

System Analysis and Control Системный анализ и управление

Research article

DOI: <https://doi.org/10.18721/JCSTCS.18313>

UDC 004.021



DEVELOPMENT OF A DUAL-LOOP METHOD OF INTELLIGENT TRAFFIC LIGHT CONTROL BASED ON REINFORCEMENT LEARNING AND HOURLY DISTILLATION OF PHASE STRATEGIES

A.M. Sazanov , *V.P. Shkodyrev*, *S.M. Ustinov* 

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

 arseny.sazanov@gmail.com

Abstract. With increasingly complex urban dynamics, as well as increasing demands for the sustainability of urban mobility and introduction of cognitive technologies into transport infrastructure, the paper proposes a dual-loop method for intelligent traffic light control based on reinforcement learning and phase strategy distillation procedures. The first level implements real-time control through an RL-agent, while the second one generates backup hourly plans based on statistics of its behavior. The method is based on a system-discrete model taking into account stochastic traffic parameters and permissible control constraints. The simulation conducted in SUMO for a real intersection demonstrates a significant reduction in average transport delay compared to classical control, confirming the efficiency, sustainability and scalability of the approach. The obtained results substantiate the possibility of practical implementation of the model within the framework of intelligent transport systems of large cities and for laying the engineering foundation for hybrid urban mobility management architectures.

Keywords: reinforcement learning, intelligent traffic light control, dual-loop control architecture, traffic light controller, traffic management and control

Citation: Sazanov A.M., Shkodyrev V.P., Ustinov S.M. Development of a dual-loop method of intelligent traffic light control based on reinforcement learning and hourly distillation of phase strategies. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 144–153. DOI: 10.18721/JCSTCS.18313

Научная статья

DOI: <https://doi.org/10.18721/JCSTCS.18313>


УДК 004.021



РАЗРАБОТКА ДВУХКОНТУРНОГО МЕТОДА ИНТЕЛЛЕКТУАЛЬНОГО СВЕТОФОРНОГО РЕГУЛИРОВАНИЯ НА ОСНОВЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ И ПОЧАСОВОЙ ДИСТИЛЛЯЦИИ ФАЗОВЫХ СТРАТЕГИЙ

А.М. Сазанов  , В.П. Шкодырев, С.М. Устинов 

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

 arseny.sazanov@gmail.com

Аннотация. На фоне усложняющейся урбанистической динамики, а также возрастающих требований к устойчивости городской мобильности и внедрения когнитивных технологий в транспортную инфраструктуру в работе предлагается двухконтурный метод интеллектуального регулирования светофоров на основе обучения с подкреплением и процедур дистилляции фазовых стратегий. Первый уровень реализует управление в реальном времени через RL-агента, второй – формирует резервные почасовые планы на основе статистики его поведения. Метод опирается на системно-дискретную модель с учетом стохастических параметров трафика и допустимых ограничений управления. Проведенное моделирование в SUMO для реального перекрестка демонстрирует существенное снижение средней задержки транспорта по сравнению с классическим управлением, подтверждая эффективность, устойчивость и масштабируемость подхода. Полученные результаты обосновывают возможность практического внедрения модели в рамках интеллектуальных транспортных систем крупных городов и заложения инженерной основы для гибридных архитектур управления городской мобильностью.

Ключевые слова: обучение с подкреплением, интеллектуальное управление светофором, двухконтурная архитектура управления, контроллер светофора, управление и контроль дорожного движения

Для цитирования: Sazanov A.M., Shkodyrev V.P., Ustinov S.M. Development of a dual-loop method of intelligent traffic light control based on reinforcement learning and hourly distillation of phase strategies // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 144–153. DOI: 10.18721/JCSTCS.18313

Introduction

At the turn of the third technological revolution, associated with the widespread penetration of digital intelligence into the material infrastructure of cities, an accelerated convergence of cognitive control systems, neural-like algorithms and the sociotechnical environment that shapes the transport mobility of megacities is being observed. One of the most critical vectors of urban transformation is the intellectualization of traffic flow regulation, where the introduction of machine learning methods, reinforcement learning (RL) algorithms in particular, opens new paradigms in the management of adaptive traffic light cycles. This is not just a modernization of signaling devices, but a radical revision of the philosophy of interaction between the transport infrastructure and its subjects, coupled with the idea of a predictive, self-learning traffic control environment.

