

УДК 004.932

DOI: [10.26102/2310-6018/2025.48.1.007](https://doi.org/10.26102/2310-6018/2025.48.1.007)

Использование нейронных сетей для определения пылевого загрязнения вблизи области открытой добычи угля по данным дистанционного зондирования Земли

Ю.А. Озарян, Т.В. Кожевникова, К.С. Цыгулёв, В.Е. Окладников

Хабаровский Федеральный исследовательский центр Дальневосточного отделения Российской академии наук, Хабаровск, Российская Федерация

Резюме. В статье рассматривается вопрос использования нейронных сетей для определения пылевого загрязнения вблизи области открытой добычи угля по данным дистанционного зондирования Земли. В работе участвовали территории угольных разрезов, находящихся в различных субъектах Российской Федерации. В качестве исходных данных использовались снимки спутника Sentinel-2, которые обрабатывались в ГИС-программе Quantum GIS. Разработан алгоритм формирования обучающей выборки, в котором используются каналы видимого и ближнего инфракрасного спектров со спутниковых снимков. Технология создания масок в разработанном алгоритме строится на использовании индекса Enhanced Coal Dust Index и его дальнейшей кластеризации. В качестве модели нейронной сети используется U-Net. Обученная модель апробирована на валидационной выборке. Точность распознавания составила 59,3 % по метрике Intersection over Union, 78,9 % по метрике Precision, 80,6 % по метрике F1 и 95,5 % по метрике Accuracy. Этот уровень точности обусловлен ограниченным объемом данных для обучения. Обсуждается возможность улучшения точности за счет увеличения объема выборки в сочетании с настройками параметров используемой нейронной сети. Полученный результат позволяет оценивать последствия воздействия работ по добыче угля на окружающую среду и на основании этих данных разрабатывать меры по обеспечению экологической безопасности.

Ключевые слова: пылевое загрязнение, дистанционное зондирование Земли, машинное обучение, кластеризация, нейронная сеть.

Благодарности: Исследование выполнено за счет гранта Российского научного фонда № 24-11-20029, <https://rscf.ru/project/24-11-20029/>. В создании научной статьи принимал участие М.Э. Добровольский (ДВГУПС) – анализ спутниковых данных и форматирование рисунков.

Для цитирования: Озарян Ю.А., Кожевникова Т.В., Цыгулёв К.С., Окладников В.Е. Использование нейронных сетей для определения пылевого загрязнения вблизи области открытой добычи угля по данным дистанционного зондирования Земли. *Моделирование, оптимизация и информационные технологии*. 2025;13(1). URL: <https://moitvvt.ru/ru/journal/pdf?id=1756> DOI: 10.26102/2310-6018/2025.48.1.007

Using neural networks to determine dust pollution near open-pit coal mining areas based on Earth remote sensing data

Y.A. Ozaryan, T.V. Kozhevnikova, K.S. Tsygulev, V.E. Okladnikov

Khabarovsk federal research center of the Far Eastern Branch of the Russian Academy of Sciences, Khabarovsk, the Russian Federation

Abstract. The article examines the use of neural networks for detecting dust pollution near open-pit coal mining areas based on remote sensing data. The study involved coal mining sites located in various regions of the Russian Federation. Satellite images from the Sentinel-2 mission served as the primary data source and were processed using Quantum GIS software. An algorithm for forming the training

dataset was developed, utilizing the visible and near-infrared spectral bands from the satellite imagery. The mask creation technology in the developed algorithm is based on the Enhanced Coal Dust Index and its subsequent clustering. U-Net is used as a neural network model. The trained model was tested on a validation dataset. The recognition accuracy was 59.3% for the Intersection over Union metric, 78.9% for the Precision metric, 80.6% for the F1 metric, and 95.5% for the Accuracy metric. This level of accuracy is attributed to the limited volume of training data. The potential for improving accuracy through increasing the sample size in conjunction with optimizing the parameters of the neural network is discussed. The results obtained provide a basis for assessing the environmental impacts of coal mining activities and for developing measures to ensure ecological safety based on these findings.

