УДК 631.417.2:631.46:574.4:004.94 *https://doi.org/10.31861/biosystems2026.02.358*

МОНІТОРИНГ ДИНАМІКИ ВМІСТУ ГУМУСУ В ҐРУНТАХ АГРОЦЕНОЗІВ ЗА ДАНИМИ СУПУТНИКОВОГО СПОСТЕРЕЖЕННЯ

Д. І. РОБУЛЯК[1], І. С. СМАГА[1], М. В. ГУНЧАК[2]

[1] *Чернівецький національний університет імені Юрія Федьковича,*
вул. Коцюбинського, 2, м. Чернівці, 58022, Україна
robuliak.dmytro@chnu.edu.ua, i.smaga@chnu.edu.ua
[2] *Чернівецька філія державної установи «Інститут охорони ґрунтів України»*
58029, вул. Героїв Майдану, 194-А, м. Чернівці, Україна
Українська науково-дослідна станція карантину рослин ІЗР НААН
60321, вул. Наукова, 4, с. Бояни, Чернівецька область, Україна
chernivtsy_grunt@ukr.net

Гумус утворюється внаслідок мікробної трансформації рослинних решток, є основним депо органічного вуглецю в агроекосистемах, трофічним та енергетичним базисом функціонування ґрунтової біоти. Вміст гумусу – консервативний показник гумусного стану ґрунтів, багаторічна динаміка якого відображає баланс процесів біогенного надходження та мікробної мінералізації органічної речовини, а пошук способів її дистанційної оцінки є актуальним завданням моніторингу ґрунтів. Мета дослідження – прослідкувати динаміку вмісту гумусу в ґрунтах агроекосистем колишнього Глибоцького району Чернівецької області за 2002–2017 рр. і перевірити можливість визначення вмісту гумусу в ґрунтах окремих полів з використанням спектральних ознак «оголеного» ґрунту за даними знімків Landsat і Sentinel-2. Використано результати агрохімічного обстеження ґрунтового покриву 232 полів, обстежених в чотирьох турах (2002, 2007, 2012, 2017 рр.), кожне з яких представлене однією опорною точкою. Межі полів виділено в середовищі Google Earth Engine за трисезонними композитами Sentinel-2 2017 р. (рік XI туру) шляхом SLIC-сегментації та нарощування області від опорної точки з вибором межі в діапазоні ±15 % облікової площі поля; параметри підбирали на 164 калібрувальних полях і перевіряли на 68 контрольних, геометричну точність оцінено за вручну оцифрованими еталонними контурами 40 контрольних полів того самого року. Для кожної з 8 комбінацій «сенсор-тур» побудовано медіанні композити «оголеного» ґрунту за даними Landsat 5/7/8 і Sentinel-2 у вікні ±1 рік навколо базового року туру обстеження з відбором пікселів за NDVI і NBR2 (протокол GEOS3) та сезонним обмеженням; моделі Ridge (з підбором коефіцієнта регуляризації у вкладеній перехресній перевірці) і Random Forest на дев'яти спектральних ознаках (шість каналів, зокрема SWIR2, та індекси NDVI, BSI, NBR2), розрахованих у полігонах, звужених на 20 м), оцінено випадковою та просторовою 5-кратною перехресною перевіркою. Медіанний індекс Intersection over Union (IoU) фінальних полігонів становив 0,44 (95 % ДІ 0,34–0,59; IoU ≥ 0,7 – у 7 з 40 полів), вільного нарощування – 0,33, базової сегментації SNIC – 0,22, тобто полігони придатні лише як наближені контури полів.

Середній вміст гумусу в ґрунтах досліджуваної території знизився з 2,13 % у 2002 р. до 2,00 % у 2017 р. (різниця −0,131 %, 95 % ДІ −0,150…−0,112 %; тест Фрідмана $p < 0,001$) за вузького діапазону значень (1,4–2,9 %, стандартне відхилення 0,20–0,22 %, коефіцієнт варіації 9,5–11,0 %), характерного для фонових у районі дерново-підзолистих і сірих лісових ґрунтів; міжпольова мінливість вмісту гумусу за величиною близька до допустимої похибки його лабораторного визначення. Коефіцієнт детермінації за перехресної перевірки не перевищував 0,055 (за випадкової перевірки Ridge – від −0,067 до 0,016, Random Forest – від −0,107 до 0,055), а зниження RMSE порівняно з прогнозом середнім значенням не перевищувало 0,010 %, тобто було в кілька разів меншим за похибку лабораторного методу. За наведеними показниками прогностична здатність спектральних ознак на рівні поля не підтверджується. Збіг середніх прогнозованих і лабораторних значень вмісту гумусу за турами обстеження тривіальний: його так само відтворює наївна модель, а зміни на рівні окремих полів дані моделі не відтворюють. Результати апробації дистанційної оцінки вмісту гумусу дозволили виявити обмеження індексного підходу за використаних даних і діапазону вмісту гумусу, які зумовлені вузькою міжпольовою варіацією його вмісту, сумірною з похибкою лабораторного методу. Для моніторингу гумусного стану ґрунтів дистанційні дані в такій постановці можуть мати лише допоміжне значення, а лабораторні визначення залишаються незамінними.

Ключові слова: гумус ґрунту, гумусний стан, агроекосистеми, дистанційне зондування Землі, Sentinel-2, Landsat, оголений ґрунт, виділення меж полів, Google Earth Engine, спектральні індекси, перехресна перевірка

Вступ. Ефективне управління земельними ресурсами сільськогосподарського призначення потребує актуальної інформації щодо просторового поширення меж окремих полів та об'єктивних показників агроекологічного стану й родючості ґрунту, особливо вмісту гумусу.

Традиційні методи інвентаризації земель, які передбачають встановлення меж полів на основі наземного знімання та ручної векторизації є трудомісткими й економічно затратними у випадку обстеження великої кількості об'єктів на певній території. Визначення вмісту гумусу в ґрунтах вимагає систематичного відбору й лабораторного аналізу ґрунтових зразків, що обмежує періодичність і просторову щільність спостережень.

З біологічного погляду гумус є не лише показником родючості ґрунту, а й ключовим компонентом агроекосистеми. Органічна речовина ґрунту утворюється внаслідок розкладання та трансформації рослинних решток ґрунтовою біотою. Сучасні уявлення розглядають її як континуум органічних сполук на різних стадіях мікробного перероблення, стабільність яких визначається насамперед фізичним захистом в агрегатах і взаємодією з мінеральними поверхнями (Lehmann, Kleber, 2015). Значна частка стабільної органічної речовини має мікробне походження – це некромаса та продукти метаболізму мікроорганізмів, тобто гумусоутворення є біохімічним процесом, у якому ключову роль відіграє мікробний анаболізм (Liang et al., 2017). Водночас гумус є матрицею, джерелом енергії та елементів живлення для мікробних угруповань і ґрунтової фауни, а діяльність дощових черв'яків та інших педобіонтів сприяє формуванню гумусових горизонтів і накопиченню вуглецю, що встановлено в чорноземах (Yakovenko et al., 2025). Ґрунти є найбільшим наземним депо органічного вуглецю, тому зменшення вмісту гумусу в агроекосистемах означає не тільки втрату родючості, а й вивільнення вуглецю в атмосферу та зниження екосистемних послуг ґрунту (Wiesmeier et al., 2019). Валовий вміст гумусу як консервативний пул із часом оновлення в десятки–сотні років інтегрує багаторічний баланс надходження рослинних решток, їх мікробної трансформації та мінералізації. Тому його багаторічна динаміка характеризує стан трофічної та енергетичної бази педобіоценозів агроекосистем, а стійке зниження вмісту гумусу свідчить про деградацію цієї бази.

Розвиток хмарних платформ геопросторового аналізу, зокрема Google Earth Engine (GEE; Gorelick et al., 2017), та можливість вільного доступу до архівів багатоспектральних супутникових знімків Sentinel-2 і Landsat відкривають можливості для автоматизації зазначених процесів (Gayibov, 2025). Алгоритми сегментації зображень, такі як SLIC (Simple Linear Iterative Clustering; Achanta et al., 2012) і SNIC (Simple Non-Iterative Clustering; Achanta, Süsstrunk, 2017), дозволяють виділяти контури однорідних ділянок земної поверхні за спектральними ознаками без залучення додаткових картографічних джерел, а часові ряди супутникових знімків – відстежувати динаміку стану ґрунтового покриву. У період «оголеного» ґрунту спектральна відбивна здатність є найбільш чутливою до властивостей верхнього шару ґрунту, зокрема вмісту гумусу (органічної речовини ґрунту). Для використання засобів дистанційного зондування ґрунтового покриву необхідно вивчити характер взаємозв'язків властивостей ґрунтів з їх спектральними характеристиками. Це дасть змогу екстраполювати отримані результати на території ландшафтів з однотипними ґрунтами. Встановлена залежність спектральних характеристик ґрунту від гранулометричного складу та вмісту гумусу (Шатохін та ін., 1996) дала підстави запропонувати технологію аерокосмічного контролю за гумусним станом ґрунтів (Шатохін та ін., 1999) та оцінювати диференціацію ґрунтового покриву (Биндич, 2019).

Відбивна здатність поверхні ґрунту значною мірою визначається властивостями, обумовленими біологічними чинниками. Гумусові речовини, що утворюються внаслідок мікробної трансформації рослинних решток, зумовлюють темне забарвлення ґрунту та, як наслідок зниження відбиття у видимому й ближньому інфрачервоному діапазонах; агрегатна структура, яка формується за участі коренів, грибного міцелію, мікробних метаболітів і дощових черв'яків, визначає шорсткість поверхні та розсіювання випромінювання; вологість ґрунту, пов'язана зі структурою та вмістом органічної речовини, знижує відбиття в усьому спектрі, особливо в короткохвильовому інфрачервоному діапазоні (Stenberg et al., 2010; Angelopoulou et al., 2019). Таким чином, спектр «оголеного» ґрунту опосередковано відображає довготривалі результати біогенного ґрунтоутворення, але водночас відображає й вплив вологості, шорсткості поверхні ґрунту після обробітку та рослинних решток, що ускладнює виокремлення власне сигналу гумусу за супутниковими даними (Demattê et al., 2018). Для зменшення цих впливів у протоколі GEOS3 поряд із порогом NDVI застосовують поріг індексу NBR2 (на основі каналів SWIR1 і SWIR2) для відсікання нефотосинтезуючої рослинності (стерні, сухих решток) (Demattê et al., 2018; Safanelli et al., 2020).

