



ISSN 1607–2855

Том 10 • № 1 • 2014 С. 70 – 74

УДК 504.064 : 535.361.2 : 519.6

Оцінювання біохімічних компонент в рослинності на основі методів статистичного навчання та даних дистанційного зондування

О.В. Семенів

Інститут космічних досліджень НАН України та Державного космічного агентства України

У роботі представлено підхід до оцінювання стану рослинності шляхом визначення концентрації хлорофілу в листі рослин за спектральними даними із застосуванням методу статистичного навчання. Проведено описання методу побудови моделі на основі рішення двоїстої задачі оптимізації. Представлено процедуру одержання експериментальних даних і результати комп'ютерних розрахунків та порівняльного аналізу.

ОЦЕНИВАНИЕ БИОХИМИЧЕСКИХ КОМПОНЕНТ В РАСТИТЕЛЬНОСТИ НА ОСНОВЕ МЕТОДОВ СТАТИСТИЧЕСКОГО ОБУЧЕНИЯ И ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ, Семенов О.В. — В работе представлен подход к оцениванию состояния растительности путем определения концентрации хлорофилла в листьях растений по спектральным данным с применением метода статистического обучения. Проведено описание метода построения модели на основе решения двойственной задачи оптимизации. Представлены процедура получения экспериментальных данных и результаты компьютерных расчетов, и сравнительного анализа.

ESTIMATION BIOCHEMICAL COMPONENTS IN VEGETATION BASED ON STATISTICAL LEARNING METHODS AND REMOTE SENSING DATA, by Semeniv O.V. — An approach for vegetation state estimation is presented in the paper. It is based on the determination of the chlorophyll content in the plant leaves using spectral data and statistical learning methods. The problem of model identification is presented as dual optimization problem. Also the experimental data obtaining procedure, numerical results and comparative analysis are shown.

Ключевые слова: ДЗЗ; оценивание; метод опорных векторов; состояние растительности; идентификация модели.

Key words: remote sensing; estimation; SVM; vegetation state; model identification.

1. ВСТУП

В даний час більшість території земної поверхні зазнає екологічних проблем, пов'язаних із збільшенням транспортного потоку, інтенсивною забудовою, підвищенням використанням водних ресурсів в промислових цілях, спорудженням об'єктів енергопостачання і т.п. [1, 2, 4, 6]. Виникає необхідність створення та відпрацювання експресних методів оцінювання екологічного стану різних природно-територіальних об'єктів, особливо зон підвищеного антропогенного навантаження. Одними із найбільш чутливих індикаторів забруднення навколишнього середовища є рослини, а саме, їх фотосинтетичний апарат. Наявність різного роду забруднень (важкі метали, нафтопродукти, нітрати, пестициди та ін.) призводять до їх накопичення в рослинах, що викликає стрес. Це проявляється у затримці росту, зменшенні біомаси, концентрації пігментів та води в листках рослин, зміні спектральних характеристик рослин, що дозволяє використовувати стан рослин як оцінку забруднення навколишнього середовища.

Для аналізу стану рослин з метою визначення ступеню забруднення використовується цілий ряд засобів та методів екологічного контролю. Такі засоби поділяються на контактні, дистанційні та біологічні [6]. Контактні методи представлені класичними методами хімічного аналізу та сучасними методами інструментального аналізу, що поділяються на фізичні та фізико-хімічні [6]. Останніми десятиліттями проводяться активні дослідження в напрямку розроблення універсальних методів діагностики стану рослин в умовах стресу на основі застосування даних дистанційних спостережень. Так, отримано залежності спектральних показників на певних довжинах хвиль від концентрації елементів важких металів у складі листків рослин [1, 2, 4, 10, 16]. Широке застосування знаходить дистанційна діагностика і при вирішенні завдань сільського господарства, наприклад, визначення водного балансу рослин, визначення концентрації хлорофілу та азоту, виявлення зон ерозії ґрунтів, знаходження ділянок зараження рослин грибами, паразитами і т.п. [3, 7, 8, 11–14, 17, 19, 20].

Однак із збільшенням інструментальних можливостей дистанційних спостережень (збільшення просторових та спектральних роздільних здатностей) виникають потреби розвитку та імплементації сучасних методів аналізу даних і оптимізації для підвищення ефективності обробки інформації та якісної її інтерпретації.

