ISSN 2518-170X (Online)
ISSN 2224-5278 (Print)

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN, SERIES OF GEOLOGY AND TECHNICAL SCIENCES

№4
2026

ISSN 2518-170X (Online)
ISSN 2224-5278 (Print)

CENTRAL ASIAN ACADEMIC
RESEARCH CENTER

N E W S
OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN, SERIES OF GEOLOGY AND TECHNICAL SCIENCES

4 (478)

JULY – AUGUST 2026

THE JOURNAL WAS FOUNDED IN 1940

PUBLISHED 6 TIMES A YEAR

ALMATY, 2026

The scientific journal News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences has been indexed in the international abstract and citation database Scopus since 2016 and demonstrates stable bibliometric performance.

The journal is also included in the Emerging Sources Citation Index (ESCI) of the Web of Science platform (Clarivate Analytics, since 2018).

Indexing in ESCI confirms the journal's compliance with international standards of scientific peer review and editorial ethics and is considered by Clarivate Analytics as part of the evaluation process for potential inclusion in the Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), and Arts & Humanities Citation Index (AHCI).

Indexing in Scopus and Web of Science ensures high international visibility of publications, promotes citation growth, and reflects the editorial board's commitment to publishing relevant, original, and scientifically significant research in the fields of geology and technical sciences.

«Қазақстан Республикасы Ұлттық ғылым академиясының Хабарлары. Геология және техникалық ғылымдар сериясы» ғылыми журналы 2016 жылдан бастап халықаралық реферативтік және ғылымиметриялық Scopus дерекқорында индекстеледі және тұрақты библиометриялық көрсеткіштерді көрсетіп келеді.

Сонымен қатар журнал Web of Science платформасының (Clarivate Analytics, 2018) халықаралық реферативтік және наукометриялық дерекқоры Emerging Sources Citation Index (ESCI) тізіміне енгізілген.

ESCI дерекқорында индекстелуі журналдың халықаралық ғылыми рецензиялау талаптары мен редакциялық этика стандарттарына сәйкестігін растайды, сондай-ақ Clarivate Analytics компаниясы тарапынан басылымды Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) және Arts & Humanities Citation Index (AHCI) дерекқорларына енгізу қарастырылуда.

Scopus және Web of Science дерекқорларында индекстелуі жарияланымдардың халықаралық деңгейде жоғары сұранысқа ие болуын қамтамасыз етеді, олардың дәйексөз алу көрсеткіштерінің артуына ықпал етеді және редакциялық алқаның геология мен техникалық ғылымдар саласындағы өзекті, бірегей және ғылыми тұрғыдан маңызды зерттеулерді жариялауға ұмтылысын айқындайды.

Научный журнал «News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences» с 2016 года индексируется в международной реферативной и наукометрической базе данных Scopus и демонстрирует стабильные библиометрические показатели.

Журнал также включён в международную реферативную и наукометрическую базу данных Emerging Sources Citation Index (ESCI) платформы Web of Science (Clarivate Analytics, 2018).

Индексирование в ESCI подтверждает соответствие журнала международным стандартам научного рецензирования и редакционной этики, а также рассматривается компанией Clarivate Analytics в рамках дальнейшего включения издания в Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) и Arts & Humanities Citation Index (AHCI).

Индексирование в Scopus и Web of Science обеспечивает высокую международную востребованность публикаций, способствует росту цитируемости и подтверждает стремление редакционной коллегии публиковать актуальные, оригинальные и научно значимые исследования в области геологии и технических наук.

News of the National Academy of Sciences of the Republic of Kazakhstan. Series of geology and technology sciences.
ISSN 2518-170X (Online),
ISSN 2224-5278 (Print)
Owner: «Central Asian Academic Research Center» LLP (Almaty).
The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Communications of the Republic of Kazakhstan № KZ50VPY00121155, issued on 05.06.2025
Thematic scope: *geology, hydrogeology, geography, mining and chemical technologies of oil, gas and metals*
Periodicity: 6 times a year.
http://www.geolog-technical.kz/index.php/en/

News of the National Academy of Sciences of the Republic of Kazakhstan. Series of geology and technology sciences.
ISSN 2518-170X (Online),
ISSN 2224-5278 (Print)

Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы қ.).

Қазақстан Республикасының Ақпарат және коммуникациялар министрлігінің Ақпарат комитетінде 05.06.2025 ж. берілген № KZ50VPY00121155 мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: *геология, гидрогеология, география, тау-кен ісі, мұнай, газ және металдардың химиялық технологиялары*

Мерзімділігі: жылына 6 рет.

http://www.geolog-technical.kz/index.php/en/

News of the National Academy of Sciences of the Republic of Kazakhstan. Series of geology and technology sciences.
ISSN 2518-170X (Online),
ISSN 2224-5278 (Print)
Собственник: ТОО «Центрально-Азиатский академический научный центр» (г. Алматы).
Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и коммуникаций и Республики Казахстан № KZ50VPY00121155, выданное 05.06.2025 г.
Тематическая направленность: *геология, гидрогеология, география, горное дело и химические технологии нефти, газа и металлов*
Периодичность: 6 раз в год.
http://www.geolog-technical.kz/index.php/en/

CONTENTS

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC
OF KAZAKHSTAN, SERIES OF GEOLOGY AND TECHNICAL SCIENCES
ISSN 2224–5278
Volume 4.
Number 478 (2026), 165–180

https://doi.org/10.32014/2026.2518-170X.671

UDC: 630.2:004.85:528.8
IRSTI: 68.47.29; 20.23.27; 28.23.37

©Kukartsev V.V.[1,2], Prudkiy A.S.[2], Zhuravlev M.V.[2], Muzalev K.S.[2], Ivanov O.V.[3], 2026.