Водночас, застосовність дистанційних

методів оцінки вмісту гумусу залежить від контрастності ґрунтового покриву досліджуваної території. Якщо вона охоплює один ґрунтово-кліматичний район із відносно однорідним набором ґрунтових відмін (наприклад, дерново-підзолисті та сірі лісові ґрунти Передкарпаття), варіативність вмісту гумусу в них може бути невисокою, що ускладнює надійне розрізнення діапазонів варіації за допомогою супутникових даних із просторовою роздільною здатністю 10–30 м. В прикладних дослідженнях, зазвичай, наводяться лише успішні випадки застосування регресійних моделей, а зазначене методологічне обмеження не розглядається (Castaldi et al., 2019; Vaudour et al., 2019).

Мета дослідження – прослідкувати динаміку вмісту гумусу в ґрунтах агроекосистем колишнього Глибоцького району Чернівецької області за 2002–2017 рр. і перевірити можливість визначення вмісту гумусу в ґрунтах окремих полів з використанням спектральних ознак «оголеного» ґрунту за даними знімків Landsat і Sentinel-2.

Матеріали та методи. Для зіставлення лабораторних і супутникових даних апробовано автоматизоване виділення меж полів навколо поодиноких опорних точок обстеження. В дослідженні використано матеріали агрохімічного обстеження ґрунтового покриву сільськогосподарських полів, кожне з яких представлене однією опорною точкою (внутрішньою точкою поля), у межах колишнього (до 2020 р.) Глибоцького адміністративного району Чернівецької області (рис. 1). Із 447 полів, для яких наявні дані хоча б одного туру обстежень, до аналізу залучено 232 поля з отриманими значеннями вмісту гумусу в усіх чотирьох турах. Така панельна вибірка дає змогу отримати динаміку показника на одній і тій самій сукупності полів. До орних земель в Прут-Сіретському межиріччі Буковинського Передкарпаття залучено, переважно, дерново-підзолисті поверхнево-оглеєні, сірі й темно-сірі опідзолені ґрунти з вмістом гумусу в гумусовому горизонті близько 2 %.

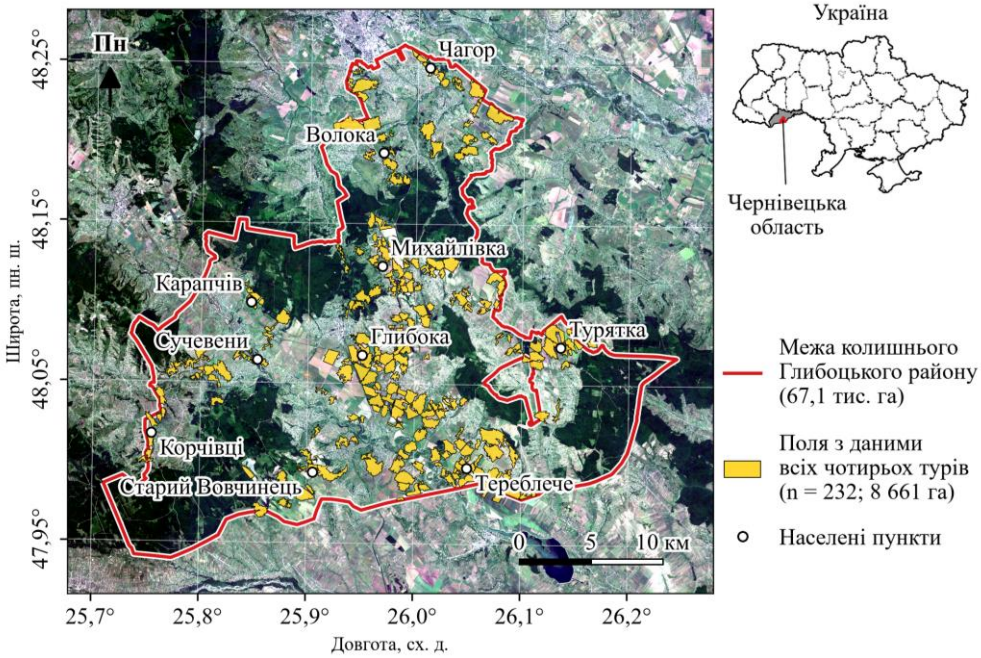


Рис. 1. Територія дослідження: колишній Глибоцький район Чернівецької області та 232 поля з даними всіх чотирьох турів агрохімічного обстеження (підкладка – медіанний композит Sentinel-2 L1C за 01.05–15.09.2017 р., RGB; межа району – об'єднання 22 сільських і селищної рад, geoBoundaries ADM3; Runfola et al., 2020)

Fig. 1. Study area: the former Hlyboka district, Chernivtsi region, and 232 fields with data from all four agrochemical survey rounds (background – Sentinel-2 L1C median composite, 1 May – 15 September 2017, RGB; district boundary – union of 22 village and settlement councils, geoBoundaries ADM3; Runfola et al., 2020)

Для кожної точки в атрибутивній таблиці зафіксовано ідентифікатор поля (ID), назву населеного пункту (City_name), площу (Square_H, га) та вміст гумусу (%) за результатами чотирьох турів агрохімічного обстеження ґрунтів (VIII тур, 2002 р.; IX тур, 2007 р.; X тур, 2012 р.; XI тур, 2017 р.). Для всіх 232 полів вибірки наявні значення в кожному з чотирьох турів; у повному масиві з 447 полів кількість визначень становила 253, 384, 396 та 444 відповідно, оскільки в різні тури обстежували різну кількість полів.

Відбір і аналіз ґрунтових зразків виконано в

межах планових турів агрохімічного обстеження співробітниками Чернівецької філії ДУ «Інститут охорони ґрунтів України». Зразки відбирали з орного шару (0–20 см); вміст гумусу визначали за методом Тюріна (ДСТУ 4289:2004). Допустима розбіжність між паралельними визначеннями за цим методом за вмісту гумусу близько 2 % становить орієнтовно 0,15–0,20 %. Вміст гумусу в базі даних наведено з точністю 0,1 %.

Виділення меж полів. Межі полів виділено в середовищі Google Earth Engine (Gorelick et al., 2017) за знімками 2017 року (XI тур обстеження) за такою процедурою: 1) з колекції Sentinel-2 (COPERNICUS/S2_HARMONIZED, рівень обробки L1C, оскільки продукти L2A для території дослідження за 2017 р. доступні не повністю) сформовано медіанні композити трьох сезонів 2017 р. – весняного (квітень–травень; 13 сцен), літнього (червень–липень; 18) та осіннього (15 серпня – жовтень; 30) – з маскуванням хмар за Cloud Score+ (cs_cdf ≥ 0,6; Pasquarella et al., 2023); медіанна кількість безхмарних спостережень на піксель становила 8, 9 і 12 відповідно; 2) для кожного сезону розраховано відбивну здатність у каналах B2, B3, B4, B8, B11, B12, нормалізований вегетаційний індекс NDVI = (B8 − B4) / (B8 + B4) та індекс оголеного ґрунту BSI = [(B11 + B4) − (B8 + B2)] / [(B11 + B4) + (B8 + B2)] (Rikimaru et al., 2002) – разом 24 ознаки; 3) ліси, забудовані землі та водні об’єкти (класи 10, 50 і 80 ESA WorldCover 2020, колекція ESA/WorldCover/v100) використано як бар’єри, через які межа поля не може проходити; у радіусі 20 м навколо опорної точки бар’єр знімали, щоб точка, яка потрапила на піксель лісосмуги чи деревної рослинності на краю поля, не зміщувала початок нарощування на сусідній суперпіксель; 4) багатосезонний стек сегментовано алгоритмом SLIC (Achanta et al., 2012) на суперпікселі площею близько 1 га (compactness = 0,05, з умовою зв’язності); 5) від суперпікселя, що містить опорну точку, область нарощували приєднанням суміжних суперпікселів у порядку зростання «вартості» (середньоквадратична спектральна відстань + w × контраст межі) в межах комірки діаграми Вороного, побудованої для всіх 447 опорних точок, щоб полігони сусідніх полів не перетиналися; 6) зупинку нарощування визначали в діапазоні ±15 % облікової площі поля Square_H – на кроці з найсильнішою межею (фінальний полігон); паралельно розраховано варіант вільного нарощування, зупинка якого визначалася лише порогом контрасту межі t без використання Square_H (w = 2,0, t = 3,35 – значення з найменшою медіанною відносною похибкою площі на калібрувальних полях); 7) полігони згладжено морфологічним закриттям (радіус 1 піксель), заповнено внутрішні пустоти і спрощено з допуском 5 м; якщо після спрощення опорна точка опинялася поза полігоном, зберігали неспрощений контур. Як базовий метод порівняння (односезонна суперпіксельна сегментація без обмеження площі) використано полігони, отримані сегментацією SNIC (Achanta, Süsstrunk, 2017) комбінованого шару 2 × NDVI + 2 × BSI медіанного композиту Sentinel-2 L1C за травень 2017 р. (5 сцен з хмарністю < 10 %; size = 30, compactness = 0,6, connectivity = 8, neighborhoodSize = 64) з векторизаційною функцією reduceToVectors (scale = 10 м) і відбором полігонів, що містять опорні точки; оскільки проєкцію обробки явно не задавали, сегментацію виконано в сітці EPSG:4326 (піксель ≈ 6,7 × 10,0 м), тому розмір полігонів визначався розміром суперпікселя, а не розміром поля.

Параметри процедури підбирали на калібрувальній частині вибірки (164 поля), а 68 контрольних полів, відібраних випадково, використовували лише для перевірки. Площі полігонів обчислено в проєкції WGS 84 / UTM 35N (EPSG:32635). Сумарна площа 232 фінальних полігонів становить 8660,9 га (сума облікових площ – 8900,3 га), медіанна – 30,2 га. Межі виділено за знімками року XI туру (2017 р.); оскільки для турів 2002–2012 рр. конфігурація полів могла відрізнятися, спектральні ознаки розраховували у внутрішній частині полігонів (див. нижче).

