboard components operate in synergy to ensure adaptability, resilience, and ability to evolve for swarm systems.

Within the structure of an autonomous AI-platform for UAV swarms, the DT module performs an asynchronous yet strategically significant function. Its primary role is not to ensure continuous communication during missions, but rather to prepare, analyze, and update behavioral strategies during periods of time, when no combat missions are carried out. This approach aligns with the requirements of autonomous navigation in combat scenarios and under EW system activity, when communication with the ground center may be unavailable or undesirable.

Before the start of a mission, DT in the ground center allows for the testing of scenarios, adaptive strategies, and behavioral models [11], which are subsequently uploaded to each drone's onboard system. During flight, the drones operate fully autonomously, relying solely on local sensors, the cognitive core, and adaptive algorithms. However, if communication is available, they exchange data with the ground center. All data about behavior, telemetry, and decisions made are recorded in internal buffers for further analysis. Table 1 presents the formalization of the components of the UAV interface module.

Table 1. Formalization of the UAV interface module'functions

Function name	Description	Call time	Data exchange format
Telemetry buffering	Collection and storage of data for subsequent transmission	After each control cycle	JSON / ROS message
SLAM or odometry module	Construction of a local environmental map or spatial orientation using camera image analysis	Real-time	Local database
Fail-safe monitor	Analysis of internal system parameters	Every minute or upon event	Log file / Signal system
Behavior controller	Adaptive switching between branches of the BT	Upon event / As planned	Internal FSM (finite state machine) state
Swarm state synchronizer	Exchange of critical information with neighboring drones	Optionally, peer-to-peer	DDS / RTPS

Information is transmitted to the ground center's DT after mission completion or at designated evacuation checkpoints. This enables in-depth analysis, model retraining, and updating the knowledge that are utilized in subsequent missions. In this way, DT ensures swarm evolution without interfering with the autonomy of task execution. The UAV interface module is responsible for data buffering, access to the latest strategies, partial scenario simulation in fallback modes, and asynchronous updates whenever the situation allows it. Its presence in the system architecture is essential, as it provides autonomous interaction with the ground center's DT and local support without violating the decentralized principle of swarm control.

DT APPLICATION DIRECTIONS IN TASKS OF AUTONOMOUS UAV SWARM NAVIGATION

Considering the specific requirements of developing a cognitive AI-platform for decentralized control of UAV swarms under EW activity conditions, DTs application seems justifiable at multiple stages of the system's life cycle. Within our project, the most relevant mathematical models for implementing DTs for the ground center are as follows.

1. UAV swarm model (system level). The objective of this model is to simulate and verify swarm behavior of drones within a virtual environment, taking into account dynamics, communication losses, external disturbances, and changes in swarm lineup. This model enables:

- test decentralized control strategies (including BT) prior to its deployment;
- analyze the stability of UAV swarm interaction under various agent loss scenarios;
- debug DDS/RTPS-based communication [12] between agents;
- train reconfiguration algorithms without risk to physical drones.

2. Individual drone model (agent level). For each type of UAV, a corresponding model is created that includes aerodynamics, navigation sensors, decision-making modules, and an interface with the autopilot. Its use allows to:

- precisely test software–hardware interaction;
- simulate sensor degradation, Global Navigation Satellite System (GNSS) disruptions, and the impact of EW effects;
- predict potential failures and transitions to fail-safe modes;
- adapt controller (e.g., PID or MPC) parameters to mission-specific conditions.

3. Environment model. The creation of a virtual 3D environment, which incorporates models of obstacles, threats, magnetic anomalies, and signal loss zones enables to:

- generate scenarios for training and testing swarm adaptation capabilities [13];
- verify functionality of local planners (e.g., SLAM, obstacle avoidance system [14]);
- develop maps for pre-flight mission simulation and risk analysis.

4. Model of mission carry out. This involves the computer-aided design and simulation of specific scenarios (e.g., patrol, evacuation, object detection), which allows to:

- optimize the initial mission BT (BT definition) according to the context;
- identify critical nodes and failure points, prepare fallback behavior branches;
- automatically evaluate the effectiveness based on key performance indicators (KPIs).

5. Telemetry model. Simulation of real-time data exchange with the ground center enables to:

- verify telemetry quality;
- configure WebUI and ROSBridge protocol;
- detect potential delays, data losses, or transmission errors.

Table 2 presents a comparison of simulation environments for the ground center's DT.

Table 2. Comparison of simulation environments for ground center DT

№	Environment	Advantages	Disadvantages
1	Gazebo + ROS 2	ROS 2 support, realistic physics, open-source	Higher configuration complexity
2	Ignition Gazebo	Enhanced graphics, DT support	Relatively new, limited plugin ecosystem
3	AirSim (Microsoft)	Realistic aerodynamics, integration with Unreal Engine/Unity	High system requirements
4	Unity + BT.CPP	Full flexibility, BT visualization support	Requires custom infrastructure

Thus, the deployment of DT is recommended in environments Gazebo + ROS 2 [15], Ignition Gazebo, AirSim, or Unity/Unreal Engine based emulators integrated with BehaviorTree.CPP (see Table 2). It is especially appropriate to implement the "Learn–Simulate–Deploy–Adapt" cycle, which integrates simulation-based learning with the gradual transfer of behavior logic to the real swarm. In this way, DT models become a key component not only in the R&D phase, but also in training, testing, certification, and operational support of the system during mission.

TASK ALLOCATION BETWEEN GROUND CENTER DT AND UAV ONBOARD SYSTEMS

The overall logic of task distribution between the ground infrastructure and the onboard UAV systems is as follows:

- the ground center is responsible for simulation, training, strategic planning, and post-mission analysis;
- the UAV onboard system provides a secure wireless interface for communication between UAVs, also is it responsible for autonomous diagnostics, real-time adaptation and navigation of each drone independently of the ground center.

This division ensures optimal utilization of computational resources, flexibility and resilience of the system under conditions of limited connectivity and dynamic operational environment.

Tasks of DT executed at the ground center

In modern multi-layered architectures of autonomous swarm systems, DT which is implemented at the ground center, plays a pivotal role in providing effective modeling, testing, mission planning, and adaptation of UAV swarms to complex and dynamic environments. Its models offload a significant portion of computational load from onboard UAV platforms to the ground infrastructure, while preserving strategic coordination, behavioral predictability, and operational flexibility of the swarm. At the system level, the UAV swarm model enables simulation of global swarm behaviors in various scenarios, testing decentralized control algorithms and robustness of DDS/RTPS protocols, which is critically important in environments with intermittent communication or in case of individual agents' failure. Not only does this allow to identify system's potential vulnerabilities, but also enhance the swarm's resilience to catastrophic events.

The environment model allows to construct complex terrain representations with natural and artificial obstacles, as well as electromagnetic anomalies, which is a critical factor for planning operations in areas with active EW interference. Model generates scenarios that ensure high realism in training AI agents and effective pre-deployment testing of autonomous navigation algorithms. The mission area visualization provided by the DT serves as a foundation for tactical decision-making by operators or command centers. At the mission level, the DT supports simulation of strategic transitions between scenarios, BT design, and the definition of KPIs, enables mission evaluation not only in terms of task completion, but also in terms of the achievement of qualitative objectives.

The telemetry model focuses on simulation and verification of communication interfaces, delays, and telemetry data losses, as well as post-mission analysis of swarm and individual drones' behavior. This is particularly important for optimizing information exchange between agents and the control center, as well as for developing a knowledge base for future missions. Equally important is the individual drone model, operating within the simulator, as it enables detailed configuration of behavior logic at the level of a single AI agent, testing responses to environmental changes, training UAV operators, and improving the onboard AI agents installed on drones. The summary of DT mathematical models' basic functions, implemented at the ground center, are given below.

1. UAV swarm model (system level):

- modeling global swarm behavior under various scenarios;
- testing decentralized control algorithms;
- testing DDS/RTPS communication protocols;
- analyzing the impact of communication losses and agent failures on swarm integrity;
- simulating faults and catastrophic events.

2. Environment model:

- creating terrain, obstacle, and magnetic anomaly models;
- generating mission scenarios under complex conditions (including EW);
- preparing training data for preliminary situations modeling;
- visualizing the mission area for tactical planning.

3. Model of mission carry out:

- designing and testing the mission tree (Behavior Tree (BT) Definition);

- defining strategic transitions and fallback scenarios;
- defining mission success criteria (in the form of KPI);
- analyzing probable trajectories and synthetic tasks.

4. Telemetry model:

- testing WebUI / ROSBridge interfaces;
- simulating delays and data loss during transmission;
- Analyzing swarm and individual drone behavior logs (offline mode).

5. Individual drone model. (agent level):

- configuring behavior logic at the level of a single agent;
- training AI agents or real operators in simulated environments.

Therefore, the deployment of DT's mathematical models within the ground center serves not merely as a simulation tool but as a foundational component of adaptive, safe, and strategically coordinated operation of swarm systems. Functioning as a virtual proving ground [16], these models facilitate iterative testing and refinement of algorithms, significantly reducing operational risks and resource expenditures in real world.

Tasks of interface module executed onboard UAV (in autonomous navigation mode)

Within the architecture of an autonomous swarm system, each UAV is equipped with interface module that plays a critical role in ensuring local adaptation, flight control, and interaction between agents under conditions of partial or complete loss of communication with the ground center. These modules are designed to maintain the UAV's operability as an autonomous, cognitively capable agent within a localized segment of the overall system. The central element of the interface module is the local AI agent, responsible for monitoring the UAV's internal technical parameters: power supply voltage, temperature conditions, and sensor integrity. Based on this data, the system implements hardware degradation forecasting, generates alerts for transition into a protected (fail-safe) mode, and defines threshold conditions for potential mission withdrawal. This approach allows each UAV not only to detect critical deviations but also to autonomously assess its operational readiness for further task execution.

Integral to this functionality is the local environmental reconstruction. Embedded odometry or SLAM-algorithms allow each UAV to generate an up-to-date local map, identify obstacles, hazardous areas, landscape alterations, and predict potential collisions. This spatial representation serves as the basis for real-time reactive route planning, which is essential for survival and successful task execution in dynamic and often hostile environments. Importantly, such an AI agent enables the UAV not merely to respond to the current operational context but also to anticipate its change, making its behavior closely to a smart device rather than a conventionally algorithm-driven system.

Another critical component of autonomy is the module of mission tree evaluation and control. This subsystem manages local execution of behavioral branches, monitors task completion success, and can adaptively switch between operational modes in response to environmental changes or variations in the UAV's internal parameters. This eliminates the limitations of rigid, preprogrammed scenarios and facilitates decision-making under uncertainty. Concurrently, each agent maintains an individual log of critical events and, whenever possible, transmits it to other swarm members, establishing the foundation for the system's collective memory.

The system architecture also incorporates a synchronization agent — a compact communication and analytical add-on responsible for maintaining a locally consistent representation of the swarm's operational state, data synchronization among neighboring agents and, in cases of data loss or corruption, initiates localized reconfiguration of behavioral strategies. Synchronization agent provides swarm's decentralized response to the loss of one or more UAVs or to data distortion within specific system segments. Such a design enhances the swarm's resilience, self-recovery capacity, and mission accomplishment potential, even under unforeseen disruptive influences. The primary onboard functions of the interface module, operating in autonomous navigation mode, are summarized as follows.

1. Local AI agent:

- continuous monitoring of the UAV's internal state parameters (power supply voltage, temperature conditions, and sensor integrity);
- prognosis of hardware component degradation and initiation of fail-safe operational modes;
- determination of threshold conditions that necessitate mission cancelation;
- self-assessment of operational eligibility for continuing mission execution.

2. Local environment reconstruction (SLAM monitoring):

- local spatial map construction (SLAM, obstacles, hazardous zones);
- prediction of potential collisions and implementation of reactive path planning;
- detection of environmental changes (such as emergence of new obstacles, threats, etc.);

3. Behavior evaluation and transitions between mission tree branches (behavior monitoring):

- mission tree execution and task monitoring;
- adaptive switching between behavioral modes;
- logging of critical events and, when possible, transmission of this information to the swarm.

4. Embedded synchronization agent:

- maintenance of a coherent local representation of the swarm's operational state;
- exchange of situational data with neighboring UAVs;
- localized reconfiguration of behavioral strategies in response to UAV loss or system faults.

Thus, the AI agents embedded within the UAVs not only enhance the functional capabilities of individual drones but also establish the foundations for their subjectivity, self-reflection, adaptive interaction, and coordinated behavior within the swarm collective. This transforms each UAV from a mere of rigidly predefined algorithms executor to an active participant in a complex, flexible, and evolving behavioral system, and is necessary condition for the transition from strictly programmed to self-learning swarm architectures.

AI agents and DTs within the architecture of autonomous swarm systems have both civilian and military applications, which significantly enhances the flexibility and universality of their deployment [17], [18]. In the military domain, such systems allow to organize autonomous combat patrols, convoy escort operations, and the evacuation of wounded personnel from active combat zones, minimizing risks to human operators. In the civilian sector, their functional capabilities can be repurposed for search-and-rescue missions under challenging conditions (e.g., post-natural disaster scenarios), wildfire monitoring, and the inspection of critical infrastructure such as bridges, gas pipelines, and power transmission lines. Such dual-use ensures the maximization of technological potential in both peacetime and wartime.

A particular focus is ensuring the cyber resilience of AI agents and DTs, as they operate within environments with potentially high risks of external interference [19]. To address these challenges, the system architecture incorporates robust protective mechanisms, including end-to-end communication channels encryption according to DDS/RTPS protocols, guaranteeing the confidentiality of transmitted data. To protect against data tampering, spoofing, or cyberattacks, data authenticity verification is implemented using digital signatures. Additionally, threat detection algorithms based on AI are employed that work through real-time identification of anomalies and atypical behaviors. In case of communication loss or corruption, fallback modes are activated, enabling the system to maintain functionality and complete its mission despite partial isolation of individual elements. Collectively, these measures establish a reliable foundation for deploying DT in complex informational, technological, and combat environments.

EXAMPLE OF PRACTICAL APPLICATION OF A DT IN UAV SWARM NAVIGATION

As a part of demonstration scenario of critical infrastructure patrol under communication jamming (EW activity), a computer simulation of the DT for mathematical models of swarm, the environment, and individual drones was conducted. During the mission preparation phase, the ground-based DT modelled a 3D map of the operational area, which included magnetic anomalies, physical obstacles, and signal loss zones. This model was used to generate the route traverse scenarios for swarm groups, that take into account the limited availability of GNSS signals.

After uploading the BT and waypoints into the onboard systems, the UAVs were executing the mission autonomously. During the experiment, Inertial Measurement Unit (IMU) sensor failure for one of the agents was purposely simulated. The UAV's AI agent detected the corresponding anomaly, initiated a fallback stabilization mode, and excluded the affected agent from coordinated interaction, notifying the other UAVs via the synchronization agent.

Quantitative mission parameters: during the simulation scenario, a virtual swarm consisting of 3 UAVs was patrolling 100 × 100 m area, as illustrated in Fig. 2. Within the operational zone, the following conditions were modeled:

- 3 obstacles (representing buildings or infrastructure objects);
- one EW zone with a diameter of 40 m, centered at coordinates (60, 60) point;
- 5 route waypoints shared by all UAVs;

- mission duration: 9 minutes;
- average UAV velocity: 4.2 m/s;
- maximum positioning error within the EW zone: up to 3.6 m;
- telemetry transmission delay (simulated): up to 2.5 s;
- IMU sensor failure detection time (drone 1): 0.8 s;
- transition time to fail-safe mode: 1.1 s from the moment of anomaly detection;
- communication packet loss rate (in EW zone): up to 18%.

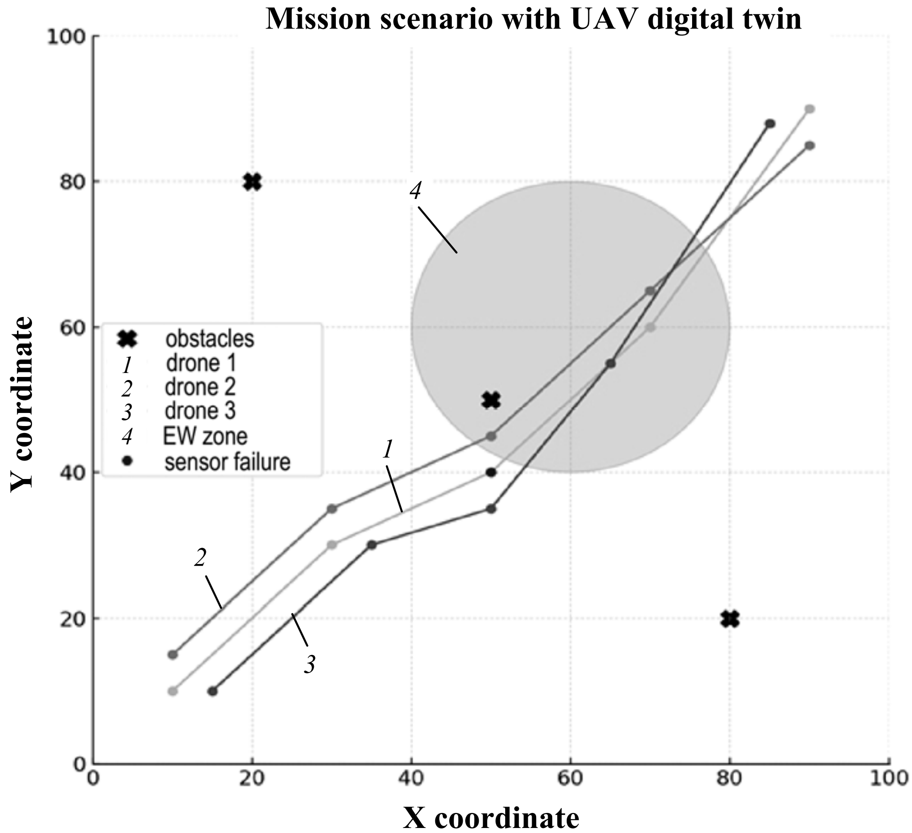


Fig. 2. Scenarios of patrolling missions by UAV swarm with DT

Therefore, the UAV AI agent identified a degradation via sensor temperature prediction, isolated drone 3 from the interaction zone, and broadcasted a status update to the remaining drones. Subsequently, the DT in the ground center updated the failure prediction model using the recorded logs. Upon mission completion, all telemetry buffers' content was transmitted to the ground center, where the DT performed trajectories visualization, calculated KPI, evaluated decision-making effectiveness, and retrained the fault detection model. Obtained data were utilized to update strategic adaptation modules for further missions.

Thus, this case exemplifies how DT can not only be a training and preparation tool, but also actively participate in the autonomous control process, enhancing swarm safety and adaptability in real time. The technical capabilities of the DT were demonstrated, as well as DT role in ensuring fault tolerance, post-mission learning, and deployment of self-adaptive swarm architectures, which is critical for the development of next-generation dual-use AI systems.

CONCLUSION

1. The ground center's DT is a critical element of the cognitive AI-platform for an autonomous UAV swarm, providing a closed cycle of adaptive training, simulation, deployment and improvement of the swarm's behavior in the conditions of a real combat environment and EW systems operation. Its functionality allows to effectively combine strategic planning and local autonomy. The ground center's DT performs modeling, training and verification, while the onboard system implements adaptation, self-control and reconfiguration in real time. The AI agent interface module on the UAV board provides asynchronous mission support, autonomous data buffering, partial environment reconstruction and determines behavior strategies without dependence on stable communication.

2. The classification of mathematical models for implementing the ground center's DT was proposed — models of the swarm, individual agents, environment, mission, and telemetry — which provides comprehensive simulation coverage of all aspects of swarm navigation. This contributes to the reliability, fault tolerance, and adaptability of the system. The ground center's DT functions as a virtual proving ground, where autonomous navigation algorithms, BT, and swarm coordination mechanisms are tested, verified, and refined, and SLAM algorithm parameters are configured. AI agents on board of each drone provide UAV autonomy, enabling it to independently assess its state, predict malfunctions, adapt behavior, and interact with other agents even under critical conditions.

3. The demonstration scenario of patrolling under EW systems activity confirmed the effectiveness of the DT in failure detection, anomaly adaptation, and swarm coordination restoration, thereby proving its role in providing self-learning and fault-tolerant architectures. The implementation of the DT-based "Learn–Simulate–Deploy–Adapt" cycle is strategically important for transforming autonomous swarms into evolving, intelligent dual-use systems.

4. During mission execution, the ground control and training station (depending on the presence and intensity of EW interference) can operate in several modes: as a DT (when connection with the UAV swarm is fully available); as a digital shadow (in the case of limited connection with swarm elements); or in combat task simulation mode (if connection with the UAVs is completely unavailable). To enhance mission performance, UAV AI agents may be equipped with a data relay function to support the communication with the ground center. In this mode, one drones are assigned to relay communication, and the other focus on mission execution, thereby increasing the overall efficiency of the system.

REFERENCES

1. F. Tao, H. Zhang, A. Liu, A.Y. Nee, "Digital twin in industry: State-of-the-art," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2018. doi: 10.1109/TII.2018.2873186
2. Lang Dai, "Intelligent manufacturing digital twin creation based on BIM and reinforcement learning," *Proc. SPIE 13691, Third International Conference on Advanced Materials and Equipment Manufacturing (AMEM 2024)*, 1369134, 8 July 2025. doi: https://doi.org/10.1117/12.3070521
3. W. Kritzinger, M. Karner, G. Traar, J. Henjes, W. Sihn, "Digital Twin in manufacturing: A categorical literature review and classification," *Ifac-PapersOnline*, vol. 51, no. 11, pp. 1016–1022, 2018. doi: 10.1016/j.ifacol.2018.08.474

4. M.W. Grieves, "Product lifecycle management: the new paradigm for enterprises," *International Journal of Product Development*, vol. 2, no. 12, pp. 71–84, 2005. doi: 10.1504/IJPD.2005.006669
5. "The Industrial Internet Reference Architecture," *An Industry IoT Consortium Foundational Document*, 2022. Available: https://www.iiconsortium.org/wp-content/uploads/sites/2/2022/11/IIRA-v1.10.pdf
6. E. Kayacan, M.A. Khanesar, O. Kaynak, *Fuzzy Neural Networks for Real Time Control Applications*. Butterworth-Heinemann, 2015.
7. S.K. Zhou, D. Rueckert, G. Fichtinger, *Machine Learning for Medical Image Analysis: Advances in Models and Methods*. Academic Press, 2019.
8. Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015. doi: 10.1038/nature14539
9. S.M. Vovk, V.V. Hnatushenko, M.V. Bondarenko, *Image processing methods and computer vision*. Dnipropetrovsk: Lira, 2016, 148 p.
10. I. Nevlyudov, S. Novoselov, K. Sukhachov, "Method of simultaneous localization and mapping for construction of 2.5D maps of the environment using ROS," *Innovative Technologies and Scientific Solutions for Industries*, no. 2 (24), pp. 145–160, 2023. doi: https://doi.org/10.30837/ITSSI.2023.24.145
11. Y. Bengio, A. Courville, P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013. doi: 10.1109/TPAMI.2013.50
12. Eclipse Foundation, "DDS for eXtremely Resource-constrained Environments (XRCE-DDS)," *OMG Specification*, 2021. [Online]. Available: https://www.omg.org/spec/DDS-XRCE/
13. Z. Hou, I. Fantoni, "Distributed Leader-Follower Formation Control for Multiple Quadrotors with Weighted Topology," *10th System of Systems Engineering Conference (SoSE), 2015*. doi: https://doi.org/10.1109/SYSOSE.2015.7151924
14. M. Norbelt, X. Luo, J. Sun, U. Claude, "UAV Localization in Urban Area Mobility Environment Based on Monocular VSLAM with Deep Learning," *Drones*, 9(3), 171, 2025. doi: https://doi.org/10.3390/drones9030171
15. M. Quigley et al., "ROS: an open-source Robot Operating System," *ICRA Workshop on Open Source Software,* vol. 3, no. 3.2, 2009.
16. E.H. Glaessgen, D. Stargel, "The digital twin paradigm for future NASA and US Air Force vehicles," in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, 2012.
17. J. Tang, G. Liu, Q. Pan, "A review on representative swarm intelligence algorithms for solving optimization problems: Applications and trends," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 10, pp. 1627–1643, 2021. doi: https://doi.org/10.1109/JAS.2021.1004129
18. T.I. Zohdi, "A machine-learning enabled digital-twin framework for tactical drone-swarm design," *Computer Methods in Applied Mechanics and Engineering*, vol. 442, 117999, 2025. doi: https://doi.org/10.1016/j.cma.2025.11799919
19. X. Wang et al., "A Survey on Security of UAV Swarm Networks: Attacks and Countermeasures," *ACM Computing Surveys*, vol. 57, no. 3, 2025. doi: https://doi.org/10.1145/3703625

Received 08.08.2025

INFORMATION ON THE ARTICLE

Michael Z. Zgurovsky, ORCID: 0000-0001-5896-7466, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: zgurovsm@hotmail.com

Nataliya D. Pankratova, ORCID: 0000-0002-6372-5813, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: natalidmp@gmail.com

Igor M. Golinko, ORCID: 0000-0002-7640-4760, Educational and Scientific Institute of Nuclear and Thermal Energy of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: golinko.igor@lll.kpi.ua

Kostiantyn D. Grishyn, ORCID: 0009-0006-5950-3739, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, student, e-mail: constantine1223h@gmail.com

ЦИФРОВІ ДВІЙНИКИ В ЗАДАЧАХ АВТОНОМНОЇ РОЙОВОЇ НАВІГАЦІЇ ДРОНІВ ПІД УПРАВЛІННЯМ ШТУЧНОГО ІНТЕЛЕКТУ / М.З. Згуровський, Н.Д. Панкратова, І.М. Голінко, К.Д. Грішин

Анотація. Розглянуто концепцію та запропоновано архітектуру цифрових двійників у задачах автономної ройової навігації безпілотних літальних апаратів (БПЛА), керованих штучним інтелектом. Показано, що ефективне функціонування рою дронів в умовах відсутності стабільного зв'язку з наземним центром можливе завдяки розподілу функцій цифрового двійника наземної станції і ШІ-агентами бортового рівня. Математичні моделі наземного ЦД забезпечують стратегічне моделювання, навчання, симуляцію місій і аналіз результатів, тоді як бортові ШІ-агенти зосереджені на локальній адаптації, діагностиці, реконструкції середовища й когнітивному управлінні поведінкою дронів. Особливу увагу приділено інтерфейсному модулю ШІ-агента БПЛА, що забезпечує асинхронну взаємодію з наземною інфраструктурою. Запропоновано функціональний поділ математичних моделей ЦД на моделі рою, середовища, місії, телеметрії та окремого ШІ-агента. Обгрунтовано доцільність використання циклу «Learn–Simulate–Deploy–Adapt» для безперервного вдосконалення ройових систем в умовах дії РЕБ і динамічного бойового середовища. Результати частково підтримано Національним фондом досліджень України, грант № 2025.06/0022 «Платформа штучного інтелекту з когнітивними сервісами для скоординованої автономної навігації розподілених систем, що складаються з великої кількості об'єктів».

Ключові слова: цифровий двійник, ройовий інтелект, автономна навігація, безпілотні літальні апарати, платформа штучного інтелекту, децентралізоване управління, імітаційне моделювання, метод локалізації та картографування (SLAM), поведінкові дерева, РЕБ.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ВИСОКОПРОДУКТИВНІ КОМП’ЮТЕРНІ СИСТЕМИ

UDC 004.41: 621.396:004.7:656.2
DOI: 10.20535/SRIT.2308-8893.2025.3.03

RESEARCH AND DEVELOPMENT OF METHODS TO IMPROVE THE QUALITY OF MOBILE COMMUNICATION AND MOBILE INTERNET IN HIGH-SPEED TRAINS

N.V. SHTEFAN, S.V. ZHIGLO

Abstract. This paper proposes effective methods and means to enhance the quality of mobile communication and mobile Internet in high-speed trains. The current issues related to achieving enhanced mobile communication and Internet quality in high-speed trains are discussed within this thematic scope. The practical research examines the metrological features of the proposed new combined methodologies for improving mobile communication and Internet quality in high-speed trains at a model-complex level. It has been established that the methodology combining methods (LTE + Wi-Fi + 5G) shows the best results due to the combination of low-latency and jitter technologies. Metrological measurements confirm its effectiveness through lower latency and jitter values compared to other methodologies. Methodology 3 (5G + Micro-grids) offers high local indicators but is limited in bandwidth. Metrological data confirm the reduced latency and jitter.

Keywords: comprehensive model, quality standards, integration testing, modular testing, technological challenges, micro-grids, 5G, digital communications.

RELEVANCE

Improving the quality of mobile communication and mobile internet in high-speed trains is a relevant issue in modern society, as mobile technologies have become an integral part of people’s daily lives. The increasing number of mobile device users and the growing demand for high-speed connections during travel make this topic particularly important. Mobile communication in high-speed trains often faces challenges such as unstable signals, high latency, jitter, and limited bandwidth, which reduce the quality of services for passengers. High-speed trains pose unique technical challenges related to their high speeds, changing network zones, and frequent handovers between base stations. These factors affect the continuity and stability of the connection.

Moreover, the relevance of this topic is reinforced by the necessity of providing passengers with reliable internet access for work, entertainment, and communication during trips, enhancing their comfort and satisfaction with the services of transport companies.

According to the work [1], improving the quality of mobile communication in mobile transport environments is critically important for the development of modern communication infrastructure, as it directly impacts user satisfaction and service efficiency. Studies by the authors [2] confirm that the combined use of technologies such as LTE, Wi-Fi, and 5G minimizes issues related to quality degradation, which is a key factor for the stable operation of mobile internet during travel. According to data presented in studies [3], the implementation of modern data transmission technologies in high-speed networks significantly reduces latency and improves connection quality, as confirmed by metrological measurements. Research into current methods for improving mobile communication and internet in high-speed trains, particularly the methods combining LTE, Wi-Fi, and 5G, is a necessary step to ensure high-quality connectivity under constantly changing network conditions. These studies enable technological advancements and contribute to the development of efficient communication systems in transport infrastructure, meeting modern market demands and passenger needs.

Thus, the relevance of this article is determined by the need to develop and implement innovative solutions to improve the quality of mobile communication and internet in high-speed trains. These efforts will address several technical and infrastructural issues, contributing to enhanced passenger service quality.

ANALYSIS OF RECENT PUBLICATIONS

Branković N., et al. (2021) [2] conducted an in-depth analysis of the development of mobile communication systems for high-speed railways. The researchers emphasize the importance of efficient communication in dynamic environments, where high train speeds necessitate significant improvements in data transmission technologies. The development of such systems is a critical factor in ensuring the quality of software solutions, as any delays or packet losses can affect communication reliability. Their findings highlight the need for new performance prediction models and resource optimization to enhance software reliability. Similar studies [1, 5, 6, 8, 10] focus on current data transmission technologies, such as 4G and 5G, but often overlook emerging technologies or alternatives that may soon enter the market. Furthermore, while the researchers propose innovative performance prediction models, their findings require further validation in real-world high-speed scenarios, where unpredictable factors could impact communication quality. The limited amount of experimental data in these studies also affects the accuracy and reliability of the results.

Dakulagi V. and Alagirisamy M. (2020) [3] explore adaptive beamforming systems for high-speed mobile communication. They propose an approach that reduces interference and improves signal quality in dynamic conditions. This method is particularly relevant for optimizing data transmission models, minimizing losses, and enhancing communication stability, all of which are crucial for control systems. However, the proposed approach is effective only in specific scenarios, and its efficiency in large networks with high user density requires further investigation.

Gunasekar A., et al. (2023) [4] introduce an innovative optical data transmission system for providing broadband internet access on high-speed trains. Their approach relies on a cooperative triple-hop system utilizing FSO-FSO-VLC technologies. This research underscores the importance of high-speed, stable connections, which contribute to software quality improvement by ensuring communication stability and enhanced performance. However, integrating new technologies such as FSO-FSO-VLC may require the development of new protocols and standards, which could delay implementation and present practical challenges.

Studies [5–15] examine data loading quality from mobile devices on high-speed trains. These studies focus on analyzing energy efficiency in mobile devices, an essential factor in ensuring software quality. Energy optimization extends system autonomy and reliability, which is particularly important in challenging operational conditions. However, these studies often neglect external factors such as environmental noise and interference, which can significantly influence measurement outcomes.

This analysis highlights the ongoing efforts to address the challenges of mobile communication and internet quality in high-speed trains, emphasizing the importance of balancing technological innovations with real-world constraints to develop effective solutions.

PROBLEM STATEMENT

The aim of the work is to study effective software systems, methods and means of improving the quality of mobile communication and the Internet in speed trains.

Achieving the goal is to solve the following tasks:

- conducting a generalized analysis of topical issues related to the research of modern methods of improving the quality of mobile communication and the Internet in speed trains;
- conducting key mobile and Internet quality parameters in high -speed trains;
- conducting an analysis of errors when measuring mobile and and Internet quality measurements in speed trains;
- investigation of the use of new combined methods to improve the quality of mobile communication and the Internet in speed trains.

MAIN PART

Table 1 shows the results of the analysis of modern methods of improving the quality of mobile communication and the Internet in speed trains.

Let's mathematically analyze the methods presented in Table 1 that can be applied to improve mobile communication and mobile internet quality in high-speed trains. According to the work [5], the optimization of mobile communication and mobile internet quality in high-speed trains, achieved through the use of a dynamic resource management system, can be mathematically described by the expression:

$$R_{\text{optimal}} = \frac{C_{\text{Available}}}{1+\lambda_{\text{load}}},$$

where R_{optimal} — optimal use of the resource; $C_{\text{Available}}$ — The channel is available ; λ_{load} — load ratio.

Table 1. The results of the analysis of modern methods of improving the quality of mobile communication and the Internet in speed trains

Method	Description	Countries of application	Advantages	Disadvantages	Metrological aspects
Dynamic resource management systems	Adapt a network resources in real time based on load	Germany, Japan	Reducing delays, increasing bandwidth	High complexity of settings	The need for accurate measurement of load and bandwidth of network
Expanded antenna system (DAS)	Using antennas to improve signal quality in trains	France, China	Improving the quality of the signal, reducing the zone of dead zones	High cost and complexity of realization	The need to calibrate antennas and measure the signal
Adaptive network calibration	Swimming Network Settings in Real Time	USA, Australia	Reducing systematic errors, improving communication quality	Difficulty in setting up	Requires accurate measurement of systematic errors and their correction
Network load forecasting	Using algorithms to predict load	South Korea, UK	Optimization of resources, reducing delays	The need for constant updating of algorithms and data	The need for accurate measurement of current loading and precision accuracy
Mobile Ratranslators	Using Ratranslators to improve the quality of the signal	Italy, Switzerland	Improving the quality of the signal, ensuring continuous coating	High cost of equipment, the ability to increase delays	The need to measure the efficiency of repeaters and adjust them
Coherent signal association	Reduction of noise and improving data rate	Japan, the Netherlands	Increasing data rate, reducing the impact of noise	High complexity of implementation, requires accurate adjustment	The need to measure the noise level and accuracy of the merging of signals
Reduction of jitter by buffering	Using buffering to reduce jitter and improve communication quality	Finland, Sweden	Reduction of the effect of variability of delays, improving video quality and audio flows	Possible increase in delays, requires effective management of buffers	The need to measure jitter and the efficiency of buffering
Internet roaming	Use roaming to ensure continuous coating through several operators	Germany, Switzerland	Ensuring a stable communication within several operators	Possible problems with network integration and roaming contract restrictions	Need to accurately measure the quality of communication between networks
Introduction of satellite technologies	Using satellites to provide communication in remote areas	Australia, Canada	Providing coverage in remote areas where there are no traditional networks	High delay, requires accurate satellite connections	The need to measure the delay of satellite connection and its quality
Multi - channel technology (MIMO)	Using multiple antennas to improve data transmission and signal quality	Singapore, South Korea	Increasing data rate, reducing interference	High cost of equipment, complexity of sale	Need to measure MIMO efficiency and its impact on data rate

Here are practical examples of applying optimization of mobile communication and mobile internet quality in high-speed trains:

- Germany: Deutsche Bahn uses dynamic resource management systems to optimize bandwidth and reduce delays in train networks [4].
- Japan: JR East implements adaptive technologies for network management in Shinkansen high-speed trains [3].

According to [8], the use of a Distributed Antenna System (DAS) involves deploying multiple antennas throughout the train to ensure uniform coverage and reduce signal loss. The formula for calculating signal coverage in high-speed trains using the DAS methodology can be computed using the formula:

$$S_{\text{coverage}} = \frac{P_{\text{antenna}}}{d^2},$$

where S_{coverage} — level of coverage; P_{antenna} — antenna power; d — distance to the observation point.

Practical examples of DAS application:

- France: SNCF implemented DAS on high-speed TGV trains to improve signal quality [10];
- China: Chinese Railways use DAS to ensure stable coverage on high-speed trains [11].

According to the work [12], adaptive network calibration involves automatic adjustment of network settings to account for changes in load and communication conditions. At the mathematical level, the formula for calculating the above-mentioned correction is given by equation:

$$X_{\text{adjusted}} = X_{\text{Measurement}} + \Delta_{\text{corrective}},$$

where X_{adjusted} — adjusted value; $\Delta_{\text{corrective}}$ — corrective coefficient.

Currently, this approach is actively used in the USA and Australia:

- USA: Amtrak implements adaptive calibration systems to improve signal quality in its high-speed trains [7];
- Australia: Australian Rail Track Corporation uses adaptive calibration systems to ensure stable communication [5].

Network load forecasting, according to the work [9], involves the use of machine learning algorithms for predicting network load and adaptive resource management:

$$L_{\text{projected load}} = \alpha L_{\text{previous}} + \beta \Delta T,$$

where $L_{\text{projected load}}$ — projected load; L_{previous} — the previous load value; ΔT — changes in time; α і β — adaptation coefficients.

In the course of the analysis, it was determined that the current methodology for network load forecasting is actively applied in South Korea and the United Kingdom:

- South Korea: Korail implements machine learning algorithms to forecast the load in KTX trains [11].
- United Kingdom: Network Rail utilizes forecasting technologies to optimize the network in high-speed trains [4].

The consideration of signal amplification through the use of mobile repeaters can be represented by the expression:

$$S_{\text{enhanced}} = S_{\text{output}} + G_{\text{amplification}},$$

where S_{enhanced} — enhanced signal; S_{output} — output signal; $G_{\text{amplification}}$ — amplification of the repeater.

Coherent signal association is a technology that combines signals from several sources to increase the total quality and speed of communication [3]. Mathematically taking into account the coherent association can be represented in the form of expression:

$$S_{\text{coherent}} = \frac{1}{N}\sum_{i=1}^{N} S_i,$$

where S_{coherent} — coherent signal; S_i — individual signals; N — number of signal sources.

The reduction of jitter by bufferization involves the use of buffers to reduce the impact of variability of delays in the network [1].

Mathematically, the effect of buffering within the reduction of jitter can be calculated by means of expression:

$$J_{\text{reduced}} = \frac{1}{n-1}\sum_{i=1}^{n}(T_i - T_{\text{average delay value}})^2,$$

where J_{reduced} — reduced jitter; T_i — individual delays; $T_{\text{average delay value}}$ — the average delay value; n — number of measurements.

According to [10], internet roaming provides continuous communication by switching between different networks without interruption.

Mathematically assessing the quality of internet roaming can be described using formula:

$$Q_{\text{roaming}} = \frac{T_{\text{switching}}}{T_{\text{connection}}},$$

where Q_{roaming} — the quality of roaming; $T_{\text{switching}}$ — time of switching networks; $T_{\text{connection}}$ — total connection time.

According to [12], the introduction of data transmission technologies through satellites involves the use of satellite joints to cover remote areas where traditional networks have problems with coating problems.

Mathematically evaluation of satellite compound can be made using a formula:

$$S_{\text{satellite signal}} = \frac{P_{\text{satellite signal}}}{1 + D_{\text{satellite signal}}},$$

where S_{sat} — satellite signal; P_{sat} — Satellite signal power; D_{sat} — delayed satellite connection.

In accordance with [8], the expanded use of multi-channel technology (MIMO) involves the use of multiple antennas to send and receive a signal that allows you to increase the data rate and improve communication quality:

$$R_{MIMO} = \log_2\left(1 + \frac{P_{\text{signal}}}{N_0 B}\right),$$

where R_{MIMO} — data transmission speed; P_{signal} — signal power; N_0 — spectral noise density; B — the width of the channel.

Analysis of the drawbacks and advantages of existing solutions:

1. Mobile communication technologies:

- 3G: While 3G provides good compatibility and wide coverage, its speed and latency do not meet modern requirements for high-speed trains.
- 4G LTE: Provides significant improvements in speed and latency compared to 3G, but it may have coverage issues in high-speed trains, especially in remote areas.
- 5G: Offers the best characteristics for high-speed mobile communication, but its deployment is expensive and complex, requiring new antennas and equipment.

2. Mobile communication enhancement technologies in moving objects:

- Mobile repeaters: Improve signal quality, but their cost and maintenance can be significant. Their installation may also require coordination with operators.
- Antenna repeater systems: Improve coverage but have high costs and installation complexity. They may also require specific standards for integration.
- Dynamic resource management: Adapts to changes in load and increases the efficiency of resource use, but it can be challenging to configure and may require new software solutions.

3. Metrological methods:

- Latency measurements: Allow quick and easy assessment of system response time, but may not account for all influencing factors.
- Data transmission speed measurements: Provide an accurate view of network bandwidth, but may be affected by other users.
- Signal quality assessment (RSRP, RSRQ): Enables evaluation of signal quality, but results may vary depending on motion and real-world conditions.

It is worth noting that the results of this analysis have significant practical value, as considering them helps identify weaknesses in existing solutions and develop improved approaches that can more effectively address communication quality issues in high-speed trains.

Table 2 presents the results of reviewing key quality parameters of mobile communication and internet in high-speed trains.

Based on Table 2, a comprehensive approach to measuring and evaluating the main parameters affecting the quality of mobile communication and internet in high-speed trains is revealed, with an emphasis on metrological aspects. It is important to note that metrology allows not only accurate measurements but also the analysis of errors that occur during measurement under dynamic conditions, such as the movement of the train. Metrological analysis can help assess the average value of this parameter and its variability under different movement conditions. Suggestions for improvement:

- Introduction of dynamic measurements using automated monitoring systems in real movement conditions, which will provide more accurate data.
- Optimization of data collection methods, considering the train's movement and potential changes in signal characteristics depending on the landscape and weather conditions.

- Use of artificial intelligence to analyze large data sets and predict potential signal loss, allowing for early adaptation of network parameters to moving conditions.

Table 2. Results reviewing key mobile and Internet quality parameters in high speed trains

Parameter	**Description**	**Measurement methods**	**Standards / standards**	**Factors of influence**	**Metrological criteria**
Latency	The time required to transfer the package from the source to the recipient	Ping tests, delay measurements using GPS	ITU-T Y.1541, ETSI EN 301 908	Speed, the quality of infrastructure	Measurement accuracy in high speed conditions
Transmission speed	Maximum Data boot speed	Speedtest, Measuring Complexes for mobile networks	3GPP TS 36.521, ITU-T Y.1564	Network load, number of users	Measurement error depending on the terrain
Signal quality (RSRP)	The force of the signal obtained from the base station	Metrological devices to estimate the level of signal	ETSI TS 136 133, 3GPP TS 38.133	Distance to the base station, obstacles on the route	Repeatability of measurements in different sections of the route
Package loss	Percentage of lost packages during data transmission	Wireshark, Ping tests	ITU-T G.1050, RFC 791	Network traffic jams, changing the terms of receiving signal	Measuring losses in real traffic motion conditions
Delayed variations (jitter)	Deviation of packet delay during their transmission over the network	Measurement of traffic monitoring tools	ITU-T Y.1540, ETSI EN 301 908	Network load, changes in speed	High measurement accuracy in random conditions
Signal instability	Measurement of frequency interruption, or transition between base stations	Signal Monitoring Tools (Cellmaper)	ETSI EN 302 307-1, ITU-R M.2135	The speed of movement of the train, the density of the coating	Measurement accuracy with route tracking
Connection time	The time required to establish a connection between the client and the network	Mobile device logs, simulation tools	ETSI TS 102 232, ITU-T Y.1564	Network load, number of users	Definition of average values and uncertainty

According to [5], metrological analysis of mobile communication parameters involves assessing measurement errors that occur under high-speed movement conditions. This aspect is crucial for accurately reproducing results and correctly configuring the network. Errors may be caused by a range of factors, including:

- Dynamic changes in signal intensity during movement.
- Delay fluctuations due to changes in route and infrastructure.
- Interference and signal overlap from different base stations.

To address this, it is essential to clearly define the types of errors that occur and assess their impact on measurement performance. Classification of measurement errors:

1. **Systematic errors**: Related to the specifics of measuring equipment and network conditions:

- Caused, for example, by data transmission delay under low signal strength conditions.
- Can be corrected through equipment calibration.

2. **Random errors**: Resulting from changes in transmission environment conditions.

- Arise due to train speed fluctuations or signal level variations.
- Their assessment requires statistical approaches.

3. **Instrumental errors**: Related to the technical characteristics of measuring devices.

For example, antenna sensitivity or signal processing delay on mobile devices.

4. **Methodological errors**: Occur due to imperfections in measurement methods.