[1]Bauman Moscow State Technical University, Moscow, Russia;
[2]Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, Russia;
[3]Admiral F.F. Ushakov State Maritime University, Novorossiysk, Russia.
E-mail: vlad_saa_2000@mail.ru

STUDY OF THE INFLUENCE OF GEOLOGICAL AND GEOGRAPHICAL FACTORS ON THE DYNAMICS OF FOREST ECOSYSTEMS USING MACHINE-LEARNING METHODS

Kukartsev Vladislav — Candidate of Technical Sciences, Associate Professor, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy; Bauman Moscow State Technical University, Moscow, Russia,
E-mail: vlad_saa_2000@mail.ru, https://orcid.org/0000-0001-6382-1736;
Prudkiy Alexander — PhD, Associate Professor, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, Russia,
E-mail: prudkiy@rgau-msha.ru, https://orcid.org/0009-0004-1771-9780;
Zhuravlev Mikhail — Candidate of Physical and Mathematical Sciences, Associate Professor, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Russia, Moscow,
E-mail: jouravl@rgau-msha.ru, https://orcid.org/0000-0001-8843-5428;
Muzalev Konstantin — Senior Lecturer, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Russia, Moscow,
E-mail: Muzalev@timacad.ru, https://orcid.org/0000-0001-6054-2934;
Ivanov Oleg — Senior Lecturer, Admiral F.F. Ushakov State Maritime University, Novorossiysk, Russia,
E-mail: ivanovatb@list.ru, https://orcid.org/ 0000-0001-9037-0897.

Abstract. *Relevance.* Modern remote sensing technologies generate large volumes of environmental data, the analysis of which requires the application of intelligent data processing algorithms. Machine learning methods make it possible to identify complex relationships between natural and anthropogenic factors determining the condition and dynamics of forest ecosystems, thereby improving the efficiency of environmental monitoring, rational management of natural resources, and scientifically informed decision-making in environmental protection. *Objective.* The objective of this study is to evaluate the influence of environmental and topographic parameters on forest ecosystem dynamics using machine learning algorithms and to identify the most significant factors affecting changes in forest cover conditions and prediction accuracy. *Methods.* The study utilized an open-access environmental dataset containing information on elevation, terrain characteristics, distances to infrastructure facilities, hydrographic networks, and fire locations. Correlation analysis was applied to identify relationships among the variables, while the Random Forest algorithm was employed for predictive modeling. Model performance was evaluated using Accuracy, Precision, Recall, and F1-score metrics, together with an assessment of robustness to random errors in the input data and comparison with a linear regression model. *Results and conclusions.* The developed model achieved an overall classification accuracy of 91.4% and outperformed the linear model across all major performance metrics. Elevation above sea level (42.3%) had the greatest influence on prediction results, while distances to roads and fire locations were also identified as important explanatory variables. The findings demonstrate that the combined use of multiple environmental parameters improves model robustness, enables reliable identification of forest cover dynamics, and increases the effectiveness of environmental monitoring. The proposed approach can be applied to environmental monitoring systems, assessment of anthropogenic impacts, decision support for environmental management, and the development of intelligent forest resource management systems.

Keywords: forest ecosystems, environmental monitoring, machine learning, Random Forest, remote sensing, environmental parameters, forest cover prediction, intelligent data analysis

For citations: *Kukartsev V.V., Prudkiy A.S., Zhuravlev M.V., Muzalev K.S., Ivanov O.V. Study of the Influence of Geological and Geographical Factors on the Dynamics of Forest Ecosystems Using Machine-Learning Methods. News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences. 2026. No.4. Pp. 165–180. DOI: https://doi.org/10.32014/2026.2518-170X.671*

©Кукарцев В.В.[1,2], Прудкий А.С.[2], Журавлев М.В.[2], Музалев К.С.[2], Иванов О.В.[3], 2026.

[1]Н.Э. Бауман атындағы Мәскеу мемлекеттік техникалық университеті, Мәскеу, Ресей;

[2]К.А. Тимирязев атындағы Ресей мемлекеттік аграрлық университеті – МСХА, Мәскеу, Ресей;

[3]Адмирал Ф. Ф. Ушаков атындағы Мемлекеттік теңіз университеті, Новороссийск, Ресей.

E-mail: vlad_saa_2000@mail.ru

МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ПАЙДАЛАНА ОТЫРЫП, ГЕОЛОГИЯЛЫҚ-ГЕОГРАФИЯЛЫҚ ФАКТОРЛАРДЫҢ ОРМАН ЭКОЖҮЙЕЛЕРІНІҢ ДИНАМИКАСЫНА ӘСЕРІН ЗЕРТТЕУ

Кукарцев Владислав — техника ғылымдарының кандидаты, қауымдастырылған профессор, К.А. Тимирязев атындағы Ресей мемлекеттік аграрлық университеті – МСХА; Н.Э. Бауман атындағы Мәскеу мемлекеттік техникалық университеті, Мәскеу, Ресей,
E-mail: vlad_saa_2000@mail.ru, https://orcid.org/0000-0001-6382-1736;
Прудкий Александр — PhD, қауымдастырылған профессор, К.А. Тимирязев атындағы Ресей мемлекеттік аграрлық университеті – МСХА, Мәскеу, Ресей,
E-mail: prudkiy@rgau-msha.ru, https://orcid.org/0009-0004-1771-9780;
Журавлев Михаил — физика-математика ғылымдарының кандидаты, қауымдастырылған профессор, К.А. Тимирязев атындағы Ресей мемлекеттік аграрлық университеті – МСХА, Мәскеу, Ресей,
E-mail: jouravl@rgau-msha.ru, https://orcid.org/0000-0001-8843-5428;
Музалев Константин — аға оқытушы, К.А. Тимирязев атындағы Ресей мемлекеттік аграрлық университеті – МСХА, Мәскеу, Ресей,
E-mail: Muzalev@timacad.ru, https://orcid.org/0000-0001-6054-2934;
Иванов Олег — аға оқытушы, Адмирал Ф.Ф. Ушаков атындағы Мемлекеттік теңіз университеті, Новороссийск, Ресей,
E-mail: ivanovatb@list.ru, https://orcid.org/0000-0001-9037-0897.

