

І. КОТУЛЯК

Словацький технічний університет, Братислава, Словаччина.

В.В. ХИЛЕНКО

Національний університет біоресурсів і природокористування України, Київ, Україна,
e-mail: *vkhilenko@ukr.net*.

Р.М. БАСАРАБ

Національний університет біоресурсів і природокористування України, Київ, Україна.

М. РІЇС

Словацький технічний університет, Братислава, Словаччина.

ЗАСТОСУВАННЯ АЛГОРИТМІВ ДЕКОМПОЗИЦІЇ ДЛЯ ПРИСКОРЕННЯ ОБРОБЛЕННЯ ВЕЛИКИХ МАСИВІВ ДАНИХ У ГЕОІНФОРМАЦІЙНІЙ СИСТЕМІ

Анотація. Запропоновано технологію та декомпозиційний алгоритм прискорення оброблення геоінформаційних даних на основі розподілу вибірок динамічних та квазістатичних даних з використанням аналізу власних чисел матриць, отриманих за допомогою ітераційного обчислення за методом Хиленка. Алгоритм спрямовано на опрацювання масивів геоінформаційних даних великої розмірності. Наведено порівняльні результати модельних обчислень з використанням відомих обчислювальних методів.

Ключові слова: геоінформаційні дані, оброблення супутникових зображень, декомпозиція, ітераційне обчислення власних чисел матриць, метод Хиленка.

Розв'язання задач для аналізу інформації та прогнозування в геоінформаційних системах (ГІС) ускладнюється через роботу з великими та надвеликими масивами інформації. До того ж розв'язання багатьох типових задач оброблення супутникових зображень, зокрема задачі зворотного фотограмметричного засічення, стикаються з проблемою незадовільної зумовленості [1, 2]. Математична та обчислювальна складність у разі сукупності зазначених проблем потребує пошуку нових рішень для подальшого вдосконалення програмно-алгоритмічного забезпечення та підвищення якості роботи ГІС. Посилювання вимог до точності (детального опису) моделей передбачає отримання знімків, зроблених на різних довжинах хвиль електромагнітного спектра [3, 4], а необхідність отримання різних інформаційних характеристик геопросторових зон зумовлює кореляційний аналіз різних комбінацій каналів [5]. Все це зумовлює подальше збільшення розмірності та складності моделей. Отже, розвиток ГІС — систем оброблення супутникових знімків має тенденцію до подальшого ускладнення проблем, що розглядаються. У цій роботі запропоновано двокрокову технологію роботи математичних алгоритмів оброблення інформації в ГІС на підставі взаємопов'язаного розв'язання зазначених проблем із застосуванням математичного апарату системного аналізу, зокрема методу зниження порядку [6–8], для кластеризації та декомпозиції інформаційних моделей.

Для значної кількості задач оброблення геопросторової інформації динаміка змін в окремих геопросторових зонах відбувається із суттєво різною швидкістю, що дає змогу на першому етапі проводити аналіз та прогнозування динаміки виділених фрагментів на основі використання відповідних кластерних (декомпозиційних) моделей. Для переходу від повних математичних моделей до спрощених (декомпозиційних) моделей потрібно заздалегідь зазначити кластери, сформовані відповідно до вимог конкретних підзадач на основі

