

КИЇВСЬКИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

На правах рукопису

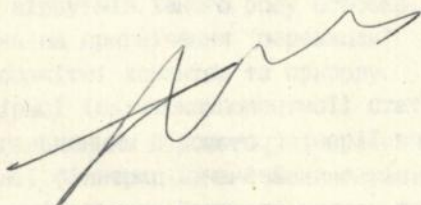
УДК 550.8.05;518.8

Жуков Микола Никанорович

МЕТОД БАГАТОВИМІРНОЇ СТАТИСТИЧНОЇ ФІЛЬТРАЦІЇ
РІЗНОВИДОВОЇ ГЕОЛОГІЧНОЇ ІНФОРМАЦІЇ ДЛЯ
ВИРІШЕННЯ ЗАДАЧ КАРТУВАННЯ ТА ПРОГНОЗУ

Спеціальність 04.00.05 - геоінформатика

Автореферат
дисертації на здобуття наукового ступеня
доктора геологічних наук



Київ - 1997



00752271 (O)

Дисертацією є рукопис.

Робота виконана на кафедрі мінералогії, геохімії та петрографії геологічного факультету ім. Тараса Шевченка.

Офіційні опоненти:

1. Доктор геолого-мінералогічних наук Б.О.Горлицький;
2. Доктор геолого-мінералогічних наук Е.Я. Жовинський;
3. Доктор фізико-математичних наук С.В.Мостовий.

Провідна установа - Державна гірнича академія України
(м.Дніпропетровськ)

Захист дисертації відбудеться 19 червня 1997 р. об 11 годині на засіданні спеціалізованої вченої ради геологічного факультету Київського університету ім. Тараса Шевченка Д.01.01.25 за адресою:

252022, Київ-22, вул.Васильківська, 90.Факс: (044)-266-23-14.

З дисертацією можна ознайомитися у бібліотеці Київського університету ім. Тараса Шевченка.

Автореферат розісланий 14 травня 1997 р.

Вчений секретар
спеціалізованої вченої ради
доктор геолого-мінералогічних наук
професор

В.Г.Молянко

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність проблеми.

Ресурс родовищ, які могли б бути відкриті класичними геологічними методами, сьогодні вичерпано. Успіх геологорозвідки пов'язується перш за все з інтеграційним системним аналізом різномірної, різномірневої, різнодостовірної інформації та її оптимізацією.

Проблеми, що відносяться до застосувань математичних методів в геології, активно розробляються на протязі останніх десятиріч зусиллями багатьох авторів. Це А.М. Бугаєць, Б.С. Бусигін, Н.Н. Боровко, Л.С. Галецький, С.В. Григорян, Б.О. Горлицький, А.І. Гаврішин, Г.Л. Гомберг, Р.І. Дубов, В.О. Дядюра, В.В. Ломтадзе, О.О. Нікітін, О.Є. Кулінкович, Р.І. Коган, С.В. Мостовий, Д.А. Родіонов, Б.І. Смирнов, О.П. Соловов, М.І. Толстой, Е.Н. Черемісіна та багато інших. Тема дисертації відноситься до методів та технологій інтегрованого (комплексного) аналізу різномірневої та різнопараметрової інформації. Передумовою інтенсифікації досліджень у цьому напрямку стали об'єктивні фактори.

Суть проблеми полягає у методології оптимального використання різномірної геологічної інформації для задач прогнозу та геологічного картування. Головна її трудність у комплексуванні даних, які реєструються в різних шкалах: номінальний, бінарний, числовий детермінований, числовий з елементом невизначеності. Її вирішено у цій роботі методологічно і практично - у вигляді діючої інформаційно-аналітичної системи на основі розробленого методу багатовимірної статистичної фільтрації різномірної інформації.

Як відомо, ідея фільтрації полягає у обробці даних таким чином, щоб результат обробки - відгук фільтру - був би при наявності корисного сигналу якомога більш помітним по відношенню до тих випадків, коли сигнал відсутній. Такого роду обробка, що називається фільтрацією, направлена на пригнічення "перешкоди" ("завади"), яка має часом вкрай різноманітний характер та природу.

Метод багатовимірної (багатокомпонентної) статистичної фільтрації є певним узагальненням відомого у теорії випадкових функцій методу статистичної фільтрації на ознаки з різнотипними шкалами, у тому числі з номінальними. В основі методу подальший розвиток та узагальнення ідей, запропонованих у свій час проф. О.О. Нікітіним для фільтрації геофізичних даних; методології комп'ютерного геопрогнозу В.В. Марченко, Б.А. Чумаченко, неметричного шкалювання

А. М. Бугайця, методології підготовки інформації для розпізнавання Н. Н. Боровка. Запропонована назва методу пояснюється таким.

У класичній трактовці статистична фільтрація геологічних даних розглядається як метод обробки просторово прив'язаних спостережень, метою якої служить виявлення тих чи інших складових спостережуваного поля. Процедури обробки, запозичені з теорії часових рядів та випадкових послідовностей, як правило зводяться до пригнічення завади шляхом згладження поля спостережень тим чи іншим способом. Цей традиційний підхід викладено у широко відомих монографіях Дж. Девіса, Т. Лаудона, У. Крамбейна та Ф. Грейбіла, Р. Мілера та Дж. Кана, у публікаціях В. І. Аронова, Г. Н. Боганіка, І. Д. Савинського та багатьох інших. Проте надії на збільшення корисного виходу з геологічної інформації давно вже пов'язуються з комплексуванням різних методів вивчення земної кори, що веде до необхідності залучення узагальнених просторово-ознакових моделей, які були б орієнтовані на ту ж плідну методологію оптимального виділення корисного сигналу на фоні завади. Цю ідею висловлювали О. О. Нікітін, О. В. Кузнецов, Е. Н. Черемісіна та ряд інших авторів. Метод, спрямований на вирішення цієї задачі, має оперувати узагальненими моделями сигналів у багатовимірному просторі різновидових ознак і виділяти їх на фоні завади у тому ж просторі. У цьому його відміна від класичної статистичної фільтрації випадкових рядів та полів. З іншого боку, термін статистична фільтрація вжитий у зв'язку з тим, що як і класична фільтрація, його головною концепцією є виділення корисних сигналів на фоні завади.

Метод, що претендує на нові досягнення у пошуках корисних копалин за рахунок більш ефективного використання геологічної інформації, має бути на рівні сучасних комп'ютерних технологій. Інструмент такого типу мусить сполучати ретельну підготовку задач із зручними, лаконічними засобами їх формулювання; можливість спільного використання різних за характером та способом реєстрації первинних геологічних даних з оптимальним алгоритмом їх перетворення для металогенічного прогнозу та розв'язання супутних задач. З цих мотивів було вирішено реалізувати метод у вигляді інформаційно-аналітичної системи з потужною базою різнопланових даних, яка б забезпечувала оперативне створення (на період виконання запиту) об'єктних та ознакових структурованих моделей з базових елементів, що зберігаються у БД.

Одною з найбільш трудомістких задач є розробка системи реєстрації різновидової інформації як необхідна умова дієвості інфор-

маційно-аналітичного інструменту. Потрібно було розробити такі нові форми реєстрації, які б надали інформації нової якості - системного характеру та прийнятності для комплексного аналізу математичними методами і засобами обчислювальної техніки. Одночасно ці форми мали підняти і якість інформації ще на етапі її збору, утримуючи певний рівень вимог за рахунок втілення в них досвіду найбільш кваліфікованих фахівців.

Мета роботи полягає у створенні нових автоматизованих аналітичних засобів для вирішення задач металогенічного прогнозу та геологічного картування на основі інтеграції систематизованої різновидової геологічної інформації, а також у підвищенні корисного виходу з геологічних матеріалів на цій основі.

Основні завдання досліджень.

1. Розробити теоретичні основи методу багатовимірної статистичної фільтрації різновидової геологічної інформації для виділення слабких альтернативних корисних сигналів у багатовимірному ознаковому просторі в умовах:

- будь-яких шкал реєстрації ознак;
- різного рівня достовірності ознак;
- залежності чи незалежності ознак.

2. Розробити систему програмних засобів з такими функціями та властивостями:

- створення у середовищі БД оперативних (на час виконання запиту) структурованих об'єктно-ознакових моделей;
- аналіз інформативності ознак у поточній задачі з єдиною мірою інформативності незалежно від шкали ознаки;
- формування багатокомпонентного фільтру та запис його відгуку у БД;
- супроводження результатів фільтрації характеристиками достовірності виділення сигналів;
- підготовка задачі за структурованим пошуковим образом;
- оптимальність виділення сигналів (у розумінні мінімуму ймовірностей помилок);
- автоматичне настроювання фільтру на оптимальний режим з можливістю ручної підстройки параметрів;
- оцінка ефективності фільтру шляхом тестування виділення сигналів;
- можливість оперативної інтерпретації результатів, графічне або квазіграфічне супроводження.

3. Розробити методику систематизованого опису (систематиза-

ції) різновидової геологічної інформації якісного та кількісного характеру згідно вимог методу.

4. Довести на практичних задачах з різними видами геологічних даних ефективність методу.

Методи досліджень:

- багатовимірний статистичний аналіз;
- метод статистичної фільтрації;
- метод максимальної правдоподібності оцінки параметрів;
- метод відношення правдоподібності перевірки гіпотез;
- математико-статистичне моделювання на ЕОМ;
- геохімічні та петрохімічні методи дослідження речовинного складу порід;
- статистичний аналіз локальних характеристик фізичних полів;
- промислово-геофізичні дослідження свердловин.

Наукова новизна.

1. Головним теоретичним результатом є вирішення проблеми інтегрованого аналізу трьох різновидів геологічної інформації при розв'язанні задач прогнозу та картування: описової з номінальними шкалами реєстрації; кількісної детермінованої; кількісної з проявами факторів невизначеності. Завдяки цьому створено умови для максимально повної реалізації її потенціалу. Зокрема, в обіг оперативного багатовимірного аналізу залучається така нетрадиційна для нього інформація, як петрографічні описи, структурно-вікові, формаційні, морфологічні ознаки тощо. В сучасних умовах, коли повністю вичерпано екстенсивні джерела інформаційного забезпечення геолого-розвідувальних робіт, запропонований шлях його інтенсифікації набуває пріоритетного значення. Новими теоретичними результатами роботи є:

- методологія спільного використання у багатовимірному аналізі міри достовірності детермінованих та недетермінованих величин, що містяться в базі даних. Вона забезпечує принципову можливість процедур інтеграційного (комплексного) аналізу різновидових ознак. Компоненти різного характеру - якісні або кількісні, статистичні оцінки або детерміновані значення мають у цих процедурах рівноправні позиції. При цьому всі вони доступні для кількісного аналізу, висновки якого теж супроводжуються мірою своєї достовірності;

- кореляційна теорія геологічних ознак з номінальними та порядковими шкалами реєстрації (т.з. якісних та кількісних показни-

ків);

- оптимізовані алгоритми багатовимірної фільтрації та класифікації, що можуть працювати з усіма шкалами реєстрації геологічних даних;

- методика автоматичного супроводження характеристиками достовірності даних, що вводяться у БД, на основі статистичних моделей, одержаних узагальненням галузевих аналітичних даних;

- моделювання розподілів структурних одиниць, зформованих запитом на основі композиції локальних розподілів;

- критерій згоди за методом максимуму правдоподібності;

- метод оцінки інформативності, побудований на основі ймовірностей помилок 1-го та 2-го роду;

- методика систематизації різновидової геологічної інформації згідно вимог методу багатовимірної фільтрації.

2) Головним практичним результатом є введення у експлуатацію дослідний зразок інформаційно-аналітичної системи "Фільтр", що складається з бази даних (БД), програмних засобів її ведення та перетворення; процедур багатовимірного аналізу (оцінка інформативності, аналіз груп, класифікація і т.д.) з використанням будь-яких ознак бази даних незалежно від шкали реєстрації. Система пройшла функціональні випробування при виконанні кількох НДР, а також шляхом розв'язання задач на реальних матеріалах геолого-геохімічних та промислово-геофізичних досліджень

Основні положення, що захищаються.

1) Метод багатовимірної статистичної фільтрації різновидової геологічної інформації на ЕОМ та створена на його основі інформаційно-аналітична система - новий ефективний засіб вирішення задач прогнозу корисних копалин та задач геологічного картування.

2) Розроблена кореляційна теорія різновидової інформації узагальнює метод кореляційного аналізу на ознакове середовище, що охоплює ознаки з номінальними та порядковими шкалами; її аналітичним апаратом розв'язуються задачі прогнозу та супутні питання.

Інші положення, на яких базується метод;

3) методологія спільного використання у багатовимірному аналізі міри достовірності детермінованих та недетермінованих величин, що містяться в базі даних;

4) методика автоматичного супроводження характеристиками достовірності даних, що вводяться у БД;

5) метод моделювання розподілів структурних одиниць, зформованих запитом;

6) метод оцінки інформативності, побудований на основі ймовірностей помилок 1-го та 2-го роду;

7) методика систематизації різновидової геологічної інформації.

Практична цінність і впровадження результатів роботи.

1) Завдяки впровадженню запропонованої методики реєстрації стають доступними для оперативного аналізу накопичені масиви інформації, що переважно знаходяться поза сферою тих сучасних комп'ютерних технологій, які орієнтуються на багатовимірний аналіз. Йдеться, зокрема про макро- та мікроописи порід, їхні властивості, формаційні, морфологічні ознаки, структурно-текстурні особливості тощо, які практично недоступні для оперативного аналізу на ЕОМ через неприйнятність форм реєстрації і відсутність відповідних програмних засобів. Завдяки цьому істотно збільшиться інформативна цінність і в багато разів - рівень використання геологічної інформації, що є головним критерієм ефективності роботи галузі, надбанням якої є саме інформаційні ресурси.

2) Метод доведений до реально працюючої інформаційно-аналітичної системи, повністю готової для впровадження. Функціональними випробуваннями доведено його високу ефективність при металогенічному прогнозі, у промисловій геофізиці, геологічному картуванні у різних масштабах робіт.

3) Метод з успіхом застосовано у різних регіонах при вирішенні задач прогнозу та картування.

- локальне прогнозування оловорудних родовищ за комплексом геолого-геохімічних та геофізичних ознак (Далекосхідне Примор'я);

- локальне прогнозування перспективних площ Халчерангінського рудного поля (Центральне Забайкалля);

- комплексна інтерпретація промислово-геофізичних даних (на прикладі теригенних відкладів Волгоградської області);

- класифікація порід Волинського плутону на петрографічні різновиди за геохімічними даними;

- прогноз нікеленосності на території Українського щита;

- кореляція інтрузивних магматичних та ультраметаморфічних утворень Українського щита.