2. МЕТОД ОЦІНЮВАННЯ ВМІСТУ ХЛОРОФІЛУ В ЛИСТКАХ РОСЛИН НА ОСНОВІ ОПОРНИХ ВЕКТОРІВ

Розглянемо загальний принцип побудови моделі оцінювання вмісту хлорофілу, що базується на використанні методу опорних векторів [5, 9, 18]. Нехай вектор $\vec{x} = (x_1, \dots, x_N)$ є спектр відбиття листків рослини, де x є значення коефіцієнту відбиття на довжині хвилі λ , а N — значення довжини хвилі. Ідентифікація моделі проводиться на основі мінімізації функції втрат

$$f(y, (\vec{x}, \vec{w})) = \begin{cases} 0 & \text{при } |y - f(\vec{x}, \vec{w})| \leq \varepsilon, \\ |y - f(\vec{x}, \vec{w})| - \varepsilon, & \end{cases} \quad (1)$$

де y — концентрація хлорофілу, яка відповідає рослині із спектром відбиття \vec{x} ; f — деяка функція; \vec{w} — вектор невідомих параметрів; ε — допустима похибка оцінювання. Назвемо для зручності модель — регресією опорних векторів, пошук якої полягає в одночасній мінімізації функції втрат (1) та норми параметрів $\|\vec{w}\|^2$ [9, 18].

Припустимо, що проведено серію експериментальних досліджень (отримано спектри відбиття та концентрації хлорофілу), які в повній мірі характеризують ймовірні стани рослини в різні фази її розвитку. Із отриманої множини даних відберемо деяку підмножину і назвемо її навчальним набором даних із J спектрів відбиття. Припустимо, що результуючий набір є репрезентативним. Тоді для розв'язання задачі (1) введемо додаткові змінні $\xi_j, \xi_j^* \geq 0, j = 1, \dots, J$, які характеризують віддаленість навчальної вибірки від ε -нечутливої зони (зони допустимої похибки). Графічне відображення процедури пошуку регресії опорних векторів на основі застосування обмежень на допустиму похибку оцінювання концентрації хлорофілу представлено на рис. 1.

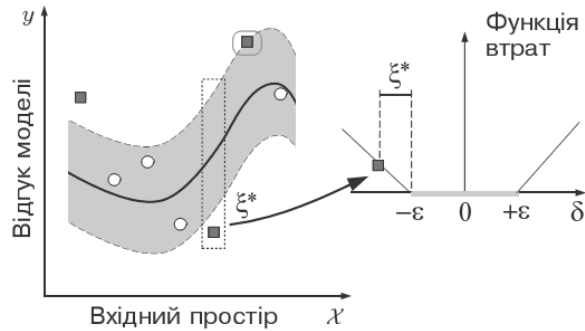


Рис. 1. Схематичне зображення ε -нечутливої смуги регресії опорних векторів

Розв'язок задачі (1) можна отримати, мінімізуючи функціонал

$$\min \frac{1}{2} \|\vec{w}\|^2 + C \sum_{j=1}^J (\xi_j + \xi_j^*), \quad (2)$$

$$\begin{cases} y_j - \vec{w} \cdot \vec{x}_j - b \leq \varepsilon + \xi_j \\ -y_j + \vec{w} \cdot \vec{x}_j + b \leq \varepsilon + \xi_j^* \\ \xi_j, \xi_j^* \geq 0, j = 1, \dots, J \end{cases} \quad (3)$$

де C — константа, що характеризує взаємозв'язок між адекватністю моделі та рівнем великих відхилень.

Для отримання більш точної та стійкої до шумів моделі оцінювання концентрації хлорофілу доцільно використовувати нелінійну регресію опорних векторів. Для її опису проведемо відображення простору вхідних даних розмірністю d^* в багатовимірний простір D . Іншими словами, задача побудови нелінійної регресії опорних векторів можна привести до розв'язання лінійної форми

$$f(\vec{x}, \vec{w}) = \langle \vec{w}, \tilde{\Phi}(\vec{x}) \rangle_D + b, \quad \vec{w} \in D_{\vec{w}}, \quad b \in R, \quad (4)$$

і відповідно оптимізаційної задачі

$$\min \frac{1}{2} \|\vec{w}\|_D^2 + C \sum_{j=1}^J (\xi_j + \xi_j^*), \quad (5)$$