Keywords: dust pollution, Earth remote sensing, machine learning, clustering, neural network.

Acknowledgements: This research was funded by Russian Science Foundation, project number 24-11-20029, <https://rscf.ru/project/24-11-20029/>. The scientific article includes contributions from M.E. Dobrovolsky (FESTU) in the analysis of satellite data and the formatting of figures.

For citation: Ozaryan Y.A., Kozhevnikova T.V., Tsygulev K.S., Okladnikov V.E. Using neural networks to determine dust pollution near open-pit coal mining areas based on Earth remote sensing data. *Modeling, Optimization and Information Technology*. 2025;13(1). (In Russ.). URL: <https://moitvvt.ru/ru/journal/pdf?id=1756> DOI: 10.26102/2310-6018/2025.48.1.007

Введение

Дистанционное зондирование Земли (ДЗЗ) играет ключевую роль в мониторинге окружающей среды, сельском хозяйстве и управлении природными ресурсами. С использованием спутников и воздушных судов, оснащенных радиометрами, сканерами и камерами, возможно получать высококачественные данные в различных спектральных диапазонах. Это позволяет детально анализировать состояние как природных, так и антропогенных объектов, что делает ДЗЗ незаменимым инструментом для различных отраслей науки и промышленности.

Актуальность темы определения пылевого загрязнения обусловлена необходимостью разработки алгоритмов для мониторинга состояния объектов открытой добычи угля и их воздействия на окружающую среду, так как пылевое загрязнение, возникающее в результате угледобывающих работ, представляет собой серьезную экологическую проблему, требующую оперативного и точного определения для разработки мер по снижению негативного воздействия [1].

Основные загрязнители воздуха, образующиеся при открытой добыче полезных ископаемых – это взвешенные частицы. При добыче угля эти частицы образуются в процессе бурения, взрывных и вскрышных работ, погрузки угля и разгрузки, эксплуатации карьерных дорог и поверхности отвалов вскрышных пород. Снижение качества воздуха может вызвать заболевание органов дыхания. Кроме этого, запыленность влияет на растительный и животный мир [2].

Анализ методов обработки данных ДЗЗ для мониторинга состояния объектов открытой добычи угля позволяет использовать информацию с космических снимков для визуализации объектов открытой добычи угля [3, 4]. В качестве информационной технологии анализа данных ДЗЗ рассматриваются методы машинного обучения и технологии обработки изображений. Апробация алгоритма и оценка его точности на реальных данных.

Целью данной работы является исследование возможности использования нейронной сети для определения пылевого загрязнения вблизи области открытой добычи угля по данным ДЗЗ.

Решаются следующие задачи: выбор территории исследования, содержащей объекты открытой добычи угля; проведение обработки спутниковых данных для формирования обучающей и валидационной выборки; разработка алгоритма и программы для детектирования пылевого загрязнения; апробация разработанного алгоритма и оценка его точности на валидационной выборке.

Методы решения задач включают использование нейронных сетей, географических информационных систем и технологий обработки изображений. Новизна работы заключается в разработке специализированного программного обеспечения для мониторинга пылевого загрязнения, куда интегрированы современные нейронные сети и методы обработки данных ДЗЗ.

В связи с этим, исследование направлено на решение важной экологической проблемы и имеет значительный потенциал для дальнейшего развития и применения в практике мониторинга окружающей среды [5].

Материалы и методы

В качестве исходных данных использовались снимки спутника Sentinel-2. Преимуществом использования этих снимков является свободный доступ, а также высокое пространственное разрешение (10 м). Выбор Sentinel-2 был также обусловлен его уникальными характеристиками. Во-первых, спутник обеспечивает высокую частоту обновления данных, что позволяет получать свежие изображения Земли каждые 3–5 дней. Это крайне важно для задач, требующих отслеживания пылевого загрязнения. Кроме того, камеры Sentinel-2 обладают 13 спектральными каналами, каждый из которых имеет свою специфическую чувствительность к определенным типам поверхностей или явлений [6].