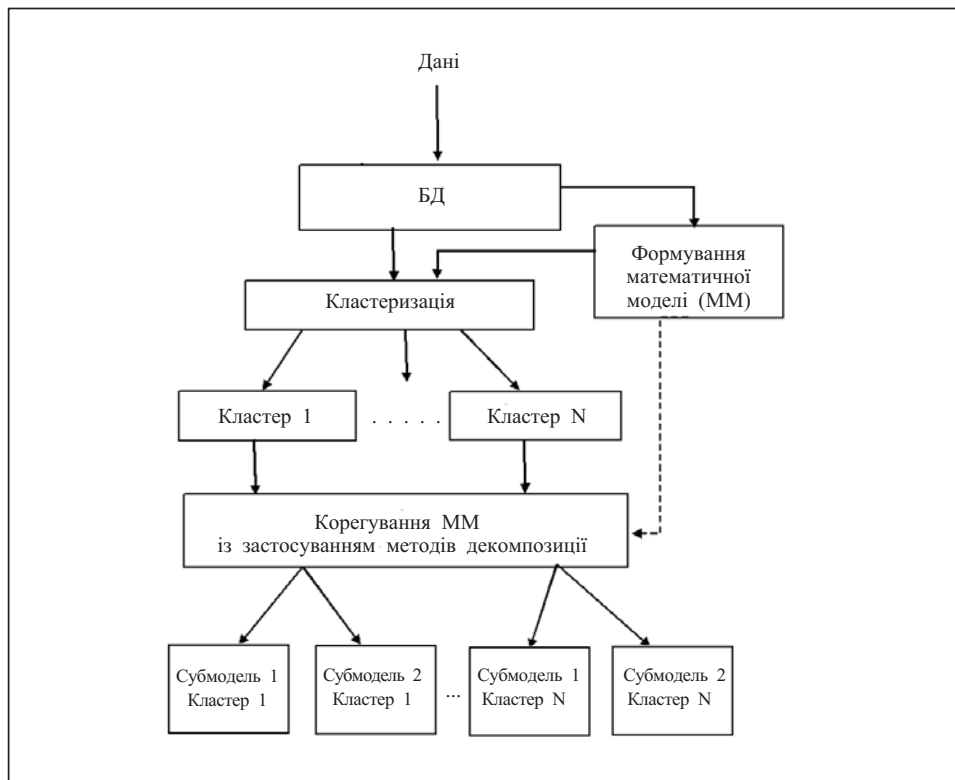


Рис. 1

фільтрації спектральної області. Зокрема, для розв'язання задач оброблення інформації за об'єктами з певними характеристиками, наприклад за заданими геометричними розмірами, потрібно визначення скінченної детермінованої множини даних об'єктів. У цьому випадку використання фільтрації в спектральній області дає змогу формалізувати процедуру визначення потрібних кластерів [3–5, 9]. Структурна схема процесу оброблення даних та побудови декомпозиційних моделей у цьому випадку матиме вигляд, представлений на рис. 1.

Зазначимо, що суттєва різниця в швидкості змін окремих параметрів (груп параметрів), змінних або складових змінних може мати місце і в межах однієї геопросторової зони для параметрів окремих пікселів (груп пікселів), що належать до інформаційного блоку. У визначених фрагментах геопросторових зон наявність параметрів, що швидко змінюються, може мати як природний характер, так і бути під впливом екзогенних факторів (природні катастрофи, діяльність людини і т.п.). Виявлення природного чи неприродного характеру змін параметрів геопросторової зони, необхідне для прийняття подальших адекватних управлінських рішень, можна здійснювати як на основі програм-аналізаторів зміни пікселів часових рядів [3–5], так і з використанням відповідних математичних моделей.

Прикладом фіксації наявності різношвидкісних процесів у окремих зонах (кластерах) під час використання програм-аналізаторів зміни пікселів часових рядів може бути порівняння послідовності знімків зміни ареолу лісових поверхонь [5]. Чисельну оцінку рознесеності швидкісних характеристик даних супутникових знімків, що досліджуються і мають чітке математичне представлення, запишемо у вигляді

$$\begin{aligned}
 \|U_k^j(t_1) - U_k^j(t_2)\| &\geq \varepsilon_1, \\
 \|U_l^i(t_1) - U_l^i(t_2)\| &\geq \varepsilon_2, \\
 \varepsilon_1 &\gg \varepsilon_2,
 \end{aligned} \tag{1}$$

де $U_m^n(t_q)$ — деяка l -та характеристика n -го пікселя, наприклад затемненість у момент часу t_q , а нормою може бути вибрано середньоквадратичне відхилення.

Формування кластерних субмоделей відповідно до співвідношення (1) є першим етапом спрощення та переходу до застосування математичного апарату декомпозиційних методів.