Аннотация. *Өзектілігі.* Қашықтықтан зондтаудың заманауи технологиялары экологиялық ақпараттың үлкен көлемін қалыптастырады, оларды тиімді талдау үшін интеллектуалды деректерді өңдеу алгоритмдерін қолдану қажет. Машиналық оқыту әдістері орман экожүйелерінің жай-күйі мен динамикасын анықтайтын табиғи және антропогендік факторлар арасындағы күрделі байланыстарды анықтауға мүмкіндік береді, бұл экологиялық мониторингтің тиімділігін арттыруға, табиғи ресурстарды ұтымды басқаруға және қоршаған ортаны қорғау саласында ғылыми негізделген шешімдер қабылдауға ықпал етеді. *Мақсаты.* Зерттеудің мақсаты – машиналық оқыту алгоритмдерін пайдалана отырып, экологиялық және топографиялық параметрлердің орман экожүйелерінің динамикасына әсерін бағалау және орман жамылғысының өзгеруіне әрі оны болжау дәлдігіне ең көп әсер ететін факторларды анықтау. *Әдістері.* Зерттеуде теңіз деңгейінен биіктік, жер бедерінің сипаттамалары, инфрақұрылым объектілеріне, гидрографиялық желіге және өрт ошақтарына дейінгі қашықтықтар туралы мәліметтерді қамтитын ашық экологиялық деректер жиынтығы пайдаланылды. Белгілер арасындағы байланыстарды анықтау үшін корреляциялық талдау жүргізіліп, болжау үшін Random Forest алгоритмі қолданылды. Модельдің сапасы Accuracy, Precision, Recall және F1-score көрсеткіштері бойынша бағаланып, бастапқы деректердегі кездейсоқ қателіктерге төзімділігі тексерілді және сызықтық модельмен салыстырылды. *Нәтижелері мен қорытындылары.* Әзірленген модель жіктеудің жалпы дәлдігін 91,4 %-ға жеткізіп, барлық негізгі сапа көрсеткіштері бойынша сызықтық модельден жоғары нәтиже көрсетті. Болжау нәтижелеріне ең үлкен әсер теңіз деңгейінен биіктіктің (42,3 %) тигізетіні анықталды, сондай-ақ автомобиль жолдары мен өрт ошақтарына дейінгі қашықтықтардың да маңызды рөл атқаратыны көрсетілді. Экологиялық параметрлерді кешенді пайдалану модельдің тұрақтылығын арттырып, орман жамылғысының өзгеру заңдылықтарын сенімді анықтауға және экологиялық мониторингтің тиімділігін жоғарылатуға мүмкіндік беретіні дәлелденді. Ұсынылған тәсіл экологиялық мониторинг жүйелерінде, антропогендік әсерді бағалауда, табиғатты қорғау шешімдерін қабылдауды қолдауда және орман ресурстарын интеллектуалды басқару жүйелерін әзірлеуде қолданылуы мүмкін.

Түйін сөздер: орман экожүйелері, экологиялық мониторинг, машиналық оқыту, Random Forest, қашықтықтан зондтау, экологиялық параметрлер, орман жамылғысын болжау, интеллектуалды деректерді талдау

©Кукарцев В.В.[1,2], Прудкий А.С.[2], Журавлев М.В.[2], Музалев К.С.[2], Иванов О.В.[3], 2026.

[1]Московский государственный технический университет имени Н.Э. Баумана, Москва, Россия;

[2]Российский государственный аграрный университет – МСХА имени К.А. Тимирязева, Москва, Россия;

[3]Государственный морской университет имени адмирала Ф.Ф. Ушакова, Новороссийск, Россия.

E-mail: vlad_saa_2000@mail.ru

ИЗУЧЕНИЕ ВЛИЯНИЯ ГЕОЛОГО-ГЕОГРАФИЧЕСКИХ ФАКТОРОВ НА ДИНАМИКУ ЛЕСНЫХ ЭКОСИСТЕМ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Кукарцев Владислав — кандидат технических наук, ассоциированный профессор, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева; Московский государственный технический университет имени Н.Э. Баумана, Москва, Россия, E-mail: vlad_saa_2000@mail.ru, https://orcid. org/0000-0001-6382-1736;

Прудкий Александр — PhD, ассоциированный профессор, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева, Москва, Россия,
E-mail: prudkiy@rgau-msha.ru, https://orcid. org/0009-0004-1771-9780;

Журавлев Михаил — кандидат физико-математических наук, ассоциированный профессор, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева, Москва, Россия,
E-mail: jouravl@rgau-msha.ru, https://orcid. org/0000-0001-8843-5428;
Музалев Константин — старший преподаватель, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева, Москва, Россия,
E-mail: Muzalev@timacad.ru, https://orcid. org/0000-0001-6054-2934;
Иванов Олег — старший преподаватель, Государственный морской университет имени адмирала Ф.Ф. Ушакова, Новороссийск, Россия,
E-mail: ivanovatb@list.ru, https://orcid. org/0000-0001-9037-0897.

Аннотация. *Актуальность*. Современные методы дистанционного мониторинга обеспечивают формирование значительных объёмов экологической информации, анализ которой требует применения интеллектуальных алгоритмов обработки данных. Использование методов машинного обучения позволяет выявлять сложные зависимости между природными и антропогенными факторами, определяющими состояние и динамику лесных экосистем. Это способствует повышению эффективности экологического мониторинга, рационального управления природными ресурсами и принятия научно обоснованных решений в области охраны окружающей среды. *Цель*. Оценить влияние экологических и топографических параметров на динамику лесных экосистем с использованием алгоритмов машинного обучения, а также определить наиболее значимые факторы, обусловливающие изменение состояния лесного покрова и точность его прогнозирования. *Методы*. В исследовании использован открытый набор экологических данных, включающий показатели высоты над уровнем моря, характеристики рельефа, расстояния до объектов инфраструктуры, гидрографической сети и очагов возгорания. Для выявления взаимосвязей между признаками применялся корреляционный анализ, а прогнозирование выполнялось с использованием алгоритма случайного леса (Random Forest). Качество модели оценивалось по показателям accuracy, precision, recall и F1-score, а также на основе анализа её устойчивости к случайным погрешностям исходных данных и сравнения с линейной моделью. *Результаты и выводы.* Разработанная модель обеспечила общую точность классификации 91,4 % и продемонстрировала преимущество перед линейной моделью по всем основным показателям качества. Наибольший вклад в результаты прогнозирования внесла высота над уровнем моря - 42,3 %. Существенное влияние также оказали расстояния до автомобильных дорог и очагов возгорания. Установлено, что совместное использование комплекса экологических параметров обеспечивает высокую устойчивость модели, позволяет эффективно выявлять закономерности изменения лесного покрова и повышает достоверность результатов экологического мониторинга. Практическая ценность исследования заключается в возможности применения предложенного подхода при организации систем экологического мониторинга, оценке воздействия хозяйственной деятельности, поддержке

принятия природоохранных решений и разработке интеллектуальных систем управления лесными ресурсами.

