

МАТЕМАТИЧНІ МОДЕЛІ ОЦІНЮВАННЯ ЩІЛЬНОСТІ БІОМАСИ ЛІСІВ УКРАЇНИ НА ОСНОВІ ГЕОПРОСТОРОВИХ ДАНИХ

О. В. Матушевський^{1,2,a}, А. Ю. Шелестов¹

¹ Навчально-науковий Фізико-технічний інститут

² Anhalt University of Applied Sciences

Анотація

Представлена методика оцінювання щільності надземної біомаси (AGBD) лісів України на основі даних лідарної місії GEDI L4A, мультиспектральних знімків Sentinel-2 та радарних даних Sentinel-1 з використанням алгоритму Random Forest. Порівнюються два підходи до формування ознак: прямі спектральні та радарні предиктори (S1S2) і латентні вектори представлень на основі супутникових ембедингів Google (EMB). Підхід EMB показує значно вищу точність: $R^2 = 0,71$ проти $R^2 = 0,52$ для S1S2 на загальній вибірці та $R^2 = 0,53$ проти $R^2 = 0,20$ на лісовій підвибірці. За отриманими моделями побудовано карти AGBD лісів України з просторовим розрізненням 10 м за 2020 — 2025 рр.

Ключові слова: надземна біомаса лісу, GEDI, Sentinel-1, Sentinel-2, Random Forest, ембединги, дистанційне зондування, Google Earth Engine.

Вступ

Ліси відіграють ключову роль у глобальному вуглецевому циклі. Близько 50% сухої маси деревини — це вуглець, тому точна карта щільності надземної біомаси (Aboveground Biomass Density, AGBD) є основою для оцінки вуглецевих запасів і виконання зобов'язань країни у сфері кліматичної звітності [1, 2].

Збройна агресія Росії з 2022 року завдала безпрецедентних пошкоджень лісовому фонду України: пожежі від обстрілів, механічне пошкодження стовбурів і крон, несанкціоновані вирубки на окупованих територіях [3]. При цьому значна частина постраждалих лісів недоступна для польових досліджень — через активні бойові дії або мінне забруднення. За таких умов єдиним надійним інструментом оцінки стану і втрат лісових екосистем залишаються методи дистанційного зондування [4].

Космічна лідарна місія GEDI (Global Ecosystem Dynamics Investigation) отримує вертикальні профілі лісового намету по дискретних треках і надає оцінки AGBD для кожного окремого вимірювання (footprint) діаметром 25 м [1]. Подібний принцип уже використовувався для побудови глобальних карт висоти лісового намету: вимірювання GEDI поєднувалися з архівами Landsat або іншими супутниковими ознаками, після чого модель поширювала точкові вимірювання на суцільне растрове покриття [5, 6]. Щоб перетворити точкові спостереження GEDI на суцільну карту біомаси, їх поєднують із щільними

растровими предикторами Sentinel-1 і Sentinel-2 через методи машинного навчання [7, 8, 9].

Для України подібні дослідження досі не проводились, що й визначає наукову новизну та практичну цінність роботи.

Мета роботи — побудувати та порівняти два підходи до оцінювання AGBD лісів України з даних GEDI і Sentinel, а також створити просторово неперервні карти AGBD за 2020 — 2025 рр.

1. Дані та матеріали

1.1. Лідарні дані GEDI (цільова змінна)

Продукт **GEDI L4A** (версія 2.1) містить оцінки AGBD [Мг/га] для кожного footprint-вимірювання, отримані на основі повнохвильових лідарних сигналів та алометричних рівнянь [2]. Спостереження відбираються за двома критеріями якості: прапорець якості алгоритму (quality flag = 1) та правило 2σ для видалення аномальних значень. Після фільтрації навчальна вибірка налічує $N_{\text{train}} = 28\,657$ спостережень по всій лісовій зоні України.