In the Russian scientific and technological context, the relevant areas are institutionally enshrined in documents such as the Strategy for the Development of the Transport Industry of the Russian Federation until 2030 with a Forecast until 2035, the Concept of the Intelligent Transport System of the Russian Federation, as well as in the priorities of the Digital Economy and National Technology

Initiative (NTI) programs. In particular, the NTI “AvtoNet” roadmap emphasizes the need to develop and implement intelligent algorithms for coordinating traffic flows under conditions of heterogeneous traffic involving traditional, automated and autonomous vehicles. These regulatory and programmatic grounds determine the high level of relevance of tasks related to the intellectualization of traffic light regulation as a systemic element of the digital transformation of urban mobility.

The paradigm of traffic light control considered in this paper is based on the integration of the RL methodology into a structurally discrete coordination model, where each traffic light node is interpreted as an agent in a stochastic environment, endowed with the ability to minimize global or local delay metrics. It is objectively necessary to move from homogeneous control systems to heterogeneous adaptive environments, in which interaction between agents occurs at the level of heuristic cooperation, mediated by utility and long-term reward functions.

The evolution of such approaches, which began with academic experiments and moved to the stage of applied integration in the largest urban agglomerations (Copenhagen, Singapore, Seoul, Beijing), reflects the global trend of decentralization of control and the transition from hierarchical systems to network architectures, in which each traffic light can act as an autonomous intelligent agent. Such a paradigm is rooted in the context of the Fourth Industrial Revolution, where autonomy, self-learning and integrative capacity of systems at the real-time level are becoming key coordinates.

Analysis of the subject area

Until recently, traffic lights, as basic elements of transport infrastructure, were considered mainly within the paradigm of rigidly deterministic regulation, based on preset time cycles and program plans, lacking the ability for emergent response to multivariant and stochastic scenarios of real road traffic.

International and domestic researches [1–5] indicate that classical control schemes – from fixed time plans to coordinated systems such as the “green wave” – have exhausted their capabilities under conditions of dynamic growth in traffic density, variability in the behavior of road users and the emergence of new forms of transport mobility (e.g., unmanned vehicles and micromobility agents). The methodological limitations of traditional approaches are expressed in their inability to account for traffic nonlinearity, lack of self-learning and adaptation [2] and also in their high dependence on centralized control systems subject to single points of failure.

An alternative to these systems is artificial intelligence methods, in particular, RL, as a cognitive model of decision making under uncertainty and limited information [3]. In recent years, there has been an explosive growth in the number of studies devoted to the application of Q-learning, Deep Q-Network (DQN), Actor-Critic algorithms and their modifications to the problem of traffic light phase control. Prototypes of the systems are implemented in simulation environments (SUMO, CityFlow, AIMSUN) and in individual pilot zones of smart cities, demonstrating a decrease in average traffic delay of up to 40–60% compared to traditional algorithms.

Thus, in [4], a comprehensive study of Q-learning effectiveness for adaptive control of a traffic light object at an intersection on the Stanford University campus was conducted. Using the SUMO platform and data on synthetic flow intensity, they implemented agent training based on state clustering using the k-means method and stochastic action selection policies. The results of a 50-day simulation showed a 40% efficiency increase compared to a fixed policy, albeit with slow convergence. The work confirms the potential of RL in urban conditions, requiring minimal sensor infrastructure.

In [5], an end-to-end simulation framework for evaluating RL algorithms in traffic control was developed and validated, including a realistic SUMO environment and an interface to RL libraries. The study showed that DQN and A2C agents, trained on dynamic intersection states, outperform traditional algorithms in terms of delay and travel time metrics. The work highlights the importance of high-level architecture and synchronization in emulation to achieve stable and reproducible results.

In [6], the authors implemented and compared the effectiveness of Q-learning, SARSA and Expected SARSA algorithms for traffic light control at a four-way intersection, using SUMO and the Python environment. The results showed that Q-learning exhibits the lowest delay and average waiting time compared to other methods. All algorithms outperformed the traditional fixed control, confirming the potential of RL for urban traffic optimization.

