



ISSN 2411–6602 (Online)

ISSN 1607–2855 (Print)

Том 12 • № 1 • 2016 С. 53 – 56

УДК 504.064 : 535.361.2 : 519.6

## Аналіз інформативних компонент для оцінювання концентрації біохімічних параметрів рослинності за спектром відбиття

О.В. Семенів

Інститут космічних досліджень НАН України та Державного космічного агентства України, м. Київ

*В роботі представлено реалізацію метода незалежних компонент в задачі оцінювання стану рослинності за спектром відбиття листя рослин. Проведено дослідження взаємозв'язку між значеннями інформаційних ознак та сумарною концентрацією хлорофіла. Представлено результати чисельного моделювання та порівняльного аналізу з іншими методами та експериментальними даними.*

*АНАЛИЗ ИНФОРМАТИВНЫХ КОМПОНЕНТ ДЛЯ ОЦЕНКИ КОНЦЕНТРАЦИИ БИОХИМИЧЕСКИХ ПАРАМЕТРОВ РАСТИТЕЛЬНОСТИ ПО СПЕКТРУ ОТРАЖЕНИЯ, Семенов О.В. — В работе представлена реализация метода независимых компонент в задаче оценивания состояния растительности по спектру отражения листьев растений. Проведено исследование взаимосвязи между значениями информационных признаков и суммарной концентрацией хлорофилла. Представлены результаты численного моделирования и сравнительного анализа с другими методами и экспериментальными данными.*

*ANALYSIS OF THE INFORMATIVE COMPONENTS FOR ESTIMATION OF VEGETATION BIOCHEMICAL PARAMETERS USING THE REFLECTION SPECTRA, by Semenov O.V. — An implementation of the independent component analysis for vegetation state estimation using the plant leaves' reflection spectra is presented. An interrelation between the informative parameters and the total concentration of chlorophyll in the leaves is investigated. Results of numerical modeling and comparative analysis with the experimental data and other methods are presented.*

**Ключевые слова:** ДЗЗ; независимый компонентный анализ; корреляционный анализ; регрессионный анализ; состояние растительности.

**Key words:** remote sensing; independent component analysis; correlation analysis; regression analysis; vegetation state.

### 1. ВСТУП

На сьогоднішній день проведено багато досліджень, присвячених проблемі оцінюванню стану біологічних об'єктів, зокрема визначенню концентрації біохімічних компонентів в рослинності. Рослини є необхідним компонентом живого світу, а фотосинтетичний апарат зелених листків може бути використаний як чутливий індикатор стресових ситуацій. Останнє обумовлено тим, що стреси різного походження впливають, насамперед, на вміст хлорофілу, а цей параметр можна визначити безконтактним методом, заснованим на використанні характеристик відбиття листків в оптичному діапазоні. Існує велика кількість публікацій, які вказують на реакцію пігментного апарата рослин у відповідь на дію стресу [1–8, 11, 13–15]. Не дивлячись на значні досягнення в цьому напрямку, розвиток засобів ДЗЗ обумовлюють ріст досліджень у напрямку створення системних методів та обчислювальних засобів для визначення концентрації біохімічних компонентів у рослинності з використанням методів оптимізації, класифікації та оцінювання з урахуванням типу рослин, проективного покриття та ґрунту.

Вплив різних стресів, перш за все, впливає на зміни вмісту пігментів в листках рослин. Це явище дозволяє розвивати ефективні підходи для швидкого виявлення стресів у фітоценозі різних типів. На сьогодні для обчислення концентрації хлорофілу в листках з різними спектральними коефіцієнтами відбиття існують численні методи, які ґрунтуються на вегетаційних індексах та регресійних співвідношеннях [2, 4, 5, 8, 11, 13–15]. Згадані підходи завжди ефективні, коли концентрація хлорофілу виміряна в лабораторних умовах, проте незначна кількість з них дозволяють оцінювати вміст хлорофілу для системи “ґрунт–рослинність”. Головною причиною цього є вплив проективного покриття на дані спостережень. Існує кілька підходів до мінімізації впливу такого шуму шляхом використання комплексних індексів рослинності, які містять компоненти для врахування впливу землі [7, 11, 15]. На практиці необхідно використовувати додаткові параметри системи, такі як листковий індекс та спектральний коефіцієнт відбиття землі. Дана умова робить процедуру малоефективною і, крім того, проблема не вирішується повністю внаслідок того, що відбиття від землі відносний показник, наприклад, він сильно залежить від вологості. Дана робота спрямована на доопрацювання та вдосконалення програмно-алгоритмічного забезпечення для обробки та виділення корисної інформації з наземних спектральних та гіперспектральних даних.