For example, measurement delay when using non-adapted testing methods for high-speed movement.

Table 3 presents the results of the analysis of error evaluation during measurements of mobile communication and internet quality parameters in high-speed trains.

Table 3. The results of the analysis assessment of errors during measuring the quality parameters

Parameter	Type of error	Method of evaluation of the error	The magnitude of the error	Factors of influence
Delay (Latency)	Systematic and accidental	Statistical packet delay analysis	±10–50 ms	Train speed, network load
Transmission speed	Accidental	Comparison of average values with standards	±5–20 Mbps	Signal, network load
Signal quality (RSRP)	Systematic and instrumental	Calibration of devices, multiple measurements	±2–5 dBm	Changing location, obstacles
Package loss	Accidental	Analysis of losses through package trackers	±0.1–2%	Traffic jams on the network, the quality of the route
Delayed variations (jitter)	Systematic and accidental	Statistical analysis of delay variations	±5–20 ms	Load on the network, route of traffic
Signal instability	Systematic and methodological	Monitoring of frequencies of signal interrupts	±2–10 Cases per hour	Distance to base stations, train speed

As seen in Table 3, the main types of errors for each of the key parameters of mobile communication and mobile internet quality are outlined. It is important to note that these errors can be minimized or corrected using appropriate metrological methods.

1. **Latency:**

- The main sources of errors are variations in the speed of the train and network load. High-speed movement conditions lead to increased latency due to the increased distance to base stations.
- Error estimation method: statistical analysis of latency at different sections of the route to average the results and correct systematic errors.

2. **Data Transfer Speed:**

- The measurement of data transfer speed may vary depending on signal quality and network load.
- To assess errors, multiple tests are conducted under different conditions, followed by comparison of the results with normative values.

3. **Signal Quality (RSRP):**

- Errors may be caused by the instrumental features of measuring devices, especially over long distances between the train and base stations.
- Regular calibration of measuring equipment can reduce systematic errors related to signal strength.

4. **Packet Loss:**

- Random packet losses may occur due to network congestion or interference along the train's route.
- Error assessment is performed by analyzing packet trackers to identify loss frequency and determine average values.

5. **Jitter:**

- Random errors vary depending on network load and the train's route. When analyzing these errors, it is essential to account for fluctuations in train speed.
- Statistical analysis using a large number of samples allows for the evaluation of average values and jitter fluctuations.

6. **Signal Instability:**

- Systematic errors occur due to frequent switching between base stations, particularly on sections of the route with poor coverage.
- Error assessment is performed by monitoring signal interruptions and comparing the frequency of these interruptions with norms.

Table 4 presents methodological suggestions for reducing errors during the measurement of mobile communication and internet quality parameters in high-speed trains.

Table 4. Methodological proposals to reduce errors during measurements of measurements of mobile and Internet quality parameters in high -speed trains

Method	Description	Expected result	Formula
Noise filtration	Using digital filters to eliminate noise in the signal	Reduction of random errors	$\Delta_{\text{Reduction}} \to 0$
Calibration of measuring devices	Regular calibration of equipment to adjust systematic errors	Reduction of systematic errors	$\Delta_{\text{sys}} \to 0$
Parameters forecasting	Predicting conditions for dynamic adjustment of parameters	Increasing measurement accuracy and reducing errors	$\Delta_{\text{Reduction}} = f(\text{Conditions})$
Medium smoothing	Averaging the results of several measurements to reduce random errors	Reduction of random oscillations in measurement results	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
Using modern standards of communication	Switching to 5G standards to reduce delay and improve network bandwidth	Reduction of errors due to more stable and faster connection	Increasing the regulatory values of quality parameters (speed, delay, jitter)

As seen in Table 4, error assessment is a critically important step to ensure high measurement accuracy of mobile communication and internet quality in high-speed trains. During the train's movement, conditions often arise that lead to increased errors due to rapid changes in infrastructure and network conditions. Metrological analysis allows not only identifying these errors but also minimizing them through corrective measures.

MODELING AND TESTING

Suggested Methods:

1. Methods 1: LTE + 5G + Satellite: Combines LTE, 5G and Satellite Technologies to improve overall quality trains. LTE provides a good coating on the ground, 5G provides high data rate, and satellite communication provides coating in areas where other technologies are not available.

2. Method 2: LTE + Wi-Fi + 5G: integrates LTE, Wi-Fi and 5G to ensure improved communication quality. Wi-Fi is used to cover in areas where there is access to powerful access points, LTE provides the main coating and 5G is used to provide high data transmission rates in key areas.

3. Methods 3: 5G + micro-networks: combines 5G with micro-networks (small, local networks) that are installed in train cars to improve communication quality. Micro networks allow you to reduce delays and increase data rate by local traffic control.

4. Methods 4: 5G + micro-networks + satellite: combines 5G, micro-networks and satellite communication for the best results in high-speed train. It provides high data transmission, delays and constant bonds in speed.

5. Method 5: LTE + DAS + Wi-Fi: Uses Distributed Antenna Systems (DAS) together with LTE and Wi-Fi to improve communication quality. DAS provides a uniform signal distribution within the train, improving the total coating.

Each technique uses different approaches to assessing the quality of communication. Basic formulas: the bandwidth (*B*) is calculated according to the expression:

$$B = \frac{R}{T},$$

where R — the amount of data transmitted; T — transmission time.

Delay (*D*) is calculated according to expression:

$$D = T_{\text{total}} - T_{\text{processing}},$$

where T_{total} — total data transfer time; $T_{\text{processing}}$ — data processing time.

Jitter (*J*) is calculated according to expression:

$$J = \frac{1}{N-1}\sum_{i=1}^{N}\left|T_i - T_{avg}\right|,$$

where N — Number of measurements; T_i — time of individual measurements; T_{avg} — the average value of time.

Batch loss (*L*) is calculated according to expression:

$$L = \frac{N_{\text{lost}}}{N_{\text{total}}} \times 100\%,$$

where T_{lost} — the number of lost packages; T_{total} — the total number of packages.

The efficiency of buffering (*E*) is calculated according to the expression:

$$E = \frac{B_{\text{buffer}}}{B_{\text{total}}} \times 100\%,$$

where B_{buffer} — buffering data; B_{total} — total data.

Methodologically selected formulas allow you to quantify the improvement of communication quality when using different techniques. Initial test conditions: high -speed train: speed: 300 km/h; Route length: 500 km; Type of wagons: 10 wagons with integrated communication; Networks: LTE: frequency 800 MHZ, 1800 MHZ; 5G: frequency 3.5 GHZ Satellite ligament: LEO satellites. The transmission of video files in size 25–150 MB was tested. In Table 5 shows the results of testing existing basic methods.

Table 5. The results of testing existing basic methods

Method	Chain	Bandwidth, Mbps	Delay, ms	Jitter, ms	Batch loss, %	Buffering efficiency, %
Method 1: LTE	LTE	50	40	5	0.5	90
Method 2: 5G	5G	150	20	2	0.1	95
Method 3: Satellite	Satellite	20	150	30	2	70
Method 4: DAS	LTE/5G	80	35	4	0.3	93
Method 5: Wi-Fi	Wi-Fi	70	50	6	1.0	85
Method 6: Wi-Fi + LTE	Wi-Fi + LTE	90	30	3	0.4	88
Method 7: LTE + 5G	LTE + 5G	160	25	3	0.2	96
Method 8: Micro-networks	Micro-networks	75	45	5	0.6	90
Method 9: Satellite + 5G	Satellite + 5G	140	50	8	1.0	85
Method 10: Mobile roaming	Mobile roaming	60	70	12	1.5	80

The results of testing according to the new proposed methods are presented in Table 6.

Table 6. The results of testing are presented according to the new proposed methods

Method	Chain	Bandwidth, Mbps	Delay, ms	Jitter, ms	Batch loss, %	Buffering efficiency, %
Proposed Methodology 1	LTE + 5G + Satellite	180	30	4	0.3	94
Proposed Method 2	LTE + Wi-Fi + 5G	140	28	3	0.2	92
Proposed Methodology 3	5G + Micro-networks	170	25	3	0.1	95
Proposed Methodology 4	5G + Micro net-works + satellite	200	20	2	0.1	97
Proposed Methodology 5	LTE + DAS + Wi-Fi	120	35	4	0.3	91

From Table 5 and 6 it is clear that: Methodology 1: LTE + 5G + Satellite: The results confirm the research data that 5G provides the highest capacity, while the satellite is much lower. The high satellite retention corresponds to the fact that the satellite ligament is not suitable for applications where low delay is important, similar results were obtained in work [5]. Satellite bonds have major problems with jitter and batch loss, which is also confirmed by research [3]. Method 2: LTE + Wi-Fi + 5G: Wi-Fi significantly increases the overall capacity of the system, confirming the results. Wi-Fi provides good delays and jitter, which corresponds to research where Wi-Fi has less Wi-Fi delays also shows a lower level of packet loss than LTE, which confirms its effectiveness [6]. Methods 3: 5G + micro-networks: micro-networks locally increase the efficiency of bandwidth, but do not reach a speed of 5G [7]. Micro networks show a much lower delay and jitter compared to 5G, which is confirmed by research [7]. Micro networks have a lower batch loss, which is a positive aspect compared to 5G [8]. Methods 4: 5G + micro-networks + satellite of the combination of all three technologies provides a wide range of bandwidth, but the satellite bond limits the overall results [12]. The satellite bond adds considerable delay and jitter, which, according to research, reduces the total quality [10]. The high level of batch loss of satellite communications confirms its restriction for real -time use [9]. From the above it is evident that the technique 2 (LTE + Wi-Fi + 5G) provides the best combination of speed, delay and quality of communication for most applications by making Wi-Fi, which improves local performance. Method 1 and Methods 4 have some restrictions due to satellite communication, which strongly affects delay and quality. Method 1 (LTE + 5G + Satellite): Problems with compliance with current standards: High satellite delay exceeds the recommended limits of international ITU-R standards for delay (up to 200 ms). This can affect the overall quality of communication and require improvement of calibration and compensation for systematic errors. Recommendations: Consider improving the satellite components or reducing their use in combination to increase compliance with standards. Methodics 2 (LTE + Wi-Fi + 5G): compliance: compliance: meets the requirements of 3GPP standards for LTE and 5G, as well as IEEE for Wi-Fi. Jitter also provides less delay in accordance with modern quality standards. Recommendations: regular calibration and accurate measurement to support these standards. Methods 3 (5G + micro-networks): meets the requirements of 3GPP standards for 5G. Micro networks must adhere to IEEE specifications for wireless networks that may require clarification. Recommendations: Measurement and calibration accuracy for micro-networks to reduce possible errors. Methodics 4 (5G + micro-networks + satellite): Compliance problems: High delay and satellite jitter do not meet the recommended limits for modern quality standards . Requires comprehensive metrological control. Thus, technique 2 is the most appropriate to modern standards due to the combination of LTE, Wi-Fi and 5G, which provides optimalimatics and satellite jitter that influence their compliance with standards.

CONCLUSIONS

The analysis of the current state of raised in the article showed that the evaluation of errors is a critical step in ensuring high accuracy of measurements of mobile and Internet quality measurements in speed trains. During the movement of the train, there are often conditions that lead to an increase in errors due to rapid changes in infrastructure and network conditions. From the proposed techniques for improving the quality of mobile communications and the Internet in speed trains: Method 2 (LTE + Wi-Fi + 5G): shows the best results by combining technologies with low delay and griter. Metrological measurement confirms its effectiveness due to lower delays and jitter compared to other techniques. Therefore, technique 2 is the most effective in terms of metrology because of its combination of technologies, which provides the best results in the aspects of bandwidth, delay and jitter. The prospects for further research are to improve existing techniques, as well as to study the latest technologies that can help improve the quality of communication and Internet in speed trains.

REFERENCES

1. W. Bai, H. Dong, J. Lu, Y. Li, "Event-triggering communication based distributed coordinated control of multiple high-speed trains," *IEEE Transactions on Vehicular Technology*, 70(9), pp. 8556–8566, 2021. doi: https://doi.org/10.1109/tvt.2021.3099529
2. N. Branković, A. Kalem, A. Medić, "Development of mobile communication systems for high-speed railway," *Science, Engineering and Technology*, 1(1), pp. 29–34, 2021. doi: https://doi.org/10.54327/set2021/v1.i1.2
3. V. Dakulagi, M. Alagirisamy, "Adaptive beamformers for high-speed mobile communication," *Wireless Personal Communications*, 113(4), pp. 1691–1707, 2020. doi: https://doi.org/10.1007/s11277-020-07287-1
4. A. Gunasekar, L.B. Kumar, P. Krishnan, R. Natarajan, D.N.K. Jayakody, "All-Optical uav-based triple-hop FSO-FSO-VLC cooperative system for high-speed broadband internet access in high-speed trains," *IEEE Access*, 11, pp. 124228–124239, 2023. doi: https://doi.org/10.1109/access.2023.3330236
5. X. Ma, J. Liu, H. Jiang, "Energy-Efficient mobile data uploading from high-speed trains," *Mobile Networks and Applications*, 17(1), pp. 143–151, 2011. doi: https://doi.org/10.1007/s11036-011-0297-3
6. K. Qian, Z. Hou, Q. Sun, Y. Gao, D. Sun, R. Liu, "Evaluation and optimization of sound quality in high-speed trains," *Applied Acoustics*, 174, 107830, 2021. doi: https://doi.org/10.1016/j.apacoust.2020.107830
7. V. Riihimäki, T. Väärämäki, J. Vartiainen, T. Korhonen, "Techno-economical inspection of high-speed Internet connection for trains," *IET Intelligent Transport Systems*, 2(1), pp. 27–37, 2008. doi: https://doi.org/10.1049/iet-its:20070014
8. S. Sun, S. Zhang, W. Wang, "A new monitoring technology for bearing fault detection in high-speed trains," *Sensors*, 23(14), 6392, 2023. doi: https://doi.org/10.3390/s23146392
9. O.S. Trindade, T. Berisha, P. Svoboda, E. Bura, C.F. Mecklenbrauker, "Assessment of treatment influence in mobile network coverage on board high-speed trains," *IEEE Access*, 8, pp. 162945–162960, 2020. doi: https://doi.org/10.1109/access.2020.3021647
10. G. Tsiachtsiras, D. Yin, E. Miguelez, R. Moreno, "Trains of thought: High-speed rail and innovation in china," *SSRN Electronic Journal*, 2023. doi: https://doi.org/10.2139/ssrn.4331236
11. V. Vahidi, "High speed trains communication systems in 5G cellular networks," *Digital Signal Processing*, 115, 103075, 2021. doi: https://doi.org/10.1016/j.dsp.2021.103075
12. V. Vahidi, E. Saberinia, "Downlink data transmission for high-speed trains in 5G communication systems," *IET Communications*, 14(18), pp. 3175–3183, 2020. doi: https://doi.org/10.1049/iet-com.2020.0123
13. J. Wang, H., Zhu, N.J. Gomes, "Distributed antenna systems for mobile communications in high speed trains," *IEEE Journal on Selected Areas in Communications*, 30(4), pp. 675–683, 2012. https://doi.org/10.1109/jsac.2012.120502
14. X. Yao, B. Zhao, X. Li, S. Li, "Distributed formation control based on disturbance observers for high-speed trains with communication delays," *IEEE Transactions on Intelligent Transportation Systems*, 25(5), pp. 3457–3466, 2024. doi: https://doi.org/10.1109/tits.2023.3330536
15. J. Zhang, H. Du, P. Zhang, J. Cheng, L. Yang, "Performance analysis of 5G mobile relay systems for high-speed trains," *IEEE Journal on Selected Areas in Communications*, 38(12), pp. 2760–2772, 2020. doi: https://doi.org/10.1109/jsac.2020.3005492

Received 25.11.2024

INFORMATION ON THE ARTICLE

Natalia V. Shtefan, ORCID: 0000-0001-7926-8437, Kharkiv National University of Radio Electronics, Ukraine, e-mail: natalya.shtefan@nure.ua

Serhii V. Zhyhlo, ORCID: 0009-0006-4080-3811, Kharkiv National University of Radio Electronics, Ukraine, e-mail: serhii.zhyhlo@nure.ua

ДОСЛІДЖЕННЯ ТА РОЗРОБЛЕННЯ МЕТОДІВ ПОКРАЩЕННЯ ЯКОСТІ МОБІЛЬНОГО ЗВ'ЯЗКУ ТА МОБІЛЬНОГО ІНТЕРНЕТУ В ШВИДКІСНИХ ПОТЯГАХ / Н.В. Штефан, С.В.Жигло

Анотація. Запропоновано розгляд ефективних методів та засобів для забезпечення покращення якості мобільного зв'язку та мобільного Інтернету у швидкісних потягах. У спектрі даної тематики розглянуто актуальні питання, пов'язані з досягненням покращення якості мобільного зв'язку та Інтернету у швидкісних потягах. У ході практичного дослідження на модельно-комплексному рівні розглянуто метрологічні особливості запропонованих нових комбінованих методик щодо покращення якості мобільного зв'язку та Інтернету у швидкісних потягах. Установлено, що методика, яка передбачає комбінацію методів (LTE + Wi-Fi + 5G), показує найкращі результати за рахунок комбінації технологій з низькою затримкою і джитером. Метрологічне вимірювання підтверджує її ефективність через менші значення затримки і джиттера порівняно з іншими методиками, методика 3 (5G + Мікромережі): пропонує високі локальні показники, але обмежена пропускна здатність. Метрологічні дані підтверджують зменшену затримку і джиттер.

Ключові слова: комплексна модель, стандарти якості, інтеграційне тестування, модульне тестування, технологічні виклики, мікромережі, 5G, цифрові комунікації.

UDC 519.87; 535.345.67
DOI: 10.20535/SRIT.2308-8893.2025.3.04

SELECTION OF TARGET FUNCTION IN OPTICAL COATINGS SYNTHESIS PROBLEMS

O.V. MITSA, P.I. STETSYUK, S.S. ZHUKOVSKYI, O.M. LEVCHUK, V.I. PETSKO, I.V. SHAPOCHKA

Abstract. The article presents general information on the use of optical coatings in various industries and analyzes the main approaches to optimizing optical filter structures. An approach to solving a class of optical coating synthesis problems is proposed, based on the formation of a new optimization model. The primary attention is paid to the formalization and analysis of the target function. To determine the quality of the optical coating, the deviation of the spectral characteristics from the required ones was estimated using the least squares, least absolute deviation, and minimum criteria. As a result, both smooth and two non-smooth target functions are proposed and analyzed. The peculiarities of their application in solving optimization problems related to optical coating synthesis are described, and corresponding numerical experiments are presented.

Keywords: optical coatings synthesis, wide bandpass filters, mathematical modeling, optimization, r-algorithm.

INTRODUCTION

Optical layered coatings have been used in a vast array of applications across different industries for many decades. They are used to modify the behaviour of light, enhancing the performance of optical devices in several ways. These coatings are commonly made of thin films of different materials that are deposited onto a substrate using various techniques, including sputtering, evaporation, and chemical vapour deposition [1]. One of the most prominent applications of optical layered coatings is in the field of optics. Optical lenses, filters, and mirrors are coated with thin layers of materials such as titanium dioxide, silicon dioxide, and aluminium to modify their refractive index, reflectivity, and transmission properties. These coatings help to reduce unwanted reflections, increase the light transmission, and improve colour accuracy, resulting in sharper, clearer images [2]. The film industry also relies heavily on optical coatings to improve the performance of cameras and lenses. Antireflective coatings applied to camera lenses reduce lens flare and ghosting, leading to crisper, higher-quality images. Similarly, polarizing filters are used to eliminate reflections and glare, resulting in better contrast and richer colours in the final footage. Optical layered coatings are also crucial in the medical field [3]. They are used to improve the performance of various medical devices, such as endoscopes, surgical lasers, and imaging systems. These coatings help to increase light transmission, reduce unwanted reflections, and improve the resolution and contrast of medical images, resulting in more accurate diagnoses and better treatment outcomes. In the field of electronics, optical layered coatings are used in the production of various displays, including LCDs and OLEDs [4]. These coatings help to increase the brightness and contrast of displays, reduce glare and reflections, and improve colour accuracy. They are also used in the production of solar panels to increase the efficiency of light absorption and conversion into electricity [5].

There are various approaches to optimizing the structures of optical layered coatings [6]. The trial-and-error method [7] involves manually adjusting the thickness and refractive index of each coating layer until the desired optical performance is achieved. However, this method can be time-consuming and does not always lead to an optimal coating design. Analytical methods use mathematical equations to calculate the thickness and refractive index of each coating layer. Some common analytical methods [8] are based on quarter-wave structures or structures that use bandwidth matching. These methods are relatively easy to use but do not always result in the optimal coating structure.

Numerical methods use computer algorithms to model the behavior of light waves within the coating and optimize the structure based on predefined criteria. Some common numerical methods include the transfer matrix method and the reverse wave analysis (RWA) method. The transfer matrix method [9] does not provide a natural way to model these optical properties, making it insufficient for synthesizing optical coatings. This method also assumes linear transformations, which do not account for light dispersion as it passes through materials. In optical coatings, materials are typically used where dispersion is a significant factor and must be considered in the design. The RWA method [10] can be very sensitive to initial conditions or input data. Even minor errors or inaccuracies in measurements or models can lead to incorrect results. However, these methods can be highly accurate and consider a wide range of structural criteria, but they are computationally complex [11].

Genetic algorithms [12] can be effective for the synthesis of optical coatings, but they may require a significant amount of computational resources and can be quite slow. The method of microstructured surfaces [13] uses structured microelements on the surface to create the desired optical properties. However, their production can be complex and require high-precision processing. Optical coatings created using the phase mask method [14] can be sensitive to changes in temperature, humidity, and mechanical stresses, leading to changes in their optical properties.

When using numerical methods, the choice of the objective function plays an important role. This work proposes several objective functions that can be used to optimize the parameters of optical coatings. One smooth and two non-smooth objective functions are presented. The effectiveness of their use is demonstrated with an example of a non-smooth objective function.

PROBLEM STATEMENT AND MATHEMATICAL MODEL

Multilayer optical coatings represent a structure consisting of N layers. The j-th layer is characterized by two parameters: the refractive index (n_j) and the geometric thickness (d_j) (Fig. 1). There are two main tasks associated with them. The first task, known as the direct or analysis task, involves determining the spectral characteristics (transmission, reflection, and absorption coefficients) of a known multilayer thin-film system based on the known characteristics of the coating. The task of calculating the characteristics of an interference coating is based on solving the stationary wave equation in the plane wave approximation. To date, a large number of computational schemes have been developed for calculating optical coatings. Perhaps the most common approach is based on calculating the tangential components of the electric and magnetic field vectors sequentially at all layer boundaries that form the coating. Introducing the matrix form of recording equations that connect the field amplitudes at adjacent boundaries allowed for a compact and consistent consideration of interference effects in layered structures of all types.

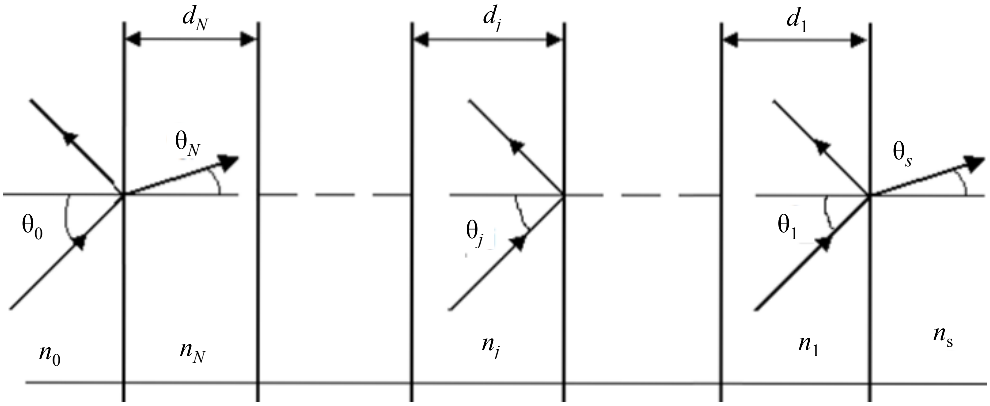


Fig. 1. Scheme of a light transmission through a multilayer optical structure

The second task, known as the inverse or synthesis task, involves determining the parameters of the multilayer optical structure that would optimally reproduce its predetermined spectral characteristics. In other words, the synthesis problem is to find such parameters of multilayered optical coating — refractive indices $\vec{n} = (n_1, n_2, \ldots, n_N)$, and geometric thicknesses of layers $\vec{d} = (d_1, d_2, \ldots, d_N)$ (N — number of layers), — under which, function, chosen to estimate transmittance factor quality, will be minimal in a given wavelength range $[\lambda_1, \lambda_2]$:

$$F^* = F(\vec{n}^*, \vec{d}^*) = \min_{\vec{n}, \vec{d}} F(\vec{n}, \vec{d}), \tag{1}$$

subject to

$$n_i^{\min} \le n_i \le n_i^{\max}, \; i = 1, 2, \ldots, N, \tag{2}$$

$$d_i^{\min} \le d_i \le d_i^{\max}, \; i = 1, 2, \ldots, N, \tag{3}$$

where F^* — minimum value of a coating target function.

Constrains (2), (3) have been imposed on the following parameters of multilayered optical coating — refractive indices and optical thicknesses. The refractive indices have been selected from the available coating-forming materials. Different sets of them can be created based on the spectral ranges of materials.

For visible and infrared rages, as a rule, the refractive index does not exceed 2.6. For the ultraviolet rage, materials with a higher refractive index can be used. Constraints (3) have been imposed on the geometric thickness of coating. The lower limit is tied to the application process, the upper limit, in the process of making multilayered optical coatings, as a rule, does not exceed the operating wavelength λ_0.

The value of the energy transmittance index for the electromagnetic wavelength λ through the multilayer optical structure should light fall on the surface at an angle θ_0 (Fig.1) has been calculated through the coefficients of the characteristic matrix $M(\vec{n},\vec{d},\lambda)$ as follows:

$$T(\vec{n},\vec{d},\lambda,\theta_0) =$$

$$= \frac{4}{2+\frac{p_0}{p_s}M_{11}^2(\vec{n},\vec{d},\lambda,\theta_0)+\frac{p_s}{p_0}M_{22}^2(\vec{n},\vec{d},\lambda,\theta_0)+p_0 p_s M_{12}^2(\vec{n},\vec{d},\lambda,\theta_0)+\frac{1}{p_0 p_s}M_{21}^2(\vec{n},\vec{d},\lambda,\theta_0)},$$

where $p_0 = n_0\cos\theta_0$ and $p_s = n_s\cos\theta_s$ — for TE wave (s-polarization); $p_0 = \frac{n_0}{\cos\theta_0}$ and $p_s = \frac{n_s}{\cos\theta_s}$ — for TE wave (p -polarization); θ_0 — angle of incidence; θ_s — angle of reflection; n_0, n_s — refractive indices of an environment and a substrate, accordingly.

The characteristic matrix of the *N*-layer structure is equal to the product of the matrices of each of the layers [15]:

$$M(\vec{n},\vec{d},\lambda,\theta_0) =$$

$$= M(n_N,d_N,\lambda,\theta_N)M(n_{N-1},d_{N-1},\lambda,\theta_{N-1})\cdots M(n_2,d_2,\lambda,\theta_2)M(n_1,d_1,\lambda,\theta_1),$$

where the characteristic matrix of the layer equals

$$M(n,d,\lambda,\theta) = \begin{matrix} \cos\delta(n,d,\lambda,\theta) & -\frac{i}{n}\sin\delta(n,d,\lambda,\theta), \\ -in\sin\delta(n,d,\lambda,\theta) & \cos\delta(n,d,\lambda,\theta), \end{matrix}$$

$\delta(n,d,\lambda,\theta) = \frac{2\pi nd\cos\theta}{\lambda}$ — phase thickness of the layer; θ —angle of incidence.

Angles of incidence for each layer follow the Snell's law and can be easily calculated according to the ratio:

$$n_0\sin\theta_0 = n_1\sin\theta_1 = n_2\sin\theta_2 = \ldots = n_j\sin\theta_j = \ldots = n_N\sin\theta_N = n_s\sin\theta_s.$$

If $\theta_0 = 0$, then the value of transmittance factor for the *N*-layer optical structure can be calculated using the following formula

$$T(\vec{n},\vec{d},\lambda) =$$

$$= \frac{4}{2+\frac{n_0}{n_s}M_{11}^2(\vec{n},\vec{d},\lambda)+\frac{n_s}{n_0}M_{22}^2(\vec{n},\vec{d},\lambda)+n_0 n_s M_{12}^2(\vec{n},\vec{d},\lambda)+\frac{1}{n_0 n_s}M_{21}^2(\vec{n},\vec{d},\lambda)},$$

where the characteristic matrix of the *N*-layer structure is written as

$$M(\vec{n},\vec{d},\lambda) = M(n_N,d_N,\lambda)M(n_{N-1},d_{N-1},\lambda)\cdots M(n_2,d_2,\lambda)M(n_1,d_1,\lambda),$$

and characteristic matrix of one layer is given by

$$M(n,d,\lambda) = \begin{matrix} \cos\frac{2\pi nd}{\lambda} & -\frac{i}{n}\sin\frac{2\pi nd}{\lambda} \\ -in\sin\frac{2\pi nd}{\lambda} & \cos\frac{2\pi nd}{\lambda} \end{matrix}.$$

It should be noted, that characteristic matrix of the multilayered optical structure meets following condition

$$det(M(\vec{n},\vec{d},\lambda)) = 1. \quad (4)$$

This follows from the fact that the characteristic matrix of each layer has the same property

$$det(M(n_i, d_i, \lambda)) = 1, \; i = 1, 2, \ldots, N.$$

Property (4) has a simple physical meaning. If an electromagnetic wave propagates in N media that do not absorb its energy, then an arbitrarily combined (of these N media) medium will not absorb the energy of the electromagnetic wave.

OPTICAL COATING TARGET FUNCTIONS AND THEIR USE

The following coating target functions can be chosen to solve the synthesis problem (1)–(3):

$$F_1(\vec{n},\vec{d}) = \frac{1}{L}\sum_{i=1}^{L} w_i (T(\vec{n},\vec{d},\lambda_i) - T_{ideal}(\lambda_i))^2, \quad (5)$$

$$F_2(\vec{n},\vec{d}) = \frac{1}{L}\sum_{i=1}^{L} w_i \left|T(\vec{n},\vec{d},\lambda_i) - T_{ideal}(\lambda_i)\right|, \quad (6)$$

$$F_3(\vec{n},\vec{d}) = \max_{\mathrm{i=1,\ldots,L}} w_i \left|T(\vec{n},\vec{d},\lambda_i) - T_{ideal}(\lambda_i)\right|, \quad (7)$$

where w_i — weighting coefficients, which determine the input on the objective function at wavelength λ_i; L — the number of grid points on the spectral interval between λ_1 and λ_2; $T(\vec{n},\vec{d},\lambda_i)$ — the value of the transmission index for parameters $(\vec{n},\vec{d})$ and at wavelength λ_i; $T_{ideal}(\lambda_i)$ — the value of the transmission index at wavelength λ_i.

Coating target functions (5)–(7) have been described below. Function $F_1(\vec{n},\vec{d})$ sets the weighted standard deviation of the transmittance indices from the required for the selected L values of wavelengths. This function is smooth, so gradient methods, quasi-Newton methods and zero-order methods (use only the values of the objective function) can be used to minimize it. Function $F_2(\vec{n},\vec{d})$ sets the weighted sum of deviations from the mean with respect to the selected L. Function $F_3(\vec{n},\vec{d})$ specifies deviation under minimax control (Chebyshev criterion). The functions $F_2(\vec{n},\vec{d})$ and $F_3(\vec{n},\vec{d})$ are non-smooth, so Shore r-algorithms and zero-order methods can be used to minimize them.

When solving the antireflective coating substrate problem, the values of $T_{ideal}(\vec{n},\vec{d},\lambda_i)$ are constand and equal to unity. With regard to afford mentioned, the objective functions takes the form:

$$F_1(\vec{n},\vec{d}) = \frac{1}{L}\sum_{i=1}^{L} w_i (T(\vec{n},\vec{d},\lambda_i) - 1)^2,$$

$$F_2(\vec{n},\vec{d}) = \frac{1}{L}\sum_{i=1}^{L} w_i \left|T(\vec{n},\vec{d},\lambda_i) - 1\right|,$$

$$F_3(\vec{n},\vec{d}) = \max_{\mathrm{i=1,\ldots,L}} w_i \left| T(\vec{n},\vec{d},\lambda_i) - 1 \right|.$$

Given that the value of transmittance factor is less than unity, the function $F_2(\vec{n},\vec{d})$ can be expressed in the following form

$$F_2(\vec{n},\vec{d}) = \frac{1}{L}\sum_{i=1}^{L} w_i \left| T(\vec{n},\vec{d},\lambda_i) - 1 \right| = \frac{1}{L}\sum_{i=1}^{L} w_i - \frac{1}{L}\sum_{i=1}^{L} w_i T(\vec{n},\vec{d},\lambda_i),$$

and will be smooth, when solving the antireflective coating substrate problem. In the similar fashion, the function $F_3(\vec{n},\vec{d})$ can be expressed in the following form

$$F_3(\vec{n},\vec{d}) = \max_{i=1,\ldots,L} w_i \left| T(\vec{n},\vec{d},\lambda_i) - 1 \right| = \max_{i=1,\ldots,L} w_i (1 - T(\vec{n},\vec{d},\lambda_i)),$$

But in contrast to the function $F_2(\vec{n},\vec{d})$, it`s non-smooth.

If all $w_i = 1$, we obtain following objective functions:

$$F_1(\vec{n},\vec{d}) = \frac{1}{L}\sum_{i=1}^{L} (T(\vec{n},\vec{d},\lambda_i) - 1)^2,$$

$$F_2(\vec{n},\vec{d}) = \frac{1}{L}\sum_{i=1}^{L} \left| T(\vec{n},\vec{d},\lambda_i) - 1 \right| = 1 - \frac{1}{L}\sum_{i=1}^{L} T(\vec{n},\vec{d},\lambda_i),$$

$$F_3(\vec{n},\vec{d}) = \max_{\mathrm{i=1,\ldots,L}} \left| T(\vec{n},\vec{d},\lambda_i) - 1 \right| = \max_{\mathrm{i=1,\ldots,L}} (1 - T(\vec{n},\vec{d},\lambda_i)).$$

In a number of studies problems of wide bandpass optical coatings synthesis have been reviewed as maximization problems for similar deviations, and not for the maximum transmittance, but for the minimum possible, i.e. zero value of the transmittance [16]. For weighted standard deviation, there is an alternative, where the maximization problem can be described as

$$\max_{\vec{n},\vec{d}} \left(\mathrm{F}(\vec{n},\vec{d}) = \frac{1}{L}\sum_{i=1}^{L} T^2(\overrightarrow{n(\lambda_i)},\vec{d},\lambda_i) \right), \tag{8}$$

subject to (2) and (3).

In a similar way, for weighted sum of deviations from the mean this problem can be described as

$$\max_{\vec{n},\vec{d}} \left(\mathrm{F}(\vec{n},\vec{d}) = \frac{1}{L}\sum_{i=1}^{L} T(\overrightarrow{n(\lambda_i)},\vec{d},\lambda_i) \right), \tag{9}$$

subject to (2) and (3).

And for deviation under minimax control (Chebyshev criterion) is as follows

$$\max_{\vec{n},\vec{d}} \left(\mathrm{F}(\vec{n},\vec{d}) = \min_{\mathrm{i=1,\ldots,L}} T(\overrightarrow{n(\lambda_i)},\vec{d},\lambda_i) \right). \tag{10}$$

subject to (2) and (3).

For these models, which use target functions (8)–(10), it is assumed that there may be a refractive index dispersion. Accordingly, the value of the refractive index is a function of wavelength and function is defined using approximation Zellmeier formula

$$n_i(\lambda) = \sqrt{A_i + \frac{B_i}{\lambda^2} + \frac{C_i}{\lambda^4} + D_i\lambda^2 + E_i\lambda^4} \ ,$$

where A_i, B_i, C_i, D_i, E_i — parameters for refraction index model in the presence of dispercion. Optical materials can be described either by the values of the dispersion formula coefficients, or directly by the values of the refractive index for different wavelengths. For many optical materials, this information is available in databases. Also, during the study, one layer can be considered smooth or partially inhomogeneous [17].

Problem (1)–(3) is multiextremal. It contains $2N$ variables, where the first N variables are the refractive indices of the layers, the second N variables are the geometric thicknesses of the layers. Bilateral constraints on variables are set by conditions (2)–(3). The local minima of the problem (1)–(3) often provide the required approximation accuracy and have implementable coating parameters. Such solutions are often called quasi-optimal. In this work we decided to follow up on the suggested term, so by quasi-optimal solutions we will always mean such local extremums of problem (1)–(3), for which the found coating parameters are practically feasible.

Problem (1)–(3) can be modeled as unconstrained optimization by using transition from one variables to another

$$x_j = x_j^{\min} + (x_j^{\max} - x_j^{\min})\sin^2 z_j, \tag{11}$$

$$x_j = \frac{x_j^{\max} z_j^2 + x_j^{\min}}{z_j^2 + 1}, \quad j = 1, \ldots, N. \tag{12}$$

Thus, a solution for each parameter can be found at infinity. An objective function has been complicated by this. Formula variables (11) provide a smoother change of the formed surface and have less abrupted transition in comparison to another formula (12). On the other hand, the transition to unconstrained optimization by formula (11) requires the calculation of the value of $\arcsin(x)$, which is a rather time-consuming operation. For the approach used in this paper, this applies to both the values of geometric thicknesses and refractive indices. To do this, the minimum and maximum refractive indices must be selected.

As the number of layers increases, more parameters for reduction of the target functions $F(\vec{n}, \vec{d})$ value in the optimization problems of optical coatings synthesis, can be obtained. Therefore, it is necessary to clarify the criterion for termination of the search process for solving optimization problem (1)–(3). This goal can be archived by looking for ε solution:

$$\left| F(\vec{n}_\varepsilon^*, \vec{d}_\varepsilon^*) - F^* \right| \le \varepsilon. \tag{13}$$

In case of minimization problem — it will be inequation $F(\vec{n}_\varepsilon^*, \vec{d}_\varepsilon^*) - F^* \le \varepsilon$, and in the case of maximization problem — $F^* - F(\vec{n}_\varepsilon^*, \vec{d}_\varepsilon^*) \le \varepsilon$.

The introduction of inequality (13) into the optimization model has been caused by two factors. First, there are a large number of quasi-optimal solutions that can have a design implementation. Secondly, it is often impossible to achieve an exact approximation of predetermined spectral characteristics. The spectral characteristics of the optical coating are analytical functions and can be differentiated an infinite number of times [11]. Accordingly, if the idealized characteristic is constant or has gaps, then exact approximation cannot be obtained. From a practical point of view, the definition of the problem should also include a condition of limiting the number of layers, which would serve as a criterion for termination of the search process, and can serve for correction of a sufficiently small value of ε.

An additional condition associated with the manufacture of optical coating selects one design from a variety of solutions that meet criterion (13), and accordingly, the second Hadamard condition will be met. This condition must also take into account the characteristics of the selected materials, their interaction with each other.

The application of the Monte Carlo method allows choosing the most fault tolerant design solutions [18]. Therefore, for the chosen optical coating, the condition must be met that a slight change in the input parameters will also satisfy criterion (8), and, respectively, will satisfy the third Hadamard condition.

COMPUTATIONAL EXPERIMENT

The developed approach has been applied to improve the behaviour of existing wide bandpass coatings. For this purpose, we used Shor's R-algorithm [11; 19] with coating target function represented as

$$F(\vec{n},\vec{d}) = \min_{\lambda_1 \le \lambda \le \lambda_2} \sum_{i=1}^{L} \left| 1 - T(\vec{n},\vec{d},\lambda_{(i)}), \right|$$

where $[\lambda_1,\lambda_2]$ — wavelength range under study; L — number of points in the wavelength range from λ_1 to λ_2. In this section, the chosen value of L equals $\lambda_2 - \lambda_1 + 1$, i.e. in the objective function, each integer-value of the interval was considered $[\lambda_1,\lambda_2]$.

Let us demonstrate application of the proposed optimization approach on a practical example. For this, we will use three optical coatings known in the industry.

In wavelength range between 450 and 800, value of the first coating target function $F(\vec{n},\vec{d}) = 1.404$ (curve 1 — parameters of the optical coating known in the industry $3.76 \cdot n_1 d_1 = 3.76 \cdot n_2 d_2 = 0.455 \cdot n_3 d_3 = n_4 d_4 = 0.25 \cdot \lambda_0$, $n_1=2.0$, $n_2=1.37$, $n_3=2.0$, $n_4=1.37$), and for the second — $F(\vec{n},\vec{d})$ = 0.838 (curve 2 — parameters, which have been calculated in this article $6.58 \cdot n_1 d_1 = 4.06 \cdot n_2 d_2 = 0.441 \cdot n_3 d_3 = 0.944 \cdot n_4 d_4 = 0.25 \cdot \lambda_0$, $n_1=2.1$, $n_2=1.35$, $n_3=1.9$, $n_4=1.35$). Accordingly, value of the coating target function $\mathrm{F}(\vec{n},\vec{d})$ has been improved by 40% (Fig. 2).

Graph of the coating target function can be easily assessed, if we will fix all parameters, except two (except geometric thicknesses of third and fourth layers, have been fixed, for optical coating with parameters $0.153 \cdot n_1 d_1 = 0.25 \cdot n_2 d_2 = 0.25 \cdot \lambda_0$, $n_1=1.35$, $n_2=1.9$, $n_3=1.35$, $n_4=2.1$ in the case of antireflection coating application with refractive index $n_s = 1.52$). As can be seen in the Fig. 3, even the part of the graph let us assume that this graph has a ravine-type shape. Let's consider the sevenlayer antireflection coating, consisting of alternating layers

(1.35 and 2.1), for which layer optical depths in respect to λ_0 are as follows — 0.05 : 0.071 : 0.062 : 0.257 : 0.018 : 0.12 : 0.2, for which all derived optimal parameters, except geometric thicknesses of sixth and seventh layers, has been fixed. Resulting graph (Fig. 6) clearly shows that graph of the estimated target function has, indeed, a ravine-type shape. It has fixed all the optimal parameters, except geometric thicknesses of sixth and seventh layers, have been fixed, for sevenlayer antireflection coating, consisting of alternating layers with refractive indices 1.35 and 2.1, layer optical depths of the first five layers with respect to λ_0 are as follows — 0.05 : 0.071 : 0.062 : 0.257 : 0.018.

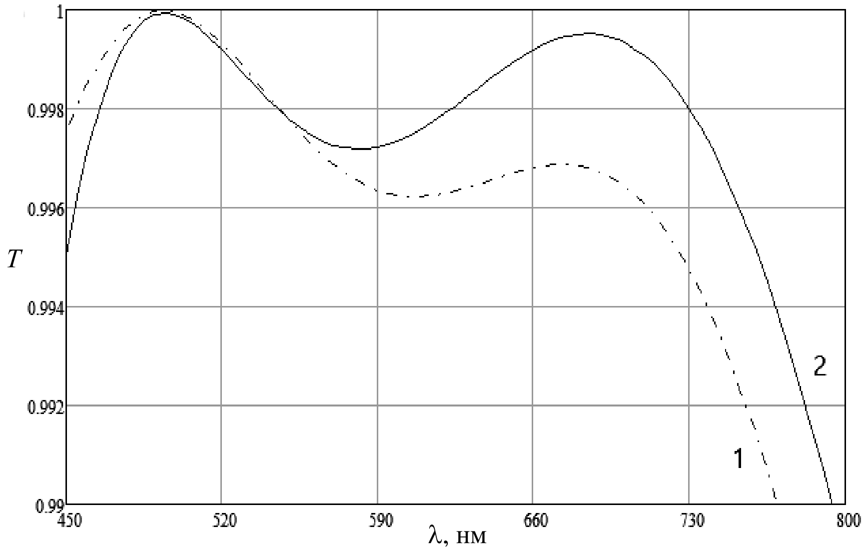


Fig. 2. Wide bandpass filter transmittance curve in the case of antireflection coating application with refractive index n_s=1.51

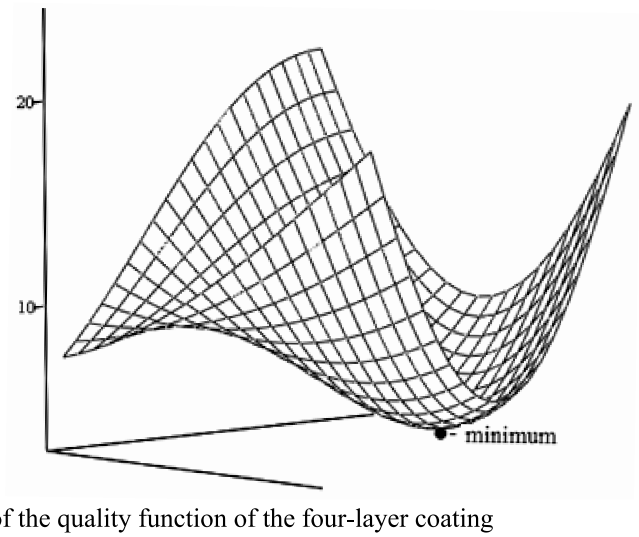


Fig. 3. Graph of the quality function of the four-layer coating

In wavelength range between 450 and 750, value of the first coating target function $F(\vec{n},\vec{d}) = 0.665$ (curve 1 — parameters of the optical coating known in the industry, layer optical depths with respect to λ_0 are as follows — 0.064 : 0.038 : 0.401 : 0.032 : 0.084 : 0.459 : 0.229), and for the second —

$F(\vec{n},\vec{d}) = 0.324$ (curve 2 — parameters, which have been calculated in this article, layer optical depths with respect to λ_0 are as follows — 0.087 : 0.03 : 0.315 : 0.043 : 0.113 : 0.48 : 0.22). Accordingly, value of the coating target function $F(\vec{n},\vec{d})$ has been improved by more than 50% (Fig. 4).

