

МЕТОД БАГАТОВИМІРНИХ АНАЛОГІВ ДЛЯ КІЛЬКІСНОЇ ІНТЕРПРЕТАЦІЇ ДАНИХ ПРОМИСЛОВОЇ ГЕОФІЗИКИ: ДОПОВНЕННЯ ЧИ АЛЬТЕРНАТИВА ДО МЕТОДУ РЕГРЕСІЙНОГО АНАЛІЗУ

© М.Н. Жуков, К.Л. Яковець, 2009

*Київський національний університет імені Тараса Шевченка, Київ, Україна
ДП "Науканафтогаз" НАК "Нафтогаз України", Київ, Україна*

The article presents arguments for applying the method of multicomponent analogues for quantitative interpretation of well-logging data as a mean of upgrading a prognosis. Advantages and disadvantages of regressive analysis and the method of multicomponent analogues for interpretation of well-logging data are analysed. Estimation of prospects is given in regard to the development of GRID-technologies.

Істотне значення має ефективність та якість методів, що використовують для відтворення моделі геологічного середовища за результатами обробки полів різної фізичної природи. Ця ланка в процесі інтерпретації набуває великого значення внаслідок вагомого впливу на достовірність кінцевих результатів. Точністю останніх значною мірою визначається достовірність оцінки запасів родовища. Від неї, в свою чергу, залежить результат подальшої інвестиційної діяльності, спрямованої на освоєння покладів.

У петрофізичних моделях виділяють цільові та робочі параметри (А.Є. Кулінкович). Цільовими називають ті параметри, оцінка яких є головним завданням інтерпретаційної процедури. Це насамперед, підрахункові параметри – коефіцієнт ефективної пористості, коефіцієнт нафто- чи газонасиченості, ефективна потужність. У деяких випадках як цільові можна вважати й інші параметри – глинистість, густина, проникність тощо. Під час комплексної кількісної інтерпретації даних ГДС як робочі використовують геофізичні параметри – показники, отримані свердловинними вимірюваннями. Робочі параметри використовують для визначення цільових.

В загальному випадку маємо деяку послідовність з p цільових параметрів Y :

$$Y_p = \{y_1, y_2, \dots, y_p\},$$

та k робочих параметрів X :

$$X = \{x_1, x_2, \dots, x_k\}.$$

В узагальненому вигляді параметричну петрофізичну модель можна зобразити у вигляді

$$y_p = F_p(x_1, x_2, \dots, x_k, a_1, a_2, \dots, a_m) \pm A_q^{(p)},$$

де x_1, x_2, \dots, x_k – значення робочих параметрів; $A = \{a_1, a_2, \dots, a_m\}$ – параметри настройки; $A_q^{(p)}$ –

похибка непрямого виміру за довірчої ймовірності q [1].

Розв'язком прямої задачі є побудова петрофізичної моделі, тобто визначення вигляду функції F та встановлення параметрів настройки. Якщо вид функції F відомий, то розв'язок прямої задачі зводиться до визначення параметрів настройки.

Для настроювання петрофізичної або петрогеофізичної моделі потрібна еталонна колекція – набір значень цільових параметрів для конкретних зразків гірської породи (у випадку настройки петрофізичної моделі) або для конкретного розрізу, якщо виконують настроювання петрогеофізичної моделі. До еталонної колекції крім значень цільового параметра входять значення робочих параметрів. Відповідно, під час настроювання петрогеофізичної моделі мають бути відомі значення геофізичних параметрів для кожного інтервалу розрізу.

З розширенням комплексу методів ГДС за кількісної оцінки підрахункових параметрів набув широкого застосування регресійний аналіз. Для оцінки рівняння регресії існує багато програм як загального призначення, так і спеціалізованого з урахуванням задач ГДС. Здебільшого вони входять до комплексу інтерпретаційних систем.