Для работы использовались каналы видимого спектра (красный – RED, синий – BLUE и зеленый – GREEN) и ближнего инфракрасного спектра – NIR. У снимков Sentinel-2 – 4, 3, 2 и 8 каналы.

После скачивания снимков с космического аппарата Sentinel-2 требуется выполнить ряд действий для подготовки данных к анализу в ГИС-системе Quantum GIS (QGIS).

Исходные данные представлены в форматах JPEG2000 (jp2) для изображений и XML для метаданных. Для конвертации изображений в удобный для работы формат GeoTIFF (tif) можно воспользоваться различными инструментами, включая модуль Semi-Automatic Classification Plugin (SCP) для QGIS.

В качестве модели нейронной сети используется U-Net. Одним из ключевых преимуществ использования U-Net в ДЗЗ является ее архитектура, которая позволяет эффективно извлекать пространственные и контекстуальные признаки из изображений. С помощью механизма пропуска (skip connections) U-Net сохраняет важные детали и текстуры, которые могут быть потеряны при традиционном сжатии изображений [7].

Для составления обучающей выборки рассматривались следующие угольные разрезы: Краснобродский разрез, разрез «Восточный», Сибиргинский разрез, разрез «Северный Маганак», разрез Харанорский, Калтанский угольный разрез и Распадский разрез. В валидационную выборку вошли угольные разрезы Ургальского и Нерюнгринского месторождений.

Формирование масок пылевого загрязнения на исследуемых объектах проводилось с помощью расчета индекса Enhanced Coal Dust Index (ECDI) и кластеризации.

ECDI – метод картирования пылевого загрязнения, предложенный в 2022 году. Он позволяет идентифицировать участки, загрязненные угольной пылью, отделяя их от других объектов. Формула его расчета выглядит следующим образом:

$$ECDI = \frac{SWIR1 - NIR + SWIR2}{SWIR1 + NIR - SWIR2}, \quad (1)$$

где SWIR1 и SWIR2 – 11 и 12 каналы спутникового снимка Sentinel-2, охватывающие коротковолновые инфракрасные спектральные диапазоны [8, 9].

Для кластеризации использован метод поиска восхождением к вершине. Это итеративный алгоритм, используемый для нахождения локального максимума функции. Для кластеризации данный метод может быть адаптирован для оптимизации функций, связанных с разбиением данных на кластеры. Его использование обусловлено тем, что данный алгоритм реализован в QGIS в качестве функции модуля геоинформационной системы System for Automated Geoscientific Analyses (SAGA GIS) [10].

Алгоритм формирования обучающей выборки состоит из двух этапов.

На первом этапе первичная обработка данных спутникового снимка. Для этого, каналы, представленные в формате jp2, преобразуются в формат tif с последующей нормализацией (Рисунок 1) и обрезкой до размеров исследуемого участка (Рисунок 2а). Данный процесс происходит с помощью плагина SCP в QGIS. Для его корректной работы необходимо избавиться от приписок, указывающих на его разрешение – «_10m» и «_20m», в конце имени файлов каналов.