Оцінка якості виділення меж полів. Якість меж оцінено прямо – за еталонними контурами – та непрямо. Для прямої оцінки з 68 контрольних полів відібрано 40 (стратифікований випадковий відбір, по 10 полів у кожному квартилі облікової площі; 4,1–128 га, 16 населених пунктів), межі яких оцифровано вручну в Google Earth Engine Code Editor за тими самими сезонними композитами Sentinel-2 2017 р. без відображення автоматизованих полігонів. Для кожного поля розраховано індекс Intersection over Union (IoU – відношення площі перетину полігона й еталона до площі їх об’єднання), коефіцієнт Дайса, граничний F1 (середнє гармонічне часток межі полігона та межі еталона, що лежать не далі 10 або 20 м від іншої межі), частку еталона поза полігоном, частку полігона поза еталоном та відстань між центроїдами; 95 % довірчі інтервали медіан отримано бутстрепом (5000 повторів), відмінності між методами перевіряли критерієм Вілкоксона. Непряму оцінку для всіх 232 полів виконано: а) за положенням опорних точок (власна точка всередині полігона; наявність у полігоні інших опорних точок із повного масиву

447 полів як ознака об'єднання суміжних полів); б) за взаємним перекриттям полігонів; в) за відношенням площі полігона до облікової площі поля Square_H – медіаною відношення, медіанною абсолютною відносною похибкою площі (MdAPE = медіана |площа / Square_H − 1| × 100 %), частками полів у межах ±25 % і ±50 % та з відношенням < 0,5 і > 1,5, а також коефіцієнтом рангової кореляції Спірмена між площами. Для фінальних полігонів порівняння із Square_H є коловим, оскільки Square_H використано для вибору межі, тому незалежною від облікової площі вважали лише оцінку варіанта вільного нарощування. Ті самі показники розраховано для полігонів базового методу SNIC. Різницю абсолютних відхилень між варіантами перевіряли критерієм Вілкоксона, між калібрувальними і контрольними полями – критерієм Манна–Вітні.

Формування різночасових композитів «оголеного» ґрунту. Для узгодження архівів супутникових даних із чотирма турами агрохімічного обстеження ґрунтів (2002, 2007, 2012, 2017 рр.) визначено відповідність «сенсор-тур» з урахуванням фактичного часового покриття: для турів 2002 та 2007 рр. використано дані Landsat 5 (LANDSAT/LT05/C02/T1_L2) та Landsat 7 (LANDSAT/LE07/C02/T1_L2); для туру 2012 р. – Landsat 7 та Landsat 8 (LANDSAT/LC08/C02/T1_L2, доступний з квітня 2013 р., тому для туру 2012 р. композит фактично охоплює лише 2013 р.); для туру 2017 р. – Sentinel-2 COPERNICUS/S2_SR_HARMONIZED) та Landsat 8. Для кожної комбінації «сенсор-тур» сформовано медіанний композит у часовому вікні ±1 рік навколо базового року туру (наприклад, 2001–2003 рр. - для туру 2002 р.) лише за сценами березня–травня та серпня–жовтня, коли поверхня полів найчастіше не вкрита рослинністю, без зимових сцен зі сніговим покривом і перезволоженою ріллею (для Sentinel-2 – сцени з часткою хмарних пікселів < 34 %), із маскуванням хмар, тіней та снігу (біти 3–5 QA_PIXEL для Landsat; для Sentinel-2 залишено класи SCL 2, 4–7). Пікселі «оголеного» ґрунту відбирали за протоколом GEOS3 (Demattê et al., 2018) з порогами, застосованими для орних земель Європи (Safanelli et al., 2020): $-0{,}05 \le \text{NDVI} \le 0{,}30$ (відсікання зеленої рослинності) і $-0{,}15 \le \text{NBR2} \le 0{,}15$ (відсікання сухих рослинних решток і стерні), де NBR2 = (swir1 − swir2) / (swir1 + swir2); після цього обчислено медіану відбивної здатності каналів і попіксельних індексів. До композитів увійшло від 25 до 121 сцени (Landsat 5: 39 і 80; Landsat 7: 95, 62 і 94; Landsat 8: 25 і 121; Sentinel-2: 86). Для кожного композиту розраховано дев'ять спектральних ознак: відбивну здатність у каналах blue, green, red, nir, swir1, swir2 (Landsat 5/7 – SR_B1, SR_B2, SR_B3, SR_B4, SR_B5, SR_B7; Landsat 8 – SR_B2–SR_B7; Sentinel-2 – B2, B3, B4, B8, B11, B12; масштабування Landsat Collection 2 Level-2: × 0,0000275 − 0,2; Sentinel-2: / 10 000) та індекси NDVI, BSI і NBR2; канал SWIR2 (2,1–2,3 мкм) охоплює смуги поглинання глинистих мінералів та органічних сполук. Щоб зменшити вплив змішаних пікселів на межах і можливих змін конфігурації полів між турами, зональні середні розраховано у внутрішній частині полігонів 2017 р., звуженій на 20 м від межі (якщо у звуженій зоні було менше 10 валідних пікселів, – у всьому полігоні; так розраховано 175 із 1856 записів «поле-композит»), функцією reduceRegions (tileScale = 8) у проєкції EPSG:32635 на нативних сітках сенсорів (30 м для Landsat, 20 м для Sentinel-2) разом із кількістю валідних пікселів. Записи вилучали з моделювання, якщо зона містила менше 10 валідних пікселів, валідні пікселі охоплювали менше половини її площі або ознаки вказували на залишки хмар чи снігу (NDVI < 0 або відбивна здатність у каналі blue > 0,3); після цього обсяг вибірки становив від 60 (X тур/Landsat 8) до 208 полів (табл. 2).

Регресійне моделювання вмісту гумусу. Зв'язок між вмістом гумусу за лабораторними визначеннями та дев'ятьма спектральними ознаками моделювали окремо для кожної з 8 комбінацій «сенсор-тур»; у кожній моделі одне поле представлене одним записом. Використано лінійну регресію з L2-регуляризацією (Ridge) зі стандартизацією ознак, у якій коефіцієнт регуляризації α обирали з 71 значення в діапазоні 10^{-2}–10^{5} ефективною перехресною перевіркою з виключенням по одному спостереженню лише на навчальній частині кожного фолду (RidgeCV у складі Pipeline, тобто у вкладеному циклі), та ансамблевий метод Random Forest (200 дерев, максимальна глибина 4, мінімум 5 спостережень у листі; Pedregosa et al., 2011). Якість моделей оцінено 5-кратною перехресною перевіркою з випадковим розподілом полів між фолдами (KFold, shuffle = True, random_state = 42) за коефіцієнтом детермінації (R^2), середньоквадратичною похибкою (Root Mean Square Error – RMSE) та середньою абсолютною похибкою (Mean Absolute Error – MAE), розрахованими за прогнозами для відкладених (out-of-fold) спостережень. Оскільки сусідні поля можуть мати подібні ґрунтові та спектральні властивості, а випадковий розподіл полів між фолдами може завищувати оцінки якості (Roberts et al., 2017; Ploton et al., 2020), стійкість оцінок перевірено за чотирма додатковими схемами:

випадковою 5-ти кратною з 10 повторами, просторовою блоковою за сіткою 2 × 2 км і 5 × 5 км (по 5 повторів) та за п’ятьма просторовими кластерами координат опорних точок (k-means). Для кожної схеми розраховано також наївну модель, яка прогнозує середнє навчальної вибірки. Для ілюстрацій використано прогнози для відкладених спостережень моделі Ridge для сенсора з вищим R^2 у кожному турі (2002 і 2012 рр. – Landsat 7, 2007 р. – Landsat 5, 2017 р. – Landsat 8). Як аналіз чутливості моделі повторено на всіх записах з наявними ознаками, без вилучення непридатних. Комбінацію X тур/Landsat 8 (n = 60; композит охоплює лише 2013 р.) наведено для повноти, але через малий обсяг вибірки її метрики не порівнюються безпосередньо з іншими (n = 152–208).

Статистична обробка даних. Для вмісту гумусу в кожному турі розраховано описові статистики (n, середнє, стандартне відхилення, коефіцієнт варіації CV, мінімум і максимум; табл. 1). Зміни між турами оцінено тестом Фрідмана для повторних вимірювань з коефіцієнтом конкордації Кендалла W; різницю між 2017 і 2002 рр. – парним t-тестом (середня різниця з 95 % довірчим інтервалом, ДІ) і критерієм знакових рангів Вілкоксона; лінійний тренд – лінійною змішаною моделлю «вміст гумусу ~ рік + (1 | поле)» (REML). Оскільки лабораторні значення вмісту гумусу округлено до 0,1 %, стійкість висновків перевіряли рандомізацією значень у межах ±0,05 % (1000 повторів). Якість регресійних моделей характеризували коефіцієнтом детермінації R^2, який показує частку дисперсії вмісту гумусу, пояснену моделлю; за перехресної перевірки від’ємне значення R^2 означає, що прогноз моделі гірший за прогноз середнім значенням. Зв’язок окремих спектральних ознак із вмістом гумусу оцінювали коефіцієнтами кореляції Пірсона (r) і Спірмена (ρ) з поправкою на множинні порівняння за процедурою Бенджаміні–Хохберга. Відтворення тренду моделями оцінювали за прогнозами для відкладених спостережень на однаковому наборі полів з придатними ознаками в усіх турах: зміну 2002→2017 рр. за прогнозом порівнювали з лабораторною зміною та зі зміною, яку дає наївна модель (95 % бутстреп-ДІ, 2000 повторів), а узгодженість змін на рівні окремих полів – коефіцієнтом Спірмена. Розрахунки виконано мовою Python з бібліотеками scikit-learn (Pedregosa et al., 2011), SciPy і statsmodels.

Результати та обговорення. Результати виділення меж полів. Апробація методу виділення меж в середовищі Google Earth Engine методом сегментації SNIC (Simple Non-Iterative Clustering) за комбінованим шаром NDVI+BSI, сформованим із безхмарного медіанного композиту Sentinel-2, з подальшою векторизацією та просторовим з’єднанням (point-in-polygon) із перенесенням атрибутів обстеження на полігони показала, що п’ять точок опинялися поза полігонами (на відстані 0,4–18,7 м). Покрокова перевірка довела, що чотири з них потрапляли на пікселі класу «деревна рослинність» ESA WorldCover (лісосмуги на краю поля), через що нарощування починалося із сусіднього суперпікселя, а одна (0,4 м) лежала біля межі, зміщеної під час спрощення контуру; обидві причини усунено (див. «Виділення меж полів»), тому жодне поле з моделювання не вилучали. Жоден полігон не містить інших опорних точок із повного масиву 447 полів (для базових полігонів SNIC – 5 з 232), тобто ознак об’єднання суміжних полів не виявлено. Власна опорна точка лежить усередині фінального полігона для всіх 232 полів. Полігони практично не перекриваються: сумарна площа перекриттів становить 0,66 га (0,008 % площі полігонів). Незалежну від облікової площі оцінку дає варіант вільного нарощування (рис. 2, б): медіанне відношення площі полігона до Square_H становить 0,97, MdAPE – 53,2 %; у межах ±25 % облікової площі перебувають 55 із 232 полів (23,7 %), у межах ±50 % – 108 (46,6 %), відношення < 0,5 мають 56 полів (24,1 %), > 1,5 – 68 (29,3 %); ρ Спірмена між площами – 0,37 ($p < 0,001$). На контрольних полях показники близькі до калібрувальних (медіанне відношення 0,98 проти 0,96; у межах ±50 % – 44,1 % проти 47,6 % полів; критерій Манна–Вітні для абсолютних відхилень $p = 0,88$), тобто перенавчання параметрів не виявлено. Базові полігони SNIC систематично занижують площу (рис. 2, а): медіанне відношення 0,30, MdAPE 73,3 %, у межах ±50 % – 56 полів (24,1 %), відношення < 0,5 – 165 (71,1 %), $\rho = -0,05$ ($p = 0,48$), оскільки їхній розмір визначався розміром суперпікселя (медіана 8,5 га), а не поля; вільне нарощування ближче до облікової площі для 134 з 232 полів, проте ця різниця статистично не значуща (критерій Вілкоксона, $p = 0,24$). Фінальні полігони з обмеженням ±15 % мають медіанне відношення 1,05 і MdAPE 11,9 % (у межах ±50 % – 227 з 232 полів; рис. 2, в), проте ця узгодженість закладена в алгоритм; п’ять фінальних полігонів мають площу менше половини облікової через обмеження коміркою Вороного.