Регресійний аналіз являє собою класичний інструмент параметричного методу інтерпретації в промисловій геофізиці. Не піддаючи сумніву ефективність цього підходу, зупинимось на деяких проблемах параметричного методу.

З теорії методу найменших квадратів випливає: щоб запропонувати адекватну регресійну модель, треба апіорі задати параметричну сім'ю, якій вона належить. Параметричні сім'ї, що зазвичай використовують, виходять із загальної теорії апроксимації функції, а не з фізико-хімічних моделей. Можна вважати, що в цій сфері домінує емпіричний підхід. Не менш важливим є питання закону розподілу. Тут переважно вжи-

вається запозичена з математичної статистики модель нормального розподілу. На широко відомий факт неузгодження реальних промислово-геофізичних матеріалів з гіпотезою нормального розподілу доводиться закривати очі у сподіванні, що він не надто зіпсує результат.

У випадках ускладнених ситуацій – відхилення від стандартних регресійних моделей, розриви залежностей або зміна їх характеру, метод регресійного аналізу знижує дієвість. Доведення цього висновку подано у публікаціях [2–4]. Для одержання об'єктивних моделей треба охопити багатовимірний діапазон можливих станів показників. Крім того, є проблеми з реалізацією нерівноточної моделі, пов'язані із визначенням вагових коефіцієнтів за мінімізацією суми квадратів відхилень. Вихід здебільшого шукають емпірично у розширенні переліку базових функцій, подрібненні петрофізичних моделей, введенні додаткових складових. Попри загальну невизначеність, завадою для подібних трансформацій є обмеженість вибірок – проблема, вирішення якої часто залежить не від інтерпретатора, а від економічних чинників.

Непараметричний підхід виходить з того, що параметричні моделі відіграють у ньому другорядну роль, а подекуди взагалі відмовляється від них. На думку авторів, його потенціал у промисловій геофізиці далеко не вичерпаний. Винятком є успішне застосування його у класифікаційних процедурах – виділенні колекторів та оцінці їх характеру насичення. Результат розв'язання таких задач носить якісний характер, що не повною мірою розкриває можливості методу, його сферу застосування. Мається на увазі, що цей метод для вирішення кількісних завдань промислової геофізики не використовують [5].

Пропонуємо новий підхід для кількісної інтерпретації даних ГДС, зокрема для розв'язання задачі посередніх вимірювань, який виходить із сучасних можливостей геоінформаційних технологій. Він полягає у визначенні основних параметрів, необхідних для підрахунку запасів нафти і газу за даними ГДС у комплексі з петрофізичними дослідженнями керна, прямими методами тощо на основі застосування методу багатовимірних аналогів (МБА) з подальшою статистичною фільтрацією ряду мір відмінності аналогів. Для цього використано базу даних (БД) специфічного формату, що дає змогу враховувати різний ступінь достовірності даних, які потім застосовано у процедурах пошуку багатовимірних аналогів. Система, що обрана для розв'язання поставленої задачі, містить процедури пошуку багатовимірних аналогів у середовищі БД з недетермінованими значеннями ознак.

У БД створюємо окремий масив надійних визначень підрахункових параметрів колекторів нафти і газу (“спостережені дані”, що характери-

зують реальний стан об'єкта дослідження – керна чи отримані за прямими методами дані; “аналітичні дані”, що характеризують істинний стан об'єкта, змодельовані чи розраховані теоретично з урахуванням всіх можливих нюансів) та суміжна інформація про родовище, яке ці дані характеризують. Вказана частина БД утворює еталонну колекцію (ЕК). Спеціальний модуль відповідає за розвиток ЕК під час експлуатації БД.

Кожен елемент розрізу (інтервал розрізу з однаковими фізичними властивостями) описується в БД набором ознак. Останніми можуть бути дані геофізичних методів досліджень свердловин та суміжна інформація про родовище. Для інтервалу розрізу, що підлягає інтерпретації, процедурою пошуку багатовимірних аналогів визначається послідовність з найближчих аналогів ЕК на основі міри відмінності об'єктів [2–4], упорядкована за її зростанням від поточного об'єкта. Відповідні можливості забезпечують різні шкали реєстрації.

