

С. А. Станкевич, О. В. Титаренко, С. В. Шкляр

## Ефективна обробка даних польового спектрометрування в природоресурсних задачах

*(Представлено академіком НАН України В. І. Ляльком)*

*Запропоновано математичні моделі та алгоритми ефективного аналізу тонких спектральних відмін наборів прецизійних спектрограм, отриманих у ході польового спектрометрування при розв'язанні природоресурсних задач. Застосування запропонованих алгоритмів дозволяє скоротити потрібні обчислювальні витрати в більш ніж 50 разів та підвищити вірогідність вдвічі.*

**Застосування польового спектрометрування.** Сучасні дані дистанційного зондування використовуються для отримання характеристик природних утворень — рослинності, гірських порід та ґрунтів, водних поверхонь. Вимоги до якості й точності характеристик зростають з підвищенням складності створюваних біофізичних моделей, удосконаленням методів наукових досліджень і технологій дистанційного зондування.

Для більш ефективного використання даних дистанційного зондування необхідні спектри природних утворень, що отримані за допомогою польових спектрометрів, потрібних для з'ясування оптимальних технічних характеристик аерокосмічних сенсорів.

Польова спектрометрія — це кількісне вимірювання яскравості, освітленості, відбиття або пропускання об'єктів земної поверхні в польових умовах як функції довжини хвилі. Виміри проводили за допомогою спектрометрів, які дають змогу вивчати зразки поза традиційних лабораторій, що потребують дорогих методів тестування. Крім того, польову спектрометрію можна застосовувати і як самостійний інструмент для виявлення процесу або речовини (матеріалу). Це представляє інтерес, тому що пряма ідентифікація речовин у полі важливіша за збір зразків для подальшого аналізу в лабораторії.

Існує велика кількість областей застосування спектральних польових даних у сільському господарстві, екології і лісництві, при розвідці корисних копалин, у гірській промисловості, ґрунтознавстві. В деяких випадках спектральну інформацію використовували як довідкові дані для інтерпретації спектральних зразків, в інших випадках — як безпосередню інформацію. Завдяки спектральному аналізу досліджено взаємозв'язок між спектральною інформацією та хімічними і біофізичними властивостями вказаних об'єктів.

**Розділення спектрограм.** Розділення близьких спектрограм здійснюється за тонкими спектральними відмінностями, а отже, проводиться по результатах робастної статистичної обробки. Досить надійні результати дає використання як показника роздільності ймовірності  $\varepsilon$  переплутання спектрограм, що розділяються [1].

Виміряні спектрограми можна вважати реалізаціями ергодичних випадкових процесів [2]. Імовірність  $\varepsilon_{12}$  переплутування спектрограм (1) й (2) визначається законами розподілу густини ймовірності їх сигналів за спектром  $p_1(\lambda)$  й  $p_2(\lambda)$ :

$$\varepsilon_{12} \simeq \exp(-n \cdot D_{12}), \quad (1)$$

де  $n$  — обсяг статистичної вибірки;  $D_{12}$  — дивергенція Кульбака–Лейблера [3].

На практиці зазвичай виконується серія спектрометрувань, яка потім, за результатами сумісної статистичної обробки, поділяється на об'єктові та фонові вимірювання. Для кожного з  $n$  можливих поділів визначається ймовірність  $\varepsilon(i)$  і далі (при відсутності додаткової інформації) на підставі принципу максимальної правдоподібності можна знайти номер  $i^*$  найкращого в розумінні статистичної значущості поділу. На жаль, пряма обробка безпосередніх результатів польового спектрометрування, як правило, демонструє досить високі значення ймовірності помилки  $\varepsilon(i^*) \geq 0,4$ . Покращити достовірність розділення спектрограм можна за допомогою селекції найбільш інформативних спектральних відліків [4].