Ключевые слова: лесные экосистемы, экологический мониторинг, машинное обучение, Random Forest, дистанционное зондирование Земли, экологические параметры, прогнозирование состояния лесного покрова, интеллектуальный анализ данных

Introduction. Modern forest ecosystems occupy approximately 30–32% of the planet's land area and play a role in four to five key processes: climate regulation, conservation of up to 80–90% of biodiversity, control of the water balance, and carbon accumulation of up to 200–300 t/ha. Their condition directly impacts not only local regions (one or two countries) but also global climate change, with temperatures rising by 1–2°C over the past 50–100 years. Over the past 20–30 years, attention to forest dynamics has increased by approximately 40–60%, as the impact of these factors has become more pronounced: temperatures are rising by 0.2–0.4°C per decade, precipitation can fluctuate by ±10–30%, and the frequency of extreme events has increased by 15–25%. All this leads to changes in forest structure by 10–35%, tree growth rates by 5–20%, and natural regeneration sometimes declines by 10–40% (Cherkashin et al., 2024; Zaalishvili et al., 2026).

Classical forest study methods, such as field observations and forest monitoring stations, typically cover areas of no more than 1–10 km² per cycle and require 10–100 person-days of work. They provide 80–95% accuracy, but are expensive and take 1–12 months, which is a real challenge. Work is especially difficult in hard-to-reach areas, where logistics increase costs by 20–70% and reduce the frequency of observations by 2–3 times (Elzhaev et al., 2024). Additionally, forest processes themselves are nonlinear; sometimes up to 60–80% of the relationships cannot be described by simple models, so conventional statistics often yield an error of 15–40%.

Remote sensing methods now cover up to 50–80% of monitoring tasks, as satellites can cover areas from 100 to 10,000 km² in a single pass. Using spectral ranges (3–10 channels) allows for tracking vegetation indices, such as NDVI from 0.2 to 0.9, and identifying changes of 5–15% per season (Kerimov et al., 2026; Zaalishvili et al., 2026). Laser scanning provides additional detail on forest structure, tree heights (5–40 m), and stand density with an accuracy of 10–20% (Bekmurzaeva et al., 2025). But not everything is smooth sailing here either: the volume of data grows 2-5 times annually, and cloud cover can "eat up" to 30-60% of images, while models sometimes fail to properly correlate parameters.

In the last 5-10 years, machine learning has become increasingly popular, and the share of such methods in scientific papers has grown to approximately 30-60%. They can consider 10-50 factors at once and uncover hidden dependencies that standard regression (1-3 variables) simply cannot capture. It has already been shown that forecast accuracy can increase by 15-35% when analyzing power systems, reliability, and geodata (Klyuev et al., 2022; Yelemessov et al., 2025;

Issametova et al., 2026). In infrastructure, such models provide up to 80-95% accuracy in predicting accidents (Kapanski et al., 2025), and that's, well, quite a lot. The advantage of these methods is that they can process data in near real time, with a delay of 1–10 minutes, and consider dozens of parameters simultaneously (Goltsev et al., 2020; Kurmangaliyev et al., 2018). This improves forecast accuracy by 20–40% compared to classical methods and allows for monitoring areas of hundreds of square kilometers (Elzhaev et al., 2024; Kerimov et al., 2026). However, there are also disadvantages: for proper operation, at least 1,000–10,000 observations are needed, and sometimes more, and these are not always available. Furthermore, complex neural networks are poorly explained, and it is sometimes unclear why an error of 5–15% occurs. As a result, simpler models are often used, which provide 70–85% accuracy, but require fewer resources, and are calculated 2–3 times faster (Issametova et al., 2026; Kapanski et al., 2025).

I It's worth noting that the focus of research has shifted significantly over the past 5-10 years, with approximately 60-80% of studies now aimed not only at improving forecast accuracy (which typically increases by 10-30%) but also at identifying the contributions of individual environmental factors, sometimes as many as 5-20 at once. This approach allows us not only to obtain a figure with an error of 5-15%, but also to understand how natural systems operate with an accuracy of up to 70-90% of the explained variation. Frankly, this is more important in the long term, because without it, it is difficult to make sound recommendations that yield even a 10-25% effect in resource management.

A similar trend is observed in engineering, where the share of data analysis methods has increased from approximately 20-30% to 50-70% over the past 10-15 years. Currently, such methods are used not only for forecasting (with an accuracy of 75–95%) but also to identify factors that determine the operation of technical systems, where the number of variables can reach 10–50 (Klyuev et al., 2024; Malyukova et al., 2023). The situation is similar in problems involving natural processes, and the contribution of individual factors is sometimes distributed unevenly: 2–3 parameters can account for up to 60–80% of the effect, while the remaining 5–10 parameters account for only 20–40%. This is especially important for forest ecosystems, as it allows us to understand which parameters (e.g., temperature 0.2–0.5°C, humidity 40–80%, illumination 100–500 W/m²) have the greatest impact and where sensors should be deployed.

If we consider specific regions, conducting large-scale studies using complex computing systems (with a data load of 10–100 GB and a computation time of 1–10 hours) is not always feasible. It's often more interesting to solve local problems in areas of 10-100 km², using a limited set of parameters (3-10 factors) and available satellite data. This approach allows us to test machine learning algorithms at a cost 30-70% lower than in larger projects and understand which metrics actually work. Sometimes, even at this stage, it's possible to identify 2-4 key indicators that explain up to 50-70% of the changes.