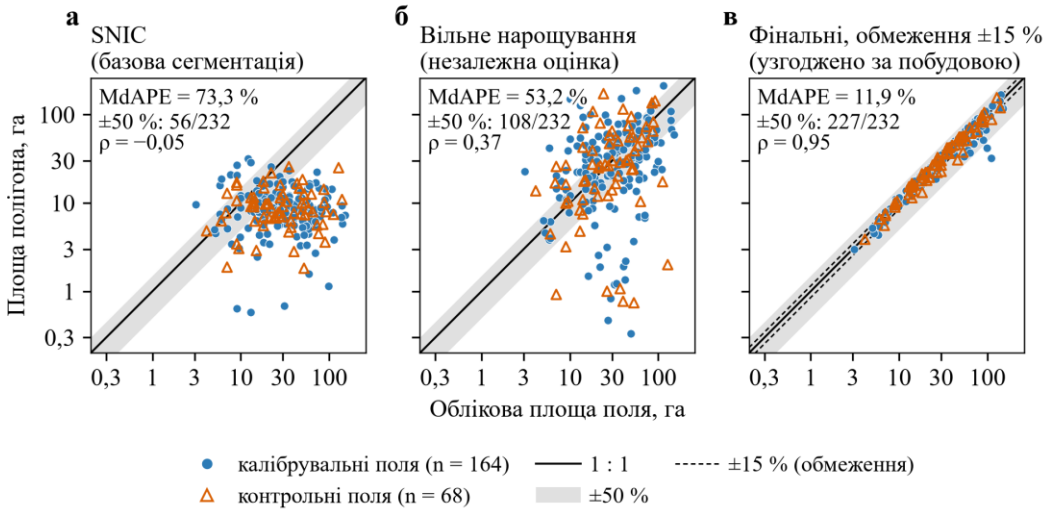


Рис. 2. Площа полігонів порівняно з обліковою площею полів (n = 232; логарифмічні шкали): а – базова сегментація SNIC; б – вільне нарощування області (незалежна оцінка); в – фінальні полігони з обмеженням ±15 % облікової площі (площу узгоджено за побудовою); суцільна лінія – 1 : 1, сіра смуга – ±50 %, пунктир – ±15 %; кола – калібрувальні, трикутники – контрольні поля

Fig. 2. Polygon area versus recorded field area (n = 232; log scales): a – baseline SNIC segmentation; b – unconstrained region growing (independent estimate); c – final polygons constrained to ±15 % of the recorded area (agreement by construction); solid line – 1 : 1, grey band – ±50 %, dashed lines – ±15 %; circles – calibration, triangles – validation fields

Порівняння з еталонними контурами 2017 р. (40 контрольних полів; рис. 3) показує, що геометрична точність усіх варіантів низька. Медіанний IoU фінальних полігонів становить 0,44 (95 % ДІ 0,34–0,59), коефіцієнт Дайса – 0,61; IoU ≥ 0,5 мають 17 із 40 полів, ≥ 0,7 – 7. Для вільного нарощування медіанний IoU становить 0,33 (0,25–0,43), для базової сегментації SNIC – 0,22 (0,16–0,27); фінальні полігони точніші за обидва варіанти (критерій Вілкоксона p = 0,009 і p < 0,001), тобто вибір межі у вікні ±15 % облікової площі в середньому покращує відповідність контурів. Водночас фінальні полігони частіше захоплюють сусідні землі: медіанна частка площі полігона поза еталоном становить 0,42 проти 0,28 для вільного нарощування, тоді як частка еталона поза полігоном – 0,34 проти 0,41. Граничний F1 низький для всіх методів (0,12 за допуску 10 м і 0,21–0,22 за допуску 20 м), медіанна відстань між центроїдами фінального полігона й еталона – 143 м. Систематичного зсуву не виявлено: оптимальне зміщення полігонів у межах ±60 м підвищує медіанний IoU лише до 0,54, тож низька точність зумовлена формою контурів – їх зубчастістю та виходом за межі поля там, де межа між посівами нечітка, – а не похибкою прив'язки. Оскільки полігони й еталони побудовано за знімками одного року, ці показники характеризують похибку самого алгоритму. Точність зростає з розміром поля (ρ Спірмена між обліковою площею та IoU – 0,40, p = 0,010). Медіанне відношення площі еталона до Square_H становить 0,98 (медіана абсолютного відхилення 11,7 %), тобто облікові площі добре відповідають межам полів станом на 2017 рік. Отже, отримані полігони придатні лише як наближені контури для просторового зіставлення даних щодо вмісту гумусу, а не як точні межі полів; для точнішого відтворення меж перспективними є методи на основі глибокого навчання (Persello et al., 2019; Waldner, Diakogiannis, 2020).

Часова динаміка вмісту гумусу за даними обстежень

На 232 полях, обстежених в усіх чотирьох турах, середній вміст гумусу у верхньому шарі ґрунту послідовно знижувався: з 2,13 % у 2002 р. до 2,10; 2,05 і 2,00 % у 2007, 2012 і 2017 роках відповідно (табл. 1). Зміни між турами статистично значущі (тест Фрідмана: $\chi^2 = 268{,}0$; df = 3; p < 0,001; W Кендалла = 0,39). У 2017 р. порівняно з 2002 р. вміст гумусу в середньому менший на 0,131 % (різниця −0,131 %, 95 % ДІ −0,150…−0,112 %; t = −13,33; p < 0,001; критерій Вілкоксона p < 0,001), тобто приблизно на 6 % від початкового рівня; на 183 полях зафіксовано зниження, на 30 – зміни відсутні, а на 19 – було зростання. За лінійною змішаною моделлю швидкість зниження становить 0,088 % за 10 років (95 % ДІ 0,078–0,097 %; p < 0,001). Висновки стійкі до округлення лабораторних значень: за рандомізації в межах ±0,05 % медіана різниці 2017 і 2002 рр. становила −0,131 % (2,5–97,5 перцентилі −0,136…−0,126 %). Вузький

діапазон значень вмісту гумусу (1,4–2,9 %) і низька мінливість (стандартне відхилення 0,20–0,22 %, коефіцієнт варіації 9,5–11,0 %) (табл. 1) відповідають генетичним особливостям ґрунтового покриву району, сформованому переважно дерново-підзолистими поверхнево-оглеєними, сірими і темно-сірими опідзоленими ґрунтами (Гунчак з співавт., 2025).

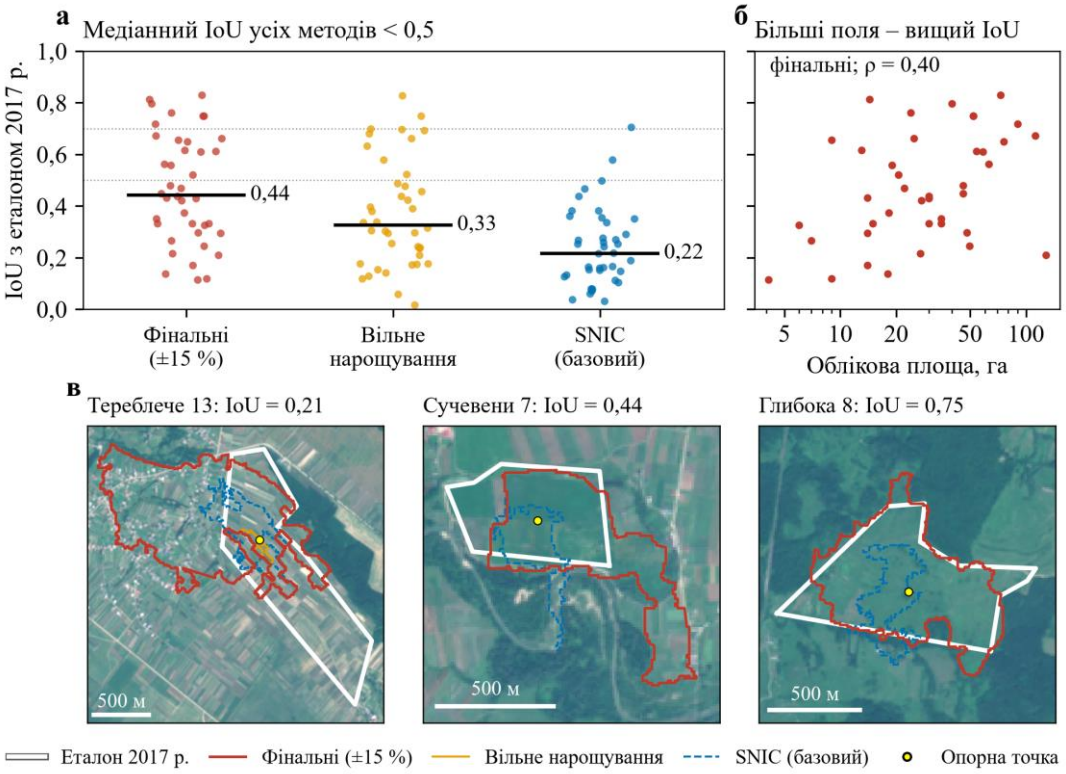


Рис. 3. Порівняння автоматизованих полігонів з еталонними контурами 2017 р. (n = 40 контрольних полів): а – IoU за методами (точки – окремі поля, горизонтальна лінія – медіана, пунктир – IoU = 0,5 і 0,7); б – IoU фінальних полігонів залежно від облікової площі поля (логарифмічна шкала); в – приклади полів з IoU фінальних полігонів, близьким до 10-го, 50-го і 90-го перцентилів (підкладка – медіанний композит Sentinel-2 L1C за червень–липень 2017 р., RGB)

Fig. 3. Comparison of automated polygons with 2017 reference boundaries (n = 40 validation fields): a – IoU by method (dots – individual fields, horizontal line – median, dotted lines – IoU = 0.5 and 0.7); b – IoU of the final polygons versus recorded field area (log scale); c – example fields with final-polygon IoU close to the 10th, 50th and 90th percentiles (background – Sentinel-2 L1C median composite, June–July 2017, RGB)

Таблиця 1.