Величина інформаційної дивергенції буде залежати від складу підібраних спектральних відліків, які можна задати булівським вектором-селектором  $W$ , а разом з нею — й імовірність помилки (1), причому  $D(i, \mathbf{0}) \equiv 0$  та відповідно  $\varepsilon(i, \mathbf{0}) \equiv 1$ . Тому обов'язково існує певний оптимальний вектор-селектор  $W^*$ :

$$W^* : \varepsilon(i^*, W^*) \rightarrow \min. \quad (2)$$

Задача полягає в знаходженні оптимального за виразом (2) вектора-селектора. Розв'язок цієї задачі у випадку великої кількості відліків неможливо отримати прямим перебором. Для кількох сот спектральних відліків добре себе зарекомендував алгоритм псевдоградієнтного пошуку на регулярній решітці в просторі змінних вектора  $W$  [5]. Але для кількох тисяч спектральних відліків, які продукують сучасні польові спектрометри, обчислювальні витрати навіть псевдоградієнтного пошуку занадто високі. Потрібно зниження вхідної розмірності.

**Оптимальна лінеаризація спектрограм.** Ефективним інструментом зниження розмірності спектрометричних даних є лінеаризація. Особливість лінеаризації при обробці даних польового спектрометрування полягає в необхідності визначення єдиного спільного набору абсцис вузлових точок лінеаризації для всієї множини спектрограм, що обробляються. Це викликано потребою подальшого коректного порівняння різних лінеаризованих спектрограм.

Для кожної окремої спектрограми математична задача зводиться до знаходження кусково-лінійної функції ( $m$ -вимірну функцію однієї змінної — спектрального відліку), всі значення якої лежать у наперед заданому околі таблично заданої функції — значень спектрограми.

Шукаємо кусково-лінійну функцію  $f: [1 \dots m] \rightarrow \mathbb{R}^n$ , абсциси точок зламу якої — цілі точки, а ординати лежать у межах  $L_i(j) \leq f_i(j) \leq H_i(j)$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ .

Це робиться в два етапи:

- 1) знаходяться можливі абсциси точок зламу;
- 2) для кожного  $i = 1, \dots, n$  визначається одновимірна кусково-лінійна функція  $f_i$  зі знайденими на першому етапі абсцисами точок зламу (тобто зі знайденими інтервалами лінійності), значення якої лежать у межах  $L_i(j) \leq f_i(j) \leq H_i(j)$ ,  $j = 1, \dots, m$ .

Рис. 1 ілюструє можливість оптимальної лінеаризації типової спектрограми рослинності, отриманої за допомогою польового спектрометра FieldSpec 3FR. Практика обробки даних польового спектрометрування демонструє можливість лінеаризації набору спектрограм з кількох тисяч спектральних відліків 30–40 лінійними сегментами з  $\pm 0,5$  %.

**Розрахунок інформаційної дивергенції.** Після проведення лінеаризації спектрограм стає можливим аналітично визначити інтеграл, що описує інформаційну дивергенцію. Розглянемо лінійні сегменти двох спектрограм, які описуються рівняннями  $a_1 \lambda + b_1$

