

# Определение содержания органического вещества почвы Минской области Беларуси на основе классификаторов градиентного бустинга спутниковых данных

А. Н. Червань, Б. Чжао

Белорусский государственный университет, Минск, 220030, Беларусь  
E-mails: chervanalex@mail.ru, geozhao@outlook.com

Одним из основных показателей проявления процессов деградации почв сельскохозяйственных земель является потеря содержания органического вещества (ОВ) почвы. Поэтому актуальны методики контроля пространственного и временного распределения ОВ как с экологической, так и с экономической точки зрения. В настоящее время данные дистанционного зондирования, в частности набор спутниковых снимков Sentinel-2, могут быть использованы для определения содержания ОВ в поверхностном гумусово-аккумулятивном горизонте почв пахотных земель. Целью данного исследования является использование данных Sentinel-2 и данных полевого почвенного картирования в Беларуси для анализа пространственной точности оценки содержания ОВ и результатов автоматизированного дешифрирования спутниковых снимков на примере Минской области Беларуси. Использован ряд аналитических методов, включая выбор репрезентативных спектральных каналов в качестве входных данных модели, с применением корреляционного анализа Спирмена и классификаторов градиентного бустинга для моделирования в геоинформационных системах. Предлагаемый подход оценивается с позиции достижения высокоточной и быстрой инверсии и пространственного анализа содержания ОВ в генетических типах почв. Точность инверсионной модели оценивалась с использованием независимой базы данных валидации. Результаты пространственного анализа показали, что в модели, основанной на снимках Sentinel-2 в диапазонах B6, B7, B8, B8A и B12, наиболее эффективно применяются производные второго порядка. Модель инверсии отразила самую высокую точность (минимум 93,8 % при среднем 96,2 %), ошибка RMSE составила 0,31, коэффициент Каппа — 0,985. Самое низкое содержание ОВ в пахотных почвах характерно для Березинского района, а самое высокое — для Любанского и Солигорского районов, отличающихся значительной долей гидротехнически мелиорированных сельскохозяйственных земель.

**Ключевые слова:** дистанционное зондирование, Sentinel-2, содержание органического вещества в почве, данные дистанционного зондирования, классификатор градиентного бустинга

Одобрена к печати: 17.03.2025

DOI: 10.21046/2070-7401-2025-22-3-121-135

## Введение

Органическое вещество (ОВ) почвы — важнейший компонент, влияющий на её производительную способность и секвестрацию органического углерода. Оно обеспечивает сельскохозяйственные культуры рядом питательных веществ, одновременно оказывая значительное влияние на элементарные эпигенетические геохимические характеристики (Zhang et al., 2021). По мере интенсификации сельского хозяйства тренды изменения содержания ОВ почвы во всём мире претерпевают постепенное изменение. Для сельскохозяйственных земель особо актуальна оценка текущего содержания ОВ, в том числе гумуса в почве для управления продуктивной способностью в агроландшафтах. Вместе с тем содержание ОВ является одной из трёх подцелей для почвенных ресурсов в соответствии с Целями устойчивого развития Организации Объединённых Наций. Оценка фактического содержания ОВ почвы особенно важна в условиях, наиболее изменённых хозяйственной деятельностью, — гидротехнически осушенных мелиорированных землях с искусственным регулированием водного режима.

Традиционный подход к мониторингу ОВ подразумевает сбор значительного количества полевых образцов почвы и последующий химический лабораторный анализ. Этот метод по своей сути довольно трудоёмкий, а его точность зависит от плотности собранных образцов. Следовательно, сложно удовлетворить требования быстрого развития современного

сельского хозяйства в получении качественной пространственной оценки агрохимического состояния почвенного покрова (Liu et al., 2020). По мере развития и уточнения технологии дистанционного зондирования определение содержания ОВ посредством спектральной дифференциации становится всё более эффективным приёмом.

Дистанционная оценка содержания ОВ включает два основных направления исследований: обработку и выбор спектральной информации и построение моделей инверсии. Полученные ранее результаты (Diek et al., 2017; Habibi et al., 2019; Yang et al., 2015) указывают на широкое использование многоспектральных изображений дистанционного зондирования для картирования и оценки содержания ОВ. Спектральные характеристики органической части, полученные по изображениям Landsat в видимом и ближнем инфракрасном диапазонах, могут быть использованы для оценки её содержания (Guo et al., 2013; Habibi et al., 2019). Исследования (Ahmed, Iqbal, 2014; Yang et al., 2015) показывают, что каналы Landsat 1, 2, 3, 4, 6 и 7 особенно показательны для моделирования и картирования ОВ. Аналогично, благодаря улучшенному пространственному разрешению Sentinel-2 по сравнению с Landsat, он также демонстрирует потенциал для мониторинга и картирования ОВ (Li et al., 2021; Tang et al., 2022; Wang et al., 2021; Zhang et al., 2021; Zhou et al., 2020). Тем не менее точность оценок ОВ по данным Sentinel-2 может зависеть от множества переменных. Большой канал пропускания, низкое пространственное разрешение и низкое отношение сигнал/шум каналов SWIR (*англ.* short wave infrared) (B11 и B12) на Sentinel-2 по сравнению с лабораторными данными могут привести к снижению прогностического эффекта. Кроме того, точность картирования ОВ с использованием снимков Sentinel-2 зависит от влажности почвы, остатков урожая и состояния рабочих участков (Castaldi et al., 2019). Вместе с тем время и частота сбора данных Sentinel также способны воздействовать на эффективность оценки ОВ по отношению к вегетативным стадиям формирования урожая разных культур (Pan et al., 2019). Вдобавок дата сбора и связанные с ней погодные и почвенные условия могут влиять на точность прогнозов относительно содержания ОВ в верхнем горизонте почв сельскохозяйственных земель (Vaudour et al., 2019). Для решения этих проблем с целью улучшения пространственной достоверности оценки ОВ используется сочетание данных из нескольких источников, например Sentinel-1 и Sentinel-2 (Yin et al., 2023). Алгоритмы машинного обучения также задействуются для картирования ОВ с применением изображений Sentinel-2 (Wang et al., 2021; Zhang et al., 2021; Zhou et al., 2020).

Классификаторы градиентного бустинга (ГБ) — это тип алгоритма машинного обучения, который всё чаще используется для определения содержания ОВ из-за его способности обрабатывать сложные наборы данных и повышать точность прогнозирования. Эти модели особенно полезны в контексте экологических и сельскохозяйственных исследований, где точное картирование свойств почвы имеет решающее значение для устойчивого управления сельскохозяйственной деятельностью. Модели ГБ продемонстрировали превосходную производительность в прогнозировании ОВ по сравнению с альтернативными моделями, такими как метод случайного леса (*англ.* Random Forest). Например, XGBoost показал (Ye et al., 2021) самую высокую точность в исследованиях, использующих дистанционное зондирование GF-6 (*англ.* GaoFen-6), с коэффициентом детерминации ( $R^2$ ) 0,771. По сравнению с альтернативными моделями, ГБ обычно демонстрирует превосходную точность и вычислительную эффективность в отличие от традиционных методов, таких как обычный кригинг и регрессионный кригинг. Они повышают точность прогнозирования среднеквадратичной ошибки (RMSE) на 6–9 % и  $R^2$  на 250 % (Tziachris et al., 2019) в сопоставлении с альтернативными методологиями машинного обучения. В отдельном исследовании (Chen et al., 2019) классификатор ГБ продемонстрировал самую низкую среднеквадратичную ошибку, что делает его надёжным методом при картировании ОВ на больших территориях.

