

УДК 553.98:681.3

© И.П. Долинский, А.П. Лобасов

*И.П. Долинский, А.П. Лобасов*

# ПРИМЕНЕНИЕ АЛГОРИТМОВ ПРОСТРАНСТВЕННОЙ КЛАССИФИКАЦИИ И КЛАСТЕРИЗАЦИИ В ЗАДАЧЕ ПРОГНОЗА НЕФТЕГАЗОНОСНОСТИ

Прогнозирование месторождений полезных ископаемых, в частности прогноз нефтегазоносности, – та область геологии, где математические методы исследований впервые начали применяться. Сначала числовая обработка проводилась над фактологическими базами данных определенного целевого назначения. Каждому объекту такой базы соответствовал вектор значений в пространстве параметров. На современном этапе с появлением и развитием геоинформационных технологий появилась возможность оперировать картографическими базами пространственных данных. Учитывая, что практически вся геологическая информация является пространственной, такой подход представляется более адекватным, а следовательно, должен быть более эффективным. Задача прогноза нефтегазоносности, таким образом, сводится к задаче классификации или кластеризации выделенных определенным образом объектов в многомерном пространстве геолого-геофизических параметров. Сами параметры при этом также являются пространственно распределенными.

Среди алгоритмов классификации (кластеризации) различают статистические и эвристические алгоритмы. Условием применения первых является распределение значений параметров на объектах по нормальному закону или другому закону, который можно свести к нормальному путем определенной математической операции над параметрами. На практике это условие часто не удовлетворяется. К тому же часть параметров не являются случайными величинами, то есть они физически обусловлены и применение к их анализу статистических методов вообще выглядит некорректным. В связи с этим эвристические методы исследования, не накладывающие никаких условий на исходные данные, имеют более общее значение. Учитывая также, что и параметры, и объекты являются пространственно распределенными, задача должна решаться средствами геоинформационных технологий в соответствующей программной среде.

## Методика исследования

И задача кластеризации, и задача классификации начинаются с определения многомерного пространства параметров. В терминах геоинформационных технологий определение множества параметров имеет значительные отличия от традиционного «табличного» подхода, где вектор параметров является набором скалярных величин, которые для дискретных объектов в пределах области исследования определяют их свойства, важные с точки зрения целевой задачи.

В геоинформационной системе параметры делятся на две группы. Первая представлена гридами (регулярными числовыми моделями) физических величин, полученными путем аппроксимации значений этих величин на дискретно распределенных в пространстве объектах, полигональных, линейных, точечных. Эта группа параметров соответствует информационному пространству при традиционном подходе. Вторая группа свойственна только для геоинформационного подхода и содержит гриды расстояний от каждого узла регулярной модели к объектам различной природы (в нефтегазовой геологии это тектонические нарушения разного порядка, генезиса и азимута простираения, соляные штоки, оси тектонических структур, другие элементы геологической карты).

Пространственная кластеризация состоит в выделении в геологической среде объектов, однородных в многомерном пространстве параметров. В нашей постановке задачи кластеризации подлечит каждый узел регулярной сети, которая покрывает всю область исследования. Для решения задачи пространственной кластеризации мы используем два эвристических алгоритма: кластеризация «по углу» и кластеризация «по расстоянию» в многомерном пространстве параметров. Неявно предполагается, что параметры независимы, т.е. оси координат в многомерном пространстве ортогональны. Независимость параметров должна обеспечиваться специалистами предметной области на содержательном уровне.

Альтернативная формальная ортогонализация состоит в трансформации исходного пространства параметров в искусственное пространство новых параметров с диагональной корреляционной матрицей. Новые параметры при этом представляются линейными функциями исходных параметров, которые в общем случае не поддаются содержательной геологической интерпретации, что собственно и препятствует широкому практическому применению математической ортогонализации пространства параметров, известной в теории распознавания образов как метод главных компонент или метод факторного анализа.