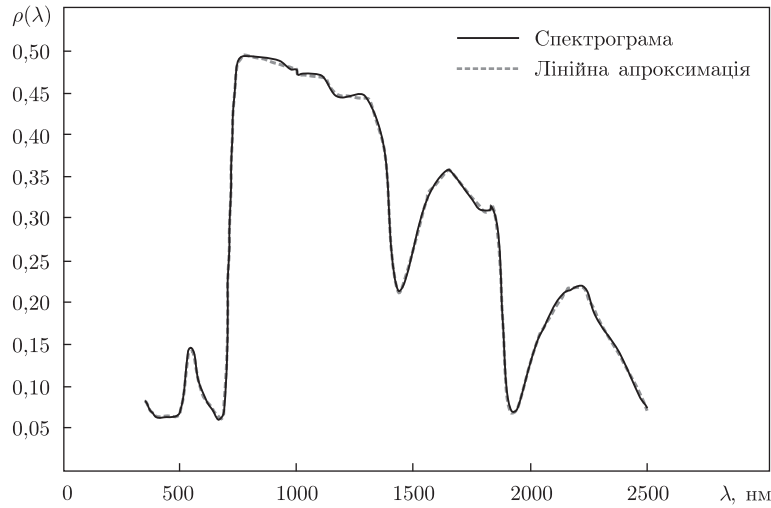


Рис. 1. Оптимальна лінеаризація типової прецизійної спектрограми

та  $a_2\lambda + b_2$ , де  $a$  і  $b$  — константи. Закони ймовірнісних розподілів у межах сегмента можна записати як

$$p_1(\lambda) = \frac{a_1\lambda + b_1}{s_1}, \quad p_2(\lambda) = \frac{a_2\lambda + b_2}{s_2}, \quad (3)$$

де  $s_i = \sum_{j=1}^m \int_{\lambda_j}^{\lambda_{j+1}} (a_{ij}\lambda + b_{ij}) d\lambda$  — нормувальний множник.

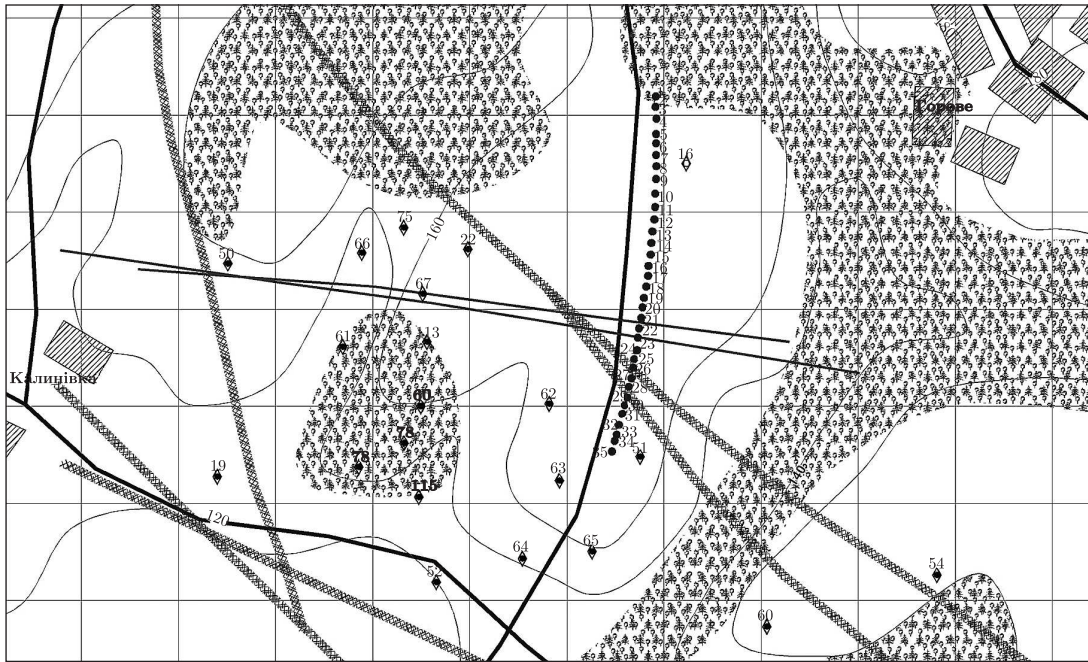
Повна дивергенція Кульбака–Лейблера має вигляд

$$D_{12} = \sum_{j=1}^m \int_{\lambda_j}^{\lambda_{j+1}} \frac{a_{1j}\lambda + b_{1j}}{s_1} \ln \frac{s_2 (a_{1j}\lambda + b_{1j})}{s_1 (a_{2j}\lambda + b_{2j})} d\lambda. \quad (4)$$