Семенів Олег Володимирович; ✉ [oleg.semeniv@gmail.com](mailto:oleg.semeniv@gmail.com)

*Вісник Астрономічної школи, 2016, том 12, № 1*

53

## 2. ВИДІЛЕННЯ ІНФОРМАТИВНИХ ОЗНАК СПЕКТРУ ВІДБИТТЯ НА ОСНОВІ НЕЗАЛЕЖНОГО КОМПОНЕНТНОГО АНАЛІЗУ

При оціненні вмісту біохімічних компонент (хлорофілу) за спектральними характеристиками, виникає ціла низка задач, вирішення яких в основному залежить від якості, достовірності та розмірності експериментальних даних. Добре відомо, що для ефективної обробки багатовимірних даних перш за все потрібно понизити розмірність набору експериментальних вимірів з подальшим виділенням найбільш інформативних ознак, які характеризують те чи інше явище. В даній роботі розглядається метод оцінювання концентрації хлорофілу на основі незалежного компонентного аналізу (МНК) [10, 12]. На відміну від методу головних компонент [12], в якому проводиться пошук некорельованих між собою компонент, в МНК здійснюється пошук статистично незалежних між собою інформативних параметрів. Даний підхід базується на оптимізації критерію “незалежності” між експериментальними вимірами та розрахунком лінійних не ортогональних інформативних ознак [12].

В ході попередніх досліджень [7, 8] показано, що для аналізу спектру відбиття рослин та отримання доволі високої точності оцінювання хлорофілу можна використовувати спектральний діапазон 670–750 нм, що дозволяє зменшити розмірність вхідних даних. Представимо спектр відбиття вищезгаданого діапазону у вигляді вектору  $\hat{\mathbf{x}} = (x_1, \dots, x_M)$ , де  $x$  є значення коефіцієнту яскравості на довжині хвилі  $\lambda$ , крок дискретизації становить 1 нм. Коли маємо справу з множиною вимірювань, значення вектору  $\hat{\mathbf{x}}$  міняються у відповідності до зміни часового параметру  $t$ . Потрібно отримати більш просте описання сімейства векторів. Перш за все зсунемо початок системи координат в “центр ваги”, використовуючи формулу

$$\bar{\mathbf{x}} = \frac{1}{L} \sum_{l=1}^L \hat{\mathbf{x}}^l, \quad l = \overline{1, L}, \quad (1)$$

де  $L$  — число спектральних кривих, виміряних протягом деякого часу. Щоб провести нормування вхідного вектору, введемо нову змінну  $\mathbf{x}$

$$\mathbf{x}(t) = \hat{\mathbf{x}}(t) - \bar{\mathbf{x}}(t). \quad (2)$$

Припустимо, що існує така рівність:

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \quad (3)$$

де матриця  $\mathbf{A}$  є скалярною матрицею, що складається із коефіцієнтів перетворення, а  $\mathbf{s}$  є вектором інформативних ознак. Припустимо, що компоненти  $s_i$ ,  $i = 1, \dots, N$ ,  $N \ll M$ , вектору  $\mathbf{s}$  є статистично незалежними та характеризуються негауссівським розподілом. Метою є знаходження матриці  $\mathbf{W} = \mathbf{A}^{-1}$ , яка задовольняє умову  $\mathbf{s} = \mathbf{W}\mathbf{x}$ . Набір компонент  $s_i$  є незалежними, коли вектор компонент  $s_j$  не несе в собі жодної інформації про інші вектори інформативних ознак вибірки при  $i \neq j$ . Технічно незалежність можна представити у вигляді функція розподілу  $p(s_1, \dots, s_i)$  величини  $s_i$ , або через  $p_i(s_i)$  маргінальну функцію розподілу величини  $s_i$ . Тоді припустимо, що  $s_i$  є незалежними між собою тоді і тільки тоді, коли загальна функція розподілу рівна