Використання ГІС для прийняття управлінських рішень передбачає наявність інформації про причини, що зумовлюють швидкі зміни в геопросторовій зоні. Питання про природну або неприродну швидку зміну параметрів геоінформаційної зони потребує додаткового аналізу. Формування бази даних протягом тривалого часу спостереження дає змогу за відсутності аналітичних залежностей використовувати методи регресійного аналізу та прогнозувати природні зміни динаміки поширення лісової зони, контролюючи її природний стан. Особлива увага натомість приділяється ділянці межі лісу та контролю її змін у часі. Зокрема, під час оцінювання стану мангрового лісу (Азія) зоною підвищеної уваги є межа області, що охороняється від незаконного вирубування лісу. Цю межу візуально чітко можна бачити на відповідному супутниковому знімку з чітким контрастом зони лісу і зони відсутності лісу [5]. На рис. 2 наведено зазначений фрагмент супутникового знімку. Фіксовані за допомогою геопросторового зондування послідовності дискретних даних дають змогу відобразити природний розвиток конкретної геопросторової зони і формують базу даних, необхідну для її аналізу та прогнозування динаміки у разі відсутності штучних (неприродних) факторів, що є причиною збурення. Використання програм-аналізаторів зміни пікселів часових рядів [3–5] створює основу формування моделей за допомогою методів регресійного аналізу [10, 11]. При цьому швидкі зміни межі лісової зони, що не відповідають прогнозним розрахункам, дають змогу терміново виявити проблемні ділянки незаконного вирубування. Подальше відстеження динаміки проблемної зони передбачає оперативне оновлення даних супутникових знімків і внесення відповідних змін до матричних математичних моделей, отриманих у разі лінеаризації вихідних нелінійних залежностей. Це поєднує характеристики параметрів геоінформаційної зони (змінних, які обчислюються). Математично зміни стану геопросторової зони, її окремих параметрів відображаються і легко виявляються за варіаціями елементів матриць, сформованих у лінеаризованих моделях нелінійних залежностей, закладених у моделях, що застосовують в ГІС.

Використання чіткого математичного підходу для виявлення характеру змін геопросторової зони і прогнозування її динаміки вимагає формування математичної моделі у вигляді

$$\dot{X} = f(x, t, \zeta), \quad (2)$$

де $x = (x_1, \dots, x_n)^T$ — вектор шуканих змінних, ζ — вектор випадкових параметрів за відомим законом розподілу.

Розв'язання значної кількості задач для аналізу супутникових знімків і прогнозування змін стану геопросторових зон передбачає додавання до загальної моделі опису процесів, що як безпосередньо, так і опосередковано впливають на зміну

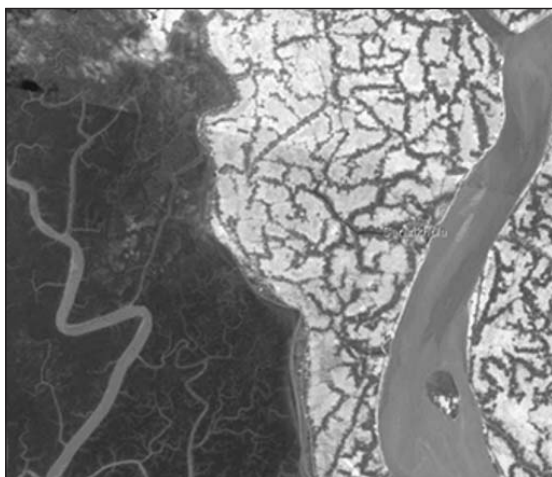


Рис. 2

характеристики зон. Так, для прогнозування стану просторової зони мангрових лісів потрібно враховувати такі взаємно корельовані фактори, як зміна рівня моря, солоність води, забрудненість довкілля, стан буферних зон тощо [5]. Прикладом незалежних від екосистеми чинників, що впливають на її стан, насамперед є результати діяльності людини.