In [7], a traffic light control system using Q-learning in the SUMO environment was developed, demonstrating a significant reduction in vehicle delay time compared to fixed cycles. The algorithm adapts to traffic density, improving the throughput and responsiveness of the system in real-world intersection scenarios.

In [8–10], the authors conducted a systematic review of approaches based on RL and deep learning for traffic light control. Single-traffic-light, multi-traffic-light and centralized schemes, Q-learning, DQN and Actor-Critic algorithms were considered. The review showed a robust superiority of intelligent methods over classical ones, especially under conditions of uncertainty and multidimensional traffic dynamics.

A comprehensive analysis of scientific sources, including comparative empirical studies, shows that there is a pressing need to formalize and develop a new intelligent method for controlling traffic lights, based on deep integration of RL into the topology of the transport network. The method should account for the specifics of Russian transport infrastructure, allow for step-by-step modular implementation and ensure transitions to both adaptive and planned stable behavior of the control system, considering the possibility of scaling.

Problem statement

Under conditions of a complexifying urban environment, intensified traffic flows and the emergence of a multi-component structure of road users (including automated vehicles, micromobility platforms and AI agents), issues of adaptive and intelligent control of traffic light infrastructure are acquiring the character of a systemic priority in the context of the formation of cyber-physical transport systems. Modern challenges, associated with spatio-temporal heterogeneity of traffic, the stochastic nature of traffic flows and the need for predictive control based on incomplete data, require going beyond traditional algorithms focused on fixed time plans or centralized coordination scenarios.

Based on the analysis of advanced scientific and applied solutions, as well as the specifics of operating traffic light objects under conditions of limited telemetry and hardware resources, the task is set to develop an intelligent method for controlling traffic light objects based on RL, implementing not only the ability to adaptively select phases in real time, but also ensuring a stable, redundant and operationalizable control strategy at the level of hourly traffic light plans.

Thus, a dual-loop model for controlling traffic light objects is created: the first loop is intelligent, operating in real time using an RL agent; the second is backup and planned, based on hourly strategies extracted from the trained behavior of the neural network. This allows for achieving a double effect: on the one hand, the system uses the advantages of adaptive and predictive regulation, characteristic of RL models; on the other hand, it ensures fault tolerance and operational controllability in case of failure or unavailability of AI loops (for example, in case of communication loss, disruption of computing infrastructure, attack or model failure).

Within the framework of the set task, it is assumed:

1. To develop an architecture of an intelligent agent for controlling a traffic light object, capable of learning in the SUMO environment using RL algorithms (Q-learning, DQN, A2C etc.).
2. To develop a methodology for logging and subsequent aggregation of the phase control strategy adopted by the agent over time intervals (hourly breakdown).
3. To build a mechanism for transforming the agent's behavior into static hourly cycles suitable for implementation in standard traffic light controllers.

4. To evaluate the effectiveness of the developed dual-loop control model under variable traffic conditions, including simulation of emergency modes.

5. To prepare technical regulations and a description of the process of transferring from the RL module to backup planned control.

The proposed problem statement is aimed not only at solving the problem of traffic light regulation optimization, but also at ensuring the principle of engineering reliability and continuity of operation of intelligent transport systems. The implementation of such a model can lay the foundation for the creation of hybrid urban mobility systems that combine intelligent adaptability and structural stability of control.

Physical statement of the problem in terms of system control of transport dynamics

Modern transport hubs of a megalopolis, especially such as the Jan Rainis blvd. and Geroev-Panfilovtsev st. intersection, are highly organized non-linear dynamic systems with perturbed inputs and multi-criteria target functions. Their functioning is subject not only to local laws of transport flows, but also to system logic formed at the intersection of physical realities of urban infrastructure, intensification of auto traffic, as well as requirements for safety and throughput.

According to the object's passport (Fig. 1), traffic light regulation in this unit is implemented in a locally adaptive mode, which already indicates the presence of elements of sensory assessment of the traffic situation and primitive logic of changing phases depending on input signals, for example, from inductive loop detectors located on key entry lanes. However, adaptability here is apparently implemented in the form of discrete choice between a limited set of predefined programs, which fundamentally limits control capabilities under conditions of highly stochastic traffic flows.