1.2. Мультиспектральні дані Sentinel-2

Sentinel-2 надає 13 спектральних каналів із просторовим розрізненням 10–20 м. Для кожного вегетаційного сезону (червень — серпень) будується безхмарний медіанний композит за допомогою алгоритму маскування хмар s2cloudless у середовищі Google Earth Engine (GEE). Ця хмарна платформа зручна для обробки великих багаторічних архівів

^amatush-mm24@lil.kpi.ua

супутникових зображень і вже активно використовуються для задач класифікації та моніторингу земного покриття [10]. Серед використаних каналів — видимий спектр (B2, B3, B4), ближній інфрачервоний (B8), red-edge (B5, B6, B7, B8A) та SWIR (B11, B12). На їх основі обчислюються вегетаційні індекси за формулою (1):

$$\begin{aligned} \text{NDVI} &= \frac{B8 - B4}{B8 + B4}, \\ \text{EVI} &= 2,5 \cdot \frac{B8 - B4}{B8 + 6 \cdot B4 - 7,5 \cdot B2 + 1}. \end{aligned} \quad (1)$$

1.3. Радарні дані Sentinel-1

Sentinel-1 надає дані синтетичної апертури (SAR) у двох поляризаціях: VV (вертикальна — вертикальна) та VH (вертикальна — горизонтальна). Медіанні композити формуються для восьми рівномірних часових зрізів вегетаційного сезону, що дозволяє врахувати сезонну динаміку зворотного розсіювання. Хвилі С-діапазону частково проникають у намет і несуть інформацію про об'єм лісу, працюючи незалежно від хмарності [11].

2. Методологія

2.1. Математична постановка задачі

Нехай $D \subset \mathbb{R}^2$ — досліджувана область, що відповідає лісовому покриттю України. Кожній просторовій точці $s \in D$ відповідає вектор p предикторних ознак:

$$\mathbf{x}(s) = (x_1(s), x_2(s), \dots, x_p(s))^T \in \mathbb{R}^p,$$

де $x_i(s)$ — значення i -го предиктора (спектральний канал або SAR-поляризація) у точці s .

Цільова змінна — щільність надземної біомаси:

$$y(s) = \text{AGBD}(s) \text{ [Мг/га]},$$

яка є неспостережуваною неперервною просторовою функцією на D . Безпосереднє вимірювання $y(s)$ можливе лише у дискретних точках $\{s_i\}_{i=1}^N$ — footprint-вимірюваннях місії GEDI. Навчальна множина:

$$\mathcal{D} = \{(s_i, y_i)\}_{i=1}^N, \quad N_{\text{train}} = 28\,657,$$

де y_i — значення AGBD із продукту GEDI L4A у точці s_i .

Необхідно побудувати відображення $f: \mathbb{R}^p \rightarrow \mathbb{R}$ таке, що для довільної точки $s \in D$:

$$\hat{y}(s) = f(\mathbf{x}(s)) \approx y(s), \quad [\text{Мг/га}]. \quad (2)$$

Задача (2) є нелінійною багатофакторною регресією. Складність і просторова нестационарність залежностей між супутниковими предикторами і біомасою унеможливує застосування параметричних лінійних моделей, що обумовлює використання ансамблевих методів машинного навчання. Отримана функція f застосовується до суцільного растрового покриття всієї лісової зони України для отримання просторово неперервної карти AGBD.

2.2. Два підходи до формування ознак

Підхід S1S2. Простір предикторів формується з $p = 21$ ознак. П'ять спектральних каналів Sentinel-2 (B2, B3, B4, B8, B8A) агреговані у безхмарний медіанний композит вегетаційного сезону разом із вегетаційними індексами за формулою (1). SAR-стек містить 16 предикторів — значення зворотного розсіювання σ_{VV}^0 і σ_{VH}^0 для восьми рівномірних часових зрізів ($k = 1, \dots, 8$):

$$\mathbf{x}^{S1}(s) = (\sigma_{VV,k}^0(s), \sigma_{VH,k}^0(s))_{k=1}^8 \in \mathbb{R}^{16}.$$

Отже, повний вектор ознак підходу S1S2: $\mathbf{x}(s) \in \mathbb{R}^{21}$.

Підхід EMB. Вектор ознак — 64-вимірні латентні представлення з набору супутникових ембедингів Google Satellite Embedding V1. Це ознаки, попередньо навчені методами самонавчання на глобальних мультисенсорних супутникових архівах; вони кодують просторово-спектральну структуру сцени у компактному латентному просторі, недоступному при прямій обробці спектральних каналів. До них додається часова змінна day, тож $\mathbf{x}(s) \in \mathbb{R}^{65}$ ($p = 65$).

2.3. Random Forest та логарифмічна трансформація

Обидві моделі навчаються методом Random Forest — ансамблю $M = 100$ дерев рішень, кожне з яких будується на бутстрап-вибірці з $\mathcal{D}$ із випадковим вибором підмножини ознак у кожному вузлі розгалуження. Прогноз ансамблю:

$$\hat{F}(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^M T_m(\mathbf{x}),$$

де T_m — m -те дерево регресії. Усереднення по бутстрап-вибірках знижує дисперсію прогнозу без збільшення зміщення, що є ключовою перевагою при нелінійних залежностях із шумними предикторами.