Результати впроваджено у ряді НДР та НТП:

1) НДР на замовлення Держкомітету СРСР з науки та техніки "Розробити та передати для впровадження у Мінгео УРСР методику використання кількісних характеристик геолого-геофізичних та структурно-речовинних ознак для кількісного зіставлення та коре-

ляції геологічно однорідних утворень" (наук. кер. М. І. Толстой, автор звіту та відповідальний виконавець М. Н. Жуков, 1980-1982 р. р.);

2) НДР за республіканською міжвідомчою програмою РН.50.01 "Вивчити склад, фізичні властивості провідних петротипів плутонічних формацій південної частини Ущ для їх кількісного співставлення та кореляції при регіональному стратиграфічному аналізі та оцінці металоносності" (наук. кер. М. І. Толстой, 1982-1985 р. р.);

3) бюджетна НДР "Розробка методу багатовимірної статистичної фільтрації для аналізу геологічних даних" (науковий керівник - М. Н. Жуков, 1991-1992 р. р.);

4) бюджетна НДР "Розробка програмного забезпечення методу багатовимірної статистичної фільтрації геологічних даних на ПЕОМ" (науковий керівник - М. Н. Жуков, 1994-1996 р. р.).

Матеріали досліджень впроваджено у ряді організацій:

- Центральна тематична експедиція Держкомгеології України; (нині ДГП "Геопрогноз");
- ВГО "Читагеологія" Мінгео СРСР;
- ВГО "Північукргеологія" Держкомгеології України;
- Київський університет ім. Тараса Шевченка.

Апробація роботи.

Основні положення дисертації та результати досліджень доповідалися на міжнародних, всесоюзних, республіканських конференціях та нарадах, головні з яких такі:

- "Второе Сибирское совещание по применению математических методов и ЭВМ в геологии и геофизике". Новосибирськ, 1967.
- Міжнародний геологічний конгрес, XXIII сесія. "Доклады советских геологов. Математические методы в геологии". Москва, 1968.
- Всесоюзна наукова конференція вузів СРСР за участю науково-дослідних інститутів "Физика горных пород и процессов". Москва, 1-3 лютого 1971 р.
- Всесоюзна конференція "Системный подход в геологии (теоретические и прикладные аспекты)". Москва, 17-19 травня 1983 р.
- 4-а всесоюзна нарада "Теория и практика геохимических поисков в современных условиях". Ужгород, 10-12 жовтня 1988 р.
- Науково-технічна конференція "Проблемы создания универсальных баз данных в геологии". Київ, 12-14 червня 1990 р.
- Міжнародний симпозіум "Применение математических методов и компьютерных технологий при решении задач геохимии и охраны окружающей среды". Львів, 26-29 травня 1992 р.
- Міжнародна конференція "Глибинна будова літосфери та нет-

радиційне використання надр Землі". Київ, 14-18 травня 1996 р.

- 3-й міжнародний симпозіум "Застосування математичних методів і комп'ютерних технологій при розв'язанні задач геохімії і охорони навколишнього середовища". Львів, 1996.

Результати роботи обговорювалися на щорічних наукових конференціях професорсько-викладацького складу Київського університету ім. Тараса Шевченка (1982 - 1996 р.р.).

Подяки. Автор висловлює щирю вдячність колегам, які допомагали у проведенні досліджень на протязі більш ніж 20-річної роботи: А. Ю. Серзі, який був помічником автора багато років і якого вже немає з нами; проф. М. І. Толстому, який був ініціатором багатьох досліджень та відповідної тематики; проф. В. Г. Молявко, з яким автор співпрацює практично весь час роботи в Київському університеті; інженеру-програмісту В. І. Теплюку за його цінну допомогу у розробці інформаційно-аналітичної системи.

Дисертація складається з вступу, шести розділів та заключення. Виконана на кафедрі геохімії, мінералогії та петрографії Київського університету ім. Тараса Шевченка.

Основний зміст роботи

У першому розділі описано склад та оператори інформаційно-аналітичної системи "Фільтр", що є інструментальним засобом запропонованого методу. Передбачене автоматичне супроводження числових значень, які заносяться в БД, характеристиками достовірності. Середні квадратичні відхилення (СКВ) похибок розраховані на основі узагальнених залежностей "похибки - значення", отриманих за виборками контрольних спостережень для порід різного складу. Засоби інтерфейсу забезпечують можливість оперативного створення (на період виконання запиту) об'єктних і ознакових структурованих моделей з базових елементів, які містяться у БД. Засоби спілкування не вимагають від користувача спеціальної підготовки; роботі з системою студенти геологічного факультету навчаються за кілька годин. Передбачене графічне або квазіграфічне супроводження відгуків на запити, що робить сприйняття результатів проведеного аналізу наочним і оперативним. Передбачена можливість роботи з базою даних у супроводі багатоварової електронної карти, зокрема оперативне відображення результатів проведеного аналізу на картографічній основі.

Система має засоби для оперативного створення ознаково-об'

ектних моделей на час виконання запиту, їх оптимізації та збереження у зовнішньому середовищі. Широкий вибір варіантів виконання завдань, засобів автоматичного настроювання на оптимальний режим. При виконанні завдань з комплексуванням інформації - автоматичне зведення до єдиного масштабу ознак, які виражаються у різних шкалах (числових і номінальних). Можливість виконання завдань інформаційною системою (ІС) в умовах неповної визначеності ознак. Забезпеченість засобами настроювання на довільну структуру БД. Оперативне створення та ініціація БД із заданою структурою.

Область застосування - інформаційно-аналітичне супроводження прогностно-пошукових робіт на рудні родовища, промислово-геофізичних досліджень, геологічного картування.

Передбачено 8 шкал для реєстрації ознак. Шкали визначають спосіб реєстрації. Поділяються на дві групи: номінальні та порядкові (числові). В номінальній розгалуженій шкалі "А" стан ознаки фіксується кодом, який складається з кількох символів у відповідності з передбаченими систематичним укажчиком рівнями розгалуження. Стани кожного рівня утворюють повну групу. Значенням ознаки в простій номінальній шкалі ("В" та "У") є символ, що служить кодом одного з станів даної ознаки. Всі стани цієї ознаки також утворюють повну групу. Значення ознаки у бінарній шкалі "У" ("є-нема") фіксується одиницею - "є" та нулем - "нема". Значенням ознаки в шкалі сполук є код довжиною до 12 символів, який вказує номери станів часткових ознак, спеціально призначених для реєстрації в цій шкалі. Вона використовується, наприклад, для опису різновидів мінералів, фракцій, включень, локальних структур. У шкалі "число" ("W") фіксується одне число, яке служить значенням ознаки, причому вважається, що її визначають без похибки. У статистичній шкалі з укажчиком закону розподілу (шкала "F") фіксується 3 значення: m - оцінка середнього значення; b - оцінка середнього квадратичного відхилу (СКВ); u - укажчик закону розподілу. Якщо закон розподілу нормальний або обсяг виборки був недостатнім для оцінки закону розподілу, то $u = -1$ або $u = 0$. При логнормальному законі $u = -2$. У статистичній шкалі без укажчика закону розподілу (шкала "D") фіксується 2 значення: m та b . Статистичні шкали супроводжують значеннями b результати реєстрації тих ознак, які піддані впливу варіації через аналітичні похибки та локальні неоднорідності порід. Для значень, які не мають статистично забезпеченої оцінки СКВ, ІС дає змогу автоматичного супроводження - на основі моделей, розрахованих за репрезентативними виборками статистичних даних.

Кожний запит складається з не менш як одного оператора. Всього передбачено 7 операторів, які визначають інформаційне поле у БД для виконання запиту та створюють структуру цього поля на період виконання запиту; процедуру та її варіант; взаємодію між послідовними однотипними операторами; адресу для виведення результатів; засоби програмування потоку запитів. Загальний вигляд оператора:

<ім'я оператора>*l=p\$<список_1>&<список_2>&...&<список_n>",

де *l - ліва специфікація оператора; l - символ з фіксованого для даного оператора переліку; параметр l необов'язковий; p\$ - права специфікація із фіксованого для даного оператора переліку; параметр p\$ необов'язковий; специфікації l та p діють незалежно; * - роз'єднувач імені оператора і специфікації; \$ - роз'єднувач специфікації та списків оператора; = - роз'єднувач лівої і правої частин оператора; в одному операторі допускається тільки один такий символ; <список> - перелік елементів, роз'єднаних комами; & - роз'єднувач списків. Елементи, що входять в <список_1>, <список_2>, ..., <список_n> утворюють ряди (ансамблі). Роз'єднувачі & допустимі лише в двох типах операторів - О і П, утворюючи в поточному запиті, відповідно, ряди (ансамблі) об'єктів і ряди ознак. Використання рядів залежить від завдання.

Мова користувача передбачає 7 операторів.

"О" ("ОБ'ЄКТИ") - визначає об'єкти БД, що залучаються до процедури обробки, вказаної оператором ЗАВДАННЯ, а також структуру, у якій слід використати перелік цих об'єктів. Крім того, оператор ОБ'ЄКТИ може вказувати спосіб зв'язку з оператором ОБ'ЄКТИ попереднього запиту.

"П", "Р" ("ОЗНАКИ") - визначає ознаки БД, що залучаються до процедури обробки, заданої оператором ЗАВДАННЯ, а також структуру переліку цих ознак. Оператор ОЗНАКИ може вказувати спосіб зв'язку з оператором ОЗНАКИ попереднього запиту.

"З", "З" ("ЗАВДАННЯ") - вказує завдання (процедуру обробки), його специфікацію та варіант. Завдання виконується з об'єктами та ознаками, вказаними в операторах О і П, а в разі наявності зв'язку з попереднім запитом - з об'єктами або ознаками, які залучаються до запиту цим зв'язком.

"У", "У" ("УСТРОЇ") - вказує устрої вводу-виводу.

"К" ("КІНЕЦЬ ЗАПИТУ") - означає кінець тексту запиту та служить вказівкою розпочати виконання запиту.

"C", "S", "E" ("СТОП") - служить ознакою завершення потоку запитів і вказівкою до завершення роботи з ІС.

"I", "I", "H" ("ІНСТРУКЦІЇ") - виводить скорочені інструкції та довідки користувачеві про роботу з ІС.

Оператор O має вигляд (обов'язкові елементи підкреслені):

ОБ'ЄКТИ*l=p\$<список 1>&<список 2>&...&<список n>".

Допустимі значення параметра *l: *N, *S, *D, *M, *R, *W, пусто. В разі специфікації *N або відсутності параметра *l встановлюється цифрова специфікація. Це означає, що елементами <списків> можуть бути номери об'єктів або переліки номерів вигляду n1-n2. Значення l=M встановлює специфікацію "Моделі", при якій елементами списків служать імена, за якими закріплені певні переліки об'єктів. При l=S встановлюється символна специфікація. Тоді елементами списків служать укажчики станів ознак (UCO), за якими в подальшому виконується пошук об'єктів у всій БД. При l=D встановлюється також символна специфікація, з тією різницею, що зону пошуку задає попередній оператор O. Специфікації *R, *W діють як *S.

Оператор O, будучи по відношенню до потоку запитів глобальним, розповсюджує свою дію на всі наступні запити до того запиту, в якому буде введено новий оператор O.

Параметр p\$ - поєднувач, який вказує спосіб з'єднання списків поточного оператора O із списками попереднього оператора O. Відсутність поєднувача означає, що попередній оператор O ігнорується, замінюючись поточним оператором O. Символ \$ - роз'єднувач між p і списками поточного оператора O. Допустимі значення p: "&", ".", "+", "-".

Оператор П (ОЗНАКИ) має вигляд (обов'язкові елементи підкреслені):

ПОКАЗНИКИ*l=p\$<список 1>&<список 2>&...&<список n>".

Параметри l та p (специфікації) приймають ті ж значення, що у оператора O.

Довільний оператор O або П може бути зафіксований на диску з тим, щоб при необхідності його можна було відтворити. Передбачена можливість запису операторів O та П у файл з заданим ім'ям, а також читання з такого файлу.

Оператор "УСТРОЇ" вживається для призначення устроїв вводу запитів та виводу результатів у потоці запитів. Має вигляд :

УСТРОЇ = <ім'я_1>, <ім'я_2>"

(підкреслені обов'язкові елементи; замість символу У допускається U). <Ім'я_1> означає устрій вводу, а <ім'я_2> - устрій виводу.

Перший запит описують не менш ніж чотирма операторами: "0" ("Об'єкти"), "П", "Р" ("Ознаки"), "З", "Z" ("Завдання"), "К" ("Кінець запиту"). Решта запитів можуть складатися з одного або більше операторів. Для окремих процедур один з операторів "0" або "П" необов'язковий. Кожен оператор розповсюджує свою дію на всі наступні запити, аж поки черговий однотипний оператор не відмінить або змінить його дію.

Оператор "ЗАВДАННЯ" призначає процедуру обробки та її варіант. Має вигляд :

ЗАВДАННЯ*l=p\$<ім'я завдання>" або Z*l=p\$<ім'я завдання>"

(обов'язкові елементи підкреслені). Значення *l залежать від конкретного завдання. Специфікація p\$ зазначає варіант завдання. За умовчанням, тобто коли параметра p\$ в тексті оператора немає, p=1 (основний варіант). Варіанти завдань дозволяють варіювати зміст процедури, вводити додаткову інформацію, виводити результати в скороченому вигляді або навпаки, з додатковими подробицями.

У **другому розділі** подано зміст та теоретичне обґрунтування процедур аналізу, що входять у склад системи. Розроблені процедури стосуються, в першу чергу, інтегрованого аналізу ознак, що реєструються в якісних та кількісних шкалах. Запропоновано концептуально новий алгоритм виявлення корисного сигналу на ознаковому полі складної структури, яке породжено кількома сотнями ознак з 8 шкалами реєстрації. Його дієвість ґрунтується на введеній мірі, у якій вклад кожної ознаки залежить від співвідношення між мірою розбіжності значень та критичною границею, яку визначає відповідний тест порівняння.

Функціональність інформаційної системи була б неповною без можливості вживати у всіх процедурах разом з базовими ознаками результатів тих чи інших перетворень інформації, що міститься у БД. Це стосується, зокрема, відгуків фільтрів, розрахованих за певними наборами ознак. Адже, лише завдяки прийнятій концепції побудо-

ви БД - супроводу збереженої інформації характеристиками достовірності - вдалося вирішити згадану проблему комплексування. Тому виникла потреба в супроводі результатів перетворень розрахунковими характеристиками достовірності. Ця задача, в тому числі і для нелінійних перетворень, вичерпно вирішується розробленою процедурою, яка реалізуючи самі перетворення, автоматично супроводжує результати характеристиками достовірності і фіксує їх у БД.