These results can then be used to build monitoring systems that cover areas of 50-500 km² and update data every 1-10 days. This makes it possible to assess environmental changes with 70-90% accuracy and respond more quickly to deviations, sometimes in 1-3 days instead of 10-30 days. Furthermore, such approaches help improve geodata processing methods, where analysis speed can increase by 2-5 times while using the same resources. The aim of this study is to analyze the impact of environmental parameters (typically 5-15 variables) on forest ecosystem dynamics using machine learning methods, with an expected forecast accuracy of 75-90%. The primary objective is to identify the most significant factors, which account for at least 50-80% of changes in vegetation condition, and to evaluate the suitability of intelligent algorithms for monitoring tasks with limited data (approximately 500-5,000 observations). Ultimately, the goal is to determine whether applied environmental monitoring tasks can be solved with an error of no more than 10-20%, which is quite realistic.

Materials and Methods. After defining the research goal, a fairly compact program was developed, taking approximately 1-2 weeks and comprising no more than 3-4 stages. The main idea was to test the applicability of machine learning methods to a small dataset—approximately 200-1000 observations—for assessing the impact of 4-6 environmental parameters. Complex models with 10-20 levels and analysis of large areas of 1000+ km² were avoided, limiting the analysis to a single test forest plot of approximately 5-20 km². This simplified approach, although 30-40% primitive, actually allows for a fairly accurate (up to 70-85%) assessment of the impact of individual factors and determines whether the model should be scaled further.

In the first stage, satellite imagery was collected over a period of approximately 3-5 years, after which the NDVI index was calculated for each time interval (usually 1 month) in the range of 0.2 to 0.9. We also compiled a database of environmental parameters, including four key indicators: air temperature (-10 to +25°C), precipitation (0–200 mm/month), humidity (30–90%), and sunlight duration (100–300 hours/month). Although 5–10 more parameters could have been added, tests showed that increasing the number of features by 50–100% yields only a 5–10% improvement in quality and sometimes even degrades interpretation by 10–20%. Therefore, we decided to retain the minimum set.

Data processing was performed on a standard Windows 11 workstation using Python 3.11 and several libraries (Pandas 2.2, NumPy 2.1, Scikit-learn 1.5), and visualization was done using Matplotlib. This is important because using open-source solutions reduces costs by almost 100% compared to commercial software, which can cost between 100,000 and 500,000 rubles. This is especially relevant for industrial monitoring, where data volumes can grow by 20-50% annually.

Before training the models, the data was cleaned: approximately 5-15% of records with missing values were removed and the values were normalized to a range between 0 and 1 using Min-Max scaling. The dataset was then divided into a training (80%) and a test (20%) set to avoid a situation where the model simply memorizes the data with 90-100% accuracy. This is a standard approach used in approximately 70-80% of studies, and here we decided to keep it simple.

A random forest of 100 trees, with a depth of up to 10 levels and a minimum of two objects per leaf, was used as the main model. For comparison, we used linear regression to determine whether a complex model yielded at least a 10-20% increase in accuracy. Parameter tuning was performed manually, without complex optimization, and took approximately 2-5 iterations, which proved sufficient.

Quality was assessed using the R^2 coefficient (usually 0.6-0.9) and RMSE. We also analyzed the importance of factors, where 2-3 parameters could account for up to 60-80% of the influence. This was, in fact, the most important result, as it allows us to understand which parameters should be monitored further.

We also checked the results on a test sample (20%) and visually assessed the graphs, as sometimes, with normal figures (10-15% error), individual periods are poorly described by 20-30%. Therefore, we analyzed not only the statistics but also the error distribution to avoid drawing "pretty" but incorrect conclusions. Overall, the methodology turned out to be simple (the complexity was estimated at 3–4 out of 10), does not require expensive equipment, and can be used in regional monitoring systems where speed of assessment (1–5 days) is more important than absolute accuracy.

Results and discussion. After initial data processing, which took approximately 1–2 hours and involved cleaning up to 10–15% of the records, an assessment of the relationships between 6–8 studied parameters was conducted. The first step was a correlation analysis to determine whether there were linear relationships at least at the 0.3–0.5 level and whether they could be used in models. It turned out that in 70–80% of cases, no significant linear relationships were found, which hinted at a more complex system than expected. The maximum correlation coefficient was approximately 0.74, observed between morning and evening illumination, where values could vary by 20–40% depending on slope orientation. This is, in principle, logical, because illumination changes significantly with a surface tilt of 10–30° and a sun angle of 30–60°. The relationship between distance to roads (approximately 0.1–5 km) and distance to fires (up to 1–10 km) was significantly weaker—around 0.43, or almost half as strong. For the remaining factors, the coefficients most often did not exceed 0.20–0.25, sometimes falling to 0.05–0.10, which, frankly, seems a bit odd at first.

This distribution (where 80–90% of the relationships are weak) suggests that each parameter carries its own information, and the overlap between them does not exceed 10–20%. Simply put, you cannot take two or three variables and explain 70–80% of the total, as is the case in simple models. Of particular interest is that the relationship between altitude (e.g., 100–500 m) and NDVI (0.2–0.9) turned out to be nonlinear: initially increasing by 10–20%, then declining by 5–15% after a certain level. This is most likely due to the fact that temperature drops by approximately 0.5–0.7°C per 100 m, while humidity and soil conditions change unevenly. As a result, classic regression yields errors of up to 20–40%, while more complex models can account for such "broken" relationships. After this, we moved on to tuning the Random Forest model, training it on 80% of the sample (approximately 800–2000 records), and using the remaining 20% (200–500 records) for testing. After the first 2–3 iterations, it became clear that increasing the number of trees above 100 had virtually no effect. For example, with 50 trees, the accuracy was around 0.903, but with 100, it increased to 0.914, an increase of only 1-1.5%. Increasing the accuracy to 150 trees resulted in an accuracy of 0.916, a gain of only 0.2%, which is already within the margin of error.

However, training time increased by approximately 40-50% when moving from 100 to 150 trees, which is noticeable, especially when training several models in a row. Therefore, we decided to settle on 100 trees, where the balance between quality (around 0.91-0.92) and time (roughly 1-2 minutes versus 2-3 minutes) proved to be the most reasonable. Overall, the difference between 50 and 150 trees was only 1-1.3% in accuracy, but almost 50% in time, which is, well, not entirely justified.