з обмеженнями

$$\begin{cases} y_j - \langle \vec{w}, \tilde{\Phi}(\vec{x}_j) \rangle_D - b \leq \varepsilon + \xi_j \\ -y_j + \langle \vec{w}, \tilde{\Phi}(\vec{x}_j) \rangle_D + b \leq \varepsilon + \xi_j^* \\ \xi_j, \xi_j^* \geq 0, j = 1, \dots, J. \end{cases} \quad (6)$$

Для досягнення умовного екстремуму в (5) із урахуванням (6) використано множники Лагранжа $a_j, \zeta_j, a_j^*, \zeta_j^*$. Таким чином розв'язок оптимізаційної задачі можна представити у вигляді функції Лагранжа

$$L_G = \frac{1}{2} \langle \vec{w}, \vec{w} \rangle_{D_{\vec{w}}} + C \sum_{j=1}^J (\xi_j + \xi_j^*) + \sum_{j=1}^J \alpha_j (y_j - \langle \vec{w}, \tilde{\Phi}(\vec{x}_j) \rangle_D - b - \varepsilon - \xi_j) + \sum_{j=1}^J \alpha_j^* (-y_j + \langle \vec{w}, \tilde{\Phi}(\vec{x}_j) \rangle_D + b - \varepsilon - \xi_j^*) - \sum_{j=1}^J (\zeta_j \xi_j + \zeta_j^* \xi_j^*), \quad (7)$$

де $a_j, \zeta_j, a_j^*, \zeta_j^* \geq 0$.

Оптимум (7) досягається в сідловій точці при умові рівності нулю часткових похідних

$$\begin{cases} \partial_b L_G = \sum_{j=1}^J (\alpha_j^* - \alpha_j) = 0, \\ \partial_{\vec{w}} L_G = \vec{w} - \sum_{j=1}^J (\alpha_j - \alpha_j^*) \tilde{\Phi}(\vec{x}_j) = 0, \\ \partial_{\xi_j} L_G = C - \alpha_j - \zeta_j = 0, \\ \partial_{\xi_j^*} L_G = C - \alpha_j^* - \zeta_j^* = 0. \end{cases} \quad (8)$$

Підставивши (8) в (7), отримаємо

$$W(\alpha, \alpha^*) = -\frac{1}{2} \sum_{j,i=1}^J (\alpha_j - \alpha_j^*)(\alpha_i - \alpha_i^*) \langle \tilde{\Phi}(\vec{x}_j), \tilde{\Phi}(\vec{x}_i) \rangle_D - \varepsilon \sum_{j=1}^J (\alpha_j + \alpha_j^*) + \sum_{j=1}^J y_j (\alpha_j - \alpha_j^*), \quad (9)$$

з обмеженнями

$$\sum_{j=1}^J (\alpha_j - \alpha_j^*) = 0, \quad i \neq j, \quad \alpha_i, \alpha_i^* \in [0; C]. \quad (10)$$

Спираючись на [5, 9, 18], розв'язок оптимізаційної задачі (5) з урахуванням (9) та (10) приведемо до розв'язку двоїстої задачі

$$f(\vec{\alpha}) = \sum_{j,i=1}^J (\alpha_j - \alpha_j^*) \langle \tilde{\Phi}(\vec{x}_j), \tilde{\Phi}(\vec{x}_i) \rangle_D + b, \quad (11)$$

де значення параметрів змінюються в діапазонах $0 \leq \alpha_j \leq C$, $0 \leq \alpha_j^* \leq C$. Із виразу (11) слідує, що із навчального набору даних залишається невелика кількість значень (опорних векторів), які використовуються в подальшому для побудови моделі. Для розв'язку нелінійної задачі пошуку регресії опорних векторів проведено процедуру переходу з вхідного простору в простір вищого порядку

$$f(\vec{\alpha}) = \sum_{j,i=1}^J (\alpha_j - \alpha_j^*) k(\vec{x}_j, \vec{x}_i) + b, \quad (12)$$

де $k(\vec{x}_j, \vec{x}_i)$ — радіально базисна керн-функція вигляду

$$k(\vec{x}_j, \vec{x}_i) = \exp(-\gamma \|\vec{x}_j - \vec{x}_i\|^2), \quad (13)$$

де $\gamma = \frac{1}{2\varepsilon^2}$.