Таблиця 1

Імена завдань

N п/п	Імена завдань			Процедура обробки
	повне ім'я (кирилиця)	Скорочене ім'я кири- лиця	лат. шрифт	
1	Обз	Обз	Obz	Огляд об'єктів БД
2	Об2	Об2	Ob2	Огляд 2 (ознак БД)
3	Спр	Спр	Spr	Довідка по об'єктах БД
4	Спп	Спп	Spp	Довідка по ознаках
5	Кзб	Кзб	Kzb	Карта заповнення БД
6	Кбд	Кбд	Kbd	Корекція БД
7	Пре	Пре	Pre	Перетворення ознак
8	Бдф	Бдф	Bdf	Виведення "БД - файл"
9	Фбд	Фбд	Fbd	Запис "файл - БД"
10	Упорядкування	Упо	Upo	Упорядкування об'єктів
11	Сортування	Сор	Sor	Сортування об'єктів
12	Статистика	Ста	Sta	Статистика; формування оптимального нормованого фільтру
13	Інп	Інп	Inp	Інформативність параметрична
14	Хап	Хап	Har	Характерні ознаки
15	Класифікація	Кла	Kla	Класифікація параметрична (при гіпотезі незалежності ознак)
16	Клн	Клн	Kln	Класифікація непараметрична (при гіпотезі залежності ознак)
17	Тек	Тек	Tek	Тест класифікації параметричної
18	Ткн	Ткн	Tkn	Тест класифікації непараметричної
19	Аналоги	Ана	Ana	Аналоги
20	Анс	Анс	Ans	Аналоги за методом найближчих зразків сигналу
21	Групування	Гру	Gru	Групування об'єктів
22	Зображення	Ізо	Izo	Зображення в координатах XOY в текстовому режимі екрана
23	Дбл	Дбл	Dbl	Дублікати
24	Діа	Діа	Dia	Діаграма у трикутній системі координат
25	Дск	Дск	Dsk	Відображення у декартовій системі координат
26	Мар	Мар	Mar	Відображення на геологічних картах, поєднаних з БД

Крім процедур аналізу петрологічної інформації предметом детальної розробки були численні службові функції, без яких робота з системою була б надто трудомісткою, а для професійно не підготовленого користувача - мабуть і неприйнятною з цієї причини. Йдеться про ведення та підримку БД, одержання довідкової інформації про її стан, пересування даних в БД та обмін між файлами різних БД.

Загальний перелік імен завдань наведений в табл. 1.

Завдання "Перетворення ознак (Pre)" виконує одно- та двомісні перетворення над ознаками, що виражаються в довільних шкалах. Одномісна операція виконується над сумою (в разі шкали Y - над логічною сумою, а при модифікації M - над добутком ознак, перелічених в першому ряду оператора П. Результат розміщується в ознаках, перелічених в другому ряду. Двомісна операція виконується над сумами (в разі шкали Y - над логічними сумами ознак, а при модифікації M - над добутками ознак), перелічених в першому та другому рядах оператора П. Результат розміщується в ознаках 3-го ряду.

Оператор O зазначає об'єкти БД, які слід піддати перетворенню. Символ операції ототожнюється з варіантом завдання і тому записується в правій специфікації, тобто між знаками "=" і "\$" оператора Z (3), де і зазначають варіант завдання.

Варіанти :

- 1 (перенесення); * (множення); E (Exp(x)); V (об'єднання);
- + (додавання); / (ділення); L (Ln(x)); O (заперечення);
- (віднімання); D (1/x); ^ (перетин); K (кодування);
- % (ділення та множення на 100).

При модифікації завдання "S" ($Z \cdot S = \text{PRE}$, $3 \cdot S = n \cdot \text{PRE}$) перетворюються суми ознак, взятих з коефіцієнтами перерахунку із файлу, що задається (для шкал F, D, W). При модифікації "M" перетворюються не суми, а добутки ознак. Для виконання перерахунків із сталими коефіцієнтами (наприклад, вмісти породоутворюючих оксидів в атомні кількості або в головні компоненти) використовуються файли цих коефіцієнтів. Переліченими варіантами вичерпуються структури фільтрів, які можуть мати реальне практичне значення. Зокрема, адитивний фільтр вигляду

$$Z = c_1 x_1 + c_2 x_2 + \dots + c_n x_n + b$$

реалізується одним завданням вигляду $Z \cdot S = \text{PRE}$ відразу по всій БД з використанням файлу коефіцієнтів Isprekoe.tmp. Мультиплікативний фільтрації відповідає фільтр вигляду

$$z = c_1 \ln x_1 + c_2 \ln x_2 + \dots + c_n \ln x_n + b$$

або

$$z = \exp(c_1 \ln x_1 + c_2 \ln x_2 + \dots + c_n \ln x_n + b).$$

Перетворення типу петрохімічних коефіцієнтів, геохімічних відношень, які є нічим іншим, як різновидами мультиплікативних фільтрів, утвореними на просторі або ознак БД або лінійних фільтрів підмножин ознакового простору БД. Новим елементом є супроводження результатів перетворень середніми квадратичними відхилами (СКВ), які автоматично обчислюються при виконанні перетворень. Без цього аналітичні засоби системи втратили б свою працездатність. У табл. 2 наведені формули, за якими обчислюються СКВ результатів перетворень над рядами числових ознак. Вжиті такі позначення: x, y - ознаки, що підлягають перетворенню; z - ознака, що є результатом перетворення; X, Y, Z - математичні сподівання ознак x, y, z ; C - константа; r - коефіцієнт кореляції між величинами x та y ; коли x та y

Таблиця 2

Формули для обчислення середніх квадратичних відхилів ознак при виконанні процедури ПРЕ

Перетворення	середній квадратичний відхил результату
$z = x + C$	$\sigma(z) = \sigma(x)$
$z = C \cdot x$	$\sigma(z) = C \cdot \sigma(x)$
$z = x/C$	$\sigma(z) = 1/C \cdot \sigma(x)$
$z = 1/x$	$\sigma(z) = \sigma(x)/X^2$
$z = \text{Exp}(x)$	$\sigma(z) = Z \cdot \sigma(x)$
$z = \text{Ln}(x)$	$\sigma(z) = \sigma(x)/X = V(x)$
$z = x + y$	$\sigma(z) = (\sigma^2(x) + 2r \cdot \sigma(x)\sigma(y) + \sigma^2(y))^{0.5}$
$z = x - y$	$\sigma(z) = (\sigma^2(x) - 2r \cdot \sigma(x)\sigma(y) + \sigma^2(y))^{0.5}$
$z = x \cdot y$	$\sigma(z) = (Y^2 \sigma^2(x) + 2rXY \cdot \sigma(x)\sigma(y) + X^2 \sigma^2(y))^{0.5}$
$z = x/y$	$\sigma(z) = (Y^2 \sigma^2(x) - 2rXY \cdot \sigma(x)\sigma(y) + X^2 \sigma^2(y))^{0.5}$
$z = \sum_{i=1}^n c_i x_i + b$	$\sigma(z) = (\sum_{i=1}^n (c_i \sigma(x_i))^2 + 2 \sum_{i=1}^n \sum_{j=1}^n c_i c_j r_{ij} \sigma(x_i) \sigma(x_j))^{0.5}$
$z = \prod_{i=1}^n x_i$	$\sigma(z) = Z \cdot (\sum_{i=1}^n V_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n r_{ij} V_i V_j)^{0.5}$

незалежні, то $r = 0$; якщо одна з величин x чи y є перетворенням іншої або обидві є результатами перетворень тієї самої величини, то $r = 1$ або $r = -1$ залежно від того, прямий чи обернений зв'язок між x та y ; r_{ij} - коефіцієнт кореляції між величинами x_i та x_j ; $\bar{b}(x)$, $\bar{b}(y)$, $\bar{b}(z)$ - середні квадратичні відхили величин x, y, z ; $V(x), V_j$ - коефіцієнти варіації величин x, x_j .

Завдання "Статистика (Sta)" призначене для порівнювальної статистичної обробки рядів об'єктів, перелічених у операторі "O", за ознаками, указаними оператором "P", а також для визначення оптимальних лінійних фільтрів. Передбачена можливість автоматичного введення в середовище БД даних про новоутворені об'єкти, що є статистичними моделями заданих рядів об'єктів. Останні можна оперативно формувати засобами ІС: структурованим пошуком у модифікаціях *S, *D оператора "O"; залученням фіксованих моделей оператора "O" через його специфікацію *M; завданням "Сортування" тощо.

Щільність розподілу статистичної ознаки G оцінюється для i -го ряду об'єктів у вигляді композиції:

$$p_i(x) = \sum_{k=1}^n p_{ik}(x)/n$$

де $p_{ik}(x)$ - щільність розподілу ознаки G для k -го об'єкта з i -го ряду Q_i . Якщо укажчик розподілу $u_{ik} = -1$, то

$$p_{ik}(x) = 1/(\sqrt{2\pi}\bar{b}_{ik}) \cdot \exp(-(x - m_{ik})^2/(2\bar{b}_{ik}^2));$$

m_{ik} та $\bar{b}_{ik}$ - математичне сподівання та СКВ ознаки G для k -го об'єкту з i -го ряду, значення яких разом з u_{ik} зберігаються у БД. Якщо $u_{ik} = -2$ і $x < 0$, $p_{ik}(x) = 0$. При $x > 0$ і $u = -2$

$$p_{ik}(x) = 1/(\sqrt{2\pi}\delta_{ik}x) \cdot \exp(-(\ln x - \mu_{ik})^2/(2\delta_{ik}^2));$$

μ_{ik} та δ_{ik} - математичне сподівання та СКВ величини $\ln G$ для вказаного об'єкта, значення яких обчислюються за формулами:

$$\delta_{ik} = \sqrt{\ln(\bar{b}_{ik}^2/m_{ik}^2 + 1)}, \quad \mu_{ik} = \ln m_{ik} - 0.5\delta_{ik}^2.$$

Якщо СКВ ознаки G дорівнює нулю, то щільність розподілу подається у вигляді δ -функції $p_{ik}(x) = \delta(x - m_{ik})$, для якої

$$\int_R \delta(x - m_{ik}) dx = 1, \quad \int_R x \delta(x - m_{ik}) dx = m_{ik}.$$

Відображення оцінок щільностей розподілу подається у зіставленні рядів об'єктів на спільному інтервалі або об'єднанні інтервалів.

У варіанті процедури з записом параметрів розподілів перевіряється гіпотеза про нормальний розподіл при альтернативі логнормального розподілу для кожного з рядів. Для цього вживається розроблений критерій згоди за методом відношення правдоподібності. За цим критерієм приймається гіпотеза про логнормальний розподіл, якщо виконується умова $K_L > 1$, де $K_L = 6/(\delta \cdot \exp(\mu))$, δ - оцінка СКВ за n спостереженнями $x_1, x_2, \dots, x_n$ вигляду:

$$\delta = \left(\sum_{i=1}^n (x_i - m)^2 / n \right)^{0.5}, \quad m = \sum_{i=1}^n x_i / n;$$

μ та δ обчислюються у вигляді:

$$\delta = \left(\sum_{i=1}^n (\ln x_i - \mu)^2 / n \right)^{0.5}, \quad \mu = \sum_{i=1}^n \ln x_i / n.$$

Ймовірність того, що розподіл логнормальний при альтернативі нормального розподілу -

$$P_L = 1 / (1 + K_L^n).$$

Для оцінки СКВ номінальної ознаки запропоновано формулу:

$$\sigma(G) = \sqrt{D(G)}, \quad D(G) = \sum_{i=1}^n \sum_{k=i+1}^n \rho(x_i, x_k) / (n(n-1)),$$

де $\rho(x_i, x_k) = 1$, якщо стани ознаки x_i та x_k різні; $\rho(x_i, x_k) = 0$, якщо ці стани однакові; n - кількість об'єктів ряду. Згідно з доведеною властивістю, $D(G)$ буде незміщеною оцінкою дисперсії статистичної ознаки, якщо покласти $\rho(x_i, x_k) = (x_i - x_k)^2$. Це означає, що наведена формула є загальною для ознаки будь-якого типу.

Фільтр настроюється на корисний сигнал, представлений об'єктами 2-го ряду, відносно фону, представленого об'єктами 1-го ряду. Якщо в операторі "0" тільки один ряд об'єктів, то фільтр буде настроєно тільки на модель, представлену об'єктами цього ряду. Коефіцієнти фільтру нормуються з тим розрахунком, щоб в межах фону модуль відгуку не перевищував 1 при коефіцієнті довіри, що відповідає рівню значущості 0.001. Передбачається, що згодом фільтр

вживається для фільтрації об'єктів БД завданням ПРЕ у специфікації *S. Для статистичних ознак коефіцієнти фільтру f , оптимального за відношенням "сигнал/завада", обчислюється за відомою формулою:

$$c_j = (m_{sj} - m_j) / \sigma_j^2 = \Delta m_j / \sigma_j^2, \quad b = - \sum_{j=1}^n c_j m_j,$$

де n - кількість тих ознак, які беруть участь у фільтрації; x_j - значення j -ї ознаки у об'єкта, що підлягає фільтрації; m_j , m_{sj} - математичні сподівання j -ї ознаки при відсутності та наявності сигналу відповідно; $\Delta m_j = m_{sj} - m_j$; σ_j - середній квадратичний відхил j -ї ознаки при відсутності сигналу; f_j - складова відгуку, що вноситься j -ю ознакою: $f_j = c_j x_j$. За рахунок b фільтр приведено до нульового математичного сподівання при відсутності сигналу: $Mf=0$.

Нормуванням коефіцієнтів фільтру - $C_j = c_j / \sigma(f)$, $B = b / \sigma(f)$ досягають одиничної дисперсії в межах фону:

$$F = \sum_{j=1}^n C_j x_j + B; \quad D(F) = 1; \quad MF = 0.$$

Для більшої наочності відгук фільтру F віднесено до квантиля u_q (0,1)-нормального розподілу, який відповідає фіксованому рівню значущості $\alpha = 1-q$ статистичного критерію перевірки гіпотези про відсутність сигналу: $F_q = F / u_q$. Отже, для фонових об'єктів модуль відгуку фільтру як правило не перевищуватиме 1, на відміну від проявів корисного сигналу.

Для ознаки G_i з тих ознак $G_1, G_2, \dots, G_L$, які фіксуються у номінальних шкалах і використовуються у фільтрі, відгук фільтру, оптимального за відношенням "сигнал/завада", після проведених перетворень має вигляд:

$$f_i = P_{js} / P_j - (1 - P_{js}) / (1 - P_j) = (P_{js} - P_j) / (P_j (1 - P_j)),$$

якщо ознака G_i в j -му стані.

Математичне сподівання та дисперсія відгуку при відсутності сигналу -

$$Mf_1 = \sum_{k=1}^n \frac{P_{ks} - P_k}{1 - P_k}; \quad Df_1 = \sum_{k=1}^n \frac{(P_{ks} - P_k)^2}{P_k (1 - P_k)^2} - (Mf_1)^2$$

Відгук фільтру за всіма ознаками, які беруть участь у фільтрації

$$f = \sum_{k=1}^L f_k.$$

де f_k обчислюється за наведеними формулами для ознак, що реєструються у статистичних та номінальних шкалах відповідно. Приведення фільтру до нульового фону та одиничної дисперсії виконується за формулою:

$$F = (f - \sum_{k=1}^L Mf_k) / (\sum_{k=1}^L Df_k)^{0.5}.$$

Критерій для виділення сигналу при рівні значущості $\alpha = 1 - q$:

$$F_q = F/u_q \geq 1.$$

Завдання "Інформативність параметрична" призначене для оцінки значень показника інформативності - ймовірності належності об'єктів до класів, зформованих у середовищі БД. Інформативність статистичної ознаки G відносно класу Q_j об'єктів при альтернативі класів $Q_1, Q_2, \dots, Q_{j-1}, Q_{j+1}, \dots, Q_T$ має зміст імовірності вірного рішення при класифікації на 2 апіорно рівноймовірні класи: Q_j - не Q_j : $I_j(G) = 0.5(P_{jj} + P_{\Gamma(j), \Gamma(j)})$, де P_{jj} імовірність правильної класифікації об'єкта з класу Q_j за ознакою G , з тією умовою, що вживається оптимальний критерій класифікації Бейеса; $P_{\Gamma(j), \Gamma(j)}$ - імовірність невіднесення об'єкта до класу Q_j , коли він в дійсності до нього не належить. У припущенні, що класи, які складають цю альтернативу, апіорі однаково ймовірні,

$$I_j(G) = 0.5 \left(\frac{TP_{jj}}{T-1} + 1 - \frac{1}{T-1} \sum_{k=1}^T P_{kj} \right).$$

У наведеній формулі: T - кількість класів; P_{kj} - імовірність віднесення до класу Q_j об'єкта, який апіорі належить до класу Q_k .