$$p(s_1, \dots, s_i) = \prod p_{s_i}(u_i). \quad (4)$$

Розрахунок зводиться до знаходження вагових коефіцієнтів матриці  $\mathbf{W}$ , таких, щоб вектор  $\mathbf{s} = \mathbf{W}\mathbf{x}$  вмщав як можна менше гауссівських компонент. Ось чому потрібна міра негауссівського критерію. Для цього використаємо величину відмінну від ентропії — негентропію [12]. Ентропія довільного вектору  $\mathbf{x}$  з математичним розподілом  $p_{\mathbf{x}}(\mathbf{u})$  визначається наступним виразом:

$$H(p_{\mathbf{x}}) = - \int p_{\mathbf{x}}(\mathbf{u}) \log p_{\mathbf{x}}(\mathbf{u}) d\mathbf{u}. \quad (5)$$

$H(p_{\mathbf{x}})$  набуває свого максимального значення тоді, коли довільний вектор  $\mathbf{x}$  має гауссівську природу. Негентропія визначається із різниці між ентропією вектору  $\mathbf{x}$  та ентропією довільного гауссівського вектору  $\mathbf{x}_g$ , що належить до тієї ж ковераційної матриці, що й  $\mathbf{x}$

$$J(\mathbf{x}) = H(p_{\mathbf{x}_g}) - H(p_{\mathbf{x}}). \quad (6)$$

Коефіцієнт різності Кульбака–Лейблера дозволяє провести розрахунок відстані між двома довільними розподілами  $p_{\mathbf{v}}$  та  $p_{\mathbf{w}}$

$$k(p_{\mathbf{v}}, p_{\mathbf{w}}) = - \int p_{\mathbf{v}}(\mathbf{u}) \log [p_{\mathbf{v}}(\mathbf{u})/p_{\mathbf{w}}(\mathbf{u})] d\mathbf{u}. \quad (7)$$

В ході перестановки можна показати, що негентропія  $J(\mathbf{x})$  рівна величині коефіцієнту різності Кульбака–Лейблера між векторами  $\mathbf{x}$  та  $\mathbf{x}_g$

$$J(\mathbf{x}) = k(p_{\mathbf{x}}, p_{\mathbf{x}_g}). \quad (8)$$

Проте на практиці провести розрахунки виразу (8) складно, тому використаємо вираз розрахунку негентропії, запропонований в [12]

$$J(\mathbf{x}) = \{E[G(\mathbf{x})] - E[G(\mathbf{x}_g)]\}^2, \quad (9)$$

де  $G$  є деяка неквадратична функція. Максимізація величини  $J(\mathbf{x})$  виразу (9) потребує використання оптимізаційного алгоритму. Використаємо так званий Fast-ICA алгоритм [12], що базується на схемі почергового розрахунку окремої компоненти:

1. Вибір початкового (довільного) вектору ваги  $\mathbf{w}$ .
2. Тоді нехай  $\mathbf{w}^+ = E\{xg(\mathbf{w}^T \mathbf{x})\} - E\{g'(\mathbf{w}^T \mathbf{x})\}\mathbf{w}$ , де  $g(\mathbf{u}) = \text{th } x(\mathbf{u})$ .
3.  $\mathbf{w} = \mathbf{w}^+ / \|\mathbf{w}^+\|$ .

Після кожної ітерації над векторами  $\mathbf{w}$  та  $\mathbf{x}$  проводиться процедура декореляції, використовуючи симетричну матрицю

$$\mathbf{W} = (\mathbf{W}\mathbf{W}^T)^{-1/2}\mathbf{W}, \quad (10)$$

де матриця  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^T$  складається з векторів  $\mathbf{w}_i$ .

### 3. ІДЕНТИФІКАЦІЯ МОДЕЛЕЙ ТА ТЕСТУВАННЯ ТОЧНОСТІ ОЦІНЮВАННЯ ХЛОРОФІЛУ

З точки зору застосування МНК для оцінювання концентрації біохімічних компонент, кожна (або деякій набір) із інформативних ознак (ІО) відповідає вмісту одній із біохімічної компоненті (наприклад хлорофілу). Сукупність внесків біохімічних компонент у спектр відбиття забезпечує специфічну форму спектральної кривої. Чим вище (нижче) концентрації певного пігменту, тим більш виражена форма кривої в певному діапазоні. Враховуючи, що біохімічних компонент є велика кількість, постає задача виявлення найбільш інформативних ознак, що відносяться до конкретного пігменту.

