

Под эгидой Правительства Российской Федерации
при поддержке Кабинета Министров Республики Татарстан

Under the auspices of the Government of the Russian Federation
With the support of the Cabinet of Ministers of the Republic of Tatarstan

Международный форум KAZAN DIGITAL WEEK – 2025

Proceedings of the International Forum KAZAN DIGITAL WEEK – 2025

**I том
Volume I**

**Казань 2025
Kazan 2025**

УДК 004.67; 504.064.36

КОМПЛЕКСНАЯ ОЦЕНКА ЗАРАСТАЕМОСТИ ВОДОЁМОВ С ПРИМЕНЕНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Валиев В.С., старший научный сотрудник;

ORCID: 0000-0002-8848-5326;

Зиганшин И.И., к.г.н., старший научный сотрудник;

ORCID: 0000-0003-4449-3444;

Александрова А.Б., к.б.н., старший научный сотрудник,

ORCID: 0009-0002-2303-6156;

Иванов Д.В., д.г.н., заместитель директора,

ORCID: 0000-0003-2822-4623;

Хасанов Р.Р., научный сотрудник Института проблем экологии и недропользования Академии наук РТ, г. Казань, Россия;

ORCID: 0009-0006-9385-7290

COMPREHENSIVE ASSESSMENT OF WATER BODIES OVERGROWTH USING MACHINE LEARNING METHODS

Valiev V.S., Senior Researcher;

ORCID: 0000-0002-8848-5326;

Ziganshin I.I., Candidate of Geographical Sciences, Senior Researcher;

ORCID: 0000-0003-4449-3444;

Aleksandrova A.B., Candidate of Biological Sciences, Senior Researcher;

ORCID: 0009-0002-2303-6156;

Ivanov D.V., Doctor of Geographical Sciences, Deputy Director;

ORCID: 0000-0003-2822-4623;

Khasanov R.R., Researcher, Research Institute for Problems of Ecology and Mineral Wealth Use of Tatarstan Academy of Sciences, Kazan, Russia;

ORCID: 0009-0006-9385-7290

Аннотация

Представлены результаты комплексного исследования зарастаемости озёр и водохранилищ Республики Татарстан, оценка величины которой осуществлялась методами машинного обучения. Отмечена перспективность комплексного подхода, сочетающего гидрохимический мониторинг, анализ состава донных отложений, мультиспектральный анализ спутниковых снимков и современные методы машинного обучения. Метод Gradient Boosting обеспечивает высокую точность прогнозирования зарастаемости, что критически важно для управления водными ресурсами.

Ключевые слова: водоемы, высшая водная растительность, зарастание Республики Татарстан, комплексная оценка, машинное обучение

Abstract

In this article the authors present the results of a comprehensive study of the overgrowth of lakes and reservoirs in the Republic of Tatarstan, the assessment of which was carried out using machine learning methods. The prospects of an integrated approach combining hydrochemical monitoring, bottom sediment analysis, multispectral analysis of satellite images with NDVI calculation and modern machine learning methods are noted. The Gradient Boosting method provides high accuracy in forecasting overgrowth, which is critical for water resources management.

Keywords: water bodies, higher aquatic vegetation, overgrowth Republic of Tatarstan, comprehensive assessment, machine learning

Зарастание водоёмов макрофитами и водорослями представляет собой многофакторный экологический процесс, который при чрезмерной интенсивности приводит к значительному ухудшению качества воды, нарушению баланса водных экосистем, снижению биологического разнообразия и, в результате, к экономическим потерям (снижение рекреационной ценности, затраты на очистку). Современные методы оценки зарастания сочетают морфометрические, гидрохимические и гидробиологические подходы, дополненные дистанционным зондированием и машинным обучением (ML).

Исследования проводились на озерах и водохранилищах Республики Татарстан в 2021–2024 гг. Их актуальность обусловлена усилением процессов зарастания водоемов региона под воздействием климатических изменений, антропогенного эвтрофирования, нарушениями гидрологического режима водоёмов.

Для решения поставленных в работе задач использовали следующие данные: морфометрические параметры водоемов: площадь литоральной зоны, максимальные глубины, степень зарастания; результаты анализа 930 проб воды по показателям: ионы аммония, нитраты, нитриты, фосфаты, БПК₅, ХПК, марганец, железо, pH, жёсткость, содержание взвешенных веществ; результаты анализа 630 проб донных отложений по показателям: гранулометрический состав (доля частиц <0,01 мм), содержание органического вещества, подвижных форм марганца, цинка, железа, меди, валовых фосфора и азота.

