

## ОЦЕНКА ПОДВЕРЖЕННОСТИ ТЕРРИТОРИИ ЦЕНТРАЛЬНОГО ЧЕРНОЗЕМЬЯ К РАЗВИТИЮ ОВРАЖНОЙ ЭРОЗИИ С ПРИМЕНЕНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

© СС ВУ Корецкий Н. А.<sup>1</sup>, Горбунов А. С.<sup>2</sup>, Бевз В. Н.<sup>3</sup>

<sup>1</sup> Воронежский государственный университет  
394018, г. Воронеж, Университетская площадь, д. 1. Российская Федерация  
e-mail: nkoretskiy98@bk.ru; ORCID: 0009-0003-6451-4350

<sup>2</sup> Воронежский государственный университет  
394018, г. Воронеж, Университетская площадь, д. 1. Российская Федерация  
e-mail: gorbunov.ol@mail.ru; ORCID: 0000-0002-2132-3453

<sup>3</sup> Воронежский государственный университет  
394018, г. Воронеж, Университетская площадь, д. 1. Российская Федерация  
e-mail: snark61@mail.ru; ORCID: 0000-0002-6024-1207

Поступила в редакцию 31.01.2025

После доработки 30.05.2025

Принята к публикации 18.06.2025

### *Аннотация*

**Цель.** Продемонстрировать новый подход для оценки предрасположенности территории к развитию овражной эрозии на основе алгоритмов машинного обучения.

**Процедура и методы.** Ключевыми исследовательскими методами стали: моделирование на основе машинного обучения (с использованием метода сглаженной многоуровневой оценки) и геоинформационное моделирование. Для анализа существующих подходов к оценке предрасположенности территории к развитию овражной эрозии использован анализ литературных источников. Непосредственно для проведения самой оценки авторы применили один из современных методов машинного обучения — *CatBoost*, который предусматривает создание ансамбля из нескольких моделей машинного обучения (ансамблирование моделей).

**Результаты.** Предложен новый метод — сглаженной многоуровневой оценки предрасположенности территории к развитию овражной эрозии на примере Воробьёвского района Воронежской области. Рассмотрены сходства и различия предлагаемого подхода с уже существующими методами, базирующимися на идее ансамблирования. Обоснованы 2 новые метрики для оценки точности работы предлагаемого метода — RF1 и NDF. Введены понятия мягкого, жёсткого и взвешенного уровней моделирования, позволившие оценить вклад морфометрии рельефа в развитие овражной эрозии. Установлено, что наибольшее влияние на этот процесс оказывают абсолютные и относительные высоты местности, LS-фактор, водосборная площадь и экспозиция склона. Вместе они объясняют 95% площадей существующей овражной эрозии в регионе. По итогам моделирования на территории Воробьёвского района Воронежской области было выявлено 2 853 га земель с высокой и очень высокой предрасположенностью к развитию овражной эрозии. В ландшафтном отношении им соответствуют остепненные долинно-балочные склоны южных экспозиций, отличающиеся наибольшей протяжённостью, крутизной и высотой, имеющие вогнутый поперечный профиль и значительные водосборные площади.

**Теоретическая и/или практическая значимость.** Предложен новый подход на основе методов машинного обучения по оценке территории на предмет её предрасположенности к развитию овражной эрозии. С практической точки зрения, предлагаемые в работе идеи и сам методический подход могут быть использованы для получения более надежного результата по сравнению с рядом классических методов и технологий машинного обучения, что особенно актуально при анализе крупных регионов с высокой неравномерностью распределения овражных форм рельефа.

**Ключевые слова:** CatBoost, ансамблевый подход, геоинформационные системы, эрозионный потенциал, Черноземье, Воронежская область

**Благодарности.** Исследование выполнено за счёт гранта Российского научного фонда № 24-27-20122.

**Для цитирования:**

Корецкий Н. А., Горбунов А. С., Бевз В. Н. Оценка подверженности территории Центрального Черноземья к развитию ображной эрозии с применением методов машинного обучения // Географическая среда и живые системы. 2025. № 2. С. 74–91. DOI: 10.18384/2712-7621-2025-2-74-91

Original Article

## ASSESSING THE VULNERABILITY OF THE CENTRAL BLACK EARTH REGION TO GULLY EROSION USING MACHINE LEARNING METHODS

N. Koretskiy<sup>1</sup>, A. Gorbunov<sup>2</sup>, V. Bevz<sup>3</sup>

<sup>1</sup> Voronezh State University

Universitetskaya pl. 1, Voronezh 394018, Russian Federation  
e-mail: nkoretskiy98@bk.ru; ORCID: 0009-0003-6451-4350

<sup>2</sup> Voronezh State University

Universitetskaya pl. 1, Voronezh 394018, Russian Federation  
e-mail: gorbunov.ol@mail.ru; ORCID: 0000-0002-2132-3453

<sup>3</sup> Voronezh State University

Universitetskaya pl. 1, Voronezh 394018, Russian Federation  
e-mail: snark61@mail.ru; ORCID: 0000-0002-6024-1207

Received 31.01.2025

Revised 30.05.2025

Accepted 18.06.2025

### Abstract

**Aim.** Demonstrate a new approach for assessing the susceptibility of an area to gully erosion based on machine learning algorithms.

**Methodology.** The key research methods were: machine learning-based modeling and geo-information modeling. A literary method was used in the analysis of existing solutions to assess the susceptibility of the territory to gully erosion. In the practical field of research, one of the most modern methods of machine learning, CatBoost, is widely used, taken as the basis of the developed approach. One of the main ideas of the proposed approach is the ensembling of machine learning models.

**Results.** A new method is proposed – a smoothed multilevel assessment of a territory's predisposition to gully erosion, using the Vorobyevsky District of the Voronezh Region as an example. The similarities and differences between the proposed approach and existing methods based on the idea of ensemble modeling are considered. Two new metrics for assessing the accuracy of the proposed method, RF1 and NDF, are justified. The concepts of soft, hard, and weighted modeling levels are introduced, allowing the contribution of relief morphometry to gully erosion development to be assessed. It has been established that the greatest influence on this process is exerted by the absolute and relative heights of the terrain, the LS factor, the catchment area, and the slope exposure. Together, they explain 95% of the areas of existing gully erosion in the region. Based on the results of modeling in the Vorobyevsky district of the Voronezh region, 2,853 hectares of land with a high and very high predisposition to gully erosion were identified. In terms of landscape, these correspond to the steppe-like valley slopes of southern exposures, which are distinguished by their greatest length, steepness, and height, and have a concave cross-section and significant catchment areas.

**Research implications.** The significance of the study lies in the proposal of a new approach to assessing the predisposition of the territory to gully erosion based on machine learning methods. From a practical point of view, the ideas proposed in the work and the method itself can be used to obtain a higher-quality result in assessing the predisposition of a territory to gully erosion, compared with a number of classical methods and machine learning technologies, especially when analyzing large territories with high uneven distribution of gully erosion.

**Keywords:** CatBoost, ensemble approach, geoinformation systems, erosion potential, Chernozemye, Voronezh Region

**Acknowledgments.** This research was supported by RSF grant no. 24-27-20122.

**For citation:**

Koretsky N. A., Gorbunov A. S., Bevz V. N. (2025). Assessing the vulnerability of the Central Black Earth Region to gully erosion using machine learning methods. In: *Geographical Environment and Living Systems*, 2, 74–91. DOI: 10.18384/2712-7621-2025-2-74-91

## ВВЕДЕНИЕ

Моделирование овражной эрозии (ОЭ) с использованием алгоритмов машинного обучения (МО) в последние годы стало перспективным направлением в исследовании деградации земель [12; 19].

Интеграция алгоритмов МО, таких как метод опорных векторов (*SVM*), случайный лес (*RF*), экстремальный градиентный бустинг (*XGBoost*) и искусственные нейронные сети на основе многослойного перцептрона (*ANN MLP*), усовершенствовала предсказание предрасположенности территории к развитию ОЭ. С помощью таких моделей устанавливаются нелинейные взаимосвязи между параметрами окружающей среды и самой ОЭ с высокой предсказательной точностью, что даёт возможность выработать стратегию по управлению земельными ресурсами, направленную на снижение рисков развития эрозии [11; 14]. Оценка ключевых факторов, влияющих на развитие эрозии, приводит к более надёжному выявлению территорий, предрасположенных к появлению промоин и оврагов. В зарубежной литературе это направление получило название *Gully Erosion Susceptibility (GES)* [15].

Использование моделей МО позволяет оценить риск развития ОЭ посредством выявления зон, которые потенциально могут быть подвержены ОЭ. На основе таких данных может и должна корректироваться политика землепользования с

целью обеспечения устойчивого развития территории [15; 16].