Завдання "Класифікація параметрична (Kla)" призначене для

оцінки ймовірностей належності об'єктів до t класів, зформованих у середовищі БД і моделюючих альтернативні сигнали.

Результат класифікації визначається порівнянням значень двокомпонентної міри належності пред'явленого об'єкта Y до кожного з класів. Першою компонентою є кількість $L(j)$ тих ознак, за якими об'єкт класифікації (Y) знаходиться у стані, який у відповідному класі (Q_j) має нульову ймовірність, якщо ознака не є статистичною (тобто номінальною, числовою) або щільність розподілу дорівнює нулю (якщо ознака статистична). Отже, $L(j)$ є мірою чужорідності об'єкта по відношенню до класу Q_j . Другою компонентою служить бейсівська ймовірність $P(Y \in Q_j)$ належності об'єкта Y до класу Q_j . Ця ймовірність обчислюється за формулою:

$$P(Y \in Q_j) = P_j / \sum_{k=1}^t P_k,$$

де $P_j = \prod_{k \in E(j)} P_{jk}(y_k)$, якщо $L(j) = L_{\min}$, $P_j = 0$, якщо $L(j) > L_{\min}$

$E(j)$ - множина номерів тих ознак серед $G_1, G_2, \dots, G_r$, для яких $P_{jk}(y_k)$ не дорівнює нулю; $L_{\min} = \min(L(1), L(2), \dots, L(r))$. Критерій класифікації складається з таких кроків. Спочатку визначають $L_{\min}$ - найменшу з величин $L(1), L(2), \dots, L(t)$ та встановлюють ті номери $j_1, j_2, \dots, j_m$, при яких $L(j) = L_{\min}$. Для класів, номери яких не входять до цього списку, покладають $P_j = 0$, а решту значень P_j обчислюють за наведеною формулою. Клас, до якого слід віднести об'єкт класифікації, визначають за найбільшою з бейсівських ймовірностей. Значення цієї ймовірності служить характеристикою вірогідності прийнятого рішення. Результати супроводжуються позначками: "внутрішня належність", якщо $L_{\min} = 0$, "зовнішня належність", якщо $L_{\min} > 0$.

Пріоритет, відданий в задачі класифікації бейсівському підходу, має очевидну підставу. Вона полягає у оптимальності бейсівського критерію класифікації у класичному розумінні - мінімуму середньої ймовірності помилок класифікації. Проте, на відміну від класичної трактовки формули Бейеса, коли під величинами $P_{jk}(x_{vk})$ розуміють значення щільності розподілу k -ї ознаки у j -му класі для точки x_{vk} , наведений вираз нами істотно узагальнений. Запропонований вираз надає можливість вживати в одному критерії класифікації статистичні і номінальні ознаки. Крім того, критерій не

втрачає дієвості, коли в окремих класах або у об'єкті Y немає інформації по тих чи інших ознаках; коли у класах є інформація по різних наборах ознак і навіть при різній кількості ознак. В умовах неповноти БД, що здебільшого є не виключна, а пересічна ситуація і відбиває реальний стан нерівномірної вивченості геологічних об'єктів, ця перевага набуває неабиякого значення.

Якщо ознака G_k статистична, то її станом x_{yk} у об'єкта Y буде локальний розподіл, описаний у БД трійкою значень (m_{yk}, b_{yk}, u_{yk}) : m_{yk} - математичне сподівання ознаки G_k об'єкта Y ; b_{yk} - її СКВ; u_{yk} - укажчик закону розподілу. Згідно з цим

$$P_{jk}(x_k) = l_k \cdot \int_R p_{yk}(x) p_{jk}(x) dx,$$

де $p_{jk}(x)$ - щільність розподілу ознаки G_k у класі Q_j ; $p_{yk}(x)$ - щільність розподілу тієї ж ознаки у об'єкті Y ; l_k - довжина області Ω_k на числовій осі, утвореної об'єднанням інтервалів $[a(j,k), b(j,k)]$ ($j=1, 2, \dots, N$):

$$\Omega_k = \bigcup_{j=1}^N [a(j,k), b(j,k)],$$

$a(y,k)$, $b(y,k)$ - $q\%$ -ні границі для ознаки (q регулюється). У програмному модулі, який реалізує параметричну класифікацію, врахована можливість участі у виразі щільності розподілу компонент типу δ -функцій ($b=0$), а також типу невизначеності ($b=\infty$). Чисельне інтегрування здійснюється за запропонованою формулою рівноймовірних інтервалів.

Якщо ознака G_k не є статистичною, то $P_{jk}(x_{yk}) = l_k \cdot p_{yk}(y_k)$, де $p_{yk}(y_k)$ - частота стану x_{yk} ознаки G_k у класі Q_j , l_k - загальна кількість різних станів ознаки, які зустрічаються у всіх класах.

Завдання "Тест класифікації параметричної" призначене для оцінки ймовірностей належності об'єктів з рядів оператора "0" до класів, представлених тими ж рядами. Тестування надає змогу оцінювати діагностичну здатність обраного набору ознак в задачі класифікації, яку згодом мають вирішувати завданням КЛА.

Завдання "Класифікація непараметрична" призначене для оцінки ймовірностей належності об'єктів до класів при гіпотезі залежності ознак, проте діє ефективно і в разі їхньої незалежності. Запропонований алгоритм має перевагу з трьох причин. Першою є оптимальність критерію Бейеса, який тут реалізовано. Друга - уживання толерантної моделі багатовимірного розподілу, яку обходять тра-

диційні труднощі параметричної класифікації за статистично залежними компонентами, зокрема, неприємна необхідність обертання кореляційних матриць і пов'язана з цим потреба у великій кількості спостережень, що робить цей шлях в буденності неприйнятним. Третя - нормальна працездатність в умовах браку інформації і навіть тоді, коли для класичного бейєсівського критерію виникає ситуація невизначеності через потрапляння значень окремих компонент у множину нульової ймовірності для всіх класів.

Результат класифікації визначається порівнянням значень двокомпонентної міри належності пред'явленого об'єкта Y до кожного з класів. Першою компонентою є кількість $L(j)$ тих ознак, за якими об'єкт класифікації (Y) знаходиться у стані, що у відповідному класі (Q_j) має нульову ймовірність. $L(j)$ служить мірою чужорідності об'єкта по відношенню до класу Q_j . Другою компонентою служить бейєсівська ймовірність $P(Y \in Q_j)$ належності об'єкта Y до класу Q_j . Ця ймовірність обчислюється за формулою:

$$P(Y \in Q_j) = P_j(y) / \sum_{i=1}^t P_i(y), \text{ де } P_i(y) = \sum_{j=1}^{b(i)} P_j^{(i)}(y) / b(i) -$$

умовна ймовірність r -вимірного стану y у класі Q_i ;

$b(i)$ - кількість об'єктів-еталонів у класі Q_i ;

$$P_j^{(i)}(y) = \begin{cases} \prod_{k \in E(i,j)} P_{jk}^{(i)}(y_k), & \text{якщо } L^{(i)}(j) = L_{\min}^{(i)}, \\ 0, & \text{якщо } L^{(i)}(j) > L_{\min}^{(i)}. \end{cases}$$

$L^{(i)}(j)$ - кількість тих ознак, для яких $P_{jk}^{(i)}(y_k)$ дорівнює нулю;

$$L_{\min}^{(i)} = \min(L^{(i)}(1), L^{(i)}(j), \dots, L^{(i)}(b(i)));$$

$E(i,j)$ - множина номерів тих ознак серед $G_1, G_2, \dots, G_r$, для яких

$P_{jk}^{(i)}(y_k)$ не дорівнює нулю; $L^{(i)}(j) = L_{\min}^{(i)}$. $P_{jk}^{(i)}(y_k)$ - ймовірність стану y_k ознаки G_k у j -го еталонного об'єкта з класу Q_i . Якщо ознака нестатистична і її стан у згаданого об'єкта y_k , то ймовірність стану дорівнює 0. Якщо стан її інший, то ймовірність стану дорівнює 1. Вираз ймовірності стану для статистичної ознаки має вигляд:

$$P_{jk}^{(i)}(y_j) = \int_R p_{vk}^{(i)}(x) \chi_{y_j}(x) dx,$$

(1)
де $p_{jk}(x)$ - щільність розподілу ознаки G_k для j -го еталонного об'єкта з класу Q_i ; $p_{vk}(x)$ - щільність розподілу G_k в стані y .

Завдання "Тест класифікації непараметричної" призначене для оцінки ймовірностей належності об'єктів до класів - еталонів альтернативних сигналів. Виконується шляхом тестової класифікації об'єктів за алгоритмом процедури КЛА.

Завдання "Аналоги" призначене для упорядкування об'єктів з 2-го ряду по останній за мірою відміни від моделі багатовимірного корисного сигналу, що після відповідного нормування забезпечує відшукування аналогів корисних сигналів у БД за сукупністю ознак. В разі такої задачі перший ряд об'єктів складають із зразків (еталонів) корисного сигналу. Наявності корисного сигналу у особі об'єкту Y відповідає відносно невелике значення відстані $\rho(Y)$, що визначає цей об'єкт як внутрішній аналог ряду еталонів. Фільтр в цьому разі буде настроєно на певний корисний сигнал. Можлива й інша постановка задачі, коли перший ряд складають, навпаки, із зразків фону, тобто фільтр настроєно на фон. Тоді внутрішнім аналогам відповідають області фону у БД, а підвищеним значенням міри $\rho(Y)$ - області відхилення від фону ("багатовимірні аномалії"). Перед виконанням процедури проводиться оптимізація моделі сигналу шляхом вилучення забруднення сигналу після завдання "Групування".

Критерій відміни встановлює максимальне значення (1) міри відміни між двома значеннями ознаки, вираженої в статистичній шкалі (F, D) в разі негативного результату перевірки гіпотези про збіжність математичних сподівань. Він еквівалентний встановленню розбіжності значень ознаки, що виражена в будь-якій іншій шкалі.

Проблему комплексування ознак з статистичними та детермінованими - числовою та номінальними шкалами вирішено шляхом введення критерію відміни станів статистичної ознаки. Різниця між оцінками m_x та m_y середніх значень ознаки G_k вважається статистично значущою при рівні значущості q , якщо $(m_x - m_y)^2 > t_q \cdot \sigma^2(G_k)$, де $\sigma(G_k)$ - оцінка середнього квадратичного відхилення ознаки G_k вигляду:

$$\sigma(G_k) = \sum_{i=1}^b \sigma_i(G_k) / b;$$

t_q - коефіцієнт розмежування, що відповідає рівню значущості q : $t_q = u_{\alpha}^2$; u_{α} - квантиль $(0,1)$ -нормального розподілу порядку $\alpha = 1 - q/2$. Механізм комплексування включається, якщо серед ознак є як статистичні, так і детерміновані. Він полягає у тому, для ста-

тистичної ознаки міра відміни її значень m_x та m_y дорівнює максимально можливому значенню 1, якщо різниця $m_x - m_y$ статистично значуща. Якщо жодна з ознак є статистичною, то потреби у згаданому механізмі немає і $\rho_k(X, Y)$ обчислюється у вигляді: $\rho_k(X, Y) = |m_x - m_y| / (u_a \cdot b(G_k))$.

Міра відміни об'єктів X та Y (віддаль між X та Y) за сукупністю ознак $G_1, G_2, \dots, G_r$ обчислюється у вигляді:

$$\rho(X, Y) = \left(\sum_{k=1}^r \alpha_k^2 \rho_k^2(X, Y) / r \right)^{0.5},$$

а віддаль між об'єктом Y та класом еталонів Q , віднесена до середньої внутрішньої віддалі ρ_0 :

$$\rho_0(Y, Q) = \rho(Y, Q) / \rho_Q = \left(\sum_{j=1}^b \rho^2(X_j, Y) / b \right)^{0.5} / \rho_Q.$$

де $\rho_Q = \left(2 \sum_{j=1}^b \sum_{i=j+1}^b \rho^2(X_j, X_i) / (b(b-1)) \right)^{0.5}$. Об'єкт Y буде вважа-

тись внутрішнім аналогом об'єктів класу Q , якщо $\rho_0(Y, Q) < 1$.

Завдання "Аналоги за методом найближчих зразків сигналу (Ans)" призначене для упорядкування об'єктів з 2-го по останній ряди за мірою близькості до об'єктів 1-го ряду.

Перевага запропонованого алгоритму у його ефективності навіть тоді, коли цей ряд забруднено завадою, а також коли клас корисних сигналів неоднорідний. Нехай Q - клас об'єктів $X_1, X_2, \dots, X_b$, до яких відшукуються аналоги; Y - об'єкт, що підлягає оцінці. Відстань об'єкта Y до класу Q обчислюється за формулою

$$\rho_{ou}(Y, Q) = \rho_u(Y, Q) / \rho_{uQ} = \left(\sum_{j=1}^{b(u)} \rho^2(X_j, Y) / b(u) \right)^{0.5} / \rho_{uQ}.$$

причому віддалі $\rho(X_1, Y), \rho(X_2, Y), \dots, \rho(X_b, Y)$ упорядковані за зростанням; $b(u)$ - округлене до найближчого цілого (≥ 1) значення $u \cdot b$; u - рівень підрізки;

$$\rho_{uQ} = (2 \sum_{j=1}^b \sum_{i=j+1}^b \rho_u^2(X_j, X_i) / b(u) / (b(u)-1))^{0.5}.$$

Завдання "Групування (Gru)" призначене для побудови ієрархічної схеми класифікації об'єктів за сукупністю ознак. Задача розв'язується шляхом описаного нижче узагальнення методу максимуму правдоподібності (ММП), зважаючи на його відомі оптимальні властивості. Як і у попередніх багатовимірних процедурах, головною проблемою було зведення у одній мірі статистичних та детермінованих ознак. Згадане узагальнення стосувалось саме її. Адже для статистичних ознак ММП приводить до мінімізації суми квадратів відхилень, тоді як для номінальних ознак алгебраїчні перетворення неможливі. Вихід було знайдено представленням суми квадратів відхилень через міру відміни пар об'єктів. Міра відміни визначена як за статистичними, так і за детермінованими (зокрема номінальними) ознаками з використанням критерію значимої різниці.