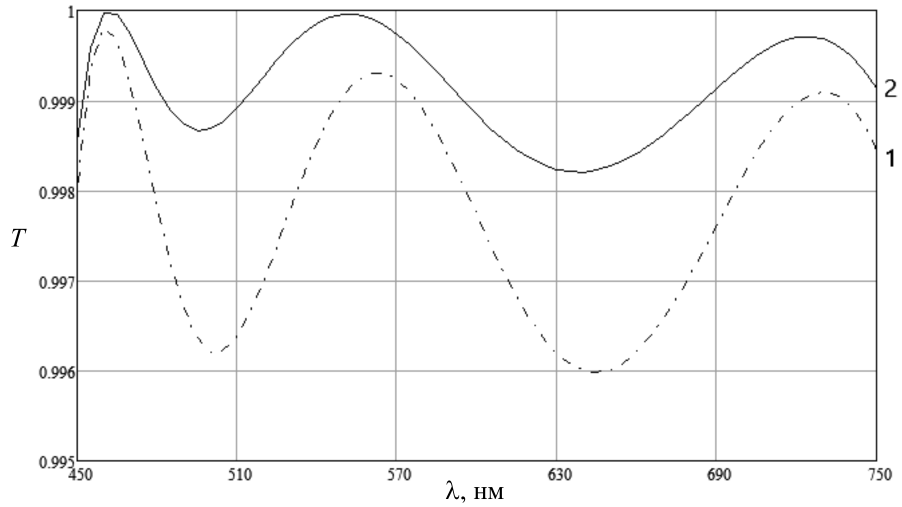


Fig. 4. Transmittance curve for sevenlayer antireflection coating, consisting of alternating layers (1.35 and 2.1) of substrate with refractive index n_s=1.52

It should be noted, that for gradient methods, the use of this objective function gives a less effective result. For these methods, one must use the target function (10).

In wavelength range between 450 and 750, value of the first coating target function $F(\vec{n},\vec{d}) = 0.953$ (curve 1 — parameters of the optical coating known in the industry, layer optical depths with respect to λ_0 are as follows — 0.06 : 0.02 : 0.35 : 0.02 : 0.07 : 0.42 : 0.21), and for the second — $F(\vec{n},\vec{d}) = 0.478$ (2 — parameters, which have been calculated in this article, layer optical depths with respect to λ_0 are as follows — 0.05 : 0.071 : 0.062 : 0.257 : 0.018 : 0.12 : 0.2). Accordingly, value of the coating target function $F(\vec{n},\vec{d})$ has been improved by almost 50% (Fig. 5).

CONCLUSIONS

This paper describes three types of target functions, which can be used for solving optimization problems of optical coatings synthesis. Their reduction to the problems of unconstrained minimization of smooth and non-smooth functions has been described and the peculiarities of the transition to new variables for each of the proposed models has been investigated. The following computer implementations can be used to accelerate solving optical coating synthesis problems: tabulation of values of trigonometric functions, fast matrix multiplication and the use of an efficient method for one-dimensional optimization.

A computational experiment has been performed, in which the target function in the form of the weighted sum of deviations from the mean was taken and spectral characteristics of the three available wide bandpass antireflection filters has been improved by using the r-algorithm for optimization. For one of the wide bandpass antireflective coatings, the target function was improved by 40%, and for the other two, the target function was improved by 50%.

Acknowledgments. The authors are grateful to colleagues from the Department of Information Management Systems and Technologies of the Uzhhorod National University and colleagues from the Department of Nonsmooth Optimization Methods of V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences for a productive discussion of the topic and the results of the work.

REFERENCES

1. P. Baumeister, Optical coating technology. SPIE press, 2004. doi: https://doi.org/10.1117/3.548071
2. A. Piegari, F. Flory (ed.), *Optical thin films and coatings: From materials to applications*. Woodhead Publishing, 2018.
3. B. Mansoor, S. Li, W. Chen, "Highly efficient antifogging/antimicrobial dual-functional chitosan based coating for optical devices," *Carbohydrate Polymers*, 296, 119928, 2022. doi: 10.1016/j.carbpol.2022.119928
4. R. Singh, K.N. Unni, A. Solanki, "Improving the contrast ratio of OLED displays: An analysis of various techniques," *Optical Materials*, 34(4), pp. 716–723, 2012. doi: 10.1016/j.optmat.2011.10.005
5. M.J. Mendes et al., "Design of optimized wave-optical spheroidal nanostructures for photonic-enhanced solar cells," *Nano Energy*, 26, pp. 286–296, 2016. doi: 10.1016/j.nanoen.2016.05.038
6. N. Nedelcu et al., "Design of highly transparent conductive optical coatings optimized for oblique angle light incidence," *Applied Physics A*, 127(8), article no. 575, 2021. doi: 10.1007/s00339-021-04726-z
7. J.A. Dobrowolski, F.C. Ho, A. Waldorf, D.F. Mitchell, "Trial-and-error optimization of optical coatings," *Applied Optics*, 32(28), pp. 5481–5490, 1993.
8. R. Willey, "Design and optimization of optical coatings and optical systems using analytical methods," *Thin Solid Films*, 287(1-2), pp. 24–31, 1996.
9. S. Larouche, L. Martinu, "OpenFilters: An Open Source Software for the Design and Optimization of Optical Coatings," in *Optical Interference Coatings, OSA Technical Digest (CD) (pp. WB6)*. Optica Publishing Group, 2007. doi: 10.1364/OIC.2007.WB6
10. H.A. Macleod, *Thin-film optical filters*. CRC Press, 2017.
11. P.I. Stetsyuk, A.V. Mitsa, "Parameter Optimization Problems for Multilayer Optical Coatings," *Cybernetics and Systems Analysis*, 41(4), pp. 564–571, 2005. doi: https://doi.org/10.1007/s10559-005-0092-x
12. S. Martin, J. Rivory, M. Schoenauer, "Synthesis of optical multilayer systems using genetic algorithms," *Applied Optics*, 34, pp. 2247–2254, 1995. doi: https://doi.org/10.1364/AO.34.002247
13. K. Haas-Santo, M. Fichtner, K. Schubert, "Preparation of microstructure compatible porous supports by sol–gel synthesis for catalyst coatings," *Applied Catalysis A: General*, 220(1-2), pp. 79–92, 2001. doi: 10.1016/S0926-860X(01)00714-1
14. V. Boominathan, J.K. Adams, J.T. Robinson, A. Veeraraghavan, "Phlatcam: Designed phase-mask based thin lensless camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(7), pp. 1618–1629, 2020. doi: 10.1109/TPAMI.2020.2987489
15. F. Abeles, "The propagation of electromagnetic waves in stratified media," *Annals of Physics*, 3(4), pp. 504–520, 1948.
16. Y.A. Pervak, A.V. Mitsa, J.I. Holovach, I.V. Fekeshgazi, "Influence of transition film-substrate layers on optical properties of the multilayer structures," *Selected Papers from the International Conference on Optoelectronic Information Technologies. International Society for Optics and Photonics*, vol. 4425, pp. 321–325, 2001. doi: 10.1117/12.429744

17. A. Mitsa, V. Mitsa, A. Ugrin, “Mathematical modeling of spectral characteristics of optical coatings with slightly inhomogeneous chalcogenide films,” *Journal of Optoelectronics and Advanced Materials*, 7(2), pp. 955–962, 2005.
18. R.Y. Rubinstein, D.P. Kroese, *Simulation and the Monte Carlo method*. *John Wiley & Sons*, 2016.
19. J.V. Burke, A.S. Lewis, M.L. Overton, “The Speed of Shor’s R-algorithm,” *IMA Journal of Numerical Analysis*, 28(4), pp. 711–720, 2008. doi: 10.1093/imanum/drn008

Received 02.04.2024

INFORMATION ON THE ARTICLE

Oleksandr V. Mitsa, ORCID: 0000-0002-6958-0870, Uzhhorod National University, Ukraine, e-mail: alex.mitsa@uzhnu.edu.ua

Petro I. Stetsyuk, ORCID: 0000-0003-4036-2543, V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences, Ukraine, e-mail: stetsyukp@gmail.com

Serhii S. Zhukovskyi, ORCID: 0000-0001-5826-0751, Zhytomyr Ivan Franko State University, Ukraine, e-mail: zss@zu.edu.ua

Oleksandr M. Levchuk, ORCID: 0000-0001-6344-9356, Uzhhorod National University, Ukraine, e-mail: alex.levchuk@uzhnu.edu.ua

Vasyl I. Petsko, ORCID: 0009-0009-7679-5288, Uzhhorod National University, Ukraine, e-mail: vasyl.petsko@uzhnu.edu.ua

Ihor V. Shapochka, ORCID: 0000-0003-0904-7879, Uzhhorod National University, Ukraine, e-mail: ihor.shapochka@uzhnu.edu.ua

ВИБІР ФУНКЦІЇ ЯКОСТІ В ЗАДАЧАХ СИНТЕЗУ ОПТИЧНИХ ПОКРИТТІВ
/ О.В. Міца, П.І. Стецюк, С.С.Жуковський, О.М. Левчук, В.І. Пецко, І.В. Шапочка

Анотація. Наведено загальні відомості про використання оптичних покриттів у різних галузях промисловості та проаналізовано основні підходи до оптимізації структур оптичних фільтрів. Запропоновано підхід до вирішення класу задач синтезу оптичних покриттів, заснований на формуванні нової оптимізаційної моделі. Основну увагу приділено формалізації та аналізу цільової функції. Для визначення якості оптичного покриття використано оцінку відхилення спектральних характеристик від необхідних за критеріями найменших квадратів, найменших абсолютних відхилень і мінімаксу. У результаті запропоновано та досліджено як гладку, так і дві негладкі цільові функції. Описано особливості їх застосування в розв’язуванні оптимізаційних задач синтезу оптичних покриттів та наведено відповідні числові експерименти.

Ключові слова: синтез оптичних покриттів, широкосмугові фільтри, математичне моделювання, оптимізація, *r*-алгоритм.

UDC 004.41
DOI: 10.20535/SRIT.2308-8893.2025.3.05

USE OF METHODS AND TOOLS FOR ENSURING SOFTWARE QUALITY

A.S. SHANTYR

Abstract. This paper proposes an examination of effective methods and tools for ensuring software quality. The scope of this topic includes current issues related to software quality assurance within the context of analyzing methods and tools used in practice to develop high-quality software. During the modeling process, a new comprehensive model for software quality assurance has been developed, combining modular testing, integration testing, and continuous integration methods. The advantage of this development is its enhanced adaptability to addressing key challenges in software quality assurance. Based on the developed model, strategies and approaches are proposed to improve configuration management processes and identify vulnerabilities in software systems.

Keywords: comprehensive model, quality standards, integration testing, unit testing, technological challenges, integration testing, continuous integration.

INTRODUCTION

In the modern digital world, often referred to as the era of software, the concept of software quality has become not only relevant but critically important. In all areas of activity, from business and science to everyday needs, software has become an integral component. However, the mere existence of a software product does not guarantee its success [1]. In this regard, it is quite appropriate to agree with the views of the authors of the work [2], which conclude that quality becomes the decisive factor influencing decision-making and meeting user needs [2].

According to the authors of the study [3], achieving high-quality software requires not only innovative technologies but also effective methods and tools for ensuring quality. These methods and tools encompass a wide range of activities, from testing and code analysis to quality management and the implementation of best practices in the software development process. Scientific research in the field of software quality assurance has been conducted for several decades, and many aspects have already been addressed, including:

- test automation: numerous tools, frameworks, and Continuous Integration / Continuous Deployment practices have been developed for automatic code quality verification [1–3];
- quality metrics: a set of metrics has been defined and implemented, such as code coverage, system response time, the number of defects per unit of code, and others [4; 5];
- software security: immunity to many types of attacks has been developed, popular vulnerabilities such as SQL injection, cross-site scripting (XSS) and others have been identified and fixed;

- development methodologies: there are many recognized methodologies, such as Scrum, Kanban, DevOps, which regulate development and quality assurance processes.

However, there are also unsolved problems and directions for further research, namely:

- intelligent testing: unsolved questions regarding the evaluation of the development of machine learning methods for automated testing that are able to adapt to changes in the code and respond to new functions. The essence of the problem of intelligent testing within the framework of quality assessment of software systems boils down to the fact that testing requires an understanding of the functionality of the program and the possibilities of automation. Ensuring the adaptability of tests to changes in the code requires intelligent systems. The complexity of solving the above-mentioned problem lies in the fact that at the current level of technological development of qualitative evaluation of software systems, there are technological difficulties in the development of optimized machine learning algorithms that can adapt to changes in the software code and detect new functionalities automatically;
- security of microservice architecture: issues related to ensuring security and integrity in the context of the rapid development of microservice systems, given the large number of interacting components, have not been fully resolved. According to [5], the significant number of interacting microservices creates a vast attack surface, complicating the detection and resolution of potential vulnerabilities. As per [6], the challenge in addressing this problem at the methodological level lies in the incomplete implementation of mechanisms that address the development of systems for detecting and monitoring potential attacks, as well as the development of secure microservice development practices;
- automation of vulnerability detection: the issues surrounding the development of effective tools for automated detection of new vulnerabilities and their subsequent remediation have not been fully addressed. According to [7], the essence of this problem is that vulnerabilities evolve, and tools need to detect new security threats. The challenge here is the constant need to develop more effective, cutting-edge tools and technologies capable of identifying new and unknown vulnerabilities through contextual understanding and advanced detection methods;
- combination of functional and learning-based testing: approaches that combine functional testing with machine learning to more effectively detect defects have not been fully developed;
- security guarantees at large scales: the development of methods and architectures to ensure the security of software systems in large, distributed, and complex ecosystems has not been fully addressed. The core of the problem is that in large ecosystems, such as cloud services, ensuring security at all levels is challenging [8]. Additionally, there are difficulties in using distributed architecture, encryption, and security standards to protect vast amounts of data and systems;
- ethical aspects in testing: the development of standards and ethical norms for automated testing systems and their impact on people has not been fully addressed.

The core of the problem is the need to solve issues related to testing that may affect user privacy and security. The complexity of this issue lies in the definition of standards and the implementation of practices that ensure ethical standards in testing.

ANALYSIS OF RECENT PUBLICATIONS

Foidl H., Felderer M. [1] considered the principles of integration of software quality models based on risk-oriented testing. The authors noted above solve the problem of defining software by using quality models. Within the scope of the raised topic, the authors propose to integrate software quality models based on risk-oriented testing, according to their position, this approach will help significantly improve the effectiveness of testing, allowing to focus on areas with high risk for software.

The advantages of this approach are more efficient use of resources: focusing testing on high-risk areas allows for more efficient use of limited resources, directing them to the most critical aspects of the software. In addition, risk-based testing, based on quality models, allows you to better consider business needs and focus on aspects of the program that have the greatest impact on business processes. The authors also note that the integration of quality models helps to identify potential problems and risks in advance, which allows developers to pay more attention to their solution in the early stages of development. In addition, the authors proved that focusing on important areas of risk improves software quality by identifying and solving problems important to users. However, the integration of software quality models into risk-based testing proposed by Foidl and Felderer may have certain drawbacks, namely:

- complexity of integration: transferring quality models into the context of risk-based testing can be challenging, especially if there are different methods and approaches to defining quality and risks;
- data heterogeneity: quality models may be based on different data sets, and risks may arise from various sources. Integrating this data can be problematic due to its heterogeneity and incompatibility;
- ambiguity in defining risks and quality: the concepts of risk and quality may have different interpretations for different stakeholders, making the integration of models difficult due to this ambiguity;
- need for large amounts of data: effective operation of integrated models requires a significant amount of data, both on software quality and risks. This can be a problem where access to such data is limited;
- system complexity: integrating models may require significant effort and resources to develop, implement, and maintain complex systems.

Thus, while integrating software quality models into risk-based testing can bring substantial benefits in identifying high-risk areas for software, it is essential to carefully study and consider the aforementioned challenges for successful application of this approach. K. Sahu and R. K. Srivastava [2] examined the foundations of software error prediction using neuro-fuzzy logic methods. Their research emphasizes software reliability and the need to find effective error prediction methods. However, the researchers did not explore the possibility of integrating error prediction methods with other quality assurance methods, such as statistical models, testing, or defect analysis. In general, recent research in the field of effective methods and tools for software quality assurance significantly improves the efficiency of quality assurance processes. According to [9], at the technological-methodological level, the issues related to the aforementioned analysis require a deep understanding of the functional and technical aspects of software systems, as well as innovative approaches to solving them. The complexity lies in combining technical expertise with creativity and flexibility in addressing diverse problems. According to [10], research on software quality assessment methods is ongoing to identify effective tools and techniques. However, according to [10], research in the field of software quality assessment requires continuous improvement of methods and tools, as well as consideration of changes in the technological landscape and user requirements. According to [11], research on effective methods and tools for software quality assurance is a crucial field in information technology. According to [12], issues related to software quality assurance can arise for various reasons, and solving them requires a systematic approach. In the framework of research on quality assurance methods and tools, the researchers from works [1–15] encountered numerous issues requiring additional comprehensive analysis. Specifically, in the area of automated testing, it is advisable to conduct research to develop the most appropriate automated tests covering various aspects of the software and its functionality to prevent malfunction in real-world conditions [1–4]. In the realm of improving quality metrics, it is important to conduct research on developing and implementing effective metrics that take into account various aspects of quality, such as reliability, performance, and security [3]. In terms of addressing software security issues, it is necessary to analyze and improve methods for identifying and eliminating potential vulnerabilities and to develop tools that enhance compliance with security requirements in a growing environment of cyber threats [4–6]. Thus, recent studies require further investigation to continue comprehensive work on resolving the identified challenges.

Setting the task. Aim of the work is to study effective methods and tools for ensuring the quality of software systems and to develop a new comprehensive model for ensuring the quality of software systems, which combines the methods of module testing, integration testing and continuous integration, which contributes to the improvement of the quality of the software product at various stages of development.

Achieving the set goal comes down to solving the following problems:

- conducting a generalized analysis of current issues related to the study of effective methods and means to ensure the quality of software systems;
- conducting a comprehensive contextual base-review generalization of the main modern qualitative and quantitative methods that are used to assess the quality of software systems;
- analysis of the main mathematical models involved in integration testing and quality assessment of software systems;
- development and research of a new complex model of quality assurance of software systems, which is based on a combination of methods of module testing, integration testing and continuous integration.

THE MAIN PART

Conducting a generalized analysis of current issues related to the study of effective methods and tools for ensuring the quality of software systems, it is appropriate to highlight a number of issues in this area, which are presented in Table 1.

Table 1 is made by the author of this article on the basis of works analysis [1–15].

Table 1. Generalized analysis of current issues related to the study of effective methods and tools to ensure the quality of software systems

The name of the generalized area of topical issues	Generalized description of the problem	Research that is needed within the framework of solving this issue
Automated testing	Lack of full test coverage, which can lead to incorrect operation of the program in real conditions	Research within the development of automated tests that cover various aspects of the program and its functionality
Quality metrics	Lack of clear metrics for measuring software quality	Research within the development and implementation of effective metrics that take into account various aspects of quality such as reliability, performance and safety
Software security	Ensuring the security of software systems in the growing environment of cyber threats	Research that involves the analysis and improvement of methods for identifying and eliminating potential vulnerabilities, as well as the development of means to improve compliance with security requirements
Development methodologies	The choice of the most appropriate methodology for the development of software systems, taking into account the trends in ensuring the quality of the software product.	Research that involves a comparative analysis of various methodologies, identifying their advantages and disadvantages, as well as developing recommendations for choosing the appropriate approach to development
Quality standards	Lack of uniform quality standards for different fields of development.	Research on the development and implementation of quality standards that take into account the specifics of various types of software systems
Version control of software systems and configuration management	Effective implementation and management of changes in software code	Research on the development of tools for automated version control and configuration management to ensure code stability and quality

Shown in Table 1 the results of a generalized visual analysis show that the quality of software systems has several key aspects that need attention and research. In particular, in the field of resolving issues on automated testing, it is advisable to conduct research on improving software testing, which is used to prevent improper operation of programs and accordingly covers the analysis of their functionality in real conditions [1–4].

With the limits of improving quality metrics, it is important to conduct research on the development and implementation of effective metrics that take into account various aspects of quality, such as reliability, productivity and security of software [3]. In the framework of adaptive development, quality metrics in our opinion are advisable to apply the definition of key quality indicators (KPIs) to measure different aspects of software quality.

Fig. 1 shows a scheme of complex contextual basis-view generalization of basic modern qualitative methods that are used to assess the quality of software systems. Analyzing in Fig. 1. Methods should be noted that today, in the field of quality assessment of software systems, specific standards do not determine the

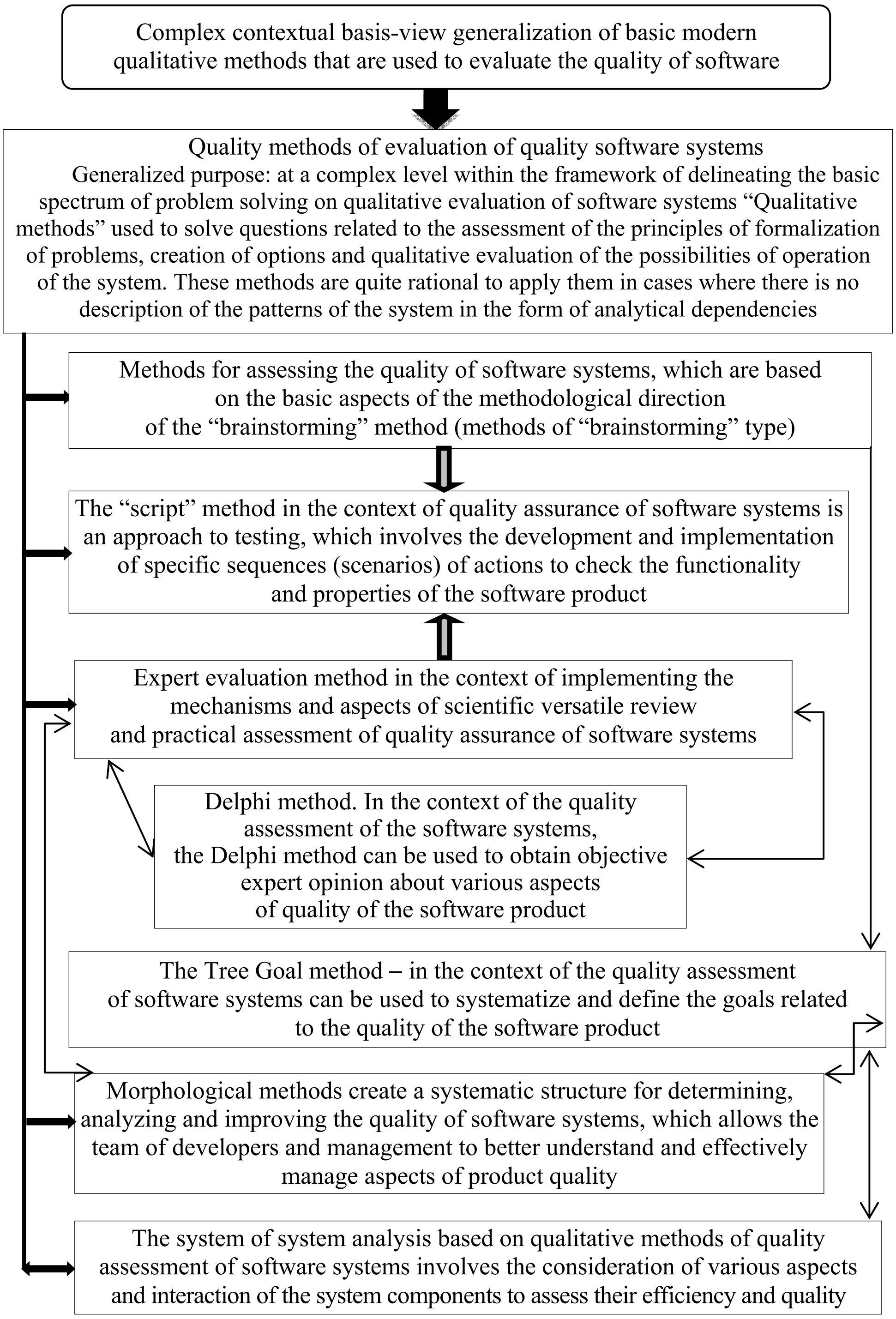


Fig. 1. Scheme of complex contextual basis-view generalization of basic modern qualitative methods that are used to assess the quality of software systems

use of specific brainstorming methods. However, brainstorming methods should be used in quality management processes and software testing, as well as in collecting requirements and design [9].

Brainstorming methods can be successfully integrated into these standardized processes to achieve better results in solving software quality problems [1−14]. Analyzing the script method in the context of quality assurance of software systems, we note that this is an approach to testing, which involves the development and implementation of specific sequences (scenarios) of actions to check the functionality and properties of the software [6]. Brainstorming methods can be used in the development of test scenarios, identification of potential defects and collecting requirements for testing.

According to ISO/IEC 9126 (Software Engineering − Product Quality): ISO/IEC 9126 is related to the quality of the software. Brainstorming methods can be used in determining the requirements for functionality, reliability, ease of use, etc.

Below are examples of how brainstorming can be used in the context of quality assessment of software systems:

- ISO/IEC 25010 (Square): ISO/IEC 25010 Standard "Systems and Software. Model of Quality and Assessment" determines the software quality model. Brainstorming methods can be used in determining the requirements, as well as in the analysis and planning of testing to ensure the completeness and variety of test scenarios;
- IEEE 730 (Standard for Software Quality Assurance Processes): The IEEE 730 Standard defines the quality assurance processes. IstQB (International Software Testing Qualifications Board): ISTQB initiative defines standards for certification of software testing professionals.

It is worth noting that the standards do not fully specify the fulfillment of the completeness of the use of brainstorming methods, they provide context and principles for the introduction of creative and collective approaches to solving problems in the quality of software systems.

In Table 2 the generalized characteristics of methodological directions of qualitative methods in the spectrum of quality of software systems are given. Brainstorming methods can be successfully integrated into these standardized processes to achieve better results in solving software quality problems [1−14].

Analyzing the script method in the context of quality assurance of software systems, we note that this is an approach to testing, which involves the development and implementation of specific sequences (scenarios) of actions to check the functionality and properties of the software [6]. Brainstorming methods can be used in the development of test scenarios, identification of potential defects and collecting requirements for testing.

According to ISO/IEC 9126 (Software Engineering —Product Quality): ISO/IEC 9126 is related to the quality of the software. Brainstorming methods can be used in determining the requirements for functionality, reliability, ease of use, etc.

Below are examples of how brainstorming can be used in the context of quality assessment of software systems:

- ISO/IEC 25010 (Square): ISO/IEC 25010 Standard "Systems and Software. Model of Quality and Assessment" determines the software quality model. Brainstorming methods can be used in determining the requirements, as well as in the analysis and planning of testing to ensure the completeness and variety of test scenarios;
- IEEE 730 (Standard for Software Quality Assurance Processes): the IEEE 730 Standard defines the quality assurance processes. IstQB (International Soft-

ware Testing Qualifications Board): ISTQB initiative defines standards for certification of software testing professionals.

It is worth noting that the standards do not fully specify the fulfillment of the completeness of the use of brainstorming methods, they provide context and principles for the introduction of creative and collective approaches to solving problems in the quality of software systems.

In Table 2 the generalized characteristics of methodological directions of qualitative methods in the spectrum of quality of software systems are given.

Quantitative methods used to evaluate the quality of software systems have the following basic generalized characteristics:

- quantitative formalization of problems: quantitative methods allow to structure and formalize problems using quantitative indicators and parameters. Provides objectivity and quantitative evaluation of system parameters;

Table 2. Generalized characteristics of methodological directions of qualitative methods in the spectrum of quality assessment of software systems

The name of the methodological direction	Description	Advantages and disadvantages
Testing	It includes the execution of programs in order to detect errors or deficiencies	Effectively detects some types of errors, helps to confirm compliance with requirements. Cannot guarantee the absence of all errors, testing costs may be high
Code analysis	Expert analysis of software code to identify potential problems or inefficiencies	Can reveal code complexity, potential vulnerabilities. Limited to the quality of the code, it is difficult to detect certain types of errors
Automated quality analysis tools	Using tools for automated code quality analysis and error detection	Reduces dependence on the human factor, quickly detects common problems. May underestimate context and complex aspects of code
Code check	Code review by the development team to identify errors and improve quality	Attracts experts, promotes knowledge exchange, is effective in identifying shortcomings. Requires time and effort from the development team
Quality modeling	Use of mathematical models	Allows to carry out quality analysis before product release

Specified in Table 2 methodological directions can be used separately, or in combination to achieve a more accurate and complete assessment of the quality of software systems.

In Fig. 2 shows a scheme of complex contextual basis-view generalization of basic modern quantitative methods that are used to assess the quality of software systems. Accordingly [3–4] quantitative methods within the framework of the implementation of the quality of software systems allow quantitative analysis and make sound solutions based on numerical data and system parameters:

- quantitative formation of variants: quantitative methods help to create different variants of system implementation, taking into account quantitative aspects. Provides objectivity and quantitative evaluation of the possibilities of different options;

- quantitative evaluation of operation options: used to quantify the efficiency and suitability of different options for the system. Allows objective analysis and comparison of different scenarios of system use;

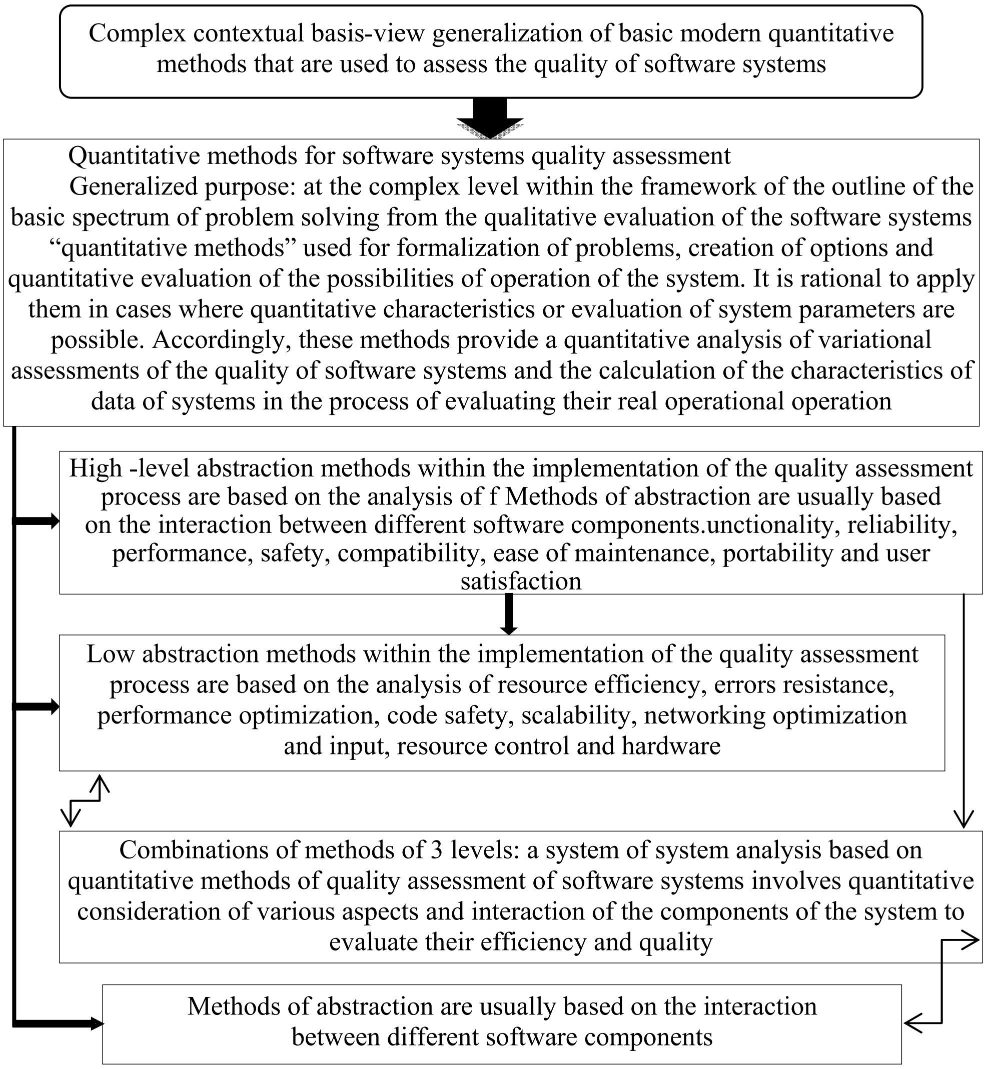


Fig. 2. Scheme of complex contextual basis-view generalization of the main modern quantitative methods that are used to evaluate the quality of software systems

- application in the presence of analytical dependencies: quantitative methods are effectively used when the possibility of expression of the system analytically allows quantitative characteristics. Provides accuracy and objectivity in the presence of quantitative data.

The modular testing method is closely linked to the concept of software testing and arose in the context of software testing. However, the specific term "modular testing" and its methodology received significant flowering through a movement known as "Extreme Programming" (XP), which began in 1990–2000 (developers: Kent Beck and his colleagues). XP is a software development methodology that defines certain practices and principles to improve the quality and speed of development. It should be noted that modular testing involves writing automated tests for each individual "module" of the code before its integration into the system. The essence of this method is to create the maximum number of automated tests for the individual components of the program and run them during development to detect errors in the early stages.

Integration testing is an important component of the quality assessment of software systems and is usually performed at the development stage when the different components of the software system are grouped into a single system. The main purpose of integration testing is to check the interaction between these components, as well as integrated system functions.

Different approaches and techniques are used in the integration testing process, such as:

- testing of modules on isolation: before integration of modules into the system, each module is tested separately to check its functionality and compliance with requirements;
- interface testing: an important step in integration testing is to check the interaction between components through their interfaces. This may include data testing between modules and verifying the data of data integrity;
- testing of the interaction of components: after the integration of the modules test the interaction between them, convincing that they work properly together and perform the expected functions;
- exception and error testing: integration testing also includes checking the system response to exceptional situations and errors, such as incorrect data or inaccessible resources;
- productivity testing: during integration testing, performance tests are also performed to ensure that the system works efficiently and is capable of processing the load caused.

The purpose of integration testing is to ensure the high quality of the software system by identifying and solving problems related to the integration of components. This allows you to ensure the correct and reliable operation of the system as a whole.

In Table 3 the results of the analysis of the main mathematical models involved in integration testing and quality assessment of software systems are presented. From Table 3 it is clear that each model has its own unique essence and purpose, which affects the process of evaluating the quality of software systems. The interaction matrix and interaction graphs allow the visualization of the structure and interaction between the components of the software system, which contributes to the identification of weaknesses and the analysis of the complexity of interaction. Models of interaction, in turn, enable a mathematical description of this interaction, which allows to carry out a deeper analysis and forecasting of the behavior of the system.

It should also be noted that when choosing models to evaluate the quality of software systems, it is important to consider their advantages and disadvantages, as well as the context of the project and the specifics of the system. The combination of different models can be useful to obtain a more complete view of the system's state and test efficiency.

The constant integration method (CI) is a key practice in the field of software development aimed at improving the quality of software systems. The main idea of the CI is to regularly and automatically combine code changes in a common repository and perform automatic tests to check the correctness of these changes. Below is a description of the constant integration method and its impact on the quality assessment of software systems:

- code changes automated: developers regularly make changes to the project code. Thanks to the CI, these changes automatically merge (integrate) into the main branch of the repository. This allows you to identify conflicts and other problems before they get into the production environment;

- automatic tests: after each change of change, the CI system automatically starts tests that check the functionality, performance, safety and other aspects of the software system. This allows you to identify possible problems during the early stages of development;

Table 3. Analysis of basic mathematical models involved in integration testing and quality assessment of software systems

Mathematical model	Essence	Appointment	Advantages	Disadvantages
Interaction matrix	Representation of interaction between components in the form of a matrix	Visualization and analysis of the interaction structure, identification of "weak places"	Ease of use, quick identification of defects	Does not take into account the dynamics of interaction, limited in use for complex systems
Interaction graph	Representation of interaction between components in the form of a graph	Visualization and analysis of complex interaction, detection of cycles and ways of interaction	Detailed overview of the interaction structure, help with architecture analysis	Difficult to use for large systems
Interaction models	Mathematical models describing the interaction between components	Analysis and prediction of interaction between components, identification and elimination of shortcomings	Deep interaction analysis, possibility of predicting system behavior	Requires a large amount of data and computing resources, difficult to implement
Probability models	Mathematical models that determine the probability of various events	Assessment of risks and effectiveness of testing, detection and elimination of shortcomings	Objective assessment of efficiency and risks, the possibility of making decisions based on probability	It is difficult to model all possible scenarios and factors affecting the probability
Testing models	Models describing the process of testing a software system	Organization of testing, determination of the most effective approaches and strategies	Systematization and structuring of the testing process, the possibility of choosing the best approaches	Requires precise definition of testing parameters and criteria, may not be flexible enough in some situations

- constant feedback: if problems are identified during the performance of tests, the developers receive notifications about it, which allows them to quickly correct mistakes and improve the quality of the code;
- automated deployment: some CI systems may automatically deploy changes to the test or production environment after successful completion of the tests. This allows you to quickly and effectively introduce new functionality and correction of errors;
- code quality metrics: some CI systems may include code analyzers and quality metrics that automatically estimate the level of readability, efficiency and other aspects of code quality.

The constant integration method allows you to create higher quality software systems, reducing the time of detection and correcting errors, facilitating the joint work of developers and increasing the reliability of software as a whole.

MODELING AND TESTING

In the range of modeling of a new complex model of quality assurance of software systems, based on the combination of modular testing, integration testing and constant integration methods, it is important to keep in mind a number of initial assumptions. In particular, given the automation of processes in our case, it is suggested that many stages of testing and integration can be automated. This means that the creation and performance of tests can be carried out without significant intervention.

Given the specifics of the modular testing, it is suggested that the individual components of the program (modules) are tested in isolation from other components to ensure that they work properly isolated from other parts of the program.

Given the specifics of integration testing, it is suggested that after successful modular testing, the components of the program are gradually combined and tested as a holistic system to believe that they are cooperating correctly.

Continuous delivery involves the assumption that the processes of correction of errors, testing and release of software are automatically and continuously to ensure rapid delivery of changes and updates.

Considering the needs of assembling the assembly and deployment at the complex level, it is suggested that the processes of collecting the software code and its automatic deployment into the test environment or the productive server must also be automated.

Considering the need for automation tools, there is an assumption that there are appropriate tools and technologies for the implementation of automated testing, integration and constant integration processes. Such tools can include media for testing, version control systems, automated testing tools and more. With the above initial assumptions, you can start modeling and implementation of a new complex model of quality assurance of software systems.

Mathematical apparatus for the developed new model of complex quality assurance of software systems, based on the combination of modular testing methods, integration testing and constant integration can be presented as follows.

1. Tasks of the modular testing procedure of the software system:

$$S_u = N_u \times P_u,$$

where S_u — the overall success of modular testing; N_u — the total number of wise tests; P_u — predicted probability of successful passage of one modular test.

2. Tasks of the procedure testing procedure:

$$S_i = N_i \times P_s,$$

where S_i — the overall success of integration tests; N_i — the total number of modular tests; P_i — The probability of successful passage of one integration test.

3. Tasks of the continuous integration procedure:

$$C = F_C \times T_C,$$

where C — total number of committees (variable code); F_C — frequency of committees; T_C — total time during which committees were carried out.

4. Tasks of the procedure of continuous integration and constant delivery:

$$R = C / T_C$$

where C — total number of committees (variable code); R — the speed of releases, which is determined by the number of committees and time; T_C — the total time during which committees were carried out.

5. The task of determining the quality of the software system:

$$Q = S_u \times S_i \times R,$$

where Q — The total quality of the software system.

Subsequently optimizing the mathematical apparatus for the model of complex quality assurance of software systems, which is based on the combination of modular testing methods, integration testing and constant integration, may include the following characteristics and methods:

Code quality metrics: used to quantify the quality of the software code. These metrics may include code coating tests, defects, test time, and more.

Statistical methods: used to analyze the results of errors, test efficiency and system stability.

Probability and statistics theory: used to calculate the likelihood of defects during testing, risk assessment and statistical significance of test results.

Optimization methods: used to increase the efficiency of testing and select the optimal test strategies taking into account resource restrictions.

Machine learning and artificial intelligence algorithms: used to automate the test analysis processes, identify anomalies and forecast possible problems.

These methods and tools allow you to create a complete and optimized complex mathematical model to ensure the quality of software systems, which takes into account the interaction of different test methods and their impact on the quality and reliability of the software.

Description of the algorithm of practical implementation of the developed model. In practice, the implementation of the developed complex model of quality assurance of software systems will include several steps and stages:

1. Planning and setting up the environment:

- Setting up the version control system (GIT example) for constant integration.
- Installation and adjustment of tools for automated testing (for the example of JUnit for modular testing, Selenium for automated integration testing).

2. Creating tests:

- Developing a set of modular tests for each component of the program.
- Writing integration tests to check the interaction between components.

3. Setting up constant integration (PI):

- Create configuration files for automatic assembly and testing after each committee.
- Setting up integration with the task management system or notification of the developers about the results of the testing.

4. Performing tests and analysis of results:

- Automatic performance of modular tests and integration tests after each committee.
- Analysis of test results and detection of errors or disadvantages in the code.

5. Control of versions and releases:

- Management of software versions through the version control system.
- Automatic software release based on test results and successful integration.

6. Monitoring and reporting:

- Monitoring the work of the continuous integration system and test results.
- Generation of code quality reports, testing and testing results for analysis and improvement of the development process. This process requires the work of developers, testers and operators, as well as the introduction of a number of tools for automation of routine testing and integration tasks.

The developed model of quality assurance of software systems, which combines modular testing, integration testing and constant integration, has several useful advantages:

Improved software quality: this model allows you to identify and correct errors in the early stages of development, providing high quality software.

Reducing the time for detecting and correcting errors: constant integration and continuous delivery allow you to identify and correct problems quickly, which reduces the time of development and improves the speed of product production.

Testing and release automation: through automated software testing and release processes, the risks of human errors can be reduced and the project stability can be reduced. Improving development efficiency: developers can focus on writing code, as many routine tasks (testing, assembly, deployment) are automated. Reduced testing costs and releases: automated quality assurance processes save time and money that is usually spent on manual testing and releases.

Increasing user confidence: high product quality and renewal speed helps to improve users' reputation and trust.

In general, this model helps to make the software development process more efficient, faster and reliable, which is critical in the modern fleeting world of software development.

In Table 4 the results of practical testing of plowing methods and the developed model in the analy sis of the quality of the application "task management system".

Table 4. Analysis of comparison of software testing methods in the analysis of the quality of the application "task management system"

Features	**Modular testing**	**Integration testing**	**Continuous integration**	**Combination of methods (unit testing + Integration + PI)**
Execution time	5 hours	8 hours	12 hours	10 hours
Number of detected errors	10	50	30	70
Level of automation	High	Average	High	High
Stability of releases	High	average	High	High
Test coverage (percentage)	90%	70%	95%	97%
Stability (scale 1-10)	9	7	8	8

From Table 4 it is observed that software quality testing by the developed method takes 10 hours, which is less than continuous integration and close to the average time required for integration testing. Thus, in terms of execution time, the combined method is quite efficient. Analyzing the number of errors detected: the combined method detects 70 errors, which is more than any other testing method. This indicates its high efficiency in detecting errors in software. Analyzing the level of automation: the combined method has a high level of automation, which allows for efficient testing without significant human involvement. Analyzing release stability and stability: the combined method has a high level of release stability, which is equal to the release stability obtained with continuous integration. This is important to ensure software quality. Analyzing test coverage: the combined method has the highest rate of test coverage — 97%. This indicates that it tests more parts of the software and detects more possible problems. So, from the standpoint of comparison with existing testing methods, the combined method (Unit Testing + Integration + PI) is very effective, as it combines the advantages of different approaches and ensures high quality and stability of the software. In turn, it should be noted that in order to obtain the results of practical testing of the described methods and the developed quality model during the quality analysis of the "task management system" application, various tools were used, including: Automated test frameworks: popular test frameworks were used for module testing and integration testing, such as like JUnit, NUnit, or PyTest. The above-mentioned test frameworks allow you to automate the execution of tests and analyze their results.

Systems for continuous integration: tools such as Jenkins, Travis CI, or GitLab CI were used for continuous integration. These systems allow you to automate the process of building, testing, and deploying software with each code change.