3. ПОПЕРЕДНЯ ОБРОБКА СПЕКТРІВ ВІДБИТТЯ ЛИСТКІВ РОСЛИН ТА РЕЗУЛЬТАТИ ЧИСЕЛЬНИХ РОЗРАХУНКІВ

Перед проведенням чисельних обрахунків здійснено відбір спектральних даних озимої пшениці. Відбір даної культури зумовлений особливостями різноманітності її архітекtonіки. Вимірювання спектрів відбиття проводилися на шістьох ділянках, кожна із яких розбивалася на три частини. Для кожної ділянки проведено 5 вимірювань, які відповідали різним фазам вегетації рослини. Для отримання спектрів відбиття використовувався спектрометр [7] із роздільною здатністю 2 нм в діапазоні 400–750 нм. Після усереднення та попереднього опрацювання даних, загальна кількість отриманих спектрів відбиття склала 140. Після вимірювань із кожної дослідної ділянки (ділянка розміру поля зору приладу) відібрано листки із різних ярусів культури для аналізу вмісту хлорофілу біохімічним методом в лабораторних умовах [7].

Отриманий в ході вимірювань набір даних розбивався на дві підгрупи, перша — навчальна, друга — тестова. Для проведення процедури навчання реалізовано алгоритми центрування та нормування вхідних даних. Перш за все початок системи координат зсувається в “центр” за формулою

$$\bar{x} = \frac{1}{J} \sum_{j=1}^J (x_{1j} + \dots + x_{Nj}) / N. \quad (14)$$

Після знаходження \bar{x} розраховувалася нормуюча величина

$$\hat{x} = \frac{1}{J} \sum_{j=1}^J \sqrt{\frac{1}{N} \sum_{n=1}^N (x_{nj} - \bar{x})^2} \quad (15)$$

що дає змогу провести процедуру автомасштабування експериментальних даних

$$\tilde{x}_{jn} = \frac{x_{jn} - \bar{x}}{\hat{x}}. \quad (16)$$

Отриманні дані в подальшому використовувалися для розв'язку оптимізаційної задачі (12) із застосуванням модифікації алгоритму Sequential Minimal Optimization [15].

При навчанні моделі використовувався набір даних із 80 спектрів відбиття, решта застосовувалася для тестування точності оцінювання. Для отримання оптимальних значень керуючих параметрів γ та C , і забезпечення стійкості моделі щодо варіативності вхідних даних, проведено процедуру крос-валідації. Також здійснено порівняння отриманих результатів оцінювання із широко вживаним деривативним методом визначення концентрації хлорофілу [3, 17]. Результати порівняння наведені в табл. 1 та продемонстровано на рис. 2. Коефіцієнт кореляції на навчальній вибірці даних становить 0,91 при середньому квадраті похибки $MSE=0,246$, а для тестової — коефіцієнт кореляції рівний 0,84 при $MSE=0,429$.

Таблиця 1. Результати порівняння розрахунку концентрації хлорофілу для регресії опорних векторів та деривативного методу

Найм. кривої	Регресія опорних векторів	Деривативний метод	Біохімічний метод
RP62.dat	2,53	2,84	2,33
RP13.dat	2,69	3,20	2,90
RP50.dat	4,28	4,30	3,75
RP4.dat	4,78	4,95	4,76
RP40.dat	5,64	5,67	6,29
RP20.dat	6,33	6,02	6,57
RP19.dat	6,6	6,67	6,82
RP58.dat	7,31	7,45	7,65
RP35.dat	7,7	7,53	8,08
SSE =	1,1539	1,7406	

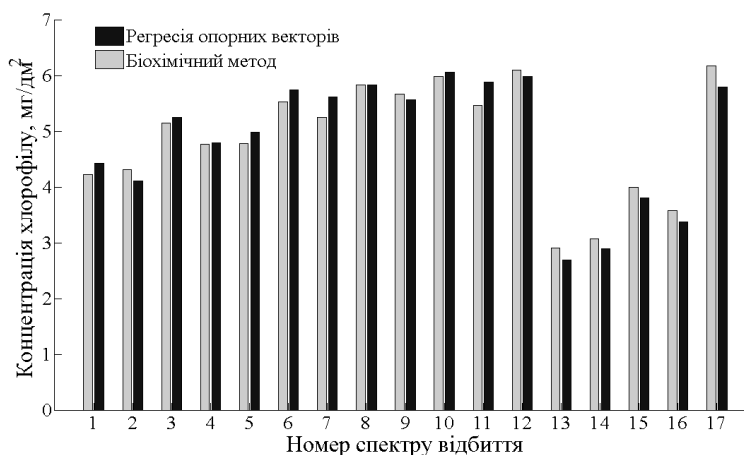


Рис. 2. Порівняння результатів отриманих на основі використання регресії опорних векторів та біохімічного методу