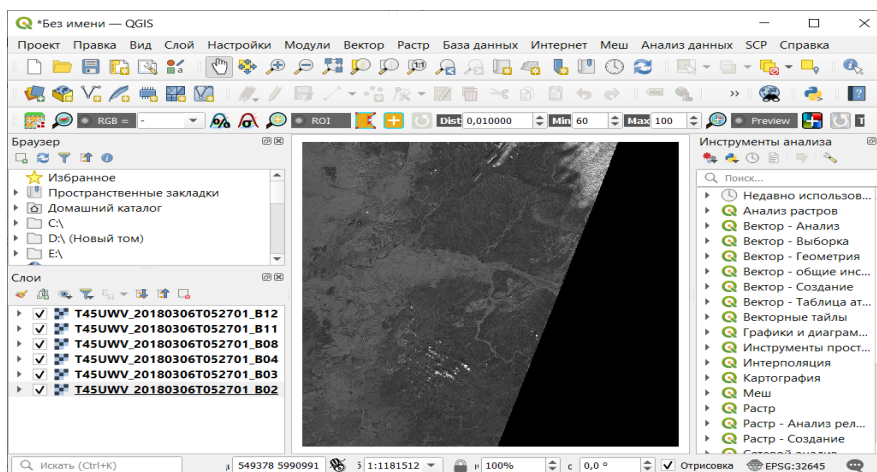


Рисунок 1 – Обработанный снимок в рабочей области программы
 Figure 1 – The processed image in the program's workspace

На втором этапе проводится формирование изображения и маски пылевого загрязнения. Чтобы получить изображения, полученные на предыдущем этапе, каналы объединяются в один файл, который сохраняется в качестве входных данных для нейронной сети. Затем, по этим же каналам проводится расчет индекса ECDI. На полученном изображении со значениями индекса (Рисунок 2б) выполняется кластеризация методом поиска восхождением к вершине для выделения пылевого загрязнения. Разделение проводится на 6 кластеров. Такое количество кластеров обеспечивает достаточно детальное разделение областей с разными уровнями загрязнения, но при этом оно не слишком избыточное или сложное для анализа. Кластеризация проводится только для упрощения и частичной автоматизации формирования масок. Сами кластеры в процессе обучения нейронной сети не учитываются.

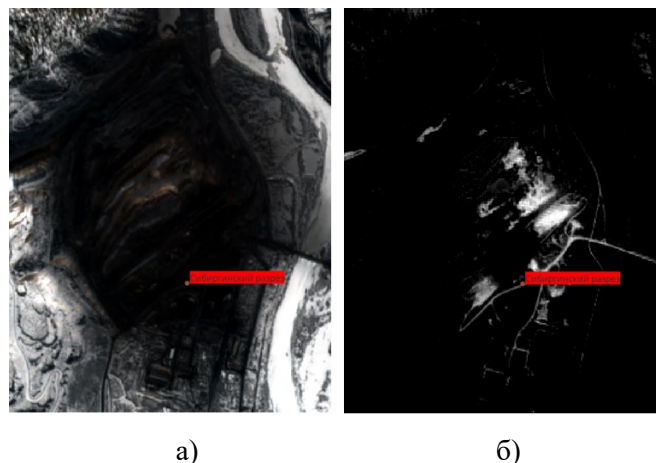


Рисунок 2 – Сибиргинский разрез: а – цветное изображение; б – изображение ECDI
 Figure 2 – Sibirginsky quarry: a – color image; b – ECDI image

После выполнения кластеризации на полученном изображении (Рисунок 3а) визуально выбираются кластеры, описывающие пылевое загрязнение. Отобрав нужные, проводится их объединение и сохранение на отдельное изображение. Полученный результат представляет собой маску, обозначающую наивысшие значения ECDI в изучаемой области разреза. Маска может содержать нежелательные элементы, которые удаляются вручную (Рисунок 3б).

Полученная маска сохраняется для последующего использования в качестве входных данных для нейронной сети.

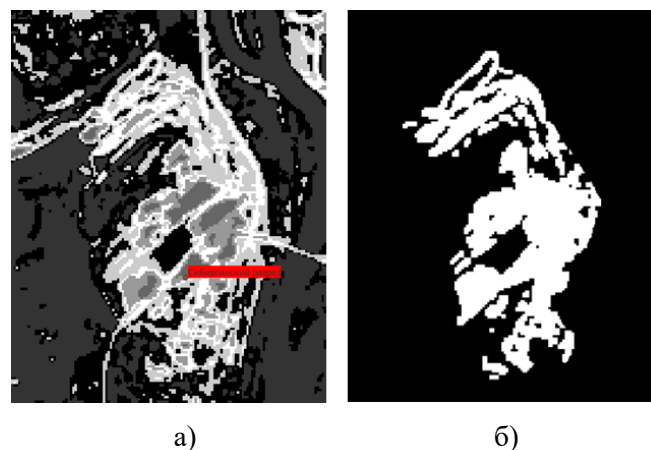


Рисунок 3 – Сибиргинский разрез: а – кластеризованное изображение; б – полученная маска
 Figure 3 – Sibirginsky quarry: a – clustered image; b – obtained mask