Несмотря на существующие достижения в применении МО для оценки *GES*, в мировом научном сообществе до сих пор отсутствует единое мнение по поводу эффективности его применения. Это связано с изменчивостью производительности моделей МО, недостатком доступных данных для обучения, проблемами применения подхода для разных типов природной среды [14; 16].

Работы отечественных учёных-эрозиологов в большей степени направлены на изучение закономерностей пространственной организации ОЭ [2; 3] и на математическое моделирование её развития на локальном уровне [7; 18]. При этом мало внимания уделяется возможности вероятностного моделирования ОЭ методами машинного и глубокого обучения.

Настоящая работа нацелена на демонстрацию нового подхода для оценки *GES* — сглаженной многоуровневой оценки предрасположенности территории к развитию ОЭ. Данный подход отличается от существующих тем, что он позволяет получать в большей степени достоверный результат (на основе оценки метода Монте-Карло) за счёт использования ансамблевого подхода с обучением на усечённой сбалансированной выборке.

В рамках данной работы построение модели для оценки предрасположенности территории к ОЭ осуществлялось на осно-

ве комплекса морфометрических величин: крутизны, экспозиция, кривизны, вертикальной расчленённости территории и др.

В качестве основных материалов для проведения исследования выступили: контуры овражных геосистем, полученные на основе дешифрирования ДДЗ (Яндекс. Спутник) [6] и цифровая модель местности (ЦММ) FABDEM v1-2 [8].

База данных включает овражные формы рельефа, длиной более 70 м [1], отнесённые к 3 типам: склоновым, приводораздельным и донным. В рамках данного исследования донные овраги не рассматривались, поскольку их генезис и морфометрические особенности существенно отличаются от водораздельных и склоновых.

Обучение, оптимизация моделей МО и математические расчёты производились на языке программирования *Python* с подключёнными библиотеками: *pandas*, *numpy*, *sklearn*, *catboost*, *optuna*, *imblearn*. Дополнительно использовались библиотеки *rasterio* для работы с растровыми изображениями и *matplotlib*, *seaborn* для построения графиков.

В качестве базовой ГИС использовался *QGIS 3.34.3*. Предварительная обработка и подготовка ЦММ *FABDEM v1-2* производилась в *SAGA GIS 9.5.1*, там же выполнялось построение морфометрических матриц-признаков.

Для оценки точности и производительности ансамбля моделей используются специально разработанные метрики – *RFI* и *NDF*, учитывающие особенности предлагаемого подхода. Повышенную достоверность результатов призван обеспечивать трёхуровневый подход к её оценке.

Территорией исследования был выбран Воробьёвский район, расположенный на востоке Воронежской области, в подзоне южной лесостепи (рис. 1). В физико-географическом отношении территория относится к северу Калачской возвышенности, представляющей собой сильно расчленённую овражно-балочную сеть пластово-денудационную равнину, сложенную мелко-мергельными породами, глинистыми отложениями палеогена и суглинками четвертичного периода. Район характеризуется густым ( $1,8 \text{ км/км}^2$ ) и глубоким (более 100 м) эрозионным расчленением,

находится в зоне влияния северо-восточного крыла активного Журавкинского неотектонического поднятия.



**Рис. 1 / Fig. 1.** Географическое положение Воробьёвского района Воронежской области / Geographical location of the Vorobyovskiy district of the Voronezh region

*Источник:* составлено авторами

Речная сеть развита слабо (густота составляет  $0,07 \text{ км/км}^2$ ), крупнейшими реками являются р. Толучеевка и р. Подгорная.

Структура ландшафтных местностей складывается из плакорного, склонового, надпойменно-террасового и пойменного типов. На междуречьях преобладают возвышенные волнистые суглинистые лесополево-степные плакоры с чернозёмами обыкновенными. По склонам долин и балок наиболее распространены сильно-наклонные суглинисто-меловые и песчано-суглинистые лесостепные поверхно-

сти со смытыми черноземными почвами и выходами коренных пород. Левобережье рек Подгорной и Толучеевки занято песчаными слабонаклонными полевыми и селитебными надпойменными террасами с песчаными черноземовидными почвами. В поймах рек преобладают пониженные плоские преимущественно луговые поверхности с аллювиальными слоисто-зернистыми почвами.

Овраги являются характерными формами рельефа района. Проведённое ранее исследование позволило выделить на территории 2 890 оврагов, общей площадью 2,8 тыс. га. Средняя густота овражной сети составляет 0,84 км/км<sup>2</sup>, плотность оврагов — 4,86 ед./км<sup>2</sup> [4; 5].

Развитию овражной сети в районе способствуют:

1. значительное вертикальное расчленение (более 100 м);
2. наличие в современной коре выветривания легкоразмываемых лессовидных суглинков;
3. активное поднятие территории (до 4 мм в год);
4. ливневый характер осадков в летнее время и интенсивное снеготаяние весной;
5. высокая распаханность (64% территории) и низкая сохранность естественных ландшафтов.

## ОЦЕНКА ПРЕДРАСПОЛОЖЕННОСТИ ТЕРРИТОРИИ К ОВРАЖНОЙ ЭРОЗИИ

Первым этапом работ являлась подготовка географической базы данных (ГБД). Общая структура её представляет собой матрицу объектов-признаков, где строки — объекты (операционно-территориальные единицы 30×30 м (ОТЕ)), которые размечены как «овражные» и «не овражные» ячейки, а столбцы — признаки (факторы). Количество ОТЕ для проведения исследования — 1,37 млн, среди которых 98,7% относятся к «не овражной» категории (все ячейки раstra, в которых нет оврага размечены как «не овраг», внутренняя структура их не рассматривалась, а 1,3% маркированы как овраги (рис. 2). Таким образом, модель анализирует морфометрию релье-

фа в каждой ячейке раstra, «видит», в каких из них есть овраги, и «ищет» близкие условия в «не овражных» ячейках, тем самым предполагает, что здесь овраг может появиться.

В качестве факторов были использованы следующие морфометрические величины: абсолютная высота, крутизна, экспозиция, топографический индекс нагрева поверхности, длина линии стока (тальвегов), водосборная площадь (площадь замкнутой фигуры, сформированной отрезком горизонтали, включающей данную точку топографической поверхности с 2 линиями тока, приходящими с вышележащей точки на концы отрезка<sup>1</sup>), превышение над базисом эрозии, LS-фактор (показатель развития плоскостного смыва), топографический индекс влажности, Гауссова горизонтальная, вертикальная, избыточная, разностная и аккумуляционная кривизны [13; 17].

В качестве объектов рассматривались ячейки раstra (30×30 м), в которых присутствует или отсутствует овражная эрозия (рис. 2).

Второй этап моделирования заключается в выборе алгоритма построения модели МО и обучения этой модели на сформированной ГБД. Задача МО в контексте оценки предрасположенности территории к ОЭ — показать вероятность, что в конкретной ячейке раstra возможно появление овражной эрозии. Это означает, что обученная модель МО на основе анализа факторов (признаков) присваивает ОТЕ значение вероятности отнесения её к классу овражной. Эта вероятность всегда имеет вид  $y \in [0, 1]$ , где 0 — нулевая вероятность принадлежности ОТЕ к оврагу, а 1 — максимальная вероятность. В качестве порога отнесения ячейки к «овражной» использовано значение 0,5, такую вероятность часто используют в подобного рода исследованиях, т. е. имеющие значение вероятности в диапазоне  $y \in [0, 0,5)$  маркируются как «не овражные территории», тогда как ОТЕ со значениями вероятности в диапазоне  $y \in [0,5, 1]$  — являются «овражными» или «условно овражными» и представляют наибольший интерес в рамках дальнейше-

<sup>1</sup> Рассчитывается для каждой точки ЦМР.



Рис. 2 / Fig. 2. Пример размеченных данных для обучения / Example of labeled training data

Источник: составлено авторами

го исследования. Вероятностные значения «условно овражных» ОТЕ интерпретируются как степень предрасположенности ОТЕ к ОЭ. Чем больше значение вероятностного вывода модели МО, тем выше предрасположенность ОТЕ к ОЭ.

В качестве основного алгоритма МО был использован алгоритм *CatBoost* – это ансамблевый алгоритм, основанный на идее градиентного бустинга, разработанный в 2017 г. [10]. Его особенность состоит в том, что машинный алгоритм последовательно строит простые модели (деревья решений) для выполнения задач классификации. В нашем случае – это отнесения или не отнесения ОТЕ к «овражной», при этом каждая следующая модель исправляет ошибки предыдущей. Бустинг обладает более высокой точностью по сравнению с технологией *Random Forest*, является одним из самых эффективных инструментов анализа табличных данных [9].

Реализация подхода сглаженной многоуровневой оценки предрасположенности территории к ОЭ подразумевает проведение трёхуровневого моделирования.