## Материалы и методы

Минская область расположена в центральной части Республики Беларусь (официальный источник: [https://gki.gov.by/ru/activity\\_branches-land-reestr/](https://gki.gov.by/ru/activity_branches-land-reestr/)), охватывая территорию в 39 900 км<sup>2</sup>. Она состоит из 22 районов. Согласно актуальным данным, предоставленным Государственным Комитетом по имуществу Республики Беларусь, общая площадь сельскохозяйственных земель составляет 1609,2 тыс. га, из которых 1199,9 тыс. га классифицируются как пашня, 402,6 тыс. га являются луговыми землями. Как отражено в *табл. 1*, осушенные сельскохозяйственные земли составляют 589,6 тыс. га, из которых 314,8 тыс. га используются под пропашные культуры. Векторные данные (границы почвенных таксонов с качественной группировкой по содержанию ОВ почвы) для представленного исследования являются частью земельно-информационных систем административных районов и результатом цифрового почвенного картирования всех сельскохозяйственных земель Беларуси для целей кадастровой оценки и налогообложения. Белорусская система цифрового картирования почв в рамках Государственного земельного кадастра играет важнейшую роль в управлении почвенными и земельными ресурсами для сельскохозяйственной деятельности и охраны окружающей среды. Она включает пространственные данные о классификации почв и землепользовании в масштабе 1:10 000, поддерживая принятие решений в сельском хозяйстве, инфраструктуре других видов экономической деятельности и охране окружающей среды. Следует отметить имеющуюся в Беларуси интеграцию цифровых почвенных данных с данными дистанционного зондирования и периодическое, не реже чем раз в 6–8 лет, обновление, что позволяет сопровождать научные исследования данными о современном физическом и агрохимическом состоянии почвенного покрова сельскохозяйственных земель.

*Таблица 1.* Реестр использования сельскохозяйственных земель в Минской области Беларуси (тыс. га)

Данные	Всего земель	Пахотные	Многолетние насаждения и сады	Луговые земли	Улучшенные луговые земли	Всего с.-х. земли	Лесные	Земли под древесно-кустарниковой растительностью	Земли под дорогами и иные
По административным границам области	3983,5	1265,9	22,8	418,0	312,0	1706,7	1692,5	148,1	68,5
По сельскохозяйственным организациям	1748,3	1149,3	5,2	381,9	292,4	1536,4	0	97,6	12,3
По осушенным землям в области	705,5	314,8	22,8	418,0	312,0	1706,7	1692,5	148,1	68,5

В соответствии с базой данных WRB (*англ.* World Reference Base) тремя наиболее распространёнными типами почв сельскохозяйственных земель являются Albeluvisols (Arenic, Ruptic), Hemic Histosols и Stagnosols (Endogleic, Umbric, Huplic) (Клебанович и др., 2011; Лапа и др., 2017; Романова, 2004; Червань и др., 2022). В настоящем исследовании территория Мядельского, Крупского, Дзержинского, Пуховичского и Солигорского районов была выбрана в качестве репрезентативных ключевых участков образцов на основе предоставленной информации о типах почв и их распределении в границах районов. Часть территории указанных районов использовалась в качестве обучающих (около 150 000 центроидов почвенных разновидностей — самых низких таксонов) и проверочных (около 1000 участков) наборов для геостатистического анализа. Конкретная информация проиллюстрирована на *рис. 1* (см. с. 124). Снимки Sentinel-2 включают изображения за апрель 2024 г. в связи с тем, что в этот период почва меньше занята снежным покровом и минимально перекрывается ещё неразвитыми озимыми культурами, при этом характерна меньшая облачность. В настоя-

щем исследовании использовались данные уровня L2A с платформы GEE (*англ.* Google Earth Engine) и был выбран поднабор с облачностью менее 1 %. Затем необработанные данные для улучшения корреляции между отражательной способностью и ОВ почвы подверглись ряду математических преобразований, включая логарифмические, степенные, обратные, обратные логарифмические, первую и вторую производную. Полосы, прошедшие тест значимости, использовались в качестве характерных зон для соответствующей инверсии, а преобразования с самыми высокими коэффициентами корреляции были объединены в качестве эталона для полос отклика. Поскольку данные, используемые в этом исследовании, являются иерархическими, результаты математических изменений на их основе не являются значимыми, но всё же предоставляют эффективную помощь в распознавании различных характеристик ОВ. Производная здесь относится к расчёту скорости изменения отражательной способности относительно длины волны, что удаляет линейный фоновый шум и выделяет точки спектральной резкости (например, местоположения красного края). После проверки с помощью программного обеспечения SPSS (*англ.* Statistical Package for the Social Sciences) был использован GEE как облачная платформа и рассчитан коэффициент отражательной способности для определения ОВ почвы.

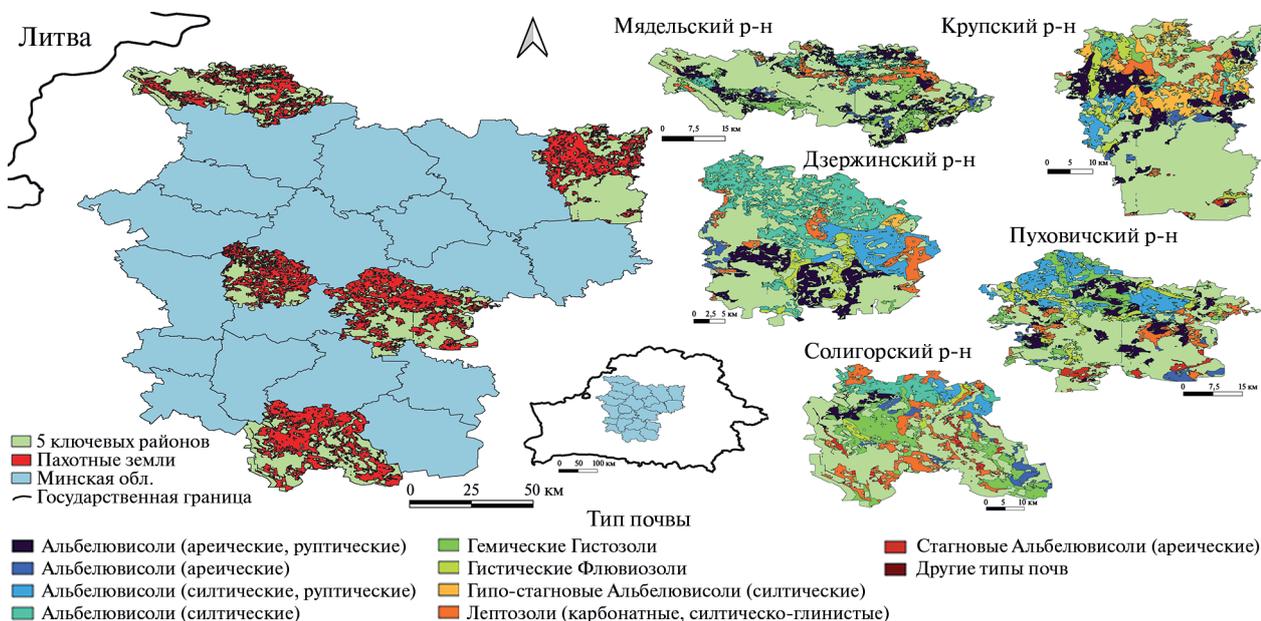


Рис. 1. Территория исследования и обучающая выборка

### Модели калибровки органического вещества почвы. Коэффициент ранговой корреляции Спирмена

Для определения спектральных зон корреляции ОВ почвы использовался коэффициент ранговой корреляции Спирмена. Это непараметрический статистический метод, применяемый для оценки монотонной связи между двумя переменными. Коэффициент ранговой корреляции Спирмена может быть использован для измерения корреляций между непрерывными или последовательными данными. Он особенно подходит для ненормально распределённых или категориальных, например, градуированных, данных (Zhang, Wang, 2023). В настоящем исследовании сведения о содержании ОВ почвы Минской области Беларуси, полученные из полевых почвенно-геоботанических экспедиций (Лапа и др., 2017), сгруппированы в пять классов со следующими уровнями: низкий, средне-низкий, средний, средне-высокий и высокий. Выполнены необходимые условия для применения коэффициента ранговой корреляции Спирмена. Подобно другим распространённым тестам корреляции, таким как корреляция Пирсона, расчёт ранговой корреляции Спирмена включает тест значимости: зна-

чения [0...1] указывают на идеальную положительную корреляцию, в то время как значения [1...0] указывают на идеальную отрицательную корреляцию.

