УДК 519.2

А.Я. Білецький, І.В. Шелевицький, В.М. Шутко, Ю.В. Юрко

АДАПТИВНА ОБРОБКА ГЕОХІМІЧНИХ АНАЛІЗІВ ЗА ДОПОМОГОЮ СПЛАЙНОВОЇ МОДЕЛІ

Запропоновано адаптивний алгоритм обробки геохімічних аналізів за допомогою сплайнової моделі. Показано можливості сплайнів виявляти різнорідні області в складі залізної руди, які, зважаючи на значну дисперсію випадкової складової, візуально помітити практично неможливо.

У прикладних задачах розвідки та видобування корисних копалин важливим є питання визначення їхньої кількості в певному обсязі. Для розв'язання цієї задачі бурять свердловини, породу з яких аналізують на вміст тих чи інших компонентів. Зрозуміло, що такий аналіз є вибірковим. Тобто вміст компонентів аналізують з певним інтервалом. В отриманих результатах суттєвою є випадкова мінливість результатів, що обумовлена багатьма факторами: як особливостями формування покладів в конкретній області, так і методами буріння та аналізу. При аналізі таких даних головними є дві задачі:

- виявлення невипадкової тенденції в розподілі компонентів у просторі (по глибині свердловини);
- виявлення областей з однорідним розподілом компонентів і границь таких областей.

У багатьох випадках зміна вмісту компонентів носить плавний неперервний характер. Навіть при наявності чітких границь процес буріння та особливості технології збирання зразків роблять відмінність плавною.

Отже, для виділення детермінованої складової бажано мати гнучку неперервну модель, здатну адекватно передавати складний характер залежності на значних інтервалах. Традиційно для розв'язку таких задач користуються алгебраїчними поліномами. Недоліки таких моделей відомі. Складний характер залежності змушує збільшувати порядок полінома, що призводить до росту обсягів розрахунків і появи осциляцій полінома. Аналогічна картина спостерігається і при використанні тригонометричних поліномів. Підбір специфічних функцій потребує індивідуального підходу до кожного набору даних, що ускладнює автоматизацію процесу і не гарантує успіху.

Хорошою альтернативою для згаданих моделей є сплайни, які поєднують простоту з гнучкістю. Однак задача виділення невипадкової складової (тренда) для сплайнової моделі також має певні труднощі.

Потрібно вибрати сплайн, який найбільше підходить. Досить відомими є глобальні сплайни, що мають неперервну першу та другу похідні. Глобальні сплайни мають властивість максимальної гладкості, яка полягає в мінімальності енергії деформації:

$$\int_a^b \frac{[S''(x)]\,dx}{\left[1+(S'(x))^2\right]^{5/2}},$$

де $S(x)$ – сплайн-функція, задана на інтервалі $[a,b]$ [1].

Однак розрахунок такого сплайна є досить складним, оскільки в його побудові беруть участь усі дані. Тому і задача згладжування такими сплайнами є досить складною. Значно зручнішими є ермітові сплайни, які мають неперервною лише першу похідну. Причому неперервність гарантується відповідністю першої похідної першій похідній залежності, що описується. Тому на формування ермітового сплайна впливають дані з найближчих сусідніх фрагментів від точки стикування. Це значно спрощує побудову сплайна і дозволяє ефективно скористатись методом найменших квадратів [2]. Вибір ермітових сплайнів обгрунтований також їхньою незначною відмінністю від глобальних виглядом залежності при інтерполяції. З іншого боку, в геологічних даних оптимальні властивості глобальних сплайнів не є обгрунтованими фізичним характером даних.

Ініша проблема полягає у виборі розташування точок стику (вузлів) сплайна. Кубічний ермітів сплайн повністю визначається положенням вузлових точок. Запишемо сплайн як функцію від числових параметрів – абсцис вузлів і ординат вузлів:

$$S(x)=\sum_{j=0}^{R} a_j \varphi_j(x,T),$$

де x – точка, що належить j-му фрагменту сплайна; a_j – значення сплайна в j-му вузлі; $\varphi_j(x,T)$ – базисна сплайн-функція, що залежить від вектора абсцис вузлових точок T.

Наслідком локальності базисних функцій є спрощення попереднього виразу до такого:

$$S(x)=a_{j-1}F1_{j-1}(x)+a_jF2_j(x)+a_{j+1}F3_{j+1}(x)+a_{j+2}F4_{j+2}(x),$$

де $F1_j(x)-F4_{j+2}(x)$ – складові відповідних базисних функцій.