*Первый (мягкий) уровень* – глобальная оценка чувствительности территории к

ОЭ. На мягком уровне происходит оценка предрасположенности территории к ОЭ в пределах всей территории исследования (вероятность оценивается для всех ячеек раstra). Для этих целей используется ансамбль из базовых моделей МО.

В настоящей работе под ансамблем понимается группа однородных моделей МО (или базовых моделей) с различной архитектурой (скоростью обучения, количеством и сложностью деревьев и др.), обученных с помощью предлагаемой технологии по индивидуальной усечённой сбалансированной выборке.

Усечённая выборка представляет собой набор из 50% ячеек раstra с оврагами, которому случайным образом назначаются 50% ячеек раstra без оврагов. Выбор равного количества целевых и нецелевых ОТЕ называется *балансировкой*. Балансировка используется в том случае если изначальное количество целевых и нецелевых ОТЕ сильно отличаются (в нашем случае количество «овражных» ячеек в 75 раз меньше, чем «не овражных»). Отсутствие балансировки классов может привести к тому, что формируемая обучающая выборка окажется нерепрезентативной, либо «овражные»

ОТЕ могут не попасть в обучающий набор данных и модель будет работать некорректно. Балансировка данных является необходимой процедурой для улучшения качества классификации.

Данный подход имеет некоторое сходство с алгоритмом *bagging* (*Bootstrap Aggregating*), который также осуществляет обучение модели на нескольких случайным образом отобранных выборках, с последующим усреднением полученных результатов. Разница состоит в том, что алгоритм *bagging* создает неравные выборки, пропорционально изначальному количеству ОТЕ разных классов, в то время как в предложенном подходе количество ОТЕ в выборках выравнивается. Эта процедура называется случайным понижением до меньшего класса или *Random Undersampling*. Сравним предлагаемый подход с алгоритмом *bagging* (рис. 3).

Рассмотрим примеры результатов моделирования предрасположенности территории к развитию овражной эрозии традиционным методом (одиночная модель) и предлагаемым методом (с использованием ансамбля из 10 моделей *CatBoost*) (рис. 4).

На этапе обучения каждый член ансамбля проходил индивидуальную гиперпараметрическую оптимизацию на основе *Tree-structured Parzen Estimator (TPE)* – метода, используемого для поиска оптимальных настроек модели МО (количества деревьев, скорости обучения, регуляризации и др.), при которых ошибка предсказания

(«овражная» ячейка или нет) была бы минимальной. Ошибка предсказания в МО называется *функцией потерь* (*Logloss*), а алгоритм оптимизации, который используется для уменьшения функции потерь получил название *градиентного спуска*.

В нашем случае использовалась логарифмическая функция потерь, позволяющая в рамках задачи классификации «овражных» и «не овражных» ОТЕ оценить не только точность, но и уверенность модели в принятых решениях. Оценка логарифмических потерь проходила кросс-валидацию (перекрестную проверку по 3 независимым выборкам). Её задача определить, насколько хорошо работает модель на новых данных.

Для оценки общей точности работы ансамбля использовалась специально разработанная метрика – *RF1*, объединяющая известные метрики *precision* (точность выделения целевых объектов (оврагов)) и *recall* (полноту выделения целевых объектов) (формулы 1–3).

Введение *RF1* связано со следующими соображениями. Задача оценки предрасположенности территории к ОЭ отличается от стандартной задачи МО сложностью оценки качества полученной модели. Например, при использовании метрики *recall* (полнота выделения оврагов) идеальный сценарий работы модели – это 100% верно классифицированных существующих в действительности овражных ОТЕ. Однако такая же точность модели может быть в

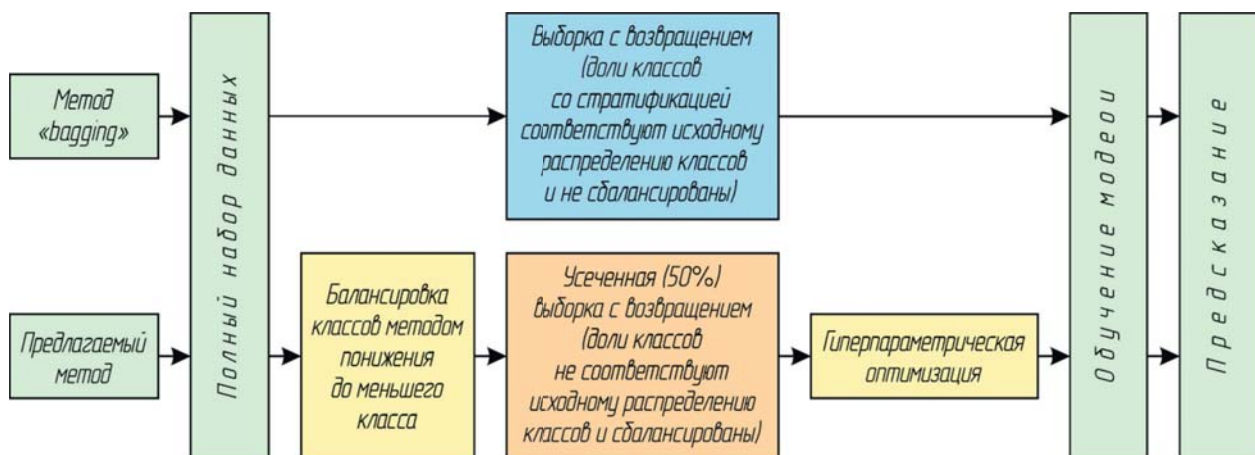
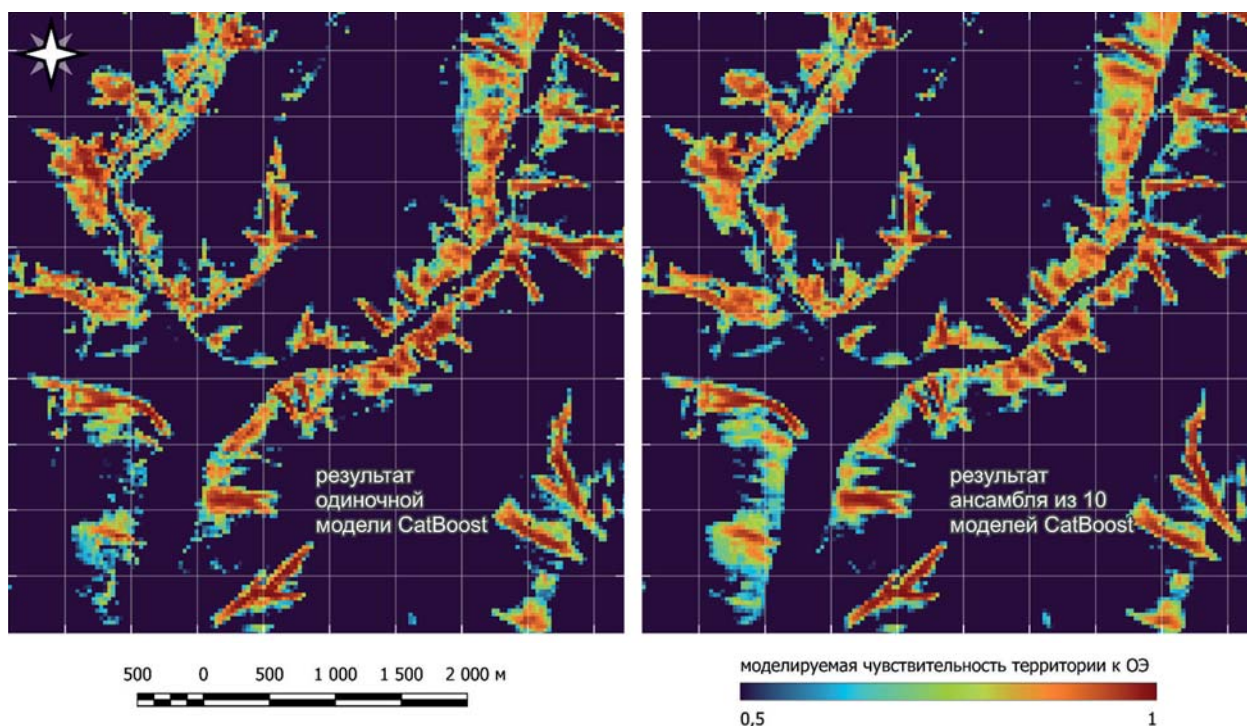


Рис. 3 / Fig. 3. Сравнение машинного обучения методом «bagging» и предлагаемым методом / Comparison of bagging machine learning and the proposed method

Источник: составлено авторами



**Рис. 4 / Fig. 4.** Сравнение результатов моделирования predisposition территории к овражной эрозии традиционным (слева) и предлагаемым (справа) методом / Compression of the results of modeling the predisposition of the territory to gully erosion using the traditional (left) and proposed (right) methods

Источник: составлено авторами

сценарии, когда модель МО классифицирует все ОТЕ исследуемой территории, как чувствительные к ОЭ – это не противоречит сущности метрики *recall*.