Зідане представлення суми квадратів відхилень для довільних числових значень $x_1, x_2, \dots, x_n$ має вигляд

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_i - x_j)^2 / n = \sum_{j>1}^n \sum_{i>1}^n \rho^2(x_i, x_j) / n,$$

де $\rho^2(x_i, x_j) = (x_i - x_j)^2$.

Позначимо Q - клас об'єктів $X_1, X_2, \dots, X_b$, які підлягають групуванню. Міра відміни об'єктів X_i та X_j за ознаками $G_1, G_2, \dots, G_r$ (віддаль) має вигляд:

$$\rho(X_i, X_j) = (\sum_{k=1}^r \alpha_k^2 \rho_k^2(X_i, X_j) / r)^{0.5}.$$

Якщо серед ознак $G_1, G_2, \dots, G_r$ є статистичні і детерміновані, то для статистичної ознаки

$$\rho_k(X_i, X_j) = |x_{ik} - x_{jk}| / (u_a \cdot \delta(G_k)), \text{ якщо } |x_{ik} - x_{jk}| / (u_a \cdot \delta(G_k)) < 1;$$

$$\rho_k(X_i, X_j) = 1, \text{ якщо } |x_{ik} - x_{jk}| / (u_a \cdot \delta(G_k)) \geq 1;$$

для детермінованої -

$$\rho_k(X_i, X_j) = 1, \text{ якщо } x_{ik} \neq x_{jk}; \rho_k(X_i, X_j) = 0, \text{ якщо } x_{ik} = x_{jk}.$$

У наведених формулах x_{ik}, x_{jk} - значення ознаки G_k об'єктів X_i та X_j відповідно. Якщо серед ознак $G_1, G_2, \dots, G_r$ немає детермінованих, а лише статистичні, то $\rho_k(X_i, X_j) = |x_{ik} - x_{jk}| / (u_a \cdot b(G_k))$.

Процес побудови класифікаційної схеми складається з $b-1$ послідовних кроків (b - кількість об'єктів, що підлягають групуванню). На n -му кроці переглядаються усі можливі об'єднання $b - n + 1$ підкласів, утворених на момент його виконання, тобто після попереднього $(n-1)$ -го кроку. Серед підкласів можуть бути й такі, що складаються з одного об'єкта. В ході перегляду визначаються такі два підкласи, при об'єднанні яких міра внутрішньої неоднорідності підкласів буде мінімальною. Такою мірою розбиття на підкласи $Q_1, Q_2, \dots, Q_t$ служить аналог квадратного кореня з суми квадратів відхилень: $S^2(Q_1, Q_2, \dots, Q_t) = S^2(Q_1) + S^2(Q_2) + \dots + S^2(Q_t)$, де

$$S^2(Q_j) = \sum_{p \in q(j), p > s \in q(j)} \rho^2(X_p, X_s) / b_j, \text{ якщо } b_j > 1,$$

$$S^2(Q_j) = 0, \text{ якщо } b_j = 1.$$

$q(j)$ - множина номерів тих об'єктів, які входять у Q_j . b_j - кількість об'єктів, які входять у Q_j . Для об'єднання підкласів Q_i та Q_j

$$S^2(Q_i \cup Q_j) = (b_i S^2(Q_i) + b_j S^2(Q_j) + \sum_{p \in q(i), s \in q(j)} \rho^2(X_p, X_s)) / (b_i + b_j)$$

Згідно з одержаною рекурентною формулою оптимальним буде об'єднання таких двох класів Q_i та Q_j , для яких різниця

$$\begin{aligned} & S^2(Q_1, Q_2, \dots, Q_i, Q_{i+1}, \dots, Q_j, Q_{j+1}, \dots, Q_t, Q_i \cup Q_j) - S^2(Q_1, Q_2, \dots, Q_t) = \\ & = \left(\sum_{p \in q(i), s \in q(j)} \rho^2(X_p, X_s) - b_i S^2(Q_i) - b_j S^2(Q_j) \right) / (b_i + b_j). \end{aligned}$$

мінімальна.

Для зручності зіставлення результатів, одержаних за різними

наборами ознак, міра внутрішньої неоднорідності підкласів кожного розбиття на підкласи віднесена до міри неоднорідності всієї сукупності залучених до задачі об'єктів: $M_0(Q_1, Q_2, \dots, Q_t) = S(Q_1, Q_2, \dots, Q_t) / S(Q)$, де

$$S(Q) = \left(\sum_{i=1}^{b-1} \sum_{j=i+1}^b \rho^2(X_i, X_j) / b \right)^{0.5}.$$

Міра $M_0(Q_1, Q_2, \dots, Q_t)$ знаходиться в межах від 0 до 1, сягаючи найбільшого значення 1 при об'єднанні всіх об'єктів в один клас на найвищому рівні класифікаційної схеми. Показник

$$K_t(Q_1, Q_2, \dots, Q_t) = M_0(Q_1, Q_2, \dots, Q_1, Q_{i+1}, \dots, Q_j, Q_{j+1}, \dots, Q_t, Q_i \cup Q_j) - M_0(Q_1, Q_2, \dots, Q_t).$$

названо контрастністю розбиття на підкласи $Q_1, Q_2, \dots, Q_t$ відносно розбиття на підкласи $Q_1, Q_2, \dots, Q_1, Q_{i+1}, \dots, Q_j, Q_{j+1}, \dots, Q_t, Q_i \cup Q_j$, утвореного з попереднього шляхом об'єднання підкласів Q_i та Q_j .

Завдання "Зображення в системі координат XOY в текстовому режимі роботи устроїв виводу (Izo)" призначене для відображення положення об'єктів у проекції на площину з двома координатними осями Ox та Oy. Завдання дає змогу давати якісні оцінки результатам фільтрації одразу кількома фільтрами у проекціях на двовимірні підпростори відгуків, візуально зіставляти моделі, побудовані з об'єктів БД, порівнювати діагностичні здатності фільтрів тощо.

Завдання "Мар - відображення на геологічній карті" включено до складу процедур з метою поліпшення інтерфейсу та робочих характеристик інформаційної системи. При проектуванні контуру графічного супроводження прийнято концепцію багат шарової карти на основі векторної графіки. Кожен з шарів має зберігатися в окремому файлі, причому карта мусить відображатися у будь-яких комбінаціях шарів. Масштаб може змінюватись у межах шарів одного рівня, а також з переходом на той рівень, що відповідає іншому масштабу листа. При наявності устрою підготовки інформації з'являється можливість безпосереднього оперування з картою для визначення об'єктів запиту. Виготовлено перший контур програмного втілення цієї концепції. Зразки виготовлених документів наведені у дисертації.

У третьому розділі описано методологію застосування інформа-

ційно-аналітичної системи до побудови кореляційної схеми геологічних утворень, уточнення положення об'єктів у кореляційній схемі, включення об'єктів до кореляційної схеми, металогенічного прогнозу. Наведено блок-схеми виконання цих функцій.

У четвертому розділі викладено основні теоретичні результати розробленої кореляційної теорії різновидової геологічної інформації. Введено такі позначення типів ознак: В - що реєструються у бінарній номінальній шкалі; N - у номінальній шкалі з більш ніж двома станами; P - у порядковій номінальній шкалі.

Коефіцієнт кореляції ознак A і X типу B має зміст:

$$r(A, X) = (P(AX) - P(A)P(X)) / (P(A)(1 - P(A))(P(X)1 - P(X)))^{0.5}$$

(P(A), P(X) - ймовірності подій A та X). Рівняння регресії A на X має вигляд:

$$A_v = P(A/X)X_v + P(A/\bar{X})\bar{X}_v + \Delta A_v = M(A_v/X)X_v + M(A_v/\bar{X})\bar{X}_v + \Delta A_v.$$

де A_v (X_v) - випадкова величина, яка приймає значення 1 або 0 залежно від того, є чи немає A (X). ΔA_v - випадкове відхилення від функції регресії; $P(A/X)$ - умовна ймовірність A при умові X; $\bar{X}$ - протилежну подія до X, $M(A_v/X)$ - умовне математичне сподівання. Доведено, що кореляційне відношення A на X має вигляд: $r(A, X) = (1 - D(\Delta A_v)/D(A_v))^{0.5} = |r(A, X)|$. Прогноз здійснюється за формулою:

$$A_p(X) = \{P(A/X)\}, A_p(\bar{X}) = \{P(A/\bar{X})\}$$

({...} - округлення до цілого). Середній квадрат відхилення від прогнозного значення -

$$M(A_v - A_p(X))^2 = \min(P(AX), P(\bar{A}/X) + P(A\bar{X}), P(\bar{A}\bar{X}))$$

збігається з середньою ймовірністю помилки прогнозу стану ознаки A за спростереженням над X.

Множинна кореляційна залежність між ознакою типу B та ознакою типу N описується рівнянням:

$$A_v = \sum_{j=1}^k P(A/X_j)X_{jv} + \Delta A_v = \sum_{j=1}^k M(A_v/X_j)X_{jv} + \Delta A_v.$$

Кореляційне відношення -

$$\rho_{A/X} = ((\sum_{j=1}^k P(A/X_j)P(X_j/A) - P(A))/(1 - P(A)))^{0.5}.$$

Прогноз А здійснюється за принципом мінімуму середньоочікуваної ймовірності помилки. Остання має вигляд:

$$\alpha_{A/X} = \sum_{j=1}^n \min(P(AX_j), P(\bar{A}X_j)).$$

а ймовірність вірного прогнозу -

$$\beta_{A/X} = \sum_{j=1}^n \max(P(AX_j), P(\bar{A}X_j)) = 1 - \alpha_{A/X}.$$

Прогноз такої якості забезпечується об'єднанням $Z = X_{j1} \cup X_{j2} \cup \dots \cup X_{jk}$ тих станів, для яких

$$\max(P(AX_j), P(\bar{A}X_j)) = P(AX_j).$$

Прогноз А виконується за станом $\bar{Z}$.

Регресія ознаки А типу N на ознаку X типу N має вигляд:

$$A_{jv} = \sum_{i=1}^k P(A_i/X_i) X_{iv} + \Delta A_{jv} \quad (j = 1, 2, \dots, m).$$

Кореляційне відношення - $\rho(A, X) = (1 - D(\Delta A_v)/D_0(\Delta A_v))^{0.5}$,

де

$$D(\Delta A_v) = \sum_{j=1}^m \sum_{i=1}^k (P(X_i) - P(A_j X_i))^2 P(A_j X_i) / P^2(X_i).$$

$$D_0(\Delta A_v) = \sum_{j=1}^m (1 - P(A_j))^2 P(A_j).$$

Прогноз ознаки А за заданим станом X_i ознаки Х із заданою довірчою ймовірністю q : $A_p(X_i) = A_{j_1} \cup A_{j_2} \cup \dots \cup A_{j_r}$, причому об'єднуються стани з найбільшими умовними ймовірностями $P(A_j/X_i)$ і такі, що $P(A_{j_1}/X_i) + P(A_{j_2}/X_i) + \dots + P(A_{j_r}/X_i) > q$. Точковий прогноз стану $A_p(X_i) = A_j^0$, де A_j^0 - стан, при якому максимальна ймовірність $P(A_j/X_i)$ при фіксованому X_i : $P(A_j^0/X_i) = \max P(A_j/X_i)$. Коefіцієнт прогнозованості обчислюється у вигляді:

$$\beta_{A/X} = (1 - \beta_A / \beta_{A/X}) / (1 - \beta_A) = (1 - \alpha_{A/X} / \alpha_A) / (1 - \alpha_{A/X}),$$

$$\text{де } \alpha_{A/X} = 1 - \sum_{j=1}^k \max P(A_j X_i), \quad \alpha_A = 1 - \max_j P(A_j).$$

Кореляційне відношення ознаки А типу В та ознаки ξ типу Р -

$$\rho(A/\xi) = ((M(f^2(\xi)) - P^2(A)) / (P(A)(1 - P(A))))^{0.5},$$

де $f(x)$ - функція регресії А на ξ , що являє собою умовне математичне сподівання вигляду: $f(x) = M(A_v/\xi=x) = P(A/\xi=x)$. Вираз $f(x)$ через умовні щільності розподілу $p_0(x)$ та $p_1(x)$ при відсутності та наявності ознаки А відповідно:

$$f(x) = P(A)p_0(x) / (P(A)p_1(x) + (1 - P(A))p_0(x)).$$

Кореляційне відношення оберненої залежності - ознаки ξ типу Р та ознаки А типу В -

$$\rho(\xi/A) = (1 - (D_1(\xi)P(A) + (1 - P(A))D_0(\xi)) / D(\xi))^{0.5};$$

$D_1(\xi), D_0(\xi)$ - умовні дисперсії ξ при наявності та відсутності А.

Кореляційні відношення ознаки А типу N та ознаки ξ типу Р в разі лінійного наближення функції регресії мають вигляд:

$$\rho(A/\xi) = (\sum_{j=1}^m \gamma_j^2 P(\bar{A}_j) P(A_j) (3P(\bar{A}_1) - 1)) / (\sum_{j=1}^m P^2(\bar{A}_j) P(A_j))^{0.5},$$

$$\rho(\xi/A) = ((1 - \sum_{j=1}^m P(A_j) D_j(\xi)) / D_j(\xi))^{0.5},$$

де $D_j(\xi)$ - умовна дисперсія величини ξ при $A = A_j$.

У п'ятому розділі описано досвід систематизації різновидової інформації на прикладі однієї з останніх редакцій регіональної бази даних (БД). У наведеному систематичному укажчику зазначені шкали реєстрації та повні переліки станів. Останні результати застосування методу, описані у розділі 6, одержані саме на цій базі.

У таблиці 3 перелічені групи ознак бази даних "Інtruзивні та метаморфічні породи УЩ" з наведенням системних імен.

Наведено приклад проектування розділу геофізичної інформації бази геологічних даних комплексного характеру, який було реалізовано у БД "Гранітоїди УЩ" масштабу 1:200000. У базі даних передбачені: загальна характеристика гравітаційного та магнітного полів - диференційованість, лінійність, характер, інтенсивність, форма; регіональні характеристики полів; локальні характеристики полів, які визначаються за локальною (місцевою) мережею - локальне середнє значення; середнє абсолютне відхилення; коефіцієнт варіації; характеристика Аббе; анізотропія поля; азимут градієнту поля; проекція одиничного вектора-нормалі до поверхні поля; середнє абсолютне відхилення проекції одиничного вектора-нормалі до поверхні поля.

У шостому розділі наведено приклади застосування методу до розв'язання задач прогнозу та картування. Технологія комп'ютерного прогнозу методом багатовимірної класифікації на основі спільного використання геологічної інформації якісного та кількісного характеру була вперше в Україні запропонована і успішно випробувана нами ще на початку 70-х років в межах Берегово-Беганського рудного поля. Ідея методу виявилась плідною, що згодом підтвердилось досвідом його ефективного застосування в різноманітних прогнозних задачах.

1. Локальне прогнозування перспективних площ Халчерангинського рудного поля (Центральне Забайкалля). Територія вивчена комплексом наземних методів у масштабі 1:50000. Об'єктами опису були локальні ділянки території - елементарні осередки, що відповідають 1 см² карти у масштабі дослідження.