4. ВИСНОВКИ

Продемонстровано підхід оцінювання стану рослинності, а саме визначення концентрації хлорофілу, як індикатора забруднень навколишнього середовища. Запропоновано застосовувати метод опорних векторів для побудови регресійної моделі оцінювання концентрації біохімічних компонент рослин. Встановлено, що існує нелінійна залежність між спектром відбиття рослини та вмістом хлорофілу, яка з високою точністю апроксимується регресією опорних векторів. Проведено алгоритмічну реалізацію моделі та її тестування на еталонних спектральних вимірюваннях озимої пшениці. На основі отриманих результатів проведено порівняльний аналіз побудованої регресійної моделі із раніше запропонованими підходами, який показав, що отримана модель має вищу точність оцінювання та забезпечує оцінку при різному рівні наявних шумів в порівнянні із лабораторним біохімічним методом. Така стійкість досягається за допомогою процедур крос-валідації та системного відбору навчальних даних.

1. Андреева А.В., Бузников А.А., Тимофеев А.А. и др. Оценка экологического состояния окружающей среды по спектрам отражения индикаторных видов растительности // Современные проблемы дистанционного зондирования Земли из космоса. — 2006. — **3**, № 2. — С. 265–270.
2. 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.
3. Бідюк П.І., Митник О.Ю. Застосування генетичного алгоритму в задачах оцінювання вмісту хлорофілу в рослинності // Наукові вісті НТУУ “КПІ”. — 2004. — № 4. — С. 65–70.
4. 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.
5. Семенов О.В., Шатохина Ю.В., Яценко В.О. Валидация моделей классификации гиперспектральных данных // Проблемы управления и информатики. — 2008. — № 3. — С. 113–119.
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. Albayrak S. Use of reflectance measurements for the detection of N, P, K, ADF and NDF Contents in sainfoin pasture // Sensors. — 2008. — № 8. — P. 7275–7286. DOI: 10.3390/s8117275
9. Bousquet O., Boucheron S., Lugosi G. Introduction to statistical learning theory // Advanced Lectures on Machine Learning. — 2003. — P. 169–207.
10. Choe E., Meer F., Ruitenbeek F., et al. Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the Rodalquilar mining area, SE Spain // Remote Sensing of Environment. — 2008. — **112**. — P. 3222–3233.
11. Delalieux S., Auwerkerken A., Verstraeten W.W., et al. Hyperspectral reflectance and fluorescence imaging to detect scab induced stress in apple leaves // Remote sensing. — 2009. — № 1(4). — P. 858–874. doi:10.3390/rs1040858
12. 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.
13. Lamb D.W., Steyn-Ross M., Schaare P., et al. Estimating leaf nitrogen concentration in rye grass (*Lolium* spp.) pasture using the chlorophyll red-edge: theoretical modelling and experimental observations // Int. J. Remote Sens. — 2002. — № 23. — P. 3619–3648.
14. Lelong C.C.D., Roger J.-M., Brégand S., et al. Evaluation of oil-palm fungal disease infestation with canopy hyperspectral reflectance data // Sensors. — 2010. — **10**. — P. 734–747. doi:10.3390/s100100734
15. Platt J.C. Fast training of support vector machines using sequential minimal optimization / In B. Scholkopf, C.J.C. Burges, A.J. Smola (Eds), Advances in kernel methods — support vector learning // Cambridge, MA: MIT Press. — 1998. — P. 185–208.
16. 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.
17. 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.
18. Smola A.J., Scholkopf B. A tutorial on support vector regression // Statistics and Computing. — № 14. — 2004. — P. 199–222.
19. Suárez L., Zarco-Tejada P.J., Berni J.A.J., et al. Modelling PRI for water stress detection using radiative transfer models // Remote Sensing of Environment. — 2009. — **113**. — P. 730–744.
20. 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 29.07.2014