Цільова змінна логарифмічно трансформується: $\tilde{y} = \log(1 + y)$, що стабілізує дисперсію і зменшує вплив викидів на функцію втрат. Прогноз відновлюється оберненим перетворенням: $\hat{y} = \exp(\hat{\tilde{y}}) - 1$.

Для коректної оцінки узагальнювальної здатності застосована просторово блокова стратегія розбиття даних: навчальна вибірка формується з трьох просторово рознесених блоків, тестова ($N_{\text{test}} = 5\,000$) — з четвертого незалежного блоку. Це усуває вплив просторової автокореляції, яка систематично завищує R^2 при звичайному випадковому розбитті [12].

Метрики якості: коефіцієнт детермінації R^2 , середньоквадратична похибка RMSE та нахил лінійної регресії спостережених відносно передбачених значень.

3. Результати

3.1. Точність моделей

Зведені метрики точності обох моделей на незалежній тестовій вибірці наведено в таблиці 1.

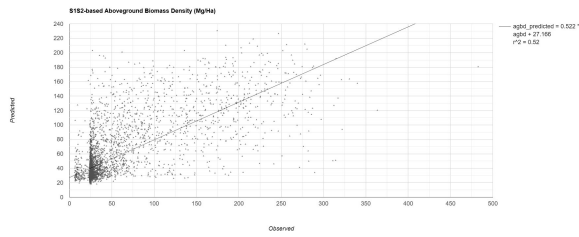
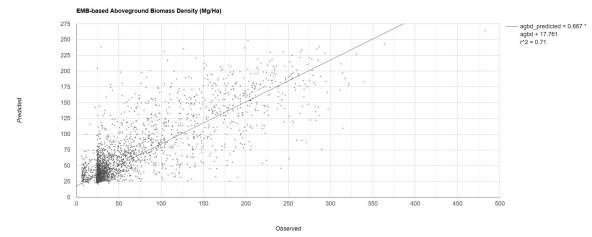
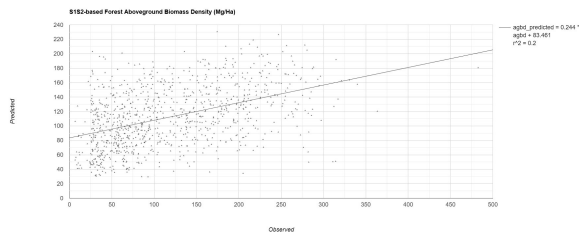
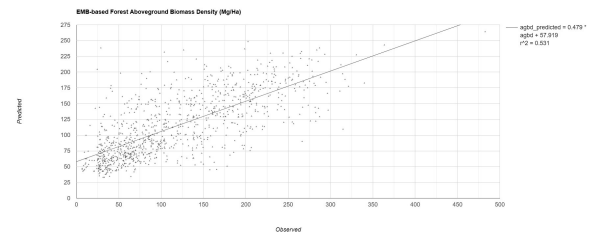
(а) S1S2, загальна вибірка ($R^2=0,52$).(б) EMB, загальна вибірка ($R^2=0,71$).(в) S1S2, лісова підвибірка ($R^2=0,20$).(г) EMB, лісова підвибірка ($R^2=0,53$).

Рис. 1. Діаграми розсіювання: спостережені проти передбачених значень AGBD для моделей S1S2 та EMB на загальній (а, б) і лісовій (в, г) вибірках.

Таблиця 1. Метрики точності моделей S1S2 та EMB на тестовій вибірці

Модель	Вибірка	R^2	RMSE, Мг/га	Нахил
S1S2	Загальна	0,52	36,64	0,522
S1S2	Ліс	0,20	68,13	0,244
EMB	Загальна	0,71	28,48	0,667
EMB	Ліс	0,53	52,07	0,479

Модель EMB перевершує S1S2 за всіма метриками. На загальній вибірці R^2 зростає з 0,52 до 0,71, RMSE знижується на 22% — з 36,64 до 28,48 Мг/га. Для лісових пікселів різниця ще суттєвіша: R^2 зростає з 0,20 до 0,53 (+0,33), а RMSE — з 68,13 до 52,07 Мг/га (–24%).