Еталонні об'єкти були обрані в межах рудопроявів олова - 13 продуктивних, з крупними та середніми родовищами олова; 13 - безруд-

Групи ознак БД "Інтрुзивні та метаморфічні породи УЩ"

Розділ	Номер групи	Ім'я групи	Текстова назва групи	К-ть ознак у гр.	Номери ознак, що вх. у групу
1	2	3	4	5	6
1	1	ГИД	<u>Розділ 1. Надпородні ознаки</u>	31	1-31
	2	ГНР	Група ІД "Ідентифікація"	4	32-35
2	3	GX1	Група "ПР - резервна у шкалі Т"		
			<u>Розділ 2. "Склад хімічний"</u>		
	4	GX2	Група Х1 "Вміст породоутворюючих оксидів"	10	36-45
	5	GX3	Група Х2 "Вміст летючих компонентів"	10	46-55
	6	GX4	Група Х3 "Вміст оксидів рідкісних елементів"	5	56-60
	7	GX5	Група Х4 "Вміст оксидів розповсюджених елементів"	9	61-69
	8	GAT	Група Х4 "Резерв у шкалі D"	9	70-78
	9	GNC	Група АТ "Атомна кількість головних породоутворюючих елементів"	16	79-94
	10	GCE	Група НС "Нормативні вмісти мінералів (за CIPW у атомно-молекулярних %)"	17	95-111
3	11	GPO	Група СЕ "Серіальна належність"	7	112-128
			Група ПО "Петрохімічні відношення"	88	129-216
			<u>Розділ 3. "Геохімія"</u>		
	12	GG1	Група Г1 "Хімічні елементи літофільні, лужні і лужно-земляні"	9	217-225
	13	GG2	Група Г2 "Хімічні елементи флюїдні"	6	226-231
	14	GG3	Група Г3 "Хімічні елементи сидерофільні"	9	232-240
	15	GG4	Група Г4 "Хімічні елементи-лантаноїди"	16	241-256
	16	GG5	Група Г5 "Халькофільні хімічні елементи"	17	257-273
	17	GG6	Група Г6 "Рідкісні і радіоактивні елементи"	19	274-292
4	18	GGO	Група Г0 "Геохімічні відношення"	55	293-347
			<u>Розділ 4. "Склад петрографічний"</u>		
	19	GTP	Група ТП "Типоморфні породоутворюючі мінерали"	5	348-352
	20	GPM	Група ПМ "Породоутворюючі мінерали, вміст (у об'ємних %)"	30	353-382
5	21	GPB	Група ПВ "Порфірові виділення та основна маса"	6	383-388
			<u>Розділ 5. "Будова породи"</u>		
	22	GAC	Група АС "Самородні елементи, сульфіди і сульфосоли"	21	389-409
	23	GAO	Група А0 "Акцесорні оксиди"	24	410-433
	24	GAP	Група АР "Акцесорні мінерали рідкоземельних і радіоактивних елементів"	15	434-448
	25	GAL	Група АЛ "Акцесорні мінерали летючих компонентів, тантало-ніобати і вольфраматів"	18	449-466

	26	GAИ	Група АИ "Акцесорні силікати і алюмосилікати"	15	467-481
	27	GAФ	Група АФ "Додаткові акцесорні мінерали"	14	482-495
	28	GTE	Група ТЕ "Текстура"	5	496-500
	29	GCT	Група СТ "Структура"	7	501-507
	30	GBЗ	Група ВЗ "Взаємостосунки породотворюючих мінералів"	22	508-529
			<u>Розділ 6. "Фізичні властивості"</u>		
6	31	GФС	Група ФС "Фізичні властивості гірських порід"	8	530-537
			<u>Розділ 7. "Вікова і формаційна належність"</u>		
7	32	GBФ	Група ВФ "Вікова і формаційна належність"	6	538-543
	330	ООП	Група "Формаційна належність об'єкта"	3	544-546
	34	GTM	Група "Тектоно-магматичний етап формування"	9	547-555
			<u>Розділ 8. "Положення, склад, форма, розміри і будова масиву"</u>		
8	35	GПТ	Група ПТ "Положення тіла в структурі"	2	556-557
	36	GMC	Група МС "Склад, будова, форма та розміри масиву"	6	558-563
			<u>Розділ 9. Спеціалізація масиву (комплексу)"</u>		
	37	GГМ	Група ГМ "Геохімічна спеціалізація масиву (комплексу) відносно кларків елементів в земній корі"	41	564-604
	38	GAЦ	Група АЦ "Акцесорна спеціалізація масиву"	12	606-617
			<u>Розділ 10. Накладені процеси</u>		
	39	GГИ	Група ГИ "Гідротермально-метасоматичні прояви, асоціючі з масивом"	38	618-655
	40	GMB	Група МВ "Метаморфізм вміщуючих порід"	7	656-662
			<u>Розділ 11. Асоціючі жильні та дайкові породи і рудопрояви"</u>		
11	41	GAЖ	Група АЖ "Асоціючі жильні та дайкові породи"	15	663-677
	42	GPM	Група РМ "Асоціючі рудопрояви та рудна мінералізація"	24	678-701
	43	GBK	Група ВК "Включення (ксеноліти, ксенокристали, рестіти, скіаліти)"	10	702-711
	44	GRD	Група RD "Резервна у шкалі F"	22	712-733
	45	GRW	Група RW "Резервна у шкалі W"	5	734-738
	46	GRT	Група RT "Резервна у шкалі T"	20	739-758
	47	GRB	Група RB "Резервна у шкалі B"	20	759-778
	48	GRA	Група RA "Резервна у шкалі A"	4	779-782
	49	GRY	Група RY "Резервна у шкалі Y"	15	783-797
	50	GRS	Група RS "Резервна у шкалі S"	3	798-800

них. Спочатку було залучено 30 геоморфологічних, геохімічних та геофізичних ознак. Після аналізу інформативності залишено ознаки, перелічені нижче.

Геологічні ознаки:

- наявність оперяючих розломів;
- наявність зон дроблення з підвищеним вмістом сурми, свинцю

або олова:

- простягання розломних зрушень.

Геоморфологічні ознаки:

- абсолютна позначка над рівнем моря;
- форма рельєфу;
- напрямок середнього ухилу.

Геохімічні ознаки:

- лінійно-логарифмічний фільтр вмістів хімічних елементів Sn, Pb, Zn, Ag, Cu, Mo, Bi

- лінійно-логарифмічний фільтр вмістів хімічних елементів Pb, Zn;

- інтервал вмісту олова у ореолі розсіювання;
- різниця лінійно-логарифмічних фільтрів вмістів хімічних елементів Pb, Zn, Ag і Cu, Bi, Mo.

Геофізичні ознаки:

- середня віддаль до поверхні інтрузії;
- форма поверхні інтрузивного тіла;
- напрям середнього ухилу підземного рельєфу;
- величина середнього ухилу;
- простягання аеромагнітних аномалій.

Після проведеного тесту найбільш інформативними виявились:

- середня віддаль до поверхні інтрузії;
- напрям середнього ухилу підземного рельєфу;
- простягання розломних зрушень;
- абсолютна позначка над рівнем моря;
- лінійно-логарифмічний фільтр вмістів хімічних елементів

Sn, Pb, Zn, Ag, Cu, Mo, Bi

- інтервал вмісту олова в ореолі розсіювання.

В ході експериментів шляхом тестування з'ясувалося, що з двох процедур - КЛА та КЛН більш ефективна КЛН. При використанні процедури КЛА ймовірність належності рудних об'єктів до свого класу становила 0.85, нерудних - 0.92. При використанні процедури КЛН тест показав, що обидві ймовірності близькі до 1.

Зона еталонних рудних об'єктів значно розширена. Крім того, визначена нова не виявлена традиційними методами рудна область, приурочена, як з'ясувалося, до зони дроблення з піротиновою мінералізацією. Прогнозні об'єкти були рекомендовані як пріоритет-

ні для постановки розвідувальних робіт на олово та поліметали.

Халчерангінський дослід показав високу ефективність методу багатовимірної статистичної фільтрації для локального прогнозування.

2. Прогнозування родовищ олова у оловорудному районі Далеко-східного Примор'я з класифікацією за промисловим значенням.

Програмні засоби оцінки інформативності та класифікації показали високу ефективність не тільки для локалізації продуктивних об'єктів, але й у вирішенні більш загальної задачі - прогнозі їх промислового значення. Об'єктами прогнозування служили елементарні осередки, що відповідають 1 см^2 карти масштабу 1:50000. З огляду на формаційні типи та промислове значення виділено 5 класів об'єктів:

Q_1 - з каситерит-силікатними та каситерит-сульфідними родовищами промислового значення (крупними та середніми);

Q_2 - з каситерит-кварцевими дрібними родовищами промислового значення;

Q_3 - з кварц-вольфрамітовими (з каситеритом) дрібними родовищами промислового значення;

Q_4 - з каситерит-силікатними та каситерит-сульфідними непромисловими рудопроявами;

Q_5 - без ознак мінералізації.

Вибір еталонів на формаційній основі обумовлений тим, що формаційні типи відрізняються геологічною будовою, речовинним складом, зв'язком з інтрузивними утвореннями, завдяки чому мусять виразно відокремлюватися у ознаковому просторі. Крім того, формаційна належність родовища здебільшого визначає його промислове значення. Оскільки класи Q_4 та Q_5 складали альтернативу до класів Q_1 - Q_3 , потрібно було забезпечити дієвість методу насамперед саме на цих двох типах сигналів. При виборі еталонів класів Q_1 - Q_4 відчутно давався взнаки їхній дефіцит, що не могло не відбитися на підсумкових результатах проведеного дослідження.

З геологічних ознак найбільш інформативними у відношенні локалізації родовищ виявилися:

- 1) напрямок складчастих структур;
- 2) відношення віддалі від Колумбійського розлому до віддалі від Центрального шову;
- 3) наявність штоків та дайок різного складу;
- 4) ступінь регіонального метаморфізму порід.

Що стосується штоків та дайок, то з цього боку об'єкти характеризувалися такими номінальними ознаками:

- наявність дайок будь-якого складу;

- наявність дайок середнього складу;
- наявність дайок кислого складу;
- наявність штоків;
- наявність штоків кислого складу;
- наявність штоків середнього складу.

Для ділянок локалізації родовищ олова характерний розвиток дайок та штоків кислого та середнього складів.

Залежно від ступеню регіонального метаморфізму виділено три зони:

- 1) ороговикованих порід (біотитів);
- 2) перехідна зона пропілітовиднозмінених порід з наявністю кварцю, серициту, хлориту, епідоту, польового шпату;
- 3) слабо змінені породи.

Для родовищ олова найбільш сприятливою виявилася перехідна зона.

Серед геоморфологічних ознак інформативними виявилися:

- абсолютна позначка (для об'єктів з родовищами характерні відносно високі позначки - 900 м і більше);
- розчленованість рельєфу (для ділянок родовищ характерні значення 3 - 5 умовних одиниць за шкалою, що охоплює 0 - 5 одиниць);
- напрямок ізогіпс рельєфу (для ділянок родовищ характерні значення 60° - 90° , що, вірогідно, відповідає напрямку рудоконтролюючих структур).

Серед геофізичних ознак інформативними виявилися:

- щільність аномалій природного електричного поля (для ділянок родовищ 1 - 4 шт/км з інтервалу 0 - 6 шт/км);
- співвідношення площ додатніх та від'ємних магнітних аномалій. Станами цієї номінальної ознаки були:
- аномалії магнітного поля відсутні;
- є тільки додатні аномалії;
- переважають додатні аномалії;
- додатніх та від'ємних аномалій однакова кількість;
- переважають від'ємні аномалії;
- тільки від'ємні аномалії.

Для ділянок родовищ була характерною переважність від'ємних аномалій.

Найбільш інформативні гідрохімічні та геохімічні ознаки:

- вміст суми металів у водах (для ділянок родовищ >30 мг/л);

- вміст сульфат-іонів у водах (для ділянок родовищ $>16 \text{ мг/л}$);

- вміст свинцю та олова у вторинних ореолах розсіювання (для ділянок родовищ $>0.06\%$).

Кількість еталонів складала: 18 - класу Q_1 ; 10 - класу Q_2 ; 11 - класу Q_3 ; 18 - класу Q_4 ; 26 - класу Q_5 . Тестування проводилося по контрольних об'єктах, що не входили у еталонні виборки. Помилки у виділенні сигналів класів $Q_1 - Q_3$ склали 11%, що за умов низької репрезентативності та досить розмитих границь між цими класами можна вважати досягненням. Ступінь локалізації перспективних ділянок висока: до класу Q_1 віднесено 8%, до Q_2 - 4%, до Q_3 - 10%, до Q_4 - 20% території. Решта території (58%) віднесена до класу Q_5 , що свідчить про неперспективність цієї її частини. У другу чергу до них слід додати 20% території, віднесеної до класу Q_4 , отже перспективна площа локалізована до 22%. Першочерговими для геологорозвідувальних робіт слід вважати ділянки, віднесені до класу Q_1 . Високими ймовірностями належності до цього класу відзначені тестові об'єкти - відомі родовища олова, що не брали участі у еталонних виборках. Крім того, виділено ще ряд перспективних ділянок у відношенні локалізації крупних родовищ олова.

Було проведено порівнювальну оцінку запропонованого методу з технологією розпізнання образів, у якій діагностика здійснюється за т.з. інформативними сполуками ознак. Подібні алгоритми передбачають попереднє кодування кількісних значень ознак за кількома інтервалами. З точки зору теорії фільтрації таке кодування є введенням даних до загрубленого лінійного фільтру, коефіцієнти якого обернено пропорційні розмаху значень ознак. Такі коефіцієнти, як показано, не є оптимальними, хоча в деяких випадках можуть бути близькими до них; крім того, результат залежить від суб'єктивного фактору - вибору інтервалів (градацій) для кодування.

Зіставлення результатів показало помітно кращу локалізацію перспективних ділянок за методом багатовимірної фільтрації. До того ж результати більш впевнено підтверджуються геологічними матеріалами. Геологічний аналіз результатів показав, що практично безпомилково відділяються об'єкти каситеріт-силікатно-сульфідної формації (класи Q_1 та Q_4) від об'єктів каситеріт-кварцевої та кварц-вольфрамітової формацій (класи Q_2 і Q_3). Досить чітко відділяються від інших класів об'єкти класу Q_1 . Зокрема, впевнено діагностуються всі еталонні та контрольні об'єкти. Крім того, 6 об'єктів, що були раніше віднесені геологами до непромислових, класифі-

ковані як промислові (Q_1) і ця оцінка підтверджена згодом геолого-геохімічними дослідженнями. В свою чергу, одне з родовищ класифіковане лише як рудопрояв (клас Q_4), що також підтверджено результатами бурових робіт. Разом з тим, три об'єкти з 12, віднесені до класу Q_1 , протирічили геологічним оцінкам, хоча прямого доказу їх неперспективності не було, оскільки бурові роботи на них не проводились. В цілому проведений дослід прогнозу був цілком успішним, зважаючи на об'єктивну складність поставленої задачі. Адже йшлося про виділення відразу 4 різних сигналів, границі між якими у ознаковому просторі не тільки розмиті, а й певною мірою умовні.