У загальному випадку чітке формування вектора змінних стану моделі (1) виходить поза рамки цієї статті. Зазначимо лише складність як безпосереднього виконання завдання побудови моделей, що визначають динаміку екосистем, так і складність самих сформованих моделей. Адекватність цього класу моделей вимагає коректного обліку низки факторів, що впливають на екосистему і разом з тим частина з них є корельованими або взаємно корельованими величинами.

Розв'язання системи (1) пов'язане із значними обчислювальними складнощами, які зумовлені, насамперед, особливостями класу моделей, що розглядаються, зокрема незадовільною зумовленістю і великою розмірністю. Жорсткість, як показує досвід розв'язання практичних задач оброблення супутникових знімків, практично завжди є загальною властивістю математичних моделей, які враховують сукупність різних процесів, що поступово змінюють стан геопросторової зони [1, 2, 12, 13]. Характерним прикладом є дослідження впливу поширення забруднених вод через підземне поховання рідких радіоактивних відходів, наведене у [12]. Відповідна математична модель геопросторової зони повинна враховувати сукупність процесів, що мають складний взаємозв'язок, і бути представлена в ГІС за необхідності прогнозування динаміки шуканих параметрів. Динаміку змінних з використанням чисельного моделювання аналізу впливу поширення забруднених вод після підземного поховання рідких радіоактивних відходів графічно представлено на рис. 3 [12]. Математична модель, що описує зображені на рис. 3 процеси, містить швидко і повільно мінливі складові і відповідно належить до класу жорстких. Вони присутні у різних співвідношеннях як в x_1 , так і в x_2 .

Обчислювальні складнощі, які спричиняє ступінь зумовленості, можуть посилюватися під час розрахунків динамічних моделей сканерних супутникових знімків, коли для їхнього зовнішнього орієнтування використовують колінеарні

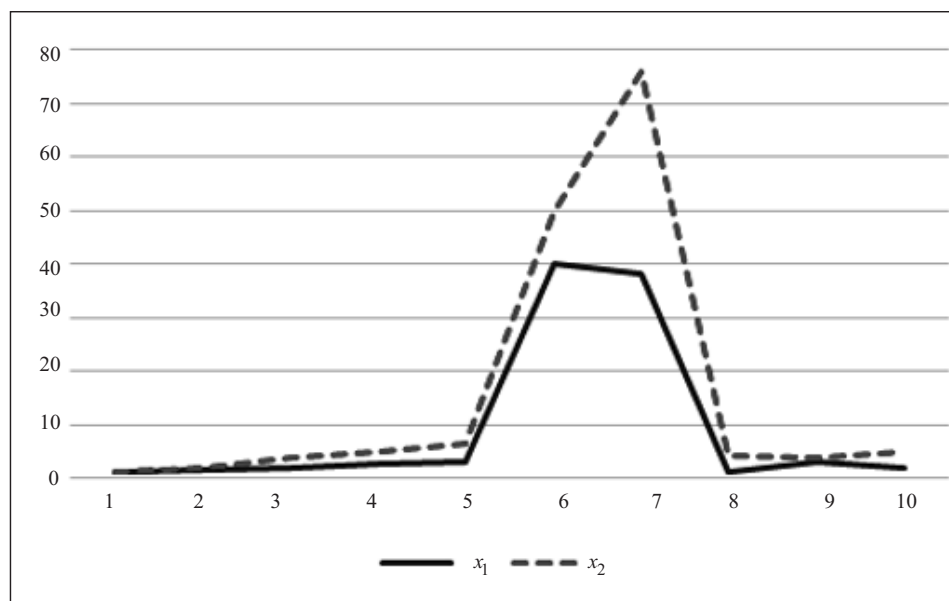


Рис. 3

залежності [14]. Зокрема, під час розрахунку зворотного фотограмметричного засічення незадовільна зумовленість системи рівнянь поправок призводить до суттєвих обчислювальних проблем [1, 2]. Система нормальних рівнянь, що використовується в цьому випадку, має вигляд [2]

$$B^T P B \Delta T + B^T P L = 0, \quad (3)$$

де B — матриця частинних похідних функцій, що розглядаються, за елементами зовнішнього орієнтування; ΔT — вектор виправлення до початкових (наближених) значень невідомих та вимірювальних координат точок; L — вектор вільних членів; T — символ транспонування; P — діагональна матриця ваг вимірювань.