Якщо дані можна описати як суму детермінованої складової U та некорельованих випадкових чисел Ξ ($Y=U+\Xi$), то значення сплайна у вузлових точках можна визначити методом найменших квадратів [2]:

$$A=(F^T F)^{-1}F^T Y,$$

де F – блочно-діагональна матриця.

Стовпці матриці F є значеннями відповідних базисних функцій. Специфічний вигляд цієї матриці забезпечує обчислювальну ефективність методу найменших квадратів для моделі з ермітовим кубічним сплайном.

Однак результат суттєво залежатиме від розміщення вузлів на осі абсцис. Задача оптимального розміщення вузлів є нелінійною задачею оптимізації багатовимірної функції багатьох змінних з обмеженнями. Розв'язок такої задачі градієнтними методами потребує досить великого обсягу обчислень, а успіх залежатиме від ряду суб'єктивних факторів.

Пропонується будувати сплайн поступово, адаптуючи його до даних. Тобто побудова сплайна починається з одного фрагмента – полінома третього степеня. Цей фрагмент продовжується до того часу, доки він адекватно описує детерміновану складову. Про адекватність моделі свідчитиме відсутність в залишках постійної складової. Як детектор наявності невипадкової складової використаємо непараметричний знаковий критерій [3].

Варто зазначити, що використання t критерію в поєднанні з методом найменших квадратів дає погані результати. Очевидно це зумовлено близькістю t критерію до критерію оптимальності методу найменших квадратів. Практично метод найменших квадратів забезпечує близьке до нуля значення t критерію.

Поява в залишках невипадкової складової є сигналом до встановлення вузлової точки і чатку наступного фрагмента сплайна. Адаптивний алгоритм побудови сплайна має такий игляд:

1) задати рівень значимості гіпотези про відсутність в залишках тренда $\alpha = 0{,}05$ начення α вказане для прикладу);

2) задати мінімальну довжину фрагмента сплайна, щоб він містив не менше чотирьох очок даних;

3) розрахувати початковий сплайн з трьох фрагментів мінімальної довжини (нехай овжина останнього фрагмента дорівнює *h)*;

4) отримати наступний відлік даних $y(t_i)$;

5) якщо $y(t_i)$ належить до останнього фрагмента, виконати п. 7;

6) перенести останній вузол сплайна в точку t_i;

7) розрахувати сплайн за методом найменших квадратів;

8) розрахувати залишки $y(t_i) - S(t_i)$ для двох останніх фрагментів сплайна;

9) за допомогою критерія серій перевірити наявність тренда в залишках;

10) якщо гіпотеза про наявність тренда відкидається, то виконати п. 4;

11) зафіксувавши останній вузол сплайна в t_{i-1} додати наступний фрагмент сплайна з станнім вузлом у точці $t_{i-1}+h$ (*h* вибираємо так, щоб остання точка належала сплайну);

12) виконати п. 4;

13) вихід.

Результати роботи алгоритму наведено на рис. 1 – 4. Виконували обробку компонентів типової свердловини. На рис. 1 – 4 знаками + показано результати хімічних аналізів, – статистичні оцінки вузлів сплайна з урахуванням довірчих інтервалів з рівнем довірчої ймовірності 0,90.

