УДК 004.8:336.76

# УДОСКОНАЛЕНИЙ ІНТЕРПОЛЯЦІЙНИЙ МЕТОД НА ОСНОВІ КОРЕЛЯЦІЙНИХ ВАГ ДЛЯ ЗАПОВНЕННЯ ПРОБІЛІВ ВІД ХМАР У AOD КАНАЛІ НА ЩОДЕННИХ ЗНІМКАХ MODIS

О. А. Кравчук[1,a], Г. О. Яйлимова[1,2]

[1] Навчально-науковий Фізико-технічний інститут

[2] Інститут космічних досліджень НАНУ та ДКАУ

## Анотація

В наш час існує багато потужних моделей глибинного навчання (моделей на основі ConvLSTM, автоенкодерів, дифузійних моделей) для реконструкції або прогнозування даних про аерозольну оптичну глибину (англ. Aerosol optical depth, далі – AOD), проте такі моделі часто потребують непошкоджених даних для навчання. В свою чергу, якщо для екваторіальних чи невеликих регіонів безхмарні знімки знайти можна, то для території у сотні тисяч квадратних кілометрів, наприклад, у Європі ця задача стає неможливою. Відповідно, пропуски від хмар необхідно заповнювати. Для вирішення цієї задачі був реалізований вдосконалений метод інтерполяції на основі кореляційних ваг (англ. Interpolation-based correlation weighting, далі – ICW).

**Ключові слова**: AOD, ICW, smoothing spline, IDW (interpolation distance weighting)

## Вступ

Аерозольна оптична глибина (AOD) є ключовим параметром для вимірювання якості повітря та моніторингу кліматичних змін. Хоча наземні станції AERONET забезпечують еталонну точність вимірювань AOD, вони обмежені своєю точковою природою та не можуть покрити всю територію земної кулі. З цієї причини використання даних AOD, отриманих із супутникових знімків, наприклад з інструментів на борту MODIS (Moderate Resolution Imaging Spectroradiometer), стає все більш популярним. Супутникові дані забезпечують цілісне і безперервне покриття Землі, дозволяючи отримати унікальну можливість для глобального моніторингу аерозольних оптичних властивостей атмосфери.

Майже всі супутникові фотографії AOD стикаються з викликом пробілів, спричинених хмарністю, що значно ускладнює, а інколи навіть робить неможливим використання цих даних для тренування глибинних нейромереж. Ці прогалини в даних не тільки знижують якість загального зображення, але й ускладнюють точне прогнозування чи реконструкцію AOD за допомогою алгоритмів машинного навчання. В результаті, перш ніж набір даних може бути використаний для тренування глибинних нейромереж, необхідно ефективно заповнити ці пробіли від хмар. Це вимагає розробки та застосування складних методів реконструкції зображень, які можуть відновлювати цілісність супутникових знімків AOD, дозволяючи сформувати високоякісний і повний набір даних для подальшого аналізу та навчання моделей.

[a]kravchuk.oleksandr@lll.kpi.ua

## 1. Аналіз існуючих підходів та їх недоліків

На сьогоднішній день у науковому співтоваристві розглядаються два основних підходи для вирішення проблеми пробілів від хмар у супутникових зображеннях AOD.

Перший підхід полягає у використанні додаткових даних для підвищення точності реконструкції. Це може включати інформацію про AOD з інших супутників, виміри AOD, отримані з наземних станцій AERONET, а також дані з наземних синоптичних станцій, такі як температура, вологість, швидкість вітру та ін. Шляхом обчислення кореляційних залежностей між різними джерелами даних можливо реконструювати AOD зображення, використовуючи всю доступну інформацію одночасно. Ця стратегія була продемонстрована у першій частині двоступеневої моделі [1], реалізованої за допомогою LightGBM.

В той час як перший підхід намагається інтегрувати максимальну кількість джерел даних, другий метод зосереджується на використанні лише одного типу даних – супутникових знімків. Цей підхід ґрунтується на обчисленні і використанні просторово-часових залежностей, характерних саме для супутникових даних про AOD з одного джерела, наприклад, з MODIS. Розробка і аплікація цього методу забезпечують можливість реконструкції зображень AOD, виходячи зі спостережень, що здійснюються протягом певного періоду часу, що дозволяє ефективно враховувати тимчасові зміни в атмосфері. Цей підхід, що став основою для цієї роботи, був представлений у дослідженні [2].

Методики тестування і оцінки точності моделей, які використовуються для заповнення пробілів від хмар у супутникових зображеннях AOD, також різняться.