Была рассчитана корреляция между содержанием ОВ почвы и отражательной способностью изображений Sentinel-2A, а также применёнными к ним преобразованиями. Коэффициенты корреляции между содержанием ОВ и спектральной отражательной способностью отрицательны при невысоких уровнях связи. При этом самые высокие коэффициенты корреляции для различных математических преобразований разных каналов наблюдаются в производных преобразованиях второго порядка. Это указывает на эффективность этого метода предварительной обработки для улучшения корреляции между спектральной яркостью и ОВ почвы по снимкам Sentinel-2A. Каналы, которые продемонстрировали статистическую значимость в различных математических преобразованиях, использовались в качестве характерных для соответствующей инверсии. Преобразования с самыми высокими коэффициентами корреляции в каждом канале затем были объединены в качестве эталона для зон отклика с исключением данных из каналов, которые не прошли тест значимости (рис. 2). Таким образом, производные второго порядка каналов В3 (центральная длина волны равна 559 нм, ширина — 46 нм), В6 (739,1 нм, 18 нм), В7 (779,7 нм, 28 нм), В8 (832,9 нм, 133 нм), В8А (864,0 нм, 33 нм), В9 (945 нм, 26 нм) и В12 (2185,7 нм, 238 нм) выбираются для следующей обработки. Вегетационный индекс NDVI (англ. Normalized Difference Vegetation Index) и водный индекс NDWI (англ. Normalized Difference Water Index) использовались параллельно с поканальными данными. Данные цифровых моделей рельефа при добавлении в качестве входных переменных ухудшили точность модели, что подтверждается другими исследованиями (Zhou et al., 2021), в связи с чем использовались только результаты дерева классификации ГБ.

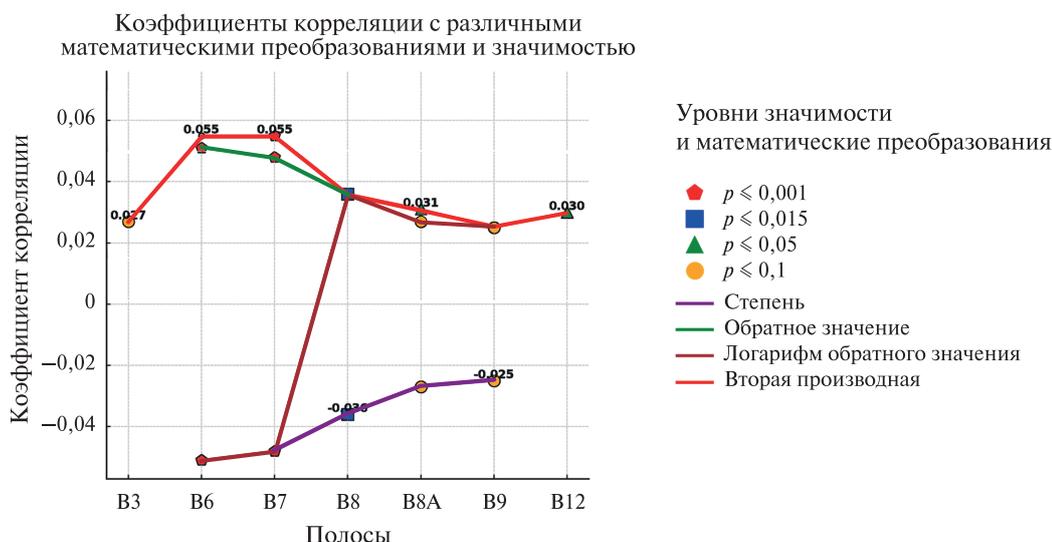


Рис. 2. Значимые коэффициенты корреляции с выделенным максимальным (абсолютным) значением канала

### Классификатор градиентного бустинга

Классификатор является интегрированным методом обучения и основан на алгоритме ГБ, в основном используемый для задач классификации данных (Demir, Sahin, 2023). Градиентный бустинг улучшает общую производительность модели путём итеративного построения нескольких слабых классификаторов (обычно деревьев решений) для постепенного уменьшения ошибки. Среди преимуществ метода выделяются следующие:

1. Высокая точность: может постепенно уменьшать ошибку с помощью итеративной оптимизации, что подходит для нелинейных задач с высокой сложностью.

2. Гибкость: поддерживает пользовательские функции потерь, которые можно адаптировать к различным сценариям (например, логарифмическая функция потерь для задач классификации).

3. Возможность предотвращения переобучения: регулируя скорость обучения, количество и глубину слабых классификаторов, ГБ может эффективно предотвращать переобучение.

4. Оценка важности признаков: может предоставить оценку важности каждого входного признака для понимания ключевых факторов, влияющих на результаты классификации.

Авторами ГБ использовался для инвертирования классов ОВ почвы (алгоритм GradientBoostingClassifier в библиотеке `sklearn.ensemble`), и по сравнению со случайным лесом и аналогичным XGBoost ГБ имеет следующие преимущества:

1. Преимущество точности: ГБ способен лучше подгонять сложные данные, минимизируя градиент функции потерь в пошаговых итерациях. В этой задаче инверсии ОВ почвы точность ГБ выше, чем у случайного леса и XGBoost, особенно при классификации среднего класса (2–4-го класса).

2. Обработка несбалансированных данных: настраивая функцию потерь и скорость обучения, ГБ может более эффективно справляться с проблемой неравномерного распределения классов ОВ почвы, в то время как случайный лес слабее при столкновении с несбалансированными данными.

3. Оценка вклада признаков более стабильна: оценка важности признаков ГБ более стабильна, что помогает идентифицировать ключевые каналы, влияющие на распределение ОВ почвы, такие как В6, В8А и В9. Напротив, случайный лес имеет некоторые колебания в расчёте важности признаков, а результаты XGBoost зависят от большей настройки параметров.

4. Производительность обобщения: благодаря более низкой скорости обучения и более слабым классификаторам ГБ может достичь лучшей производительности обобщения и снизить риск переобучения. По сравнению с аналогичным более быстрым алгоритмом XGBoost чрезмерная оптимизация способна легко привести к переобучению и повлиять на надёжность прогноза (Anantha Natarajan et al., 2022; Kang et al., 2022; Sahin, 2022).

Поэтому в этом исследовании использовался метод ГБ в функции `sklearn.ensemble` платформы GEE, а тренер был настроен в соответствии со следующими параметрами для достижения инверсии содержимого ОВ почвы для Минской области Беларуси: количество деревьев решений 200, усадка 0,1, частота дискретизации 1,0, максимальное количество узлов 16. Выбор данных параметров обоснован ниже.

Большее количество деревьев обычно улучшает обучаемость модели, но также может привести к переобучению. Поскольку данные включают в себя снимки дистанционного зондирования и инверсию ОВ почвы, данные имеют сильную нелинейную связь, и использование большего количества деревьев помогает улавливать сложные закономерности. Количество 100–500 обычно считается подходящим выбором; 200 деревьев является умеренным числом, обеспечивающим достаточную мощность моделирования, не приводя к высоким вычислительным затратам.