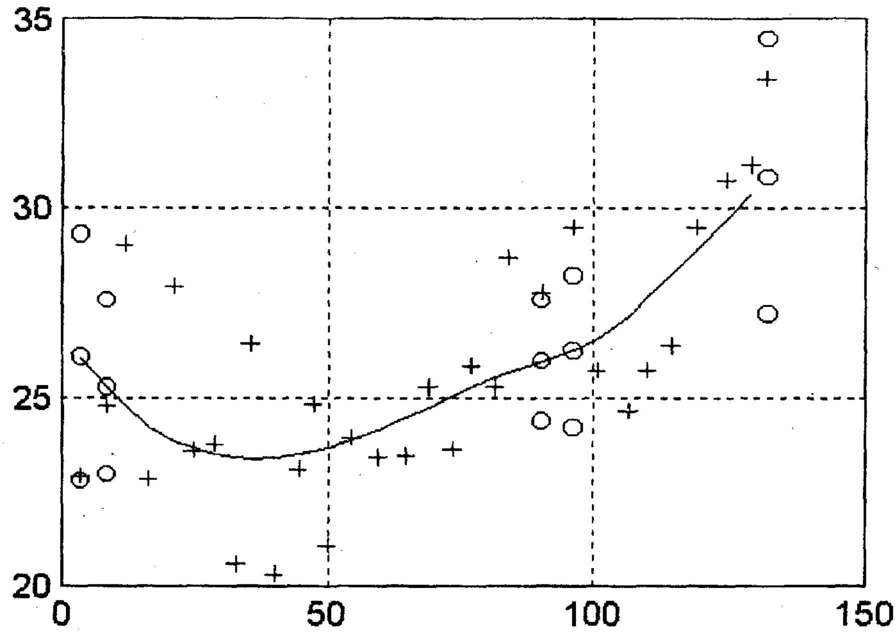

Рис. 1. Вміст магнітного заліза

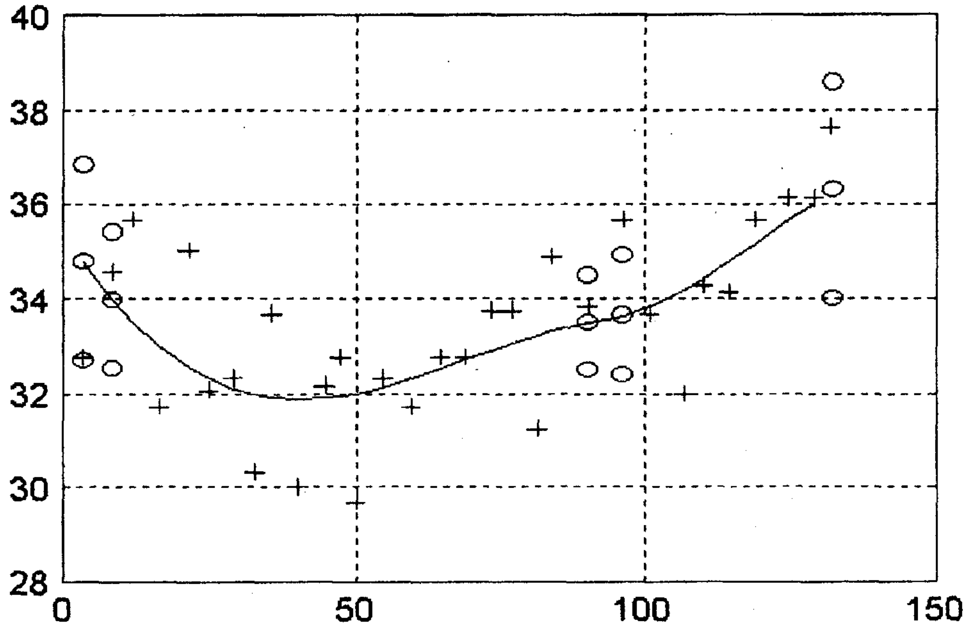

Рис. 2. Вміст заліза

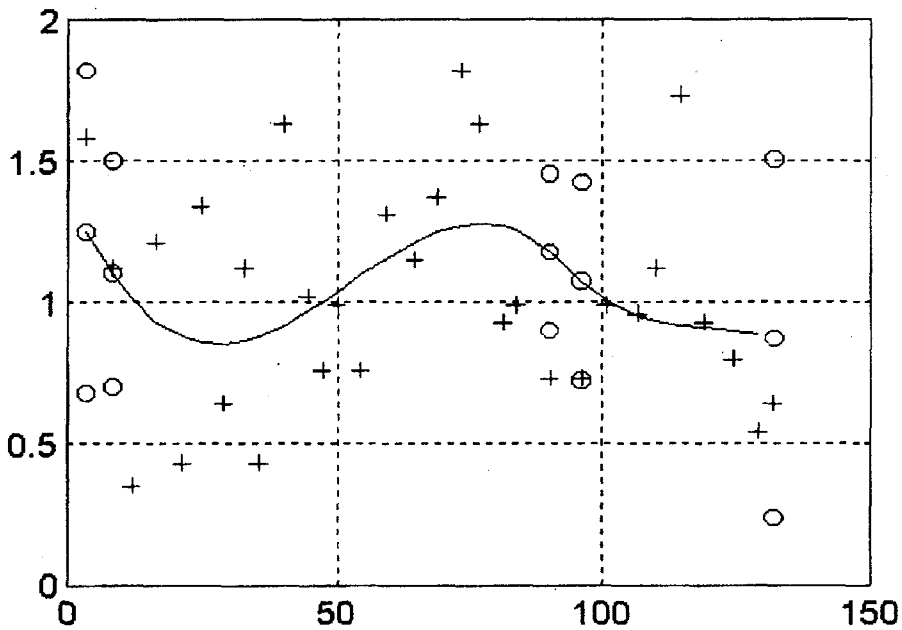

Рис. 3. Вміст алюмінію

По осі абсцис відкладено глибину свердловини, а по осі ординат – процентний вміст компонентів. Графіки містять дві пари подібних за характером поведінки компонентів. Схожі залежності вмісту заліза та магнітного заліза, алюмінію та окислу вуглецю (CO_2). Характерним для чотирьох графіків є наявність двох близьких вузлових точок на помітці 90. Це свідчить про наявність в даній області границі між різнорідними областями, що підтверджується іншими методами аналізу. З огляду на значну дисперсію випадкової складової візуально помітити на графіку таку відмінність практично неможливо.

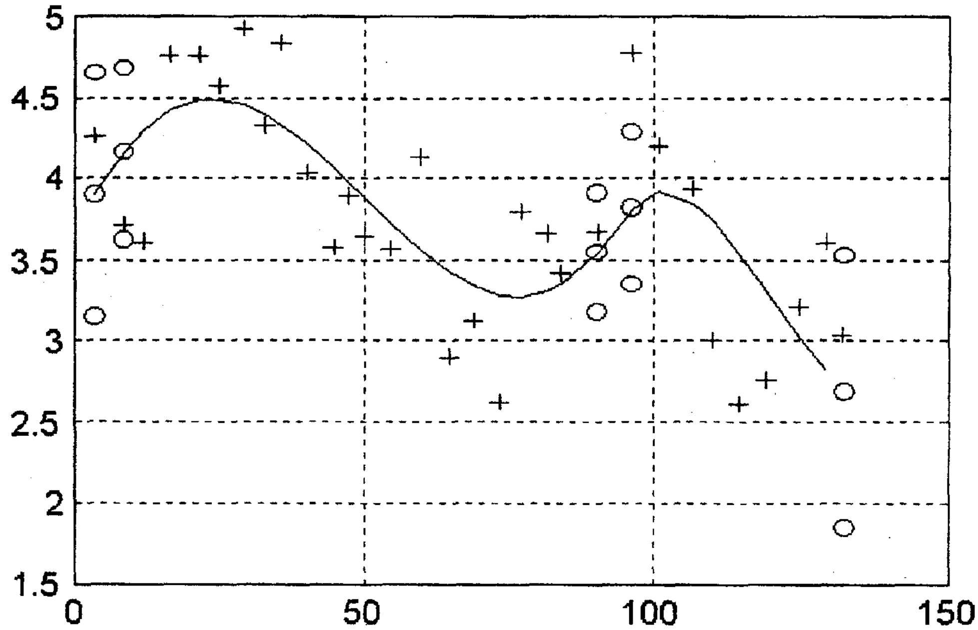

Рис. 4. Вміст окислу вуглицю (CO_2)

Обчислювальна ефективність алгоритму пояснюється також можливістю його рекурентної реалізації. Статистичні оцінки параметрів сплайна легко отримати за допомогою рекурентного методу найменших квадратів та його модифікації для послідовного збільшення числа оцінюваних параметрів [2]. Особливістю сплайнової моделі є відсутність залежності числа обумовленості системи нормальних рівнянь від числа оцінюваних параметрів. Число обумовленості практично відповідає аналогічному для алгебраїчного полінома третього степеня.

Список літератури

1. *Корнейчук Н.П.* Сплайны в теории приближения. – М.: Наука, 1984. –142 с.

2. *Бойко И.Ф., Шелевицкий И.В., Шутко В.Н.* Рекуррентный алгоритм построения сплайнов методом наименьших квадратов // Статистические методы обработки сигналов в авиационных радиоэлектронных системах: Сб.науч.тр. – К.: КМУГА, 1995. – С. 84.– 88.

3. *Бендат Дж., Пирсол А.* Прикладной анализ случайных данных: – М.: Мир, 1989. – 268 с.

Стаття надійшла до редакції 4 листопада 1999 року.