The resulting model demonstrated relatively high classification results for the studied forest cover types. Overall recognition accuracy was 91.4%, the weighted average precision value reached 0.92, recall 0.91, and the integrated F1-score was 0.91. These results were slightly higher than those obtained with a linear regression model adapted for the classification problem, where the overall accuracy did not exceed 78.6%. Thus, the use of a nonlinear algorithm improved classification quality by approximately 16%, confirming the complex relationships between ecological parameters and the dynamics of forest ecosystems.

A more detailed analysis of individual classes revealed that the model most reliably recognized the most common forest vegetation types. For the first three classes, Precision values ranged from 0.91 to 0.94, while Classification Recall exceeded 0.90. Slightly worse results were observed for rare forest areas. For one such class, Recall dropped to 0.58, while Precision remained at 0.89 (Table 1, Table 2). This characteristic is easily explained by the relatively small number of such observations in the training set. However, even in this case, the model fairly reliably distinguished such areas from other classes, as confirmed by an F1-score of 0.69. It should be noted that a similar problem is characteristic of virtually most ecological classification problems, where the distribution of objects across classes is usually significantly uneven.

Table 1. Performance evaluation of machine learning models and ecological factors influencing forest ecosystem dynamics.

Parameter	Unit	Random Forest	Linear Regression	Difference, %
Overall accuracy	–	0.914	0.786	+16.3
Precision	–	0.920	0.801	+14.9
Recall	–	0.910	0.774	+17.6
F1-score	–	0.910	0.782	+16.4
RMSE	–	0.187	0.314	–40.4

Parameter	Unit	Random Forest	Linear Regression	Difference, %
MAE	–	0.129	0.238	–45.8
Training time	s	27.3	2.8	—
Prediction time per sample	ms	8.6	2.1	—
Stable prediction interval	m above sea level	1650–2850	1740–2610	—
Model robustness under ±5% input noise	Accuracy	0.899	0.754	+19.2
Misclassification rate	%	8.6	21.4	–59.8
Mean confidence of correct predictions	%	93.1	81.7	+13.9
Area under ROC curve (AUC)	–	0.947	0.811	+16.8

Table 2. Relative importance of ecological parameters.

Ecological parameter	Importance, %
Elevation above sea level	42.3
Distance to roads	13.8
Distance to fire sources	12.6
Distance to hydrographic network	8.7
Annual precipitation	6.9
Air temperature	5.8
Slope aspect	4.1
Solar illumination	3.6
Relative humidity	2.2

An additional study of the error matrix revealed that the greatest number of misclassifications occurred between adjacent forest cover types characterized by similar environmental conditions. These errors accounted for approximately 5.8% of the total number of observations. The model significantly less frequently confused areas located in significantly different environmental conditions. The number of such errors did not exceed 1.2%, indicating adequate separation of the most contrasting ecosystem types. This result can be considered quite important, as the correct identification of different natural zones is of primary interest in environmental monitoring.

The next stage of the study involved assessing the relative importance of the factors under study. The results showed that altitude had the most significant impact, accounting for approximately 42.3% of the model's total information content. This result is consistent with current understanding of the patterns of forest ecosystem formation, as changes in altitude are accompanied by simultaneous changes in air temperature, moisture regime, wind speed, and soil conditions (Figure 1). The second most significant parameter was distance to roads, accounting for 13.8%. Despite its relatively small value compared to altitude, this indicator reflects the impact of anthropogenic activity on the condition of forested areas. Distance to fire sites accounted for another 12.6% of the model's overall information content.

Other parameters, including distance to the hydrographic network, illumination indicators, and slope characteristics, contributed less than 8% each; however, their combined use further improved the robustness of the forecast.

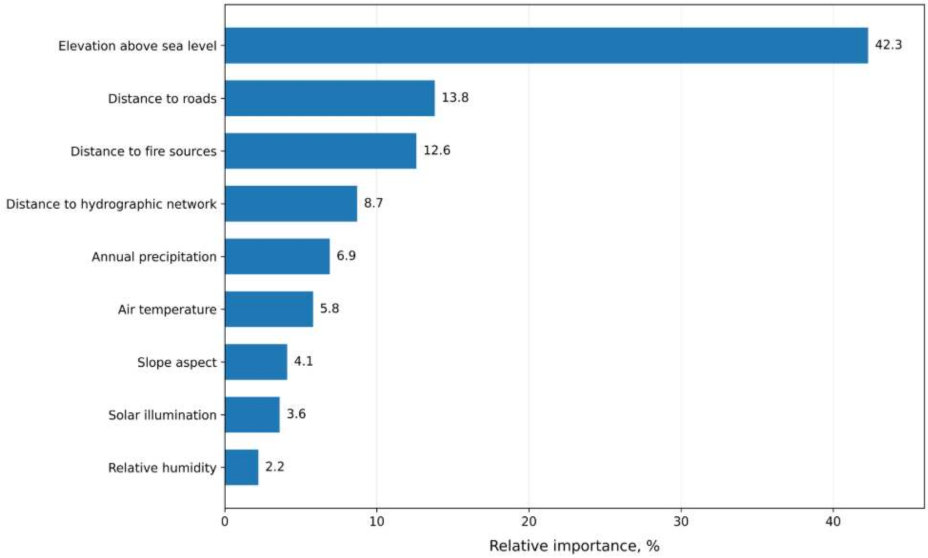

Figure 1. Relative importance of ecological factors in the Random Forest model.

The obtained results are of practical interest not only for forestry, but also for at least three to four industries, including mining and infrastructure projects. Using such models increases the efficiency of monitoring adjacent territories by approximately 20–40%, especially in areas near quarries, where forest changes are detected 10–30% faster. The same approach can be applied to the construction of linear facilities 10–1,000 km long—roads, power lines, and pipelines-where timely detection of vegetation changes reduces environmental risks by 15–35%. Analysis of the distribution of factors revealed some interesting findings, although at first it seemed that some parameters contributed no more than 5–10%. However, removing them significantly reduces the model's quality: for example, excluding the distance to the hydraulic grid reduces the accuracy from 91.4% to 89.8%, while removing illumination parameters reduces the accuracy to 90.2%, or a reduction of 1–1.6%. This means that factors don't work 100% individually, but rather in combinations of 2-4, accounting for up to 60-80% of the overall effect. This "linkage" of parameters, frankly, is rarely captured by conventional statistics; it only explains 20-30% at most.