Code quality monitoring and analysis: code quality monitoring tools such as SonarQube or CodeClimate were used. These tools allow you to identify potential problems in your code, such as code duplication, improper use of variables, and other anomalies. Resource Usage Analyzers: profilers and performance analyzers such as VisualVM have been used to analyze the application's resource usage, and YourKit is also possible. These tools allow you to identify performance and memory usage issues in an application. Bug trackers: used bug trackers like Jira, Bugzilla, GitHub Issues to track the issues found and fix them later. These tools contributed to obtaining practical results of testing and quality analysis of the "task management system" application using the described testing methods and the developed quality assurance model. The application of this strategy and approaches is quite expedient in terms of improving configuration management processes and identifying vulnerabilities in software systems.

CONCLUSIONS

The analysis of the current state of the issue raised in the article showed that an important task is to further improve the methods and means of quality assurance of software systems in the context of rapid technology development. Research confirms that effective methods and quality assurances are a key element of successful development of software systems. In the course of modeling, a new complex model of quality assurance of software systems was developed, which is based on the combination of modular testing methods, integration testing and constant integration. The advantage of this development is to increase its adaptation to solving the main tasks for the quality of software systems. The results of practical research have shown that the integrated use of methods, such as modular testing, integration testing and continuous integration, better facilitates the detection and correction of errors in the early stages of software development. The prospects for further research are to improve existing techniques, as well as to study the latest technologies that can help improve the quality of software in the future.

REFERENCES

1. H. Foidl, M. Felderer, "Integrating software quality models into risk-based testing," *Software Quality Journal*, vol. 26, no. 2, pp. 809–847, 2016. doi: https://doi.org/10.1007/s11219-016-9345-3
2. K. Sahu, R.K. Srivastava, "Predicting software bugs of newly and large datasets through a unified neuro-fuzzy approach: reliability perspective," *Advances in Mathematics: Scientific Journal*, vol. 10, no. 1, pp. 543–555, 2021. doi: https://doi.org/10.37418/amsj.10.1.54
3. K. Sahu, Fahad A. Alzahrani, R.K. Srivastava, R. Kumar, "Evaluating the Impact of Prediction Techniques: Software Reliability Perspective," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 1471–1488, 2021. doi: https://doi.org/10.32604/ cmc.2021.014868
4. S. Sackey, D.-E. Lee, B.-S. Kim, "Duration Estimate at Completion: Improving Earned Value Management Forecasting Accuracy," *KSCE Journal of Civil Engineering*, vol. 24, issue 3, pp. 693–702, 2020. doi: https://doi.org/10.1007/s12205-020-0407-5
5. P. Sharma, A.L. Sangal, "Building and Testing a Fuzzy Linguistic Assessment Framework for Defect Prediction in ASD Environment Using Process-Based Software Metrics," *Arabian Journal for Science and Engineering*, vol. 45, issue 12, pp. 10327–10351, 2020. doi: https://doi.org/10.1007/s13369-020-04701-5

6. Y. Hassouneh et al., “Boosted Whale Optimization Algorithm with Natural Selection Operators for Software Fault Prediction,” *IEEE Access*, vol. 9, pp. 14239–14258, 2021. doi: https://doi.org/10.1109/access.2021.3052149
7. Rafa Al-Qutaish, “Quality Models in Software Engineering Literature: An Analytical and Comparative Study,” *Journal of American Science*, vol. 6, pp. 166–175, 2010.
8. J. Estdale, E. Georgiadou, “Applying the ISO/IEC 25010 Quality Models to Software Product,” in Larrucea X., Santamaria I., O’Connor R., Messnarz R. (eds) Systems, Software and Services Process Improvement. EuroSPI 2018. Communications in Computer and Information Science, vol. 896, 12 p., 2018. Springer, Cham. doi: https://doi.org/10.1007/978-3-319-97925-0_42
9. C. Zhang B. Li, L. Wang, H. Xu, T. Shao, “A Hierarchical Model for Quality Evaluation of Mixed Source Software Based on ISO/IEC 25010,” *International Journal of Software Engineering and Knowledge Engineering*, 33(02), pp. 1–25, 2022. doi: https://doi.org/10.1142/S021819402250070X
10. Elder Bruno Evaristo Correa, Jeffson Celeiro Sousa, Antônio Jorge Gomes Abelém, Sandro Ronaldo Bezerra Oliveira, “An evaluation of Security Features based on Iso/Iec 25023 for a Distributed Autonomic Scientific Publisher Tool on a Permissioned Blockchain,” *Journal of Information Systems and Technology Management*, vol. 19, 2022. doi: https://doi.org/10.4301/S1807-1775202219020
11. J.-X. Chen, “Overall performance evaluation: new bounded DEA models against unreachability of efficiency,” *Journal of the Operational Research Society*, vol. 65, issue 7, pp. 1120–1132, 2014. doi: https://doi.org/10.1057/jors.2013.57
12. M.-A. Filz, C. Herrmann, S. Thiede, “Simulation-based Assessment of Quality Inspection Strategies on Manufacturing Systems,” *Procedia CIRP*, vol. 93, pp. 777–782, 2020. doi: https://doi.org/10.1016/j.procir.2020.04.069
13. A. Golabchi, S. Han, S. AbouRizk, “A simulation and visualization-based framework of labor efficiency and safety analysis for prevention through design and planning,” *Automation in Construction*, vol. 96, pp. 310–323, 2018. doi: https: //doi.org/10.1016/j.autcon.2018.10.001
14. P. Han, L. Wang, P.X.K. Song, “Doubly robust and locally efficient estimation with missing outcomes,” *Statistica Sinica*, 26(2), 2016. doi: https://doi.org/ 10.5705/ss.2014.030
15. L. Hund et al., “Distinguishing between model- and data-driven inferences for high reliability statistical predictions,” *Reliability Engineering & System Safety*, vol. 180, pp. 201–210, 2018. doi: https://doi.org/10.1016/j.ress.2018.07.017

Received 04.04.2024

INFORMATION ON THE ARTICLE

A.S. Shantyr, ORCID: 0000-0002-0466-3659, State University of Information and Communication Technologies, Ukraine, e-mail: shantyr.a@duikt.edu.ua

ВИКОРИСТАННЯ МЕТОДІВ ТА ІНСТРУМЕНТІВ ДЛЯ ЗАБЕЗПЕЧЕННЯ ЯКОСТІ ПРОГРАМНИХ СИСТЕМ / А.С. Шантир

Анотація. Запропоновано розгляд ефективних методів та засобів для забезпечення якості програмних систем. У спектрі даної тематики розглянуто актуальні питання, пов’язані із забезпеченням якості програмних систем у межах реалізації аналізу методів та засобів, які застосовуються на практиці для забезпечення розроблення високоякісного програмного забезпечення. У ході моделювання розроблено нову комплексну модель забезпечення якості програмних систем, яка ґрунтується на поєднанні методів модульного тестування, інтеграційного тестування та постійного інтегрування. Перевагою цієї розробки є підвищення її адаптування до вирішення основних завдань із забезпечення якості програмних систем. На базі розробленої моделі запропоновано стратегії та підходи щодо вдосконалення процесів управління конфігурацією та виявлення вразливостей програмних систем.

Ключові слова: комплексна модель, стандарти якості, інтеграційне тестування, модульне тестування, технологічні виклики, інтеграційне тестування, неперервне інтегрування.

UDC 62.50
DOI: 10.20535/SRIT.2308-8893.2025.3.06

AUTOMATED CONTROL OF DYNAMIC SYSTEMS FOR ENSURING UKRAINE'S SECURITY USING COGNITIVE MAP IMPULSE PROCESS MODELS. PART 1. DEMOGRAPHIC SECURITY

V. ROMANENKO, Y. MILIAVSKYI

Abstract. The paper provides a cognitive map (CM) of demographic security and a dynamic model of CM impulse processes described as a difference equations system (Robert's equations). The external control vector for the CM impulse process is implemented by means of varying the CM nodes' coordinates. A closed-loop control system for the CM impulse process is proposed. It includes a multivariate discrete controller designed based on an automated control theory method, which generates the chosen control actions. We solve a discrete controller design problem for automated control of dynamic processes to ensure demographic security. The controller suppresses external and internal disturbances during CM impulse processes control based on the invariant ellipsoids method. The paper presents an algorithm for CM weights identification based on the recurrent least squares method. We present the results of a qualitative research study on dynamic processes related to demographic security in Ukraine under various disturbances during martial law.

Keywords: cognitive map, demographic security, invariant ellipsoid, linear matrix inequalities, impulse process.

INTRODUCTION

To study the dynamic processes for system ensuring of Ukraine's demographic security we use cognitive modelling, which is one of the most relevant areas of scientific and practical research of complex systems of different nature now. Cognitive modeling is based on the notion of a cognitive map (CM), which is a weighted directed graph, its nodes reflect coordinates (factors, concepts) of the complex system and weighted edges (arcs) of the graph describe interrelations between CM nodes. When disturbances affect CM nodes, we can observe impulse transitional process, its dynamics is described by the difference equation [1]:

$$\Delta y_i(k+1) = \sum_{j=1}^{n} a_{ij} \Delta y_j(k), \tag{1}$$

where $\Delta y_i(k) = y_i(k) - y_i(k-1), i = 1,2,...,n$, a_{ij} — weight of an edge connecting the j-th node and the i-th one. Equation (1) describes the free motion of the i-th node of CM without external control impact. We can write this equation in vector-matrix form:

$$\Delta \overline{Y}(k+1) = A\Delta \overline{Y}(k), \qquad (2)$$

where $\Delta \overline{Y}(k) = \overline{Y}(k) - \overline{Y}(k-1)$, A is a weighted adjacency matrix of the CM of size $n \times n$.

In order to implement control of the CM impulse process (2) based on modern control theory it is necessary to be able to physically change some coordinates of CM nodes as control actions. Then we can describe the forced motion of the CM impulse process under external control as the difference equation:

$$\Delta \overline{Y}(k+1) = A\Delta \overline{Y}(k) + B\Delta \overline{U}(k), \qquad (3)$$

where $\Delta \overline{U}(k) = \overline{U}(k) - \overline{U}(k-1)$ — vector of controls increments with size $m \le n$. The operator fills the control matrix $B(n \times m)$ and in its simplest form uses ones and zeros.

If the CM has unmeasurable coordinates, they can be included into equation (3) as disturbances. In such a case the impulse process (3) will be written as

$$\Delta \overline{Y}(k+1) = A\Delta \overline{Y}(k) + B\Delta \overline{U}(k) + \Psi\Delta\overline{\xi}(k), \qquad (4)$$

where $\Delta\overline{\xi}(k) = \overline{\xi}(k) - \overline{\xi}(k-1)$ — vector of unmeasurable coordinates (disturbances).

PROBLEM STATEMENT

The first problem is to create a controlled dynamic model of CM impulse process describing multivariate demographic process in Ukraine. The second problem is to research and develop the system for suppressing constrained internal and external disturbances by means of control of the demographic security CM impulse process during martial law. The third problem is to implement an adaptive CM impulse process control under unknown or unmeasurable coefficients of the adjacency matrix *A*; this control should combine procedures of the matrix *A* elements' estimation during the transient process and usage of these estimates of the matrix $\hat{A}$ for a control vector design. The forth problem is to perform a simulation of the designed closed-loop control system and to research dynamic processes' quality with respect to ensuring demographic security of Ukraine under different disturbances during martial law.

CREATION OF A DEMOGRAPHIC SECURITY COGNITIVE MAP

Fig. 1 represents the schema of the CM of demographic security of Ukraine, developed based on cause and effect relations during martial law. The CM nodes have the following meaning:

0 — state support of families with children; 1 — average salary of a worker; 2 — consumer price index; 3 — export volume; 4 — import volume; 5 — population in Ukraine; 6 — real GDP of Ukraine; 7 — inflation rate; 8 — migration out of Ukraine; 9 — birth rate; 10 — unemployment rate; 11 — death rate; 12 — military events, spends on the war.

The following CM nodes coordinates can be varied as control actions:

- state support of families with children ($\Delta u_1(k)$);

- average salary of a worker ($\Delta u_2(k)$);
- export volume ($\Delta u_3(k)$);
- import volume ($\Delta u_4(k)$).

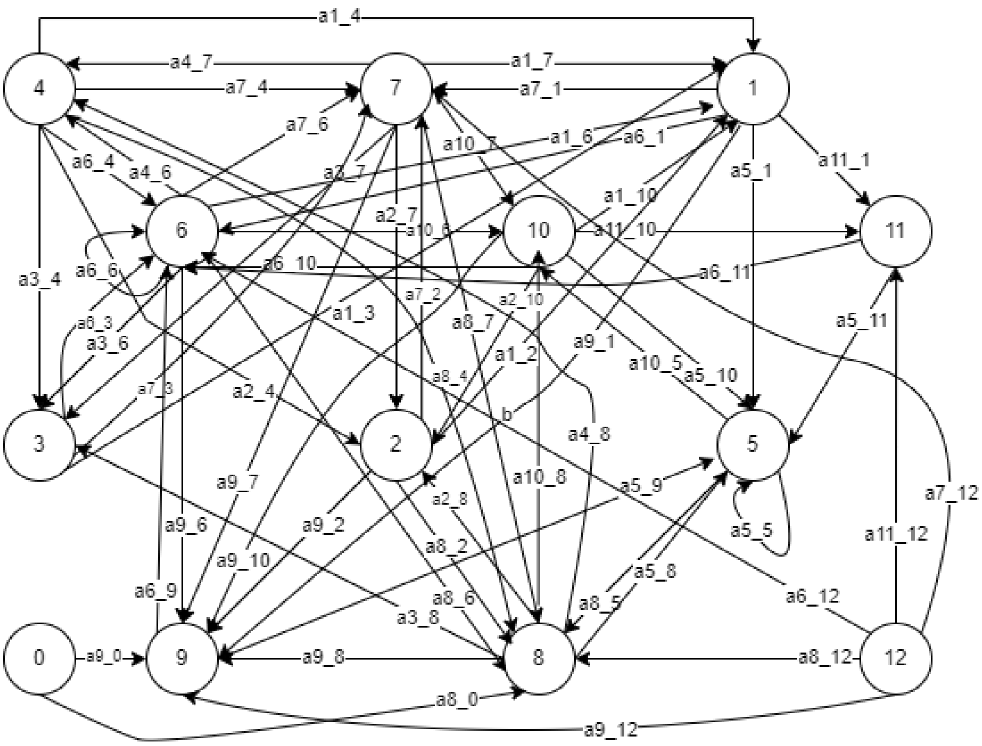


Fig. 1. Demographic security CM

Adjacency matrix A of the CM impulse process has the following form:

$$A=\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0.4 & 0.3 & -0.3 & 0 & 0.6 & 0.65 & 0 & 0 & -0.7 & 0 & 0\\
0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0.8 & 0.3 & 0 & -0.3 & 0 & 0\\
0 & 0 & 0 & 0 & -0.2 & 0 & 0.4 & 0 & -0.2 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -0.1 & -0.7 & 0.2 & 0 & 0 & 0 & 0\\
0 & 0.05 & 0 & 0 & 0 & 0.1 & 0 & 0 & -0.7 & 0.8 & -0.1 & -0.8 & 0\\
0 & 0.7 & 0 & 0.5 & -0.4 & 0 & 0.1 & 0 & 0 & 0.4 & -0.4 & -0.4 & -0.3\\
0 & 0.3 & 0.4 & -0.35 & 0.35 & 0 & -0.05 & 0 & 0 & 0 & 0 & 0 & 0.2\\
-0.3 & 0 & 0.3 & -0.1 & 0.15 & 0.1 & -0.3 & 0.2 & 0 & 0 & 0 & 0 & 0.5\\
0.7 & 0.15 & -0.2 & 0 & 0 & 0 & 0.2 & -0.15 & -0.35 & 0 & -0.2 & 0 & -0.5\\
0 & 0 & 0 & 0 & 0 & 0.1 & -0.45 & 0.5 & -0.15 & 0 & 0 & 0 & 0\\
0 & -0.1 & 0 & 0 & 0 & 0 & -0.2 & 0 & 0 & 0 & 0.3 & 0 & 0.5\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

Consider main disturbances affecting the demography, to implement control of the demographic security under martial law:

1. Mass migration of population out of Ukraine because of the threat to life under missile attacks on civilian targets in all regions and because of occupation of the territories.

2. High death rate because of military actions at the front and because of missile attacks over all territory of Ukraine.

3. Low birth rate because of young men population in the army, migration, unemployment rate increase and general uncertainty about the future affecting willingness of people to have children.

All these disturbances are practically impossible to describe mathematically using probabilistic indicators, specifically, to find their distributions, research their stationarity, analyse their fluctuations calculating their variance, find correlations etc. We can only set up limitations on their amplitude when describing the disturbances.

PROBLEM OF SUPPRESSING CONSTRAINED INTERNAL AND EXTERNAL DISTURBANCES DURING CONTROL OF DEMOGRAPHIC SECURITY COGNITIVE MAP IMPULSE PROCESS

Studies [3–5] present the theoretical foundations on the suppression of arbitrary constrained external disturbances in terms of invariant ellipsoids based on the design of a static state feedback, which minimizes the size of the invariant ellipsoid of the dynamical system. In this case, we implemented a robust control, where the analysis and synthesis problems are reduced to equivalent conditions in the form of linear matrix inequalities (LMI), solved numerically on the basis of semi-definite programming. In [5] we solve the problem of suppression of constrained external disturbances based on the invariant ellipsoids approach in the implementation of a closed-loop control system of impulse processes in CM of cryptocurrency on financial markets.

The general model of the dynamics of impulse processes (2) is decomposed into two interrelated systems of difference equations:

$$\Delta\overline{X}(k+1) = A_1\Delta\overline{X}(k) + D\Delta\overline{Z}(k); \tag{5}$$

$$\Delta\overline{Z}(k+1) = A_2\Delta\overline{Z}(k) + \Psi\Delta\overline{X}(k)\,. \tag{6}$$

Here $\overline{X}$ is the vector of measurable coordinates of CM nodes which are to be stabilized later; $\overline{Z}$ is the vector of CM coordinates considered as disturbances. The matrices A_1, D, A_2, Ψ are parts of the adjacency matrix of the initial model (2). The matrices D, Ψ show the relationships between the first (5) and the second (6) parts of the initial CM (2). The increments of coordinates $\Delta\overline{Z}(k)$ are taken into account as external constrained disturbances with unknown probabilistic characteristics in the first system of equations (5) of the CM model.

We designed a control vector to suppress constrained perturbations $\Delta\overline{Z}(k)$ by implementing static state controller in the feedback loop

$$\Delta\overline{U}(k) = -K_p\Delta\overline{X}(k)\,, \tag{7}$$

which acts directly on the measured nodes coordinates $\overline{X}$ of the first impulse process equations system (5) according to the controlled model:

$$\Delta\overline{X}(k+1) = A_1\Delta\overline{X}(k) + B\Delta\overline{U}(k) + D\Delta\overline{Z}(k) \tag{8}$$

The control is performed by changing the resources of the CM nodes, which are affected by the vector $\Delta\overline{U}(k)$.

In this paper, the change in the weight coefficients $\Delta A_1(k)$ with respect to the known estimated values of the matrix $\hat{A}_1$ is proposed to be considered as the internal perturbations in the CM impulse process model (5) of the demographic situation. For this purpose, in [4, 5] we modify the model (5) as follows:

$$\Delta\overline{X}(k+1) = A_1\Delta\overline{X}(k) + \Delta A_1\Delta\overline{X}(k) + D\Delta\overline{Z}(k)\,, \tag{9}$$

where $\Delta A_1 = A_1 - A_{1_{\text{var}}}(k)$ is the change in the adjacency matrix of CM (5) during the sampling period, $A_{1_{\text{var}}}(k)$ is the real unknown value of the matrix A_1, which changes as the demographic system evolves.

Let us denote the increment of internal perturbations in (9) as $\Delta A_1(k)\Delta\overline{y}(k) = \Delta\overline{w}(k)$. Then the equation of the uncontrolled impulse process (9) will be written as:

$$\Delta\overline{X}(k+1) = A_1\Delta\overline{X}(k) + (I_1 \quad D)\begin{bmatrix}\Delta\overline{w}(k)\\ \Delta\overline{Z}(k)\end{bmatrix}, \tag{10}$$

where the vectors and matrices have the following dimensions: $\dim\Delta\overline{X} = n$; $\dim\Delta\overline{Z} = p$; $\dim\Delta\overline{w} = n$; $A_1(n\times n)$; $D(n\times p)$, I is a unit matrix of dimension $n\times n$. We assume that the internal and external perturbations are jointly constrained by the norm l_∞, so that:

$$\left\|\begin{bmatrix}\Delta\overline{w}(k)\\ \Delta\overline{Z}(k)\end{bmatrix}\right\|_\infty = \sup\left\{[\Delta\overline{w}^T(k) \quad \Delta\overline{Z}^T(k)]\begin{bmatrix}\Delta\overline{w}(k)\\ \Delta\overline{Z}(k)\end{bmatrix}\right\}^{1/2} \le 1\,. \tag{11}$$

In [3] invariant ellipsoids on state variables are proposed to describe the characteristic of the effect of disturbances of the type (11) on the trajectory of a dynamic discrete system (10). For the CM they take the form:

$$\varepsilon_{\Delta\overline{X}} = \{\Delta\overline{X}(k)\in R^n : \Delta\overline{X}^T P^{-1}\Delta\overline{X} \le 1\},\ P>0\,, \tag{12}$$

if from $\Delta\overline{X}(0)\in\varepsilon_{\Delta\overline{X}}$ the condition $\Delta\overline{X}(k)\in\varepsilon_{\Delta\overline{X}}$ follows for all discrete moments of time $k = 1, 2, 3,\ldots$. Then the matrix P is called the matrix of the ellipsoid $\varepsilon_{\Delta\overline{X}}$.

In [4; 5] the condition of invariance of the ellipsoid (12) under disturbances (11) is proven. According to it, invariance is guaranteed when the following LMI is met:

$$\frac{1}{\alpha}A_1PA_1^T - P + \frac{I_1 + DD^T}{(1-\alpha)} \le 0\,. \tag{13}$$

ALGORITHM FOR A STATE CONTROLLER DESIGN FOR THE COGNITIVE MAP IMPULSE PROCESS

The state equation of the controlled CM impulse process (10) under additional internal disturbances $\Delta w(k)$ takes the form:

$$\Delta\overline{X}(k+1) = A_1\Delta\overline{X}(k) + B\Delta\overline{U}(k) + \left(I_1 \quad D\right)\begin{bmatrix}\Delta\overline{w}(k)\\ \Delta\overline{Z}(k)\end{bmatrix}. \tag{14}$$

When the state controller (7) is applied, the equation of the closed-loop CM impulse process control system is written as follows:

$$\Delta\overline{X}(k+1)=(A_1-BK_p)\Delta\overline{X}(k)+(I_1 \quad D)\begin{bmatrix}\Delta\overline{w}(k)\\ \Delta\overline{Z}(k)\end{bmatrix}. \qquad (15)$$

It is assumed that the pair $(A_1 \quad B)$ in the model (14) is controllable. Then the LMI (13) for the closed-loop system looks like:

$$\frac{1}{\alpha}(A_1-BK_p)P(A_1-BK_p)^T-P+\frac{I_1+DD^T}{(1-\alpha)}\le 0. \qquad (16)$$

We consider the minimization of the trace of the ellipsoid matrix (12) as the optimality criterion for the design of the controller (7):

$$trP(\alpha)\to\min,\ \alpha^*\le\alpha<1, \qquad (17)$$

This ensures minimization of the size of the invariant ellipsoid (12) with the largest suppression of disturbances $\begin{bmatrix}\Delta\overline{w}(k)\\ \Delta\overline{Z}(k)\end{bmatrix}$, which are constrained only by the maximum range (11). After multiplying the factors in the inequality (16), we obtain:

$$\frac{1}{\alpha}(A_1PA_1{}^{\mathrm T}-BK_pPA_1{}^{\mathrm T}-A_1PK_p{}^{\mathrm T}B^{\mathrm T}+BK_pPK_p{}^{\mathrm T}B^{\mathrm T})-P+\frac{I_1+DD^{\mathrm T}}{(1-\alpha)}\le 0. \qquad (18)$$

Inequality (18) is nonlinear with respect to P and K_p, which need to be optimized. In [3] a replacement $L=K_pP$ and introduction of an additional constraint is done:

$$\begin{bmatrix}R & L\\ L^T & P\end{bmatrix}\ge 0, \qquad (19)$$

where $R=R^T$. This inequality is equivalent to $R\ge LP^{-1}L^T=K_pPK_p{}^T$ according to the Schur's formula at $P>0$. Then to meet inequality (18) it is sufficient that:

$$\frac{1}{\alpha}(A_1PA_1{}^T-BLA_1{}^T-A_1L^TB^T+BRB^T)-P+\frac{I_1+DD^T}{(1-\alpha)}\le 0. \qquad (20)$$

Minimization of criterion (17) under constraints (19), (20) is performed with respect to variables P, L, R using semi-definite programming method by using Matlab-based SeDuMi Toolbox. Then the matrix $\hat{K}_p$ of the optimal state controller (7) is defined as:

$$\hat{K}_p=\hat{L}\hat{P}^{-1} \qquad (21)$$

with the estimated values of $\hat{\alpha},\hat{P},\hat{L},\hat{R}$, providing minimization of criterion (17) under constraints (19), (20).

PROBLEM OF THE COGNITIVE MAP WEIGHTS IDENTIFICATION BASED ON THE RECURRENT LEAST SQUARES METHOD

The model of the controlled CM impulse process (3) of the "input-output" type can be represented as:

$$(I-A_1q^{-1})\Delta\overline{Y}(k)=Bq^{-1}\Delta\overline{U}(k). \qquad (22)$$

Weighting coefficients of the adjacency matrix A are usually determined by applying expert estimates based on cause-and-effect relations. In the process of evolving of the demographic situation, these coefficients in the model (22) will change over time, depending on changes in the influence of the CM nodes on each other. So the problem of adaptive control of the CM implulse process appears, when both estimation of the parameters (coefficients) of the adjacency matrix A_1 and design of the control vector $\overline{U}(k)$ must be performed simultaneously.

Let us describe the equation (22) coordinate-wise (for each CM node):

$$\Delta y_i(k) = \sum_{j=1}^{n} a_{ij} \Delta y_j(k-1) + b_i \Delta u_i(k-1) + \xi_i(k). \tag{23}$$

It is assumed that the disturbances $\xi_i(k)$, caused by inaccurate measuring of the CM nodes coordinates and inaccurate knowledge of the model coefficients, are white noise. This assumption is plausible because $y_i(k), u_i(k)$ in model (23) are presented in the form of the first differences, i.e. increments. It should also be taken into account that the structure of the matrix A_1 is known and some of the coefficients a_{ij} are obviously equal to zero (in those cases when there are no connections between the corresponding CM nodes).

Let us write model (23) as follows:

$$\Delta y_i(k) - b_i \Delta u_i(k-1) = \overline{X}_i^T(k)\overline{\Theta}_i + \xi_i(k), \tag{24}$$

where $\overline{\Theta}_i = [a_{ij_1} \ldots a_{ijP_1}]^{\mathrm{T}}$ consists of the non-zero coefficients in the *i*-th row of matrix A_1, $\overline{X}_i^T(k) = [\Delta Y_{j1}(k-1), \ldots, \Delta Y_{jP_i}(k-1)]$ is a vector of measured CM nodes coordinates.

The current estimate of the vector $\overline{\Theta}_i$ is denoted by $\hat{\overline{\Theta}}_i(k)$. To estimate the weight coefficients of the matrix A_1 we apply the recurrent least squares method [6–9]:

$$\hat{\overline{\Theta}}_i(k) = \hat{\overline{\Theta}}_i(k-1) + K_i(k)(\Delta y_i(k) - b_i \Delta u_i(k-1) - \overline{X}_i^T(k)\hat{\overline{\Theta}}_i(k-1));$$

$$K_i(k) = \frac{1}{1 + \overline{X}_i^T(k) P_i(k-1) \overline{X}_i(k)} P_i(k-1) \overline{X}_i(k) \overline{X}_i^T(k) P_i(k-1); \tag{25}$$

$$P_i(k) = P_i(k-1) - \frac{1}{1 + \overline{X}_i^T(k) P_i(k-1) \overline{X}_i(k)} P_i(k-1) \overline{X}_i(k) \overline{X}_i^T(k) P_i(k-1).$$

The recurrent procedure (25) should be performed for each CM node $\Delta y_i(k)$, $i = 1, 2, \ldots, n$ at each sampling period. We use the obtained estimates $\hat{\overline{\Theta}}_i(k)$ as the coefficients values of the adjacency matrix A_1 at the current sampling period in the control algorithm (7), (20), (21). For parametric identification of the adjacency matrix A_1, we can also apply non-recurrent identification methods outlined in [10].

EXPERIMENTAL RESEARCH OF THE SYSTEM SUPPRESSING CONSTRAINED INTERNAL AND EXTERNAL DISTURBANCES DURING CONTROL OF THE DEMOGRAPHIC SECURITY COGNITIVE MAP IMPULSE PROCESS

To ensure demographic security in Ukraine it is reasonable to stabilize the following coordinates $\overline{X}$ of the CM on Fig. 1: 0, 1, 2, 3, 4, 5, 6, 9. The following coordinates $\overline{Z}$ are considered as disturbances affecting the demographic security: 7, 8, 10, 11, 12. After decomposition of the model (2) into models (5) and (6) we conclude that although model (2) is unstable, state equations (5) and (6) are stable. Control actions $\Delta u_1(k)$, $\Delta u_2(k)$, $\Delta u_3(k)$, $\Delta u_4(k)$ are fed to the nodes 0, 1, 3, 4 respectively. So matrix B in the controlled impulse process equation (8) is the following:

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}^{\mathrm{T}}$$

During simulation of closed-loop control system dynamics of the CM impulse process based on the proposed method, we applied step impulse with unit amplitude as an external disturbance at the initial time moment fed to the node (12) — military events, spends on the war. Internal disturbances are generated as following: at each sampling period non-zero coefficients of the matrix A_1 are varied under the formula $A_{1\,\mathrm{var}}(k) = A_1 \xi(k)$, where $\xi(k)$ is a normal random variable (Gaussian white noise) for the control only values of A_1 are used while $A_{1\,\mathrm{var}}$ are applied as unknown internal disturbance. Initial levels of all the CM nodes coordinates are taken equal to zero for simplicity.

Fig. 2 shows the transient processes of the CM nodes coordinates, Fig. 3 — their increments. Here solid lines denote transient process under control, dashed lines — without control. Fig. 4 shows control actions changes.

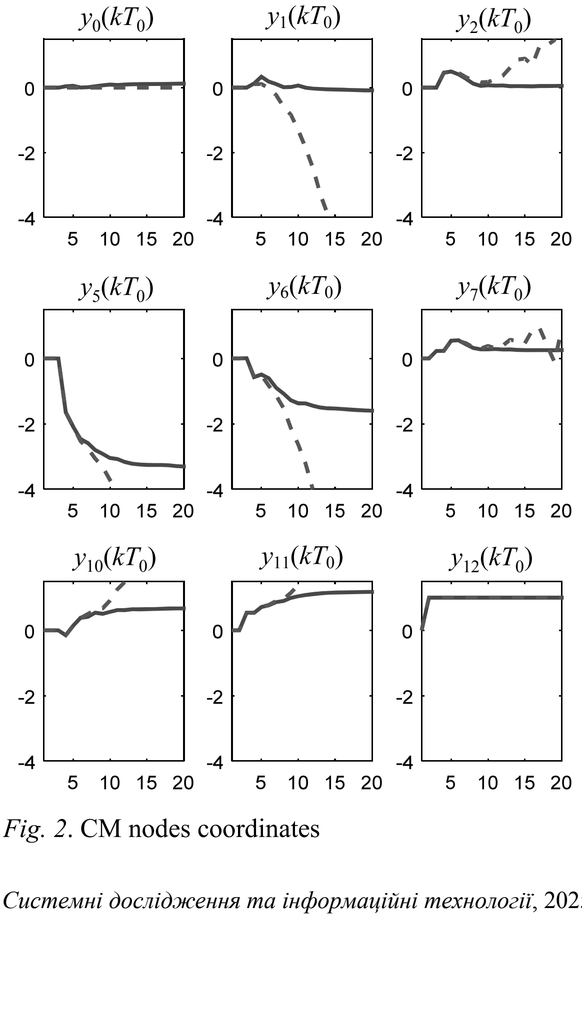


Fig. 2. CM nodes coordinates

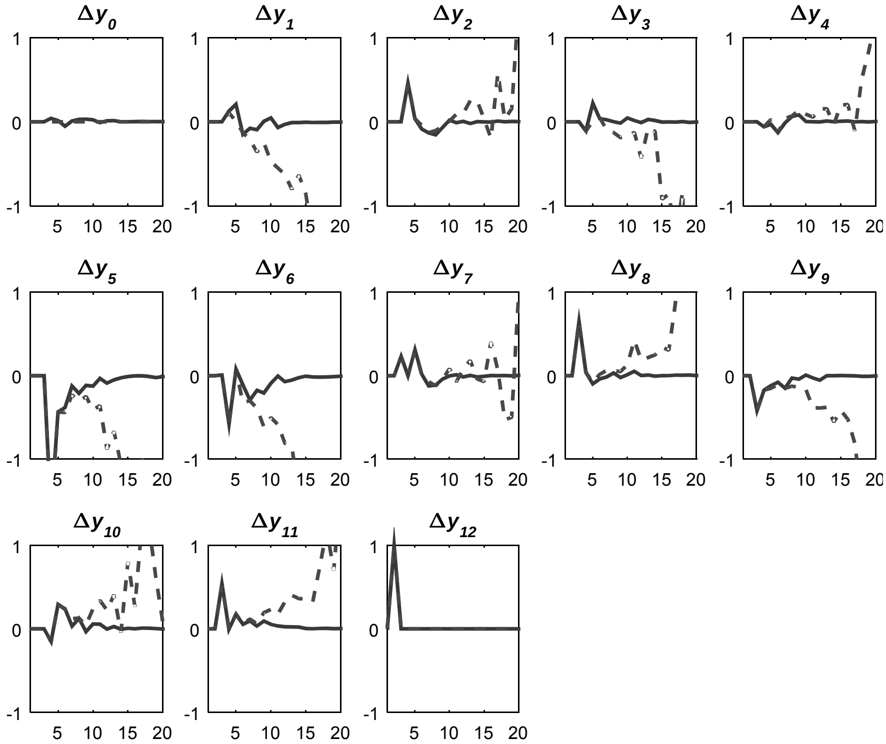


Fig. 3. CM nodes coordinates increments

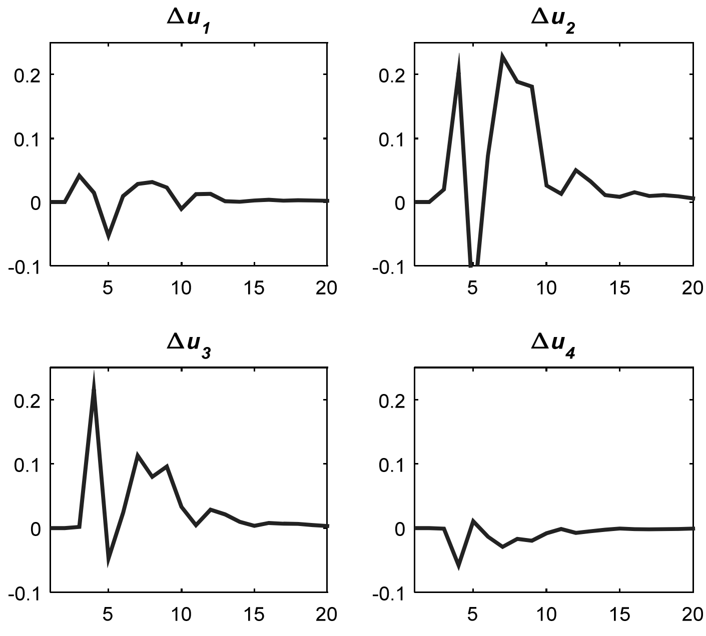


Fig. 4. Control actions

CONCLUSIONS

The paper considers important problem of demographic security under martial law in Ukraine. Possible approach to solve this problem was suggested based on the CM impulse processes modelling and control. Specifically, the CM of demographic security was created and the control method for suppressing disturbances based on invariant ellipsoids was applied. As a result, the control system was designed and the simulation was performed.

Based on the simulation results, we can conclude that the suggested approach will help to stabilize very dangerous and unstable demographic process initiated by increase of the military spends and the military events intensity. Without control this process leads to the catastrophic depopulation of Ukraine. Under the suggested control, the simulation demonstrates that despite significant decrease of the population at the beginning, we are able to stabilize it at some level and stop this process. Main control actions the government should apply are: export increase and import decrease, average salary increase and support of the families with children. The latter two of them are necessary to increase birth rate and decrease migration, while the former two actions are necessary to prevent inflation and stabilize economy.

REFERENCES

1. F. Roberts, *Discrete Mathematical Models with Applications to Social, Biological and Environmental Problems*. Prentice-Hall, Englewood Cliffs, 1976.
2. V.D. Romanenko, Yu.L. Milyavskiy, "Ensuring the sustainability of pulse processes in cognitive maps on the basis of the models in the states space," *System Research and Information Technologies*, no. 1, pp. 26–42, 2014.
3. V.D. Romanenko, Yu.L. Milyavskiy, "Control automation of impulse processes in cognitive maps with constrained disturbance suppression based on invariant ellipsoids method," *System Research and Information Technologies*, no. 2, pp. 29–39, 2017. doi: 10.20535/SRIT.2308-8893.2017.2.03
4. V.D. Romanenko, Yu.L. Milyavskiy, "Suppressing constrained internal and external disturbances for impulse processes control in cognitive maps of complex systems," *System Research and Information Technologies*, no. 4, pp. 104–114, 2018. doi: 10.20535/SRIT.2308-8893.2018.4.09
5. V. Romanenko, Yu. Milyavskiy, H. Kantsedal, "Automated Control Problem for Dynamic Processes Applied to Cryptocurrency in Financial Markets," in *Recent Developments in Automatic Control Systems*; Eds: Yuriy P. Kondratenko, Vsevolod M. Kuntsevich, Arkadii A. Chikrii, Vyacheslav F. Gubarev. Denmark: River Publisher, 2022, pp. 417–444.
6. R. Isermann, *Digital Control Systems*. Springer-Verlag, Berlin, Heidelberg, 1981.
7. K.J. Astrom, B. Wittenmark, *Computer Controlled Systems*. Prentice-Hall, Inc., Englewood Cliffs, 1984.
8. L. Ljung, *System Identification: Theory for the User*. Prentice Hall PTR, 1999.
9. D.W. Clarke, P.J. Gawthrop, "Self-tuning Control," *Proc. IEEE*, 126, pp. 633–640, 1979. doi: https://doi.org/10.1049/piee.1979.0145
10. V.F. Gubarev, *Modeling and Identification of Complex Systems*. Kyiv: Naukova dumka, 2019.

Received 10.12.2024

INFORMATION ON THE ARTICLE

Viktor D. Romanenko, ORCID: 0000-0002-6222-3336, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: romanenko.viktorroman@gmail.com

Yurii L. Miliavskyi, ORCID: 0000-0003-0882-3418, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: yuriy.milyavsky@gmail.com

ЗАДАЧІ АВТОМАТИЗОВАНОГО КЕРУВАННЯ ДИНАМІЧНИМИ ПРОЦЕСАМИ СИСТЕМНОГО ЗАБЕЗПЕЧЕННЯ БЕЗПЕКИ УКРАЇНИ НА ОСНОВІ МОДЕЛЕЙ ІМПУЛЬСНИХ ПРОЦЕСІВ У КОГНІТИВНИХ КАРТАХ. ЧАСТИНА 1. ЗАБЕЗПЕЧЕННЯ ДЕМОГРАФІЧНОЇ БЕЗПЕКИ / В.Д. Романенко, Ю.Л. Мілявський

Анотація. Наведено когнітивну карту (КК) демографічної безпеки, на основі якої описано динамічну модель імпульсних процесів КК у вигляді системи різницевих рівнянь (рівнянь Робертса). Виконано вибір зовнішнього вектора керувальних дій імпульсним процесом КК, який реалізується шляхом варіювання координат вершин КК. Реалізовано замкнену систему керування імпульсним процесом КК, до складу якої входить синтезований на основі методів теорії автоматичного керування багатовимірний дискретний регулятор, який формує вибрані керувальні дії. Розв’язано задачу проєктування дискретного регулятора для автоматизованого керування динамічними процесами для забезпечення демографічної безпеки. Функція регулятора полягає у приглушенні зовнішніх та внутрішніх збурень при керуванні імпульсними процесами КК на основі методу інваріантних еліпсоїдів. Наведено алгоритм ідентифікації вагових коефіцієнтів КК на основі рекурентного методу найменших квадратів. Подано результати дослідження якості динамічних процесів стосовно забезпечення демографічної безпеки в Україні в разі дії різноманітних збурень в умовах воєнного стану.

Ключові слова: когнітивна карта, демографічна безпека, інваріантний еліпсоїд, лінійні матричні нерівності, імпульсний процес.

UDC 004.8:336:796.3:519.83
DOI: 10.20535/SRIT.2308-8893.2025.3.07

ANALYSIS AND FORECASTING OF THE FINANCIAL BENEFIT FOR THE TENNIS MATCH OUTCOMES BY MACHINE LEARNING METHODS

K. SHUM, N. KUZNIETSOVA

Abstract. Tennis is one of the most popular sports in the world, attracting considerable attention from casual fans and professional analysts. The application of machine learning methods enables the accurate prediction of match results, opening up opportunities for profit through betting on likely winners. This study evaluates the financial benefits of predicting tennis match outcomes by identifying an effective sports betting strategy. The study examines various machine learning methods and auxiliary algorithms, comparing them to select the best betting strategy for maximizing the user's potential profit. In the paper, the method and algorithm for determining effective sports betting strategies were developed. This algorithm and method were tested on tennis game datasets (for both women and men), and the best tennis betting strategy was identified. As part of the study, a software product has been developed to predict the outcomes of tennis matches.

Keywords: forecasting, machine learning, betting strategies, financial benefit.

INTRODUCTION

Tennis is a dynamic and unpredictable game, combining many factors that influence the course of events during matches: players' physical conditions, psychological state, chosen tactics, anthropometry, weather conditions, and more. Each of these aspects can be decisive in achieving the desired outcome. Thanks to this versatility, tennis ranks among the most popular sports globally, captivating a broad audience of fans, from casual spectators who enjoy the thrill of the game to professional sports analysts who study the game from a scientific perspective.

Match outcome prediction holds a special place among the various aspects of sports interest. As for the standard fan, a match result is typically a topic of discussion and emotional enjoyment. However, the prediction is practical for analysts and professional bettors who place wagers on sports events [1]. Knowing the likelihood of a player's victory not only allows for more informed betting to secure financial gain but also aids in developing strategies for long-term success. In this context, the betting process goes beyond simple gambling for many professional participants in the sports betting market. Predicting tennis match outcomes becomes a critical tool for making informed decisions, assessing risks, and evaluating potential benefits, ultimately supporting the financial growth of the bettor. It also gives us the possibility to solve such tasks as understanding behavior and forecasting the gamer's outflow [2]. The players who win are motivated to stay longer in the game while they understand the game's process and can also plan their own strategy and evaluate their financial benefits.

PROBLEM STATEMENT

This research was conducted to deeply understand the betting and gambling processes by applying modern techniques and approaches. It was first decided to try machine learning algorithms for evaluating and forecasting games' outcomes and for finding hidden dependencies. Then, based on these models, we can find the most important variables that could be interpreted as some key factors for winning on some side. It means that we can also take into account some preliminary information before making a bet. Next, the strategy of effective betting should be defined. For this reason, we will develop the algorithm for defining the most effective strategy that can be used by gamblers to maximize their profit as a result of the sports betting process.

MACHINE LEARNING METHODS AND AUXILIARY ALGORITHMS FOR THE GAME OUTCOMES PREDICTION

Machine learning methods

Machine learning is currently a powerful tool for solving various tasks across different fields of human activity. The significant potential and efficiency of machine learning methods and algorithms make this technology crucial in areas where traditional approaches may fall short. Predicting the outcomes of sports events is no exception. In this study, predictive models have been developed to predict the results of men's (hereafter, M) and women's (hereafter, W) tennis matches. These models are based on logistic regression (M and W), multilayer perceptron (W), random forest (M), and extreme gradient boosting (M).

Logistic regression is a method that models the relationship between a categorical target variable and a set of independent predictor variables. Although logistic regression is a classification algorithm, it is based on a linear regression model [3; 4]. To produce categorical outcomes, it transforms the continuous output of linear regression into a range between 0 and 1 (interpreted as the probability of belonging to a specific class) using the logistic function, also known as the sigmoid function [3], which can be described by the following formula:

$$\sigma(z) = \frac{1}{1+e^{-z}} \ ,$$

where $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$ is a linear combination of independent variables x_i and their coefficients β_i, $i = \overline{1,n}$, n is a number of predictor variables and β_0 is the intercept term.