Усадка управляет долей каждого дерева, которая вносит вклад в окончательный прогноз. Если скорость обучения слишком высока (например, 0,5), модель может сходиться слишком быстро и упускать лучшие решения; если она слишком мала (например, 0,01), для достижения того же результата требуется больше деревьев. Значение 0,1 обычно используется по умолчанию для ГБ, которое обеспечивает лучшее обобщение в большинстве случаев.

Частота дискретизации, равная 1,0, означает, что каждое дерево будет обучено с применением полных обучающих данных без подвыборки. Контроль доли выборок, задействуемых при обучении каждого дерева, часто используется для снижения риска переобучения.

Ограничение максимального количества конечных узлов на дерево контролирует сложность модели. Если количество слишком мало (например, 4), модель может быть недообучена и изменения в значении ОВ почвы улавливаются недостаточно. Если оно слишком велико (например, 100+), сложность каждого дерева будет высокой и вероятно переобучение модели. Значение 16 обычно поддерживает хороший баланс между моделированием сложных взаимосвязей, не делая модель слишком сложной для переобучения.

## Результаты и обсуждение

### Моделирование органического вещества почвы и валидация данных

Конкретная информация о корреляции между каналами и классами ОВ почвы, полученная на основе корреляционного анализа Спирмена, представлена в *табл. 2*. С точки зрения разных математических вариаций корреляция между каналами и классами содержания ОВ почвы была разной. С помощью обратных и обратных логарифмических вариаций самая высокая положительная корреляция обнаружена в нескольких каналах, таких как В6 (0,051) и В7 (0,048 — самая высокая среди всех результатов). Производительность каналов В8 и В8А для вариации второй производной была хорошей, достигнув 0,036 и 0,027 соответственно, из-за того, что эти каналы расположены в ближней инфракрасной области. Она имеет сильную положительную корреляцию со спектральным откликом на содержание ОВ почвы, и значимость этой корреляции была хорошей. Что касается логарифмической и исходной форм вариации, наблюдалась общая отрицательная корреляция, особенно в каналах В6 и В7, где коэффициенты корреляции достигли  $-0,051$  и  $-0,048$  соответственно, показывая значительную отрицательную связь.

Таблица 2. Характеристика каналов, выбранных на основе корреляционного анализа

Канал	Первая производная	Логарифм	Оригинал	Взаимный	Обратный логарифм	Вторая производная
В1	0,002	$-0,009$	$-0,009$	0,009	0,009	0,011
В2	0,027	$-0,002$	$-0,002$	0,002	0,002	0,002
В3	0,016	$-0,024$	$-0,024$	0,024	0,024	0,027
В4	$-0,005$	$-0,011$	$-0,011$	0,011	0,011	0,008
В5	$-0,003$	$-0,021$	$-0,021$	0,021	0,021	0,024
В6	0,005	$-0,051^3$	$-0,051^3$	$0,051^3$	$0,051^3$	$0,055^3$
В7	$-0,003$	$-0,048^3$	$-0,048^3$	$0,048^3$	$0,048^3$	$0,050^3$
В8	0,002	$-0,036^2$	$-0,036^2$	$0,036^2$	$0,036^2$	$0,036^2$
В8А	0,008	$-0,027^2$	$-0,027^2$	$0,027^2$	$0,027^2$	$0,031^2$
В9	0	$-0,025$	$-0,025$	0,025	0,025	0,025
В10	$-0,007$	$-0,003$	$-0,003$	0,003	0,003	0,003
В11	0,007	$-0,006$	$-0,006$	0,006	0,006	0,012
В12	0,007	$-0,024^1$	$-0,024^1$	$0,024^1$	$0,024^1$	$0,030^1$

Значимость коэффициентов корреляции:  $^1 < 0,05$ ;  $^2 < 0,01$ ;  $^3 < 0,001$ .

Из-за использования иерархических данных в этом исследовании корреляция Спирмена для прерывистых данных делает коэффициент корреляции математически незначимым. Поэтому они не имеют существенной ценности при построении математических моделей для инверсии классов ОВ почвы. Однако разные каналы после различных математических изменений оказывают более существенный эффект на значение ОВ, что было использовано для улучшения результатов классификации. Корреляционный тест с использованием полевых данных является следующим шагом в построении конкретной числовой математической модели для прогнозирования ОВ почвы и выходит за рамки этого исследования.

В настоящей работе использовался корреляционный анализ Спирмена для скрининга ключевых спектральных каналов, который измеряет одномерную корреляцию между классом ОВ почвы и спектральными признаками. Однако высокая корреляция Спирмена не определяет высокий вклад канала в классификацию. Высокая корреляция может быть следствием сильной линейной корреляции данных, тогда как важность признаков, усиленных ГБ, отражает нелинейные эффекты. Например, канал может не иметь высокого коэффициента

корреляции, но он всё равно может играть важную роль в классификации в сочетании с другими каналами.

Выполнен статистический анализ эффективности ключевых каналов на разных уровнях значимости. Коэффициент корреляции канала В8 (832,9 нм, вторая производная) со значимостью  $p = 0,015$  составляет 0,036. Спектральный канал В6 (739,1 нм, обратная величина) имеет корреляцию 0,051, что указывает на его значительный вклад в инверсию содержания ОВ почвы. Среди каналов на уровне значимости  $p \leq 0,1$  выделяются канал В8А (864,0 нм, логарифм и вторая производная), где наблюдается некоторая положительная корреляция с коэффициентами корреляции  $-0,027$  и  $0,027$ , и канал В9 (945,0 нм, логарифм) с коэффициентом корреляции  $-0,025$ . Корреляция таких каналов, как В1, В2, В4 и т. д. в каждой математической форме вариации, была слабой и незначительной, с ограниченным вкладом в значение ОВ почвы.

Таким образом, каналы В6, В7 и В8 выбраны приоритетными в качестве основных входных переменных при инверсии классов содержания ОВ почвы. Каналы В8А (логарифм и вторая производная) и В9 (логарифм) использовались в качестве вспомогательных, а канал В3 (вторая производная) и В12 (вторая производная) были добавлены для дальнейшей оптимизации эффективности модели. В окончательном построении полученной модели градиентного бустинга, из-за выбора точек с применением случайной функции, средний результат пяти вычислений использовался в качестве контроля точности. Самая низкая точность составила 93,8 %, а средняя точность равна 96,2 %, при RMSE 0,31, коэффициенте Каппа 0,985. В табл. 2 приведена возможность применения метода градиентного бустинга для классификации классов содержания органического вещества почв Минской области Беларуси.

Количественная оценка ошибки классификации RMSE представляет собой размер средней ошибки между классами ОВ почвы, прогнозируемыми моделью, и истинными классами. Поскольку ОВ почвы является переменной порядковой классификации (пять уровней от низкого до высокого), RMSE можно интерпретировать как значимость ошибки прогнозирования и она характеризует следующие два основных аспекта: во-первых, количественную оценку ошибки классификации (размер средней ошибки между классами ОВ почвы, прогнозируемыми моделью, и истинными классами) и, во-вторых, оценку стабильности классификации. В исследовании модель имеет RMSE = 0,31 и точность классификации 96,2 %, что указывает на низкий уровень неправильной классификации и в значительной степени точные прогнозы уровней ОВ почвы. Однако из-за использования иерархических данных RMSE, равная 0,31 при пяти классификациях, означает, что существует большая ошибка межклассификации, и после использования её в качестве следующих трёх классификаций наблюдается значительное улучшение результатов из-за статистических характеристик классификации.