Інтеграл в правій частині рівняння (4) легко беруться заміною змінних:

$$\begin{aligned} & \int_{\lambda_j}^{\lambda_{j+1}} \frac{a_1\lambda + b_1}{s_1} \ln \frac{s_2 (a_1\lambda + b_1)}{s_1 (a_2\lambda + b_2)} d\lambda = \\ & = \frac{1}{s_1} \left[ \frac{(a_1\lambda_{j+1} + b_1)^2}{4a_1} (2 \ln(a_1\lambda_{j+1} + b_1) - 1) - \frac{(a_1\lambda_j + b_1)^2}{4a_1} (2 \ln(a_1\lambda_j + b_1) - 1) - \right. \\ & - [(2c_1(a_2\lambda_{j+1} + b_2) + 4c_2) \ln(a_2\lambda_{j+1} + b_2) - c_1(a_2\lambda_{j+1} + b_2) - 4c_2] \frac{a_2\lambda_{j+1} + b_2}{4a_2} + \\ & + [(2c_1(a_2\lambda_j + b_2) + 4c_2) \ln(a_2\lambda_j + b_2) - c_1(a_2\lambda_j + b_2) - 4c_2] \frac{a_2\lambda_j + b_2}{4a_2} + \\ & \left. + \ln \frac{s_2}{s_1} \left( a_1 \frac{\lambda_{j+1} + \lambda_j}{2} + b_1 \right) (\lambda_{j+1} - \lambda_j) \right], \quad (5) \end{aligned}$$

де  $c_1 = \frac{a_1}{a_2}$ ,  $c_2 = b_1 - b_2 \frac{a_1}{a_2}$  — допоміжні константи.



Умовні позначення: ● точки, де проводилося спектрометрування; розломи: за даними сейсмічних робіт, — за даними буріння; свердловини: продуктивна, непродуктивна; \* \* \* лісові насадження; дороги; горизонталі; населені пункти

Рис. 2. Розташування просторового профілю, вздовж якого проводилося спектрометрування рослинності

Відтепер за рівнянням (4) з урахуванням виразу (5) стає можливим безпосередньо розрахувати дивергенцію Кульбака–Лейблера для довільних двох лінеаризованих спектрограм. Крім того, стає реальним проведення селекції інформативних спектральних інтервалів (2), оскільки вектор-селектор  $W$  буде накладено не на безліч маловпливових вхідних спектральних відліків, а на набагато важливіші лінійні сегменти спектрограм, у просторі можливих комбінацій яких і буде здійснюватися псевдоградієнтний пошук.

Можна показати [6, 7], що для типового зниження розмірності в результаті лінеаризації у 50–90 разів скорочення обчислювальних витрат складе в середньому 40–270 тис. разів. Обчислювальні ж витрати на лінеаризацію можна вважати постійними, вони будуть приблизно лінійно зростати лише зі збільшенням кількості вхідних спектрограм  $n$ .

**Апробація на польових даних.** Апробацію запропонованої моделі здійснювали для визначення границі контуру покладу вуглеводнів шляхом розділення спектрограм рослинності за простором уздовж профілю, який перетинає продуктивну та непродуктивну ділянки покладу вуглеводнів. Профіль було прокладено в межах однорідних рослинних фацій ландшафту з урахуванням геологічної будови — вхрест простягання геологічної структури, між непродуктивною та продуктивною пошуковими свердловинами (рис. 2).

Прецизійне спектрометрування виконували за допомогою сучасного польового спектрометра типу FieldSpec 3FR. Відбір листя берези здійснено 17.06.2009 р. за пошуковим маршрутом уздовж лісосмуги. Проведено спектрометрування листя берези в 35 точках уздовж обраного профілю.

Результати обробки даних польового спектрометрування шляхом оптимальної лінеаризації набору отриманих спектрограм ілюструє рис. 3.