Перший підхід, представлений у дослідженні [2], полягає у порівнянні прогнозованих карт AOD з інтерпольованими даними, отриманими з наземних станцій, таких як AERONET. Цей метод дозволяє оцінити точність реконструкції AOD зображень шляхом порівняння з високоякісними, точними даними з наземних спостережень, проте втративши точність на етапі інтерполювання.

У контрасті, другий підхід, описаний у дослідженні [3], використовує методику штучного створення пробілів від хмар на якісних супутникових знімках. Після цього «зіпсовані» дані подаються на вхід моделі, і результати порівнюються з оригінальними знімками, щоб оцінити ефективність моделі у відновленні первинної інформації. Для створення штучних пробілів часто використовуються випадково модифіковані бінарні маски хмар, отримані з інших супутникових зображень, шляхом їх розтягування, зсуву та перевертання. Недоліком цього методу є порівняння результату роботи моделі з завідомо неідеальними супутниковими даними.

## 2. Набір використаних даних та пілотна територія

Як джерело даних для навчання і тестування моделей використовувалася канал AOD 0.55$\mu$m із супутникової колекції MCD19A2 Version 6.1, NASA MODIS [4]. Ця колекція є найновішим продуктом NASA в області оцінки поверхневого AOD. Вона поєднує спостереження з двох інших супутників Terra і Aqua за допомогою нового алгоритму MAIAIC (Multi-angle Implementation of Atmospheric Correction).

Навчальний набір даних був сформований із щоденних знімків з колекції MCD19A2 протягом 2023 року на території Франції (рис. 1). Зона інтересу мала квадратну форму, де лівий правий кут розташований за координатами (49.9875, 0.0193), а правий нижній – (45.2125, 6.7957). Просторове розрізнення 3км, розмір вихідного зображення 192 × 192 пікселі.

В якості тестового набору даних використовувалися випадковим чином повернуті та зміщені бінарні маски хмар, отриманих з тієї ж супутникової колекції за період 2022 року, які було накладено на найбільш непошкоджені знімки. Просторове розрізнення і інші параметри такі ж, як і в начальному наборі даних.

В якості потенційного навчального набору даних для майбутніх моделей машинного/глибинного навчання використовувалися реконструйовані моделлю знімки з початковим відсотком хмар не більше 10%.

## 3. Розроблений метод відновлення даних

Метод, описаний у цій роботі, представляє собою модифікацію підходу, вперше запропонованого в [2],

Рис. 1. Досліджувана територія

звідки використання згладжуючих сплайнів для визначення щоденних відхилень AOD від норми та ідея інтерполяції даних на основі усереднених значень у «блоці» були запозичені. Основною новацією є збільшення кількості врахованих сусідніх «блоків» для кожного пікселя з одного до будь-якої менше восьми (кількість сусідніх блоків пікселя).

### 3.1. Заміна AOD на його відхилення від норми

З метою враховувати не тільки просторову, а й часову складову при заповненні пробілів від хмар використовувався наступний прийом. Значення AOD для кожного пікселя і кожного дня року замінювалося на різницю поточного значення AOD і значення сплайну [5] в цій просторово-часовій точці (рис. 2). Таким чином в отриманому наборі даних кожна точка представляла собою відхилення AOD від норми (рис. 3).

Для побудови сплайну кількість спостережень для точки за рік має бути не менше 10. В іншому випадку, в кінці алгоритму значення цих точок замінювалося на середнє у вікні 9 × 9 навколо них.

### 3.2. Інтерполяція

Подальший алгоритм заповнення пропусків від хмар можна описати наступним чином:

1. Ділимо зображення 192 × 192 пікселі на сітку 24 × 24 блока (фактор ділення 8). Кожний блок це середнє значення по вікну 8 × 8 на оригінальному зображенні. Якщо у вікні 8 × 8 немає жодного значення, то значення блоку рахується по методу Interpolation Distance Weighting від інших непустих блоків:

$$block_{a,b} = \frac{1}{\sum_{\text{block}_{i,j} \neq NaN} \frac{1}{d_{i,j}}} \sum_{block_{i,j} \neq NaN} \frac{\text{block}_{i,j}}{d_{i,j}},$$

де $d$ – евклідова відстань між $(a, b)$ і $(i, j)$.