В рамках кластеризации «по углу» предполагается, что каждому объекту (узлу регулярной точечной сети) соответствует вектор в многомерном пространстве параметров, начало которого находится в центре координат признакового пространства, а координатами конца являются значения параметров на этом объекте. Кластеризация «по углу» состоит в выделении кластеров, объединяющих объекты, векторы которых образуют узкий пучок в пространстве параметров (рис. 1). «Угол» между векторами  $X$  и  $Y$  в многомерном пространстве параметров рассчитывается на базе формулы скалярного произведения двух векторов:

$$\cos(L) = X \cdot Y / |X| \cdot |Y|, \quad (1)$$

$X = \{X_1, X_2, \dots, X_p, \dots, X_n\}$ ;  $Y = \{Y_1, Y_2, \dots, Y_p, \dots, Y_n\}$ ;  
 $n$  – размерность пространства параметров;  
 $|X|$ ,  $|Y|$  – длины векторов  $X$  и  $Y$ .

При этом в один кластер могут попадать объекты с резко отличающимися значениями параметров, но с высокими значениями коэффициентов парной корреляции (т.е. объекты, векторы которых одинаково ориентированы в пространстве параметров).

Цифрами в кружках на рис. 1 показаны номера кластеров, малыми цифрами – номера объектов.

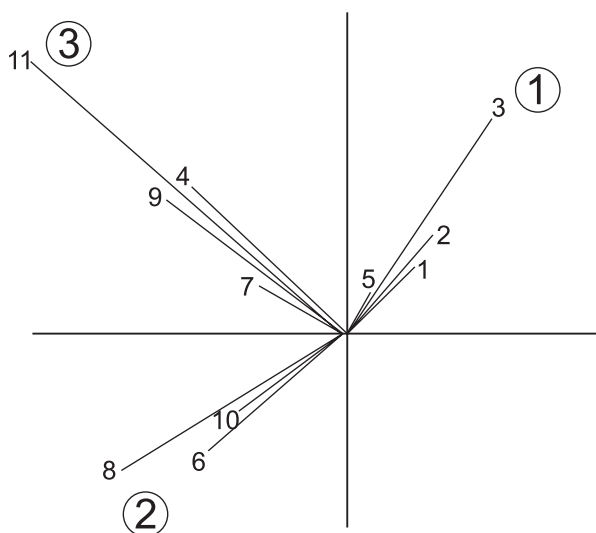


Рис. 1. Кластеризация «по углу» в плоскости двух параметров

В терминах метода кластеризации «по расстоянию» в один кластер объединяются близкие в многомерном пространстве параметров объекты (узлы регулярной точечной сети). Объекты в рамках этого подхода представляются точками в пространстве параметров. Положение объекта в многомерном пространстве параметров определяется значениями параметров на объекте. Мерой близости объектов в многомерном пространстве параметров является евклидово расстояние:

$$\Sigma(X_i - Y_i)^2, \quad (2)$$

$X_i, Y_i$  –  $i$ -е координаты векторов объектов или средних векторов кластеров  $X$  и  $Y$ .

Для контроля качества кластеризации используется расстояние Махаланобиса (известная в математической статистике метрика):

$$|X - Y| / E_x \cdot E_y, \quad (3)$$

$X, Y$  – векторы средних значений пары кластеров;  
 $E_x, E_y$  – ковариационные матрицы этих кластеров. Если в классе один объект, его ковариационная матрица превращается в скаляр равный 1.

В один кластер попадают объекты, которые образуют в пространстве параметров плотную группу (рис. 2).