Использование метрики *f1-score* (*f1*-мера) на результатах работы всего ансамбля моделей демонстрирует критически низкое качество работы ансамбля даже в том случае, когда общая доля верно распознанных ОТЕ существующей эрозии порядка 95%. Это объясняется тем, что метрика *f1-score* представляет из себя гармоническое среднее между метриками *recall* и *precision*. Множественные ложноположительные срабатывания ансамбля, к которым чувствительна метрика *precision*, такие как прогнозирование наличия оврагов, в условиях где их нет и они сформироваться не могут (плоские водоразделы, центральная пойма и др.), вызывают резкое ухудшение *f1-score*. В то же время ложноположительные срабатывания в контексте решаемой задачи (прогнозирование наличия оврагов, в условиях где их нет, но

они с высокой вероятностью могут появиться: крутые долинно-балочные склоны, ложбинообразные понижения на склонах и др.) оцениваются как качественный результат. Результат, при котором часть существующих «не овражных» ОТЕ перепределяется моделью в «овражные» ОТЕ с целью оценки степени устойчивости к ОЭ – естественный процесс. Предлагаемая метрика *RFI* позволяет одновременно учесть вышеназванные показатели точности, минимизировав возможные ошибки. Общая технология расчёта метрики точности *RFI* приведена ниже.

$$precision_i = \frac{TP_i}{TP_i + FP_i} \quad (1)$$

$$recall_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

где:

*precision<sub>i</sub>* – точность выделения оврагов *i*-й моделью из ансамбля,



$recall_i$  – полнота выделения оврагов  $i$ -й моделью из ансамбля;

$TP_i$  – овражные ОТЕ, классифицированные как овраги  $i$ -й моделью из ансамбля;

$FP_i$  – не овражные ОТЕ, классифицированные как овраги  $i$ -й моделью из ансамбля;

$FN_i$  – овражные ОТЕ, классифицированные как не овражные участки  $i$ -й моделью ансамбля.

$$recall_{ens} = \frac{TP_{ens}}{TP_{ens} + FN_{ens}} \quad (3)$$

где:

$recall_{ens}$  – полнота выделения оврагов всем ансамблем моделей  $TP_{ens}$  – овражные ОТЕ, классифицированные как овраги ансамблем моделей;

$FN_{ens}$  – овражные ОТЕ, классифицированные как не овражные ансамблем моделей.

Далее рассчитывается  $f1_{avg}$  – гармонические средние значения между полнотой и точностью для каждой  $i$ -й модели из ансамбля, усреднённые по количеству моделей в ансамбле (формула 4):

$$f1_{avg} = \frac{1}{N} \sum_{i=1}^N \frac{2 * recall_i * precision_i}{recall_i + precision_i} \quad (4)$$

Непосредственно на основе  $recall_{ens}$  и  $f1_{avg}$  выполняется расчёт метрики  $RF1$  по формуле 5:

$$RF1 = \frac{2 * recall_{ens} * f1_{avg}}{recall_{ens} + f1_{avg}} \quad (5)$$

В качестве дополнительной оценки результаты работы ансамбля моделей МО можно использовать представленную метрику  $NDF$  – нормализованную разность между  $FP$  (ложноположительными срабатываниями) и  $FN$  (ложноотрицательными срабатываниями) (формула 6). Метрика  $NDF$  имеет ограниченную информацию о качестве работы ансамбля в целом и в особенности отдельных его моделей, но позволяет оценить соотношение ложноположительных срабатываний ( $FP$ ) и ложноотрицательных срабатываний ( $FN$ ).  $FP$  – это ОТЕ, в пределах которых на момент исследования ОЭ отсутствует, однако модель оценила предрасположенность этих

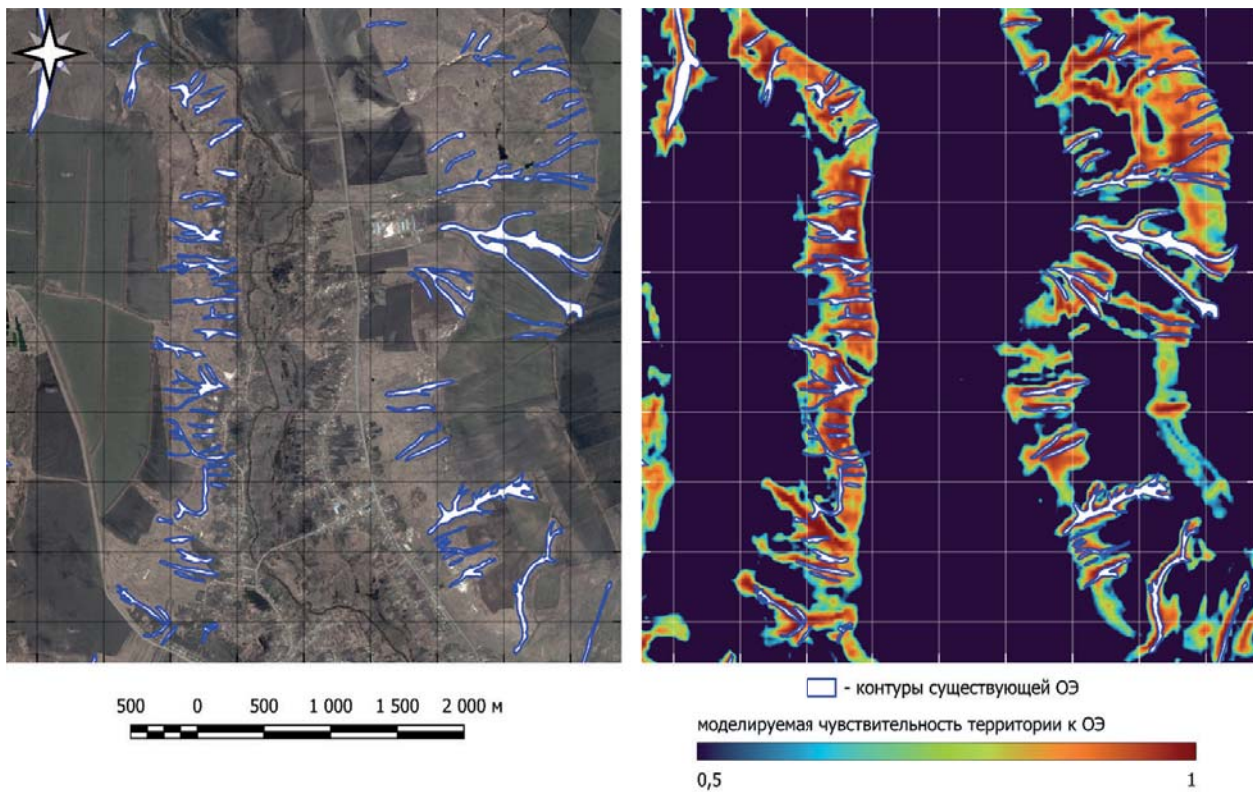
ОТЕ как большую либо равную 0,5.  $FN$  – это ОТЕ, в пределах которых на момент исследования ОЭ присутствует, однако модель оценила предрасположенность этих ОТЕ как меньшую, чем 0,5. Метрика  $NDF$  может быть полезна для оценки того, насколько сильно ОТЕ с оцененной повышенной предрасположенностью к ОЭ преобладают над ОТЕ с существующей ОЭ, в описании которых ансамбль дал ошибочное срабатывание.

$$NDF = 0,5 * \left( 1 + \frac{FP_{ens} - FN_{ens}}{FP_{ens} + FN_{ens}} \right) \quad (6)$$

Как показала практика, применение ансамблевого подхода, помимо уменьшения числа артефактных значений, плодотворно сказалось на общей точности полученного ансамбля моделей *CatBoost* мягкого уровня. Точность по метрике  $RF1$  составила 0,926, по метрике  $NDF$  0,91 по данным трёхкратной перекрестной проверки (табл. 1).

После завершения процесса обучения базовых моделей и формирования ансамбля через каждую модель ансамбля пропускаться полный набор данных (ОТЕ). В результате была создана модель общей предрасположенности территории к ОЭ, где 0 – территория не предрасположена к развитию ОЭ, а 1 – наивысшая предрасположенность территории к ОЭ (рис. 5).

Моделирование на I уровне позволяет оценить общую предрасположенность территории к ОЭ и выявить глобальные факторы, которые влияют на развитие эрозии. Оценка важности факторов на этом уровне моделирования позволила установить, что основной вклад в формирование оврагов вносят высота местности,  $LS$ -фактор и Гауссова кривизна поверхности. Изменения этих морфометрических параметров в пределах территории исследования в значительной мере коррелируют с ареалом распространения существующих оврагов. В физико-географическом отношении эти территории соответствуют ландшафтам долинно-балочных склонов и уступов надпойменных террас, имеющих значительную протяжённость, крутизну и вогнутый поперечный профиль.