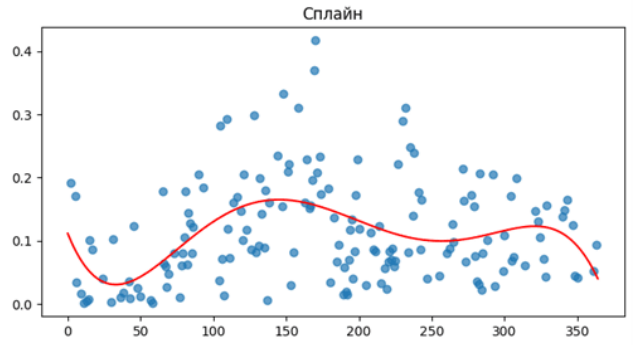


Рис. 2. Значення AOD для конкретного пікселю протягом року

Рис. 3. Похибка сплайну

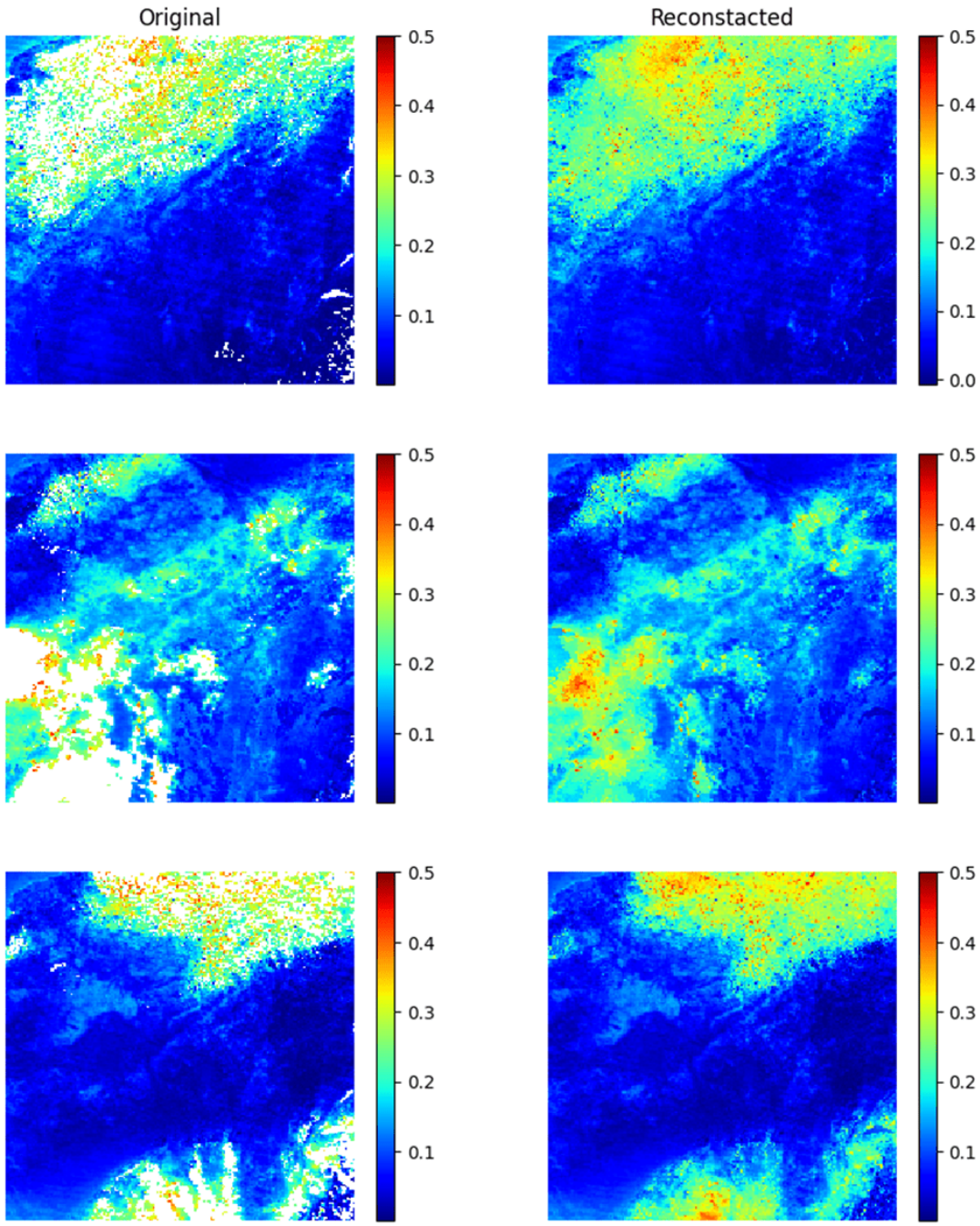


Рис. 4. Оригінальні і реконструйовані супутникові зображення

2. Далі знаходимо коефіцієнт кореляції Пірсона між кожним елементом оригінальної мапи AOD зі своїм і з кожним сусіднім до елемента блоком. Далі необхідно обрати $n$ блоків з найбільшою кореляцією (у цьому дослідженні використовувалося $n = 2$), з використанням яких надалі будуть заповнюватися пробіли.
3. Кожен невідомий елемент супутникового зображення заповнюється поєднанням метода ICW і лінійної регресії:

$$aod_{i,j} = \alpha_{i,j} + \beta_{i,j} \times \left( \sum_{k=1}^{n} \mathrm{cor}(aod_{i,j}, \mathrm{block}_k) \times \mathrm{block}_k \right),$$