The spatial configuration of the intersection, judging by the plan for the arrangement of technical equipment, demonstrates the presence of:

- multi-lane driveways with the possibility of left turns;
- pedestrian crossings with protection phases;
- partial asymmetry of flows;
- sensor infrastructure, potentially suitable for integration into a digital control architecture.

Thus, the object is an easily observable, partially controlled, multi-channel transport system operating under conditions of highly dynamic input data.

Mathematical statement of the problem within the framework of system analysis

The considered problem of developing an intelligent method for controlling traffic light objects with redundant strategies based on distillation of the behavior of a neural network agent is rooted in the field of synthesis of complex cyber-physical control systems with elements of stochastic optimal control and automata theory in a spatio-temporal transport environment. The methodological basis is system analysis, within which the transport network is presented as a hierarchical discrete-continuous dynamic system with partial observability and variable dimensionality of the phase space.

Let us assume that on a discrete time interval

$$T = \{0, \Delta t, 2\Delta t, \dots, T_{\max}\} \subset \mathbb{R}, \quad (1)$$

we observe a transport system on a limited section of the road network, including one or more controlled intersections. This means that the modeling or control of the system is carried out on a uniform time grid with a step Δt , from the initial moment (0) to the final time T_{\max} .

Each crossroad $i \in I$, where I is the set of crossroad identifiers, is equipped with a traffic light object consisting of a finite set of admissible switching phases such that:



where K_i is the number of phases for intersection i .

The traffic light control agent, at each moment of time t , selects an action $A_t \in \mathbf{A}$, where \mathbf{A} is the set of admissible actions (phase switches), in accordance with the policy $\pi: \mathbf{S} \rightarrow \mathbf{A}$, which determines the probability of selecting one or another action depending on the observed state of the system. The policy can be deterministic ($\pi(s) = a$) or stochastic, that is:

The agent’s goal is to maximize the expected total reward for an episode of length T , expressed by the functional:

where $R(S_t, A_t)$ is the instantaneous reward function, reflecting the target indicators of traffic light regulation efficiency (e.g., average delay, queue length, number of stops etc.), and $\gamma \in (0, 1)$ is the discount coefficient, which sets the significance of future rewards.

149

$$\pi^* = \arg \max_{\pi} J(\pi) \text{ with } \forall t : S_t = f(S_t, A_t, \xi_t),$$

where $f(\cdot)$ is the dynamics of the transport environment evolving under the influence of the agent's action and the random disturbance vector ξ_t , simulating the stochastics of movement.

Intelligent loop (first control level)

The first level of the proposed architecture is an intelligent agent based on the DQN algorithm. In this approach, the approximation of the optimal action value function $Q^*(s, a)$ is achieved by means of a deep neural network parameterized by weights θ . The weights are updated based on the Bellman equation.

Second loop (phase translation)

The second level of the architecture is responsible for the formation of hourly control strategies based on the procedure of distilling the RL agent's behavior into a form suitable for interpretation and execution as a stationary phase schedule. This procedure transforms the flow of stochastic decisions, made by the agent during simulation or real operation, into aggregated phase control profiles reflecting repeating patterns of system behavior. Within the framework of such translation, the following is performed:

- statistical aggregation of distributions of selected phases for each hour of the day;
- normalization and adaptation of phase durations to a fixed cyclic structure (e.g. 60 or 90 seconds);
- ensuring technical constraints (permissible minimums, multiplicities, transition intervals);
- formation of a set of hourly cycles representing a backup control plan that is resilient to failures or unavailability of the first (intelligent) level.

Thus, the second circuit implements the procedure for synthesizing a planned control policy that maintains heuristic proximity to the strategy obtained as a result of RL, but transformed into a formalized, interpretable and implemented control scheme at the level of regular controllers. Let statistics on the choice of control phases be accumulated over some observed period of the RL agent's operation.

Initial data

- Let there be the statistics of the adaptive (RL) traffic light control program operation for a certain period.
- For each main phase i , the total duration of its burning $t_{i,RL}$ is known.
- There are n main phases in total, T_{cycle} is the duration of one fixed cycle (without transition phases).
- For each phase, the minimum permissible burning time $t_{i,min}$ is specified.
- For each phase, it is assumed that it can be repeated in a cycle N_i times.
- The time for switching between phases (transition phases) is also taken into account, we will denote it by S .