Then the probability of an object's belonging to a specific class can be represented as:

$$P(y=1) = \sigma(z); \quad P(y=0) = 1 - \sigma(z) .$$

Random forest is an ensemble machine learning method that combines the predictions of multiple decision trees to improve the model's accuracy and stability. The trees are constructed on random subsets of data from the training set, and random subsets of features are used to reduce the correlation between the trees. The final prediction $\hat{y}$ is determined by majority voting, making the method robust to overfitting and effective for various tasks [5], and is determined by the following formula:

$$\hat{y} = Mode\{h_i(x)\}, \quad i = \underline{1,n} \ ,$$

where $h_i(x)$ is a prediction of the i-th tree, and n is a number of trees in the forest.

Extreme Gradient Boosting (XGBoost) is an ensemble machine learning method that implements gradient boosting with decision trees. XGBoost uses an iterative approach, where the key idea is to build an ensemble of decision trees, with each subsequent tree sequentially correcting the errors of the previous ones, thereby improving the model's overall accuracy [6]. If the prediction for the i-th sample after $k-1$ iterations is represented as $\hat{y}_i^{(k-1)}$, then at the k-th iteration, the prediction value will be updated using the following formula:

$$\hat{y}_i^{k)} = \hat{y}_i^{(k-1)} + \eta h_k(x_i),$$

where η is the learningn rate, which determines how strongly each tree influences the final prediction.

A *multilayer perceptron* (MLP) is a type of artificial neural network consisting of several layers: an input layer, one or more hidden layers, and an output layer. Each layer contains neurons that take a weighted sum of input data from the previous layer, apply an activation function to it, and pass the result to the next layer. Weighted sum is counted using the following formula:

$$z_j^{(l)} = \sum_{i=1}^{n} w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)},$$

where $z_j^{(l)}$ is the activation of neuron j in layer l; $w_{ij}^{(l)}$ is the weight connecting neuron i in the previous layer to the neuron j; $a_i^{(l-1)}$ is the activation of neuron i in the previous layer; $b_j^{(l)}$ is the bias of neuron j.

The training process is repeated over several iterations until the model converges to an optimal solution. Due to the architecture, MLP can model complex nonlinear relationships between data [7].

Auxiliary algorithms

The *time discounting method* is an approach that assigns greater significance to newer data and less to older data [8]. The idea is to apply weights relative to the time between events. In the study, this method is used to predict player statistics in the men's division, as the prediction of match performance is based on the values of relevant statistical variables. Weights are applied using an exponential function $W(t)$:

$$W(t) = \min(f^t, \ f),$$

where t represents the time in months between the scheduled match and a previously played match, and f is the discount factor, which can range from 0 to 1. The discount factor determines the extent of time discounting and is set by the researcher. The smaller the value of f, the lower the weight given to older matches.

The *data filtering algorithm* is the process of selecting subsets from a large dataset based on specified criteria. Initially, the selection conditions are defined (for example, these may be the values of variables), after which the corresponding samples are formed, which allows the efficient extraction of the most relevant data for further analysis and use.

BUILDING THE MACHINE LEARNING MODELS

For this study, we decided to use real data and develop our models for the prediction of tennis match outcomes using Python, along with relevant machine learning and data processing libraries (Sklearn, Pandas, Numpy, etc.). For this reason, we used two different datasets for men's and women's games. The dataset from user JeffSackmann's GitHub repository [9] was used to predict men's matches, and the dataset from the Tennis-data website [10] was utilized for women's matches. In both cases, the records began at the start of 2010, with a total of 153.959 and 37.731 games recorded, respectively.

For both datasets, initial preprocessing was carried out: the properties and specifics of each variable were analyzed, missing values were handled, irrelevant records and variables were removed, and the data was transformed into a format suitable for future models. Each dataset was duplicated, and the corresponding player columns were swapped to balance the number of positive and negative classes (1 for the first player's victory, 0 for loss). As a result, the final training datasets contained 209.116 and 47.816 records for men and women, respectively.

The most important features for prediction by using statistical methods were selected, such as:

- **For men:** twenty significant predictor variables were selected, including tournament seeding numbers, differences in height, ranking, ranking points, as well as percentage differences in various statistical indicators (e.g., first serve percentage, percentage of points won on return, etc.).
- **For women:** five significant predictor variables were selected, including differences in ranking points, age, and differences in the win odds set by the Pinnacle bookmaker and between maximum and average odds from other bookmakers.

The next stage involves constructing machine learning models using the methods mentioned earlier. To determine their best parameters, the grid search algorithm was applied. This algorithm selects the combination of the most effective features from a given set that ensures the highest model performance. The results obtained are presented in Tables 1–5.

Table 1. Parameters of the logistic regression model (women)

Parameter	Description	Value
test_size	The proportion of the dataset that is used for testing the model	0.1
solver	The method for determining the optimal model weights that minimize the loss function	liblinear
fit_intercept	The presence of a bias term in the model equation	False
C	Regularization strength	4.25
penalty	The type of regularization used to control the model's overfitting	L2

Table 2. Parameters of the multilayer perceptron model (women)

Parameter	Description	Value
test_size	The proportion of the dataset that is used for testing the model	0.1
solver	The method for determining the optimal model weights that minimize the loss function	lbfgs
activation	Activation function of the hidden layer	relu
alpha	L2-regularization strength	0.005
hidden_layer_ sizes	Number of neurons in the hidden layers	(100,)
learning_rate	Learning rate for weight updates	constant

Table 3. Parameters of the logistic regression model (men)

Parameter	Description	Value
test_size	The proportion of the dataset that is used for testing the model	0.1
solver	The method for determining the optimal model weights that minimize the loss function	newton-cg
fit_intercept	The presence of a bias term in the model equation	True
C	Regularization strength	1
penalty	The type of regularization used to control the model's overfitting	None

Table 4. Parameters of the random forest model (men)

Parameter	Description	Value
test_size	The proportion of the dataset that is used for testing the model	0.15
n_estimators	The number of trees in the forest	100
criterion	The function to measure the quality of a split	log_loss
max_features	The number of features to consider when looking for the best split	None
min_samples_leaf	The minimum number of samples required to be at a leaf node	2

Table 5. Parameters of the XGBoost model (men)

Parameter	Description	Value
test_size	The proportion of the dataset that is used for testing the model	0.15
n_estimators	The number of trees (iterations) of the model	100
learning_rate	Learning rate	0.05
max_depth	The maximum depth of each tree	6
reg_alpha	L1-regularization parameter	0.5
reg_lambda	L2-regularization parameter	1.5

After building the models with the specified parameters, their performance was evaluated on the training and validation datasets using standard classification quality metrics. The results are presented in Tables 6–10.

Table 6. Evaluation of the logistic regression model (women)

Sample	Quality metric					
	Accuracy	Precision	Recall	F1 Score	Roc Auc	Loss
Training	0.68966	0.68897	0.70124	0.69496	0.76102	0.57987
Validation	0.68341	0.68355	0.68202	0.68277	0.75474	0.58704

Table 7. Evaluation of the multilayer perceptron model (women)

Sample	Quality metric					
	Accuracy	Precision	Recall	F1 Score	Roc Auc	Loss
Training	0.69175	0.68903	0.70954	0.69894	0.7599	0.5803
Validation	0.68290	0.68239	0.68323	0.68274	0.7535	0.58745

T a b l e 8 . Evaluation of the logistic regression model (men)

Sample	**Quality metric**					
	Accuracy	Precision	Recall	F1 Score	Roc Auc	Loss
Training	0.98278	0.98292	0.98253	0.98272	0.99790	0.05599
Validation	0.98145	0.98146	0.98144	0.98145	0.99789	0.05618

T a b l e 9 . Evaluation of the random forest model (men)

Sample	**Quality metric**					
	Accuracy	Precision	Recall	F1 Score	Roc Auc	Loss
Training	0.98345	0.98345	0.98376	0.98381	0.99763	0.07793
Validation	0.98363	0.98362	0.98366	0.98349	0.99789	0.07484

T a b l e 10 . Evaluation of the XGBoost model (men)

Sample	**Quality metric**					
	Accuracy	Precision	Recall	F1 Score	Roc Auc	Loss
Training	0.98396	0.98501	0.98293	0.98396	0.99834	0.04733
Validation	0.98359	0.98267	0.98455	0.98361	0.99842	0.04601

After analyzing the results obtained, it can be concluded that the models predicting women's matches perform at an acceptable level but are slightly worse than those predicting men's matches. The models for men's tennis show excellent values across all metrics. However, it is important to note that their performance may decline due to the necessity of applying the time discounting method to predict statistics for future matches. None of the developed models exhibit signs of overfitting, as the quality metrics for both the training and validation datasets are very close.

To facilitate the process of predicting matches and to provide a straightforward interpretation of the results, a web interface was developed to allow users to interact with the developed models easily, input the necessary data for predictions via the keyboard, and modify it if needed. Separate prediction pages for men and women were implemented, with their interfaces shown in Figs. 1 and 2. Additionally, a database containing historical match records for men was created.

Fig. 1. Prediction page for men's match outcomes

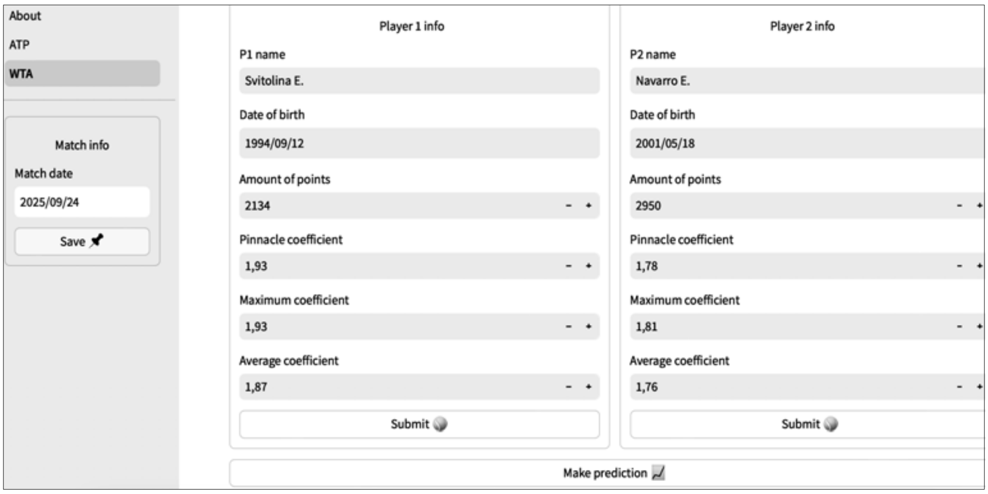


Fig. 2. Prediction page for women's match outcomes

ALGORITHM FOR DETERMINING THE BEST BETTING STRATEGIES

To determine an effective betting strategy, a method based on the ROI (return on investment) metric as the prior indicator of a bettor's success and, accordingly, the target metric of the built predictive model's effectiveness was developed and applied. An important condition is that each bet must be evenly distributed with an identical amount.

Let S_0 represent the bettor's (player's) initial capital.

$$S_0 = s\,t,$$

where s is the amount of a single bet (always the same), and t is the bettor's tolerance for losses, i.e., the number of consecutive bets he is willing to lose before ceasing to follow the strategy. The tolerance is determined by the bettor and can be adjusted during the betting process.

Then, S_i is the player's current capital after the i-th bet has been placed.

$$S_i = S_{i-1} + \mathrm{I}_i\, s\, coef_i - s\ ,$$

where $coef_i$ is the coefficient of the i-th bet, $i = \overline{1,n}$, and n is the total number of bets, while I_i is the indicator of the success of the i-th bet, which is determined as follows:

$$\mathrm{I}_i = \begin{cases} 1, & \text{if the bet won} \\ 0, & \text{otherwise} \end{cases}.$$

Let P_i be the player's profit after calculating the i-th bet:

$$P_i = S_i - S_0\,.$$

Let ROI_i be the percentage of winnings from each bet, calculated after the i-th bet, averaged over the distance:

$$ROI_i = \frac{P_i 100}{s\,i},\ i = \overline{1,n}\,.$$

Then, the betting strategy is considered effective if the following conditions are met:

1. $S_i > s,\ i = \overline{1,n}$. This condition means that the player's current capital must always be greater than the amount of one bet to be able to place it.

2. $ROI_i > 0$ при $i = \underline{h,n}$. Here, n is the total number of bets placed, and h is the minimum number of bets determined as the calculation threshold for profit, which can be adjusted by the player. This condition means that after the h-th bet, the return on investment (ROI) must always be greater than 0, demonstrating the strategy's stability and profitability over the long term.

3. $\sum_{j=0}^{t-1} 1(ROI_{k+j} < ROI_{k+j-1}) < t$,

where $k = \underline{1,n-t+1}$ for $n > t-1$, $ROI_0 = 0$, and $1(\cdot)$ is an indicator defined as follows:

$$1(\cdot) = \{1, \text{ if } ROI_{k+j} < ROI_{k+j-1} \, 0, \text{ otherwise} .$$

This condition means that the player is willing to tolerate no more than t consecutive lost bets (tolerance for loss).

The algorithm for determining an effective strategy using the described method is presented in the form of a flowchart in Fig. 3.

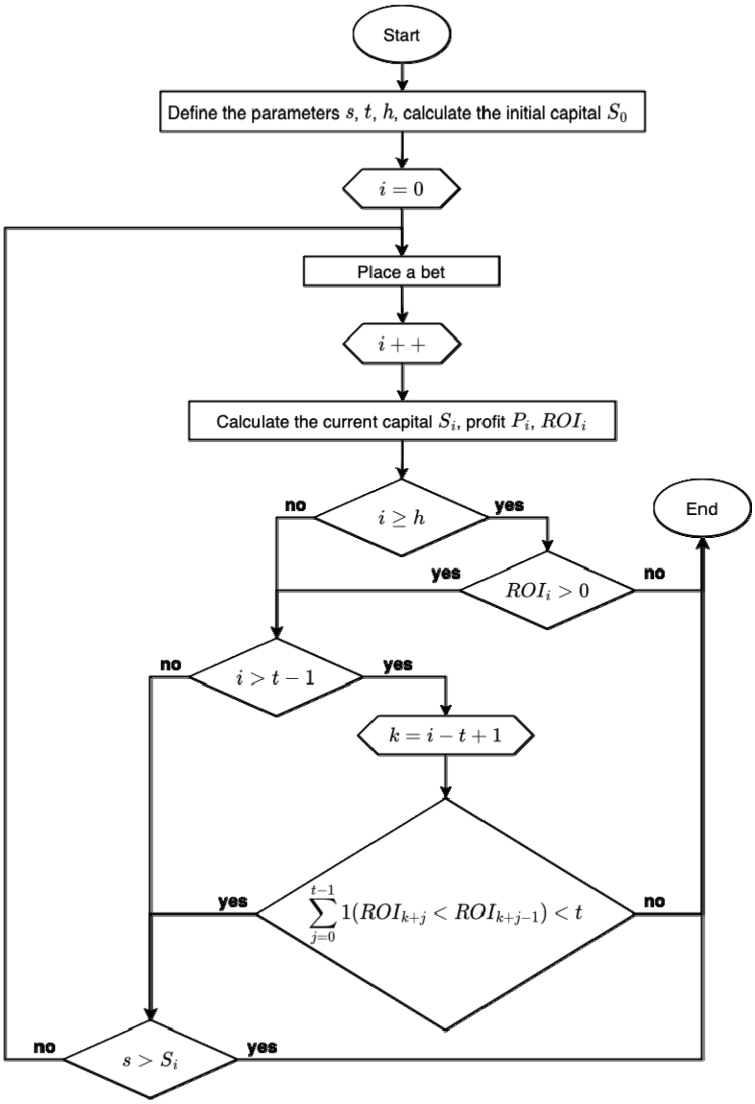


Fig. 3. Flowchart of the algorithm for determining the effectiveness of a betting strategy

DETERMINING THE BEST BETTING STRATEGIES BY DEVELOPED ALGORITHM

New datasets of predicted matches were created using the models and algorithms previously developed to determine potentially successful betting strategies. A sample of 239 predicted matches for women and 346 for men was compiled. The prediction of players' statistics for men was performed in two algorithm variations: for all court surfaces and only for a selected surface. As a result, six predictions were made for each men's match.

A filtering process was applied to the obtained data to exclude specific categories of games, thereby increasing the scope for identifying the most suitable conditions for profitability. For women, filters were considered for the minimum probability of a player's victory and the minimum odds. For men, filters included the current form (i.e., the number of matches played in the last 60 days) and the minimum odds. Tables 11–12 present the number of successful (profitable) strategies for each model.

Table 11. Successful strategies (women)

Model	Number of profitable strategies	Total number of strategies	Percentage of profitable strategies
Logistic regression	44	99	44
Multilayer perceptron	34	99	34

Table 12. Successful strategies (men)

Model	Number of profitable strategies	Total number of strategies	Percentage of profitable strategies
Logistic regression (all surfaces)	24	144	17
Random forest (all surfaces)	33	144	23
XGBoost (all surfaces)	0	144	0
Logistic regression (selected surface)	18	99	18
Random forest (selected surface)	0	99	0
XGBoost (selected surface)	7	99	7

The tables reveal that the XGBoost model with the algorithm for predicting players' statistics across all surfaces and the random forest model with the algorithm for predicting players' statistics on a selected surface did not show any profitable betting strategies.

Tables 13–14 present the most successful and effective strategies for each model with the parameters $s = 100$, $t = 5$, $h = 10$.

Table 13. Most successful effective strategies for the women's division

Model	Minimum probability threshold	Minimum coefficient threshold	Prediction ratio (correct predictions / total predictions)	Percentage of correct predictions	Increase in initial capital (%)	ROI (%)
Logistic regression	0.65	1.35	22/23	96	154.8	33.65
Multilayer perceptron	0.65	1.35	27/28	96	201.8	36.04

Table 14. Most successful effective strategies for the men's division

Model	Minimum number of matches played threshold	Minimum coefficient threshold	Prediction ratio (correct predictions / total predictions)	Percentage of correct predictions	Increase in initial capital (%)	ROI (%)
Logistic regression (all surfaces)	12	1.5	14/24	57	57.4	11.96
Random forest (all surfaces)	12	1.25	19/30	63	158.6	26.43
Logistic regression (selected surface)	8	1.6	12/19	63	82	21.58
XGBoost (selected surface)	8	1.55	14/23	61	104.2	22.48

Based on the results, it can be concluded that the two best strategies for obtaining financial gains from betting on tennis match outcomes are:

- **For women:** the multilayer perceptron model, as its strategy has a higher percentage increase in initial capital and ROI.
- **For men:** the random forest model with the algorithm for predicting players' statistics across all types of courts, as it has the highest ROI and percentage increase in initial capital.

To visualize the change in ROI from betting according to the most successful effective strategies, we generated the graphs shown in Fig. 4 and 5, illustrating the effectiveness of the chosen strategies and models for women's and men's tennis matches, respectively.

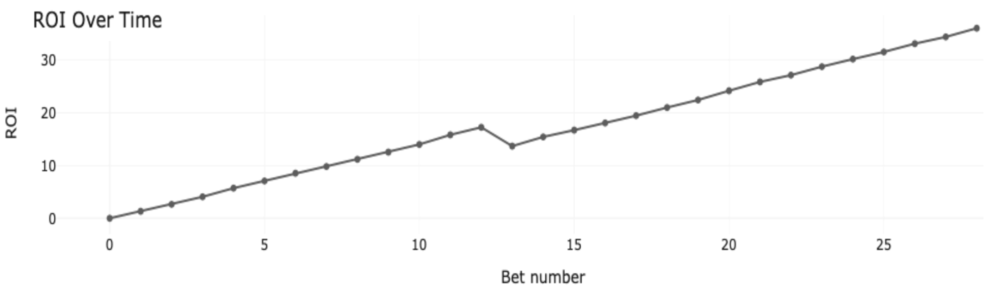


Fig. 4. Change in ROI for the most successful effective strategy based on the multilayer perceptron model for women's tennis matches

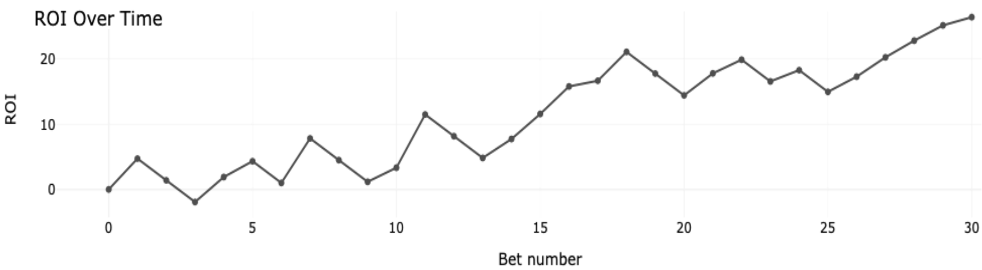


Fig. 5. Change in ROI for the most successful effective strategy based on the random forest model with statistical prediction algorithm across all types of courts for men's tennis matches

CONCLUSIONS

The first part of the conducted study was dedicated to the implementation of various machine learning methods and auxiliary algorithms for predicting the outcomes of tennis matches. The second part aimed to determine the best strategies for obtaining financial benefit from sports betting. For this work the real data both for men's and women's tennis matches were selected, processed and analyzed. Five machine learning models were developed based on logistic regression, multilayer perceptron, random forest, and extreme gradient boosting methods. Men's tennis results forecasting is based on players' statistics as predictor variables. Therefore, an algorithm that uses the time discounting method was applied, enabling the statistics forecasting for future matches based on the player's historical games. Forecasting of outcomes were made on new datasets to determine the best betting strategies. Based on the results obtained, using a filtering algorithm and the developed method for assessing strategy effectiveness, the most successful and effective betting strategies were identified for use in sports betting to maximize user profits.

A web interface was created to facilitate the use of the developed models and provide a clear interpretation of the obtained results. This interface allows users to easily manipulate input data for prediction by entering it via the keyboard or, if necessary, modifying it. In future research, we will focus on studying and using background information received from the key variables as well as modifying and proposing more different strategies for the players based on their attitude and risk tolerance.

REFERENCES

1. Gianluca Di Censo, Paul Delfabbro, Daniel King, "Examining the Role of Sports Betting Marketing in Youth Problem Gambling," *Journal of Gambling Studies*, vol.40, pp. 2005–2025, 2024. doi: http://dx.doi.org/10.1007/s10899-024-10347-x
2. N. Kuznietsova, P. Bidyuk, "Forecasting of Financial Risk Users' Outflow," *IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, Kyiv: 08–12 October, 2018, pp. 250–255. Available: https://ieeexplore.ieee.org/abstract/document/8516782
3. A.T. Fernandes, D.B. Figueiredo Filho, E.C. da Rocha, W. da S. Nascimento, "Read this paper if you want to learn logistic regression," *Revista de Sociologia e Política*, vol. 28, no. 74, 2020. doi: https://doi.org/10.1590/1678-987320287406en
4. N.V. Kuznietsova, Z.S. Chernysh, "Regression models application for analysis and forecasting of the financial activity quality indicators of the company," *System Research and Information Technologies*, no. 2, pp. 67–81, 2020. doi: https://doi.org/10.20535/SRIT.2308-8893.2020.2.05
5. M. Brown, *Applied Predictive Analytics*. 2023. [Online]. Available: https://core.ac.uk/download/568285513.pdf
6. C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, no. 3, Aug. 2020. doi: https://doi.org/10.1007/s10462-020-09896-5
7. K.Y. Chan et al., "Deep neural networks in the cloud: Review, applications, challenges and research directions," *Neurocomputing*, vol. 545, p. 126327, Aug. 2023. doi: https://doi.org/10.1016/j.neucom.2023.126327

8. M. Sipko, W. Knottenbelt, *MEng Computing -Final year project Machine Learning for the Prediction of Professional Tennis Matches*. 2015. Available: https://www.doc.ic.ac.uk/teaching/distinguished-projects/2015/m.sipko.pdf
9. "GitHub - JeffSackmann/tennis_atp: ATP Tennis Rankings, Results, and Stats," *GitHub*. Available: https://github.com/JeffSackmann/tennis_atp
10. "Tennis Betting | Tennis Results | Tennis Odds," *Tennis-data.co.uk*, 2019. Available: http://www.tennis-data.co.uk/alldata.php

Received 24.12.2024

INFORMATION ON THE ARTICLE

Kyryl I. Shum, ORCID: 0009-0009-7503-3249, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: shumkirillid@gmail.com

Nataliia V. Kuznietsova, ORCID: 0000-0002-1662-1974, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: natalia-kpi@ukr.net

ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ТЕНІСНИХ МАТЧІВ І АНАЛІЗ ФІНАНСОВИХ ВИГОД / К.І. Шум, Н.В. Кузнєцова

Анотація. Реалізовано програмний продукт, який дозволяє прогнозувати результати тенісних матчів, розроблено метод визначення ефективних стратегій спортивних ставок. Теніс є одним із найпопулярніших видів спорту у світі, який привертає значну увагу як звичайних уболівальників, так і професійних аналітиків. Використання методів машинного навчання дає змогу ефективно прогнозувати результати матчів, що відкриває можливості для отримання прибутку від ставок на ймовірних переможців. Мета дослідження – оцінювання фінансової вигоди від прогнозування результатів тенісних ігор через пошук ефективної стратегії спортивних ставок. Розглянуто різні методи машинного навчання і допоміжні алгоритми та виконується їх порівняння з метою вибору найкращої стратегії укладання ставок задля максимізації потенційного прибутку користувача. Об'єкт дослідження – прогнозування результативності тенісних матчів. Предмет дослідження – моделі, методи машинного навчання та допоміжні алгоритми прогнозування результативності тенісних ігор. Результатом дослідження є визначення найкращої стратегії тенісного беттингу.

Ключові слова: прогнозування результатів тенісних ігор, машинне навчання, спортивний беттинг, стратегії ставок.

UDC 004.5
DOI: 10.20535/SRIT.2308-8893.2025.3.08

MULTI-CRITERIA MATHEMATICAL MODEL OF CREDIT SCORING IN DATA SCIENCE PROBLEMS

O.O. PYSARCHUK, M.D. VASYLIEVA, D.R. BARAN, I.O. PYSARCHUK

Abstract. A multi-criteria optimization mathematical model of credit scoring is proposed. The model is derived using a nonlinear trade-off scheme to solve multi-criteria optimization problems, allowing for the construction of a Pareto-optimal solution. The proposed approach forms an integrated assessment of a borrower's creditworthiness based on a structured set of indicators that reflect the financial, credit, and social profile of clients. The model is designed for use in intelligent CRM and ERP systems operating on Big Data and does not rely on labeled training samples, making it applicable to unsupervised learning tasks. It can also serve as a foundational layer for further deep-learning analysis. Methodological steps for implementing the model, from indicator normalization to final decision-making, are described. A technological implementation demonstrates the model's effectiveness in automated loan decisions and fraud detection.

Keywords: Data Science, Big Data, SCORIG machine learning, decision making, multi-criteria mathematical models, intelligent CRM, ERP systems.

INTRODUCTION

The development of the modern IT industry determines the methodologies and technologies of electronic banking. This also affects the automation of intelligent decision-making processes. One of these directions is making decisions about granting loans to consumers of lending services (clients) provided by banking institutions. This process relates to the field of credit scoring (SCORIG). It is based on the analysis of a set of indicators of the client's creditworthiness and establishing an individual integrated assessment (SCORE) in order to make an informed decision on granting a loan. The practice of scoring analysis is not limited to a binary yes/no assessment of lending. Scoring analysis should ensure the formation of an adequate risk assessment that determines a specific credit program adapted in the loan life cycle to the properties of a specific client. The process of scoring analysis should ensure high economic performance indicators of the banking institution. For lending programs, this means maximizing the number of loans issued, but through programs that are adequate to the risks of non-repayment of loan funds by the client. Currently, credit scoring is implemented by automated software tools that have the properties of intelligence and are organized in the format of distributed CRM or ERP systems.

Features of credit scoring in modern CRM, ERP are, in fact, the implementation of Data Science technologies on Big Data arrays. This imposes rather strict requirements on the computational complexity of credit scoring models.

Undoubtedly, the quality of automated credit scoring decision-making is determined by the mathematical models underlying the automated software tools. Therefore, the pragmatic effort of banking institutions to increase the economic efficiency makes relevant the task of developing effective mathematical models of credit scoring in Data Science tasks on Big Data arrays.

ANALYSIS OF RESEARCH AND PUBLICATIONS

The specificity of the credit scoring task consists in considering it within the class of classification methods and models. Classical methods of machine learning are most often used [1–5]: discriminant analysis; logistic regression; decision trees; method of support vectors; naive Bayes classifier; neural networks and others. In general, the assessment task refers to the theory of system analysis [6]. Performance evaluation is carried out in the following sequence: determining factors, indicators and criteria; forming the decision-making model; interpreting the obtained result. Single-criteria and multi-criteria models of efficiency evaluation are distinguished [7, 8], with the latter being comparatively more adequate. The main drawback of traditional credit scoring approaches is the consideration of the assessment task at the level of classification and forming an integrated assessment on a discrete field of static numerical representations of many factors of a particular borrower.

Currently, digital twin technologies are rapidly developing, which involve implementing any business processes into digital virtual reality in order to automate and optimize them. Digital twin is effective not so much in the aspect of automation as in the maximum approximation of the “twin” to the real physical process. This means that the twin must have a high level of adequacy but with abstraction that allows productive processing of, for example, Big Data arrays.

The main idea of the approach proposed in the article is to approximate the problem of credit scoring to its real physical essence at the formalization level. This means that a real client should come as close as possible to the image of an unattainable ideal based on a set of indicators. In this formulation, the task of scoring analysis reflects the task of multi-criteria evaluation. This will be applied to the synthesis of a mathematical model of scoring analysis, as a multi-criteria optimization mathematical model of evaluation.

Multi-criteria formalization of the decision-making task a priori has a higher level of adequacy since it mathematically allows describing a specific practical task of natural language expressed by the scheme “how best...”. Moreover, the criterion allows describing the entire set of possible indicator values, even though represented by a set of discrete limited realizations.

Moreover, the criterion allows describing the entire set of possible values of indicators, although represented by a set of discrete limited implementations. Thus, the increased adequacy of the evaluation task at the formalization stage ensures that the “twin” closely approximates the real physical process. Therefore, we should potentially expect an increase in the efficiency of the final result of the scoring analysis.

Formulation of the problem. The aim of this article is to synthesize a multicriteria optimization mathematical model for credit scoring.

PRESENTATION OF THE MAIN MATERIAL

The achievement of the stated goal is implemented at three levels: model, methodological, and technological.

I. *Multicriteria Optimization Mathematical Model of Credit Scoring*. The synthesis of the mathematical model is implemented in stages: defining factors, indicators, and criteria; forming the decision-making model; interpreting the obtained result.

Defining factors, indicators, and criteria are the initial data for evaluation and represent the scoring card. The scoring card structure is known but may vary in the number and values of indicators according to the specific conditions of a particular banking institution.

The classical structure of the scoring card includes the following groups of indicators[1–5]: information from the banking institution—credit product; information about the borrower/client—credit history; financial; social (see Table 1).

Table 1. Scoring card — general structure

Credit product (bank): ▪ Amount; ▪ Term; ▪ The purpose of the loan; ▪ …	**Financial (borrower):** ▪ Assets; ▪ Obligations; ▪ Monthly income; ▪ Monthly expenses; ▪ …
Credit history (borrower): ▪ In the current bank; ▪ In other banks; ▪ Credit bureau data; ▪ …	**Social indicators (borrower):** ▪ Work experience; ▪ Time of residence at the current address; ▪ Marital status; ▪ …

The specificity of indicator values in the scoring card is formed as a dynamic database of client interactions.

In the classic setting, the task of scoring analysis is formalized as the task of classifying new customers based on information about existing clients.

Let the set of bank customers be given

$$\{Z_i\},\ i = 1 \ldots n. \tag{1}$$

Each client is characterized by a p-dimensional vector of heterogeneouss features.

$$X_i = [x_{i1}, \ldots x_{ip}]^{\mathrm{T}}. \tag{2}$$

It is known that each client Y_i belongs to one of two creditworthiness classes $k \le 2$:

$$Y = \begin{cases} y = 1 - \text{the client is creditworthy}, \\ y = 0 - \textit{the client is not creditworthy}. \end{cases} \tag{3}$$

New clients are characterized by a sample: $\{W_j\},\ j = 1 \ldots m$.

A sample of clients with known creditworthiness class serves as the training set — $\{Z_i^N\},\ i = 1 \ldots (n - N)$.

It is necessary to implement a scoring algorithm that classifies new clients $\{W_j\}$, $j = 1 \ldots m$, based on their feature vectors $X_i = [x_{i1}, \ldots x_{ip}]^{\mathrm{T}}$.

The specified scheme corresponds to the strategy of learning with a teacher, however, in the practice of scoring analysis, the presence of a training sample is a rather rare phenomenon. In this case, forming a training set becomes a separate, rather complex task.

Additionally, discrete indicators of the scoring card may not be informative when considered individually. This necessitates calculating secondary indicators and comparing them with other clients.

Therefore, the article considers a modified formulation of the problem of scoring analysis.

Let a set of bank clients (1) be given. Each client is characterized by a p-dimensional vector of heterogeneous features (2) $X_i = [x_{i1}, \ldots x_{ip}]^{\mathrm{T}}$. It is necessary to classify each bank client Y_i into two classes (3).

The multi-criteria scoring method is proposed to solve the classification problem formalized in this way. This decision is based on the following considerations. Scoring analysis is essentially aimed at building a digital twin of the banking institution's team, which forms the requirements for the ideal client. This enables the analysis of alternatives for binary classification of clients based on a multitude of factors, comparing the ideal image with the real client — evaluating the degree of closeness between them. This process is accompanied by considerations/doubts/analysis of many factors—often following the "best–worst" scheme. It is multicriteria scoring that allows incorporating the decision maker's considerations in transforming static indicators into dynamic requirements of criteria.

In addition, the scoring model must meet a number of technical requirements [1–5]:

1. High adequacy in dividing borrowers into two categories from the perspective of credit issuance: "positive" and "negative";

2. The scoring point is a measure of the probability of the borrower belonging to the "positive" or "negative" class;

3. The scoring model should form the average rating of "negative" borrowers significantly lower than the average rating of "positive" ones;

4. There should be a ranking of borrowers within the rating of "positive" decisions;

5. There is a cut-off point when it is unprofitable for the bank to issue loans to borrowers below a certain scoring point;

6. The scoring model should ensure detection of fraud.

Multi-criteria scoring implements the given list of requirements and has a number of unique advantages, which will be proved with a computational example [8].

Based on the structure of the scoring card (Table 1) and setting extremum requirements for its indicators, we generally obtain a system of criteria for the categories of the scoring card:

credit product (bank)

$$P = [p_i \to \text{extreme}]^{\mathrm{T}}, \; i = 1 \ldots k_p, \tag{4}$$

financial (borrower)

$$F = [f_i \to \text{extreme}]^{\mathrm{T}}, \; i = 1 \ldots k_f ,$$

credit history (borrower)

$$K = [k_i \to \text{extreme}]^{\mathrm{T}}, \; i = 1 \ldots k_k ,$$

social (borrower)

$$S = [s_i \to \text{extreme}]^{\mathrm{T}}, \; i = 1 \ldots k_s ,$$

extended vector of criteria

$$W = [w_i = p_i, f_i, k_i, s_i \to \text{extreme}]^{\mathrm{T}}, \quad i = 1 \ldots k_p + k_f + k_k + k_s .$$

The directions of the extremum (*extreme=min, max*) of each indicator of the scoring card are unique for each banking institution. This effectively reflects the bank's understanding of the image of the ideal client and considers the logic of mental deliberations "best-case scenario – worst-case scenario".

Analysis of the content and practical significance of indicators in the scoring card suggests a conflicting nature of criteria for the "ideal" borrower. Therefore, we have a multicriteria optimization problem in scoring.

The decision-making model is formed by aggregating/integrating partial criteria vectors (4) into a generalized/integrated assessment score using convolution through a non-linear trade-off scheme [8].

Compared to other aggregation schemes of partial criteria [9], convolution has a number of proven advantages [8]. The convolution uses a non-linear trade-off scheme, which allows obtaining a Pareto-optimal solution with low computational costs. The optimization problem is solved under constraints, ensuring unimodality of the generalized criterion function and guaranteeing a unique solution in any case. Convolution enables the use of a minimax approach, focusing on maximizing the dominant partial criterion of optimality. Weight coefficients of partial criteria allow consideration of subjective factors in dominating their influence on scoring results.

The convolution criterion for discretely given partial optimality criteria has the form [8]:

$$Y(\varphi_0) = \sum_{l=1}^{b} \gamma_{0l} (1 - \varphi_{0l})^{-1} \to \min , \tag{5}$$

where $l = 1 \ldots b$ — the number of partial optimality criteria included in the convolution; γ_{0l} — normalized weight coefficient; φ_{0l} — normative partial criterion.

The values of weight coefficients are assigned within a unified rating scale and normalized according to the expression:

$$\gamma_{0l} = \frac{\gamma_l}{\sum_{l=1}^{b} \gamma_l}, \tag{6}$$

where γ_l is the current (non-normalized) value of the weight coefficient.

The normalization of partial criteria aims to bring them to a single scale of change (0...1) and to the direction of minimization. Therefore, partial criteria that are minimized and those that are maximized are normalized separately.

Normalization can be implemented, for example, relative to the maximum (minimum) values characterizing the change in partial optimality criteria by expressions

$$\varphi_{0l}^{\min} = \frac{\varphi_l^{\min}}{\max \varphi_l^{\min} + \Delta}, \ \varphi_{0l}^{\max} = \frac{\min \varphi_l^{\max} - \Delta}{\varphi_l^{\max}}, \tag{7}$$

where $\max \varphi_l^{\min}$, $\min \varphi_l^{\max}$ — maximum and minimum values of the minimizing and maximizing criteria in the interval of their consideration; Δ — the reserve coefficient, which varies between 0.1 and 0.3 and ensures the elimination of the operation of division by zero for normalization of values $\max \varphi_l^{\min}$, $\min \varphi_l^{\max}$.

Convolution (5) can be presented in matrix form

$$Y(\varphi_0) = G\Phi^T,$$
$$G = [\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_l], \ l = 1 \ldots b, \tag{8}$$
$$\Phi = [(1-\varphi_{01})^{-1}, (1-\varphi_{02})^{-1}, (1-\varphi_{03})^{-1}, \ldots, (1-\varphi_{0l})^{-1}]^T, \ l = 1 \ldots b.$$

To form a multi-criteria mathematical model of bank scoring, we formalize the appearance of the scoring card in the accepted notation (4) — see table 2.

Table 2. The scoring card — formalized structure

№	*P*				*F*				*K*				*S*			
	p_1	p_2	…	$p_{i=k_p}$	f_1	f_2	…	$f_{i=k_f}$	k_1	k_2	…	$k_{i=k_k}$	s_1	s_2	…	$s_{i=k_s}$
	w_1	w_2	…	w_{k_p}	w_{1+k_p}	w_{2+k_p}	…	$w_{k_p+k_f}$	…	$w_{k_p+k_f+k_k+k_s}$						
1	$w_{1(1)}$	$w_{1(2)}$	…	$w_{1(k_p)}$	$w_{1(1+k_p)}$	$w_{1(2+k_p)}$	…	$w_{1(k_p+k_f)}$	…	$w_{1(k_p+k_f+k_k+k_s)}$						
2	$w_{2(1)}$	$w_{2(2)}$	…	$w_{2(k_p)}$	$w_{2(1+k_p)}$	$w_{2(2+k_p)}$	…	$w_{2(k_p+k_f)}$	…	$w_{2(k_p+k_f+k_k+k_s)}$						
…	…	…	…	…	…	…	…	…	…	…						
ν	$w_{\nu(1)}$	$w_{\nu(2)}$	…	$w_{\nu(k_p)}$	$w_{\nu(1+k_p)}$	$w_{\nu(2+k_p)}$	…	$w_{\nu(k_p+k_f)}$	…	$w_{\nu(k_p+k_f+k_k+k_s)}$						

Taking into account the given designations, the generalized assessment of the ν-th borrower according to the vector of criterion requirements (4) for the scoring card of Table 1 in accordance with the convolution (5) is determined by the expression:

by the extended vector of criteria in scalar form:

$$Y_\nu(w_0) = \sum_{l=1}^{k_p+k_f+k_k+k_s} \gamma_{\nu(0l)}(1-w_{\nu(0l)})^{-1} \to \min, \tag{9}$$

by the extended vector of criteria in matrix form:

$$Y_\nu(w_0) = G_\nu \Phi_\nu^{\mathrm{T}},$$
$$G_\nu = [\gamma_{(1)}, \gamma_{(2)}, \gamma_{(3)}, \ldots, \gamma_{(l)}], \ l = 1 \ldots k_p + k_f + k_k + k_s, \tag{10}$$
$$\Phi_\nu = [(1-\mathrm{w}_{\nu(01)})^{-1}, (1-\mathrm{w}_{\nu(02)})^{-1}, (1-\mathrm{w}_{\nu(03)})^{-1}, \ldots, (1-\mathrm{w}_{\nu(0l)})^{-1}]^T,$$
$$l = 1 \ldots k_p + k_f + k_k + k_s.$$

Normalization of weight coefficients, as well as partial criteria included in (9), (10), is implemented according to expressions (6), (7), taking into account the direction of the extremum.

To account a significant number of criteria in the generalized multicriteria assessment, it is advisable to use the technology of nested convolutions. This approach also allows regulating the influence of groups of scoring card indicators on the assessment result. This is implemented by sequentially (within the four groups of partial criteria (4)) reduction of partial criteria to the generalized by group and to the integrated efficiency criterion in scalar form:

$$P_{\text{v}}(p_0) = \sum_{l=1}^{k_p} \gamma_{\text{v}(0l)} (1 - p_{\text{v}(0l)})^{-1} \to \min ,$$

$$F_{\text{v}}(f_0) = \sum_{l=1}^{k_f} \gamma_{\text{v}(0l)} (1 - f_{\text{v}(0l)})^{-1} \to \min ,$$

$$K_{\text{v}}(k_0) = \sum_{l=1}^{k_k} \gamma_{\text{v}(0l)} (1 - k_{\text{v}(0l)})^{-1} \to \min ,$$

$$S_{\text{v}}(s_0) = \sum_{l=1}^{s} \gamma_{\text{v}(0l)} (1 - s_{\text{v}(0l)})^{-1} \to \min ,$$

$$Y_{\text{v}}(w_0) = \gamma^{P}_{\text{v}(0l)} (1 - P_{\text{v}0}(p_0))^{-1} + \gamma^{F}_{\text{v}(0l)} (1 - F_{\text{v}0}(f_0))^{-1} +$$

$$+ \gamma^{K}_{\text{v}(0l)} (1 - K_{\text{v}0}(k_0))^{-1} + \gamma^{S}_{\text{v}(0l)} (1 - S_{\text{v}0}(s_0))^{-1} \to \min , \qquad (11)$$

$$P_{\text{v}0}(p_0) = [\sum_{i=1}^{k_p} (1 - [\max p_{\text{v}(0l)} - \Delta])^{-1}]^{-1} ,$$

$$F(p_0) = [\sum_{i=1}^{k_p} (1 - [\max f_{\text{v}(0l)} - \Delta])^{-1}]^{-1} ,$$

$$K_{\text{v}0}(p_0) = [\sum_{i=1}^{k_p} (1 - [\max k_{\text{v}(0l)} - \Delta])^{-1}]^{-1} ,$$

$$S_{\text{v}0}(p_0) = [\sum_{i=1}^{k_p} (1 - [\max s_{\text{v}(0l)} - \Delta])^{-1}]^{-1} . \qquad (12)$$

Normalization (12) is performed relative to the worst assessment — the maximum value of the normalized indicator, which characterizes the partial criterion of the scoring card.

Similarly, matrices are formed and generalized group criteria ratings, which are part of the matrix model of multicriteria scoring (10), are normalized.

$$Y_{\text{v}}(w_0) = \gamma^{P}_{\text{v}(0l)} (1 - G_{pv0} \Phi^{T}_{pv0})^{-1} + \gamma^{F}_{\text{v}(0l)} (1 - G_{fv0} \Phi^{T}_{fv0})^{-1}$$

$$+ \gamma^{K}_{\text{v}(0l)} (1 - G_{kv0} \Phi^{T}_{kv0})^{-1} + \gamma^{S}_{\text{v}(0l)} (1 - G_{sv0} \Phi^{T}_{sv0})^{-1} \to \min ,$$

$$G_{pv0}, G_{fv0}, G_{kv0}, G_{sv0} = \text{normalization}[\gamma_{\text{v}(0l)}] , \quad l = 1 \ldots k_p, k_f, k_k, k_s$$

are the same in structure, but may have different values

$$Y_{\text{v}}(w_0) = \gamma^{P}_{\text{v}(0l)}(1 - G_{pv0}\Phi^{\text{T}}_{pv0})^{-1} + \gamma^{F}_{\text{v}(0l)}(1 - G_{fv0}\Phi^{\text{T}}_{fv0})^{-1}$$

$$+ \gamma^{K}_{\text{v}(0l)}(1 - G_{kv0}\Phi^{\text{T}}_{kv0})^{-1} + \gamma^{S}_{\text{v}(0l)}(1 - G_{sv0}\Phi^{\text{T}}_{sv0})^{-1} \to \min , \qquad (13)$$

$$G_{pv0}, G_{fv0}, G_{kv0}, G_{sv0} = \text{normalization}[\gamma_{\text{v}(0l)}] ,$$

$l = 1 \ldots k_p, k_f, k_k, k_s$ are the same in structure, but may have different values

$$\Phi_{pv0} = \text{normalization}[(1 - p_{\text{v}(0l)})^{-1}]^{\text{T}} , \; l = 1 \ldots k_p ,$$

$$\Phi_{fv0} = \text{normalization}[(1 - f_{\text{v}(0l)})^{-1}]^{\text{T}} , \; l = 1 \ldots k_f ,$$

$$\Phi_{kv0} = \text{normalization}[(1 - k_{\text{v}(0l)})^{-1}]^{\text{T}} , \; l = 1 \ldots k_k , \qquad (14)$$

$$\Phi_{sv0} = \text{normalization}[(1 - s_{\text{v}(0l)})^{-1}]^{\text{T}} , \; l = 1 \ldots k_s .$$

The interpretation of the obtained result involves bringing the value of the generalized assessment (9), (10), or in the form (11), (13) to a unified scale, for example, from 0 (the worst rating) to 1 (the best rating). This is achieved by normalizing the generalized score to the abstract worst customer score according to the expression

$$I_0 = 1 - \frac{I}{\max I}, \quad \max I = \sum_{i=1}^{5} (1 - [\max F_i - \Delta])^{-1} , \qquad (15)$$

where $\max F_l$ — the worst possible value of the partial indicator; Δ — the reserve coefficient, which ensures the avoidance of incorrect operations during normalization.