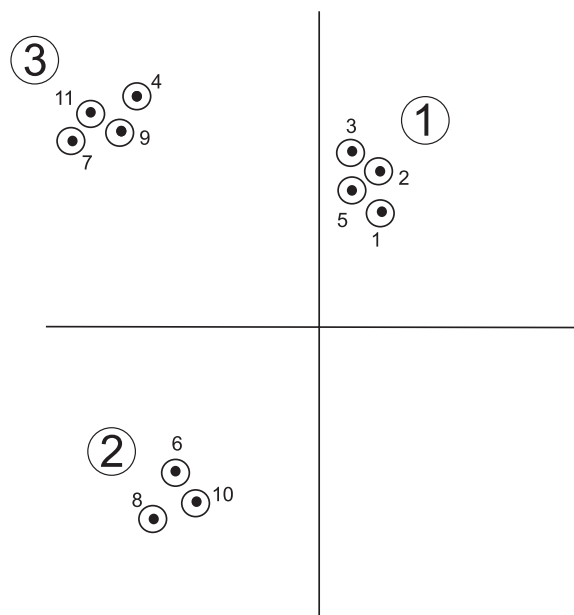


Рис. 2. Кластеризация «по расстоянию» в плоскости двух параметров

Цифрами в кружках показаны номера кластеров, малыми цифрами – номера объектов.

Для реализации процедуры кластеризации применяется известный в теории распознавания образов алгоритм иерархической кластеризации.

Исходное состояние системы – количество классов равно количеству объектов ( $N$ ), которые отождествляются с классами. Для кластеризации

объектов используется итерационный процесс. На  $i$ -й итерации два ближайших в пространстве параметров класса объединяются в один класс, для которого рассчитывается средний вектор, или центр тяжести класса (в зависимости от метода, который применяется). Число классов на  $i$ -й итерации становится равным  $N - i$ . Процесс завершается, когда количество классов становится равным заданному. Если после этого какой-либо класс состоит только из одного объекта, последний считается не кластеризованным.

В методе иерархической кластеризации классы «растут» сразу вокруг нескольких «центров кристаллизации». Основная проблема метода иерархической кластеризации – большие затраты времени на расчеты. На  $i$ -й итерации число операций по определению расстояний (углов) между парой объектов составляет  $(N - i)!$

Альтернативный метод можно назвать методом последовательной кластеризации. Центром «кристаллизации» первого класса выбирается произвольный объект. Количество объектов в нем растет путем присоединения ближайших по расстоянию (или углу) объектов. Процесс наращивания класса продолжается до тех пор, пока расстояние (угол) между центром класса и очередным ближайшим объектом оказывается больше заданного. Этот объект становится центром кристаллизации следующего класса, число объектов в котором наращивается так же, как и в предыдущем. Процесс кластеризации продолжается до полного исчерпания объектов. Так же, как и в методе иерархической кластеризации, классы, которые после завершения процесса состоят только из одного объекта, отбраковываются, а объекты, их составляющие, считаются некластеризованными.

Метод отличается от предыдущего гораздо меньшими затратами времени на вычисления. На  $i$ -й итерации число операций по определению расстояний (углов) между парой объектов составляет  $(N - i)$ , в течение всего процесса –  $N!$  итераций, то есть столько, сколько нужно только на первой итерации в методе иерархической классификации.

Результатом кластеризации в геоинформационной постановке является карта распределения на площади исследования кластеров, выделенных по одному из рассмотренных алгоритмов.

Последним шагом в задаче кластеризации является содержательная интерпретация (определение природы) выделенных кластеров.

Пространственная классификация состоит в выделении в геологической среде объектов с а priori заданными свойствами-параметрами, которые определяются координатами центров тяжести эталонных объектов, расположенных на площади исследования в многомерном пространстве параметров, или

абстрактными эталонными массивами значений параметров, полученных в результате экспертных оценок (в том числе и взятыми из литературных источников). Как и в предыдущей задаче, пространственной классификации подлежит каждый узел регулярной сети точек, которая покрывает всю область исследования, или группы точек, объединенных полигональными или линейными природными объектами. Результатом решения задачи является карта распределения объектов разных классов на площади исследования.