**Рис. 5 / Fig. 5.** Пример результата работы модели общего уровня / Example of the result of the general level model

*Источник:* составлено авторами

*Таблица 1 / Table 1*

**Сводная точность результатов моделирования первого уровня на основе полного набора данных / Summary accuracy of first level simulation results (all data)**

Метрика	Точность единичной модели из ансамбля										Общая точность на основе всего набора данных	Оценщик общей точности
	1	2	3	4	5	6	7	8	9	10		
Recall	0.933	0.925	0.936	0.930	0.935	0.933	0.931	0.932	0.927	0.921	<b>0.957</b>	Ансамбль
F1	0.900	0.898	0.895	0.901	0.895	0.903	0.900	0.897	0.895	0.893	0.897	Среднее по моделям
NDF	0.321	0.304	0.316	0.301	0.305	0.293	0.324	0.367	0.304	0.332	<b>0.917</b>	Ансамбль
RF1	-	-	-	-	-	-	-	-	-	-	<b>0.926</b>	Ансамбль

*Источник:* составлено авторами

Для уточнения результата моделирования I уровня предлагается дополнительно использовать моделирование второго II уровня.

*Второй (жесткий) уровень* – внутренняя оценка предрасположенности территории к развитию ОЭ. На этом уровне происходит оценка предрасположенности территории к развитию ОЭ в пределах тех ОТЕ, которые получили оценку предрасположенности выше 0,5 по результатам моделирования I уровня. В остальном технологический процесс моделирования II уровня аналогичен процессу моделирования I уровня (рис. 6).

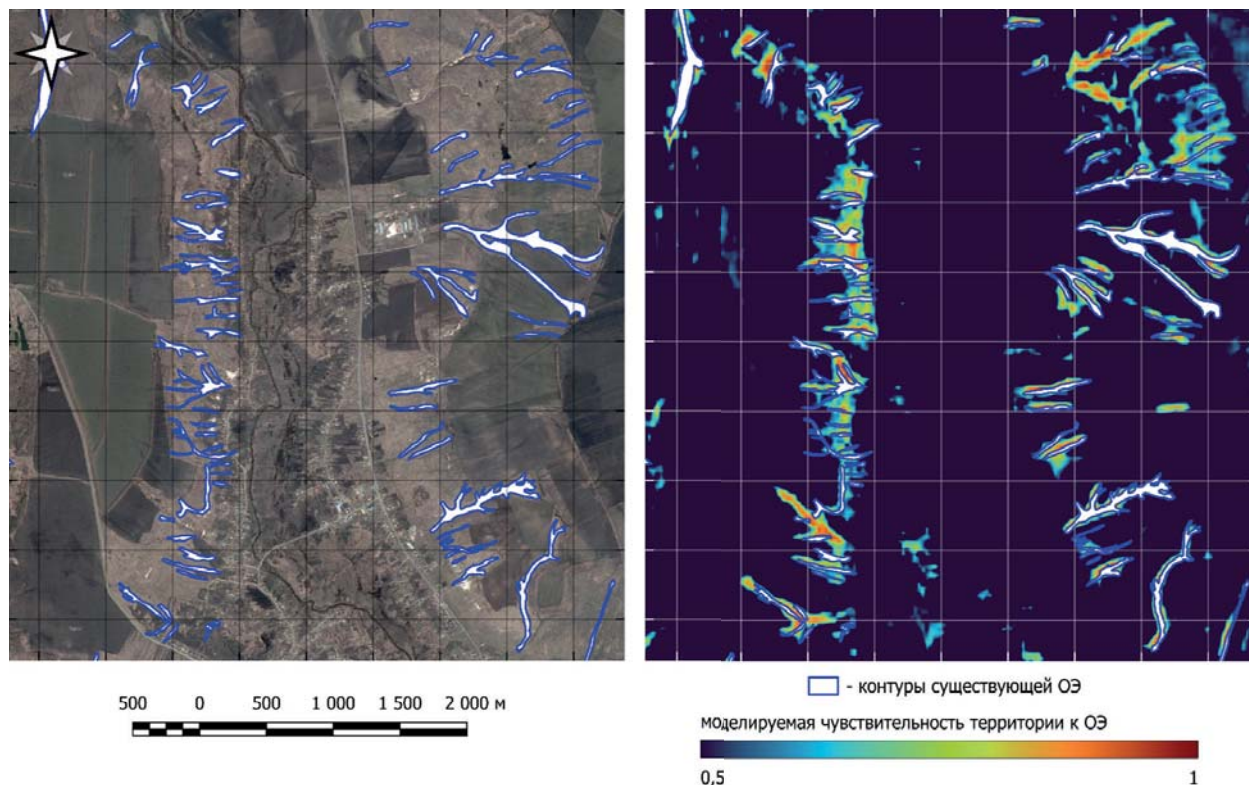
Практическая значимость моделирования жесткого уровня заключается в том, что оно позволяет минимизировать выраженное влияние на развитие ОЭ ведущих глобальных факторов, выявленных на I уровне моделирования (высота местности, *LS*-фактор и Гауссова кривизна), а также найти иные показатели, которые вносят свой вклад в формирование оврагов. В рамках текущего исследования было установлено, что на жестком уровне мо-

делирования наибольшее влияние на формирование ОИ оказывают превышение над местным базисом эрозии (перепад высот), водосборная площадь и экспозиция.

Таким образом, модель провела внутреннюю дифференциацию выделенных на I уровне склоновых ландшафтов, оценив, что наибольшей предрасположенностью к развитию овражной эрозии обладают склоны с максимальным перепадом высот, наибольшими водосборными площадями и экспозициями южных румбов, на которых происходит интенсивное снеготаяние весной и активизируются экзогенные процессы.

Также следует отметить, что результаты II уровня моделирования, как правило, являются менее точными по сравнению с моделированием I уровня (табл. 1–2) и должны использоваться не в качестве самостоятельного результата, а в роли дополняющего (табл. 2).

*Третий (взвешенный) уровень* – подразумевает балансировку результатов мягкого моделирования на основе результата жесткого моделирования.



**Рис. 6 / Fig. 6.** Пример результата работы модели частного уровня / Example of the result of the private level model

*Источник:* составлено авторами

Таблица 2 / Table 2

Сводная точность результатов моделирования II уровня на основе полного набора данных /  
Summary accuracy of second level simulation results (all data)

Метрика	Точность единичной модели из ансамбля										Общая точность на основе всего набора данных	Оценщик общей точности
	1	2	3	4	5	6	7	8	9	10		
Recall	0.677	0.673	0.662	0.671	0.647	0.665	0.686	0.676	0.661	0.648	0.727	Ансамбль
F1	0.713	0.713	0.707	0.708	0.694	0.697	0.716	0.709	0.709	0.705	0,705	Среднее по моделям
NDF	0.593	0.602	0.613	0.595	0.617	0.580	0.576	0.583	0.615	0.601	0.772	Ансамбль
RF1	-	-	-	-	-	-	-	-	-	-	0.719	Ансамбль

Источник: составлено авторами

На этом этапе не происходит обучение нового ансамбля, все расчёты выполняются на основе двух ранее полученных результатов (I и II уровней) (формула 7). Для этого используются точности предыдущих уровней моделирования по метрике *RF1*.

$$W = \max \left( 0.5 \left( \frac{\alpha_1^2}{\alpha_1^2 + \alpha_2^2} * y_1 + \frac{\alpha_2^2}{\alpha_1^2 + \alpha_2^2} * y_2 \right) \right) \quad 7)$$

где:

$y_1, y_2$  – результаты оценки предрасположенности территории к ОЭ по результатам I и II уровней моделирования соответственно;

$\alpha_1, \alpha_2$  – точность ансамблей по метрике *RF1* I и II уровней моделирования соответственно;

0,5 – поправочный коэффициент, исключающий возможность появления оценок ниже 50%.

Полученный алгоритм расчёта значений взвешенного уровня можно воспринимать как частный случай реализации фильтра Калмана, который позволяет при объединении данных сгладить итоговые результаты с помощью установления весовых коэффициентов для разных показателей.