Для детальной оценки производительности модели по каждому классу содержания ОВ в табл. 3 представлены метрики классификации: F1-мера, точность (*англ.* precision) и полнота (*англ.* recall). Модель демонстрирует почти идеальную полноту ( $\text{recall} = 1,0$ ) для класса низкий, что исключает пропуск образцов с низким содержанием ОВ. Для класса высокий точность достигает 1,0, что означает отсутствие ложноположительных прогнозов. Все классы (от низкого до высокого) имеют F1-меру выше 0,985, подтверждая сбалансированную эффективность модели.

Таблица 3. Метрики классификации модели по уровням содержания ОВ

Класс	F1-мера	Точность	Полнота
Низкий	0,985	0,971	1,000
Низко-средний		0,990	
Средний			
Высоко-средний			
Высокий	0,985	1,000	0,970

### **Пространственный анализ классов содержания органического вещества почвы**

Пространственное распределение классов содержания ОВ по административным единицам Минской области демонстрирует значительную изменчивость на основе группировки ОВ по пяти классам (низкую, средне-низкую, среднюю, средне-высокую и высокую). Низкая степень (класс 1) доминирует в общем распределении со средней долей 31,59 % по всем районам. Примечательно, что Березинский район (46,29 %) и Борисовский район (41,43 %) демонстрируют самую высокую долю почв низкой степени содержания ОВ. Аналогичным образом оценены все районы области. Средняя доля средне-низкой степени (класс 2) в среднем для области составляет 20,23 %, а такие районы, как Несвижский (27,47 %) и Дзержинский (27,59 %), выделяются своими самыми высокими пропорциями. Средняя степень (класс 3) с общей долей 18,38 % показывает относительно сбалансированное распределение по региону. Минский район выделяется как примечательная область, где почвы 3-го класса составляют 26,34 %, что отражает умеренно благоприятные условия ОВ. Между тем такие регионы, как Слуцкий район (25,28 %), также демонстрируют значительную долю почв среднего класса. Напротив, степени средне-высокая (класс 4) и высокая (класс 5), которые представляют превосходное содержание ОВ, менее распространены, со средними долями 15,15 % и 14,64 % соответственно. Солигорский район имеет самую высокую долю почв высокого класса (25,84 %), что отражает его потенциал для устойчивой продуктивности почвы, в то время как такие районы, как Борисовский и Березинский, демонстрируют относительно низкую долю почв высокого класса, что подчёркивает пространственную изменчивость содержания ОВ в пределах региона. Недостаток ОВ определяется в первую очередь их генезисом и фракционным (гранулометрическим) составом: от минимального содержания в песчаных и супесчаных почвах до максимального — в иловато-суглинистых почвах.

Общий средний класс ОВ, рассчитываемый относительно классифицированных по значению содержания гумуса площадей почв, для Минской области составляет 2,61, при этом такие районы, как Солигорский (3,12) и Любанский (2,92), достигают значения класса выше среднего. С другой стороны, Березинский (2,31) и Стародорожский (2,43) районы характеризуются содержанием ОВ ниже среднего регионального значения. Стандартное отклонение значения классов ОВ по районам колеблется от 1,23 до 1,51, причём более высокие значения указывают на большую неоднородность качества почвы. Так, наибольшую изменчивость имеет Солигорский район со стандартным отклонением 1,51, а наиболее равномерное распределение содержания ОВ в почвах — Несвижский район со стандартным отклонением 1,23. Это также обусловлено неоднородностью геоморфологического фактора — преобладанием равнинного рельефа на сельскохозяйственных землях Несвижского района и искусственных горнодобывающих объектов в рельефе Солигорского района. В пространственном отношении сельскохозяйственные земли с более высоким содержанием ОВ в почвах расположены в юго-восточном регионе Минского района. Конкретные данные приведены на *рис. 3* (см. с. 130).

Что касается модели инверсии данных дистанционного зондирования на ключевых участках, территория была дифференцирована на три класса (высокое, среднее и низкое содержание ОВ), и та же группировка была сделана на основе полевых векторных данных почвы, как показано в *табл. 4* и на *рис. 4* (см. с. 131). Объединение крайних четырёх классов относительно среднего в два класса (высокий и низкий) обусловлено геостатистически. Сравнение результатов инверсии и векторных данных по пяти классам подчёркивает заметное сходство в распределении низких, средних и высоких классов. Для категории низкий оба набора данных демонстрируют согласованные пропорции, особенно в таких регионах, как Солигорский и Пуховичский районы, где доли различаются менее чем на 2 %. Аналогично, класс средний показывает сбалансированную тенденцию, и только незначительные отклонения наблюдаются по районам, что позволяет предположить, что модель инверсии эффективно фиксирует распределение классов среднего диапазона. В классе высокий результаты инверсии немного превышают пропорции по сравнению с векторными данными в таких районах, как Мядельский и Дзержинский; однако общие закономерности остаются близкими.

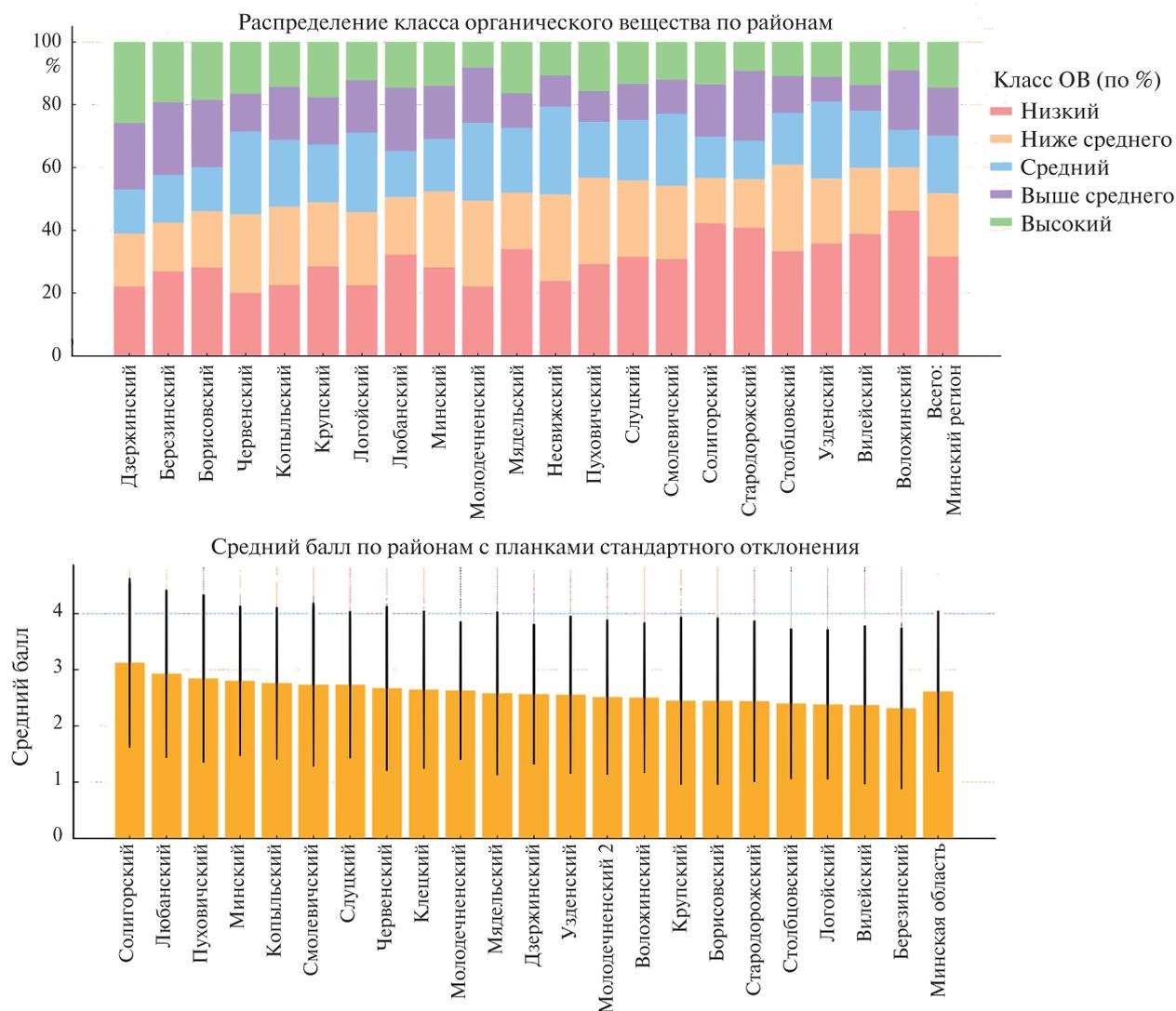


Рис. 3. Статистический анализ ОВ почвы в разрезе Минской области

Таблица 4. Сравнение результатов тройной классификации инверсивной модели и векторных данных