Для подолання цих проблем використовуватимемо декомпозиційні методи, зокрема метод зниження порядку [6–8], що дає змогу реструктуризувати повні вихідні моделі ГІС для представлення її у вигляді окремих взаємозв'язаних підмоделей, які характеризуються різною швидкістю їхньої динаміки. Переходячи згідно з алгоритмами методу зниження порядку до декомпозиційної моделі та здійснюючи її лінеаризацію, перетворюємо систему рівнянь (1) до вигляду

$$\dot{X}'_1 = A_{11}x'_1 + A_{12}x'_2 + \zeta_1, \quad (4)$$

$$\dot{X}'_2 = A_{21}x'_1 + A_{22}x'_2 + \zeta_2, \quad (5)$$

де x'_1 та x'_2 — повільні і швидкі змінні, а розмірність системи (5) визначається кількістю малих власних чисел якобіана відповідної функції f .

Визначення розмірностей субмоделей (4) і (5) виконуватимемо за допомогою алгоритмів методу Хиленка [6–8] і оцінимо (з використанням ітераційних обчислень) значення власних чисел матриці A . Отримані результати дають змогу проводити у подальшому обчислення, переходячи, за необхідності, до уточнюваних декомпозиційних моделей, побудованих за допомогою методу зниження порядку.

Обчислення власних значень матриць (яка зазначена у публікаціях [1, 12] як істотна обчислювальна проблема) необхідне також для оцінювання стійкості розв'язування системи рівнянь (4), (5) стосовно варіацій елементів матриць. Вважатимемо, що наявність таких варіацій є перманентною і спричиняється насамперед апаратними та програмними помилками знімання даних та округленнями отриманих значень. Зазначену проблему доцільно розв'язувати на основі багаторазового застосування наведених вище ітераційних алгоритмів методу Хиленка аналогічно до одноразового їхнього застосування за методом зниження порядку для визначення розмірності декомпозиційних підмоделей (4) та (5).

Аналіз чутливості субмоделей (4), (5) виконаємо на модельних матрицях аналогічно до [15] з використанням матриць розміру $n = 2 \times 2$ і $n = 10 \times 10$. Зазначимо, що у простішому випадку стосовно подальшого аналізу спрощена модель динаміки процесу, наведеного на рис. 3 [12], на визначеному інтервалі може бути подана у вигляді системи двох рівнянь з модельною матрицею розміру $n = 2 \times 2$:

$$\begin{pmatrix} -1001 & 1000 \\ 1000 & -1001 \end{pmatrix} \quad (6)$$

Матриця розміру $n = 10 \times 10$ має вигляд:

-11100	500	50	250	100	200	50	30	50	200	(7)
500	-600	50	50	250	100	80	60	150	40	
50	50	-525	200	50	40	100	70	200	25	
250	50	200	-475	10	70	40	25	30	50	
100	250	50	10	-425	100	150	40	100	250	
200	100	40	70	100	-380	100	60	30	45	
50	80	100	40	150	40	-300	80	65	30	
30	60	70	25	30	60	80	-250	100	90	
50	150	200	50	100	30	65	100	-70	200	
200	40	25	50	250	45	30	90	200	-20	

Обчислення похибки декомпозиційних моделей є необхідним для оцінювання адекватності отриманих розв'язків і дає змогу вибирати модель відповідно до вимог точності розв'язків і можливостей використовуваного апаратного забезпечення. У публікаціях, що визначають результати практичного використання наявних на ринку систем оброблення супутникових знімків, зазначено появи збоїв у їхній роботі, які періодично виникають через недостатній розмір пам'яті [9]. Застосування декомпозиційних моделей ГПС, що формуються з відносно незалежних субмоделей, які визначають динаміку окремих геопросторових зон, дає змогу суттєво послабити вимоги до характеристик обчислювального комплексу. Насамперед це дасть змогу зменшити до мінімальних значень обсяг оперативної пам'яті задіяного обчислювального комплексу. Враховуючи украй великі обсяги оброблюваних даних, це розширює вибір визначених апаратних розв'язків.