Goal

The goal is to construct a fixed program that is as close as possible to the RL statistics, i.e., select such phase durations $t_{ij} (j = 1...N_i)$, so that:

- The total burning time of each phase in the cycle is proportional to its share in the RL statistics.
- The durations of individual phases are not less than the minimum allowable ones.
- The total duration of all phases and transitions is equal to the duration of the cycle.

Formalization

Phase:

$$r_i = \frac{t_{i,RL}}{\sum_{k=1}^n t_{k,RL}}.$$

Limitations:

1. For each phase:

$$\sum_{j=1}^{N_i} t_{i,j} = r_i \cdot (T_{cycle} - S).$$

2. For each $t_{i,j}$:

$$t_{i,j} \geq t_{i,\min}.$$

3. Sum time:

$$\sum_{i=1}^n \sum_{j=1}^{N_i} t_{i,j} + S = T_{cycle}.$$

Goal function

Goal function is to minimize the spread of phase durations relative to the desired mean value:

$$\sum_{i=1}^n \sum_{j=1}^{N_i} (t_{i,j} - \bar{t}_i)^2 \rightarrow \min,$$

where

$$\bar{t}_i = \frac{r_i \cdot (T_{cycle} - S)}{N_i}.$$

Finding the optimal number of phase repetitions

- For each phase, the possible number of repetitions N_i is considered (for example, from 1 to 4).
- For each set (N_1, \dots, N_n) , a quadratic programming problem with the objective function and constraints above is solved.
- The final set (N_1, \dots, N_n) and the corresponding phase durations are selected based on the minimum value of the objective function.

Features

The time for transition phases S depends on the number of switches and the duration of each transition phase.

All phase durations are rounded to whole seconds, while the sum is preserved.

If it is impossible to satisfy the constraints (for example, too little time for all phases taking into account the minimum durations), the solution is considered impossible.

Result

Thus, we have a fixed program, where the phase durations correspond maximally to the RL statistics, satisfying all technical constraints (minimum time, discreteness, transition structure).

Evaluation of simulation results

As part of the evaluation of the efficiency of the proposed dual-loop architecture of intelligent traffic light control, a series of simulation experiments were conducted in the SUMO environment based on the real Jan Rainis blvd. and Geroev-Panfilovtsev st. intersection. The model was adapted to the topological and phase features of the object, including the number of lanes, traffic directions, presence of pedestrian crossings and sensor inputs. Two modes were tested:

- 1) classic fixed regulation corresponding to the current program;

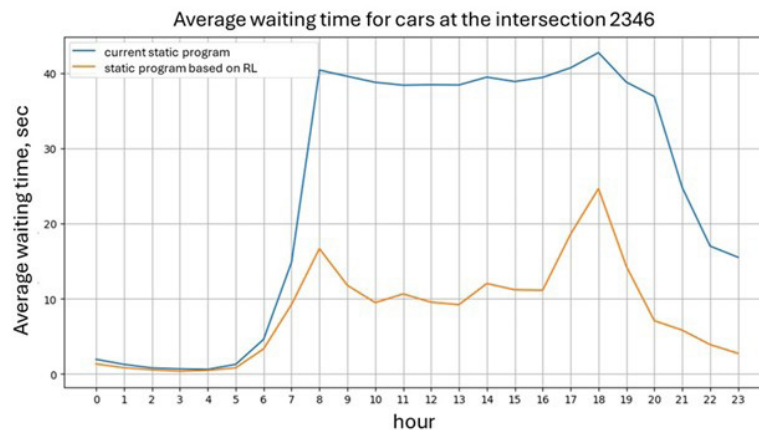


Fig. 2. Results of modeling the average waiting time of cars at an intersection

2) adaptive control using an RL agent and hourly phase distillation.

The main metric for the evaluation was the average vehicle delay (avgDelay) at the intersection entrances. To ensure statistical reliability, the simulation was carried out on a sample of 30 days, including both weekdays and weekends, with different flow intensities. The comparison graphs of queue dynamics and delay times between RL, distilled and fixed modes presented in Fig. 2 over the entire time interval demonstrate the consistent superiority of the intelligent model. A 32.7% reduction in the average number of stops was also recorded, which has a positive effect on the smoothness of the ride, fuel consumption and CO₂ emissions.