Район и метод	Низкий (класс 1), %	Средний (класс 2), %	Высокий (класс 3), %	Стандартное отклонение	Среднее	Разница
Солигорский (модель)	36,364	22,369	41,266	0,880	2,049	0,063
Солигорский (векторные данные)	37,557	26,299	36,144	0,858	1,986	
Пуховичский (модель)	43,771	20,804	35,425	0,886	1,917	0,073
Пуховичский (векторные данные)	45,576	24,511	29,912	0,855	1,843	
Мядельский (модель)	45,788	20,076	34,136	0,886	1,883	0,088
Мядельский (векторные данные)	46,835	26,813	26,353	0,831	1,795	
Дзержинский (модель)	47,132	27,876	24,992	0,820	1,779	0,053
Дзержинский (векторные данные)	46,022	35,437	18,542	0,755	1,725	
Крупский (модель)	42,893	26,575	30,533	0,848	1,876	0,193
Крупский (векторные данные)	48,963	33,715	17,322	0,750	1,684	

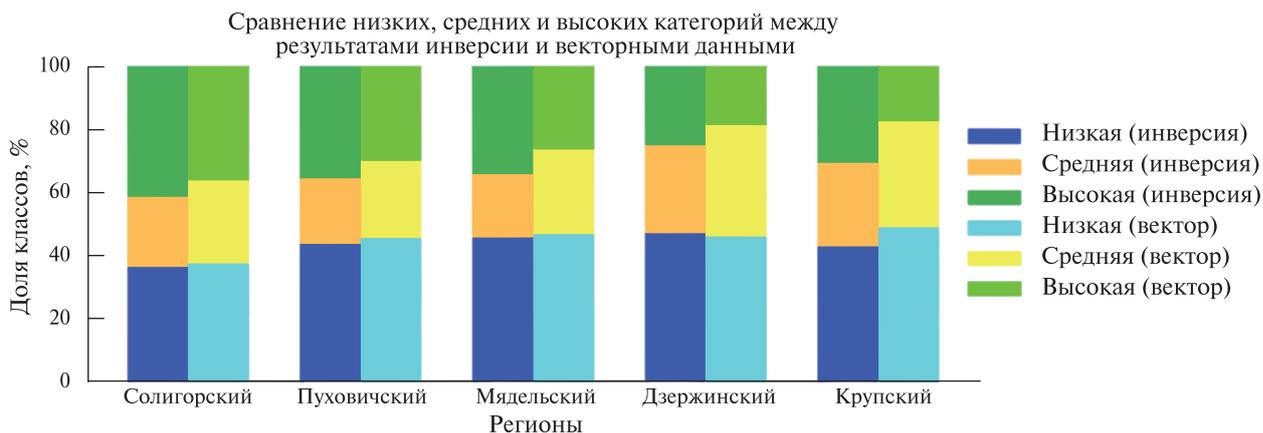


Рис. 4. Соотношение классов ОВ почвы между результатами инверсии и векторными данными по административным районам

### Заключение

В результате проведённых исследований спутниковые снимки Sentinel-2 в сочетании с данными полевого почвенного картирования были классифицированы с учётом оценки содержания органического вещества в почвенном покрове Минской области Беларуси. Используя ГБ вместе с математическими преобразованиями спектральных диапазонов, достигнута высокоточная классификация ОВ в пространстве. В частности, преобразования второй производной таких каналов, как B6, B7, B8 и B8A, показали самые высокие коэффициенты корреляции, эффективно фиксируя спектральные значения, связанные с содержанием ОВ почвы. Результаты выявили значительную пространственную изменчивость классов ОВ по всему региону. Такие районы, как Солигорский и Любанский, имели более высокое среднее содержание ОВ, что указывает на лучшее агрохимическое состояние почвы и её продуктивную способность, в то время как такие районы, как Березинский и Борисовский, отнесены к более низким классам, что требует специальных мероприятий по регулированию почвенного плодородия в агроландшафтах Минской области. Модель ГБ продемонстрировала надёжную производительность, достигнув средней точности 96,2 %, коэффициента Каппа 0,985 и среднеквадратичного отклонения 0,31, превзойдя такие альтернативные методы, как Random Forest и XGBoost, с точки зрения обработки несбалансированных данных и предоставления стабильных оценок важности признаков.

Сравнительный анализ модели инверсии данных дистанционного зондирования и полевых векторных данных почвы демонстрирует высокую степень согласованности в классификации ключевых участков на классы низкого, среднего и высокого ОВ в пяти репрезентативных районах. Модель инверсии эффективно фиксирует пространственную изменчивость и общие тенденции, отражённые в векторных данных, особенно в классах низкий и средний, где отклонения минимальны. Хотя в категории высокий наблюдаются небольшие переоценки, общее выравнивание усиливает надёжность модели инверсии. Эти результаты подтверждают полезность модели для точной пространственной классификации и подчёркивают её потенциал для более широкого применения в анализе содержания ОВ почвы на основе дистанционного зондирования.

Результаты этого исследования подчёркивают полезность интеграции изображений Sentinel-2 с передовыми алгоритмами машинного обучения для мониторинга и картирования ОВ. Этот подход обеспечивает надёжный и эффективный инструмент для оценки качества почвы в региональном масштабе, предлагая ценные идеи для устойчивого управления земельными ресурсами и сельскохозяйственного планирования в Минской области и за её пределами. Будущие исследования могут изучить включение слияния данных из нескольких источников для дальнейшего повышения надёжности модели и увеличения пространственного и временного разрешения картирования ОВ почвы.