Статистичні характеристики вмісту гумусу за турами обстеження (n = 232 поля)

Table 1.

Statistical characteristics of humus content by survey round (n = 232 fields)

Тур обстеження (рік)	Кількість визначень, n	Середнє значення, %	Стандартне відхилення, %	Коефіцієнт варіації CV, %	Діапазон, %
VIII (2002)	232	2,13	0,21	9,9	1,6–2,6
IX (2007)	232	2,10	0,20	9,5	1,5–2,5
X (2012)	232	2,05	0,21	10,2	1,4–2,5
XI (2017)	232	2,00	0,22	11,0	1,4–2,9

Водночас, стандартне відхилення вибірки за величиною близьке до допустимої розбіжності паралельних визначень гумусу в зразках ґрунту за методом Тюріна (≈ 0,15–0,20 % абс.), що обмежує інформативність міжпольових відмінностей (див. нижче). Виявлене зниження вмісту гумусу узгоджується з уявленнями про гумусний стан орних ґрунтів як результат балансу між надходженням органічних решток і їх біологічною трансформацією.

За даними дослідників, це зумовлено низкою факторів: 1) посилення мінералізації органічної речовини мікроорганізмами за інтенсивного обробітку, який руйнує ґрунтові агрегати і робить фізично захищену органічну речовину доступною для мікробного розкладу (Lehmann, Kleber, 2015; Wiesmeier et al., 2019); 2) зменшення надходження рослинних решток і органічних добрив, а отже – субстрату для мікробного анаболізму, через який формується значна частка стабільної органічної речовини ґрунту (Liang et al., 2017); 3) спрощення сівозмін і зменшення частки багаторічних трав у структурі посівів, що порушує біологічний кругообіг вуглецю та азоту; 4) підвищення температури навколишнього середовища, що прискорює мікробне розкладання органічної речовини; наприклад, для чорноземів проявляється залежність вмісту гумусу від змін клімату (Дегтярьов та Щербаков, 2023).

Якість регресійних моделей гумус ~ спектральні індекси. Результати перехресної перевірки регресійних моделей для 8 комбінацій «сенсор-тур» наведено в таблиці 2. За випадкової 5-кратної перевірки R^2 моделей Ridge становив від −0,067 до 0,016, Random Forest – від −0,107 до 0,055; RMSE моделей (0,205–0,244 %) практично дорівнювала RMSE наївної моделі, яка прогнозує середнє навчальної вибірки (0,207–0,237 %), і стандартному відхиленню вмісту гумусу (0,20–0,22 %, див. табл. 1). Підбір коефіцієнта регуляризації у вкладеному циклі усунув різко від'ємні значення R^2 моделі Ridge: медіана обраних α становила ≈ 200, а в 27 % фолдів було обрано верхню межу діапазону (10^5), за якої модель фактично зводиться до прогнозу середнім. Висновок не залежить від схеми перехресної перевірки (табл. 3): з 80 комбінацій «модель × «сенсор-тур» × схема» меншу RMSE, ніж наївна модель, мали 18, проте виграш не перевищував 0,010 % (Random Forest, IX тур/Landsat 7: R^2 = 0,055 за однократної та 0,052 за повторної випадкової перевірки, 0,013 і −0,003 за просторових блокових схем), що в кілька разів менше за похибку лабораторного методу; найнижчі значення R^2 дала кластерна просторова схема (до −0,284 для X туру/Landsat 8 за n = 60 і до −0,148 для інших комбінацій). Аналіз чутливості на всіх записах з наявними ознаками дав подібний результат (R^2 Ridge – від −0,084 до 0,030, Random Forest – від −0,164 до 0,034). Коефіцієнти парної кореляції окремих спектральних ознак з вмістом гумусу за модулем не перевищували 0,20 (r Пірсона) в усіх комбінаціях, крім X туру/Landsat 8 ($|r| \le 0{,}24$ за n = 60); номінально значущими ($p < 0{,}05$) були 7 з 72 коефіцієнтів (здебільшого для BSI, з додатним знаком), але жоден не залишився значущим після поправки на множинні порівняння.

Таблиця 2.

Метрики якості регресійних моделей (Ridge і Random Forest, 5-кратна перехресна перевірка) за сенсор-турами

Table 2.

Performance metrics of regression models (Ridge and Random Forest, 5-fold cross-validation) by sensor–round combination

Тур	Сенсор	n	R^2, Ridge	RMSE, %, Ridge	MAE, %, Ridge	R^2, RF	RMSE, %, RF	MAE, %, RF
VIII	Landsat 5	157	−0,033	0,209	0,170	−0,067	0,212	0,175
VIII	Landsat 7	202	0,016	0,215	0,176	−0,012	0,218	0,181
IX	Landsat 5	160	−0,029	0,215	0,172	−0,074	0,220	0,181
IX	Landsat 7	152	−0,067	0,218	0,176	0,055	0,206	0,166
X	Landsat 7	200	−0,025	0,212	0,169	−0,107	0,221	0,174
X	Landsat 8	60	−0,035	0,212	0,164	0,029	0,205	0,158
XI	Landsat 8	177	−0,007	0,235	0,178	−0,080	0,244	0,185
XI	Sentinel-2	208	−0,020	0,230	0,175	−0,060	0,234	0,180

Примітка: n – кількість полів з придатними спектральними ознаками (вилучено записи з < 10 валідними пікселями, валідною часткою площі < 0,5, NDVI < 0 або blue > 0,3); RF – Random Forest; RMSE наївної моделі (середнє навчальної вибірки) – 0,207–0,237 %; комбінацію X/Landsat 8 (n = 60) наведено для повноти, її метрики не є безпосередньо порівнянними з іншими.

Note. n – number of fields with valid spectral features (records with < 10 valid pixels, valid area fraction < 0.5, NDVI < 0 or blue > 0.3 excluded); RF – Random Forest; naive-model (training-set mean) RMSE – 0.207–0.237 %; the X/Landsat 8 combination (n = 60) is shown for completeness, and its metrics are not directly comparable with the others.

Таблиця 3.

Коефіцієнт детермінації регресійних моделей за різних схем 5-кратної перехресної перевірки (8 комбінацій сенсор-тур; для схем з повторами – середнє за повторами)

Table 3.

Coefficient of determination of the regression models under different 5-fold cross-validation schemes (8 sensor–round combinations; mean over repeats for repeated schemes)

Схема перехресної перевірки / Cross-validation scheme	Ridge: R^2, мін.–макс. (середнє)	Random Forest: R^2, мін.–макс. (середнє)	RMSE, %	Моделей, кращих за наївну (з 16)
Випадкова 5-кратна, 1 повтор / random, 1 repeat	−0,067…0,016 (−0,025)	−0,107…0,055 (−0,039)	0,205–0,244	5
Випадкова 5-кратна, 10 повторів / random, 10 repeats	−0,065…−0,002 (−0,032)	−0,095…0,052 (−0,044)	0,206–0,239	4
Просторові блоки 2 × 2 км, 5 повторів / 2 × 2 km blocks, 5 repeats	−0,143…−0,006 (−0,044)	−0,107…−0,003 (−0,052)	0,209–0,238	5
Просторові блоки 5 × 5 км, 5 повторів / 5 × 5 km blocks, 5 repeats	−0,087…−0,019 (−0,043)	−0,146…0,013 (−0,069)	0,210–0,243	1
Просторові кластери k-means (5) / k-means clusters (5)	−0,284…−0,008 (−0,082)	−0,135…−0,020 (−0,083)	0,208–0,242	3

Примітка. «Кращі за наївну» – моделі з меншою RMSE, ніж у наївної моделі за тієї самої схеми
Note. "Better than naive" – models with a lower RMSE than the naive model under the same scheme

Розсіювання точок на графіку демонструє відсутність вираженої залежності прогнозованих значень вмісту гумусу від фактичних, оскільки точки утворюють горизонтальну смугу навколо середнього прогнозованого значення замість групування уздовж лінії ідеального збігу (рис. 4). Аналогічно, просторовий розподіл розрахункових значень вмісту гумусу (рис. 5, б) виявляється істотно згладженим порівняно з фактичним розподілом (рис. 5, а). Це свідчить про фактичне відтворення моделлю середніх значень вибірки, а не локальної варіативності вмісту гумусу: стандартне відхилення прогнозів моделі Ridge (0,017–0,039 %) у 6–14 разів менше, ніж лабораторних значень на тих самих полях (0,22–0,24 %). Отримані результати узгоджуються з відомим методологічним обмеженням дистанційної діагностики гумусу в ґрунтах: спектральні ознаки композитів «оголеного» ґрунту з просторовою роздільною здатністю 10–30 м добре розрізняють контрастні ґрунтові відміни (наприклад, піщані та глинисті ґрунти, еродовані ділянки), проте значно менш чутливі до різниць вмісту гумусу в межах одного відсоткового пункту, характерних для однорідного ґрунтового покриву одного ґрунтово-кліматичного району (Castaldi et al., 2019; Vaudour et al., 2019; Angelopoulou et al., 2019).