Результаты

В результате работы алгоритма была сформирована обучающая выборка, состоящая из 56 снимков. В валидационную выборку вошло 14 снимков. Описание и обучение нейронной сети осуществлялось в среде Google Colab на языке программирования Python с помощью библиотек PyTorch и NumPy.

Для задания обучающей выборки `data_train` использовался метод `DataLoader`, который принимает список пар (изображение, маска). Изображения предварительно трансформируются с помощью функции `rollaxis`, чтобы изменить порядок осей, перемещая ось каналов на вторую позицию. Маски преобразуются в четырехмерные

тензоры, добавляя новую ось. Функция DataLoader упаковывает эти пары в батчи, перемешивает их и подготавливает к итеративной обработке в процессе обучения модели. Аналогичный процесс применяется к валидационной выборке data_val, создавая батчи данных для оценки модели.

На Рисунке 4 представлен результат обучения на data_train после 100 эпох. Оценка в конце каждой эпохи проводилась на data_val.

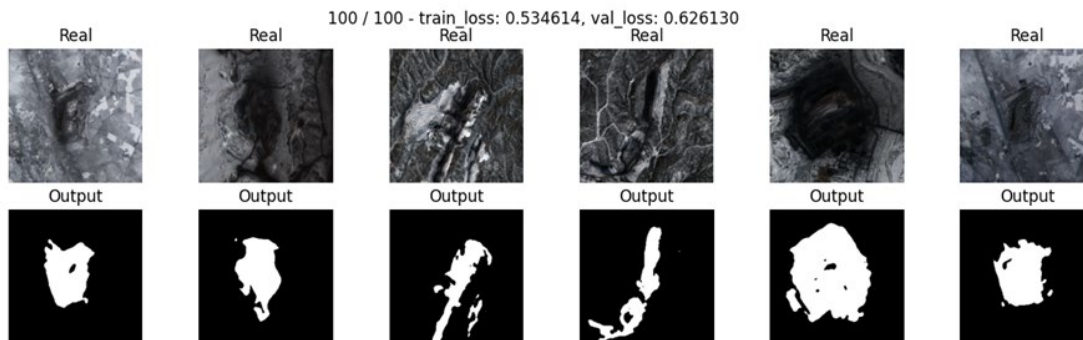


Рисунок 4 – Результат обучения после 100 эпох
Figure 4 – Training result after 100 epochs

Точность распознавания пылевого загрязнения на валидационной выборке представлена в Таблице 1.

Таблица 1 – Метрики точности распознавания модели U-Net на валидационной выборке
Table 1 – Metrics of recognition accuracy of the U-Net model on the validation dataset

Класс	Accuracy	Precision	Recall	F1	IoU
Пылевое загрязнение	0,9554	0,7887	0,8237	0,8058	0,5932

Пример распознавания представлен на Рисунке 5. Полученный уровень точности обусловлен ограниченным объемом данных для обучения – 70 снимков. Для небольшой выборки данных такая точность является приемлемым результатом, что подтверждает потенциал используемой архитектуры.



Рисунок 5 – Результат обработки снимка, не представленного в выборке
Figure 5 – The result of processing an image not presented in the sample

Заключение

В работе исследована возможность использования нейронной сети для определения пылевого загрязнения вблизи области открытой добычи угля по данным ДЗЗ. Результатом является алгоритм обработки данных ДЗЗ и обученная модель нейронной сети U-Net. Апробация модели показала потенциальную возможность ее дальнейшего использования в определении пылевого загрязнения.