## Литература

1. Клебанович Н. В., Прокопович С. Н., Харламова Е. В. Опыт составления почвенных карт Беларуси в международной системе WRB // Земля Беларуси. 2011. № 2. С. 41–47.
2. Лапа В. В., Черныш А. Ф., Азаренок Т. Н. и др. Атлас почв сельскохозяйственных земель Республики Беларусь. Минск: ИВЦ Минфина, 2017. 170 с.
3. Романова Т. А. Диагностика почв Беларуси и их классификация в системе ФАО-WRB. Минск: Ин-т почвоведения и агрохимии НАН Беларуси, 2004. 428 с.
4. Червань А. Н., Цыбулько Н. Н., Яцухно В. М. Методические подходы и практическое применение результатов оценки деградации земель/почв Беларуси // Изв. РАН. Сер. геогр. 2022. Т. 86. № 1. С. 55–68. DOI: 10.31857/S2587556622010058.
5. Ahmed Z., Iqbal J. Evaluation of Landsat TM5 multispectral data for automated mapping of surface soil texture and organic matter in GIS // European J. Remote Sensing. 2014. V. 47. No. 1. P. 557–573. DOI: 10.5721/EuJRS20144731.
6. Anantha Natarajan V., Sunil Kumar M., Tamizhazhagan V., Chevumoi R. M. Prediction of soil pH from remote sensing data using gradient boosted regression analysis // J. Pharmaceutical Negative Results. 2022. V. 13. Spec. Iss. 6. P. 29–36. DOI: 10.47750/pnr.2022.13.S06.005.
7. Castaldi F., Chabrilat S., Don A., van Wesemael B. Soil organic carbon mapping using LUCAS topsoil database and Sentinel-2 data: An approach to reduce soil moisture and crop residue effects // Remote Sensing. 2019. V. 11. No. 18. Article 2121. DOI: 10.3390/rs11182121.
8. Chen D., Chang N., Xiao J. et al. Mapping dynamics of soil organic matter in croplands with MODIS data and machine learning algorithms // Science of the Total Environment. 2019. V. 669. P. 844–855. DOI: 10.1016/j.scitotenv.2019.03.151.
9. Demir S., Sahin E. K. An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost // Neural Computing and Applications. 2023. V. 35. No. 4. P. 3173–3190. DOI: 10.1007/s00521-022-07856-4.
10. Diek S., Fornallaz F., Schaepman M. E., De Jong R. Barest Pixel Composite for agricultural areas using Landsat time series // Remote Sensing. 2017. V. 9. No. 12. Article 1245. DOI: 10.3390/rs9121245.
11. Guo Y. K., Zeng F., Ding M. Q. et al. Research on SPOT-5 image-based soil organic matter content estimation // Applied Mechanics and Materials. 2013. V. 409–410. P. 246–251. DOI: 10.4028/www.scientific.net/AMM.409-410.246.
12. Habibi L. N., Komariah K., Ariyanto D. P. et al. Estimation of soil organic matter on paddy field using remote sensing method // SAINS TANAH — J. Soil Science and Agroclimatology. 2019. V. 16. No. 2. P. 159–168. DOI: 10.20961/stjssa.v16i2.35395.
13. Kang T.-H., Choi S.-W., Lee C., Chang S.-H. Soil classification by machine learning using a tunnel boring machine's operating parameters // Applied Sciences. 2022. V. 12. No. 22. Article 11480. DOI: 10.3390/app122211480.
14. Li X., Ding J., Liu J. et al. Digital mapping of soil organic carbon using Sentinel series data: A case study of the Ebinur Lake watershed in Xinjiang // Remote Sensing. 2021. V. 13. No. 4. Article 769. DOI: 10.3390/rs13040769.
15. Liu H. J., Zhang M. W., Yang H. X. et al. Inversion of cultivated soil organic matter content combining multi-spectral remote sensing and random forest algorithm // Trans. Chinese Society of Agricultural Engineering. 2020. V. 36. No. 10. P. 134–140. DOI: 10.11975/j.issn.1002-6819.2020.10.016.
16. Pan H., Chen Z., de Wit A., Ren J. Joint assimilation of leaf area index and soil moisture from Sentinel-1 and Sentinel-2 data into the WOFOST model for winter wheat yield estimation // Sensors. 2019. V. 19. No. 14. Article 3161. DOI: 10.3390/s19143161.
17. Sahin E. K. Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping // Geocarto Intern. 2022. V. 37. No. 9. P. 2441–2465. DOI: 10.1080/10106049.2020.1831623.
18. Tang S., Du C., Nie T. Inversion estimation of soil organic matter in Songnen Plain based on multispectral analysis // Land. 2022. V. 11. No. 5. Article 608. DOI: 10.3390/land11050608.
19. Tziachris P., Aschonitis V., Chatzistathis T., Papadopoulou M. Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters // CATENA. 2019. V. 174. P. 206–216. DOI: 10.1016/j.catena.2018.11.010.
20. Vaudour E., Gomez C., Loiseau T. et al. The impact of acquisition date on the prediction performance of topsoil organic carbon from Sentinel-2 for croplands // Remote Sensing. 2019. V. 11. No. 18. Article 2143. DOI: 10.3390/rs11182143.
21. Wang X., Han J., Wang X. et al. Estimating soil organic matter content using Sentinel-2 imagery by machine learning in Shanghai // IEEE Access. 2021. V. 9. P. 78215–78225. DOI: 10.1109/ACCESS.2021.3080689.
22. Yang R., Rossiter D. G., Liu F. et al. Predictive mapping of topsoil organic carbon in an Alpine environment aided by Landsat TM // PLOS One. 2015. V. 10. No. 10. Article e0139042. DOI: 10.1371/journal.pone.0139042.

23. Ye Z., Sheng Z., Liu X. et al. Using machine learning algorithms based on GF-6 and Google Earth Engine to predict and map the spatial distribution of soil organic matter content // Sustainability. 2021. V. 13. No. 24. Article 14055. DOI: 10.3390/su132414055.
24. Yin H., Chen C., He Y. et al. Synergistic estimation of soil salinity based on Sentinel-1 image texture and Sentinel-2 salinity spectral indices // J. Applied Remote Sensing. 2023. V. 17. No. 1. Article 018502. DOI: 10.1117/1.JRS.17.018502.
25. Zhang L., Wang L. Optimization of site investigation program for reliability assessment of undrained slope using Spearman rank correlation coefficient // Computers and Geotechnics. 2023. V. 155. Article 105208. DOI: 10.1016/j.compgeo.2022.105208.
26. Zhang M., Zhang M., Yang H. et al. Mapping regional soil organic matter based on Sentinel-2A and MODIS imagery using machine learning algorithms and Google Earth Engine // Remote Sensing. 2021. V. 13. No. 15. Article 2934. DOI: 10.3390/rs13152934.
27. Zhou T., Geng Y., Chen J. et al. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms // Science of the Total Environment. 2020. V. 729. Article 138244. DOI: 10.1016/j.scitotenv.2020.138244.
28. Zhou T., Geng Y., Ji C. et al. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images // Science of the Total Environment. 2021. V. 755. Article 142661. DOI: 10.1016/j.scitotenv.2020.142661.

## Determination of soil organic matter content in Minsk Region of Belarus using gradient boosting classifiers based on satellite data

A. N. Chervan, B. Zhao

*Belarusian State University, Minsk 220030, Belarus*  
*E-mails: chervanalex@mail.ru, geozhao@outlook.com*

One of key indicators of soil degradation processes in agricultural lands is the loss of soil organic matter (SOM). Therefore, methodologies for monitoring the spatial and temporal distribution of SOM are critically important from both ecological and economic perspectives. Currently, remote sensing data, particularly Sentinel-2 satellite imagery, can be used to estimate SOM content in the surface humus-accumulative horizon of arable soils. Using Sentinel-2 data and field soil mapping data from Belarus, the study aimed to analyze the spatial accuracy of SOM content estimation and the results of automated satellite image interpretation, with a focus on Minsk Region of Belarus. A range of analytical methods was employed, including selection of representative spectral bands as model inputs using Spearman's correlation analysis and gradient boosting classifiers for GIS-based modeling. The proposed approach was evaluated in terms of achieving high-precision and rapid inversion, as well as spatial analysis of SOM content across genetic soil types. The inversion model accuracy was validated using an independent database. Results of the spatial analysis demonstrated that the model, based on Sentinel-2 imagery in bands B6, B7, B8, B8A and B12, most effectively utilized second-order derivatives. The inversion model achieved the highest accuracy (minimum 93.8 %, average 96.2 %), with an RMSE of 0.31 and a Kappa coefficient of 0.985. The lowest SOM content in arable soils was observed in the Berezino District, whereas the highest SOM levels were found in the Lyuban and Soligorsk districts that are characterized by a significant proportion of hydromeliorated agricultural lands.