The obtained numerical assessment can be converted to the linguistic category of the client's solvency according to the fundamental assessment scale of the Table 2.

Table 3. Fundamental rating scale

Integrated performance assessment I_0	**Linguistic category of efficiency**
1,0 – 0,7	High
0,7 – 0,5	Good
0,5 – 0,4	Satisfactory
0,4 – 0,2	Low
0,2 and less	Unsatisfactory

The numerical evaluation of the normalized generalized indicator (15) (see the left column of Table 3) is proportional to the probability of the client returning the loan. That is, it characterizes the risk of providing a credit loan.

Thus, expressions (9)–(15) form *a multi-criteria mathematical model of credit scoring*. The differences, advantages and features of the model are as follows. The model allows you to consider the indicators of the scoring card in terms of infological connections: factors, indicators, criteria, which contributes to increasing the adequacy of the ideal client profile of a banking institution. The model ensures obtaining a generalized client assessment as a solution to an optimization problem using a minimax approach to image requirements. The obtained solution is Pareto-optimal. Subjective priorities of scoring card indicators can be taken into account in client assessment by adjusting both partial and group criteriaweights. The model is structurally open to adding scoring card indicators. The proposed model does not require a priori data on loan issuance/refusal to clients, that is, it implements an unsupervised learning scheme. Thus, the proposed model can be a primary superstructure, acting as a highly accurate binary classifier to deep learning methods based on artificial neural networks. Undoubtedly, artificial neural networks are designed and capable of accumulating large segments of labeled data, and the proposed multi-criteria model cannot compete with these advantages. But in the context of unsupervised learning, the multi-criteria model has better potential properties than the existing approaches in the essence of the formalization of the classification task. Research presented below has proven the model's capability to detect fraud. The model fully meets the requirements for bank scoring models, which will also be proved by a computational example.

II. *The methodology of multi-criteria credit scoring* determines the sequence of actions for for performing calculations and obtaining the resulting assessment, including the following stages:

1. Establishing a set of indicators from the scoring card (Table 1) in the form of (4).
2. Normalizing criteria (4) using expressions (6), (7).
3. Formulating the generalized client assessment, expressions (9)–(14).
4. Interpreting the generalized assessment with normalization (15) and in accordance with Table 3.

III. *The technology of multicriteria credit scoring* involves practical aspects of implementing the synthesized model (9)–(15) and the methodology of its application to the architecture of the software system and to a specific script-based implementation.

The technological processes *of multi-criteria credit scoring can be represented by the structural diagram in Fig. 1, which implements the architecture of the software script of credit scoring.*

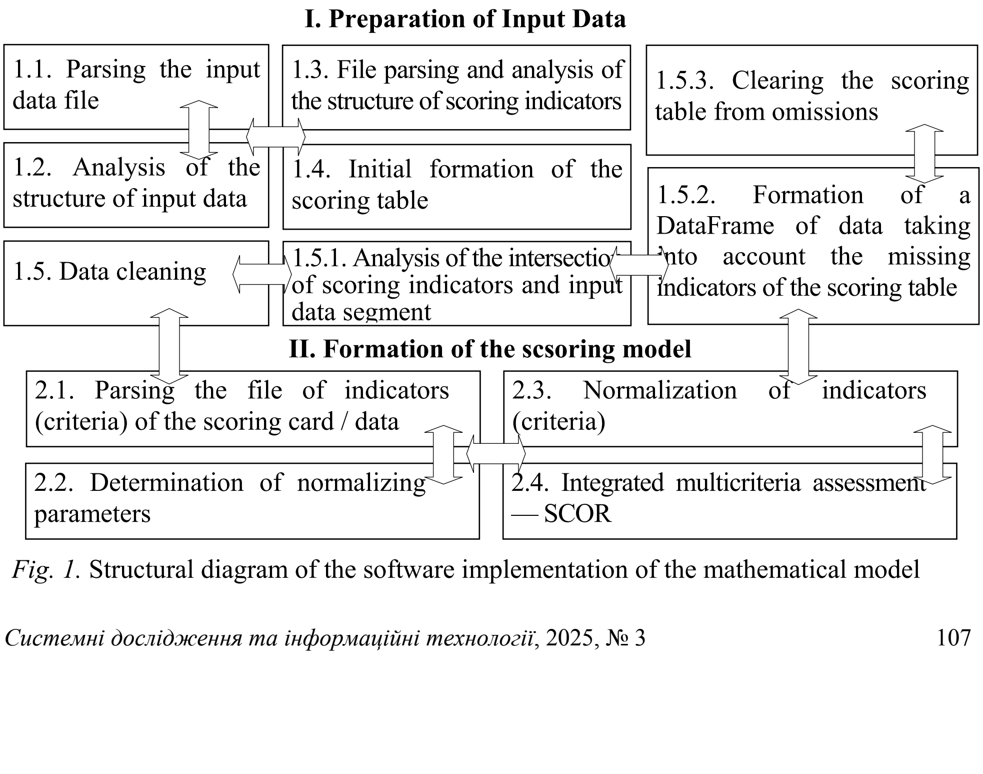


Fig. 1. Structural diagram of the software implementation of the mathematical model

Technological processes are divided into two blocks: preparation of input data and formation of the scoring model. The input data includes two files: scoring card indicators in “sample_data.xlsx” and indicator descriptions in “data_description.xlsx” (Fig. 2). In total, there are 121 scoring card indicators and 500 records of potential bank clients.

	A	B	C	D	E	G	H	I	J	K	L	M
1	Application	loan_amoun	loan_days	applied_at	gender	birth_date	Marital	children	education	fact_addr	fact_addr_start	has_immov
2	1	3000	30	01.02.2021 0:21	1	03.02.1995	2	1	5	4	14.06.2015	0
3	2	1000	7	01.02.2021 0:24	1	19.01.1984	2	2	4	0	NULL	0
4	3	1000	3	01.02.2021 0:36	2	02.08.1994	2	1	4	2	17.05.2018	0
5	4	1600	30	01.02.2021 0:34	1	12.11.1992	1	1	3	3	25.11.1992	0
6	5	2500	18	01.02.2021 23:22	2	22.10.1997	1	1	5	3	01.01.1999	1
7	6	1000	30	01.02.2021 0:38	2	12.06.1996	2	1	3	3	10.12.2009	0
8	7	1300	30	01.02.2021 0:40	1	05.04.1993	1	1	5	1	26.09.2012	1
9	8	2000	14	01.02.2021 1:00	1	17.09.1984	2	2	5	3	19.09.2012	0
10	9	2500	12	01.02.2021 0:47	1	27.09.1986	1	1	5	1	12.01.2015	0
11	10	1000	7	01.02.2021 0:50	2	15.06.1991	1	1	6	4	01.09.2014	0
12	11	2000	30	01.02.2021 0:52	2	21.08.1981	5	3	3	1	12.03.1981	0
13	12	1400	3	01.02.2021 0:50	2	04.04.1994	4	1	4	2	29.07.2010	0
14	13	2500	30	01.02.2021 1:19	1	27.06.1991	2	2	3	2	27.09.2007	0
15	14	2000	30	01.02.2021 1:04	1	11.03.1998	1	1	4	2	08.05.1998	0
16	15	1500	30	01.02.2021 2:17	1	09.08.1996	1	1	2	1	27.10.2002	0
17	16	1000	30	01.02.2021 1:40	1	25.02.1995	1	1	5	3	25.02.1995	0

a — Scoring card sample_data.xlsx

	A	B	C	D	E	F
1	**Field_in_data**	**Description_of_information**	**Filling**	**At_the_time**	**Place_of_definition**	**Note**
2	**id**	Application ID	Necessarily	Known	Determined by the application registration system at the time of submission	
3	**loan_amount**	Loan amount in the application	Necessarily	Known	Specified by the borrower	
4	**loan_days**	Loan repayment term in the application	Necessarily	Known	Specified by the borrower	
5	**applied_at**	Date and time of application acceptance	Necessarily	Known	Determined at the time of application	Local time on the user's device or application time?
6	**language**				Determined by the browser at the time of application	Always UA
7	**purpose_id**	ID of the purpose of loan use	Necessarily	Known		Always 999-999, the client enters the destination itself, previously it was possible to choose from a list
8	**purpose_other**	The purpose of using the loan	Necessarily	Known	Specified by the borrower	
9	**ip**	IP of the client from which the request came	Necessarily	Known	Determined by the browser at the time of application	
82	**loan_status_id**	Not used		Not known	Assigned by the application processing system	
83	**decision_status_id**	The status of the decision on the application		???		
84	**decision_application_**	ID in the decision-making system		???		
85	**decision_flow_id**	ID the credit policy process		???		
86	**decision_is_exported**	Export check box		???		
87	**decision_exported_at**	Export date		???	**fields are related to decision-making**	
88	**product_id**	ID of the product issued to the customer	Necessarily	???	Parameters related to this product	

b — Indicators of the scoring card data description.xlsx

Fig. 2. Structure of input data

The implementation of credit scoring technology in script form, according to the developed model and its application methodology, is carried out using the Python programming language with libraries such as NumPy and Pandas. **The original project with the code is available at** https://github.com/Pysarchuk-O/scoring.git .

Over the input data files, a series of preparatory stages are implemented, dictated by the specifics and details of the data: parsing of data files and converting them to the pandas – DataFrame format (blocks 1.1, 1.3); analyzing the structure of the input data by size, data types, and presence of missing values (block 1.2); initial formation of the scoring card and analysis of its structure for compliance with the fields of Table 1 (block 1.4); cleaning the input data from gaps using the rejection strategy in cases of a significant number of them and the impossibility of recovery based on the essence of the indicator (registration address, residence, place of work, etc.); selection of values for the scoring card based on the intersection of input data (Fig. 2, *a*) and indicators (Fig. 2, *b*) with further control of preservation of the of Table 1's structure (block 1.5). After these stages, which are classic, a content analysis of indicators is implemented (also within Block 1.5's functionality). This involves selecting objective-subjective indicators that purely characterize the borrower and are not secondary, as designated by the system (product status, product profile ID, etc.). To specify the data in Fig. 2, these are indicators from the scoring card marking fields (see Fig. 2, *b*) "borrower indication" and "parameters related to the issued product". Analysis showed that the data in Fig. 2 do not have solutions for issuing the product, there is no data structure of the "training pair" type. Thus, fields related to "decision-making" (Fig. 2, *b*) are absent in the scoring card (Fig. 2, *a*). Therefore, we are dealing with data oriented towards implementing an unsupervised machine learning model.

The result of the data preparation stage is a scoring card in general form, as shown in Fig. 2, *a*, which contains a list of indicators presented in Fig. 3. Analysis of the scoring card structure according to Fig. 2 and comparison with Table 1 confirms the correspondence and presence of the main categories for scoring.

	Field_in_data	Description_of_information	Filling	At_the_time	Place_of_definition	Note
0	loan_amount	Loan amount in the application	Necessarily	Known	Specified by the borrower	
1	loan_days	Loan repayment term in the application	Necessarily	Known	Specified by the borrower	
2	gender_id	Gender of the borrower	Necessarily	Known	Specified by the borrower	ID decryption is required
3	birth_date	Borrower's date of birth	Necessarily	Known	Specified by the borrower	Sometimes it is filled incorrectly
4	children_count_ic	The number of children of the borrower	Necessarily	Known	Specified by the borrower	ID decryption is required
5	education_id	The borrower's education level	Necessarily	Known	Specified by the borrower	ID decryption is required
6	fact_addr_owner	Residential address: property type		Known	Specified by the borrower	ID decryption is required
7	has_immovables	Owns real estate	Necessarily	Known	Specified by the borrower	

Fig. 3. Scoring card indicators after data cleaning and balancing d_segment_data_ description_cleaning.xlsx

Further, a series of stages for script deployment and application of the proposed multicriteria credit scoring model is implemented. This is implemented in accordance with the formulated methodology and is reflected in the structural diagram in Fig. 1 by blocks 2.1–2.4.

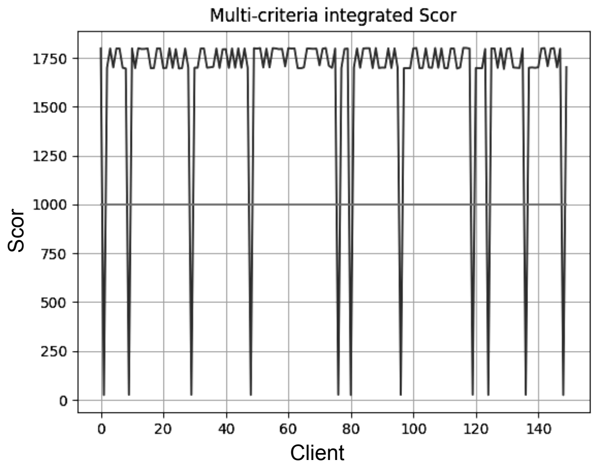
The initial step to initiate the model's operation involves parsing the indicators (criteria) file of the scoring card (block 2.1). The criteria requirements for the scoring card indicators are formulated based on the values remaining after cleaning and balancing (Fig. 3). These criteria should reflect the ideal client profile and are unique to each banking institution. It is possible, for example, to establish a system of criteria requirements shown in Fig. 4. They are used for the calculation example.

	A	B	C	D	E	F	G
1		Field_in_data	Description_of_information	Minimax	At_the_time	Place_of_definition	Note
2	0	loan_amount	Loan amount in the application	min	Known	Specified by the borrower	
3	1	loan_days	Loan repayment term in the application	min	Known	Specified by the borrower	
4	2	gender_id	Gender of the borrower	max	Known	Specified by the borrower	ID decryption is requi
5	3	birth_date	Borrower's date of birth		Known	Specified by the borrower	Sometimes it is filled incorrectly
6	4	children_count_id	The number of children of the borrower	min	Known	Specified by the borrower	ID decryption is requi
7	5	education_id	The borrower's education level	max	Known	Specified by the borrower	ID decryption is requi
8	6	fact_addr_owner_type_	Residential address: property type		Known	Specified by the borrower	ID decryption is requi
9	7	has_immovables	Owns real estate	max	Known	Specified by the borrower	
10	8	has_movables	Owns a vehicle	max	Known	Specified by the borrower	
11	9	employment_type_id	Place of work: type of employment	max	Known	Specified by the borrower	ID decryption is requi
12	10	organization_type_id	Place of work: type of organization	max	Known	Specified by the borrower	ID decryption is requi
13	11	organization_branch_id	Place of work: branch of the organizatio	max	Known	Specified by the borrower	ID decryption is requi
14	12	employees_count_id	Place of work: number of employees of	max	Known	Specified by the borrower	ID decryption is requi

Fig. 4. Criterion requirements for scoring card indicators d_segment_data_description_ cleaning_minimax.xlsx

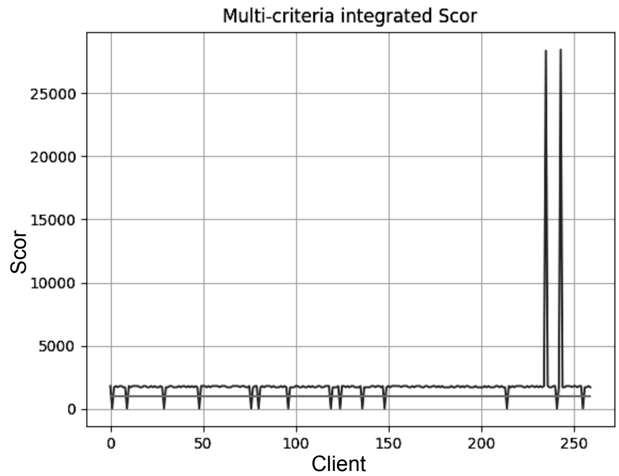
Further, the selection of minimax indicators indicated in Fig. 4 from the scoring map of Fig. 1 and the formation of an integrated multicriteria assessment (scor-py) according to model (9) is implemented. The calculation results are presented in the graphs of Fig. 5.

The graphs in Fig. 5 show the unnormalized scoring values (scor axis) for each client included in the scoring card in Fig. 2, *a* (client axis). Fig. 5, *a* presents the scoring evaluation for 150 clients without abnormally high scores. Fig. 5, *b* and 5, *c* show the scoring evaluations for 250 and all 500 clients with abnormally high score values. The red line characterizes the empirical value of dividing customers into creditworthy and non-creditworthy. The results shown in Fig. 5, *a* demonstrate a strictly binary classification of clients into the mentioned categories. Differences in the scores of creditworthy clients reflect their internal distribution, proportional to the risk characteristics of granting credit. Abnormally high scoring values presented in Fig. 5, *b* and 5, *c* indicate possible fraud in the data or incorrectly filled fields in the scoring card. Fraud analysis involves reversing the process of analyzing the criteria vector for each indicator individually and in the multicriteria scoring complex.

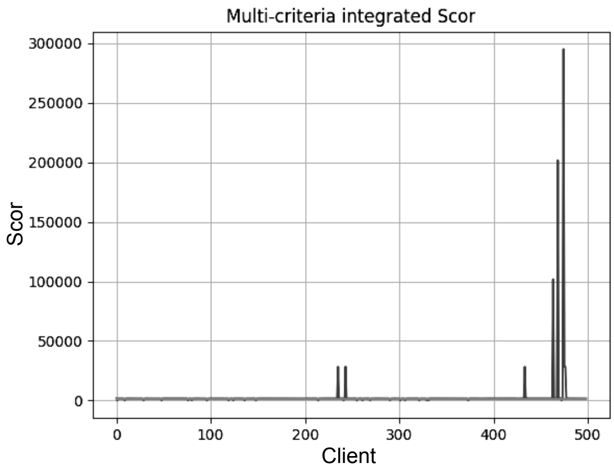


a — The scoring evaluation for 150 clients without abnormally high scores

Fig. 5. Client Creditworthiness Assessment — scor (*Beginning*)



b — The scoring evaluations for 250 clients with abnormally high score values



c — The scoring evaluations for 500 clients with abnormally high score values

Fig. 5. Client Creditworthiness Assessment — scor (*Continued*)

Comparison of the above results of calculations obtained in accordance with the proposed approach was carried out with discriminant analysis on a similar segment of data. The results of the comparison in more than 80% did not contradict each other, which indicates the effectiveness of the proposed approach.

CONCLUSIONS

In the research, a multi-criteria mathematical model of credit scoring for Data Science tasks was developed. The model provides the formation of the ideal client profile in the format of criterial requirements. The integrated assessment is formed using convolution according to a nonlinear trade-off scheme and represents the solution to an optimization problem, belonging to Pareto optimal solutions. The model operates on data which corresponds to the principles of unsupervised learning. It can be applied independently or serve as a foundational layer for deep learning methods. The practical application of the model is facilitated by its methodological and technological representations. A computational example of applying the model to real big data proved its effectiveness for the decision-making stage of customer lending and fraud detection.

REFERENCES

1. Mohamed Hani AbdElHamid, Mohamed Tawfik ElMasry, *Machine learning approach for credit score analysis: a case study of predicting mortgage loan defaults*. NOVA Information Management School Instituto Superior de Estatística e Gestão de Informação Universidade Nova de Lisboa, 2019, 60p. Available: https://run.unl.pt/bitstream/10362/62427/ 4/TEGI0439.pdf
2. Maria Fernandez Vidal, Fernando Barbon, *Credit scoring in financial inclusion. How to use advanced analytics to build credit-scoring models that increase access.* Washington: Consultative Group to Assist the Poor: World Bank, 2019, 52 p. Available: https://www.findevgateway.org/guide-toolkit/2019/07/credit-scoring-financial-inclusion
3. "Credit scoring approaches guidelines," *The World Bank Group*, 2019, 64 p. Available: https://thedocs.worldbank.org/en/doc/935891585869698451-0130022020/ original/CREDITSCORINGAPPROACHESGUIDELINESFINALWEB.pdf
4. Rory P. Bunker, M. Asif Naeem, Wenjun Zhang, "Improving a Credit Scoring Model by Incorporating Bank Statement Derived Features," *Preprint submitted to Expert Systems with Applications*, 2017. Available: https://www.researchgate.net/ publication/309606871_Improving_a_Credit_Scoring_Model_by_Incorporating_Bank_Statement_Derived_Features
5. Engku Muhammad Nazri E.A. Bakar, "Credit scoring models: techniques and issues," *Journal of Advanced Research in Business and Management Studies*, vol. 7, issue 2, pp. 29–41, 2017. Available: https://www.researchgate.net/publication/323306120_Credit_scoring_models_techniques_and_issues
6. Gerrit Muller, *System Modeling and Analysis: a Practical Approach*. University of South-Eastern Norway-NISE, 2023, 128 p. Available: https://www.gaudisite.nl/ SystemModelingAndAnalysisBook.pdf
7. Nolberto Munier, *Mathematical Modelling of Decision Problems. Using the SIMUS Method for Complex Scenarios*. Springer, 2021, 196 p.
8. Albert Voronin, *Multi-Criteria Decision Making for the Management of Complex Systems*. IGI Global, 2017, 220 p.
9. Goran Cirovi´c, Dragan Pamuˇcar, *Multiple-Criteria Decision Making*. MDPI, 2022, 313 p.

Received 10.07.2024

INFORMATION ON THE ARTICLE

Oleksii O. Pysarchuk, ORCID: 0000-0001-5271-0248, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: PlatinumPA2212@gmail.com

Maria D. Vasylieva, ORCID: 0000-0003-2217-643X, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: mdvasilieva@gmail.com

Danylo R. Baran, ORCID: 0000-0002-3251-8897, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: danil.baran15@gmail.com

Illya O. Pysarchuk, ORCID: 0000-0003-4343-0142, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: flimka134@gmail.com

БАГАТОКРИТЕРІАЛЬНА МАТЕМАТИЧНА МОДЕЛЬ КРЕДИТНОГО СКОРИНГУ В ЗАДАЧАХ DATA SCIENCE / О.О. Писарчук, М.Д. Васильєва, Д.Р. Баран, І.О. Писарчук

Анотація. Запропоновано багатокритеріальну оптимізаційну математичну модель кредитного скорингу. Модель отримано з використанням нелінійної схеми компромісів для вирішення задач багатокритеріальної оптимізації, що дозволяє побудувати Парето-оптимальне рішення. Запропонований підхід формує інтегровану оцінку кредитоспроможності позичальника на основі структурованого набору показників, що відображають фінансовий, кредитний та соціальний профіль клієнтів. Модель призначено для використання в інтелектуальних CRM та ERP системах, що працюють з великими даними, і не потребує розмічених навчальних вибірок, що робить її придатною для задач навчання без учителя. Вона також може слугувати базовим рівнем для подальшого аналізу з використанням методів глибинного навчання. Описано методологічні кроки впровадження моделі, від нормалізації показників до прийняття остаточних рішень. Технологічна реалізація демонструє ефективність моделі в автоматизованому прийнятті рішень щодо кредитування і виявленні шахрайства.

Ключові слова: Data Science, Big Data, SCORIG машинне навчання, прийняття рішень, багатокритеріальні математичні моделі, інтелектуальні CRM, ERP системи.

UDC 658.562:004.855.5
DOI: 10.20535/SRIT.2308-8893.2025.3.09

FORECASTING THE QUALITY OF TECHNOLOGICAL PROCESSES BY METHODS OF ARTIFICIAL NEURAL NETWORKS

S.S. FEDIN, O.O. ROMANIUK, R.M. TRISHCH

Abstract. A set of models of feed-forward neural networks has been created to obtain operational forecasts of the quality of mechanical engineering processes. It is established that the use of the Back Propagation of Error machine learning algorithm allows for obtaining forecasted estimates for the controlled parameter of the metal-working process with significantly smaller ranges of the mean absolute percentage error, mean square error, relative approximation error, and variance ratio criterion compared to the BFGS algorithm. It is shown that the proposed MLP neural network models can be recommended for practical applications in controlling the accuracy of the machining process of shaft-type parts.

Keywords: accuracy, details, quality, forecasting, machine learning, neural network, technological process.

INTRODUCTION

In modern conditions of information development and intellectualization of various industries, an urgent problem is the application of management methods based on quantitative assessment of quality indicators of technological processes. According to DSTU 2925-94, quality is a set of characteristics of a product (process, service) that relate to its ability to meet established and foreseeable needs [1].

In other words, it is a measure of the compliance of a certain multicriteria process with expectations or requirements, which can be presented, for example, in the form of functionally dependent statistics proposed in scientific publications [2; 3] for socio-economic systems or processes in the energy sector [4–9]. At the same time, one of the urgent problems in the modern machine-building industry is to improve the quality and efficiency of high-precision machining, the complexity of which is associated with the fact that the cutting process on well-established machines is characterized by instability and a multitude of interrelated variables [10]. Cutting conditions dynamically change randomly due to the influence of various disturbing factors: scatter of allowances, variation in the hardness and structure of the work piece metal, continuously changing cutting properties of the tool, etc. [11]. In addition, quality indicators depend on the stiffness and thermal deformation of the elements of the technological system, the nature and parameters of the relative vibrations of the tool and the work piece, etc. According to DSTU 3514-97, one of the quality indicators of technological process is accuracy, i.e. a property that determines the proximity of actual and nominal values of parameters according to their probability distribution [12].

In this case, the control of the process accuracy is reduced to forecasting the machining error at a certain point in time (or during a given machining cycle) and introducing a corrective action (a readjustment pulse) to shift the tool by the forecasted value [13].

Machining errors have systematic and random components and are essentially random variables, for the forecasting of which it is necessary to know the probable estimates of the distribution and stability characteristics over time, fixing the value of the controlled parameter of each details that is consistently manufactured [14]. These fixed values are the basis for building analytical models and control charts that can be used to estimate the components of the total machining error:

- a systematic component — to eliminate the scattering centers of the controlled parameter of the details (setting level);
- a random component — for the displacement of the dimensions of details from the centers of scattering (according to the instantaneous distribution of dimensional deviations under the constant action of external factors within controlled limits) [13].

It should be noted that the principle of using control charts and analytical models to synthesize algorithms of existing systems for controlling the accuracy of metalworking processes has some drawbacks, namely:

- the estimation of the distribution parameters should be based on the assumption that the machine setup level remains unchanged, but the center of the details size dispersion is shifted randomly;
- the use of control cards allows you to adjust the parameters of the machining equipment (machine) after evaluating the results of the previous detail before the start of production of the next detail and, thus, control is carried out off-line.

One of the ways to technologically ensure the accuracy of machining processes is to introduce corrective actions in automated machine control systems based on the results of forecasting deviations of the controlled parameters of details based on adaptive artificial intelligence models [15].

The modern approach to adaptive management requires the model to be able to automatically change its structure or algorithm of functioning. However, the effective practical application of control algorithms depends on their flexibility and learning ability [16]. Therefore, an urgent task is to improve the accuracy of the adaptation process when changing on-line technological parameters using self-tuning models. Such information models can be more adaptive due to reconfiguration (retraining) on the basis of retrospective statistical information when the parameters of the details machining process change in order to determine or adjust the control law and, as a consequence, to ensure the quality of the metalworking process. Taking into account the conditions of nonlinear dynamics of technological parameters, the efficiency of adaptive control of metalworking processes can be significantly increased by using models of rectilinear artificial neural networks (Multilayer Perceptron – MLP). At the same time, the joint use of MLP models and control cards will allow to realize the principle of information support for the accuracy of technological processes.

PROBLEM DEFINITION

The choice of the neural network modeling methodology for operational forecasting of the quality of technological processes is due to the following properties [17; 18]:

- first, neural networks are among the best methods for classifying patterns, approximating and extrapolating nonlinear functions, including non-stationary time series;
- secondly, the presence of nonlinear activation functions in a multilayer neural network ensures the effective implementation of any nonlinear mappings $X \rightarrow Y$ with a given accuracy for the identification and control of complex nonlinear technical objects;
- thirdly, the parallelism of neural networks is a prerequisite for the effective implementation of software and hardware support for neural network controllers, which allow, on the basis of quantitative retrospective information, to provide on-line control of the metalworking process based on the forecasted discrete values of the controlled parameter of the details that are sequentially manufactured.

Consider a discrete process with one input $y(t)$, for which each subsequent output value $y(t+1)$ depends only on the previous value

$$y(t+1) = f_p[y(t), y(t-1), ..., y(t-\overline{q})], \quad (1)$$

where y is an input/output, t is a discrete integer time, $\overline{q}$ is a nonnegative integer, and $f_p(\cdot)$ a function.

The task is to control an object that is described by expression (1) based on learning. The control must be performed in such a way that the output signal corresponds to a reference signal $r(t)$, subject of minimizing a certain norm $e(t) = r(t) - y(t)$. In this case, the a priori quantitative information about the control object is the value q, which is an estimate of the value $\overline{q}$ of expression (1). This task can be solved on the basis of MLP models, the adaptive properties of which allow us to consider their various architectures and configurations in the structure of a neurocontroller or neuroemulator [19].

Given a given estimate of q, an MLP-type neural network model with $n = q + 1$ inputs and one output $m = 1$ can be used to model the function $f_p(\cdot)$ of expression (1). Denoting the mapping performed by the neuroemulator of the control object as $\varphi_E(\cdot)$ and its output as y_1, we obtain

$$y_1 = \varphi_E(x_E),$$

where $x_E - n$ is an n-dimensional vector.

For case $x_E(t) = [y(t), y(t-1), ..., y(t-q)]^{\mathrm{T}}$ — the goal of neurosimulator machine learning is to minimise a norm of error $x_E(t) = [y(t), y(t-1), ..., y(t-q)]^{\mathrm{T}}$.

RESEARCH OBJECTIVE

The aim of the work is to create adaptive models of feed-forward neural networks for operational forecasting of the quality of mechanical engineering processes by the accuracy parameter of cylindrical details of the “shaft” type.

Literature review. One of the ways to solve the problem of quality assurance by increasing the accuracy and productivity of machining details is to use on-line tracking automatic control systems in machine tools. In particular, the fundamental research conducted by B.S. Balakshin made it possible to establish links between the factors acting in the process of machining details and to formalize the mechanisms of error formation [20].

In modern industrial production, technological process control is based on the use of methods and means of active control of the quality of manufactured products. In the studies of by S.S. Volosov [21] and M.S. Nevelson [22] show that the most effective means of active control are automatic or combined systems that implement the principle of adaptive control. Studies [23; 24] note that the current level of development and improvement of methods and means of active control requires the introduction of adaptive systems for monitoring technological processes based on artificial intelligence technologies, in particular, neural network modeling. For example, S.V. Bilenko [23] developed methods for identifying the state and intelligent control of the machining process, aimed at determining the optimal cutting mode with a minimum amount of a priori information for continuous correction of this mode in the face of disturbances in the dynamic system of the machine tool. In the paper A.P. Nikishechkin [24] proposed the principles of constructing neural network adaptive control systems for metalworking processes and created a method for synthesizing neural network components of an adaptive control system directly in the process of its operation. The work of P.D. Wasserman [25] shows that when choosing a neural network architecture for forecasting, several configurations with different numbers of hidden neurons are usually tested. At the same time, an effective solution to the problem of time series forecasting based on the use of MLP models is shown. An adaptive MLP model was proposed in [26] to determine the structure of the time series of deviations of the diameter of shaft-type details from the nominal size and to forecast the accuracy of the technological process of machining details by a controlled parameter. It is shown that the use of such a model under the condition of non-stationarity of the controlled parameters of product quality allows obtaining reliable information about the future state of the technological process and increasing the efficiency of quality management in real time. Paper [27] shows that for quality management at a separate stage of the technological process in the conditions of noisy input information, one of the effective methods is the use of two-layer MLPs with the Back Propagation of Error learning algorithm.

Paper [28] proposes a model of a feed-forward neural network for forecasting and controlling production, the practical application of which is aimed at implementing a mechanism for controlling continuous multi-stage production processes without intermediate outputs. Study [29] proposes a method of using neural networks to identify product defects and make corrective changes to the technological process in order to manage its quality.

It should be noted that the construction of neurocontrollers is an important area of application of neurocontrol in metalworking to ensure the quality of technological processes. Thus, in [30], a method was formalized for creating neural-index models designed to plan the process of machining rotating details based on typical examples. Study [31] proposes a controller for optimizing the milling process, in which modeling based on artificial neural networks is used to learn the correspondence between the inputs and outputs of the technological process. Tianjin University (China) has developed a milling process control technology based on the use of a three-layer neural network with the Back Propagation of

Error machine learning algorithm and performed simulations for different processing modes with experimental confirmation of the effectiveness of the proposed control technology [32].

Thus, the literature review shows the relevance of scientific and practical research aimed at managing the quality of technological processes of machining sequentially machined details based on methods for forecasting the accuracy of their manufacture using adaptive neural networks.

MATERIALS AND METHODS

The controlled parameter of sequentially machined details is the accuracy of the cylindrical surface diameter of the shaft detail, namely the deviation of the actual surface size from its nominal value within the tolerance field. The tolerance field and the nominal value are determined by the requirements of the relevant standards and design documentation.

The data for creating neural network models are presented in the form of diameter deviations of 50 consecutively machined details of the shaft type ∅50h11 made of St45 steel within the tolerance field of a controlled size of 200 μm [26]. Thus, the time series of diameter deviations contains 25 values for each of the 2 realizations of the machining process obtained between machine tool adjustments under the same roughing modes (Table 1).

Table 1. Deviation of the controlled size of details of the type shaft ∅50h11 from the nominal value of y, μm [33]

Detail number	Implementation No. 1	Implementation No. 2
1	24	38
2	36	49
3	35	55
4	44	61
5	50	76
6	55	80
7	76	71
8	75	88
9	63	93
10	84	85
11	88	105
12	80	90
13	103	101
14	90	110
15	100	92
16	105	133
17	91	125
18	129	128
19	125	152
20	115	143
21	142	166
22	149	167
23	158	165
24	183	169
25	185	173

To build neural network forecasted models, examples of the training sample were obtained using the sliding window method x_i and x_{i+1}, which moves along the time sequence of retrospective data of deviations of the controlled parameter (Table 1) with a step equal to one processing cycle (one detail). In this case, the data in the window x_i are the inputs of the neural network, and the data of the second window x_{i+n} are the outputs.

Thus, training samples are instantaneous samples of values of the controlled parameter of sequentially machined details, represented as a time series shifted relative to the initial values x_i with a lag of one machining cycle, which corresponds to the process (1).

When forming examples of training sample, it is advisable to divide the time series of the forecasted indicator by x_i into $n = 5,\ldots,20$ values, as this range characterizes the volume of instantaneous sampling of deviations of details dimensions from nominal values, accepted in mechanical engineering [13; 34]. Taking into account this range and the total volume of realizations No. 1 and No. 2, which is equal to 25 values of deviations of details dimensions, a training set with $n=6$ inputs corresponding to the values of a sequentially shifted (by five levels) time series of deviations of details dimensions was created. In this case, the output $m=1$ determines the "reference" value of the deviation of the size of each subsequent detail – y. Thus, the number of examples (facts) of the training sample from implementation No. 1 is $25-6=19$ (Table 2).

Table 2. Training sample based on data from implementation No. 1

Example number	x_1	x_2	x_3	x_4	x_5	x_6	y
1	24	36	35	44	50	55	76
2	36	35	44	50	55	76	75
3	35	44	50	55	76	75	63
4	44	50	55	76	75	63	84
5	50	55	76	75	63	84	88
6	55	76	75	63	84	88	80
7	76	75	63	84	88	80	103
8	75	63	84	88	80	103	90
9	63	84	88	80	103	90	100
10	84	88	80	103	90	100	105
11	88	80	103	90	100	105	91
12	80	103	90	100	105	91	129
13	103	90	100	105	91	129	125
14	90	100	105	91	129	125	115
15	100	105	91	129	125	115	142
16	105	91	129	125	115	142	149
17	91	129	125	115	142	149	158
18	129	125	115	142	149	158	183
19	125	115	142	149	158	183	185

To obtain forecaster estimates of the accuracy of the technological process, a second sample of 19 examples was prepared, characterizing the values of deviations in the shaft diameter of each consecutively manufactured detail from implementation No. 2 (Table 3).

Table 3. Forecasted sample based on the data from implementation No. 2

Example number	x_1	x_2	x_3	x_4	x_5	x_6	y
1	38	49	55	61	76	80	71
2	49	55	61	76	80	71	88
3	55	61	76	80	71	88	93
4	61	76	80	71	88	93	85
5	76	80	71	88	93	85	105
6	80	71	88	93	85	105	90
7	71	88	93	85	105	90	101
8	88	93	85	105	90	101	110
9	93	85	105	90	101	110	92
10	85	105	90	101	110	92	133
11	105	90	101	110	92	133	125
12	90	101	110	92	133	125	128
13	101	110	92	133	125	128	152
14	110	92	133	125	128	152	143
15	92	133	125	128	152	143	166
16	133	125	128	152	143	166	167
17	125	128	152	143	166	167	165
18	128	152	143	166	167	165	169
19	152	143	166	167	165	169	173

Given the fact that multilayer neural networks with only one hidden layer and a sigmoidal activation function can perform any nonlinear mapping between two finite-dimensional spaces with a given accuracy, we will determine a sufficient number of hidden neurons [35]. At the same time, we note that in the on-line mode, during the sequential manufacture of each detail, the training sample size increases by one example. Thus, when training MLP models to obtain a forecaster estimate of the controlled indicator of detail No. 25 from implementation No. 2 (Table 1), the training sample size will be equal to $K = 19 + 18 = 37$. Using the values of K, n, and m, we will determine the minimum $L_{\min}$ and maximum $L_{\max}$ number of neurons in the hidden layer based on the dependencies presented in [36]

$$L_{\min} = 2 \cdot \sqrt{n\,m}, \quad L_{\max} = 0.5 \cdot (n + m) + 2 \cdot \sqrt{K}. \tag{2}$$

In accordance with dependencies (2), the total number of neurons in the hidden layer was calculated as the arithmetic mean between $L_{(\min)} \approx 4.9$ and $L_{(\max)} \approx 15.7$. Thus, computational experiments are performed using MLP models with a 6:10:1 architecture and sigmoidal neuronal activation functions.

To verify the obtained results of neural network forecasting, the following statistical criteria are used: mean absolute percentage error *MAPE* (3); root mean square error *RMSE* (4); minimum and maximum relative approximation error δ (5); coefficient of determination $D(6)$; correlation coefficient $R(7)$; variance ratio $S(8)$:

$$MAPE = \frac{100}{N} \cdot \sum_{i=1}^{n} \frac{\left| y_i^{\text{out}} - y_i \right|}{y_i}, \tag{3}$$

where y^{out}, y are respectively the forecasted and actual values of the i-th example, $i = 1, \ldots, n$, $N = 19$;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i^{\text{out}} - y_i)^2}{N}}\,; \tag{4}$$

$$\delta_{\min} = \frac{RMSE}{y_{\max}}\cdot 100,\;\; \delta_{\max} = \frac{RMSE}{y_{\min}}\cdot 100\,; \tag{5}$$

$$D = \frac{\left[N\cdot\sum_{i=1}^{N}(y_i^{\text{out}}y_i) - \sum_{i=1}^{N}y_i\cdot\sum_{i=1}^{N}y_i^{\text{out}}\right]^2}{\left[N\sum_{i=1}^{N}y_i^2 - \left(\sum_{i=1}^{N}y_i\right)^2\right]\left[N\sum_{i=1}^{N}{y_i^{\text{out}}}^2 - \left(\sum_{i=1}^{N}y_i^{\text{out}}\right)^2\right]}\,; \tag{6}$$

$$R = \sqrt{D}\,; \tag{7}$$

$$S = \frac{\sigma_\Delta}{\sigma_y}\,, \tag{8}$$

σ_Δ – standard deviation of the forecast error; $\Delta = (y - y^{\text{out}})$ – forecast error; σ_y – standard deviation of the forecasted indicator.

CONDUCTING COMPUTATIONAL EXPERIMENTS USING NEURAL NETWORK MODELING METHODS

When creating MLP models, we used STATISTICA 10, a system for statistical data analysis and forecasting, as well as BrainMaker Professional 3.52, a system for modeling neural networks. The use of different software during computational experiments is due to the need to ensure the reproducibility of the forecasting results. To ensure the convergence of the forecasting results and their verification assessment, two series of computational experiments were performed in STATISTICA and BrainMaker Professional. Thus, the training of neural network models was repeated twice for each sequentially manufactured detail from implementation No. 2 (Table 3). In this case, the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm was used in the STATISTICA system, and the Back Propagation of Error algorithm was used in the BrainMaker Professional system. It should be noted that these machine learning algorithms are iterative gradient methods of numerical optimization designed to find local extrema of a nonlinear transformation function by minimizing the MLP error and obtaining the desired output — y.

Using the BFGS algorithm in the Automated Neural Networks module of the STATISTICA 10 system, one neural network model out of 50 MLP models was automatically selected for each detail from implementation No. 2 according to the criterion of minimum training and testing error. An example of the interface of the created forecasted MLP-model for the first forecasted example (Table 3), i.e. the 7th detail from realisation No. 2 (Table 1) in the STATISTICA system is shown in Fig. 1.

When using the Back Propagation of Error algorithm in the BrainMaker system, the accuracy tolerance parameter for training neural network models was set to $\text{TOL} = 0.1$. Analysis of the results of training the MLP model at the number of epochs $Run = 163$ for the 7th detail from implementation No. 2 (Table 1) and the value of $RMS \approx 0.070$ indicates high accuracy of neural network training (Fig. 2).

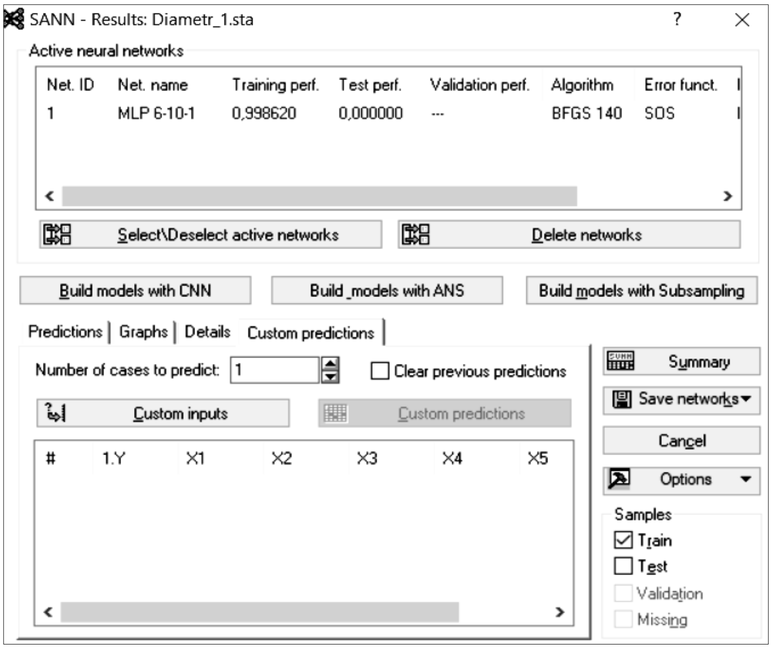


Fig. 1. Interface of the MLP model with 6:10:1 architecture created in STATISTICA 10 for the 7th detail of implementation No. 2

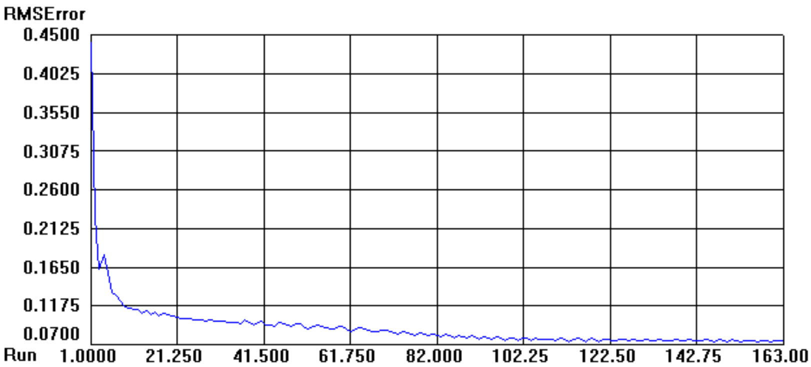


Fig. 2. Graph of changes in the *RMS Error* of training the MLP model in BrainMaker Professional 3.52 for the 7th detail from implementation No. 2 of the first series of computational experiments

An example of the interface of the trained and tested MLP model for the 7th detail from implementation No. 2, which shows the absence of unrecognized facts (Bad=0) in the BrainMaker system, is shown in Fig. 3.