де $\alpha_{i,j}$ і $\beta_{i,j}$ – це коефіцієнти лінійної регресії для кожного елемента мапи.
4. Як було описано вище, точки, які були відсіяні на етапі побудови сплайну, замінюються на середнє у вікні $9 \times 9$ навколо них.

## 4. Отримані результати

В результаті роботи моделі, описаної в попередньому розділі, вдалося заповнити пробіли від хмар з достатньою точністю для подальшого навчання нейромереж та інших моделей машинного навчання (рис. 4).

Точність моделі оцінювалася наступним чином. Обираються зображання з найменшою кількістю хмар і штучно спотворюється за допомогою бінарних масок інших хмар. Коректуємо бінарну маску таким чином, щоб її хмари не накладалися на хмари у зображені (іншими словами, щоб пікселі хмар бінарної маски були тільки на безхмарних ділянках зображення). Далі спотворене зображення подається на вхід моделі і порівнюється з оригінальним лише в тих місцях, де були хмари у зкоректованій бінарній масці. В якості метрик був обраний коефіцієнт кореляції і MAE.

$$R = \frac{\sum_{i,j \in \mathrm{mask}} (aod_{i,j}^{\mathrm{real}} - \overline{aod}^{\mathrm{real}}) \times (aod_{i,j}^{\mathrm{pred}} - \overline{aod}^{\mathrm{pred}})}{\sqrt{\sum_{i,j \in \mathrm{mask}} (aod_{i,j}^{\mathrm{real}} - \overline{aod}^{\mathrm{real}})^2 \times \sum_{i,j \in \mathrm{mask}} (aod_{i,j}^{\mathrm{pred}} - \overline{aod}^{\mathrm{pred}})^2}}$$

$$\mathrm{MAE} = \frac{1}{N} \sum_{i,j \in \mathrm{mask}} |aod_{i,j}^{\mathrm{real}} - aod_{i,j}^{\mathrm{pred}}|$$

$$\mathrm{MRE} = \frac{1}{N} \sum_{i,j \in \mathrm{mask}} \frac{|aod_{i,j}^{\mathrm{real}} - aod_{i,j}^{\mathrm{pred}}|}{aod_{i,j}^{\mathrm{real}}}$$

При рівні хмар до 10% можна досягти таких показників: MAE $\approx$ 0,02; MRE $\approx$ 0,33; R $\approx$ 0,77. Оскільки при створенні реального тренувального набору даних для майбутньої глибинної моделі будуть обиратись менш пошкоджені знімки, значення метрик може значно покращитись.

## Перелік використаних джерел


1. Handling missing data in large-scale MODIS AOD products using a two-step model / Y. Chi, Z. Wu, K. Liao, Y. Ren // Remote Sensing. — 2020. — Т. 12, № 22. — С. 3786. — URL: https://doi.org/10.3390/rs12223786.
2. Gap-filling MODIS daily aerosol optical depth products by developing a spatiotemporal fitting algorithm / T. Zhang, Y. Zhou, K. Zhao, Z. Zhu, G. R. Asrar, X. Zhao // GIScience & Remote Sensing. — 2022. — Т. 59, № 1. — С. 762—781. — URL: https://doi.org/10.1080/15481603.2022.2060596.
3. *Daniels J., Bailey C. P., Liang L.* Filling Cloud Gaps in Satellite AOD Retrievals Using an LSTM CNN-Autoencoder Model // 2022 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). — 2022. — С. 2758—2761. — URL: https://doi.org/10.1109/IGARSS46834.2022.9884482.
4. EarthData, NASA. — URL: https://search.earthdata.nasa.gov/portal/idn/search?fi=MODIS.
5. A review of spline function procedures in R / A. Perperoglou, W. Sauerbrei, M. Abrahamowicz, M. Schmid // BMC Medical Research Methodology. — 2019. — Т. 19. — С. 1—16. — URL: https://doi.org/10.1186/s12874-019-0666-3.