Візуалізація значень міри відмінності об'єктів під час визначення групи аналогів дає змогу керуватися перепадом значень міри відмінності у їх упорядкованому рядку. З наведеного випливає, що група аналогів може складатися з одного об'єкта або навіть – у разі неповноти ЕК – аналогів потрібної якості знайдено не буде. Кількісна оцінка цільового параметра (параметра, який треба визначити) виходить з відгуку фільтра, що є результатом певної обробки значень цього параметра в ряді об'єктів-еталонів. Оцінкою похибки такого непрямого вимірювання слугує довірчий інтервал, побудований з використанням параметрів достовірності еталонів, розмішених у БД.

Характерною особливістю обраної системи є спрямованість на комплексування різновидової інформації з автоматичним урахуванням достовірності складових. Останнє забезпечується наявністю статистичної шкали реєстрації, особливістю якої є супроводження зареєстрованого значення характеристикою його достовірності. Для виконання ролі такої супроводжувальної характеристики виявилось достатнім використання середнього квадратичного відхилення (с.к.в.) локального розподілу. Цю характеристику вводимо у БД разом із значенням показника. Супроводження нею кожного значення, що зберігається в БД, є обов'язковою умовою успішної роботи запропонованої технології. Проблема, що зазвичай виникає, – відсутність джерела для одержання оцінки зазначеної характеристики. Вихід полягає у автоматичному супроводженні кожного значення ознаки локальним с.к.в. – значенням, що вводиться в БД розрахунком за допомогою функції регресії.

Передумовою успіху запропонованої технології є кондиційність бази даних стосовно поставленої задачі. Близький до оптимального резуль-

тат можна очікувати в разі необхідної повноти БД, зокрема, наявності репрезентативних колекцій об'єктів, які можуть відігравати роль представників корисного сигналу, тобто охоплення максимального діапазону коливань прогнозованої характеристики. БД зі своєю експлуатацією набуває все більшої ефективності для розв'язку вказаної задачі за рахунок поповнення ЕК. Із набуттям вичерпної повноти ЕК задача визначення підрахункових параметрів вирішуватиметься оптимально, причому саме в ускладнених ситуаціях, які можна віднести до проблемних місць параметричного методу (методу регресії).

Для оптимального розв'язання задачі важливе значення має розробка комплексів параметрів-ознак, які і формують “ознаковий” простір. “Зайві” ознаки під час побудови вирішального правила, можуть завадити стійкості рішення і, отже, погіршити його прогностичні властивості.

Після проведення прогнозування отримуємо ряд із об'єктів ЕК, серед яких здійснювали пошук, що упорядковуються залежно від міри віддалі до об'єкта пошуку. Тут можливі два варіанти, від яких і залежать розв'язки. Перший варіант – спостерігається різкий перепад, починаючи з першого знайденого аналогу. Це означає, що об'єкт пошуку є “чужим” для цієї БД. Йому приписуємо значення шуканої характеристики, що має перший аналог, але це значення, певною мірою, умовне. В таких випадках відіграє свою негативну роль неповнота бази (некондиційність БД). Другий варіант – у послідовності аналогів спостерігається повільне зростання відстаней від першого аналогу. Для такого випадку пропонуємо використання фільтра. Фактично, йдеться про побудову фільтра, який здатний підвищити відношення сигнал–завада, тим самим поліпшити якість прогнозу, наприклад, за рахунок вагових коефіцієнтів, обернено пропорційних віддалям. Вікно фільтра (кількість аналогів, що підлягають фільтрації) індивідуальне для кожного об'єкта, до якого відшукують аналоги, та визначається поведінкою відстаней до цього об'єкта. За сталої кількості аналогів, що беруть участь у визначенні цільового параметра для об'єктів інтерпретації, маємо справу з тим, що пріоритетність часто залежить від порядку розташування еталонів-аналогів у БД, де здійснюється пошук. Питання оптимізації описаного фільтра повністю не вирішене – тут можливе втілення нових ідей та розробок, а відповідно, його вирішення складає потенціал удосконалення запропонованого методу.