Thus, the proposed dual-loop model demonstrated high efficiency, fault tolerance and scalability. Simulation data substantiate its feasibility for use in the conditions of the Russian road infrastructure and serve as a basis for developing prototypes for implementation in intelligent transport systems of cities.

REFERENCES

1. Seliverstov S.A., Seliverstov Ya.A., Lukomskaya O.Yu., Sazanov A.M., Shkodyrev V.P. Analysis of the functional features of intelligent traffic management systems. *Transport: nauka, tekhnika, upravlenie. Nauchnyi informatsionnyi sbornik [Transport: Science, Technology, Management. Scientific information collection]*, 2023, Vol. 12, Pp. 25–31. DOI: 10.36535/0236-1914-2023-12-3
2. Seliverstov S.A., Sazanov A.M., Lukomskaya O.Y., Nikitin K.V., Shatalova N.V., Benderskaya E.N. Analysis of modern approaches to optimizing traffic control systems. *2021 XXIV International Conference on Soft Computing and Measurements (SCM)*, 2021, Pp. 106–108. DOI: 10.1109/SCM52931.2021.9507147
3. Seliverstov S.A., Seliverstov Ya.A., Shatalova N.V., Sazanov A.M. Development of cognitive transport system architecture. *Transport: nauka, tekhnika, upravlenie. Nauchnyi informatsionnyi sbornik [Transport: Science, Technology, Management. Scientific information collection]*, 2024, Vol. 4, Pp. 8–12. DOI: 10.36535/0236-1914-2024-04-2
4. Dowling R., Skabardonis A., Alexiadis V., Hardy M. *Traffic Analysis Toolbox Volume III: Guidelines for applying traffic microsimulation modeling software*. McLean, VA: Turner-Fairbank Highway Research Center, 2004. Available: <https://highways.dot.gov/media/6916> (Accessed 09.09.2025)
5. Tan T. *Data-driven adaptive traffic signal control via deep reinforcement learning*, PhD thesis. Stanford, CA: Stanford University, 2020. Available: <http://purl.stanford.edu/fs712rs0591> (Accessed 01.07.2025).

6. Masfequier Rahman Swapno S.M., Nuruzzaman Nobel S.M., Ramachandra A.C., Babul Islam M., Haque R., Rahman M.M. Traffic light control using reinforcement learning. *2024 International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 2024, Pp. 1–7. DOI: 10.1109/ICICACS60521.2024.10498933
7. Xing Y., Shen F., Zhao J. A perception evolution network for unsupervised fast incremental learning. *2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, 2013, Pp. 1–8. DOI: 10.1109/IJCNN.2013.6706752
8. Saadi A., Abghour N., Chiba Z., Moussaid K., Ali S. A survey of reinforcement and deep reinforcement learning for coordination in intelligent traffic light control. *Journal of Big Data*, 2025, Vol. 12, Art. no. 84. DOI: 10.1186/s40537-025-01104-x
9. Seliverstov S.A., Sazanov A.M., Benderskaia E.N., Nikitin K.V., Seliverstov Ia.A. Razrabotka arkhitektury intellektual'noi sistemy upravleniia dorozhnym dvizheniem [Development of the architecture of an intelligent traffic management system]. *Soft Computing and Measurements (SCM)*, 2021, Pp. 281–285.
10. Seliverstov S., Lukomskaya O., Titov V., Vashchuk A., Khalturin A. On building the architecture of the intelligent transportation system in the Arctic region. *Transportation Research Procedia*, 2021, Vol. 57, Pp. 603–610. DOI: 10.1016/j.trpro.2021.09.089

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Arseniy M. Sazanov
Сазанов Арсений Михайлович
 E-mail: arseny.sazanov@gmail.com

Viacheslav P. Shkodyrev
Шкодырев Вячеслав Петрович
 E-mail: shkodyrev@mail.ru

Sergei M. Ustinov
Устинов Сергей Михайлович
 E-mail: usm50@yandex.ru
 ORCID: <https://orcid.org/0000-0003-4088-4798>

Submitted: 27.07.2025; Approved: 19.08.2025; Accepted: 19.08.2025.

Поступила: 27.07.2025; Одобрена: 19.08.2025; Принята: 19.08.2025.