Для улучшения точности работы алгоритма на основе предложенной модели можно рассматривать несколько подходов. Во-первых, увеличение размера обучающей выборки является основным способом улучшения результатов. Сбор дополнительных снимков с областями пылевого загрязнения позволит модели лучше распознавать загрязненные области и улучшит производительность на новых данных. Во-вторых, применение методов аугментации данных, таких как случайные повороты, сдвиги, масштабирование и зеркальные отражения, может существенно увеличить эффективный размер обучающей выборки, что способствует улучшению точности. Кроме того, оптимизация гиперпараметров, таких как параметры регуляризации, размер мини-батча и количество эпох, может помочь улучшить производительность модели. Использование предобученных энкодеров, интегрированных в архитектуру U-Net, может улучшить начальную точность и ускорить процесс обучения.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Абрамов Н.С., Макаров Д.А., Талалаев А.А., Фраленко В.П. Современные методы интеллектуальной обработки данных ДЗЗ. *Программные системы: теория и приложения*. 2018;9(4):417–442. <http://doi.org/10.25209/2079-3316-2018-9-4-417-442>
Abramov N., Makarov D., Talalaev A., Fralenko V. Modern methods for intelligent processing of Earth remote sensing data. *Program Systems: Theory and Applications*. 2018;9(4):417–442. (In Russ.). <http://doi.org/10.25209/2079-3316-2018-9-4-417-442>
2. Yu H., Zahidi I. Environmental hazards posed by mine dust, and monitoring method of mine dust pollution using remote sensing technologies: An overview. *Science of The Total Environment*. 2023;864. <https://doi.org/10.1016/j.scitotenv.2022.161135>
3. Опарин В.Н., Потапов В.П., Гиниятуллина О.Л., Андреева Н.В., Счастливцев Е.Л., Быков А.А. Оценка пылевого загрязнения атмосферы угледобывающих районов Кузбасса в зимний период по данным дистанционного зондирования Земли. *Физико-технические проблемы разработки полезных ископаемых*. 2014;(3):126–137.
Oparin V.N., Potapov V.P., Giniyatullina O.L., Andreeva N.V., Schastlivtsev E.L., Bykov A.A. Evaluation of dust pollution of air in Kuzbass coal-mining areas in winter by data of remote earth sensing. *Journal of Mining Science*. 2014;50(3):549–558. <https://doi.org/10.1134/S1062739114030168>
4. Nie X., Hu Z., Ruan M., Zhu Q., Sun H. Remote-Sensing Evaluation and Temporal and Spatial Change Detection of Ecological Environment Quality in Coal-Mining Areas. *Remote Sensing*. 2022;14(2). <https://doi.org/10.3390/rs14020345>
5. Мусина Г.А., Ожигин Д.С., Ожигина С.Б. Экологический мониторинг на основе снимков, полученных с помощью беспилотных летательных аппаратов. *Интерэкспо Гео-Сибирь*. 2019;4(2):196–204. <https://doi.org/10.33764/2618-981X-2019-4-2-196-204>
Moosina G.A., Ozhigin D.S., Ozhigina S.B. Environmental monitoring on the bases of the pictures received by of pilotless flying devices. *Interexpo GEO-Siberia*. 2019;4(2):196–204. (In Russ.). <https://doi.org/10.33764/2618-981X-2019-4-2-196-204>
6. Саворский В.П., Лупян Е.А., Горный В.И., Ермаков Д.М., Панова О.Ю., Константинова А.М. Методы и инструменты анализа данных ДЗЗ для выявления изменений растительного покрова, вызванных техногенными отходами и отвалами. *Современные проблемы дистанционного зондирования Земли из космоса*. 2019;16(6):31–47. <https://doi.org/10.21046/2070-7401-2019-16-6-31-47>
Savorskiy V.P., Loupian E.A., Gornyy V.I., Ermakov D.M., Panova O.Yu., Konstantinova A.M. Methods and information tools for analyzing remote sensing data to