To forecast the controlled parameter of the machining process of the 8th detail from implementation No. 2, the MLP model was trained using sample examples (Table 2) with the first fact from Table 3 added to it. Using this fact in the training sample allows us to implement the principle of simulation of the neurocontroller's functioning and continue the process of training the neural network model on-line.

Thus, by the method of sequentially adding facts to the training set, 19 MLP models with a 6:10:1 architecture were created in two series of computational experiments using the gradient learning algorithms BFGS and Back Propagation of

Error for all — from the 7th to the 25th consecutively manufactured details from implementation No. 2 (Table 3).

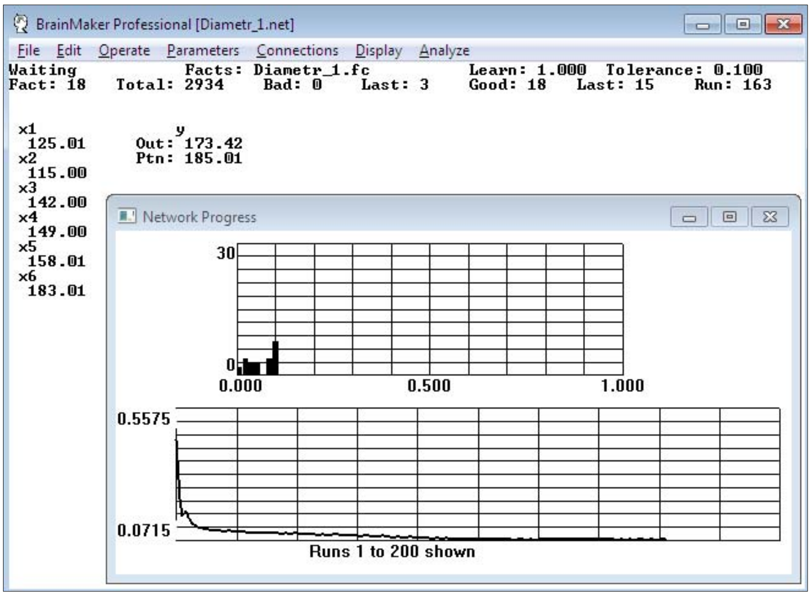


Fig. 3. The interface of the MLP model with 6:10:1 architecture created in BrainMaker Professional 3.52 for the 7th detail of the implementation No. 2 of the first series of computational experiments

RESULTS OF COMPUTATIONAL EXPERIMENTS

The results of forecasting the value of the controlled parameter for the 7th detail from implementation No. 2, obtained MLP models trained in two series of computational experiments using the BFGS algorithm are shown in Fig. 4 and Fig. 5, respectively.

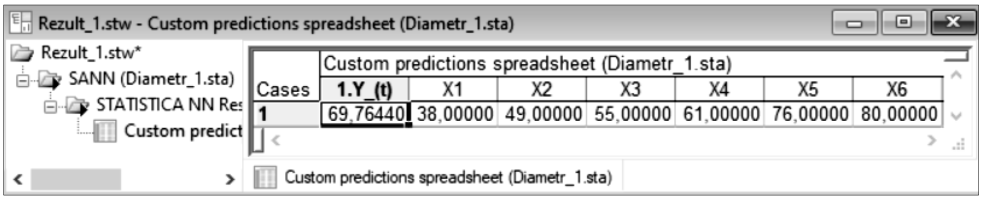


Fig. 4. The result of forecasting the controlled parameter of the 7th detail from implementation No. 2, obtained in STATISTICA 10 for the first series of computational experiments

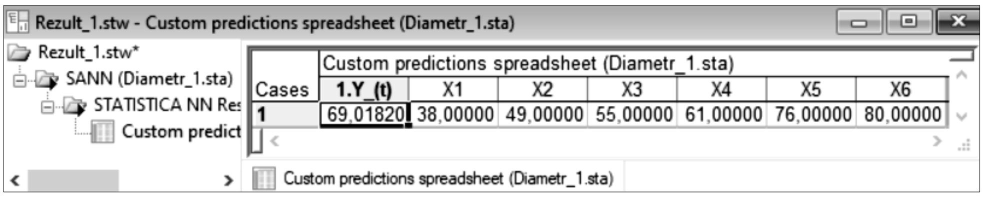


Fig. 5 The result of forecasting the controlled parameter of the 7th detail from implementation No. 2, obtained in STATISTICA 10 for the second series of computational experiments

The results of the forecasting in two series of computational experiments using the Back Propagation of Error algorithm for the 7th detail from implementation No. 2 are shown in Fig. 6 and Fig. 7, respectively.

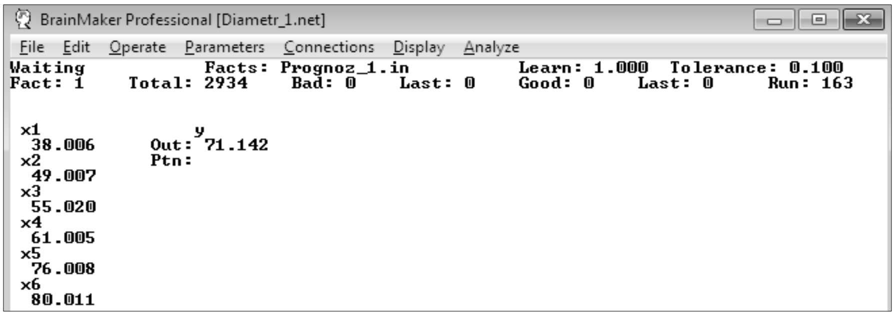


Fig. 6. The result of forecasting the controlled parameter of the 7th detail from implementation No. 2, obtained in BrainMaker Professional 3.52 for the first series of computational experiments

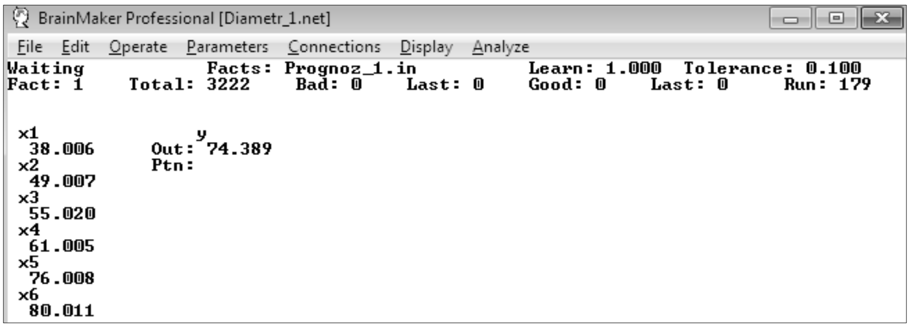


Fig. 7. The result of forecasting the controlled parameter of the 7th detail from implementation No. 2, obtained in BrainMaker Professional 3.52 for the second series of computational experiments

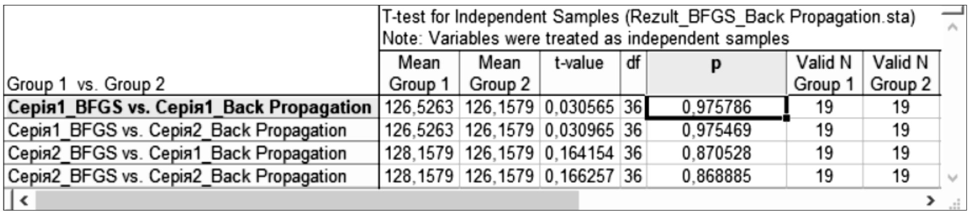
The obtained forecaster estimates of the deviation of the controlled size of shaft type details ∅50h11 from the nominal value for all sequentially manufactured details from the 7th to the 25th in two series of computational experiments conducted using MLP models with 6:10:1 architecture and trained by BFGS and Back Propagation of Error algorithms are shown in Table 4.

Table 4. The results of forecasting the controlled size of details of the type shaft ∅50h11 from the nominal value of y^{out}, μm

Example number	**Algorithm BFGS**		**Algorithm Back Propagation of Error**	
	Series 1	Series 2	Series 1	Series 2
1	70	69	71	74
2	80	83	79	77
3	84	90	81	87
4	88	90	88	88
5	103	102	104	101
6	92	86	98	96
7	100	103	99	108
8	121	118	115	111
9	96	103	109	109
10	133	123	128	121
11	145	151	125	128
12	112	118	126	129
13	152	152	155	158
14	168	172	146	145
15	166	166	172	172
16	180	183	173	173
17	168	173	172	169
18	173	174	179	175
19	173	179	177	176

DISCUSSION OF THE OBTAINED RESULTS

The reproducibility of the obtained forecasting results is confirmed by testing the statistical hypothesis that there is no significant difference between the forecasting results using different machine learning algorithms for two series of experiments (Table 4), which was performed on the basis of a *t*-test for independent variables, since the condition $p > 0.05$ is met (Fig. 8).



T-test for Independent Samples (Rezult_BFGS_Back Propagation.sta)
Note: Variables were treated as independent samples

Group 1 vs. Group 2	Mean Group 1	Mean Group 2	t-value	df	p	Valid N Group 1	Valid N Group 2
Серія1_BFGS vs. Серія1_Back Propagation	126,5263	126,1579	0,030565	36	0,975786	19	19
Серія1_BFGS vs. Серія2_Back Propagation	126,5263	126,1579	0,030965	36	0,975469	19	19
Серія2_BFGS vs. Серія1_Back Propagation	128,1579	126,1579	0,164154	36	0,870528	19	19
Серія2_BFGS vs. Серія2_Back Propagation	128,1579	126,1579	0,166257	36	0,868885	19	19

Fig. 8. Screenshot of the result of testing the statistical hypothesis that there is no significant difference between the forecasting results for two series of experiments in STATISTICA 10

The convergence of the two series of neural network forecasting results based on the machine learning algorithms BFGS and Back Propagation of Error (Table 4) is confirmed by the significant pairwise correlation coefficients $R_{\text{BFGS}} = 0.993$ and $R_{\text{Back Propagation}} = 0.995$, respectively.

Verification of the obtained results of neural network forecasting of the accuracy of the technological process of details machining was carried out using statistical criteria (3)–(8) (Table 5).

Table 5. Estimates of statistical criteria for forecast verification

Algorithm		Criterion						
		MAPE, %	*RMSE*, μm	δ_{min}, %	δ_{max}, %	*R*	*D*	*S*
BFGS	Series 1	5.322	9.668	5.589	13.617	0.969	0.939	0.282
	Series 2	6.494	11.180	6.463	15.747	0.964	0.929	0.314
Back Propagation of Error	Series 1	4.767	6.886	3.981	9.699	0.984	0.969	0.198
	Series 2	5.051	6.856	3.963	9.656	0.983	0.966	0.197

Based on the values of the coefficient of determination *D* (Table 5), we calculated the correlation coefficients (7), the value of which allowed us to classify all the results obtained by *R* (Table 5) as "Strong" – a qualitative measure of statistical relationship according to the Chaddock scale (Table 6).

Table 6. Correlation between quantitative and qualitative estimates of the correlation coefficient according to the Chaddock scale [37]

Quantitative measure statistical connection	Qualitative measure statistical connection
$0<R<0.1$	None
$0.1<R<0.3$	Weak
$0.3 <R<0.5$	Moderate
$0.5<R<0.7$	Noticeable
$0.7<R<0.9$	Close
$0.9<R<0.99$	Strong
$0.99<R<1$	Functional

Thus, the analysis of the values of all statistical criteria (Table 5) allows us to recommend the use of the Back Propagation of Error algorithm when creating neural network models to forecast the quality of machining processes for details by the parameter of shaft diameter deviation from the nominal value. It should be noted that since the proposed methods of neural network forecasting are based on the use of machine learning algorithms that allow finding local minima in the nonlinear mapping $X \to Y$, the prospect of further research may be the joint application of neural network methods and evolutionary modelling, which include genetic algorithms [38; 40].

CONCLUSIONS

1. To obtain an operational forecasting of the quality of mechanical engineering technological processes by the parameter of accuracy of machining of shaft-type details, MLP neural network models with the BFGS and Back Propagation of Error training algorithms were developed using STATISTICA 10 and BrainMaker Professional 3.52 systems, respectively.

2. As a result of two series of computational experiments, it was found that the use of the Back Propagation of Error algorithm allows to obtain forecasted estimates of the controlled process parameter with significantly smaller ranges of the mean absolute percentage error, mean square error, relative approximation error and variance ratio criterion compared to the BFGS algorithm. At the same time, sufficiently large values of the coefficient of determination and significant estimates of the correlation coefficient were obtained for both algorithms.

3. It has been established that the use of the developed adaptive MLP-models with the Back Propagation of Error algorithm allows to obtain forecasted estimations of accuracy indicators of technological processes of shaft-type details machining with 90-96 % reliability, which is confirmed by the range of values of relative approximation error (5). Thus, the created MLP-models can be recommended for application in neurocontrollers or neuroemulators for formation of control actions and prevention of discrepancies of parameters of details at control of accuracy of process of their machining in a mode on-line.

REFERENCES

1. *Product quality. Quality assessment. Terms and definitions: DSTU 2925-94. [In force since 1996-01-01]*. K.: Gosstandart of Ukraine,1995, 32 c.
2. O. Cherniak, R. Trishch, R. Ginevičius, O. Nechuiviter, V. Burdeina, "Methodology for Assessing the Processes of the Occupational Safety Management System Using Functional Dependencies," *Integrated Computer Technologies in Mechanical Engineering – 2023, ICTM 2023. Lecture Notes in Networks and Systems*, vol. 996, Springer, Cham, 2024, pp. 3–13. doi: https://doi.org/10.1007/978-3-031-60549-9_1
3. R. Trishch, O. Cherniak, D. Zdenek, V. Petraskevicius, "Assessment of the occupational health and safety manage-ment system by qualimetric methods," *Engineering Management in Production and Services*, vol. 16, no. 2, pp. 118–127, 2024. doi: 10.2478/emj-2024-0017
4. E. Khomiak, R. Trishch, O. Zabolotnyi, O. Cherniak, L. Lutai, O. Katrich, "Automated Mode of Improvement of the Quality Control System for Nuclear Reactor Fuel Element Shell Tightness," *Information Technology for Education, Science, and Technics. Lecture Notes on Data Engineering and Communications Technologies, ITEST 2024*, vol. 1, pp. 79–91, Springer, Cham. doi: https://doi.org/10.1007/978-3-031-71801-4_7
5. P. Hovorov, A. Kindinova, R. Trishch, E. Khomiak, O. Cherniak, O. Katrych, "Management of Power Grid Modes in Conditions of High Heterogeneity," *2024 IEEE 5th KhPI Week on Advanced Technology (KhPIWeek), Kharkiv, Ukraine, 2024*. doi: 10.1109/KhPIWeek61434.2024.10878032

6. P. Hovorov, R. Trishch, V. Hovorov, E. Khomiak, O. Vasilevskyi, V. Kukharchuk, "Peculiarities of Voltage Quality Control in Power Supply and Lighting Systems of Cities," *2024 IEEE 5th KhPI Week on Advanced Technology (KhPIWeek), Kharkiv, Ukraine, 2024*. doi: 10.1109/KhPIWeek61434.2024.10877979
7. P. Hovorov, V. Hovorov, M. Khvorost, A. Kindinova, R. Trishch, "Comprehensive Solution of Issues of Voltage Regulation and Compensation of Reactive Power in Power Supply and Lighting Systems of Cities," *2024 IEEE 5th KhPI Week on Advanced Technology (KhPIWeek), Kharkiv, Ukraine, 2024*. doi: 10.1109/ KhPIWeek61434.2024.10877952
8. P. Hovorov, R. Trishch, R. Ginevičius, V. Petraškevičius, K. Šuhajda, "Assessment of Risks of Voltage Quality Decline in Load Nodes of Power Systems," *Energies*, 18(7), 1579, 2025. doi: https://doi.org/10.3390/en18071579
9. E. Khomiak, R. Trishch, J. Nazarko, M. Novotný, V. Petraškevičius, "Method of Quality Control of Nuclear Reactor Element Tightness to Improve Environmental Safety," *Energies*, 18(9), 2172, 2025. doi: https://doi.org/10.3390/en18092172
10. S.M. Belyi, "Influence of cutting modes on the accuracy and surface roughness during turning," *Collection of scientific papers of Khmelnytsky National University. Series: Technical sciences*, no. 4, pp. 28–34, 2023.
11. E.V. Shvetsov, I.Y. Kharchenko, "Ensuring the accuracy of holes during cutting," *Bulletin of the Donbass State Machine-Building Academy*, no. 1(65), pp. 92–97, 2021.
12. *Statistical methods of quality control and regulation. Terms and definitions: DSTU 3514-97. [Effective from 1997-07-01]*. K.: Gosstandart of Ukraine, 1997, 48 p.
13. I.I. Plaskin, *Optimization of technical solutions in mechanical engineering*. M.: Mashinostroenie, 1982, 176 p.
14. Y.V. Shramko, O.S. Volkov, "Analysis of the influence of workpiece installation errors on the accuracy of machining details," *Bulletin of Mechanical Engineering and Transport*, no. 2(26), pp. 55–60, 2022.
15. S.S. Fedin, N.A. Zubretska, *Evaluation and forecasting of industrial products quality using adaptive artificial intelligence systems: monograph*. K.: Interservice, 2012, 206 p.
16. A.A. Yudashkin, *Application of neural networks for the construction of adaptive control systems for technological processes: Candidate of Technical Sciences (PhD): 05.13.07*. Samara State Technical University (SSTU). Samara, 1994, 145 p.
17. Roheen Qamar, Baqar Ali Zardari, "Artificial Neural Networks: An Overview," *Mesopotamian Journal of Computer Science*, vol. (2023), pp. 124–133, 2023. doi: https://doi.org/ 10.58496/MJCSC/2023/015
18. F. Fan, J. Xiong, M. Li, G. Wang, "On interpretability of artificial neural networks: A survey," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 6, pp. 741–760, 2021. doi: 10.1109/TRPMS.2021.3066428
19. K.J. Hunt, D. Sbarbaro, R. Zbikowski, P.J. Gawthrop, "Neural networks for control systems - a survey," *Automatica*, vol. 28, issue 6, pp. 1083–1112, 1992. doi: https://doi.org/10.1016/0005-1098(92)90053-I
20. *Adaptive control of machine tools*; Edited by B.S. Balakshin. M.: Mashinostroenie, 1973, 688 p.
21. S.S. Volosov, Z.Sh. Geyler, *Product quality management by means of active control*. M.: Izdatelstvo standardov, 1989, 264 p.
22. M.S. Nevelson, *Automatic control of machining accuracy on metal-cutting machines*. L.: Mashinostroenie, 1982, 184 p.
23. S.V. Bilenko, *Increasing the efficiency of high-speed machining on the basis of approaches of nonlinear dynamics and neural network modeling: Dis....dr.tekhn.sci: 05.03.01*. K.-on-A., 2006, 331 p.
24. A.P. Nikishechkin, *Improving the quality of the adaptation process when changing technological parameters using a neural network apparatus: Candidate of Technical Sciences: 05.13.06*. M.: Stankin, 2002, 187 p.
25. P.D. Wasserman, *Neural Computing: Theory and Practice*. New York, NY: Van Nostrand Reinhold, 1989, 189 p.
26. N.A. Zubretskaya, S.S. Fedin, "Neural network forecasting of the accuracy of technological processes by the quality parameters of manufactured products," *Information Processing Systems*, issue 2, pp. 17–20, 2014.
27. S.S. Fedin, R.M. Trishch, "Quality management using neural network methods," *System management methods, technology and organization of production, repair and operation of cars*, issue 15, pp. 228–230. K.: NTU, TAU, 2003.

28. Hojin Cho et al., “MMP Net: A feedforward neural network model with sequential inputs for representing continuous multistage manufacturing processes without intermediate outputs,” *IISE Transactions*, 56(10), pp. 1058–1069, 2023. doi: https://doi.org/10.1080/24725854.2023.2242434
29. Izabela Rojek, “Technological process planning by the use of neural networks,” *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 31 (1), pp. 1–15, 2017. doi: https://doi.org/10.1017/S0890060416000147
30. Huang Jin, Jiang Pin Yu, Zhao Rujia, Shen Bing, “An approach to neural index models for scheduling based on the example of machining processes,” *J. Xi’an Jiaotong Univ*., 1996, pp. 12–18.
31. Wang Chaojun, Liu Yanming, “Milling optimization controller combining genetic algorithm and neural networks,” *Contr. Theory and Appl.*, vol. 16, no. 4, pp. 607–610, 1999.
32. Yang Zheyong, Zhang Da-wei, Yuang Tian, “Milling process control using a three-layer back-propagation neural network,” *J. Tianjin Univ. Sci. and Technol*., 2000, pp. 206–210.
33. S.S. Fedin, *Artificial Intelligence Systems and Data Analysis Technologies: Workshop*; 2nd ed. K.: Interservice, 2021, 848 c.
34. A.A. Matalin, *Technology of mechanical engineering*. L.: Mashinostroenie, 1985, 496 p.
35. S. Omatu, *Neurocontrol and its applications*; Book 2. M.: IPRZHR, 2000, 272 p.
36. I.S. Astakhova, A.S. Potapov, V.A. Chulyukov, *Artificial Intelligence Systems. A practical course: a textbook*. M.: Binom, Laboratory of Knowledge, 2008, 276 p.
37. Ugur Turan, “A Correlation Coefficients Analysis on Innovative Sustainable Development Groups,” *EUREKA: Social and Humanities*, 1(1), pp. 46–55, 2020. doi: https://doi.org/10.21303/2504-5571.2020.001130
38. D. Whitley, T. Starkweather, C. Bogart, “Genetic algorithms and neural networks: Optimizing connections and connectivity,” *Parallel Computing*, vol. 14, issue 3, pp. 347–361, 1990. doi: https://doi.org/10.1016/0167-8191(90)90086-O
39. *Z. Guo, R.E. Uhrig, “Use of genetic algorithms to select inputs for neural network,” Proceedings of International Workshop on Combinations of Genetic Algorithms and Neural Networks, COGAN-92, 1992*, pp. 223–234.
40. S.S. Fedin, “Improving the accuracy of neural network exchange rate forecasting using evolutionary modeling methods,” *System Research and Information Technologies*, no. 3, pp. 7–24, 2024. doi: https://doi.org/10.20535/SRIT.2308-8893.2024.3.01

Received 07.05.2025

INFORMATION ON THE ARTICLE

Serhii S. Fedin, ORCID: 0000-0001-9732-632X, National Transport University, Ukraine, e-mail: sergey.fedin1975@gmail.com

Oksana O. Romaniuk, ORCID: 0000-0001-9774-9875, Open International University of Human Development “Ukraine”, Ukraine, e-mail: knutdromanuk@gmail.com

Roman M. Trishch, ORCID: 0000-0002-9503-8428, Ukraine National Aerospace University “Kharkiv Aviation Institute”, Ukraine, e-mail: trich_@ukr.net

ПРОГНОЗУВАННЯ ЯКОСТІ ТЕХНОЛОГІЧНИХ ПРОЦЕСІВ МЕТОДАМИ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ / С.С. Федін, О.О. Романюк, Р.М. Тріщ

Анотація. Створено комплекс моделей прямошарових нейронних мереж для отримання оперативних прогнозів якості технологічних процесів машинобудування. Установлено, що використання алгоритму машинного навчання Back Propagation of Error дає змогу отримати прогнозні оцінки контрольованого параметра процесу металообробки зі значно меншими діапазонами середньої абсолютної відсоткової похибки, середньої квадратичної похибки, відносної похибки апроксимації та критерію дисперсійного відношення порівняно з алгоритмом BFGS. Показано, що запропоновані моделі нейронних мереж типу MLP можуть бути рекомендовані для практичного застосування під час управління точністю процесу механічної обробки деталей типу вал.

Ключові слова: деталі, машинне навчання, нейронна мережа, прогнозування, технологічний процес, точність, якість.

UDC 004.942
DOI: 10.20535/SRIT.2308-8893.2025.3.10

MATHEMATICAL MODELING OF INFORMATION DIFFUSION PROCESS BASED ON THE PRINCIPLES OF THERMAL CONDUCTIVITY

V.O. RETS, E.V. IVOHIN

Abstract. Information diffusion, a fundamental process underlying societal evolution and decision-making, shares intriguing analogies with thermodynamics. This paper presents a mathematical model that bridges these domains by proposing an analogy between thermodynamics and information theory. The study introduces a solved heat equation as a foundational framework to model information diffusion within societal contexts. The specified societal conditions embedded within the solved heat equation are central to this model. These conditions encapsulate the susceptibility of a society to assimilate new information, the constraints dictating the number and nature of available information sources, and the dynamics of information distribution characterized by its aggressiveness. The relationship between information diffusion and thermodynamics lies in their inherent propensity to seek equilibrium or optimal states. Leveraging this analogy, the solved heat equation becomes a potent tool to simulate the dynamics of information spread, analogous to the flow of thermal energy within physical systems. This work aims to stimulate further inquiry into the parallels between thermodynamics and information theory, presenting a theoretical framework and software implementation that open new avenues for understanding and modeling information diffusion dynamics within complex societal systems.

Keywords: information diffusion, heat equation modeling, partial differential equations, information propagation, temperature distribution, physical process analogy, information flow analysis, system analysis, mathematical modeling.

INTRODUCTION

In today's information world, the distribution of information is a key process that affects the evolution of society, decision-making paradigms, and technological progress. The goal of information is to empower and educate consumers and help them make choices that can affect everything from choosing a book to choosing a government. However, information varies in both quality and impact. In recent years, the number of accidental and intentional consequences of using social network platforms to spread misinformation has increased [1]. If we perceive news as a description of current, real and important events [2] that affect people [3], then fake news is a whole ecosystem of information that includes both the distribution of true information and the creation and distribution of misinformation [3, 4]. Fake news is used to create disinformation articles, hoaxes, rumors, parodies, incorrect editorials, incorrect facts, etc. Such a variety of goals, channels, sources and motivations makes it difficult to understand their spread.

In order to learn how to properly use information as a tool that would play in favor of its user, you need to be able to model it. To do this, you can present a mathematical model of information distribution based on predetermined parameters, such as:

- perception of new information by society;
- imitations on the number and "aggressiveness" of information sources;
- time limit during which the news distribution process takes place.

The diffusion approach can be singled out among the most well-known and meaningful approaches used in the research of information distribution processes [5, 6]. Using the results of research on the relationship between the equations of thermodynamics and information theory, it is possible to draw a parallel and build a diagram of heat distribution in a one-dimensional rod to model the distribution of information in a linear graph of social relations, considering the linear graph as a subgraph of the tree of the social hierarchy [7].

HEAT EQUATION

The heat conduction equation is a well-known fundamental differential equation in partial derivatives, which is widely used in physics, engineering and various scientific fields. It describes how the distribution of heat changes over time in a certain region of space. Mathematically, it relates the rate of temperature change to the spatial distribution of temperature and time.

The general form of the heat conduction equation in one-dimensional space is represented by the formula [8]:

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2},$$

where the function $u(x,t)$ represents the temperature or distribution of heat in the material at a certain place and at a certain time, t denotes a time variable, x denotes a spatial coordinate, α is the coefficient of thermal conductivity, which is a constant for each material that characterizes the rate of diffusion heat through the material.

FORMULATION OF THE PROBLEM

In the real world, it is appropriate to consider the process of information distribution with several sources that have certain limitations. These can be television channels, Internet blogs, or even just people who in one way or another notify people around them about news or facts known to them.

Having set the goal of raising public awareness of a certain news (which can be both true and deliberately incorrect), it is necessary to calculate the optimal location of information sources to achieve the desired level of awareness – based on pre-established constraints, among which:

- the maximum number of sources;
- "aggressiveness" of the source, i.e., how quickly the parts of society connected with it will receive new information [9];
- maximum allotted time.

In this case, knowing the data on the attitude and receptivity of the society to new information, it is possible to draw a parallel with the processes characteristic to that of thermodynamics [10] — the distribution of information in society can be formalized on the basis of the propagation of heat in the material. At the same time, its thermal diffusivity can be considered as susceptibility to new information, and the "aggressiveness" of the sources will be determined by the maximum temperature of the heating element. For simplicity of initial calculations, it can be assumed that information is not lost in the system, that is, there is no "forgetting" effect, which means no heat exchange of the material with the surrounding environment. To conduct the research, we will consider a linear graph of social relations, the mechanistic analogue of which will be a completely isolated one-dimensional rod with a thermal conductivity coefficient α. Then the achievement of the desired level of awareness can be considered as heating the rod at least by Δt degrees with the help of heating sources of temperature T_I evenly located in the rod for the maximum time $t_{\max}$ with the initial temperature of the rod equal to zero, which is in turn equal to the temperature of the environment.

PROPOSITION

Given the given restrictions (on the number of heating elements, their maximum temperature, and heating time), we can conclude that under certain conditions it will not be possible to heat the entire rod with one source, so it is worth considering splitting the rod into equal parts, in the center of each of which a heating element can be placed. Thus, by dividing the general problem of calculating the temperature distribution $U_L(x,t)$ along a rod of length L into equal subtasks of heating parts of the rod $U_i(x,t)$, $i=\overline{1,n}$, where n is the number of subtasks, we can study only the case of heating one of them to obtain a complete picture of the experiment:

$$U_L(x,t)=\begin{cases} U_1(x,t), & x\in[x_0,x_1]; \\ U_2(x,t), & x\in[x_1,x_2]; \\ \cdots\cdots\cdots & \cdots\cdots\cdots \\ U_n(x,t), & x\in[x_{n-1},x_n], \end{cases} \quad \forall i\in\overline{1,n} \;\; x_i-x_{i-1}=\frac{L}{n}.$$

Since the heating of a part of the rod occurs in the middle equally in both directions, the subtask is decomposed into two more identical problems of calculating $U_{i/2}(x,t)$ with the heating of a part of the rod on one side, $i=\overline{1,n}$. At the same time, for each $i=\overline{2,n}$, the boundaries of the neighboring problems U_{i-1} and U_i will not have heat exchange due to the same heat dynamics (due to the same size of the neighboring subtasks' definition regions and the same power of their heating elements), which means that in the $U_{i/2}$ problems, we can consider a completely isolated rod except for the heated side. Due to the completely identical conditions for each part of the rod, the $U_{i/2}(x,t)$ problems are identical, so their solution can be denoted as $u(x,t)$. The tasks will be as follows: find the solution to the equation

$$\frac{\partial u}{\partial t}=\alpha\frac{\partial^2 u}{\partial x^2}$$

with boundary conditions

$$u(0,t) = T_1, \quad \left.\frac{\partial u(x,t)}{\partial x}\right|_{x=L} = u_x(L,t) = 0,$$

and the initial condition $u(x,0) = 0$.

SOLUTION OF THE SUBTASK

First, let's determine the steady solution u_s, which determines the equilibrium state. By solving the differential equation

$$\frac{d^2 u_s}{dx^2} = 0$$

we have

$$u_s = Ax + B u(0) = B = T_1 u'(L) = A = 0 u_s = T_1 .$$

In this case, until the moment of equilibrium, $u(x,t)$ will be influenced by the function $\nu(x,t)$, then $u(x,t)$ can be written as:

$$u(x,t) = T_1 + \nu(x,t) . \quad (1)$$

We solve a similar problem with respect to ν:

$$\frac{\partial \nu}{\partial t} = \alpha \frac{\partial^2 \nu}{\partial x^2} ,$$

with boundary conditions

$$\begin{cases} \nu(x,t); \\ \left.\dfrac{\partial \nu(x,t)}{\partial x}\right|_{x=L} = \nu_x(L,t) = 0, \end{cases}$$

and initial condition

$$\nu(x,0) = -T_1.$$

Apply the method of separating variables

$$\nu(x,t) = X(x)T(t) . \quad (2)$$

Substitute into the equation and get

$$X(x)T'(t) = \alpha X''(x)T(t) , \quad \frac{X''}{X} = \frac{T'}{\alpha T} = -\lambda .$$

We have two equations:

$$X'' + \lambda X = 0 , \quad (3)$$

$$T' + \alpha\lambda T = 0. \quad (4)$$

Solve (3):

$$\begin{cases} X'' + \alpha X = 0, \\ X(0) = 0, \\ X'(L) = 0. \end{cases}$$

Considering the following intervals

$$1.\ \lambda = \beta^2 > 0 ,\ 2.\ \lambda = 0 ,\ 3.\ \lambda = -\gamma^2 < 0 ,$$

we obtain that the solutions are nontrivial only for the first interval, and we have the following values:

$$\lambda_n = \frac{\pi\left(n+\frac{1}{2}\right)}{L}, \quad n = 0,1,2,\dots .$$

Therefore, the solutions of (3) take the following form $X_n(x) =$ $= D_n \sin\frac{\pi\left(n+\frac{1}{2}\right)}{L}x$, $n = 0,1,2,\dots$.

Let's return to equation (4) $T' + \alpha\lambda T = 0$. The solution to this equation is as follows $T(t) = Ae^{-\alpha\lambda t}$.

For each $X_n(x)$ with an eigenvalue λ_n, we set $T_n(t)$ to $T_n(t) = A_n e^{-\alpha\lambda_n t}$.

Now we can write down the general form of the solution to (2):

$$\nu(x,t) = \sum_{n=0}^{\infty} A_n \sin\frac{\pi\left(n+\frac{1}{2}\right)x}{L} e^{-\alpha\lambda_n t}.$$

Consider the initial condition for finding A_n:

$$\varphi(x) = \sum_{n=0}^{\infty} A_n \sin\frac{\pi\left(n+\frac{1}{2}\right)x}{L} = -T_1.$$

Using the orthogonality property of the eigenfunctions of the Sturm–Liouville problem [11]:

$$A_n = \frac{X_n, \varphi}{X_n, X_n}$$

from where

$$A_n = -\frac{2T_1}{L}\int_0^L \sin\frac{\pi\left(n+\frac{1}{2}\right)x}{L}dx,$$

$$A_n = -\frac{2T_1}{L}\left(-\frac{2L\cos\frac{2\pi n+\pi}{2}-1}{\pi(2n+1)}\right) = \frac{4T_1\left(\cos\frac{2\pi n+\pi}{2}-1\right)}{\pi(2n+1)}.$$

Substituting this value into (1), we have a solution to the problem for half of one part of the rod:

$$u(x,t) = \frac{4T_1}{\pi}\sum_{n=0}^{\infty}\frac{\cos\frac{2\pi n+\pi}{2}-1}{2n+1}\sin\frac{\pi\left(n+\frac{1}{2}\right)x}{L}e^{-\alpha\frac{\pi\left(n+\frac{1}{2}\right)}{L}t} + T_1$$

APPLICATION OF THE SOLUTION

Substituting $t = t_{\max}$ into the resulting formula, we can obtain the function $u(x)$ of the rod temperature distribution at the maximum permissible time point. Since in the context of the task at hand, it is necessary to raise the temperature of the rod by at least Δt, i.e., from 0 to a certain $T_2 < T_1$, it is necessary to find such $x = x_{T_2}$, at which T_2 will be achieved. To do this, we can use the dichotomy method, which will give us half the lengths of the maximum size parts of the rod. After that, the partition length L_n can be found by the formula:

$$L_p = \frac{L}{2x_{T_2}} .$$

Further, we can perform additional optimization procedures using similar approaches:

- reduce the heating time so that the temperature at the point x_{T_2} does not rise above T_2.
- reduce the temperature of the heating element so that the temperature at the point x_{T_2} does not rise above T_2 during the specified time.

IMPLEMENTATION

In the course of the study, the described algorithm was implemented in Python with visualization using the methods of the Matplotlib package and two optimization options — time and maximum temperature.

As an example, let's take the problem of heating an aluminum rod with a length of 30 centimeters, a maximum temperature of 100º, and a maximum time of 2 minutes. After performing the described calculations, we get that in order to achieve the desired result, it is necessary to break the rod into 6 equal parts. The steps of simulating the heating process using the described software can be seen in Figs. 1–3.

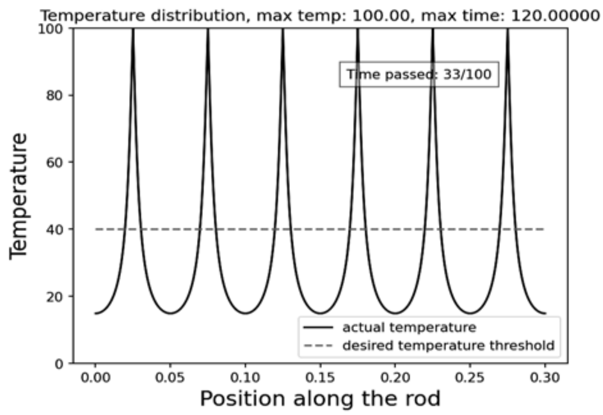


Fig. 1. The temperature distribution at the beginning of the simulation

Fig. 1 shows the temperature distribution at the beginning of the simulation, where heating took place for a third of the maximum time. It can be noted here that the distribution corresponds to the described model — at the heating points, the temperature is maximum from the beginning of the simulation, and in the neighboring regions it is distributed according to the distance from the heating elements. Also, the desired temperature along the rod is indicated in gray in the figures — it should be achieved within a given time.

In Fig. 2 shows the distribution at the time of two-thirds of the maximum time — the temperature along the entire rod has increased significantly, steadily approaching the desired level.

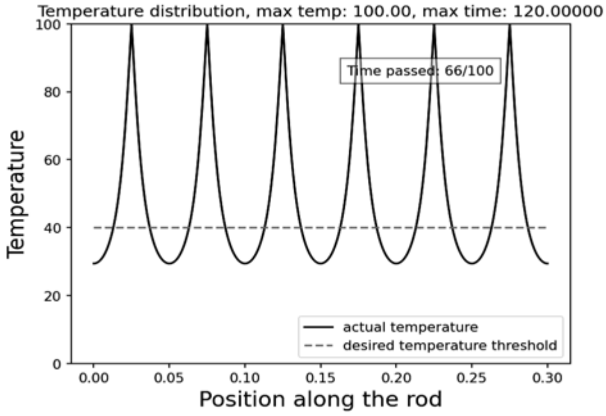


Fig. 2. The distribution at the time of two-thirds of the maximum time

Finally, Fig. 3 shows the final state — the desired temperature has been reached along the entire rod, which means that the program has completed execution.

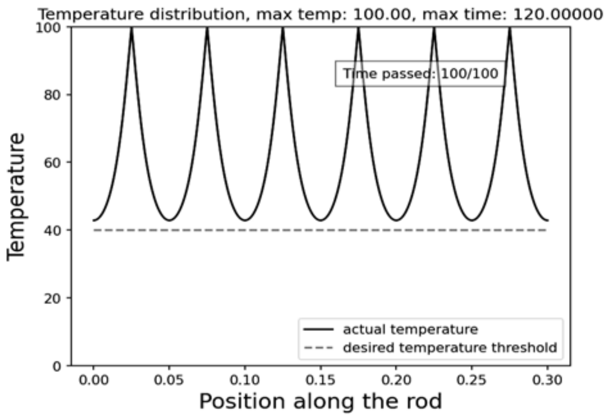


Fig. 3. The desired temperature has been reached along the entire rod

It should be noted that the temperature scale can be either in Kelvin or in Celsius, since the heating process in the context of the task depends only on the temperature difference between the heating source and the initial state, and the calculations are performed in the Celsius–Kelvin dimensional scale.

The proposed approach made it possible to study and compares the results obtained in the process of information distribution in social groups. Taking into account the assumed analogy between the process of heat exchange and information distribution, a comparison was made using the indicators of iron as a material whose thermal conductivity determines the part of society with a low level of perception of new information, and aluminum to describe the representatives of society who are prone to easy assimilation and perception of information. The length of the respective material was used to represent the conditional size of the studied society: 3 meters for small groups and 50 meters for large groups. The threshold value of heating corresponding to the level of awareness was set at 40º, and the heating source for the level of exposure at 100º. The maximum time is 20 minutes (see results in Table).

Table

Perception type \ Sources	The required number of sources	Conventional length of a separate zone of influence
High perception, small society	6	0.5
Low perception, small society	24	0.12
High perception, big society	93	0.54
Small perception, big society	389	0.13

CONCLUSION

This article investigates one of the approaches to the mechanistic application of models based on the use of the heat conduction equation to model and simulate the process of information multi-source distribution in a society. The problem is solved by analogizing a linear graph of social relations as a one-dimensional homogeneous rod, which makes it possible to find the required number of information sources to achieve the required level of selected information awareness in society within a given time. To conduct a simulation study of the found model, a software tool was created in Python and the Matplotlib visualization package, which finds the required number of sources under the specified conditions and conducts a simple visualization of the temperature distribution during the process of information distribution ("heating"). A qualitative improvement of the results obtained could be to perform similar calculations in two-dimensional space, which would allow for a more accurate modeling of connections in social groups. An important result should also be the study of the heating processes of a non-insulated rod or plate, for which the factor of information forgetting is added. This approach can be used to model real information diffusion processes to analyze existing news distribution, or to develop and simulate information campaigns for the effective distribution of desired information, such as advertising.

REFERENCES

1. "Digital Wildfires," *World Economic Forum*, 2018. Available: http://reports. wefo-rum.org/global-risks-2018/digital-wildfires/
2. James W. Kershner, *The Elements of News Writing*. Boston, MA: Pearson Allyn and Bacon, 2005.
3. Brian Richardson, *The Process of Writing News: From Information to Story*. Boston, MA: Pearson, 2007.
4. Claire Wardle, "Fake News. It's Complicated," *Medium*. Available: https://medium. com/1st-draft/fake-newsits-complicated-d0f773766c79
5. E.V. Ivohin, O.F. Voloshyn, M.F. Makhno, "Modeling of Information Dissemination Processes Based on Diffusion Equations with Fuzzy Time Accounting," *Cybernetics and Systems Analysis*, 57(6), pp. 896–905, 2021. doi: https://doi.org/10.1007/ s10559-021-00416-z
6. E.V. Ivokhin, L.T. Adzhubey, Y.A. Naumenko, M.F. Makhno, "On one approach to using of fractional analysis for hybrid modeling of information distribution proc-

esses," *System Research and Information Technologies*, no. 4, pp. 128–137, 2021. doi: https://doi.org/10.20535/SRIT.2308-8893.2021.4.10

7. M. Boguná, R. Pastor-Satorras, A. Díaz-Guilera, A. Arenas, "Models of social networks based on social distance attachment," *Physical review E*, 70(5), 056122, 2004. doi: 10.1103/PhysRevE.70.056122
8. John Rozier Cannon, "The one-dimensional heat equation," *Encyclopedia of Mathematics and its Applications (23)*. Cambridge University Press, 1984. doi: https://doi.org/10.1017/CBO9781139086967
9. M. MacKuen, "Exposure to information, belief integration, and individual responsiveness to agenda change," *American Political Science Review*, 78(2), pp. 372–391, 1984. doi: https://doi.org/10.2307/1963370
10. S. Diakonova, S. Artyshchenko, D. Sysoeva, I. Surovtsev, M. Karpovich, "On the application of the thermal conductivity equation to describe the diffusion process," in *E3S Web of Conferences*, vol. 175, p. 05050. EDP Sciences, 2020. doi: https://doi.org/10.1051/e3sconf/202017505050
11. M.A. Al-Gwaiz, *Sturm-Liouville theory and its applications*. Springer, 2008, 264 p. doi: https://doi.org/10.1007/978-1-84628-972-9

Received 15.03.2024

INFORMATION ON THE ARTICLE

Eugene V. Ivokhin, ORCID: 0000-0002-5826-7408, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: ivohin@univ.kiev.ua

Vadym O. Rets, ORCID: 0009-0007-8632-7010, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: vadym.rets@gmail.com

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ПРОЦЕСУ ПОШИРЕННЯ ІНФОРМАЦІЇ НА ОСНОВІ ПРИНЦИПІВ ТЕПЛОПРОВІДНОСТІ / В.О. Рець, Е.В. Івохін

Анотація. Поширення інформації являє собою фундаментальний процес, що лежить в основі суспільної еволюції та прийняття рішень і має доведені аналогії з термодинамікою. Запропоновано математичну модель, що дозволяє продемонструвати можливість застосування законів та принципів явища теплопровідності для моделювання процесів поширення інформації. Розглянуто оригінальну модель на основі скалярного рівняння у частинних похідних для формалізації процесів поширення інформації в суспільному контексті. Визначено характерні умови, які знайшли відображення у розв'язку розглянутого рівняння теплопровідності; відзначено сприйнятливість суспільства до засвоєння нової інформації, обмеження, що визначаються кількістю та властивостями доступних джерел інформації, а також динаміку поширення інформації, яка відзначається своєю агресивністю. Взаємозв'язок між поширенням інформації та термодинамікою полягає у властивій їм схильності шукати рівноважний або оптимальний стан. Ураховуючи цю подібність, розв'язане рівняння теплової енергії можна використовувати як потужний інструмент для моделювання динаміки поширення інформації, аналогічно потоку теплової енергії у фізичних системах. Ця робота має на меті стимулювати подальше дослідження паралелей між термодинамікою та теорією інформації, подаючи теоретичну основу та програмну реалізацію, що визначають нові шляхи для розуміння та моделювання динаміки поширення інформації в складних суспільних системах на основі використання різних моделей фізичних процесів.