3.2. Аналіз систематичних відхилень

Діаграми розсіювання (рис. 1) виявляють дві характерні закономірності, спільні для обох моделей і обох підвбірок.

Завищення при малих AGBD. При значеннях нижче 50 Мг/га (молодняки, рідколісся) обидві моделі суттєво переоцінюють біомасу. Причина — неоднорідний спектральний сигнал, де частка підстильної поверхні висока.

Заниження при великих AGBD. При значеннях понад 150 Мг/га обидві моделі систематично занижують прогноз — це ефект спектрального насичення: оптичні та SAR-предиктори С-діапазону втрачають чутливість до структури намету у щільних стиглих лісах [13]. Нахил регресії значно менший за одиницю (0,522 для S1S2 і 0,667 для EMB

на загальній вибірці), що кількісно підтверджує цей зсув.

На лісовій підвибірці (рис. 1в, г) модель S1S2 з $R^2 = 0,20$ і нахилом 0,244 по суті видає постійне значення в діапазоні 80–140 Мг/га. Підхід EMB ($R^2 = 0,53$, нахил 0,479) значно кращий: латентні ознаки кодують інформацію про структуру лісу, недоступну у прямих спектральних каналах.

3.3. Карти AGBD лісів України

На тренуванні моделі застосовані до суцільного растрового стеку ознак для всієї лісової зони України. Побудовано щорічні карти AGBD з просторовим розрізненням 10 м за 2020 — 2025 рр. в середовищі GEE (рис. 2).

Карпати мають найвищі значення AGBD (> 150 Мг/га) — тут переважають стиглі букові та ялицеві ліси. **Полісся** вирізняється однорідними значеннями 80–130 Мг/га, характерними для соснових і мішаних рівнинних лісів. **Лісостепова і степова зони** мають нижчі значення (30–80 Мг/га) — через фрагментованість і переважання молодих насаджень.

Карта EMB-моделі візуально більш диференційована і краще відображає межі між лісовими типами, тоді як S1S2-модель дає більш однорідний і «розмитий» розподіл.

Висновки

- Підхід EMB на основі супутникових ембедингів Google забезпечує значно вищу точність ($R^2 = 0,71$, RMSE = 28,48 Мг/га) порівняно з підходом S1S2 ($R^2 = 0,52$, RMSE = 36,64 Мг/га).
- Для лісових пікселів перевага EMB ще більша: $R^2 = 0,53$ проти $R^2 = 0,20$. Прямі спектраль-



(а) Модель S1S2.



(б) Модель EMB.

Рис. 2. Карти AGBD лісів України за 2025 рік: моделі S1S2 (а) та EMB (б).

ні предиктори не дозволяють диференціювати стиглі ліси за щільністю біомаси.

- Обидві моделі систематично занижують AGBD при значеннях понад 150 Мг/га через ефект спектрального насичення — системне обмеження предикторів С-діапазону.
- Побудовано серію карт AGBD лісів України для 2020 — 2025 рр., що відображають просторово диференційований розподіл біомаси: максимум — у Карпатах, мінімум — у степовій зоні.

Перспективи розвитку — залучення предикторів L-діапазону (ALOS PALSAR-2) для подолання насичення та аналіз динаміки AGBD на воєнно-постраждалих територіях без фізичного доступу.

Перелік використаних джерел

- Multi-resolution gridded maps of vegetation structure from GEDI / R. Dubayah, J. Armston, S. P. Healey, Z. Yang, P. L. Patterson, S. Saarela, G. Ståhl, Q. Du, L. Duncanson, S. Hancock [та ін.] // Scientific Data. — 2022. — Т. 9, № 1. — С. 351. — DOI: [10.1038/s41597-022-01474-3](https://doi.org/10.1038/s41597-022-01474-3).
- Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission / L. Duncanson, J. R. Kellner, J. Armston, R. Dubayah, D. M. Minor, S. Hancock, S. P. Healey, P. L. Patterson, S. Saarela, S. Marselis

[та ін.] // Remote Sensing of Environment. — 2022. — Т. 270. — С. 112845. — DOI: [10.1016/j.rse.2021.112845](https://doi.org/10.1016/j.rse.2021.112845).