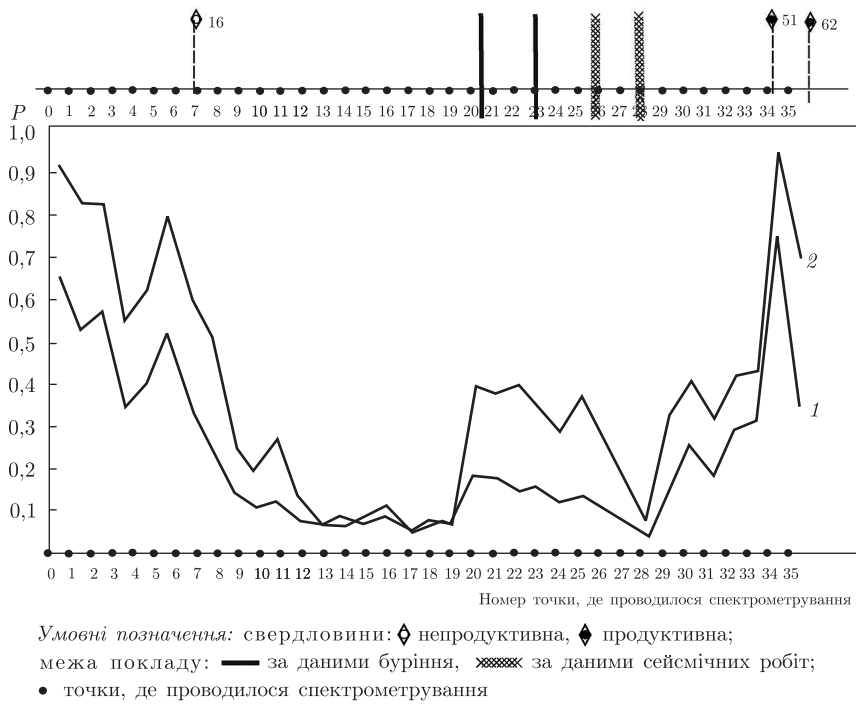


Рис. 3. Результати розділення спектрограм рослинності вздовж просторового профілю: 1 — лінеаризованих, 2 — лінеаризованих з селекцією найбільш інформативних сегментів

За результатами обробки спектрограм з використанням запропонованої методики максимальна вірогідність розділення відповідає 5-й та 34-й точкам просторового профілю, що пояснюється наявністю там розбурених свердловин (№ 16, № 51 й № 62). На цих ділянках спостерігається значне забруднення території, яке впливає на спектральні характеристики рослин. Крім того, Східнорогінцівська нафтогазоносна структура має достатньо високу неотектонічну активність, про що свідчать інтенсивні сучасні флюїдогеодинамічні процеси в межах її блоків. Збільшення вірогідності розділення спектрограм рослинності в точках 19–28 пов'язане з геологічними розломами та вказує на високу флюїдопровідність цієї ділянки внаслідок наявності зони підвищеної тріщинуватості в породах фундаменту. Отже, спектрограми рослинності в точках 1–19 можуть бути віднесені до продуктивної ділянки покладу вуглеводнів, а в точках 19–35 — до непродуктивної.

Таким чином, запропонована методика статистичного розділення спектрограм рослинності шляхом оптимальної лінеаризації з подальшою селекцією інформативних спектральних інтервалів демонструє принципову можливість виявлення закономірних змін спектральних характеристик рослинності при переході від пошукових ділянок до фонових.

Запропонована модель статистичного розділення спектрограм є досконалою теоретичною основою методики виявлення об'єктів за їх спектрами по матеріалах прецизійного польового спектрометрування і може бути успішно адаптована для розв'язання певного кола природоресурсних задач (у сільському господарстві, екології і лісництві, при розвідці корисних копалин, у гірській промисловості, ґрунтознавстві тощо).