*Означення.* Введемо вектор  $s_i$ , значення компонент якого відповідають коефіцієнту кореляції із вмістом хлорофілу  $Z_i$  по модулю вище порогового значення 0,2, і назвемо вектором найбільш інформативних ознак спектру відбиття  $x_i(\lambda)$ .

Із вибірки експериментальних даних вибрано набір спектральних кривих відбиття (17 спектрів, що відповідають різній концентрації хлорофілу) для розрахунку виразу (1) та проведено нормування. Результати розрахунків використано для знаходження матриці  $\mathbf{W}$  за допомогою вищеописаного методу. Після цього розраховано вектор незалежних компонент  $\mathbf{s}$ , з якого виділено найбільш ІО. В таблиці 1 представлено коефіцієнти кореляції між концентрацією хлорофілу та значеннями незалежних компонент із застосуванням наведеного означення. Як видно із таблиці, найвищий коефіцієнт кореляції забезпечують ІО під порядковими номерами 1 та 7. На основі їх значень проведено побудову регресійної моделі

$$z(s_1, s_7) = 5,52 - 0,992s_1 + 1,169s_7, \quad (11)$$

статистичні параметри  $R = 0,87$ ,  $R^2 = 0,71$ . Описаний вище метод та отриману модель використано для розробки чисельного алгоритму визначення концентрації хлорофілу, програмна реалізація якого виконана на мові C++ з використанням математичних бібліотек пакету Matlab.

Для визначення концентрації хлорофілу МНК використано спектри відбиття листків озимої пшениці з різним вмістом хлорофілу, концентрація якого встановлена біохімічним методом (БХМ) [9]. Результати порівняння розрахунків вмісту хлорофілу МНК, а також з використанням методу першої похідної (МПП) [2] приведені в табл. 2.

В ході тестування методу на 70 спектральних кривих розраховано коефіцієнт кореляції між експериментально встановленою концентрацією хлорофілу та розрахованими на основі МНК  $r = 0,87$ , а також виведено середній квадрат похибки, рівний 1,9399. Результати порівняння оцінок хлорофілу, отриманих МНК і хімічним методом [9], показані на рис. 1.

**Таблиця 1.** Значення коефіцієнтів кореляції

	Номер незалежної компоненти							
	1	2	3	4	5	6	7	8
Значення коеф. кореляції	-0,545	0,257	0,269	0,299	0,273	-0,235	0,642	0,216

**Таблиця 2.** Результати порівняння отриманих за допомогою моделі (11) та МПП

Найменування спектру відбиття	МНК	МПП	Біохім.
RP62.dat	2,42	2,84	2,33
RP13.dat	2,71	3,20	2,90
RP50.dat	3,98	4,30	3,75
RP4.dat	4,78	4,95	4,76
RP40.dat	5,19	5,67	6,29
RP20.dat	6,21	6,02	6,57
RP19.dat	6,74	6,67	6,82
RP58.dat	6,95	7,45	7,65
RP35.dat	8,00	7,53	8,08
Середньоквадратична похибка	1,9399	1,7406	