Вегетационный индекс NDVI (Normalized Difference Vegetation Index) [1] рассчитывался по данным спутниковых снимков Sentinel-2 [2], что позволило количественно оценить плотность растительности на водной поверхности и выделить зоны активного зарастания водоемов.

Данные разделялись на обучающую, тестовую и валидационную выборки в соотношении 60%, 20%, 20%.

Для анализа использовались четыре алгоритма со следующими базовыми значениями гиперпараметров, которые подбирались методом GridSearch по сетке диапазонов их возможных значений [3, 4]:

- градиентный бустинг (Gradient Boosting); архитектура моделей: ансамбль – 100–200 деревьев, глубина деревьев – 3–5 уровней, скорость обучения – 0,05;
- случайный лес (Random Forest); количество деревьев – 100–500, критерий разделения – MSE;
- метод опорных векторов (SVR); ядро: радиальная базисная функция, параметр регуляризации $C=1,0$;
- многослойный перцептрон (MLP); архитектура – 1–2 скрытых слоя (100 нейронов), функция активации – ReLU, оптимизатор – Adam.

Указанные методы использовались для расчета моделей множественных регрессий, среди которых выбирались наиболее эффективные. Для сравнения эффективности моделей использовались: коэффициент детерминации (R^2), оценивающий долю объяснённой дисперсии; среднеквадратичная ошибка (MSE) и средняя абсолютная ошибка (MAE), оценивающие точность моделей, кросс-валидация, оценивающая устойчивость моделей. Для оценки важности отдельных признаков применялся пермутационный анализ, фиксирующий снижение точности модели в результате перетасовки (пермутации) значений оцениваемого предиктора при неизменном состоянии значений всех остальных предикторов.

Целью машинного обучения являлось вскрытие взаимосвязей между ключевыми предикторами зарастаемости, выявленными нами ранее [5], и построение итоговой модели зарастаемости водоемов. Расчеты регрессионных моделей различными методами ML осуществля-

лись с помощью библиотеки машинного обучения Skikit-Learn в среде Python с использованием ансамбля моделей Sklearn.ensemble, Sklearn.svm и sklearn.metrics.

Первая модель оценивает содержание органического вещества в донных отложениях в зависимости от изменчивости содержания в донных отложениях азота, фосфора, подвижных форм марганца и доли пелитовой фракции размером $<0,01$ мм. Лучшей моделью для прогноза содержания органического вещества является многослойный перцептрон MLP ($R^2=0,77$, MAE=2,73). Наиболее важными предикторами модели оказались валовый азот (пермутационная важность=75,5) и доля пелита (важность=7,4).

Вторая модель вскрывает зависимость доли пелита от изменчивости значений органического вещества, pH, содержания в донных отложениях подвижных форм Cu, Zn, Mn, Fe, валового азота и фосфора. Лучшая модель регрессии получена методом Gradient Boosting ($R^2=0,68$, MAE=8,79). Главными предикторами модели являлись подвижные формы марганца (важность=250,5) и цинка (важность=133,7).

Третья модель описывает взаимосвязь концентраций фосфатов в воде водных объектов с концентрацией NH_4^+ , NO_3^- , NO_2^- , БПК₅, Mn, взвешенных веществ, жесткости и pH. Лучшая модель регрессии получена методом Gradient Boosting ($R^2=0,45$, MAE=0,07). Основным предиктором модели оказались ионы аммония (важность=0,042).

В качестве итоговой модели использовалась взаимосвязь наиболее важных предикторов и факторов зарастаемости: доля зарастаемости, доля литорали, максимальная глубина водоёма, содержание органического вещества, валового фосфора и пелита в донных отложениях, концентрация фосфатов в воде.

Лучшей итоговой регрессионной моделью является модель, рассчитанная методом Gradient Boosting ($R^2=0,98$, MAE=0,013), а наиболее важными предикторами – концентрации фосфатов в воде (важность=1397,4); доля пелита (важность=257,0) и содержание валового фосфора в донных отложениях (важность=34,1). График соотношения предсказанных и фактических значений зарастаемости представлен на рис. 1.