3. Комплексна інтерпретація промислово-геофізичних даних (на прикладі теригенних відкладів Волгоградської області). Метою експерименту було порівняння з іншими алгоритмами на реальному виробничому матеріалі при оцінці колекторських властивостей та характеру насичення за результатами свердловинних вимірювань.

У Волгоградській області промислово продуктивні горизонти теригенного девону занурені на великі глибини до 2800 - 3400 м. Породи неоднорідні за структурою, текстурою, ступенем глинистості, ущільнення, складом цементу. Шари нерідко малої потужності та невитримані по площі. Результати вимірювань у свердловині змінюються через глибоке проникнення у породи фільтрату бурового розчину, явище екранування, перешаровування. Через ці причини інтерпретація становить вельми непросту задачу, що й обумовило вибір саме цього району для проведення експерименту.

Наведені результати застосування п'яти різних методів для визначення колекторських властивостей та характеру насичення 60 шарів у межах Кудиновського родовища, Пан-Шляховської та Ново-Добринської площ області. Серед шарів 20 - нафтонасичених, 20 - водонасичених і 20 - неколекторів. Еталонами послужили також 60 шарів - теж по 20 кожного з перелічених типів. Характер насичення у всіх 120 шарів установлений безпосереднім опробуванням. До нафтонасичених віднесені шари, які дали притоки нафти; до водонасичених - які дали притоки шарової води; до неколекторів - ті, що не дали ніякого притоку. Всі дані були свого часу люб'язно надані автору промислово-геофізичною лабораторією Волгоградського НДІ нафтової та газової промисловості (керівник лабораторії - М.Г. Белоусова). Автор реконструював задачу інтерпретації цих даних з подвійною метою: продемонструвати ефективність свого методу на геофізичному матеріалі, який було раніше інтерпретовано за алгоритмами інших авторів; оцінити ефективність комплексування різних показників в

умовах задачі такої складності, котра виключала б можливість іншого шляху.

Для характеристики шарів вжиті 13 показників.

1) Уявний опір (ρ_0), що вимірювався при різних довжинах зонду: 0.55 м ($\rho_{0.55}$); 1.05 м ($\rho_{1.05}$); 2.75 м ($\rho_{2.75}$); 4.25 м ($\rho_{4.25}$); 8.5 м ($\rho_{8.5}$);

2) питомий електричний опір (ρ_n);

3) відношення уявного опору за мікрозондом до опору бурового розчину проти розмитих глин ($\rho_{мз}/\rho_n$);

4) глибина каверн ΔR ; ,

5) товщина глинистої корки $h_{тк}$;

6) показання нейтронного гама-методу у відносних одиницях, виправлених за вплив умов вимірювання (глибини каверн, глинистої корки) ΔI_{ng} ;

7) показання гама-методу у відносних одиницях, виправлених за вплив умов вимірювання (ΔI_g);

8) відношення амплітуди власної поляризації (ВП) проти досліджуваного шару до амплітуди ВП проти опорного шару ($\Delta E_{вп}$);

9) "геофізичний параметр насичення" $R_{нт}$ - відношення питомого опору проти досліджуваного шару до величини ρ_n проти непродуктивного шару того ж типу, що і досліджуваний шар.

Оцінки щільностей розподілів підтвердили об'єктивну складність поставленої задачі: в значній мірі вони накладаються одна на одну, що виключає можливість побудови простих критеріїв інтерпретації.

Із сподівань на оптимальні властивості запропонованого нами критерію для комплексної інтерпретації було вжито завдання ТКН ("Тест класифікації непараметричної") відразу для всіх показників, без будь-якої підстройки шляхом вибірки гірших з них, підбору вагових коефіцієнтів тощо. Оптимальний метод мусить настроюватися на найбільш ефективний режим автоматично. Результати проведеного експерименту показали це очевидним чином: середня ймовірність помилок становить 0.013 (1.3%); ймовірність пропуску нафтонасиченого шару - 0.04 (4%).

З метою подальшого зменшення похибок при виділенні продуктивних об'єктів проведено експеримент фільтрації з однією альтернативою - непродуктивні об'єкти. Така постановка задачі обіцяє в усякому разі не більший рівень помилок, ніж у розглянутій вище більш складній задачі виділення 3-х сигналів. Наведені результати тестування наочно демонструють ефективність запропонованого мето-

ду: помилки пропуску продуктивних об'єктів становлять 2%; хибні висновки про продуктивність практично виключені. Продemonстровано також ефективне вирішення задачі фільтрації, де корисним сигналом служать колектори, а фоном - неколектори.

Представляють інтерес результати експерименту, метою якого було порівняння ефективності різних алгоритмів інтерпретації за 60 контрольними об'єктами, про які йшлося вище (20 - водонасичених, 20 - нафтонасичених і 20 - неколекторів).

1. Алгоритм "мінімальних кодових віддалей". Вирішальне правило подібних алгоритмів ґрунтується на порівнянні середніх віддалей від досліджуваного об'єкта до еталонних об'єктів кожного класу.

2. Метод поділяючих функцій, якими ознаковий простір поділяється на області прийняття гіпотез про належність об'єкта до класів. До такого типу належить, зокрема, класичний метод багатовимірної статистики, відомий під назвою "метод дискримінантних функцій".

3. Алгоритм розпізнання за діагностичними ("інформативними") сполуками кодів ознак. Кожна ознака кодується шляхом рознесення значень по кількох заздалегідь визначених інтервалах.

Результати експерименту наведені у табл. 4.

Таблиця 4

Результати інтерпретації контрольних шарів за характером насичення. Позначення: В - водонасичені; Н - нафтонасичені; НК - неколектори.

Алгоритм	Кількість помилок			% помилок		
	Н	В	НК	Н	В	НК
Мінімальних віддалей	8	5	4	40%	25%	20%
Розділяючих функцій	4	1	—	20%	10%	—
Інформативних сполук кодів* (без $P_{нг}$)	6 (з 30)	13 (з 30)		20% (з 30)	43% (з 30)	
Комплексної статистичної фільтрації	1	1	0	5%	5%	0%

Примітка*. Розв'язувалась задача дихотомії на 2 класи - продуктивні - непродуктивні. Використано по 30 контрольних об'єктів з кожного класу.

Як видно з поданих у таблиці даних, експериментальна перевірка показала більшу ефективність методу комплексної статистичної фільтрації.

4. Застосування методу статистичної фільтрації у задачах геологічного картування (на прикладі Волинського плутону).

Волинський плутон у геологічному відношенні досить складний. Навіть мікроскопічний аналіз наштовхується на труднощі через складність геологічної будови, структурно-текстурну нерівномірність та неоднорідність. Тому було запропоновано з'ясувати можливості розчленування основних петрографічних різновидів порід Волинського плутону за результатами відносно дешевих спектральних аналізів вмісту хімічних елементів.

Еталонним матеріалом послужили виборки вмістів малих хімічних елементів у пробах, належність яких до певних петрографічних різновидів була однозначно встановлена на основі всебічного аналізу - мікроскопічного визначення шліфів та вимірів породоутворюючих оксидів. Всього було сформовано 6 еталонних виборок: 1) габро-лабрадорити; 2) олівинові габро-лабрадорити; 3) дрібно- та середньозернисті габро- та габронорити; 4) лабрадорити; 5) крупнозернисті габро та олівинові габро; 6) габро-лабрадорити північно-східної частини плутону - т.з. чеповичиський тип. Після аналізу інформативності всіх показників було залишено 13: вмісти Mg, Cr, Mn, Cu, Ti, V, Ni, Co, Pb, Mo, Zr, Zn, Ga. Попередній статистичний аналіз показав досить складні умови для виділення шести різних корисних сигналів: розподіли накладаються один на другий; досить значна аналітична похибка - до 30%; наявність кореляційних зв'язків, в тому числі і наведених при вимірюванні.

Контрольні результати свідчать про успіх проведеного експерименту, який доводить можливість вирішення задач геологічного картування за геохімічними даними. Ефективність методу за рахунок набагато менших витрат порівняно з мікроскопічним аналізом очевидна. Всього для аналізу було залучено близько 400 проб. Частина зразків пройшла мікроскопічний аналіз для визначення їхнього петрографічного різновиду. Результати виведені на кольорову карту-схему, виготовленій за процедурою MAP.

5. Застосування методу багатовимірної статистичної фільтрації для прогнозу нікеленосності на території Українського щита.

У попередніх прикладах йшлося про досвід застосування методу для локального прогнозування. Проте локалізації таких площ, яку здійснюють в ході крупномасштабних робіт, передусь прогноз на ста-

дії дрібномасштабних робіт. Завданням такого прогнозу є виявлення структурно-формаційних комплексів, масивів, перспективних у металогенічному відношенні. Для якісного прогнозування необхідно провести комплексний аналіз всієї наявної геологічної інформації, зіставити її з накопиченим досвідом, експертними оцінками та винести рішення, від якого залежить напрямок подальших геолого-розвідувальних робіт і, отже, доля тих невеликих витрат, з якими вони пов'язані. Матеріали для апробації методу були запропоновані Центральною тематичною експедицією Держгеолкомітету України під час виконання спільної науково-дослідної тематики. Для прогнозних робіт було використано процедури "Аналоги" та "Аналоги за методом найближчих зразків сигналу" з упорядкуванням об'єктів БД за близькістю до ряду еталонних металогенічних об'єктів. Результати супроводжувалися відображенням об'єктів на діаграмах у проєкціях на головні петрохімічні компоненти за допомогою процедури "DSK".

Задача розв'язувалась шляхом пошуку петрохімічних аналогів відомих нікеленосних об'єктів Кольського півострова, Воронежського масиву, Норильського родовища. Компоненти фільтру були розраховані разом із своїми середніми квадратичними відхилами і розміщені у БД процедурою ПРЕ. Зовнішніми зразками корисного сигналу служили:

1) Прутовський комплекс; 2) нікеленосні масиви Печенги; 3) нікеленосні масиви Норильського району; 4) масиви Воронежського кристалічного масиву. Багатовимірною статистичною моделлю еталонів Печенги на ознаковому просторі згаданих компонент фільтру, як показав її аналіз засобами "Фільтру", виявилася неоднорідною. Тому було проведено оптимізацію моделі сигналу.

Результати фільтрації показали високу ефективність методу, а якщо взяти до уваги, що у експерименті брали участь більше 2500 об'єктів, то винятково високий ступінь локалізації прогнозу. Ще одним підтвердженням ефективності методу стали результати прогнозу на основі зразка сигналу, породженого у середовищі БД зовнішніми еталонами Норильського комплексу.

З метою зіставлення результатів з методом головних компонент виконано відображення прогнозних об'єктів та зразків сигналу у проєкціях F1 - F2, F1 - F3, F2 - F3. Головні компоненти F1, F2, F3 були розраховані процедурою ПРЕ.

Як показали одержані процедурою DSK діаграми, велика кількість об'єктів Прутовського масиву, однозначно спрогнозованих методом багатоконцентної фільтрації як нікеленосні, на діаграмах явно випадають з поля зовнішніх еталонів нікеленосності, змішую-

чись з непродуктивними. Наведені дані з очевидністю демонструють переваги запропонованого методу у вирішенні прогностичних задач.

Заклучення

В роботі узагальнено результати багаторічних досліджень автора у розробці нового методу інтегрованого аналізу геологічної інформації для вирішення прогностично-пошукових задач. Будучи переконаним прихильником того, що науково-методичні новації мають довести свою перевагу на практичних задачах, автор поставив за мету втілити метод у діючий інструмент. Результатом цих зусиль стала інформаційно-аналітична система інтегрованої геологічної інформації "Фільтр". Передумовою її створення було вирішення низки проблем, починаючи з систематизації та формалізації геологічних даних, розробки комп'ютерних програм і закінчуючи здобуванням інформації та її підготовкою для розробленої технології. Серед залучених геологічних даних найбільший інтерес представляли ті матеріали, на яких можна було б випробувати спроможність методу при вирішенні складних прогностичних задач. У роботі наведена частина геологічних результатів, які демонструють його можливості.

Впродовж створення інформаційно-аналітичної системи предметом постійного піклування автора були засоби для зручного і оперативного спілкування з нею. Автор вбачав своїм обов'язком зробити їх такими, щоб системою в усій повноті могли б користуватися геологи, які не мають спеціальної підготовки, і навіть студенти геологічних спеціальностей під час виконання курсових та дипломних робіт. Аналітичні процедури можуть працювати з будь-якими шкалами реєстрації геологічної інформації. Всі програми розроблені автором, жоден з модулів не є "запозиченою" або одержаною законним шляхом продукцією якогось іншого автора або фірми. Аналога методу і відповідної комп'ютерної реалізації в Україні немає. Мова користувача має засоби для виконання перетворень середовища бази даних (БД), взаємодії операторів у потоці запитів, створення та збереження моделей задач, структурування інформації та її обробки в структурованому вигляді, прискореного пошуку в БД, обстеження та корекції її стану, автоматичного моделювання багатовимірних образів корисних сигналів на основі всього потенціалу БД, оптимізованих процедур прогнозу. Сприйняття повідомлень не потребує спеціальних знань. За наявності зовнішніх устроїв відповідних форматів забезпечена можливість при формулюванні задачі адресуватися безпо-

середньо до геологічної карти, координатно зв'язаної з середовищем БД. При цьому результати розв'язання задачі система також може відобразити на цій карті або вивести їх разом з самою картою.

Автор вважає своїм досягненням реалізовану можливість спільної роботи аналітичних процедур з усіма шкалами реєстрації. Це стосується і тих модулів, що базуються на кореляції вживаних ознак.

Систематизація та формалізація геологічної інформації складала проблему, без вирішення якої мета роботи була б недосяжною. Результатом стала детально розроблена нова методика реєстрації геологічних даних і створені за нею кілька версій потужної БД по території Українського щита. Цей результат роботи має самостійне методологічне значення як ефективний інструмент забезпечення якості геологічних даних, а також підвищення корисного виходу з неї. Зараз вже не підлягає сумніву, що в сьогоdnішніх умовах це єдиний реальний шлях збереження пошукової дієвості накопичених геологічних матеріалів в масштабі всієї країни.

Автор має надію, що завдяки впровадженню запропонованої системи стануть доступними для оперативного аналізу накопичені масиви тієї інформації, що переважно знаходяться поза сферою сучасних комп'ютерних технологій, які орієнтуються на багатовимірний аналіз. Йдеться, зокрема про макро- та мікроописи порід, їхні властивості, структурно-текстурні особливості, формаційні, морфологічні ознаки тощо, які практично недоступні для оперативного аналізу через неприйнятність форм реєстрації і відсутність відповідних програмних засобів. Завдяки цьому істотно збільшиться інформативна цінність і в багато разів - рівень використання геологічної інформації, що врешті-решт є головним критерієм ефективності роботи галузі, головним надбанням якої є саме інформаційні ресурси.

Автор вбачає найближчу перспективу її розвитку у тому, щоб національне надбання - геологічна інформація стала високотехнологічною продукцією. Він має надію, що його скромний вклад сприятиме позитивним зрушенням у цьому напрямку. Ці зрушення мають втілитися у впровадженні сучасних аналітичних технологій, створення крупних баз різновидових даних національного рівня з ретельно зібраними інформаційними колекціями - запорукою подальшого поповнення інформаційного багатства України.