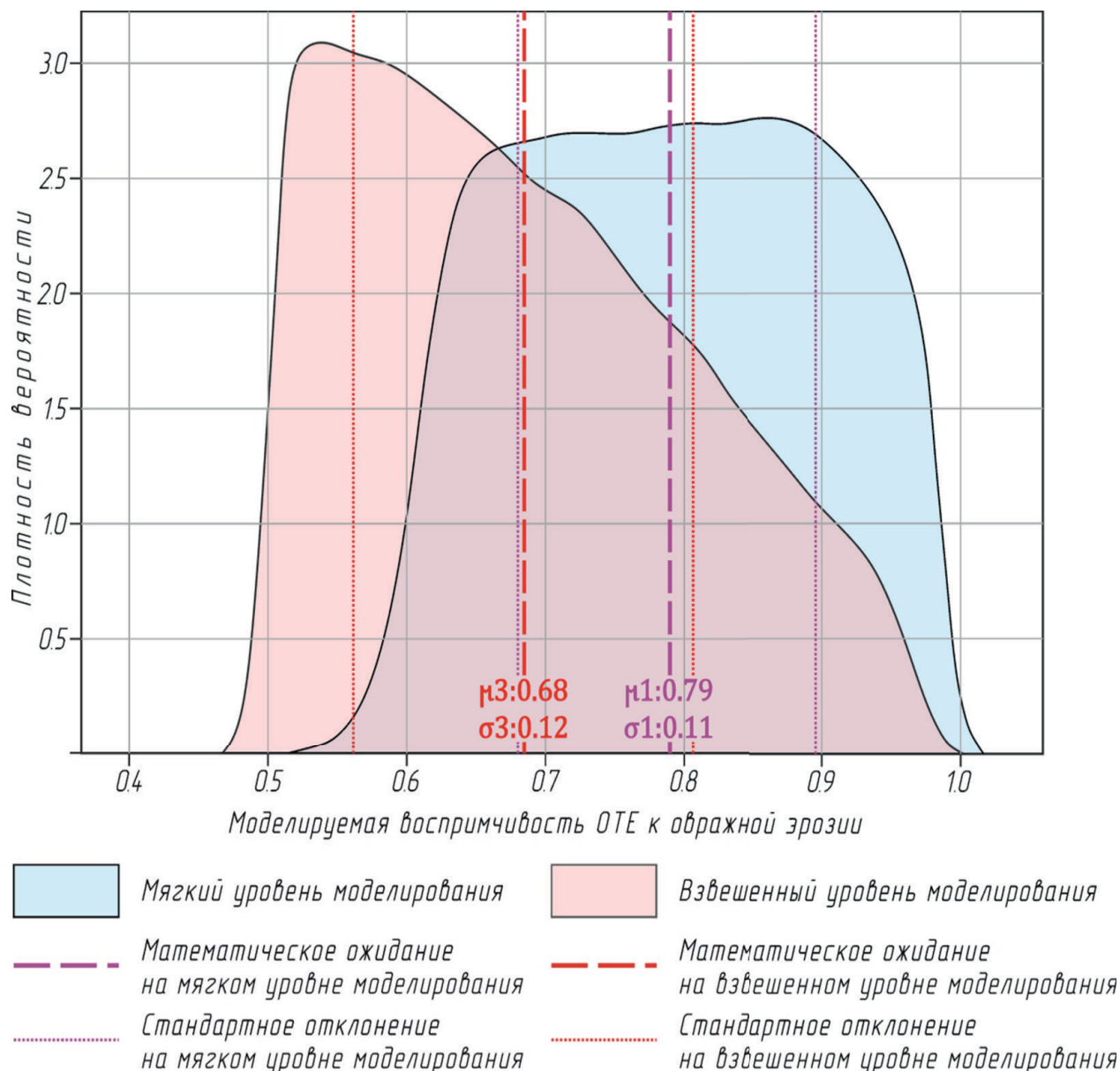
Следует отметить, что по результатам моделирования I уровня часто наблюдается формирование пространственных

кластеров с выраженной гомогенностью значений предрасположенности ОЭ к ОЭ. В рамках этих кластеров производная первого порядка от моделируемых значений практически не изменяется в пространстве, либо изменяется очень слабо. Фактически модель выделяет единые ареалы склоновых ландшафтов как наиболее предрасположенные для развития ОЭ (рис. 6). Результаты, полученные на II уровне моделирования, позволяют улучшить дифференциацию I уровня, выделив внутри склонов наиболее предрасположенные для развития эрозии территории (вогнутые склоны южных экспозиций с наибольшими перепадами высот и водосборными площадью (рис. 7).

Таким образом, полный схематический процесс производства оценки предрасположенности территории к ОЭ по предлагаемой методике приведён на рисунке 8.

Оценка точности III уровня моделирования, как и всей предлагаемой методики, производится по 3 показателям: *RF1* I и II уровней моделирования, а также *Recall* мягкого моделирования, которые в данном случае составляют 0,93, 0,72 и 0,95 соответственно.

Таким образом, разработанная авторами модель машинного обучения описывает 95% случаев наличия овражной эрозии в Воробьёвском районе Воронежской обла-



**Рис. 7 / Fig. 7.** Распределение частот предрасположенности территории к овражной эрозии на мягком и взвешенном уровнях моделирования / Frequency distribution of area susceptibility to gully erosion at soft and weighted modeling levels

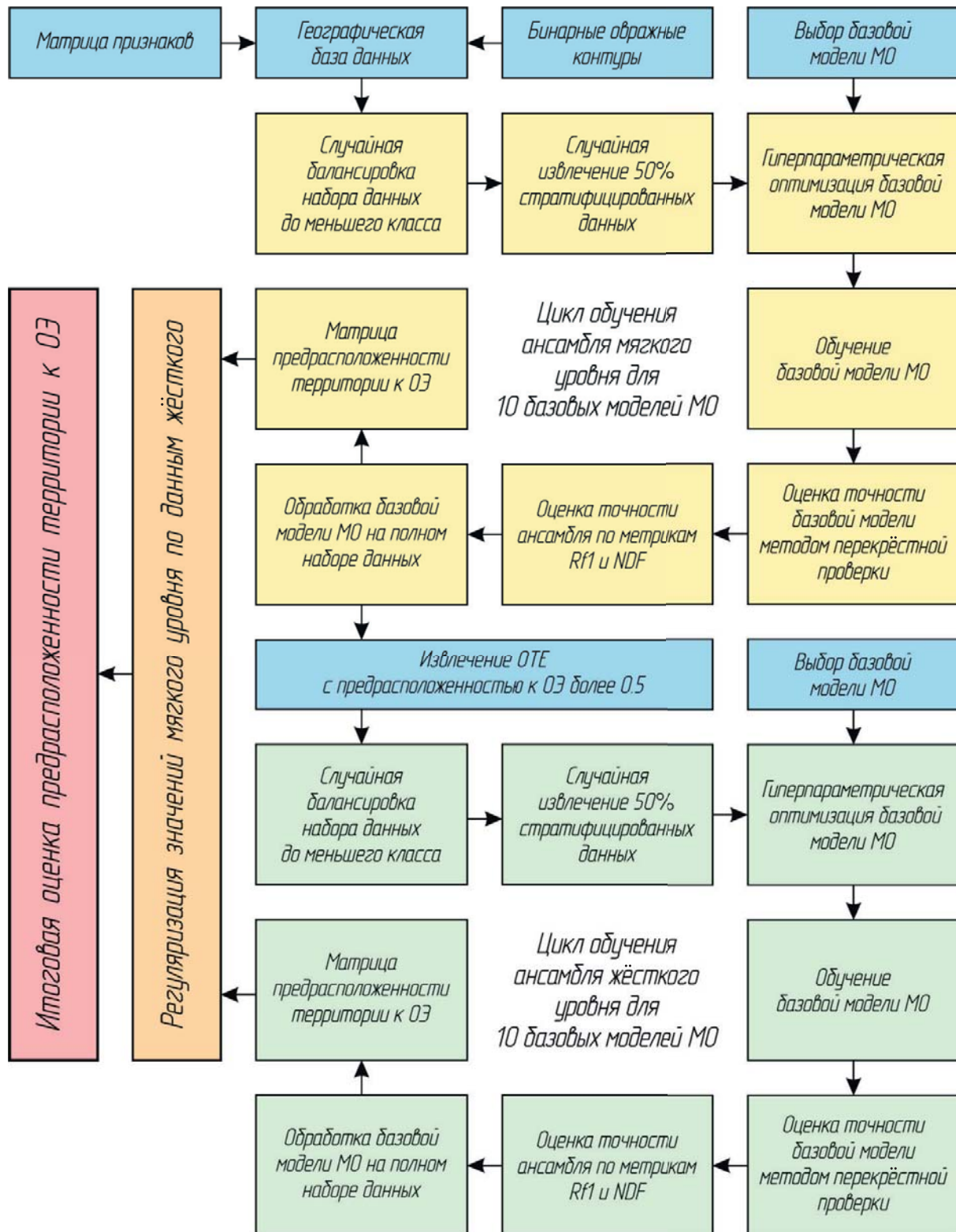
*Источник:* составлено авторами

сти. Наряду с существующими оврагами, она также способна показать территории, предрасположенные к развитию оврагов, но на которых в настоящее время процессы линейной эрозии не наблюдаются. Проведенное исследование позволило выявить такие земли и установить, что в настоящее время на территории исследования (Воробьевский муниципальный район) около 840 га относятся к категории земель с очень высокой предрасположенностью к развитию ОЭ (показатель оценки выше

0,9); 2013 га – к высокой (0,9–0,8); 3 092 га с предрасположенностью выше среднего (0,8–0,7); 3 876 га со средней (0,7–0,6); 4 303 га с низкой (0,6–0,5) (рис. 9).