Результати обрахунків власних чисел матриці (6) і найбільшого за модулем власного числа матриці (7) представлено у табл. 1. Отримання власних чисел наведених матриць свідчить про їхню погану обумовленість. Власні числа розглянутих матриць обчислювалися за методами Крилова, Ньютона–Лаверньє, Хилленка. Вибір цих методів зумовлений тим, що (як показано у роботі [15]) вони стійкі до численних обмежень допустимого діапазону зміни власних чисел.

Вважаючи, що низка природних та технологічних факторів, що мають випадкову природу (недосконалість вимірювань, наявність хмарності, шуми під час передачі сигналів та інше), додають похибку в отримані дані [16–18], випадковим чином внесемо зміни в матриці (6) і (7)

$$\begin{matrix} -1030 & 950 \\ 980 & -1020 \end{matrix} \quad (8)$$

і представимо їхні змінені елементи (з варіативністю орієнтовно 10 %) у вигляді

$$\begin{aligned} a(1,1) &= -11080; \\ a(1,4) &= 300; \\ a(2,1) &= 400; \\ a(6,6) &= -335; \\ a(8,9) &= 140; \\ a(9,9) &= -60; \\ a(10,1) &= 150. \end{aligned} \quad (9)$$

Власні числа матриці розміру 2×2 та найбільше власне число матриці розміру 10×10 після внесення похибок, що обчислені за методом Хилленка, пред-

Таблиця 1

Розмір матриці $n = 2 \times 2$				(6)
λ_{\max}	-2001,0	λ_{\min}	-1,0	
Розмір матриці $n = 10 \times 10$				(7)
λ_{\max}				-11137,5

Таблиця 2

Розмір матриці $n = 2 \times 2$				(6)
λ_{\max}	-1989,9	λ_{\min}	-60,1	
Розмір матриці $n = 10 \times 10$				(7)
λ_{\max}				-11113,2

ставлено у табл. 2. Наведені у табл. 1 та табл. 2 дані показують, що незначні (у процентному відношенні) зміни вхідних даних призводять до несуттєвих коливань великих власних чисел. Однак для малих власних чисел ті самі зміни вхідних даних призводять до принципово інших і, в більшості випадків, неприйнятних результатів.

Наведені у табл. 1 та табл. 2 дані свідчать про значно більшу стійкість обчислювального процесу насамперед за «повільними» змінними у разі переходу від повних моделей до декомпозиційних моделей. Відповідно для роботи з декомпозиційними моделями можна послабити вимоги до параметрів обчислювального комплексу, який використовують для оброблення вихідних супутникових зображень.

На підставі аналізу інформації та прогнозування в ГІС можна дійти таких висновків.

- Оцінювання власних чисел матриць у математичних моделях, що визначають стан геопросторових зон, дає змогу формалізувати та автоматизувати процес вилучення областей з нехарактерною швидкістю змін (такою, що не відповідає прогнозам, зробленим на базі математичних моделей, чи даним, які зібрано за тривалий період спостережень).

- З урахуванням великих обсягів інформації, пов'язаних з використанням ГІС, застосовано математичний апарат методів декомпозиції для перетворення вихідних моделей, що дає змогу оптимізувати витрати часу та обчислювальний ресурс для розв'язання задач геопросторового зондування Землі.

- Використання декомпозиційних моделей ГІС, що містять відносно незалежні субмоделі, які визначають динаміку окремих геопросторових зон, дає змогу суттєво послабити вимоги до низки характеристик використовуваного обчислювального комплексу, зокрема до обсягу оперативної пам'яті.