Матеріали дисертації опубліковані у 71 працях, в тому числі у 3-х одноосібних монографіях та у 4-х монографіях, виданих з співавторами. Список основних статей, книг та брошур подано нижче.

1. Жуков М.Н. Про критерій нормальності розподілу хімічних

елементів в гірських породах з альтернативою логнормального розподілу. Вісник Київського університету. Серія геологія. -К., вип. 8, 1966.

2. Жуков М.Н. До питання про визначення аномальних значень за геохімічних розшуків. Вісник Київського університету. Серія геологія. -К., вип. 9, 1967.

3. Жуков Н.Н., Нагорский В.А. О методе усредненных проб при геохимических поисках. -В сб. "Материалы по геологии геофизике и геохимии Украины, Молдавии, Казахстана и Забайкалья", вып. 4, 1968.

4. Толстой М.И., Жуков Н.Н. О сопоставлении геологических образований по количественному распределению в них химических элементов. -В сб. "Материалы по геологии геофизике и геохимии Украины, Молдавии, Казахстана и Забайкалья", вып. 4, 1968, с. 117-128.

5. Жуков М.Н., Дядюра В.О. Методика статистичного аналізу даних геохімічного випробування. Вісник Київського університету. Серія геологія. -К., вип. 10, 1968.

6. Жуков Н.Н., Толстой М.И., Нагорский В.А., Остафийчук И.М. Оценка металлогенической специализации магматических образований на основе математического анализа количественной геохимической информации. Международный геологический конгресс, XXIII сессия. Доклады советских геологов. Математические методы в геологии. 1968, с. 61-62.

7. Толстой М.И., Жуков Н.Н. О выборе количества проб при геохимической и петрофизической характеристике изверженных образований. -В сб. "Материалы по геологии геофизике и геохимии Украины, Молдавии, Казахстана и Забайкалья", вып. 5, 1969, с. 133-137.

8. Жуков М.Н. До питання про правомірність використання методів математичної статистики випадкових величин при аналізі геологічної інформації. Вісник Київського університету. Серія геологія. -К., вип. 11, 1969.

9. Шептура В.И., Криволап Л.С., Дядюра В.А., Жуков Н.Н. Применение метода распознавания образов для разделения гранитоидов на основе качественных признаков. -В сб. "Материалы по геологии геофизике и геохимии Украины, Молдавии, Казахстана и Забайкалья", вып. 5, 1970, с. 71-75.

10. Остафийчук И.М., Жуков Н.Н. Сопоставление геологических объектов на основе многомерных распределений химических элементов. // "Материалы по геологии геофизике и геохимии Украины, Молдавии, Казахстана и Забайкалья", вып. 6, 1970, с. 84-99.

11. Толстой М.И., Жуков Н.Н., Остафийчук И.М. Об особенностях

пространственного распределения химических элементов в магматических образованиях в связи с их количественной оценкой. Геохимия, 12. - М., 1971, с. 1492-1494.

12. Жуков Н.Н. О построении оптимальных способов площадного опробования геологических объектов. "Материалы по геологии геофизике и геохимии Украины, Молдавии, Казахстана и Забайкалья", вып. 7, 1971, с. 116-122.

13. Жуков Н.Н., Молявко В.Г., Остафийчук, И.М., Степченко С.Б. Определение фациальной принадлежности горных пород по данным геохимического опробования методом многомерной классификации на ЭВМ (на примере Волынского массива). Геологический журнал, т. 31, вып. 1, 1971.

14. Сухорада А.В., Жуков Н.Н. Количественная интерпретация терромагнитных кривых. Геофизический сборник. - К., вып. 46, 1972, с. 57-60.

15. Жуков Н.Н., Голуб В.Н. Выбор системы признаков для анализа неоднородностей физических полей при решении задач геологического картирования. Геологический журнал АН УССР, т. XXXIV, вып. 1, Киев, 1974.

16. Шептура В.И., Латыш В.Т., Зинченко В.Т., Жуков Н.Н., Серга А.Ю. Эндогенные ореолы рассеяния редких элементов вокруг пегматитов Волыни и их использование в поисково-разведочной практике. В сб. "Вопросы прикладной геохимии и петрофизики. - К., 1974, с. 110-115.

17. Жуков Н.Н., Серга А.Ю. Об учете природной и аналитической дисперсий при планировании способа геохимического опробования. В сб. "Вопросы прикладной геохимии и петрофизики. - К., 1974, с. 138-148.

18. Жуков Н.Н., Мишин Л.Т., Серга А.Ю., Тарасенко Н.Я. Метод количественного прогнозирования месторождений олова с помощью программы "Непараметрическая классификация". Вопросы прикладной геохимии и петрофизики. - К., 1975, с. 127-138.

19. Толстой М.И., Жуков Н.Н. Об опробовании магматических образований при их общей геохимической характеристике. Вопросы прикладной геохимии и петрофизики. - К., 1975, с. 101-112.

20. Жуков Н.Н. Вероятностно-статистические методы анализа геолого-геофизической информации. - К.: Вища школа, 1975. 303 с.

21. Жуков Н.Н., Серга А.Ю. Применение метода разделения неоднородной совокупности по комплексу признаков для расчленения вулканогенных толщ Советского Закарпатья. Вопросы прикладной геохимии и петрофизики. - К., 1976, с. 101-107.

22. Табачний Л.Я., Жуков Н.Н. Некоторые аспекты оценки точности результатов спектрального анализа. "Вопросы прикладной геохимии и петрофизики". - К., 1978, с. 120-125.

23. Толстой М. И., Ширинян К. Г., Остафийчук И. М., Адамян А. А., Жуков Н. Н., Карапетян С. Г., Молявко В. Г., Продайвода Г. Т., Сухорада А. В., Харазян Э. Х. Состав, физические свойства и вопросы петрогенезиса новейших вулканических образований Армении. - Ереван, изд. АН Армянской ССР. 1980, 322 с.

24. Жуков Н. Н. Статистические модели в геохимических исследованиях. "Вопросы прикладной геохимии и петрофизики". - К., 1980, с. 17-31.

25. Жуков Н. Н. Линейные фильтры в задаче обнаружения полиэлементных аномалий. - К., Вестник Киев. ун-та. Прикладная геохимия и петрофизика, вып. 9. - К., 1982, с. 88-92.

26. Жуков Н. Н., Молявко В. Г., Гасанов Ю. Л., Михайлова Е. Н. Принципы систематизации геологической информации для задач сопоставлений и корреляции геологических образований. Вестник Киев. ун-та. Прикладная геохимия и петрофизика, вып. 10. - К., 1983, с. 20-26.

27. Кадурын В. Н., Носырев И. В., Жуков Н. Н. Применение ЭВМ для выделения генераций акцессорных минералов изверженных и метаморфических пород и интерпретация их параметров. Доклады АН УССР, серия геологическая, 1983 г.

28. Жуков Н. Н. Информационно-логическая система фильтрации комплексной геолого-геофизической информации как средство решения задач корреляции геологических образований. Вестник Киевского университета "Прикладная геохимия и петрофизика". - К., 1984, вып. 11, с. 12-17.

29. Жуков Н. Н. Статистический контроль качества спектральных определений содержаний химических элементов в горных породах. Вестник Киевского университета "Прикладная геохимия и петрофизика". - К., 1985, вып. 12, с. 55-61.

30. Жуков Н. Н., Лобанов К. А. Группирование геологических объектов по комплексу признаков в информационной системе "Фильтр". Вестник Киевского университета "Прикладная геохимия и петрофизика". - К., 1986, вып. 13, с. 18-21.

31. Жуков Н. Н., Выжва С. А. Оценка воспроизводимости результатов измерения скоростей упругих волн импульсно-фазовым методом. Вестник Киевского университета "Геология". - К., 1986, вып. 5, с. 94-96.

32. Жуков Н. Н. Информационная система "Фильтр". База данных "Гранитоиды УЩ" (рекомендации). Мин. геологии УССР. - К., 1986, 163 с.

33. Толстой М. И., Чекунов А. В., Щербаков И. Б., Гожик А. П., Гузий Н. И., Есипчук К. Е., Жуков Н. Н., ОРСА В. И., Пашкевич И. К., Проскура Н. И., Санин Е. В., Соловьев И. В., Сухорада А. В. Петрофизика гранитоид-

дов Украинского щита. -К., "Наукова думка", 1987, 238 с.

34. Жуков Н.Н., Серга А.Ю., Проскура Н.И. Первая серия стандартных образцов состава плутонических пород УССР и их использование для повышения качества аналитических работ. Вестник Киевского университета "Прикладная геохимия и петрофизика". -К., 1987, вып. 14, с. 121-131.

35. Жуков Н.Н., Гасанов Ю.Л., Соловьёв И.В. Методологические принципы организации поля качественных геологических признаков в информационно-догической системе "Фильтр". Вестник Киевского университета "Прикладная геохимия и петрофизика". -К., 1988, вып. 15, с. 115-121.

36. Жуков Н.Н., Серга А.Ю., Бондаренко В.Ю., Проскура Н.И. Результаты изучения однородности материала первой серии стандартных образцов горных пород Украины. Вестник Киевского университета "Прикладная геохимия и петрофизика". -К., 1988, вып. 15, с. 127-135.

37. Толстой М.И., Жуков Н.Н. и др. Метод локального прогнозирования оруденения с использованием ЭВМ (на примере Ханчерангинского рудного поля Центрального Забалькалья). Вестник Киевс. ун-та "Прикладная геохимия и петрофизика". -К., 1988, Вып. 15, с. 9-16.

38. Жуков Н.Н. Корреляция геологических признаков с номинальными шкалами регистрации. Вестник Киевского университета "Прикладная геохимия и петрофизика". -К., 1988, вып. 15.

39. Толстой М.И., Жуков Н.Н. и др. Вещественно-физические особенности гранитоидов Криворожской сверхглубокой скважины и ее обрамления. Геологический журнал, 1989, N 2. -К., с. 65-75.

40. Жуков Н.Н., Павлов Г.Г., Пийяр Ю.К. Информационная система "Фильтр". База данных "Интрузивные и ультраметаморфические породы УЩ" (рекомендации). ГлавКТУ "Укргеология". -К., 1989, 108с.

41. Жуков Н.Н. Статистическая обработка геологических данных на программируемых микрокалькуляторах (справочные сведения, тексты программ, примеры применения). Методические рекомендации. Ч. 1. -К., Мингео СССР, Ин-т повыш. квалификации руковод. работников и специалистов. 1989, 24 с.

42. Жуков Н.Н. Статистическая обработка геологических данных на программируемых микрокалькуляторах (справочные сведения, тексты программ, примеры применения). Методические рекомендации. Ч. 2. -К., Мингео СССР, Ин-т повыш. квалификации руковод. работников и специалистов. 1989, 75 с.

43. Жуков Н.Н. Статистическая обработка геологических данных на программируемых микрокалькуляторах (справочные сведения, тексты

программ, примеры применения). Методические рекомендации. Ч. 3. -К., Мингео СССР, Ин-т повыш. квалификации руковод. работников и специалистов. 1989, 55 с.

44. Жуков Н. Н., Павлов Г. Г., Соловицкий В. Н., Пийяр Ю. К. Информационная система геологических данных "Фильтр" с базой данных "Интрузивные и ультраметаморфические породы УЩ". //Проблемы создания универсальных баз данных в геологии. Тезисы докл. научно-техн. конференции, 12-14 июня 1990 г. -К., 1990, с. 31-32.

45. Молявко В. Г., Жуков М. Н., Зінченко О. В., Павлов Г. Г. Досвід організації інформаційної системи для цілей формаційного аналізу магматичних утворень. Вісник Київського університету "Геологія", вип. 12, 1994, с. 57-65.

46. Жуков М. Н. Статистичний аналіз геологічних даних. -К., Випол, 1995. 551 с.

47. Жуков М. Н. Методика застосування інформаційної системи геологічних даних "Фільтр" для побудови кореляційних схем магматичних комплексів. Вісник Київського університету. Геологія, вип. 14, 1996, с. 146-158.

Жуков М. Н. Метод многомерной статистической фильтрации разновидовой геологической информации для решения задач картирования и прогноза. Диссертация на соискание ученой степени доктора геологических наук по специальности 04.00.05 - геоинформатика. Киевский университет им. Тараса Шевченко. Киев, 1997.

Изложены теоретические основы метода многомерной статистической фильтрации геологических данных, регистрируемых в шкалах разных типов - номинальных, порядковых детерминированных, порядковых с фактором неопределенности. Построена корреляционная теория признаков, выражаемых в таких шкалах. Описана информационная система "Фильтр" для IBM PC, являющаяся реализацией метода. Система содержит аналитические оптимизированные процедуры, опробована функционально на задачах прогноза и картирования. Описана методология систематизации геологических данных и систематический указатель признаков (формационных, структурно-тектурных, морфологических, минерогенических, гео- и петрохимических, геофизических), использованных при проектировании региональных баз данных. Приведены примеры применения метода и программных средств: прогноз никеленосности на территории Украинского щита; локальное прогнозирование месторождений олова в пределах Халчерангинского рудного поля; прогноз месторождений и рудопроявлений с оценкой промышленного

значения в дальневосточном Приморье; оценка характера насыщенности терригенных отложений (нефте-, водонасыщенные, неколлекторы) по промыслово-геофизическим данным (Волгоградская область).

Ключевые слова: многомерный анализ геологических данных, статистическая фильтрация, корреляционная теория, систематизация, информационно-аналитическая система, база геологических данных, прогноз месторождений, геологическое картирование.

Zhukov M.N. Multivariate statistical filtration of heterogeneous geological information method for the solution of mapping and prognostication tasks. Thesis presented for a doctor's degree. Speciality 04.00.05 - geoinformatics. Kiev Taras Shevchenko University. Theoretical basis of multivariate statistical filtration of geological data method are stated. These data are registered in different types of the scales: nominal, determined ordinal, ordinal with the indetermination. Correlational theory of signs, expressed in such scales, is created. Informational system "Filter" for IBM compatible PC-method's realization is described. System contains more than 20 analytical optimized procedures, and was functionally tested on mapping and progs. Methodology of geological data systematization and systematic schedule of signs (formational, morphological, mineragenical, geo- and petrochemical, geophysical) are adduced. These signs were used in the projecting of several hod's application and software examples: the prognosis of nickel-bearing on the Ukrainian shield's territory; the local prognosis of tin deposits within the Kharcheganski orefield's limits; prognosis of deposits and ore-disp industrial significance in the Far-Eastern Primorje; the appraisal of deposit's saturation (oil-, water-saturated, noncollectors) by industrially-geophysical data (Volgogradskaja area). Key words : multivariate analysis of geological data, statistical filtration, correlational theory, systematization, informationally-analytical system, geological data base, prognosis of the deposits, geological mapping.

Підп. до друку 2.09.97р. Формат 60х84/16. Друк офс. Папір друк. Друк арк. 2
Тираж 100 екз. Зам. 873. Друкарня Птаженко-Закладні закладів, м.Київ, вул. Лисенка, 6.

436591

AB 37.797