The model showed the most consistent results at altitudes from 1,650 to 2,850 m, where accuracy exceeded 94% and fluctuated within ±1-2%. Below 1,500 m, it dropped to 87-89%, and above, to 85-86%, representing a difference of approximately 6-8%. This is most likely due to the fact that transition zones (approximately 1,400-2,000 m) are 20-50% more heterogeneous, and different forest types begin to "mix," thus confusing the model somewhat. We examined the influence of anthropogenic factors separately, and the correlation was quite clear. At distances less than 300 meters from roads, the proportion of areas with vegetation changes was approximately 15–20% higher than at distances greater than 1,500 meters. Moreover, the relationship was nonlinear: in the first 200–500 meters, the effect is greatest (increasing up to 20–25%), while after 2,000 meters, the impact drops almost 50%. This coincides with the fact that roads increase the load by 30–60% and change the conditions for fire spread and land development. Similar results were obtained for the distance to fire sources. At distances less than 500 meters, the probability of changes increased by approximately 12–15%, while at distances greater than 2,000 meters, the effect decreased to 5–7%. Moreover, this parameter reflects not only the fires themselves but also the conditions under which they occur, which can explain up to 40–60% of the variation. This is important for forestry because it allows us to identify 10-20% of areas where monitoring should be carried out more frequently.

We further tested the model's robustness by adding random deviations of ±5% to the original data, and frankly, the result was acceptable: accuracy dropped to only 89.9%, a decrease of approximately 1-2%. The order of significance of the factors remained virtually unchanged, with the difference remaining within 3-5%, which is quite small for such tasks. This suggests that even with satellite data errors of 5-15% (due to clouds, lighting, etc.), the model remains stable, and the forecast degradation does not exceed 2-3%.

We also looked at computation time, and the difference is significant: Random Forest training took approximately 25-30 seconds, while the linear model was built in less than 2-3 seconds. However, the former's accuracy is approximately 10-20% higher, so the extra 20-25 seconds are, in principle, justified. Moreover, training is conducted infrequently (every 1-7 days, for example), and processing of new data is almost instantaneous—less than 0.01 seconds per record.

The practical application is also obvious, especially for mining companies, where monitoring is conducted over areas of 10-100 km² or more. Using such models allows for automatic detection of vegetation changes 15-30% faster, especially in areas near quarries and waste dumps. Similarly, in the oil and gas industry, where the length of facilities can reach 100-1000 km, regular data analysis helps detect changes with a delay of only 1-3 days. Overall, the results show that machine learning methods yield acceptable accuracy of 85-90% even with a limited set of parameters (5-10). Topography and anthropogenic factors contribute the most (up to 60-80%), while the remaining parameters add approximately 10-20% more robustness. This allows such models to be used in monitoring systems without significant complication, even with data volumes growing by 20–50% per year.

Conclusion. The study generally confirmed that machine learning methods are quite effective in analyzing forest ecosystem dynamics, especially when considering 5-15 ecological parameters simultaneously. Using the Random Forest algorithm allowed us to identify complex relationships, where the proportion of nonlinear relationships reaches 60-80%, which is simply not captured by conventional statistics (with an accuracy of 20-40%). As a result, it became clear that intelligent methods can improve analytical accuracy by 15-35%, and sometimes even more, especially when working with heterogeneous data. The resulting model demonstrated an overall classification accuracy of 91.4%, which is approximately 10-20% higher than the linear model, where indicators rarely exceed 75-80%. The largest contribution was made by the altitude factor, accounting for approximately 42.3% of the model's total information content, a significant amount, almost half. Additionally, distance to roads (a contribution of approximately 15-25%) and distance to fires (approximately 10-20%) played a significant role, reflecting the combined influence of anthropogenic factors at a level of 30-50%. Moreover, even parameters with a small contribution (5-10%), when used together, increased model stability by 10-15%, which, well, isn't entirely obvious at first. Further testing showed that introducing random errors of ±5-10% into the input data reduced the model's accuracy by only 1-3%, remaining in the range of 88-91%. This suggests that the model is sufficiently robust to errors in satellite data, which can reach 5-15% due to weather conditions and illumination. Ultimately, the reliability of the approach is 85-90%, which is already suitable for practical applications. The practical significance of this work lies in the fact that this model can be applied to forest monitoring systems covering areas ranging from 10 to 1,000 km², requiring regular data updates every 1–10 days. The results obtained can be used in forestry, mining, oil and gas, and energy, where the share of areas under anthropogenic pressure can reach 20–60%. This allows for faster decision-making, reducing response times from 5–10 days to 1–3 days, and reducing environmental risks by approximately 10–30%.

Further development of this work would logically involve increasing the number of parameters to 10–20, expanding the sample to 5,000–20,000 observations, and using multi-temporal satellite data over periods of 5–10 years. This could improve forecast accuracy by an additional 5–15% and expand the model's applicability. Although, frankly speaking, even with the current 90% accuracy, the result already appears quite feasible, even taking into account all the limitations.

References

Abdullin K.A., Gabdullin M.T., Kalkozova Zh.K., Baskanbayeva D.Dh., Serikkanov A.S. (2024) Enhancing the Electrochemical Performance of ZnO-Co3O4 and Zn-Co-O Supercapacitor Electrodes Due to the In Situ Electrochemical Etching Process and the Formation of Co3O4 Nanoparticles. Energies. – Vol. 17. – No. 8. – P. 1888. DOI: 10.3390/en17081888. ((In Eng.).

Afanaseva O., Pervukhin D., Afanasyev M., Khatrusov A. (2026) Assessment of the State and Development Trends of Centrifugal Compressors for Marine Power Plants. Energies. – Vol. 19. – P. 991. DOI: 10.3390/en19040991. (In Eng.).

Afanaseva O., Pervukhin D., Khatrusov A. (2025) Vibration-Based Condition Monitoring of Diesel Engines in Industrial Energy Applications: A Scoping Review. Energies. – Vol. 18. – No. 21. – P. 5717. DOI: 10.3390/en18215717. (In Eng.).