СПИСОК ЛІТЕРАТУРИ

1. Шавук В.С. Теоретическое обоснование цифровой фотограмметрической системы обработки космических снимков высокого разрешения. *Автореф. дис. ... канд. техн. наук.* Москва, 2009. 24 с.
2. Семинев А.А., Тарасова Е.И. Обратная фотограмметрическая засечка: надежность решения задачи. *Вестник СПбГУ. Сер. 7.* 2012. Вып. 4. С. 129–134.
3. Hong Hu, Liedtke J. Monitor forest change over time. URL: <https://learn.arcgis.com/ru/projects/monitor-forest-change-over-time/>.
4. Simon Woo. Подготовка изображений и растровых данных для анализа. URL: <https://learn.arcgis.com/ru/projects/prepare-imagery-and-raster-data-for-analysis/>.
5. Начало работы со спутниковыми изображениями. URL: <https://learn.arcgis.com/ru/projects/get-started-with-imagery>.

6. Khilenko V.V. Order reduction method and adequate simplification of sae models with uncertain coefficients. *Cybernetics and Systems Analysis*. 1998. Vol. 34, N 3. P. 458–461.
7. Grishchenko A.Z., Khilenko V.V. Determining the number of fast and slow components in decomposition of arbitrarily large linear dynamical models. *Cybernetics and Systems Analysis*. 1991. Vol. 27, N 6. P. 795–801.
8. Khilenko V.V. Convergence of algorithms of order-reducing method during analysis of non-linear mathematical fragment models of computer-aided systems for controlling the production processes. *Cybernetics and Systems Analysis*. 2001. N 3. P. 373–380.
9. Семин П. Как обрабатывать спутниковые снимки с помощью Sen2Cor. 2020. URL: <https://habr.com/ru/post/501188/>.
10. Дрейпер Н., Смит Г. Прикладной регрессионный анализ, 3-е изд. Москва: Диалектика, 2016. 912 с.
11. Фёрстер Э., Рёнц Б. Методы корреляционного и регрессионного анализа, 1983. 304 с.
12. Букаты М.Б. Геоинформационные системы и математическое моделирование. Уч. пособие. Томск: Томский политехнический университет, 2009. 75 с.
13. Нейронные сети и ГИС. URL: <https://scibook.net/sistemyi-geologii-geoinformatsionnyie/neuronnyie-seti-gis-48136.html>.
14. Мердок Джеймс. Нормальные формы и развертки для локальных динамических систем. Springer, 2003. 500 с. ISBN 978-0-387-21785-7.
15. Khilenko V.V., Stepanov O.V., Kotuliak I., Reis M. Optimization of the selection of software elements in control systems with significantly different-speed processes. *Cybernetics and Systems Analysis*. 2021. Vol. 57, N 2. P. 185–189.
16. Skakun S.V., Basarab R.M. Reconstruction of missing data in time-series of optical satellite images using Self-Organizing Kohonen maps. *Journal of Automation and Information Sciences*. 2014. Vol. 46. P. 19–26. <https://doi.org/10.1615/JAutomatInfScien.v46.i12.30>.
17. Савиных В.П., Цветков В.Я. Геоинформационный анализ данных дистанционного зондирования. Москва: Картгеоцентр – Геодезиздат, 2001. 228 с.
18. Djamai N., Fernandes R. Comparison of SNAP-derived sentinel-2A L2A product to ESA product over Europe. <https://doi.org/10.3390/rs10060926>.

I. Kotuliak, V.V. Khilenko, R.M. Basarab, M. Ries

APPLICATION OF DECOMPOSITION ALGORITHMS TO SPEED UP PROCESSING OF LARGE DATA SETS IN GIS

Abstract. The authors propose a technology and a decomposition algorithm for speeding up the processing of geoinformation data based on the division of dynamic and quasi-static data samples using the analysis of eigenvalues of matrices obtained by means of iterative calculation by the Khilenko method. The algorithm is aimed at processing large geoinformation data arrays. Comparative results of the model calculations using known computational methods are presented.

Keywords: geoinformation data, satellite image processing, big data, decomposition, iterative calculation of matrix eigenvalues, Khilenko’s method.

Надійшла до редакції 21.07.2022