- detect changes in vegetation caused by industrial waste and dumps. *Sovremennye Problemy Distantionnogo Zondirovaniya Zemli iz Kosmosa*. 2019;16(6):31–47. (In Russ.). <https://doi.org/10.21046/2070-7401-2019-16-6-31-47>
7. Wang Y., Zhang D., Dai G. Classification of High Resolution Satellite Images Using Improved U–Net. *International Journal of Applied Mathematics and Computer Science*. 2020;30(3):399–413. <https://doi.org/10.34768/amcs-2020-0030>
 8. Попов С.Е., Потапов В.П., Замараев Р.Ю. Разработка модуля расчёта нормализованного индекса разностной активности открытых работ на угольных месторождениях на базе радарных данных Sentinel-1. *Современные проблемы дистанционного зондирования Земли из космоса*. 2022;19(3):62–76. <https://doi.org/10.21046/2070-7401-2022-19-3-62-76>
Popov S.E., Potapov V.P., Zamaraev R.Yu. A software development approach for open pits monitoring based on Normalized Difference Activity Index and Sentinel-1 SAR data. *Sovremennye Problemy Distantionnogo Zondirovaniya Zemli iz Kosmosa*. 2022;19(3):62–76. (In Russ.). <https://doi.org/10.21046/2070-7401-2022-19-3-62-76>
 9. Hao L., Zhang Z., Yang X. Mine tailing extraction indexes and model using remote-sensing images in southeast Hubei Province. *Environmental Earth Sciences*. 2019;78(15):1–11. <https://doi.org/10.1007/s12665-019-8439-1>
 10. He D., Le B.T., Xiao D., Mao Y., Shan F., Ha T.T.L. Coal mine area monitoring method by machine learning and multispectral remote sensing images. *Infrared Physics & Technology*. 2019;103. <https://doi.org/10.1016/j.infrared.2019.103070>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Озарян Юлия Александровна, кандидат технических наук, начальник отдела Хабаровского Федерального исследовательского центра Дальневосточного отделения Российской академии наук, Хабаровск, Российская Федерация.

e-mail: ozaryanigd@gmail.com

ORCID: [0000-0002-8892-4595](https://orcid.org/0000-0002-8892-4595)

Yulia A. Ozaryan, Candidate of Engineering Sciences, Department Head of Khabarovsk federal research center of the Far Eastern Branch of the Russian Academy of Sciences, Khabarovsk, the Russian Federation.

Кожевникова Татьяна Владимировна, начальник отдела Вычислительного центра Дальневосточного отделения Российской академии наук, Хабаровск, Российская Федерация.

e-mail: ktvsl@mail.ru

ORCID: [0000-0002-2660-3104](https://orcid.org/0000-0002-2660-3104)

Tatyana V. Kozhevnikova, Department Head of Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, Khabarovsk, the Russian Federation.

Цыгулёв Кирилл Сергеевич, инженер Вычислительного центра Дальневосточного отделения Российской академии наук, Хабаровск, Российская Федерация.

e-mail: kirill.tsygulev@mail.ru

ORCID: [0000-0002-8750-1451](https://orcid.org/0000-0002-8750-1451)

Kirill S. Tsygulev, engineer of Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, Khabarovsk, the Russian Federation.

Окладников Владимир Евгеньевич, инженер Вычислительного центра Дальневосточного отделения Российской академии наук, Хабаровск, Российская Федерация.

e-mail: okladnikov03@mail.ru

Vladimir Ye. Okladnikov, engineer of Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, Khabarovsk, the Russian Federation.

*Статья поступила в редакцию 28.11.2024; одобрена после рецензирования 19.12.2024;
принята к публикации 20.01.2025.*

*The article was submitted 28.11.2024; approved after reviewing 19.12.2024;
accepted for publication 20.01.2025.*