Також є можливість інтерактивно регулювати внесок кожного знайденого аналогу у присвоєння шуканого значення цільового параметра об'єкта, що інтерпретується. Для реалізації такого підходу вводять вагові коефіцієнти до знайдених аналогів c_1, c_2, \dots, c_k . Збільшення вагового коефі-

цієнта приводить до збільшення ролі одного із знайдених аналогів. Нульовий коефіцієнт c_j виключає відповідний аналог з розгляду.

З метою експериментального підтвердження працездатності викладеного методу визначення підрахункових параметрів було змодельовано декілька різновидів парних і множинних залежностей [2–4]. Отримані результати продемонстрували високу ефективність запропонованого методу для оцінки підрахункових параметрів. Лише у теоретично “чистих” випадках, де апіорі регресійний метод діє оптимально, метод МБА дає ненабагато гіршу точність прогнозування. Також вивчали оптимальну кількість аналогів, що підлягають фільтрації у разі автоматичного (неінтерактивного) процесу інтерпретації, тобто без втручання людського фактору в процес. Проведено експерименти, коли використовували не один аналог, а ланцюг упорядкованих аналогів.

Метод передбачає можливість визначати стосовно об'єкта, що підлягає інтерпретації, чи належить він внутрішньо до середовища, охарактеризованого базою даних. Такий аналіз можна використовувати для якісної перевірки відповідності бази даних, за допомогою якої відшукуватимемо найближчі аналоги до об'єкта.

Створення глобальної проблемно-орієнтованої бази знань або інтегрування нових розробок в існуючу проблемно-орієнтовану базу даних, яка, з одного боку, відкрита для розширення, а з іншого – розрахована на включення у потужні системи інтегрованої інтерпретації, відкриває простір для поліпшення якості кількісної інтерпретації даних промислової геофізики.

Принцип роботи такої системи можна описати приблизно так. Геофізик-інтерпретатор починає інтерпретацію промислово-геофізичних даних зі сповіщення програмі всієї наявної інформації про родовище, яке досліджується. Програма порівнює ці спостереження з даними різного типу покладів, вказуючи наявність аналогічних покладів та риси подібності чи відмінності. Іншими словами, програма в результаті діалогу з геофізиком-інтерпретатором, за рахунок поступового додавання, внесення інформації, локалізує область інформації, серед якої здійснюватиметься пошук. Безпосередня інтерпретація розпочинається вже в межах знайденого осередку даних. Тут відшукуються багатовимірні еталони-аналоги (для яких відомі шукані характеристики – цільові параметри – порід-колекторів) до об'єктів інтерпретації.

Сучасний розвиток інформатизації дає підстави зробити оптимістичний прогноз щодо перспектив використання запропонованого методу для кількісної інтерпретації. Вони виходять з процесу розгортання GRID-технологій, який набирає оберти. Всесвітня GRID-мережа є найбільш досконалим на сьогодні засобом спільного викорис-

тання корпоративних інформаційних ресурсів. Оптимальне вирішення задачі кількісної інтерпретації методом багатовимірних аналогів полягає у створенні масштабної бази даних, що містить всю інформацію, набуту в промисловій геофізиці, зокрема інтерпретації каротажних даних. Значному поліпшенню стану кількісної інтерпретації в промисловій геофізиці України сприяла б мережева БД, в якій були би корпоративні учасники на рівні великих регіонів або навіть кількох країн. Є проблема конфіденційності – конкретна інформація про параметри того чи іншого родовища є комерційною таємницею і, як правило, недоступна для загального користування. Проте вона не є нездоланною, оскільки для успішного прогнозування потрібна інформація не про конкретне родовище, а про анонімні колекції, якими обмінюються на взаємовигідних засадах. З рештою, якщо учасники такої GRID-системи дійдуть до згоди щодо спільного використання інформаційних ресурсів, то це посилить якість інтерпретаційних робіт всіх учасників.