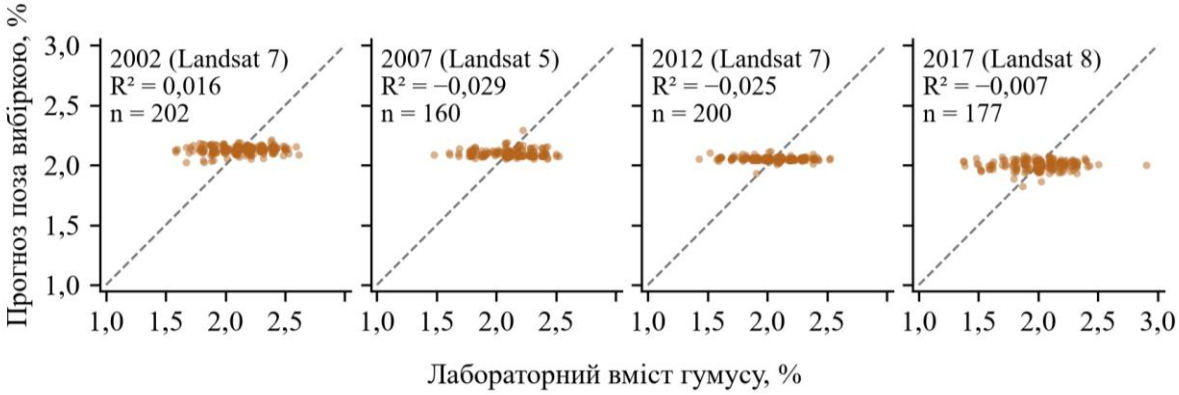


Рис. 4. Співвідношення лабораторних значень вмісту гумусу та прогнозів моделі Ridge для відкладених спостережень (out-of-fold, 5-кратна перехресна перевірка; сенсор з вищим R^2 у кожному турі); пунктирна лінія – лінія ідеального збігу (y = x)

Fig. 4. Relationship between laboratory humus content and out-of-fold predictions of the Ridge model (5-fold cross-validation; sensor with the higher R^2 in each survey round); dashed line – line of perfect agreement (y = x)

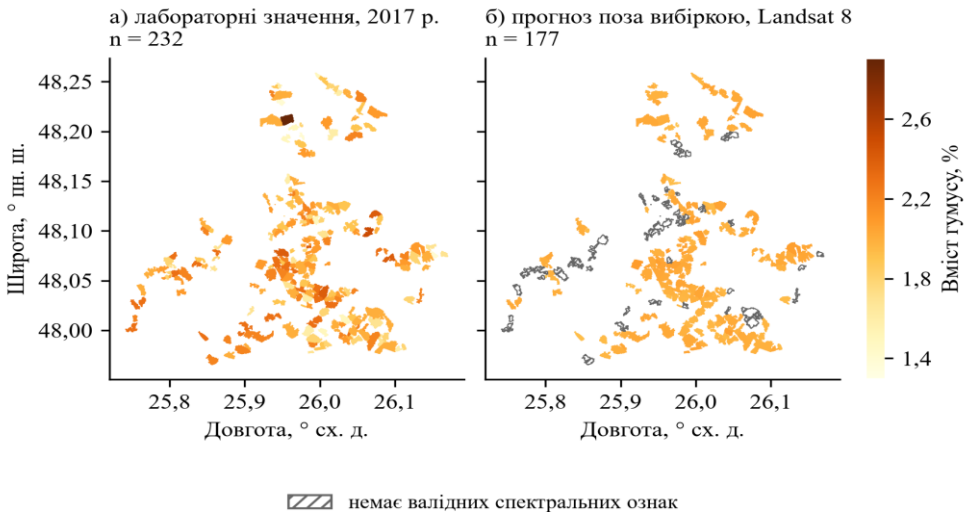


Рис. 5. Просторовий розподіл вмісту гумусу станом на 2017 р.: а – лабораторні значення (232 поля); б – прогноз моделі Ridge за даними Landsat 8 для відкладених спостережень (177 полів; заштриховано поля без придатних спектральних ознак); однакова шкала значень

Fig. 5. Spatial distribution of humus content in 2017: a – laboratory values (232 fields); b – out-of-fold prediction of the Ridge model from Landsat 8 data (177 fields; hatched – fields without valid spectral features); common colour scale

У цитованих працях моделюється вміст органічного вуглецю ґрунту (SOC), пов'язаний із вмістом гумусу через коефіцієнт перерахунку (умовно 1,724, хоча фактичне значення варіює; Pribyl, 2010), тому їх висновки щодо чутливості переносимо на гумус лише якісно. Подібність середніх прогнозованих і лабораторних значень за турами (рис. 6) не є доказом прогностичної здатності моделей. На однаковому наборі з 138 полів зміна 2002→2017 рр. за прогнозами Ridge для відкладених спостережень (−0,129 %) близька до лабораторної (−0,141 %; різниця 0,012 %, 95 % ДІ −0,015…0,039 %), але так само її відтворює і наївна модель, яка взагалі не використовує спектральних даних (−0,125 %); прогноз Random Forest також становить −0,125 %. Зміни на рівні окремих полів моделі не відтворюють: коефіцієнт Спірмена між прогнозованою і лабораторною зміною від'ємний як для Ridge (−0,19), так і для наївної моделі (−0,21), що є артефактом перехресної перевірки (виключення поля з навчальної вибірки зміщує її середнє в протилежний бік), а для Random Forest становить −0,09 (p = 0,27). Отже, збіг середніх є наслідком того, що моделі навчалися окремо для кожного туру на його власних лабораторних значеннях, і не свідчить про здатність дистанційно оцінювати динаміку саме вмісту гумусу.

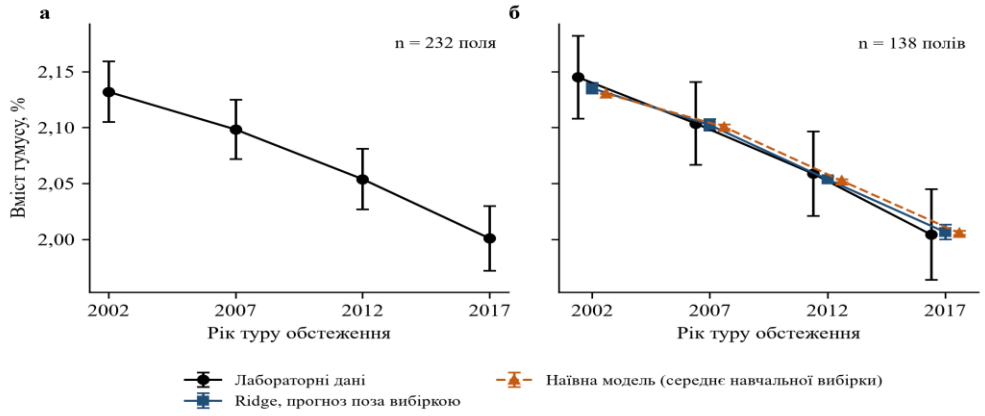


Рис. 6. Динаміка середнього вмісту гумусу за турами обстеження: а – лабораторні значення (n = 232); б – середні лабораторні значення, прогнози Ridge для відкладених спостережень (сенсор з вищим R^2 у кожному турі) та прогнози наївної моделі (середнє навчальної вибірки) на однаковому наборі полів (n = 138); планки похибок – 95 % довірчий інтервал; маркери зміщено по горизонталі для читабельності

Fig. 6. Dynamics of mean humus content by survey round: a – laboratory values (n = 232); b – mean laboratory values, out-of-fold Ridge predictions (sensor with the higher R^2 in each round) and naive-model predictions (training-set mean) on the same set of fields (n = 138); error bars – 95 % confidence interval; markers are offset horizontally for readability

Принциповим обмеженням є співвідношення «сигнал/шум» цільової змінної. Стандартне відхилення вмісту гумусу між полями в межах туру (0,20–0,22 %) сумірне з допустимою розбіжністю паралельних визначень за методом Тюріна (≈ 0,15–0,20 % абс.). Якщо розглядати цю величину як межу повторюваності (≈ 2,8 стандартного відхилення похибки), стандартне відхилення лабораторної похибки становить ≈ 0,05–0,07 %, що разом з похибкою округлення до 0,1 % (≈ 0,03 %) відповідає приблизно 8–15 % дисперсії вибірки. З урахуванням похибки відбору (одна змішана проба на поле площею 30–50 га), а також можливих міжлабораторних і міжтурових систематичних відмінностей частка шуму є істотно більшою, а частка дисперсії, яку принципово можна пояснити будь-якими предикторами, – відповідно меншою. Ще менш сприятливе співвідношення для змін на рівні окремого поля: середня зміна за 15 років (0,131 %) менша за допустиму розбіжність окремого визначення, тому невідтворення моделями змін на рівні окремих полів є очікуваним. Отже, вузька варіація цільової змінної, сумірна з похибкою лабораторного методу, є фундаментальною причиною неможливості побудови інформативної регресії на рівні поля.

Застосований протокол побудови ознак (відбір пікселів «оголеного» ґрунту за NDVI і NBR2 із сезонним обмеженням, канал SWIR2, розрахунок ознак у внутрішній частині полігонів 2017 р.) і підбір коефіцієнта регуляризації наблизили якість моделей до рівня прогнозу середнім значенням, але не забезпечили практично значущого прогнозу: максимальний R^2 не перевищив 0,055, а обрані значення α часто відповідали майже повному «вимиканню» ознак. Отже, на цій території спектральні ознаки «оголеного» ґрунту не містять виявного сигналу міжпольових відмінностей вмісту гумусу, який перевищував би шум цільової змінної.

Практичним наслідком отриманих результатів є те, що отримані за супутниковими даними значення вмісту гумусу в ґрунтах полів без проведення лабораторних вимірювань слід розглядати виключно як індикативні. Для підвищення точності дистанційної оцінки вмісту гумусу в ґрунтах в межах однорідних за ґрунтовим покривом агрокліматичних районів перспективним є розширення набору ознак, зокрема за рахунок текстурних характеристик зображення, показників рельєфу (цифрових моделей висот), багаторічних (а не лише bare-soil) композитів та глобальних ґрунтових продуктів (наприклад, SoilGrids) як допоміжних коваріатів.

Нечутливість спектральних ознак до відмінностей вмісту гумусу в межах одного відсоткового пункту повинна враховуватися в екологічному моніторингу ґрунтів. Зміна, виявлена за 15 років (близько 0,1 % в абсолютному вираженні), на порядок менша за цей поріг, тому дистанційні методи в застосованій постановці не дозволяють відстежувати поступову дегуміфікацію ґрунтів, а отже – й пов’язані з нею зміни трофічної та енергетичної бази педобіоти на рівні окремих полів. У такій постановці супутникові дані можуть слугувати лише для стратифікації території (виділення контрастних ділянок, еродованих схилів, зон з різною історією землекористування) і планування точок відбору проб, тоді як оцінка гумусного стану ґрунтів і пов’язаного з ним стану трофічної бази педобіоценозів потребує аналітичних визначень; у сучасних системах оцінки здоров’я ґрунту вміст органічної речовини розглядають разом із біологічними показниками – мікробною біомасою, «диханням ґрунту», лабільними фракціями органічної речовини (Lehmann et al., 2020). Регулярні тури агрохімічного обстеження залишаються незамінним джерелом даних для агроекологічного моніторингу.