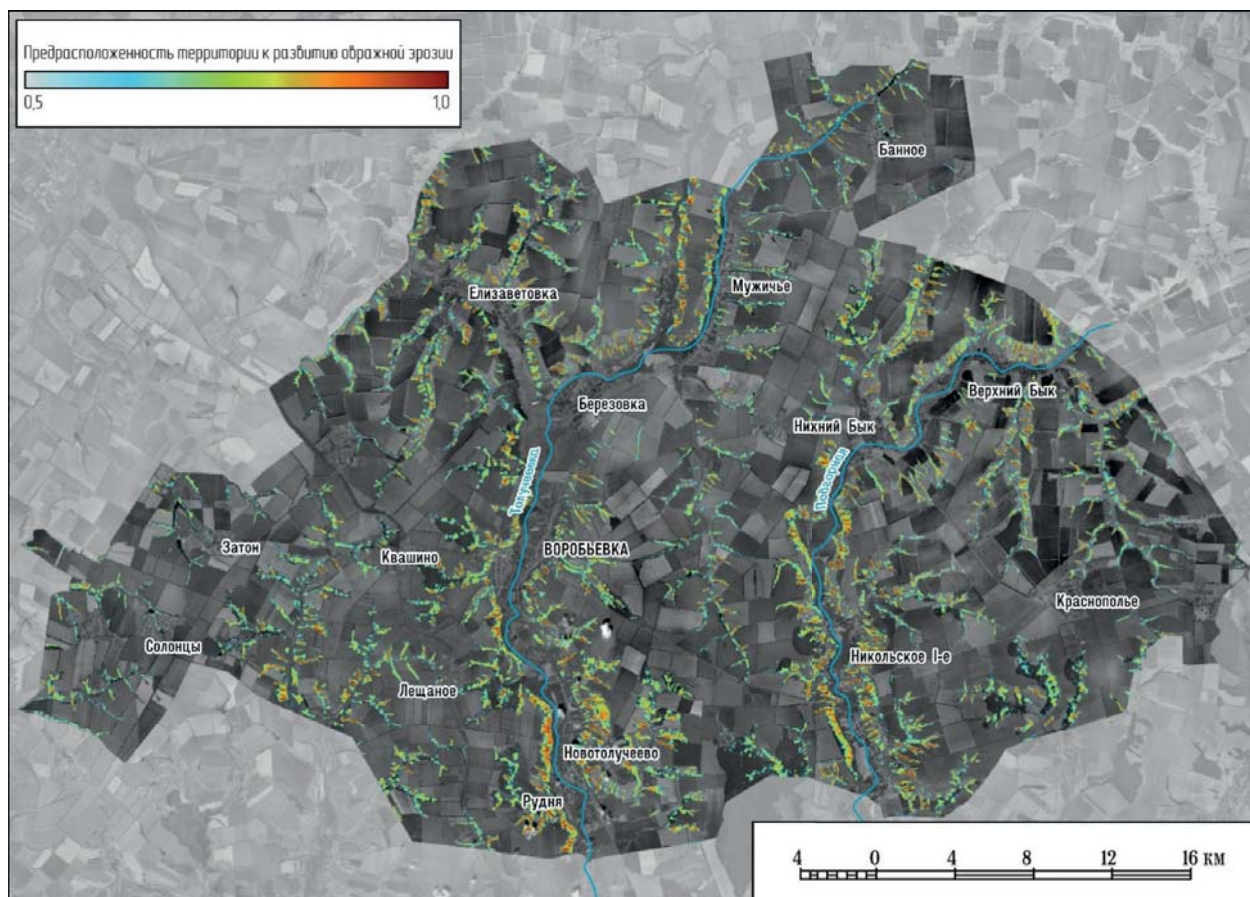
## ЗАКЛЮЧЕНИЕ

В рамках настоящей работы представлена реализация нового подхода к решению задачи оценки предрасположенности территории к развитию овражной эрозии.



**Рис. 8 / Fig. 8.** Процесс реализации метода сглаженной многоуровневой оценки predisposition территории к овражной эрозии / Workflow of implementing the method of smoothed multi-level assessment of the predisposition of the territory to gully erosion

*Источник:* составлено авторами



**Рис. 9 / Fig. 9.** Предрасположенность территории Воробьевского района Воронежской области к развитию овражной эрозии / Predisposition of the Vorobyevsky District of the Voronezh Region to gully erosion

Источник: составлено авторами

Он заключается в:

1) использовании в моделировании ансамблевого подхода с обучением на индивидуальной усечённой сбалансированной выборке;

2) расчёте специальных метрик точности ( $RFI$  для оценки качества работы ансамбля в целом и  $NDF$  для оценки «производительности» ансамбля);

3) выполнении моделирования на 3 уровнях: мягком — для оценки территории в целом; жёстком — для оценки территории с учетом минимизации влияния глобальных факторов; взвешенном — для получения более строгой оценки предрасположенности ОЭ к ОЭ.

Предлагаемый подход к ОЭ апробирован на территории исследования — Воробьевском районе Воронежской области и показал высокую точность описания существующей на данный момент ОЭ (степень согласованности — 0,95), а также вы-

сокие расчётные точности ансамблей I и II уровней по метрике  $RFI$  — 0,93 и 0,72 соответственно.

Использование многоуровневого подхода открывает возможность эффективно обнаружить несогласованность между факторами, обуславливающими овражную эрозию. Так, набор глобальных факторов (абсолютная высота местности, Гауссова кривизна, LS-фактор) не совпадает с набором внутренних факторов (относительные перепады высот, водосборная площадь и экспозиция склона).

По итогам моделирования на территории Воробьевского района Воронежской области было выявлено 2 853 га земель с высокой и очень высокой предрасположенностью к развитию овражной эрозии. В ландшафтном отношении им соответствуют остепенные долинно-балочные склоны южных экспозиций, отличающиеся наибольшей протяжённостью, крутизной и высотой,

имеющие вогнутый поперечный профиль и значительные водосборные площади (т. е. наибольшей предрасположенностью к развитию овражной эрозии обладают правые

коренные долинные склоны рр. Толучеевка и Подгорная у сёл Рудня и Никольское 1-е, уступы террас у с. Новотолучеево, склоны балок у сёл Елизаветовка и Мужичье).

