

## ОПТИМІЗАЦІЯ ПРОЦЕСІВ ОБРОБКИ ВЕЛИКИХ ДАНИХ ІЗ ЗАСТОСУВАННЯМ МЕТОДІВ РОЗПАРАЛЕЛЮВАННЯ

Л.Л. Шуміло

Інститут космічних досліджень НАН України та ДКА України

*shumilo.leonid@gmail.com*

**Вступ.** Питання оптимізації процесів обробки великих даних завжди виникають при роботі з ними через велику кількість ресурсів та часу, потрібних для цих процесів. Одним із прикладів таких процесів є побудова часового ряду супутникових знімків, необхідних для розв'язання багатьох прикладних задач таких як класифікація сільськогосподарських посівів, прогнозування врожайності та інші [1-3]. Тому виникає необхідність, враховуючи кількість супутникових даних за останні роки та час їх обробки для складання часового ряду, оптимізувати використання обчислювальних ресурсів та пришвидшити цей процес.

**Постановка задачі та її розв'язок.** В даній роботі запропоновані способи оптимізації процесу побудови часового ряду у програмах написаних на мовах python та C# за допомогою розпаралелювання, що дозволяють здійснювати роботу з кількома каналами растрів одночасно. Для даного експерименту було використано 10 супутникових знімків Sentinel-1 за 2016 рік, та проаналізовано 3 способи розпаралелювання процесів на мові python: модуль `threading`, модуль `subprocess` та модуль `multiprocessing` та 2 способи розпаралелювання на мові `csharp`: модуль `threading` та `Parallel`. Найкращим з них виявився спосіб з використанням модулю `multiprocessing`. Якщо в середньому процедура створення часового ряду на мові python працює на даній тестовій вибірці при однакових умовах 4760 секунд, тоді при розбитті її на 4 процеси за допомогою модуля `multiprocessing` ми отримуємо середній час в 4286 секунд. Основною проблемою оптимізації створення часового ряду за допомогою розпаралелювання є велика кількість звернень кожного процесу до жорсткого диску, оскільки одночасно до нього може отримати доступ тільки один процес, тим не менш було отримано хороший результат, враховуючи вигравш в часі в середньому 473,5 секунд, що є 1/10 всього часу побудови часового ряду супутникових знімків.

**Результати.** Використовуючи запропонований спосіб розпаралелювання процесів за допомогою модуля `multiprocessing`, вдалось оптимізувати процес створення часового ряду супутникових знімків та пришвидшити його на 5 відсотків при розпаралелюванні на 2 процеси та на 10 відсотків при розпаралелюванні на 4 процеси. Використання часового ряду знімків дозволило покращити точність карти класифікації, отриманої за допомогою ансамблю нейронних мереж, більше ніж на 10 відсотків у порівнянні з класифікацією по одному знімку [2].

### References

1. Kussul N. Regional scale crop mapping using multi-temporal satellite imagery / N. Kussul, S. Skakun, A. Shelestov, M. Lavreniuk, B. Yailymov, O. Kussul // International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences. – 2015. – P. 45-52.
2. Kussul N. Grid technologies for satellite data processing and management within international disaster monitoring projects / N. Kussul, A. Shelestov, S. Skakun // Grid and Cloud Database Management. – 2011. – P. 279–305.
3. Lavreniuk M. Regional Retrospective High Resolution Land Cover For Ukraine: Methodology And Results / M. Lavreniuk, N. Kussul, S. Skakun, A. Shelestov, B. Yailymov. // International Geoscience and Remote Sensing Symposium (IGARSS). 2015 IEEE International – 2015. – P. 3965-3968. - DOI: 10.1109/IGARSS.2015.7326693.
4. Lavreniuk M. Validation Techniques for Land Cover and Land Use Maps // M. Lavreniuk, N. Kussul, A. Shelestov, B. Yailymov // Worldcover 2017 conference (14–16 March 2017 | ESA-ESRIN | Frascati (Rome), Italy) – 2016. – P. 47-48.