Отримані результати мають такі обмеження: 1) межі полів для всіх турів виділено за знімками 2017 р.; можливі зміни конфігурації полів у 2002–2012 рр. частково нівельовано розрахунком ознак у внутрішній частині полігонів, але інструментально не контролювались; 2) геометрична точність полігонів низька (медіанний IoU 0,44) і оцінена на 40 полях за еталонами одного виконавця без повторної оцифровки; під час оцифровки для контролю грубих помилок була доступна облікова площа поля, що могло наблизити еталони до Square_H; 3) міжпольова варіація вмісту гумусу сумірна з похибкою лабораторного методу, що обмежує максимально досяжну якість будь-якої регресійної моделі; 4) аналіз обмежено охопленими впродовж всіх чотирьох турів обстеження 232 полями з 447 наявних, що забезпечує порівнянність турів, але зменшує обсяг вибірки для окремих комбінацій «сенсор-тур», особливо для даних Landsat 8 у 2012 р. (n = 60).

Висновки:

1. Апробовано автоматизоване виділення меж полів у середовищі Google Earth Engine за поодинокими опорними точками обстеження (SLIC-сегментація трисезонних композитів Sentinel-2 станом на 2017 рік і нарощування області з вибором межі в діапазоні ±15 % облікової площі). Полігони містять власні опорні точки, не об’єднують суміжні поля і практично не

перекриваються; за вручну оцифрованими еталонами того самого року (40 полів) медіанний IoU фінальних полігонів становить 0,44 (вільного нарощування – 0,33, SNIC – 0,22), тому отримані полігони придатні лише як наближені контури для зіставлення лабораторних і супутникових даних.

2. За результатами обстежених в усіх чотирьох турах (2002–2017 рр.) 232 полів, за 15 років середній вміст гумусу статистично значуще знизився з 2,13 до 2,00 % (−0,131 %, 95 % ДІ −0,150…−0,112 %; тест Фрідмана $p < 0,001$), що відповідає швидкості 0,088 % за 10 років. Можливими біологічними причинами дегуміфікації ґрунтів є переважання мінералізації органічної речовини над гумусоутворенням унаслідок зменшення надходження до ґрунту рослинних решток та органічних добрив і руйнування структурних агрегатів під час обробітку.

3. Регресійні моделі зв'язку спектральних ознак «оголеного» ґрунту з вмістом гумусу (Ridge, Random Forest), побудовані окремо для 8 комбінацій «сенсор-тур» (Landsat 5/7/8, Sentinel-2; медіанні композити у вікні ±1 рік від базового року обстеження), навіть після відбору пікселів «оголеного» ґрунту за NDVI і NBR2, включення каналу SWIR2 та підбору коефіцієнта регуляризації мали R^2 від −0,284 до 0,055 і зменшували RMSE порівняно з прогнозом середнім не більше ніж на 0,010 %. Прогностичну здатність моделей на рівні окремого поля не підтверджено; збіг середніх прогнозованих і лабораторних значень за турами тривіальний і не є доказом здатності досліджуваного методу відтворювати динаміку вмісту гумусу.

4. Основною причиною відсутності прогностичної здатності моделі є вузький діапазон міжпольової варіації вмісту гумусу, сумірний з похибкою лабораторного методу; додатковий вплив можуть мати вологість ґрунту, шорсткість його поверхні та рослинні рештки, а також невисока просторова роздільна здатність - 10 –30 м. Прогнозовані значення вмісту гумусу в ґрунтах мають лише індикативний характер і не замінюють лабораторних його визначень.

5. Перспективними напрямами подальших досліджень є розширення набору предикторів (текстурні характеристики зображення, показники рельєфу, глобальні ґрунтові продукти SoilGrids (Poggio et al., 2021), методи виділення меж) на основі глибокого навчання з перевіркою за еталонними контурами та поєднання дистанційних даних із результатами аналітичного визначення вмісту гумусу для моніторингу гумусного стану ґрунтів агроекосистем.

Конфлікт інтересів. *Автори заявляють, що дослідження проводилося за відсутності будь-яких комерційних або фінансових відносин, які можна було б витлумачити як потенційний конфлікт інтересів.*

Список літератури:

1. Биндич, Т. Ю. (2019). Оцінювання диференціації ґрунтового покриву за допомогою космічних зображень. *Таврійський науковий вісник, 109* (2), 162–170. https://doi.org/10.32851/2226-0099.2019.109-2.26
2. Гунчак, М. В., Пасічняк, В. І., Грищенко, О. М., Колесник, Т. М. (2025). Стан родючості ґрунтів лісостепової зони Чернівецької області. Вісник Національного університету водного господарства та природокористування, 2 (110), 80–93. https://doi.org/10.31713/vs220256
3. Дегтярьов, В. В., Щербаков, О. Ю. (2023). Уміст гумусу в чорноземах типових Лівобережного Лісостепу України у зв'язку з глобальними змінами клімату. Агрохімія і ґрунтознавство, 95, 60–68. https://doi.org/10.31073/acss95-06
4. Шатохін, А. В., Ачасов, А. Б., Линдін, М. О. (1999). Технологія аерокосмічного контролю за гумусним станом ґрунтів. У *Розробки – виробництву: збірник наукових праць* (с. 107–108). Київ: Аграрна наука.
5. Шатохін, А. В., Чічоткіна, О. А., Ніколенко, І. Б. (1996). Спектральні характеристики ґрунтів Лісостепу і Степу України та їх числова таксономія. *Агрохімія і ґрунтознавство, 58*, 45–52.
6. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (11), 2274–2282. https://doi.org/10.1109/TPAMI.2012.120
7. Achanta, R., Süsstrunk, S. (2017). Superpixels and polygons using simple non-iterative clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4895–4904). https://doi.org/10.1109/CVPR.2017.520
8. Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., Bochtis, D. (2019). Remote sensing techniques for soil organic carbon estimation: A review. Remote Sensing, 11 (6), 676. https://doi.org/10.3390/rs11060676
9. Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B. (2019). Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS Journal of Photogrammetry and Remote Sensing, 147*, 267–282. https://doi.org/10.1016/j.isprsjprs.2018.11.026
10. Demattê, J. A. M., Fongaro, C. T., Rizzo, R., Safanelli, J. L. (2018). Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sensing of Environment, 212, 161–175. https://doi.org/10.1016/j.rse.2018.04.047
11. Diek, S., Fornallaz, F., Schaepman, M. E., de Jong, R.

(2017). Barest pixel composite for agricultural areas using Landsat time series. *Remote Sensing, 9* (12), 1245. https://doi.org/10.3390/rs9121245
12. Gayibov, A. (2025). *Agricultural field boundary detection through integration of Simple Non-Iterative Clustering (SNIC) super pixels and Canny edge detection method* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.04529
13. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031
14. Lehmann, J., Bossio, D. A., Kögel-Knabner, I., Rillig, M. C. (2020). The concept and future prospects of soil health. Nature Reviews Earth & Environment, 1 (10), 544–553. https://doi.org/10.1038/s43017-020-0080-8
15. Lehmann, J., Kleber, M. (2015). The contentious nature of soil organic matter. Nature, 528 (7580), 60–68. https://doi.org/10.1038/nature16069
16. Liang, C., Schimel, J. P., Jastrow, J. D. (2017). The importance of anabolism in microbial control over soil carbon storage. Nature Microbiology, 2 (8), 17105. https://doi.org/10.1038/nmicrobiol.2017.105
17. Pasquarella, V. J., Brown, C. F., Czerwinski, W., Rucklidge, W. J. (2023). Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 2125–2135). https://doi.org/10.1109/CVPRW59228.2023.00206
18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf
19. Persello, C., Tolpekin, V. A., Bergado, J. R., de By, R. A. (2019). Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. Remote Sensing of Environment, 231, 111253. https://doi.org/10.1016/j.rse.2019.111253
20. Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nature Communications, 11 (1), 4540. https://doi.org/10.1038/s41467-020-18321-y
21. Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL, 7* (1), 217–240. https://doi.org/10.5194/soil-7-217-2021
22. Pribyl, D. W. (2010). A critical review of the conventional SOC to SOM conversion factor. Geoderma, 156 (3–4), 75–83. https://doi.org/10.1016/j.geoderma.2010.02.003
23. Rikimaru, A., Roy, P. S., Miyatake, S. (2002). Tropical forest cover density mapping. *Tropical Ecology, 43* (1), 39–47.
24. Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40 (8), 913–929. https://doi.org/10.1111/ecog.02881
25. Runfola, D., Anderson, A., Baier, H., Crittenden, M., Dowker, E., Fuhrig, S., Goodman, S., Grimsley, G., Layko, R., Melville, G., Mulder, M., Oberman, R., Panganiban, J., Peck, A., Seitz, L., Shea, S., Slevin, H., Youngerman, R., Hobbs, L. (2020). geoBoundaries: A global database of political administrative boundaries. PLOS ONE, 15 (4), e0231866. https://doi.org/10.1371/journal.pone.0231866
26. Safanelli, J. L., Chabrillat, S., Ben-Dor, E., Demattê, J. A. M. (2020). Multispectral models from bare soil composites for mapping topsoil properties over Europe. Remote Sensing, 12 (9), 1369. https://doi.org/10.3390/rs12091369
27. Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. Advances in Agronomy, 107, 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7
28. Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P. (2019). Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sensing of Environment, 223*, 21–33. https://doi.org/10.1016/j.rse.2019.01.006
29. Waldner, F., Diakogiannis, F. I. (2020). Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. Remote Sensing of Environment, 245, 111741. https://doi.org/10.1016/j.rse.2020.111741
30. Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.-J., Kögel-Knabner, I. (2019). Soil organic carbon storage as a key function of soils – A review of drivers and indicators at various scales. Geoderma, 333, 149–162. https://doi.org/10.1016/j.geoderma.2018.07.026
31. Yakovenko, V., Gorban, V., Kotovych, O., Didur, O., Poleva, J. (2025). Humus forms, earthworm bioturbation and soil organic carbon storage in Chernozems of the low-intensity land use of steppe zone of Ukraine. Geoderma Regional, 42, e00988. https://doi.org/10.1016/j.geodrs.2025.e00988

References:

1. Byndych, T. Yu. (2019). The assessment of soil cover differentiation by space imagery [Otsiniuvannia dyferentsiatsii gruntovoho pokryvu za dopomohoiu kosmichnykh zobrazhen]. *Taurian Scientific Herald, 109* (2), 162–170. https://doi.org/10.32851/2226-0099.2019.109-2.26 [in Ukrainian].