#### ЛИТЕРАТУРА

1. География овражной эрозии / под. ред. Е. Ф. Зориной. М.: МГУ, 2006. 324 с.
2. Дроздов К. А. Распространение овражной эрозии на территории Центральных черноземных областей // Региональные ландшафтные исследования. 2022. Вып. 6. С. 65–75.
3. Ермолаев О. П., Рысин И. И., Голосов В. Н. Картографирование овражной эрозии на востоке Русской равнины // Геоморфология. 2017. № 2. С. 38–51. DOI 10.15356/0435-4281-2017-2-38-51
4. Корецкий Н. А., Бевз В. Н. Естественные факторы оврагообразования и их географический анализ на территории Воробьевского района Воронежской области // Региональные ландшафтные исследования. 2022. Вып. 6. С. 20–38.
5. Корецкий Н. А., Бевз В. Н. Методические вопросы моделирования вероятности проявления овражной эрозии с использованием машинного обучения // Известия Дагестанского государственного педагогического университета. Естественные и точные науки. 2023. Т. 17/ № 4. С. 19–24. DOI: 10.31161/1995-0675-2022-16-1-19-24
6. Корецкий Н. А., Бевз В. Н. Некоторые закономерности проявления овражной эрозии на территории Воробьевского района Воронежской области // Региональные ландшафтные исследования. 2020. Вып. 4. С. 113–122.
7. Сидорчук А. Ю. Экспресс-метод оценки овражного потенциала // Вестник Московского университета. Серия 5: География. 2021. № 1. С. 54–61.
8. A 30 m global map of elevation with forests and buildings removed / L. Hawker, P. Uhe, L. Paulo, J. Sosa // Environmental Research Letters. 2022. Vol. 17. P. 024016. DOI: 10.1088/1748-9326/ac4d4f
9. Breiman L. Random Forests // Machine Learning. 2001. Vol. 45. P. 5–32. DOI: 10.1023/A:1010933404324
10. CatBoost: unbiased boosting with categorical features / L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, A. Gulin // Journal of Functional Analysis. 2017. Vol. 1706. P. 09516.
11. Comparison of machine learning models for gully erosion susceptibility mapping / A. Arabameri, Wei Chen, M. Loche, Xia Zhao, Yang Li, L. Lombardo, A. Cerda, et al. // Geoscience Frontiers. 2020. Vol. 11. P. 1609–1620. DOI: 10.1016/j.gsf.2019.11.009
12. GIS-based gully erosion susceptibility modeling, adapting bivariate statistical method and AHP approach in Gombe town and environs Northeast Nigeria / O. Igwe, U. I. John, O. Solomon, O. Ozioko // Geoenviron Disasters. 2020. Vol. 7. DOI: 10.1186/s40677-020-00166-8
13. Florinsky I. V. Digital Terrain Analysis in Soil Science and Geology. Amsterdam: Elsevier Inc., 2016. 486 p.
14. Identification of gully erosion activity and its influencing factors: A case study of the Sunshui River Basin / Fengjie Fan, Xingli Gu, Jun Luo, Bin Zhang, Hui Liu, Haiqing Yang, Lei Wang // PLoS ON. 2024. Vol. 19. DOI: 10.1371/journal.pone.0309672
15. Integrated machine learning and geospatial analysis enhanced gully erosion susceptibility modeling in the Erer watershed in Eastern Ethiopia / T. B. Gelete, P. Pasala, N. G. Abay, G. Weldu // Environmental Informatics and Remote Sensing. 2024. Vol. 12. P. 1–20. DOI: 10.3389/fenvs.2024.1410741
16. Machine Learning-Based Gully Erosion Susceptibility Mapping: A Case Study of Eastern India / S. Saha, J. Roy, A. Arabameri, J. Roy, T. Blaschke // Sensors. 2020. Vol. 20. P. 1313. DOI: 10.3390/s20051313
17. Shary P. A., Sharaya L. S., Mitusov A. V. Fundamental quantitative methods of land surface analysis // Geoderma. 2002. Vol. 107. P. 1–32. DOI: 10.1016/S0016-7061(01)00136-7
18. Sidorchuk A. The Theoretical Approach to the Modelling of Gully Erosion in Cohesive Soil // Earth. 2022. Vol. 3. P. 228–244. DOI: 10.3390/earth3010015
19. Wang Y., Zhang Y., Chen H. Gully erosion susceptibility prediction in Mollisols using machine learning models // Journal of Soil and Water Conservation. 2023. Vol. 78. P. 385–396. DOI: 10.2489/jswc.2023.00019

#### REFERENCES

1. Zorina E. F., ed. *Geografiya ovrazhnoj erozii* [Geography of gully erosion]. Moscow: MSU Publ., 2006, 324 p.
2. Drozdov K. A. [Geography of gully erosion in the Central Cernozems Region]. In: *Regional-*



- nye landshaftnye issledovaniya* [Regional landscape researches], 2022, vol. 6, pp. 65–75.
3. Yermolayev O. P., Rysin I. I., Golosov V. N. [Mapping assessment of gully erosion in the East of the Russian Plain]. In: *Geomorfologiya* [Geomorphology], 2017, vol. 2, pp. 8–51. DOI: 10.15356/0435-4281-2017-2-38-51
  4. Korecky N. A., Bevz V. N. [Natural factors of gully formation and their geographical analysis on the territory of Vorobyevsky district of Voronezh region]. In: *Regionalnye landshaftnye issledovaniya* [Regional landscape researches], 2022, vol. 6, pp. 20–38.
  5. Koretsky N. A., Bevz V. N. Methodological Issues of the Gully Erosion Probability Models in the Central Chernozem Region Using Machine Learning. In: *Izvestiya Dagestanskogo gosudarstvennogo pedagogicheskogo universiteta. Yestestvennyye i tekhnicheskiye nauki* [News of the Dagestan State Pedagogical University. Natural and Exact Sciences], 2023, vol. 17, pp. 41–55. DOI: 10.31161/1995-0675-2023-17-3-41-55
  6. Korecky N. A., Bevz V. N. [Some patterns of gully erosion on the territory of Vorobyovka district, Voronezh region]. In: *Regionalnye landshaftnye issledovaniya* [Regional landscape researches], 2020, vol. 4, pp. 113–122.
  7. Sidorchuk A. Yu. Express method of estimation of gully erosion potential. In: *Lomonosov Geography Journal*, 2021, vol. 1, pp. 54–61.
  8. Hawker L., Uhe P., Paulo L., Sosa J. A 30 m global map of elevation with forests and buildings removed. In: *Environmental Research Letters*, 2022, vol. 17, p. 024016. DOI: 10.1088/1748-9326/ac4d4f
  9. Breiman L. Random Forests. In: *Machine Learning*, 2001, vol. 45, pp. 5–32. DOI: 10.1023/A:1010933404324
  10. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A. CatBoost: unbiased boosting with categorical features. In: *Journal of Functional Analysis*, 2017, vol. 1706, pp. 09516.
  11. Arabameri A., Wei Chen, Loche M., Xia Zhao, Yang Li, Lombardo L., Cerda A., et al. Comparison of machine learning models for gully erosion susceptibility mapping. In: *Geoscience Frontiers*, 2020, vol. 11, pp. 1609–1620. DOI: 10.1016/j.gsf.2019.11.009
  12. Igwe O., John U. I., Solomon O., Ozioko O. GIS-based gully erosion susceptibility modeling, adapting bivariate statistical method and AHP approach in Gombe town and environs Northeast Nigeria. In: *Geoenviron Disasters*, 2020, vol. 7. DOI: 10.1186/s40677-020-00166-8
  13. Florinsky I. V. *Digital Terrain Analysis in Soil Science and Geology*. Amsterdam: Elsevier Inc., 2016. 486 p.
  14. Fengjie Fan, Xingli Gu, Jun Luo, Bin Zhang, Hui Liu, Haiqing Yang, Lei Wang. Identification of gully erosion activity and its influencing factors: A case study of the Sunshui River Basin. In: *PLoS ON*, 2024, vol. 19. DOI: 10.1371/journal.pone.0309672
  15. Gelete T. B., Pasala P., Abay N. G., Weldu G. Integrated machine learning and geospatial analysis enhanced gully erosion susceptibility modeling in the Erer watershed in Eastern Ethiopia. In: *Environmental Informatics and Remote Sensing*, 2024, vol. 12, pp. 1–20. DOI: 10.3389/fenvs.2024.1410741
  16. Saha S., Roy J., Arabameri A., Roy J., Blaschke T. Machine Learning-Based Gully Erosion Susceptibility Mapping: A Case Study of Eastern India. In: *Sensors*, 2020, vol. 20, pp. 1313. DOI: 10.3390/s20051313
  17. Shary P. A., Sharaya L. S., Mitusov A. V. Fundamental quantitative methods of land surface analysis. In: *Geoderma*, 2002, vol. 107, pp. 1–32. DOI: 10.1016/S0016-7061(01)00136-7
  18. Sidorchuk A. The Theoretical Approach to the Modelling of Gully Erosion in Cohesive Soil. In: *Earth*, 2022, vol. 3, pp. 228–244. DOI: 10.3390/earth3010015
  19. Wang Y., Zhang Y., Chen H. Gully erosion susceptibility prediction in Mollisols using machine learning models. In: *Journal of Soil and Water Conservation*, 2023, vol. 78, pp. 385–396. DOI: 10.2489/jswc.2023.00019

#### ИНФОРМАЦИЯ ОБ АВТОРАХ

*Корецкий Никита Александрович* – преподаватель кафедры физической географии и оптимизации ландшафта факультета географии, геоэкологии и туризма Воронежского государственного университета;  
email: nkoretskiy98@bk.ru

*Горбунов Анатолий Станиславович* – кандидат географических наук, доцент кафедры физической географии и оптимизации ландшафта факультета географии, геоэкологии и туризма Воронежского государственного университета;  
email: gorbunov.ol@mail.ru

*Бевз Валерий Николаевич* – кандидат географических наук, доцент кафедры физической географии и оптимизации ландшафта факультета географии, геоэкологии и туризма Воронежского государственного университета;  
email: snark61@mail.ru

#### INFORMATION ABOUT THE AUTHORS

*Nikita A. Koretsky* – lecturer, Department of Physical Geography and Landscape Optimization, Faculty of Geography, Geoecology and Tourism, Voronezh State University;  
email: nkoretskiy98@bk.ru

*Anatoliy S. Gorbunov* – PhD (Geography), Associate Professor, Department of Physical Geography and Landscape Optimization, Faculty of Geography, Geoecology and Tourism, Voronezh State University;  
email: gorbunov.ol@mail.ru

*Valeriy N. Bevz* – PhD (Geography), Associate Professor, Department of Physical Geography and Landscape Optimization, Faculty of Geography, Geoecology and Tourism, Voronezh State University;  
email: snark61@mail.ru