Для ефективного вирішення задачі інтерпретації геологічних розрізів за даними ГДС запропонованим методом треба оперувати репрезентативними еталонними колекціями. Такі колекції можуть бути створені як за умови їх безперервного накопичення впродовж часу, так і зацікавленими науковими та виробничими осередками. Здебільшого такі бази даних в геофізичних орга-

нізаціях існують, питання лише в ефективній технології спільного використання.

З теорії методу випливає, що для пошуку адекватних аналогів для кожного конкретного інтервалу розрізу необхідно мати кондиційну колекцію. Кондиційність буде забезпечена широтою об'єднаного інформаційного ресурсу корпоративних користувачів, можливо, розміщених у різних частинах світу.

1. Жуков Н.Н. Статистический анализ геологических данных. – Киев, 1995. – 552 с.
2. Довженко К.Л. Статистичне прогнозування підрахункових параметрів методом багатовимірних аналогів // Наук. вісн. Нац. гірн. ун-ту. – 2006. – № 3. – С. 41–44.
3. Жуков М.Н., Яковець К.Л. Прогнозування підрахункових параметрів методом багатовимірних аналогів. Експериментальна перевірка ефективності шляхом моделювання // Геоінформатика. – 2006. – № 4. – С. 46–51.
4. Жуков М.Н., Яковець К.Л. Метод багатовимірних аналогів для оцінки підрахункових параметрів вуглеводнів // Вісн. Київ. ун-ту. Сер. Геологія. – 2007. – Вип. 41. – С. 51–55.
5. Жуков М.Н., Яковець К.Л. Оцінка параметрів нафтогазонасиченості засобами багатовимірної статистики: стан та актуальні задачі // Моніторинг небезпечних геологічних процесів та екологічного стану середовища: VII Всеукр. наук. конф., Київ. – К., 2006. – С. 274–276.

Надійшла до редакції 21.03.2008 р.

М.Н. Жуков, К.Л. Яковець

МЕТОД БАГАТОВИМІРНИХ АНАЛОГІВ ДЛЯ КІЛЬКІСНОЇ ІНТЕРПРЕТАЦІЇ ДАНИХ ПРОМИСЛОВОЇ ГЕОФІЗИКИ: ДОПОВНЕННЯ ЧИ АЛЬТЕРНАТИВА ДО МЕТОДУ РЕГРЕСІЙНОГО АНАЛІЗУ

У статті розглянуто аргументи щодо застосування під час кількісної інтерпретації даних промислової геофізики методу багатовимірних аналогів як засобу підвищення якості прогнозу цільових параметрів. Проаналізовано переваги й недоліки регресійного аналізу та методу багатовимірних аналогів для інтерпретації промислово-геофізичних даних. Наведено оцінки перспективності методу у зв'язку з розвитком GRID-технологій.

Н.Н. Жуков, Е.Л. Яковець

МЕТОД МНОГОМЕРНЫХ АНАЛОГОВ ДЛЯ КОЛИЧЕСТВЕННОЙ ИНТЕРПРЕТАЦИИ ДАННЫХ ПРОМЫШЛЕННОЙ ГЕОФИЗИКИ: ДОПОЛНЕНИЕ ИЛИ АЛЬТЕРНАТИВА К МЕТОДУ РЕГРЕССИОННОГО АНАЛИЗА

В статье приводятся аргументы на применение при количественной интерпретации данных промышленной геофизики метода многомерных аналогов как средства повышения качества прогноза. Анализируются преимущества и недостатки регрессионного анализа и метода многомерных аналогов для интерпретации промышленно-геофизических данных. Приведены оценки перспективности в связи с развитием GRID-технологий.