Ключові слова: дифузія інформації, моделювання рівнянь теплопровідності, диференціальні рівняння в частинних похідних, поширення інформації, розподіл температури, аналогія фізичного процесу, аналіз інформаційних потоків, системний аналіз, математичне моделювання.

UDC 004.8:681.5:629.735
DOI: 10.20535/SRIT.2308-8893.2025.3.11

METHODS OF SWARM ARTIFICIAL INTELLIGENCE IN AUTONOMOUS NAVIGATION TASKS OF UAVS

M.Z. ZGUROVSKY, Yu.P. ZAYCHENKO, A.M. TYTARENKO, O.V. KUZMENKO

Abstract. This paper presents a comparative analysis of nine swarm intelligence (SI) methods in terms of their suitability for onboard AI platforms in autonomous unmanned aerial vehicle (UAV) swarms. A set of key criteria is defined, including computational complexity, scalability, latency, robustness to agent loss, and adaptability. Decentralized Behavior Trees (BTs) are identified as the most balanced approach for the reactive behavior layer, while the global swarm optimization method GBestPSO proves effective for high-level planning. A hybrid two-layer cognitive architecture is proposed that integrates BTs and GBestPSO, with functional separation between layers and communication based on DDS/RTPS protocols. The architecture exhibits high autonomy, fault tolerance, modularity, and suitability for real-time embedded systems operating in dynamic or adversarial environments. The results were partially supported by the National Research Foundation of Ukraine, grant No. 2025.06/0022 “AI platform with cognitive services for coordinated autonomous navigation of distributed systems consisting of a large number of objects”.

Keywords: swarm intelligence, UAV, autonomous navigation, behavior trees, GBestPSO, ROS 2, DDS, cognitive architecture.

INTRODUCTION

In the current environment of increasing complexity and dynamism in both military and civilian settings, autonomous swarms of unmanned aerial vehicles (UAVs) are increasingly being seen as an effective tool for reconnaissance, patrol, escort, and rapid response missions [1; 2]. Their advantages include high mobility, the ability to cover large areas, and the ability to operate in a decentralized mode [3]. However, to achieve true autonomy, each agent in the swarm must have the cognitive ability to perceive the environment, assess the situation, predict the consequences of its actions, and interact with other agents without centralized control [4].

Building such a system requires the deployment of an onboard AI platform with cognitive services capable of functioning in conditions of partial or complete loss of communication, in an informationally complex environment, or under the influence of electronic warfare [5]. Architecturally, this platform includes several functional components: a cognitive core, a sensory-analytical layer, a navigation controller, a communication module, a security module, a digital twin, and reconfiguration services [6]. The cognitive core plays a decisive role in this architecture, implementing the agent's intellectual subjectivity. It is based on the swarm artificial intelligence (AI) method, which forms mechanisms of perception, planning, coordination, and adaptation [7].

In this context, a key scientific and practical task arises – the choice of a swarm AI method suitable for implementation on board a drone. Such a method must meet strict limitations on computing resources (limited processor performance, small memory capacity), ensure real-time operation (low decision-making latency), support scaling to tens or hundreds of agents, be resistant to swarm element losses, and be adaptive to dynamic environmental changes [8].

In addition, the method must be integrated into the open-source modular platform Robot Operating System (ROS) using the Data Distribution Service (DDS) [9; 10], and must support working with simulation digital twins, which are critical for validating and training behavioural models in a safe virtual environment [11].

This challenge has no trivial solution. Traditional approaches — such as centralized planning, ant algorithms, and reinforcement learning methods — although highly effective in certain aspects, are usually overly resource-intensive, dependent on stable connectivity, or too inert to adapt to environmental changes [12; 13].

Therefore, this research focuses on finding, comparing, and justifying the most acceptable swarm AI method for embedding into an onboard AI platform with cognitive services, capable of not only ensuring the autonomous behaviour of an individual drone, but also implementing a holistic, adaptive, self-reconfiguring swarm system of a new generation.

CRITERIA FOR THE CHOICE OF SWARM METHODS OF AN ON-BOARD AI PLATFORM WITH COGNITIVE SERVICES

The successful implementation of autonomous swarm navigation for drones requires careful selection of a swarm artificial intelligence method that not only provides the necessary behavioural complexity but also meets the limitations of the onboard AI platform. The basis for this selection is a set of criteria that consider both the technical and functional requirements for an autonomous multindrone system. Below is a list of key criteria and their rationale:

1. **Computational complexity.** This criterion assesses how much the swarm AI method loads the processor when the number of agents increases. Quadratic or cubic complexity (e.g., $O(n^2)$, $O(n^3)$) significantly limits scalability and performance when running on embedded processors, especially in environments with limited computing resources (e.g., STM32, Raspberry Pi CM4, Jetson Nano). Methods with linear or logarithmic complexity ($O(n)$, $O(n \log n)$) scale better and are more suitable for decentralized autonomy [8; 18].

2. **Memory Requirements.** Methods that require large buffers, tables, and historical data are not suitable for implementation on compact boards without additional GPU or expanded memory. The total amount of memory required to store internal variables, sensor maps, behavioural patterns, etc. is considered [6].

3. **Latency.** Latency is critical for real-time applications: the system must respond to changes in the environment or commands with a delay of no more than

50–70 ms. Methods with high latency do not allow the drone to manoeuvre safely or make timely decisions in combat or rescue operations [13].

4. **Scalability.** This indicator determines how effectively the method maintains performance as the number of agents increases. For example, the Boids method scales well to 100+ agents, while ACO scales to no more than 30. High scalability is a prerequisite for complex scenarios with dozens or hundreds of drones [3].

5. **Robustness to Agent Loss.** In military and unstable environments, drones may be lost. The swarm AI method must automatically adapt to a decrease in the number of swarm members. Centralized approaches or Leader-Follower models are vulnerable to such losses, while decentralized behaviour trees (BTs) remain functional even when individual nodes are lost [17].

6. **Environmental Adaptivity.** True autonomy requires the ability to respond to changes in the environment: obstacles, changes in terrain, dynamic threats. Methods with low levels of sensory integration (e.g., classical PSO) have limited adaptability. In contrast, behavioural trees or DRL with sensory context integration are capable of dynamically reconfiguring behaviour [14].

7. **Onboard Real-Time Suitability.** This criterion refers to the ability of the algorithm to operate without the need for an external computing centre or cloud services. The method must function autonomously on the drone's hardware platform, considering real time, energy efficiency, and resource limitations [9]. Algorithms that already have known libraries under ROS 2, RTOS, or support DDS protocols are particularly valued.

These seven criteria form the basis of a multi-criteria model for evaluating and selecting a swarm AI method. They not only formalize the decision-making process, but also directly influence the architecture of the AI platform, its stability, adaptability, and viability in a real-world environment.

ANALYSIS AND SELECTION OF SWARM METHODS FOR USE IN AN AI PLATFORM WITH COGNITIVE SERVICES

Developing an effective AI platform for an autonomous swarm of unmanned aerial vehicles involves choosing a swarm method that not only meets functional requirements (adaptation, coordination, decentralization) but also complies with the strict limitations of the real environment: limited computing resources, the need for real-time decision-making, loss of communication, or individual agents. That is why it is crucial to compare existing methods in terms of computational complexity, memory requirements, latency, scalability, resilience to agent loss, adaptability, and suitability for onboard implementation (Table 1).

Classic swarm optimization methods, particularly Method 1 and its global modification (Method 9), demonstrate moderate complexity ($O(n^2)$) and scale well to 50 agents. They are suitable for search and global positioning tasks, but depend on one or more leader agents, which is a vulnerability in real-world environments.

Method 2, on the other hand, shows better adaptation to complex environments through a pheromone amplification mechanism, but has significantly higher complexity and latency, which makes it unsuitable for real-time tasks on typical onboard processors.

Both methods (1 and 2) are suitable for centralized or periodic optimization tasks but are limited in scenarios where continuous adaptation to the external environment is required.

Methods 3 and 5 are characterized by the lowest computational complexity ($O(n)$), high scalability (100+ agents), and low latency (<30 ms), which makes them technically attractive. However, they exhibit limited cognitive ability: Boids does not support goal-oriented planning, and stigmergy requires precise signal tuning and has limited adaptation to unpredictable scenarios.

Table 1. Comparative table of swarm AI methods

N	Swarm AI method	Computational complexity	Memory Requirements	Latency	Scalability	Robustness to Agent Loss	Environmental Adaptivity	Onboard Real-Time Suitability
1	Particle Swarm Optimization (PSO) [16]	Medium ($O(n^2)$)	Medium	< 50 ms	Good (up to ~50 agents)	Medium	Limited	Yes
2	Ant Colony Optimization (ACO) [3]	High ($O(n^2)$ log n)	High	100–200 ms	Limited (up to 30 agents)	High	Good	Limited
3	Boids (Reynolds' Rules) [21]	Low ($O(n)$)	Low	< 20 ms	High (100 + agents)	Low	Low	Yes
4	Consensus-based Algorithms [19]	Medium ($O(n \log n)$)	Medium	50–100 ms	Medium (up to 50 agents)	Medium	Medium	Yes
5	Stigmergy-based Models [20]	Low ($O(n)$)	Low	< 30 ms	High (100+ agents)	High	Good	Yes
6	Deep Reinfor-cement Learning (DRL) [22]	High ($O(n^3)$)	High	> 200 ms	Limited (20–30 agents)	Medium	High	No
7	Decentralized Behaviour Trees (BTs) [18]	Medium ($O(n \log n)$)	Medium	30–70 ms	High (50–100 agents)	High	High	No
8	Leader-Follower [17]	Low ($O(n)$)	Low	< 20 ms	Limited (~10–20 drones in classic implementation)	Low–medium (loss of leader → critical risk)	Limited (mainly in modified versions)	Yes, especially with low data intensity
9	Global swarm optimization method (GBestPSO) [16,23]	Medium ($O(n^2)$)	Medium	< 50 ms	Good (up to 50 agents)	Medium (dependence on leader)	Limited	Limited

Consensus-based algorithms (Method 5) ensure consistency within the swarm with moderate complexity and delays. However, they also do not allow modeling complex agent behavior or changing the mission structure during its execution. In turn, the Leader-Follower method, despite its simple implementation and low latency, has a fatal flaw – critical dependence on the leader. If the leader is lost, coordinated behavior is destroyed, which is unacceptable for combat or emergency scenarios.

The deep reinforcement learning method (Method 6), on the contrary, demonstrates the highest level of adaptability: models can self-learn and generalize knowledge about new environments. However, the inference delay exceeds 200 ms even on easy tasks, and the hardware requirements include a GPU (Jetson Xavier, Orin, or Nvidia RTX), which is beyond the capabilities of typical onboard platforms.

The most balanced method in terms of all criteria is the decentralized behavior tree method (Method 7). It has moderate computational complexity ($O(n\log n)$), operates with an average latency of 30–70 ms even on weak processors, scales up to 100 agents, demonstrates high adaptability, resilience to losses, and is suitable for implementation in ROS 2 with DDS protocols. BTs provides autonomous decision-making logic for each agent based on local context, sensor information, and a predefined behavior structure. In addition, it supports tree reconfiguration during a mission, which is especially important for tasks with variable goal or role structures in a swarm system.

Therefore, although Method 7 is the best choice for building the basic cognitive architecture of the agent, it is advisable to use a hybrid approach in which the global swarm optimization method (Method 9) plays a supporting role in global or semi-global optimization tasks. This division of functions allows BTs to be responsible for reactive behavior, real-time decision-making, and fault tolerance, while GBestPSO is responsible for strategic tasks, such as finding optimal agent locations, distributing subtasks, or optimizing configuration at the beginning of a mission.

This combination allows for the implementation of a two-level cognitive architecture: behavioral level – BTs, planning level – GBestPSO. Interaction between these levels can be carried out via DDS messages, which will ensure deterministic exchange between autonomous agents without loss of synchronization.

Figure 1 illustrates the two-level cognitive architecture of an autonomous swarm system that combines two complementary approaches: behavior trees (BTs) at the behavioral level and global swarm optimization GBestPSO at the planning level. The architecture is designed to ensure autonomous, adaptive, and fault-tolerant behavior of the UAV swarm in dynamic environments and in the event of communication loss.

The lower block of the figure reflects the logic of the behavioral level, constructed in the form of a decision tree. The ROOT node initiates the execution of the tree, CONDITION defines the situational conditions (for example, the presence of an obstacle or loss of communication), and ACTION represents the drone's reactive actions in response to these conditions. This structure allows each agent to make decisions autonomously based on the local context. This level is implemented locally on board the drone and provides the ability to react immediately, dynamically reconfigure the tree, and operate in real time.

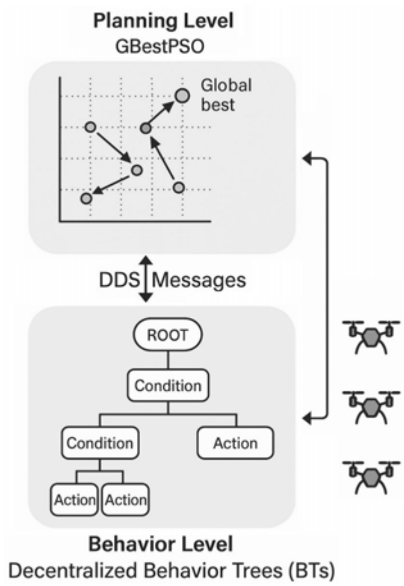


Fig. 1. Two-level cognitive architecture: planning level — GBestPSO, behavioral level — BTs

The upper block represents the planning level — global swarm optimization using the GBestPSO method, in which agents collectively search for optimal locations or configurations. Here, GLOBAL BEST means the best position found by all agents, which is transmitted to others as a strategic reference point. This benchmark is transmitted to the behavioral level in the form of mission conditions via DDS/RTPS protocols. In case of communication loss, the planning level can be disabled, and the behavioral level will continue to operate using the last received gBest.

The communication between levels is provided through DDS messages, which guarantees reliable, deterministic communication without centralized control. This integration allows for a flexible balance between the local autonomy of each drone and the global coordination of the entire swarm, ensuring high system efficiency in complex scenarios, including military or rescue operations.

Therefore, the proposed hybrid approach combines the modularity, adaptability, and speed of BTs with the global optimization capabilities of GBestPSO, creating a flexible, scalable, and technically feasible next-generation AI platform for UAV swarms.

THE BASIC ARCHITECTURE OF THE HYBRID APPLICATION OF METHODS BTS + GBESTPSO

To respond to the challenges of autonomous swarm navigation in conditions of limited resources and a highly dynamic environment, a hybrid two-level architecture has been proposed that combines the declarative-reactive approach of behavioral trees (BTs) with the evolutionary-optimization strategy of the global particle swarm (GBestPSO). Such integration allows for the simultaneous achievement of real-time performance, adaptability, scalability, and initial strategic coordination of swarm behavior.

The structure of hybrid architecture is divided into two functional levels:

- **Planning level (GBestPSO-Level):** implements global or semi-global optimization at the beginning of the mission or periodically in task replanning mode. It determines the selection of scenarios for current tasks, the strategic positions of agents, task clustering, and the distribution of roles or target points.
- **Behavioral level (BT-Level):** provides reactive decision-making by each agent in real time [18], including analysis of the local context, avoidance of obstacles, adaptation to role changes, loss of communication, or loss of an agent.

Communication between levels is carried out via DDS/RTPS protocols [9] in the form of semantically described messages (Table 2).

Data exchange between subsystems of two levels is organized via DDS/RTPS protocols [9], which allow semantically structured messages to be transmitted between agents. For example, the Swarm Optimizer module, which operates at the GBestPSO level, periodically calculates the globally best swarm location strategy (gBest) and transmits it via DDS as an optimized configuration or task. At the behavioral level, such information is interpreted as external mission conditions, which are automatically reflected in the updated behavioral logic through the Condition → Action tree branches. Importantly, even if communication with the planning level is lost, the behavioral level continues to autonomously execute the mission based on the last received strategic guideline. Mutual adaptation is ensured by QoS DDS channels, which provide priority delivery of important messages, such as changes in behavior tree statuses, sensor events, or strategic instructions. The above-mentioned relationships between modules can be represented by the following list:

T a b l e 2 . Main modules and functions of the basic hybrid architecture 7. BTs + 9. GBestPSO

Level	Module	Functions
BT level	Behavior Manager	Interprets behavior of trees in real time, activates local actions, reconfigures the tree in response to external changes
	Perception Module	Processing sensor data, recognizing situations, building a local map
	Communication Module	DDS/RTPS exchange of states, positions, parts of the BT tree
	Local Planner	Low-level planning (SLAM, obstacle avoidance, actions)
	Adaptation Module	Dynamic tree reconfiguration in case of environmental changes or losses
GBestPSO level	Swarm Optimizer	Implements a global PSO algorithm; forms a gBest position that sets a strategic landmark
	Global Task Allocator	Distribution of subtasks, goals, or roles among agents based on the gBest position
	Mission Planner (optional)	Scenario planning for complex missions or dynamic restarts
Both levels	Telemetry and Diagnostics	Status exchange, logging, communication with ground station

- The Swarm Optimizer module (at the GBestPSO level) periodically calculates the global best strategy (gBest) and transmits it via DDS in the form of tasks or configurations.
- The behavioral level (BT-Level) perceives these tasks as "external mission conditions" that are reflected in the corresponding branches of the behavior tree.
- Change in **gBest** → change in agent behavior through the reaction of **Condition → Action** nodes.
- Reaction to loss of communication: BT-Level works autonomously, with the last accepted gBest, allowing the agent to continue the mission without external optimization.
- Mutual adaptation is provided through QoS DDS channels, which prioritize behavioral commands, sensor events, tree statuses, and strategic instructions.

The software and hardware implementation of this architecture is based on ROS 2 (Foxy or Humble versions) and RTOS for real-time systems such as STM32H7. The communication layer is implemented using Fast DDS [9] or Cyclone DDS with QoS support. At the library level, BehaviorTree.CPP is used for BT modules and our own implementation of the GBestPSO algorithm based on the classical approach [14]. Jetson Orin Nano, Xavier NX, Raspberry Pi CM4 with Coral TPU, as well as lightweight STM32H7 controllers are used as hardware platforms. The main components of the software and hardware architecture are:

- Operating system: ROS 2 (Foxy/Humble), RTOS (on STM32H7).
- Communication: Fast DDS / Cyclone DDS (with QoS support).
- Libraries: BehaviorTree.CPP (BT-level), own implementation of GBestPSO (based on [14]).
- Platforms: Jetson Orin Nano / Xavier NX, Raspberry Pi CM4 with Coral TPU, STM32H7.

Among the advantages of the proposed architecture, its autonomy, flexibility, and fault tolerance should be noted. The behavioral level ensures local adaptability and the ability to respond to changes in the environment without the need for global control. The planning level, in turn, ensures strategic coordination between agents, forming a single target configuration of the system. The use of DDS protocols eliminates dependence on a central node, which significantly increases fault tolerance. In the event of a loss of the global reference point (gBest), the system automatically switches to fully local mode without losing functionality. The modular structure of this architecture ensures its scalability and adaptability to different mission classes: from reconnaissance and escort to combat or search and rescue operations. Thus, the main advantages of hybrid architecture are:

- BT-Level provides local adaptability and autonomy.
- GBestPSO-Level adds strategic coordination and global coordination.
- Interconnection via DDS eliminates dependence on a central node.
- High fault tolerance: if gBest is lost, the architecture switches to fully local mode.
- Modularity support allows architecture to be scaled for different mission classes (reconnaissance, escort, attack).

Thus, the proposed hybrid architecture, combining behavioral trees (BTs) and global particle swarm optimization (GBestPSO), demonstrates an effective combination of local autonomy and strategic coordination in swarm navigation tasks. Its two-level structure allows it to simultaneously achieve adaptability, responsiveness, and energy efficiency while maintaining a high degree of resilience to failures and communication losses. The behavioral level provides an immediate response to environmental dynamics, while the planning level defines global landmarks and roles, which increases the integrity of swarm behavior in complex conditions. The use of DDS/RTPS protocols allows for reliable, distributed communication between agents without centralized dependency. Such architecture not only has high application potential in the military and civilian spheres but also provides a scalable foundation for the further evolution of swarm intelligence systems towards the self-learning and self-coordinated platforms of the future.

ALGORITHM

The movement of each drone is described by two main phases:

1. **Phase 1: Approach to the Target.** Drones move toward target Y^* according to the velocity equation with the local best position. The influence of cognitive (C_1) and social (C_2) components depends on the distance to the leader (Y_L) and target, respectively:

$$v_{ij}(t+1) = v_{ij}(t) + C_1(d(Y_L(t), X_i(t)))r_1(t)[Y_{L_j}(t) - X_{ij}(t)] +$$
$$+ C_2(d(Y^*(t)X_i(t)))r_2(t)[Y_j^*(t) - X_{ij}(t)], \quad (1)$$

where coefficients C_1 and C_2 are linearly dependent on distance, which allows adjusting the influence of the leader and the target on the behavior of the swarm:

$$C_1(d(Y_L(t), X_i(t))) = \frac{C_{1\max} - C_{1\min}}{Y_L(0) - X_i(0)}(Y_{Lj}(t) - X_{ij}(t)) + C_{1\min}; \quad (2)$$

$$C_2(d(Y^*(t), X_i(t))) = \frac{C_{2\max} - C_{2\min}}{Y^*(0) - X_i(0)}(Y_j^*(t) - X_{ij}(t)) + C_{2\min}. \quad (3)$$

2. **Phase 2: Firing Ring for Killing.** After reaching the target's circumference with a radius of r_0, the drones switch to polar coordinates and begin to rotate

around the target. The coordinates of the leader Y_L are described using the radius r_0 and angle $\varphi_L(t)$:

$$\begin{cases} y_1^L(t) = r_0 \cos(\varphi_L(t)), \\ y_2^L(t) = r_0 \sin(\varphi_L(t)). \end{cases} \quad (4)$$

The angular position is updated using the following formula:

$$\varphi_L(t+1) = \varphi_L(t) + \omega_L(t)t, \quad (5)$$

where $\omega_L(t)$ — angular velocity of rotation.

Each drone is controlled by its own BT, which cyclically processes the tick. The tree consists of control nodes (Selector, Sequence) and execution nodes (Condition, Action). Behavior Tree diagram:

- **Root:** initiates tree execution.
- **Sequence:** main sequence of actions for each drone.
- **Parallel:** used for simultaneous execution of tasks independent of the main movement cycle, such as modeling the impact of electronic warfare and drone loss.
  - **Action: Model_Jamming()** — simulates the probability of communication loss (P_{10}).
- **Action: Model_Drone_Loss()** — simulates the probability of drone loss (P_2).
- **Selector (Main mission logic):** Checks the mission status and switches between phases.
  - **Branch 1: Approach to the Target**.
    - **Condition: Is_Far_From_Target(r_0)** – checks whether the distance to the target is $> r_0$.
    - **Action: Run_LBestCombinedPSO_Approach_Phase()** – calculates the new speed and position.
    - **Action: Elect_New_Leader()** – this node is activated in case of Failure of the **Is_Leader_Connected()** node, which implements the self-organization mechanism.
  - Branch 2: **Firing Ring for Killing.**
    - **Condition: Is_Close_To_Target(r_0)** – checks whether the distance to the target is $\le r_0$.
    - **Sequence: Firing_Ring_Sequence.**
      - **Action: Convert_To_Polar()** – transition to polar coordinates.
      - **Decorator: Sync_Attack()** – uses the Decorator
      - **Decorator: Sync_Attack()** – uses the Decorator mechanism for synchronization. The node allows the execution of the child node only after $|ready\_agents| \ge M_{Y^*}$.
      - **Action: Attack_Target()** – attack the target.

A key element of the system is its ability to be adaptive:

- **Loss of leader.** If the leader drone loses connection with other agents (simulated by the **Model_Jamming** node), this initiates **Failure** in the leader verification conditions. The **Selector** in BT automatically switches to the branch that triggers self-organization. A new leader is selected based on the criterion $Y_{Lnew}(t) = X_i(t)$, where $d(X_i(t), Y^*) = \min_i d(X_i(t), Y^*)$.

Synchronization. The **Decorator** node ensures that the ring firing phase does ot start until enough drones (M_Y) reach the target area, which is critical for coordinating the attack.

Inertial component. In the equation, the velocity $v_{ij}(t)$ acts as an inertial component, preventing sudden changes in trajectory and ensuring smooth movement.

EXAMPLE

Let us consider a practical simulation experiment using a two-level cognitive architecture consisting of a group of drones and integrating behavioral trees (BTs) and global optimization using the GBestPSO method. According to this experiment:

- **The behavioral level (BT-level)** functioned locally on board each drone and provided reactive adaptation to sensory events such as obstacle detection, role changes, or loss of communication. This level was implemented using the BehaviorTree.CPP library with a built-in dynamic tree reconfiguration mechanism. Condition → Action nodes allowed for the rapid transformation of behavioral logic in response to changes in external mission conditions received from the planning level.
- **The planning level based on GBestPSO** was responsible for forming the strategic configuration of the swarm through global or local optimization. It periodically calculated the globally best position or scenario (gBest), which was broadcast as strategic guidelines to each agent. Data transfer between levels was carried out using DDS/RTPS protocols, which guaranteed quality of service (QoS), buffering of critical messages, and resistance to temporary communication losses.
- **The simulation experiment** was conducted in a configuration of 5–10 drones equipped with Jetson Orin Nano or Raspberry Pi CM4 paired with STM32H7 (Table 3).

Table 3. Quantitative characteristics of the simulation experiment

No	Parameter	Value
1	Number of drones in the swarm group	5–10 (Jetson Orin Nano or Raspberry Pi CM4 + STM32H7)
2	Coverage area	100 × 100 m
3	Number of control points	6
4	Mission type	Search and patrol with reconfiguration elements
5	BTs algorithm	6–10 conditional nodes, 3–4 adaptive branches
6	GBestPSO algorithm	Classic method [23]
7	Number of PSO iterations before start	30
8	BT module response delay	~45 ms (on Jetson Nano)
9	gBest transmission delay via DDS	~10–15 ms
10	Tree adaptation time when changing gBest	< 150 ms
11	Simulated communication losses	up to 30% of DDS packets
12	Mission success rate	> 95% (all drones completed the route or selected fallback actions)

The mission covered an area of 100 × 100 m, contained six control points, and was of a search-and-patrol nature with elements of reconfiguration. The BTs algorithm included 6 to 10 conditional nodes and 3–4 adaptive branches responsible for changing behavior in conditions of uncertainty. The GBestPSO algorithm was implemented in its classical form [23] with typical parameters: space dimension $d = 2$, inertial weight $\omega = 0.8$, attraction coefficients $C_1 = C_2 = 1.497$. Before the start of the mission, 30 PSO iterations were performed to find the initial strategic configuration. The BT module's response delay to an event was about 45 ms, gBest transmission via DDS was 10–15 ms, and behavior tree adaptation when gBest changed took up to 150 ms. Even with a simulated loss of up to 30% of DDS messages, the mission success rate exceeded95%.

In critical situations, the system demonstrated high adaptability. In the event of a single drone failure, other agents automatically restructured the behavior tree to compensate for the loss. If the planning level lost communication, the BT modules continued to operate autonomously based on the last strategic reference point. When obstacles were detected, the SLAM module formed a new route, and the corresponding branch of the behavior tree activated avoidance actions. The total mission time exceeded the baseline scenario by no more than 12%. The average delay between events and response was about 90 ms. The positioning error in areas without GNSS remained within 2.8 m, and swarm synchronization was maintained even with the loss of up to 40% of DDS messages.

The diagram (Fig. 2) shows a typical scenario for the implementation of a UAV swarm mission for the proposed architecture with hybrid integration of BTs and GBestPSO. This implementation is viable and technically effective for complex search and patrol missions in conditions of limited communication and a dynamic environment.

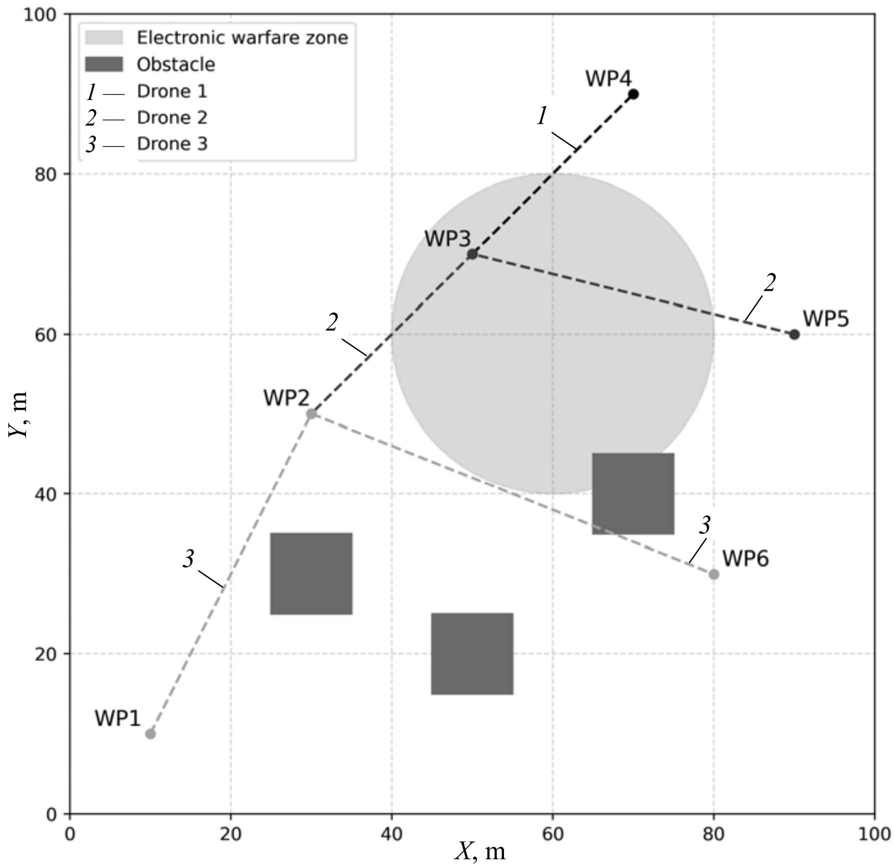


Fig. 2. Scenario for implementing the mission of a swarm of UAVs for the hybrid architecture of BTs and GBestPSO

The following images are used in Fig. 2: blue dots (WP1–WP6) – mission route control points; red circle – electronic warfare (EW) zone; gray squares — obstacles on the route (buildings, infrastructure objects); lines with markers — trajectories of three drones moving autonomously using behavioral trees (BTs), responding to obstacles and conditions broadcast from the global level of GBestPSO.

This example demonstrates the realistic implementation of a two-level cognitive architecture for swarm drones in a simulation environment with digital twin elements under threatening conditions and limited connectivity. It confirms the feasibility of the BTs + GBestPSO hybrid approach in complex, dynamic, or hostile conditions.

CONCLUSIONS

1. The article analyzes modern swarm artificial intelligence methods from the perspective of their suitability for implementation as part of an onboard AI platform for autonomous unmanned aerial vehicles. Considering the requirements for limited computing resources, real-time operation, resistance to agent loss, and adaptability, a list of key criteria for selecting a swarm AI method was formulated.

2. A comparative analysis of nine leading methods allowed us to identify decentralized behavioral trees (BTs) as the most balanced approach for the basic cognitive architecture of an agent. BTs combine low latency, resilience to losses, adaptability to the environment, and suitability for onboard implementation in ROS 2. At the same time, the GBestPSO method, as a classic global swarm optimization tool, proved to be useful for performing strategic tasks at the higher level of the system, for the initial configuration of the swarm, task distribution, and search for optimal agent locations.

3. The hybrid architecture proposed in the article, which combines BTs and GBestPSO into a two-level structure, allows for a compromise between local responsiveness and global coordination. This architecture is characterized by high autonomy, fault tolerance, scalability, and real-time adaptability. It provides a flexible division of responsibilities: BTs for immediate behavioral response, GBestPSO for planning optimization. The communication foundation, implemented through DDS/RTPS, eliminates dependence on central control and guarantees deterministic data exchange between agents.

4. A practical simulation experiment using digital twins and the BTs + GBestPSO hybrid architecture demonstrated its effectiveness in complex dynamic conditions. Even with the loss of up to 30% of DDS network messages, the system maintained coordinated swarm behavior, and the response time between a sensor event and tree restructuring averaged ~90 ms. Adaptation to new obstacles, role changes, resistance to agent failures, and the ability to operate without a global reference point demonstrated the high viability of the architecture. Mission success exceeded 95%, confirming the practical feasibility of the proposed approach.

5. Thus, the results of the study not only confirm the theoretical validity of hybrid architecture but also demonstrate its real implementation in conditions close to combat or emergency rescue situations. The proposed system can serve as a basis for building a new generation of swarm AI platforms with the potential for further evolution towards self-learning, self-coordinated, and safe-to-use autonomous systems in real environments.

REFERENCES

1. D. Floreano, R.J. Wood, "Science, technology and the future of small autonomous drones," *Nature*, 521(7553), pp. 460–466, 2015. doi: https://doi.org/10.1038/nature14542
2. M.J. Krieger, J.-B. Billeter, "The call of duty: Self-organised task allocation in a population of up to twelve mobile robots," *Robotics and Autonomous Systems*, 30(1–2), pp. 65–84, 2000. doi: 10.1016/S0921-8890(99)00065-2
3. M. Brambilla, E. Ferrante, M. Birattari, M. Dorigo, "Swarm robotics: a review from the swarm engineering perspective," *Swarm Intelligence*, 7, pp. 1–41, 2013. doi: 10.1007/s11721-012-0075-2
4. G.-Z. Yang et al., "The grand challenges of Science Robotics," *Science Robotics*, 3(14), eaar7650, 2018. doi: 10.1126/scirobotics.aar7650
5. Y. Mekdad, A. Arış, L. Babun, A. El Fergougui, M. Conti, R. Lazzeretti, S. Uluagac, "A survey on security and privacy issues of UAVs," *Computer Networks*, vol. 224, 109626, 2023. doi: 10.1016/j.comnet.2023.109626
6. B. Barricelli, E. Casiraghi, D. Fogli, "A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications," *IEEE Access*, pp. 167653–167671, 2019. doi: 10.1109/ACCESS.2019.2953499
7. V. Trianni, A. Campo, *Fundamental Collective Behaviors in Swarm Robotics*, pp. 1377–1394, 2015. doi: 10.1007/978-3-662-43505-2_71
8. H. Hamann, *Swarm Robotics: A Formal Approach.* Springer, Cham, 2018. doi: https://doi.org/10.1007/978-3-319-74528-2
9. Y. Maruyama, S. Kato, T. Azumi, *Exploring the performance of ROS2*, pp. 1–10, 2016. doi: 10.1145/2968478.2968502
10. "Data Distribution Service (DDS) Specification. Version 1.4.", *Object Management Group (OMG)*. 2015. Available: https://www.omg.org/spec/DDS
11. S. Boschert, R. Rosen, "Digital twin—The simulation aspect," in *Mechatronic Futures*, pp. 59–74. Springer, 2016. doi: 10.1007/978-3-319-32156-1_5
12. M. Dorigo, M. Birattari, M. Brambilla, "Swarm robotics and artificial swarm intelligence, in *Handbook of Artificial Intelligence*, 2014.
13. M. Hüttenrauch, A. Šošić, G. Neumann, "Deep Reinforcement Learning for Swarm Systems," *Journal of Machine Learning Research*, vol. 20, pp. 1–31, 2019.
14. M. Colledanchise, P. Ögren, *Behavior Trees in Robotics and AI: An Introduction.* CRC Press, 2018. doi: https://doi.org/10.1201/9780429489105
15. M. Dorigo, V. Maniezzo, A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE Transactions on Evolutionary Computation, 1(1)*, pp. 53–66, 1996. https://doi.org/10.1109/4235.585892
16. R.C. Eberhart, J. Kennedy, "A new optimizer using particle swarm theory," *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pp. 39–43. IEEE, 1995. doi: https://doi.org/10.1109/MHS.1995.494215
17. W. Li, H. Dong, H. Gao, "Leader-follower formation control of multi-UAVs with collision avoidance," *IEEE Transactions on Industrial Electronics*, 64(4), pp. 3184–3196, 2017. doi: https://doi.org/10.1109/TIE.2016.2638818
18. A. Marzinotto, M. Colledanchise, C. Smith, P. Ögren, "Towards a unified behavior trees framework for robot control," *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5420–5427. doi: https://doi.org/10.1109/ICRA.2014.6907656
19. R. Olfati-Saber, J.A. Fax, R.M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, 95(1), pp. 215–233, 2007. doi: https://doi.org/10.1109/JPROC.2006.887293
20. V. Parunak, M. Purcell, R. O'Connell, "Digital Pheromones for Autonomous Coordination of Swarming UAV's," in *1st UAV Conference, 2002*. doi: 10.2514/6.2002-3446

21. C.W. Reynolds, “Flocks, herds and schools: A distributed behavioral model,” *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH ‘87)*, pp. 25–34. doi: https://doi.org/10.1145/37401.37406
22. O. Vinyals et al., “Grandmaster level in StarCraft II using multi-agent reinforcement learning,” *Nature*, 575(7782), pp. 350–354, 2019. doi: https://doi.org/10.1038/s41586-019-1724-z
23. M. Zgurovsky, Y. Zaychenko, *The Fundamentals of Computational Intelligence: System Approach*. Springer, 2016. doi: https://doi.org/10.1007/978-3-319-35162-9

Received 24.08.2025

INFORMATION ON THE ARTICLE

Michael Z. Zgurovsky, ORCID: 0000-0001-5896-7466, Educational and Scientific Complex “Institute for Applied System Analysis” of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: zgurovsm@hotmail.com

Yuriy P. Zaychenko, ORCID: 0000-0001-9662-3269, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: zaychenkoyuri@ukr.net

Andrii M. Tytarenko, ORCID: 0000-0002-8265-642X, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: titarenkoan@gmail.com

Oleksii V. Kuzmenko, ORCID: 0000-0003-1581-6224, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: oleksii.kuzmenko@ukr.net

МЕТОДИ РОЄВОГО ШТУЧНОГО ІНТЕЛЕКТУ В ЗАВДАННЯХ АВТОНОМНОЇ НАВІГАЦІЇ БПЛА / М.З. Згуровський, Ю.П. Зайченко, А.М. Титаренко, О.В. Кузьменко

Анотація. Подано порівняльний аналіз дев’яти методів ройового інтелекту (РІ) з точки зору їхньої придатності для бортових платформ ШІ в автономних роях безпілотних літальних апаратів (БПЛА). Визначено набір ключових критеріїв, включаючи обчислювальну складність, масштабованість, затримку, стійкість до втрати агентів та адаптивність. Децентралізовані дерева поведінки (ДП) визначені як найбільш збалансований підхід для реактивного рівня поведінки, тоді як глобальний метод оптимізації рою GBestPSO виявляється ефективним для високорівневого планування. Запропоновано гібридну двошарову когнітивну архітектуру, яка інтегрує ДП та GBestPSO, із функціональним розділенням між шарами та зв’язком на основі протоколів DDS/RTPS. Архітектура демонструє високу автономність, відмовостійкість, модульність та придатність для вбудованих систем реального часу, що працюють у динамічних або змагальних середовищах. Результати частково підтримано Національним фондом досліджень України, грант № 2025.06/0022 «Платформа штучного інтелекту з когнітивними сервісами для скоординованої автономної навігації розподілених систем, що складаються з великої кількості об’єктів».

Ключові слова: ройовий інтелект, БПЛА, автономна навігація, дерева поведінки, GBestPSO, ROS 2, DDS, когнітивна архітектура.

ВІДОМОСТІ ПРО АВТОРІВ

Баран Данило Романович,
аспірант кафедри обчислювальної техніки факультету інформатики та обчислювальної техніки КПІ ім. Ігоря Сікорського, Україна, Київ

Васильєва Марія Давидівна,
старший викладач кафедри обчислювальної техніки факультету інформатики та обчислювальної техніки КПІ ім. Ігоря Сікорського, Україна, Київ

Голінко Ігор Михайлович,
доцент, кандидат технічних наук, доцент кафедри автоматизації теплоенергетичних процесів НН ІАТЕ КПІ ім. Ігоря Сікорського, Україна, Київ

Грішин Костянтин Дмитрович,
аспірант НН ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Жигло Сергій Вікторович,
аспірант кафедри інформаційно-вимірювальних технологій Харківського національного університету радіоелектроніки, Україна, Харків

Жуковський Сергій Станіславович,
доцент, кандидат педагогічних наук, доцент кафедри комп'ютерних наук та інформаційних технологій Житомирського державного університету імені Івана Франка, Україна, Житомир

Зайченко Юрій Петрович,
професор, доктор технічних наук, професор кафедри математичних методів системного аналізу НН ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Згуровський Михайло Захарович,
академік НАН України, професор, доктор технічних наук, науковий керівник ННК «ІПСА» КПІ ім. Ігоря Сікорського, Україна, Київ

Івохін Євген Вікторович,
професор, доктор фізико-математичних наук, професор кафедри системного аналізу та теорії прийняття рішень Київського національного університету імені Тараса Шевченка, Україна, Київ

Касьянов Павло Олегович,
член-кореспондент НАН України, професор, доктор фізико-математичних наук, директор ННК «ІПСА» КПІ ім. Ігоря Сікорського, Україна, Київ

Кузнецова Наталія Володимирівна,
доцент, доктор технічних наук, професор кафедри математичних методів системного аналізу НН ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Кузьменко Олексій Віталійович,
аспірант кафедри системного проектування НН ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Левчук Олександр Миколайович,
кандидат технічний наук, доцент кафедри інформаційних управляючих систем та технологій Ужгородського національного університету, Україна, Ужгород

Мілявський Юрій Леонідович,
доктор технічних наук, доцент кафедри математичних методів системного аналізу НН ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Міца Олександр Володимирович,
професор, доктор технічних наук, завідувач кафедри інформаційних управляючих систем та технологій Ужгородського національного університету, Україна, Ужгород

Палійчук Лілія Сергіївна,
кандидат фізико-математичних наук, науковий співробітник ННК «ІПСА» КПІ ім. Ігоря Сікорського, Україна, Київ

Панкратова Наталія Дмитрівна,
член-кореспондент НАН України, професор, доктор технічних наук, заступник директора ННК «ІПСА» КПІ ім. Ігоря Сікорського, Україна, Київ

Писарчук Ілля Олексійович,
студент магістратури кафедри обчислювальної техніки факультету інформатики та обчислювальної техніки КПІ ім. Ігоря Сікорського, Україна, Київ

Писарчук Олексій Олександрович,
професор, доктор технічних наук, професор кафедри обчислювальної техніки факультету інформатики та обчислювальної техніки КПІ ім. Ігоря Сікорського, Україна, Київ

Пецко Василь Іванович,
кандидат технічних наук, доцент кафедри інформаційних управляючих систем та технологій Ужгородського національного університету, Україна, Ужгород

Рець Вадим Олександрович,
аспірант кафедри системного аналізу та теорії прийняття рішень Київського національного університету імені Тараса Шевченка, Україна, Київ

Романенко Віктор Демидович,
професор, доктор технічних наук, заступник директора з науково-педагогічної роботи НН ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Романюк Оксана Олександрівна,
доцент, кандидат технічних наук, доцент кафедри автомобільного транспорту та сучасної інженерії Відкритого міжнародного університету розвитку людини «Україна», Україна, Київ

Савченко Ілля Олександрович,
доктор технічних наук, доцент кафедри математичних методів системного аналізу НН ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Стецюк Петро Іванович,
член-кореспондент НАН України, доктор фізико-математичних наук, завідувач відділу методів негладкої оптимізації Інституту кібернетики імені В.М. Глушкова НАН України, Київ

Титаренко Андрій Миколайович,
доктор філософії, асистент кафедри математичних методів системного аналізу НН ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Тріщ Роман Михайлович,
професор, доктор технічних наук, завідувач кафедри мехатроніки та електротехніки Національного аерокосмічного університету «Харківський авіаційний інститут», Україна, Харків; Mykolas Romeris University (Company VAT code LT119517219, Data is collected and kept in the Register of the Legal Entities, code 111951726), Lithuania, Vilnius

Федін Сергій Сергійович,
професор, доктор технічних наук, професор кафедри інформаційних систем і технологій Національного транспортного університету, Україна, Київ

Шантир Антон Сергійович,
кандидат технічних наук, доцент кафедри штучного інтелекту Державного університету інформаційно-комунікаційних технологій, Україна, Київ

Шапочка Ігор Валерійович,
кандидат фізико-математичних наук, доцент кафедри інформаційних управляючих систем та технологій Ужгородського національного університету, Україна, Ужгород

Шовкопляс Тетяна Володимирівна
кандидат фізико-математичних наук, асистент кафедри загальної математики ММФ КНУ ім. Тараса Шевченка, Україна, Київ

Штефан Наталія Володимирівна,
доцент, кандидат технічних наук, доцент кафедри інформаційно-вимірювальних технологій Харківського національного університету радіоелектроніки, Україна, Харків

Шум Кирил Ігорович,
магістр НН ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