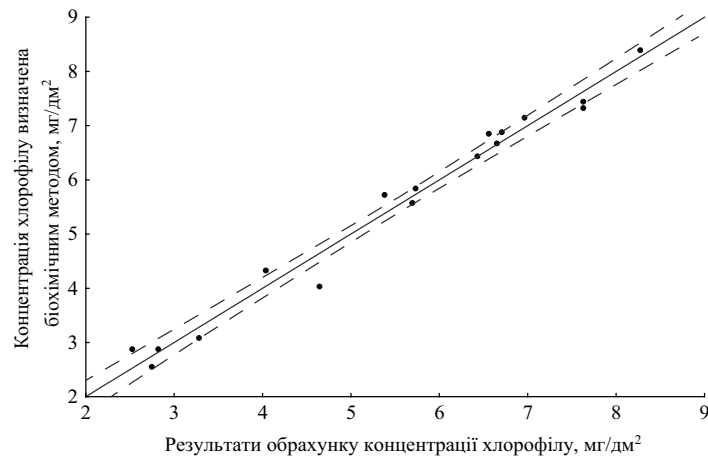


Рис. 1. Результати оцінок хлорофілу з використанням моделі (11) та БХМ

#### 4. ВИСНОВКИ

Запропоновано новий системно-інформаційний підхід до оцінювання вмісту біохімічних компонентів у рослинності з урахуванням таких факторів невизначеності, як тип рослини, тип ґрунту та проективне покриття. Розроблено метод на основі компонентного аналізу та проведено реалізацію в алгоритмі обробки зашумлених спектральних даних. Встановлено, що в ближньому інфрачервоному діапазоні спектру відбиття рослинності дві інформативні ознаки значимо корелюють із вмістом хлорофілу, що підтверджує отримані результати методами першої похідної та головних компонент. Побудовано лінійну регресійну модель, що демонструє значиму точність оцінювання хлорофілу. Підвищити точність можна шляхом побудови нелінійної регресійної моделі та вибору оптимального набору навчальних даних. Даній задачі будуть присвячені наступні дослідження.

1. Андреева А.В., Бузников А.А., Скрябин С.В. и др. Исследование характера изменения оптических характеристик растительности под воздействием тяжелых металлов для разработки метода дистанционной диагностики загрязнения // Современные проблемы дистанционного зондирования Земли из космоса. — 2007. — **4**, № 2. — С.175–182.
2. Кочубей С.М. Оценка основных параметров сельскохозяйственных посевов по спектру отражения растительности в оптическом диапазоне // Космічна наука і технологія. — 2003. — № 9. — С.185–190.
3. Кынчева Р., Илиев И., Борисова Д., Горгиев Г. Раннее обнаружение физиологического стресса растительности по многоспектральным данным // Современные проблемы дистанционного зондирования Земли из космоса. — 2011. — **8**, № 4. — С.319–326.
4. Лялько В.І., Попов М.О. Багатоспектральні методи дистанційного зондування землі в задачах природокористування. — Київ: Наукова думка. — 360 с.
5. Семенів О.В. Оцінювання стану рослинності на основі Фур'є-аналізу даних дистанційного спостереження // Вісник Астрономічної школи. — 2015. — **11**, № 1. — С.56–60.
6. Якунина И.В., Попов Н.С. Методы и приборы контроля окружающей среды. Экологический мониторинг. — Тамбов: Изд-во Тамб. гос. техн. ун-та, 2009. — 188 с.
7. Яценко В.О., Кочубей С.М., Хандрига П.А. и др. Новый метод дистанционного оценивания содержания хлорофилла в растительности и его программно-аппаратная реализация // Космічна наука і технологія. — 2007. — № 3(13). — С.35–45.
8. Яценко В.О., Семенів О.В. Динамічний підхід до оцінювання параметрів біохімічних компонент в рослинності // Вісник Київського університету. Сер.: Фізико-математичні науки. — 2009. — № 1. — С.163–168.
9. Arnon D.I. Copper enzymes in isolated chloroplasts. Polyphenoloxidase in *Beta vulgaris* // Plant Physiol. — 1949. — **24**. — P.1–15.
10. Comon P. Independent component analysis, A new concept? // Signal processing. — 1994. — **36**. — P.287–314.
11. Feret J.B., François C., Asner G.P., et al. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments // Remote Sensing of Environment. — 2008. — № 112. — P.3030–3043.
12. Hyvarinen A., Karhunen J., Oja E. Independent Component Analysis. — New York: J. Wiley & Sons, Inc, 2001. — 481 p.
13. Rosso P.H., Pushnik J.C., Mui Lay, Ustin S.L. Reflectance properties and physiological responses of *Salicornia virginica* to heavy metal and petroleum contamination // Environmental Pollution. — 2005. — **137**. — P.241–252.
14. Smith K.L., Steven M.D., Colls J.J. Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks // Remote Sensing of Environment. — 2004. — **92**. — P.207–217.
15. Zarco-Tejada P.J., Berni J.A.J., Suárez L., et al. Imaging chlorophyll fluorescence with an airborne narrow-band multispectral camera for vegetation stress detection // Remote Sensing of Environment. — 2009. — **113**. — P.1262–1275.

Надійшла до редакції 31.07.2016  
Прийнята до друку 5.09.2016