Оптимальна лінеаризація є ефективним інструментом зниження розмірності спектрометричних даних. Попередня лінеаризація даних польового спектрометрування дозволяє описати набір спектрограм з кількох тисяч спектральних відліків (кожна з 30–40 лінійними сегментами), що, в свою чергу, дозволяє значно скоротити потрібні обчислювальні

витрати та створити можливість ефективної обробки спектрометричних даних саме в польових умовах.

Методика, технічні рішення та практичні рекомендації, які можна розробляти на підставі викладених підходів, спроможні значно розширити функціональність та підвищити ефективність існуючих апаратно-програмних комплексів обробки даних польового спектрометрування. Наприклад, відбір найбільш інформативних лінійних сегментів дає змогу підвищити вірогідність правильного розділення спектрограм у 2–3 рази.

Подальші дослідження мають бути спрямовані на розробку алгоритмів ідентифікації більш складних просторових розподілів результатів польового спектрометрування, на теоретичне обґрунтування створення нових ефективних інваріантів для аналізу прецизійних спектрограм, а також на залучення додаткових наземних геологічних, геофізичних та дистанційних даних для підвищення ефективності розв'язку природоресурсних задач [8].

1. *Архинов А. И., Станкевич С. А., Титаренко О. В.* Определение границы контура залежи углеводородов по признаку статистической делимости спектрограмм растительного покрова // Материалы Всеросс. конф. “Дегазация Земли: геодинамика, геофлюиды, нефть, газ и их парагенезы”. – Москва: ИПНГ РАН, 2008. – С. 33–36.
2. *Станкевич С. А.* Кількісне оцінювання інформативності гіперспектральних аерокосмічних знімків при розв'язанні тематичних задач дистанційного зондування Землі // Доп. НАН України. – 2006. – № 10. – С. 136–139.
3. *Martinez-Usó A., Pla F., Sotoca J. M., Garcia-Sevilla P.* Clustering-Based Hyperspectral Band Selection Using Information Measures // IEEE Transact. on Geosci. and Rem. Sensing. – 2007. – **45**, No 12. – P. 4158–4171.
4. *Станкевич С. А., Титаренко О. В.* Оптимізація даних наземного спектрометрування при картуванні меж фітоіндикаційних аномалій // Наук. вісн. Нац. аграрн. ун-ту. – 2008. – Вип. 128. – С. 312–319.
5. *Станкевич С. А.* Оптимизация состава спектральных каналов гиперспектральных аэрокосмических изображений при решении тематических задач дистанционного зондирования Земли // Косм. наука і технологія. – 2007. – **13**, № 2. – С. 25–28.
6. *Floudas C. A.* Deterministic Global Optimization: Theory, Algorithms and Applications. – Oxford: Oxford Univ. Press, 1995. – 650 p.
7. *Попов М. А., Станкевич С. А.* Методы оптимизации числа спектральных каналов в задачах обработки и анализа данных дистанционного зондирования Земли // Современные проблемы дистанционного зондирования Земли из космоса. – Вып. 3. Т. 1. – Москва: ИКИ РАН, 2006. – С. 106–112.
8. *Станкевич С. А., Титаренко О. В.* Методика інтеграції дистанційних та геолого-геофізичних даних при пошуку нафти та газу // Уч. зап. Таврич. нац. ун-та ім. В. И. Вернадского. – 2009. – **22 (61)**, № 1. – С. 105–113.

*ДУ “Науковий центр аерокосмічних досліджень Землі  
Інституту геологічних наук НАН України”, Київ*

*Надійшло до редакції 26.03.2010*

**S. A. Stankevich, O. V. Titarenko, S. V. Shklyar**

### **Field spectrometry effective data processing in the natural resource management**

*Mathematical models and algorithms for the effective analysis of fine spectral differences between the sets of precision spectrograms registered during a field spectrometry in the natural resource management are offered. These algorithms reduce the necessary computing expenses by more than 50 times and increase the reliability twice.*