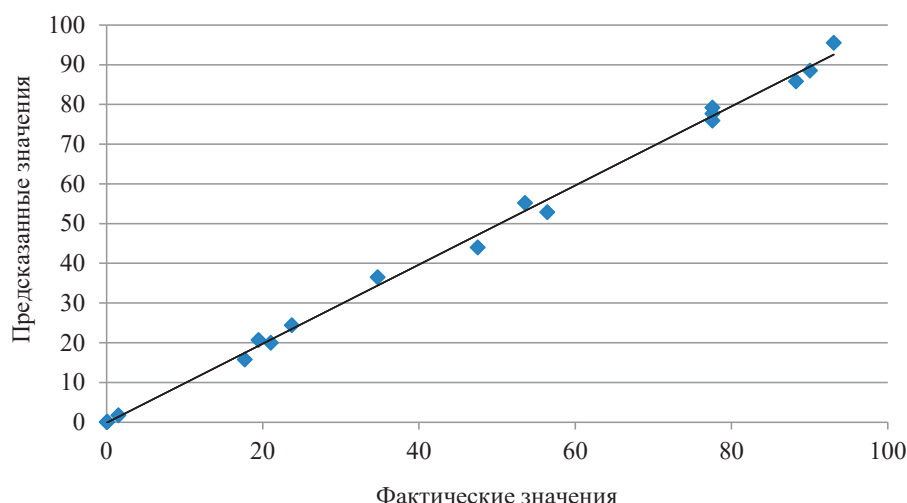


Рис. 1. Качество итоговой регрессионной модели прогноза зарастаемости водоёмов Республики Татарстан

Лучшие параметры Gradient Boosting {'learning_rate': 0,05, 'max_depth': 5, 'n_estimators': 200}: количество деревьев – 200, глубина деревьев – 5, скорость обучения – 0,05, функция потерь – MSE.

Проведенное исследование продемонстрировало высокую эффективность метода Gradient Boosting для построения моделей прогнозирования зарастания водоёмов Республики Татарстан ($R^2=0,98$). Отмечена и статистически подтверждена важная роль высоких кон-

концентраций фосфатов в воде и мелкодисперсных фракций донных отложений в процессе зарастания водоёмов. Ключевым событием интенсификации процессов зарастания является одновременная фиксация повышенных концентраций фосфатов в воде, а также накопление азота, аккумуляция частиц <0,01 мм и рост содержания подвижных форм марганца и цинка в донных отложениях.

Следует отметить перспективность комплексного подхода, сочетающего гидрохимический мониторинг, анализ донных отложений, мультиспектральный анализ спутниковых снимков с расчетом NDVI и современные методы машинного обучения. Комбинация гидрохимических данных и методов машинного обучения (особенно Gradient Boosting) обеспечивает высокую точность прогнозирования зарастаемости, что критически важно для управления водными ресурсами.

Полученные модели могут быть интегрированы в систему экологического мониторинга для прогнозирования и управления процессами зарастания водоёмов.

Список литературы

1. NDVI – Theory and Practice Electronic resource. Available at: <https://gis-lab.info/qa/ndvi.html> (дата обращения: 13.06.2025).
2. Sentinel-2A, 2B Electronic resource. <https://innoter.com/sputniki/sentinel-2a-2b> (дата обращения: 13.06.2025).
3. Тюрин, А. С. Развитие методов градиентного бустинга / А. С. Тюрин // Нано-биотехнологии. Теплоэнергетика. Математическое моделирование / Сборник статей международной научно-практической конференции. – Липецк, 2024. – С. 293–296.
4. Колосов, В. С. Возможности оптимизации гиперпараметров для разработки моделей градиентного бустинга на примере открытого набора данных / В. С. Колосов // Медицина будущего – 2025 / Материалы Всероссийского научного форума с международным участием. – Тюмень, 2025. – С. 149–150.
5. Хасанов, Р. Р. Показатели зарастания озёр Саралинского участка Волжско-Камского заповедника / Р. Р. Хасанов, И. И. Зиганшин, В. С. Валиев А. Б. Александрова // Биосфера. – 2025. – Т. 17, №2. – С. 75–82.

УДК 338.24:004

ЦИФРОВАЯ ИНТЕГРАЦИЯ КАК ИНСТРУМЕНТ УСТОЙЧИВОГО РАЗВИТИЯ

Вишнякова О.Н., д.э.н., и.о. заведующего кафедрой экономики и управления в спорте ФГБОУ ВО «Поволжский государственный университет физической культуры, спорта и туризма», г. Казань, Россия;
ORCID: 0000-0003-3069-2178;
E-mail: olga_911@mail.ru

DIGITAL INTEGRATION AS A TOOL FOR SUSTAINABLE DEVELOPMENT

Vishnyakova O.N., Doctor of Economics, Acting head of the department of economics and management in sports, Volga State University of Physical Culture, Sports and Tourism, Kazan, Russia;
ORCID: 0000-0003-3069-2178;
E-mail: olga_911@mail.ru

Аннотация

Развитие экономической интеграции и автоматизация процессов управления осуществляются поэтапно на основе принципов устойчивого развития и в рамках Государственной