**Keywords:** remote sensing, Sentinel-2, soil organic matter content, remote sensing data, gradient boosting classifier

Accepted: 17.03.2025

DOI: 10.21046/2070-7401-2025-22-3-121-135

## References

1. Klebanovich N. V., Prokopovich S. N., Kharlamova E. V., Experience of compiling soil maps of Belarus in the international WRB system, *Zemlya Belarusi*, 2011, No. 2, pp. 41–47 (in Russian).
2. Lapa V. V., Chernysh A. F., Azarenok T. N. et al., *Atlas pochv sel'skokhozyastvennykh zemel' Respubliki Belarus'* (Atlas of agricultural soils of the Republic of Belarus), Minsk: IVC Minfina, 2017, 170 p. (in Russian).
3. Romanova T. A., *Diagnostika pochv Belarusi i ikh klassifikatsiya v sisteme FAO-WRB* (Diagnostics of soils of Belarus and their classification in the FAO-WRB system), Minsk: Institute of Soil Science and Agrochemistry of the National Academy of Sciences of Belarus, 2004, 428 p. (in Russian).
4. Chervan A. N., Tsybulko N. N., Yatsukhno V. M., Methodological approaches and practical application of the results of land/soil degradation assessment in Belarus, *Izvestiya Rossiiskoi akademii nauk. Ser. geographicheskaya*, V. 86, No. 1, pp. 55–68 (in Russian), DOI: 10.31857/S2587556622010058.
5. Ahmed Z., Iqbal J., Evaluation of Landsat TM5 multispectral data for automated mapping of surface soil texture and organic matter in GIS, *European J. Remote Sensing*, 2014, V. 47, No. 1, pp. 557–573, DOI: 10.5721/EuJRS20144731.
6. Anantha Natarajan V., Sunil Kumar M., Tamizhazhagan V., Chevumoi R. M., Prediction of soil pH from remote sensing data using gradient boosted regression analysis, *J. Pharmaceutical Negative Results*, 2022, V. 13, Spec. Iss. 6, pp. 29–36, DOI: 10.47750/pnr.2022.13.S06.005.
7. Castaldi F., Chabrillat S., Don A., van Wesemael B., Soil organic carbon mapping using LUCAS topsoil database and Sentinel-2 data: An approach to reduce soil moisture and crop residue effects, *Remote Sensing*, 2019, V. 11, No. 18, Article 2121, DOI: 10.3390/rs11182121.
8. Chen D., Chang N., Xiao J. et al., Mapping dynamics of soil organic matter in croplands with MODIS data and machine learning algorithms, *Science of the Total Environment*, 2019, V. 669, P. 844–855, DOI: 10.1016/j.scitotenv.2019.03.151.
9. Demir S., Sahin E. K., An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost, *Neural Computing and Applications*, 2023, V. 35, No. 4, pp. 3173–3190, DOI: 10.1007/s00521-022-07856-4.
10. Diek S., Fornallaz F., Schaepman M. E., De Jong R., Barest Pixel Composite for agricultural areas using Landsat time series, *Remote Sensing*, 2017, V. 9, No. 12, Article 1245, DOI: 10.3390/rs9121245.
11. Guo Y. K., Zeng F., Ding M. Q. et al., Research on SPOT-5 image-based soil organic matter content estimation, *Applied Mechanics and Materials*, 2013, V. 409–410, pp. 246–251, DOI: 10.4028/www.scientific.net/AMM.409-410.246.
12. Habibi L. N., Komariah K., Ariyanto D. P. et al., Estimation of soil organic matter on paddy field using remote sensing method, *SAINS TANAH — J. Soil Science and Agroclimatology*, 2019, V. 16, No. 2, pp. 159–168, DOI: 10.20961/stjssa.v16i2.35395.
13. Kang T.-H., Choi S.-W., Lee C., Chang S.-H., Soil classification by machine learning using a tunnel boring machine's operating parameters, *Applied Sciences*, 2022, V. 12, No. 22, Article 11480, DOI: 10.3390/app122211480.
14. Li X., Ding J., Liu J. et al., Digital mapping of soil organic carbon using Sentinel series data: A case study of the Ebinur Lake watershed in Xinjiang, *Remote Sensing*, 2021, V. 13, No. 4, Article 769, DOI: 10.3390/rs13040769.
15. Liu H. J., Zhang M. W., Yang H. X. et al., Inversion of cultivated soil organic matter content combining multi-spectral remote sensing and random forest algorithm, *Trans. Chinese Society of Agricultural Engineering*, 2020, V. 36, No. 10, pp. 134–140, DOI: 10.11975/j.issn.1002-6819.2020.10.016.
16. Pan H., Chen Z., de Wit A., Ren J., Joint assimilation of leaf area index and soil moisture from Sentinel-1 and Sentinel-2 data into the WOFOST model for winter wheat yield estimation, *Sensors*, 2019, V. 19, No. 14, Article 3161, DOI: 10.3390/s19143161.
17. Sahin E. K., Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping, *Geocarto Intern.*, 2022, V. 37, No. 9, pp. 2441–2465, DOI: 10.1080/10106049.2020.1831623.
18. Tang S., Du C., Nie T., Inversion estimation of soil organic matter in Songnen Plain based on multispectral analysis, *Land*, 2022, V. 11, No. 5, Article 608, DOI: 10.3390/land11050608.
19. Tziachris P., Aschonitis V., Chatzistathis T., Papadopoulou M., Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters, *CATENA*, 2019, V. 174, pp. 206–216, DOI: 10.1016/j.catena.2018.11.010.
20. Vaudour E., Gomez C., Loiseau T. et al., The impact of acquisition date on the prediction performance of topsoil organic carbon from Sentinel-2 for croplands, *Remote Sensing*, 2019, V. 11, No. 18, Article 2143, DOI: 10.3390/rs11182143.
21. Wang X., Han J., Wang X. et al., Estimating soil organic matter content using Sentinel-2 imagery by machine learning in Shanghai, *IEEE Access*, 2021, V. 9, pp. 78215–78225, DOI: 10.1109/ACCESS.2021.3080689.

22. Yang R., Rossiter D. G., Liu F. et al., Predictive mapping of topsoil organic carbon in an Alpine environment aided by Landsat TM, *PLOS One*, 2015, V. 10, No. 10, Article e0139042, DOI: 10.1371/journal.pone.0139042.
23. Ye Z., Sheng Z., Liu X. et al., Using machine learning algorithms based on GF-6 and Google Earth Engine to predict and map the spatial distribution of soil organic matter content, *Sustainability*, 2021, V. 13, No. 24, Article 14055, DOI: 10.3390/su132414055.
24. Yin H., Chen C., He Y. et al., Synergistic estimation of soil salinity based on Sentinel-1 image texture and Sentinel-2 salinity spectral indices, *J. Applied Remote Sensing*, 2023, V. 17, No. 1, Article 018502, DOI: 10.1117/1.JRS.17.018502.
25. Zhang L., Wang L., Optimization of site investigation program for reliability assessment of undrained slope using Spearman rank correlation coefficient, *Computers and Geotechnics*, 2023, V. 155, Article 105208, DOI: 10.1016/j.compgeo.2022.105208.
26. Zhang M., Zhang M., Yang H. et al., Mapping regional soil organic matter based on Sentinel-2A and MODIS imagery using machine learning algorithms and Google Earth Engine, *Remote Sensing*, 2021, V. 13, No. 15, Article 2934, DOI: 10.3390/rs13152934.
27. Zhou T., Geng Y., Chen J. et al., High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms, *Science of the Total Environment*, 2020, V. 729, Article 138244, DOI: 10.1016/j.scitotenv.2020.138244.
28. Zhou T., Geng Y., Ji C. et al., Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images, *Science of the Total Environment*, 2021, V. 755, Article 142661, DOI: 10.1016/j.scitotenv.2020.142661.