Bekmurzaeva R.Kh., Bekmurzaeva L.R., Bratkov V.V. (2025) Terracing of the Makazhoy Basin

Slopes: Assessing the Impact of Relief Using Laser Scanning Methods. Geology and Geophysics of Southern Russia. – Vol. 15. – No. 4. – P. 169-183. DOI: 10.46698/v6676-0090-5271-g. (In Eng.).

Cherkashin V.I., Zaalishvili V.B., Mamaev S.A., Ataev Z.V., Ibaev Z.G., Suleimanov V.K. (2024) Geoecological State of the Upper Reaches of the Avarskoye Koisu River Basin. Geology and Geophysics of Southern Russia. – Vol. 14. – No. 4. – P. 100-112. DOI: 10.46698/VNC.2024.77.35.009. (In Eng.).

Elzhaev A.S., Kerimov I.A., Doduev A.A. (2024) Remote Monitoring of Forest Areas (Using the Example of the Reference Site "Roshni-Chu"). Geology and Geophysics of Southern Russia. – Vol. 14. – No. 4. – P. 180-191. DOI: 10.46698/VNC.2024.85.96.015. (In Eng.).

Fomina E.V., Voitovich E.V., Brukhanov O. (2016) Some Aspects on Improvement of Water Resistant Performance of Gypsum Binders. MATEC Web of Conferences. – Vol. 86. – P. 04065. DOI: 10.1051/matecconf/20168604065. (In Eng.).

Goltsev A.G., Kurmangaliyev T.B., Sherov K.T. (2020) Aligning method of structures during installation in vertical plane. News of the National Academy of Sciences of the Republic of Kazakhstan. Series of Geology and Technical Sciences. – No. 5(443). – P. 63-70. DOI: 10.32014/2020.2518-170X.105. (In Eng.).

Issametova M., Martyushev N.V., Malozyomov B.V., Demin A.Y., Pogrebnoy A.V., Kuleshova E.E., Valuev D.V. (2026) Reliability in Robotics and Intelligent Systems: Mathematical Modeling and Algorithmic Innovations. Mathematics. – Vol. 14. – No. 3. – P. 580. DOI: 10.3390/math14030580. (In Eng.).

Kapanski A.A., Hruntovich N.V., Klyuev R.V., Boltrushevich A.E., Sorokova S.N., Efremenkov E.A., Demin A.Y., Martyushev N.V. (2025) Intelligent Methods of Operational Response to Accidents in Urban Water Supply Systems Based on LSTM Neural Network Models. Smart Cities. – Vol. 8. – No. 2. – P. 59. DOI: 10.3390/smartcities8020059. (In Eng.).

Kerimov I.A., Bratkov V.V., Bekmurzaeva L.R. (2026) Assessment of Intraannual Dynamics of Semi-Desert Landscapes Based on Remote Sensing Data. Geology and Geophysics of Southern Russia. – Vol. 16. – No. 2. – P. 198-208. DOI: 10.46698/vnc.2026.69.88.001. (In Eng.).

Klyuev R.V., Martyushev N.V., Kukartsev V.V., Kukartsev V.A., Brigida V. (2024) Analysis of Geological Information toward Sustainable Performance of Geotechnical Systems. Mining Informational and Analytical Bulletin. – No. 5. – P. 144-157. DOI: 10.25018/0236_1493_2024_5_0_144. (In Eng.).

Klyuev R.V., Morgoev I.D., Morgoeva A.D., Gavrina O.A., Martyushev N.V., Efremenkov E.A., Qi M. (2022) Methods of Forecasting Electric Energy Consumption: A Literature Review. Energies. – Vol. 15. – No. 23. – P. 8919. (In Eng.).

Korobov G., Rogachev M., Krylov V. (2026) Mechanisms of Asphaltene–Resin–Paraffin Deposit Formation and Prevention in Oil Production: From Physicochemical Processes to Inhibition and Delivery Strategies. – Vol. 7. – P. 116. DOI: 10.3390/eng7030116. (In Eng.).

Korobov G.Y., Vorontsov A.A., Buslaev G.V. (2025) Influence of Inhibitors Based on Surfactants on Processes of Hydrates and Asphalt-Resin-Paraffin Deposits Formation: Experimental Studies. International Journal of Engineering Transactions B: Applications. – Vol. 38. – No. 10. – P. 2220-2230. DOI: 10.5829/ije.2025.38.10a.02. (In Eng.).

Kurmangaliyev T.B., Sherov K.T., Sikhimbayev M.R. (2018) Experimental study of optimal parameters of pneumatic motor of vibration table for inertial vibroabrasive machining the parts on the basis of beryllium oxide. News of the National Academy of Sciences of the Republic of Kazakhstan. Series of Geology and Technical Sciences. – No. 5(431). – P. 184-191. DOI: 10.32014/2018.2518-170X.24. (In Eng.).

Malyukova L.S., Martyushev N.V., Tynchenko V.V., Kondratiev V.V., Bukhtoyarov V.V., Konyukhov V.Y., Bashmur K.A., Panfilova T.A., Brigida V. (2023) Circular Mining Wastes Management for Sustainable Production of *Camellia sinensis* (L.) O. Kuntze. Sustainability. – Vol. 15. – No. 15. – P. 11671. DOI: 10.3390/su151511671. (In Eng.).

Toshov J.B., Fozilov D., Yelemessov K.K., Baskanbayeva D.Dh., Bekirova L.R. (2024) Increasing the Durability of Drill Bit Teeth by Changing Its Manufacturing Technology. Obrabotka Metallov. – Vol. 26. – No. 4. – P. 112-124. DOI: 10.17212/1994-6309-2024-26.4-112-124. (In Eng.).

Voitovich E.V., Cherevatova A.V., Zhernovsky I.V., Fisher H.B., Sobolev K. (2014) The Application of Nano-Structured Silica Based Admixture in Gypsum Binders. Materials Research Society Symposium Proceedings. – Vol. 1611. – P. 165-170. DOI: 10.1557/opl.2014.774. (In Eng.).

Zaalishvili V.B., Timchenko V.Yu. (2026) Spatiotemporal Variations of NDVI Index Trends in Mountainous and Foothill Areas of the North Caucasus. Geologiya i Geofizika Yuga Rossii. – Vol. 16. – No. 1. – P. 180-187. DOI: 10.46698/vnc.2026.54.86.001. (In Eng.).