2. Hunchak, M. V., Pasichniak, V. I., Hryshchenko, O. M., Kolesnyk, T. M. (2025). The state of soil fertility in the forest-steppe zone of Chernivtsi region [Stan rodiuchosti gruntiv lisostepovoi zony Chernivetskoi oblasti]. Bulletin of the National University of Water and Environmental Engineering, 2 (110), 80–93. https://doi.org/10.31713/vs220256 [in Ukrainian].
3. Dehtiarov, V. V., Shcherbakov, O. Yu. (2023). Humus content in chernozems typical of the Left-Bank Forest-Steppe of Ukraine in connection with global climate change [Umist humusu v chornozemakh typovykh Livoberezhnoho Lisostepu Ukrainy u zviazku z hlobalnymy zminamy klimatu]. Agrochemistry and Soil Science, 95, 60–68. https://doi.org/10.31073/acss95-06 [in Ukrainian].
4. Shatokhin, A. V., Achasov, A. B., Lyndin, M. O. (1999). Technology of aerospace monitoring of soil humus status [Tekhnolohiia aerokosmichnoho kontroliu za humusnym stanom gruntiv]. In *Developments for production: Collection of scientific papers [Rozrobky – vyrobnytstvu: zbirnyk naukovykh prats]* (pp. 107–108). Kyiv: Ahrarna nauka. [in Ukrainian].
5. Shatokhin, A. V., Chichotkina, O. A., Nikolenko, I. B. (1996). Spectral characteristics of soils of the Forest-Steppe and Steppe of Ukraine and their numerical taxonomy [Spektralni kharakterystyky gruntiv Lisostepu i Stepu Ukrainy ta yikh chyslova taksonomiia]. *Agrochemistry and Soil Science, 58*, 45–52. [in Ukrainian].
6. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (11), 2274–2282. https://doi.org/10.1109/TPAMI.2012.120
7. Achanta, R., Süsstrunk, S. (2017). Superpixels and polygons using simple non-iterative clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4895–4904). https://doi.org/10.1109/CVPR.2017.520
8. Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., Bochtis, D. (2019). Remote sensing techniques for soil organic carbon estimation: A review. Remote Sensing, 11 (6), 676. https://doi.org/10.3390/rs11060676
9. Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B. (2019). Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS Journal of Photogrammetry and Remote Sensing, 147*, 267–282. https://doi.org/10.1016/j.isprsjprs.2018.11.026
10. Demattê, J. A. M., Fongaro, C. T., Rizzo, R., Safanelli, J. L. (2018). Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sensing of Environment, 212, 161–175. https://doi.org/10.1016/j.rse.2018.04.047
11. Diek, S., Fornallaz, F., Schaepman, M. E., de Jong, R. (2017). Barest pixel composite for agricultural areas using Landsat time series. *Remote Sensing, 9* (12), 1245. https://doi.org/10.3390/rs9121245
12. Gayibov, A. (2025). *Agricultural field boundary detection through integration of Simple Non-Iterative Clustering (SNIC) super pixels and Canny edge detection method* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.04529
13. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031
14. Lehmann, J., Bossio, D. A., Kögel-Knabner, I., Rillig, M. C. (2020). The concept and future prospects of soil health. Nature Reviews Earth & Environment, 1 (10), 544–553. https://doi.org/10.1038/s43017-020-0080-8
15. Lehmann, J., Kleber, M. (2015). The contentious nature of soil organic matter. Nature, 528 (7580), 60–68. https://doi.org/10.1038/nature16069
16. Liang, C., Schimel, J. P., Jastrow, J. D. (2017). The importance of anabolism in microbial control over soil carbon storage. Nature Microbiology, 2 (8), 17105. https://doi.org/10.1038/nmicrobiol.2017.105
17. Pasquarella, V. J., Brown, C. F., Czerwinski, W., Rucklidge, W. J. (2023). Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 2125–2135). https://doi.org/10.1109/CVPRW59228.2023.00206
18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf
19. Persello, C., Tolpekin, V. A., Bergado, J. R., de By, R. A. (2019). Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. Remote Sensing of Environment, 231, 111253. https://doi.org/10.1016/j.rse.2019.111253
20. Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nature Communications, 11 (1), 4540. https://doi.org/10.1038/s41467-020-18321-y
21. Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL, 7* (1), 217–240. https://doi.org/10.5194/soil-7-217-2021
22. Pribyl, D. W. (2010). A critical review of the conventional SOC to SOM conversion factor. Geoderma, 156 (3–4), 75–83. https://doi.org/10.1016/j.geoderma.2010.02.003
23. Rikimaru, A., Roy, P. S., Miyatake, S. (2002). Tropical forest cover density mapping. *Tropical Ecology, 43* (1), 39–47.

24. Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40 (8), 913–929. https://doi.org/10.1111/ecog.02881
25. Runfola, D., Anderson, A., Baier, H., Crittenden, M., Dowker, E., Fuhrig, S., Goodman, S., Grimsley, G., Layko, R., Melville, G., Mulder, M., Oberman, R., Panganiban, J., Peck, A., Seitz, L., Shea, S., Slevin, H., Youngerman, R., Hobbs, L. (2020). geoBoundaries: A global database of political administrative boundaries. PLOS ONE, 15 (4), e0231866. https://doi.org/10.1371/journal.pone.0231866
26. Safanelli, J. L., Chabrillat, S., Ben-Dor, E., Demattê, J. A. M. (2020). Multispectral models from bare soil composites for mapping topsoil properties over Europe. Remote Sensing, 12 (9), 1369. https://doi.org/10.3390/rs12091369
27. Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. Advances in Agronomy, 107, 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7
28. Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P. (2019). Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sensing of Environment, 223*, 21–33. https://doi.org/10.1016/j.rse.2019.01.006
29. Waldner, F., Diakogiannis, F. I. (2020). Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. Remote Sensing of Environment, 245, 111741. https://doi.org/10.1016/j.rse.2020.111741
30. Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.-J., Kögel-Knabner, I. (2019). Soil organic carbon storage as a key function of soils – A review of drivers and indicators at various scales. Geoderma, 333, 149–162. https://doi.org/10.1016/j.geoderma.2018.07.026
31. Yakovenko, V., Gorban, V., Kotovych, O., Didur, O., Poleva, J. (2025). Humus forms, earthworm bioturbation and soil organic carbon storage in Chernozems of the low-intensity land use of steppe zone of Ukraine. Geoderma Regional, 42, e00988. https://doi.org/10.1016/j.geodrs.2025.e00988

MONITORING THE DYNAMICS OF HUMUS CONTENT IN AGROCENOSIS SOILS BASED ON SATELLITE OBSERVATION DATA

D. I. Robuliak[1], I. S. Smaha[1], M. V. Hunchak[2]

[1] Yuriy Fedkovych Chernivtsi National University,
Ukraine, 58022, Chernivtsi, Kotsiubynsky 2 Str.
robuliak.dmytro@chnu.edu.ua, i.smaga@chnu.edu.ua
[2] Chernivtsi Branch of the State Institution «Soil Protection Institute of Ukraine»,
194-A Heroiv Maidanu St., Chernivtsi, 58029, Ukraine
Ukrainian Research Plant Quarantine Station of the Institute of Plant Protection of NAAS,
4 Naukova St., Boiany, Chernivtsi region, 60321, Ukraine
chernivtsy_grunt@ukr.net

Humus is formed through the microbial transformation of plant residues, is the main pool of organic carbon in agroecosystems and provides the trophic and energy basis for soil biota. Total humus content is a conservative indicator of soil humus status whose long-term dynamics reflects the balance between biogenic organic matter inputs and microbial mineralisation, i.e. the state of the trophic basis of soil biota communities, and the search for methods of its remote estimation is a relevant task of soil monitoring. The aim of the study was to assess the dynamics of soil humus content in agroecosystems of the former Hlyboka district, Chernivtsi region, in 2002–2017 and to test whether bare-soil spectral features derived from Landsat and Sentinel-2 data allow humus content to be estimated at the level of individual fields. The results of four rounds of agrochemical survey (2002, 2007, 2012, 2017) of 232 fields surveyed in all rounds were used (humus determined by the Tyurin method); each field is represented by a single reference point. Field boundaries were delineated in Google Earth Engine from three-season Sentinel-2 composites of 2017 (the year of the XI survey round) by SLIC segmentation and region growing from the reference point, with the boundary selected within ±15 % of the recorded field area; parameters were tuned on 164 calibration fields and checked on 68 validation fields, and geometric accuracy was assessed against manually digitised reference boundaries of 40 validation fields for the same year. For each of 8 «sensor–round» combinations, median bare-soil composites were built from Landsat 5/7/8 and Sentinel-2 data within a ±1-year window around the survey round, with GEOS3-style NDVI and NBR2 pixel screening and a seasonal restriction; Ridge (with the regularisation parameter tuned in nested cross-validation) and Random Forest regression models based on nine spectral features (six bands including SWIR2, NDVI, BSI and NBR2) computed within polygons shrunk by 20 m were evaluated by random and spatial 5-fold cross-validation. The median Intersection over Union (IoU) was 0.44 for the final polygons (95 % CI 0.34–0.59; IoU ≥ 0.7 for 7 of 40 fields), 0.33 for unconstrained region growing and 0.22 for the baseline SNIC segmentation, so the polygons are suitable only as approximate field

outlines. Mean humus content decreased from 2.13 % in 2002 to 2.00 % in 2017 (difference −0.131 %, 95 % CI −0.150 to −0.112 %; Friedman test p < 0.001) within a narrow range of values (1.4–2.9 %, standard deviation 0.20–0.22 %, coefficient of variation 9.5–11.0 %) typical of the predominantly sod-podzolic and grey forest soils of the district; between-field variability is close in magnitude to the permissible error of laboratory humus determination. The cross-validated coefficient of determination did not exceed 0.055 (under random cross-validation, Ridge −0.067 to 0.016, Random Forest −0.107 to 0.055), and the reduction in RMSE relative to a prediction by the mean did not exceed 0.010 %, several times smaller than the laboratory error; hence, the predictive ability of the spectral features at the field level was not confirmed. The agreement between mean predicted and laboratory values by survey round is trivial: a naive model reproduces it equally well, and the models do not reproduce changes at the level of individual fields. The results characterise the limitations of the index-based approach for the data and humus range used, not the fundamental impossibility of remote humus estimation; the main reason is the narrow between-field variation of humus content, comparable to the error of the laboratory method. For monitoring soil humus status, remote sensing data in this setting can play only an auxiliary role, while laboratory determinations remain indispensable.

Keywords: soil humus, humus status, agroecosystems, remote sensing, Sentinel-2, Landsat, bare soil, field boundary delineation, Google Earth Engine, spectral indices, cross-validation

Отримано редколегією 02.09.2026 р.
Підписано до друку 15.10.2026 р.
Дата публікації 31.10.2026 р.

ORCID ID
Дмитро Робуляк: https://orcid.org/0009-0001-0816-0151
Іван Смага: https://orcid.org/0000-0002-9000-3832
Михайло Гунчак: https://orcid.org/0000-0002-3521-8531