- Decrease in forest above-ground biomass in war-damaged forests of Ukraine: A case study using GEDI / V. Myroniuk, A. Bilous, M. Lesiv, D. Schepaschenko // Folia Forestalia Polonica, Series A – Forestry. — 2026. — DOI: [10.2478/ffp-2026-0004](https://doi.org/10.2478/ffp-2026-0004).
- Remote sensing of forest degradation: a review / Y. Gao, M. Skutsch, J. Paneque-Gálvez, A. Ghilardi // Environmental Research Letters. — 2020. — Т. 15, № 10. — С. 103001. — DOI: [10.1088/1748-9326/abaad7](https://doi.org/10.1088/1748-9326/abaad7).
- Mapping global forest canopy height through integration of GEDI and Landsat data / P. Potapov, X. Li, A. Hernandez-Serna, A. Tyukavina, M. C. Hansen, A. Kommareddy, A. Pickens, S. Turubanova, H. Tang, C. E. Silva [та ін.] // Remote Sensing of Environment. — 2021. — Т. 253. — С. 112165. — DOI: [10.1016/j.rse.2020.112165](https://doi.org/10.1016/j.rse.2020.112165).
- A high-resolution canopy height model of the Earth / N. Lang, W. Jetz, K. Schindler, J. D. Wegner // Nature Ecology & Evolution. — 2023. — Т. 7. — С. 1778—1789. — DOI: [10.1038/s41559-023-02206-6](https://doi.org/10.1038/s41559-023-02206-6).
- State-wide forest canopy height and aboveground biomass map for New York with 10 m resolution, integrating GEDI / H. Tamiminia, B. Salehi, M. Mahdianpari, T. Goulden // Ecological Informatics. — 2024. — Т. 79. — С. 102404. — DOI: [10.1016/j.ecoinf.2023.102404](https://doi.org/10.1016/j.ecoinf.2023.102404).
- Integrating spaceborne LiDAR and Sentinel-2 images to estimate forest aboveground biomass in Northern China / S. Luo, J. Liu, C. Zhang, J. Tao, X. Lou, X. Chen // Carbon Balance and Management. — 2022. — Т. 17, № 1. — С. 15. — DOI: [10.1186/s13021-022-00212-y](https://doi.org/10.1186/s13021-022-00212-y).
- Fusing GEDI with earth observation data for large area aboveground biomass mapping / S. Saarela, A. Wästlund, S. Holm, A. Grafström, A. A. Mensah, M. Nilsson, M. Holmberg, G. Ståhl // International Journal of Applied Earth Observation and Geoinformation. — 2022. — Т. 115. — С. 103108. — DOI: [10.1016/j.jag.2022.103108](https://doi.org/10.1016/j.jag.2022.103108).
- Exploring Google Earth Engine platform for big data processing: classification of multi-temporal satellite imagery for crop mapping / A. Shelestov, M. Lavreniuk, N. Kussul, A. Novikov, S. Skakun // Frontiers in Earth Science. — 2017. — Т. 5. — С. 17. — DOI: [10.3389/feart.2017.00017](https://doi.org/10.3389/feart.2017.00017).
- Estimation of Aboveground Biomass for Different Forest Types Using Data from Sentinel-1, Sentinel-2, ALOS PALSAR-2, and GEDI / Z. Zhang, M. Liu, X. Liu, G. Zhou // Forests. — 2024. — Т. 15, № 1. — С. 215. — DOI: [10.3390/f15010215](https://doi.org/10.3390/f15010215).
- Subedi P. B., Zurqani H. A. Multi-sensor forest aboveground biomass estimation using GEDI //

- Advances in Space Research. — 2026. — Т. 77, № 3. — С. 2784—2806. — DOI: [10.1016/j.asr.2025.12.032](https://doi.org/10.1016/j.asr.2025.12.032).
13. Wang J., Xiang C., Liang A. Estimation of Forest Aboveground Biomass in China Based on GEDI and Sentinel-2 Data: Quantitative Analysis of Optical Remote Sensing Saturation Effect and Terrain Compensation Mechanisms // Remote Sensing. — 2025. — Т. 17, № 20. — С. 3437. — DOI: [10.3390/rs17203437](https://doi.org/10.3390/rs17203437).
- Роботу виконано в рамках проекту Horizon Europe «FUTUREFOR — Застосунки Copernicus для моніторингу лісів наступного покоління» (грант № 101180278), а також у межах проекту «Сервіси Copernicus наступного покоління: інтелектуальний цифровий двійник для аналізу лісового покриття в Україні» (наказ Міністерства освіти і науки України від 09.01.2026 № 23).*