Для реализации задачи пространственной классификации нами используются эвристические алгоритмы классификации «по углу» и классификации «по расстоянию» в многомерном пространстве параметров, аналогичные алгоритмам кластеризации «по углу» и кластеризации «по расстоянию», рассмотренным выше. Класс для узла определяется минимальным углом (минимальным расстоянием) между ним и средним вектором класса из множества эталонных классов.

### Объект исследования

В качестве объекта исследования взят участок площадью 2740 км<sup>2</sup> в пределах южной прибортовой части Днепровско-Донецкой впадины. В пределах участка в отложениях верхневизейского подъяруса каменноугольной системы открыты 10 месторождений, из которых 9 – газоконденсатные и одно – газовое. 43 объекта отнесены к непродуктивным структурам (рис. 3).

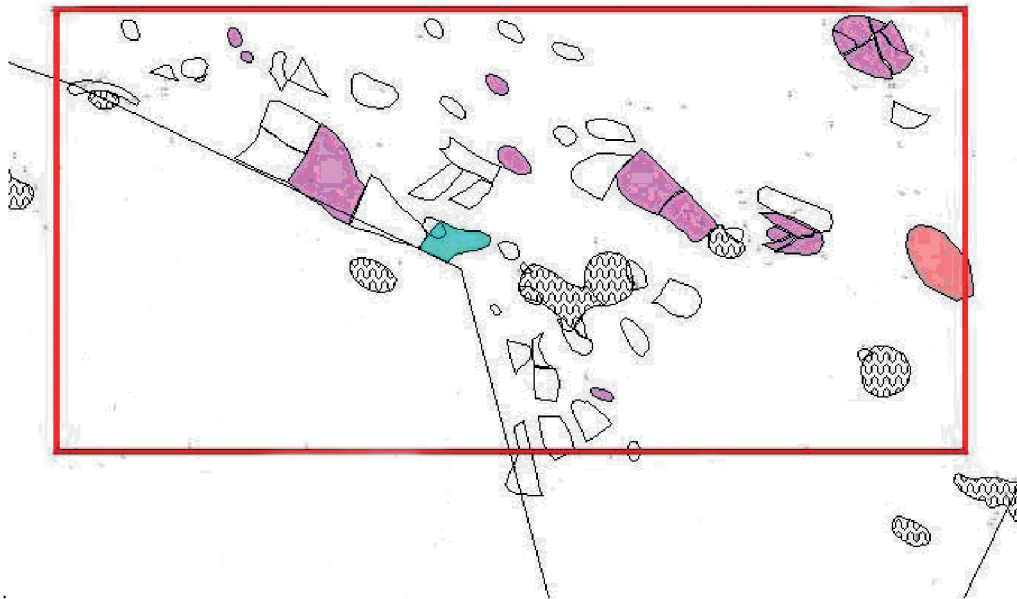
Для апробации алгоритмов в признаковое пространство задачи включены три параметра (пространственное распределение их значений показано на рис. 4-6):

- 1) абсолютные отметки подошвы верхневизейского подъяруса, км (рис. 4);
- 2) локальная составляющая подошвы подъяруса (превышение над региональным фоном), км (рис. 5);
- 3) палеотемпература максимального прогрева толщи подъяруса на протяжении геологической истории, градусы Цельсия (рис. 6).

Карты (регулярные модели), представленные на рис. 4, 5, построены по скважинным данным средствами программного комплекса Geomapping [1]. Карта, представленная на рис. 6, заимствована из работы [2].

Кластеризации и классификации подлежали узлы регулярной точечной сети, покрывающей области исследования (рис. 7, 8, 9, 10). Результаты кластеризации показаны на рис. 7, 8.

Кластер, обозначенный на рис. 7, 8 красным цветом, отвечает продуктивным объектам. В этот



Газовые
  Газоконденсатные
  Непродуктивные структуры
  Соляные штоки

Рис. 3. Распределение месторождений и непродуктивных структур на участке исследования

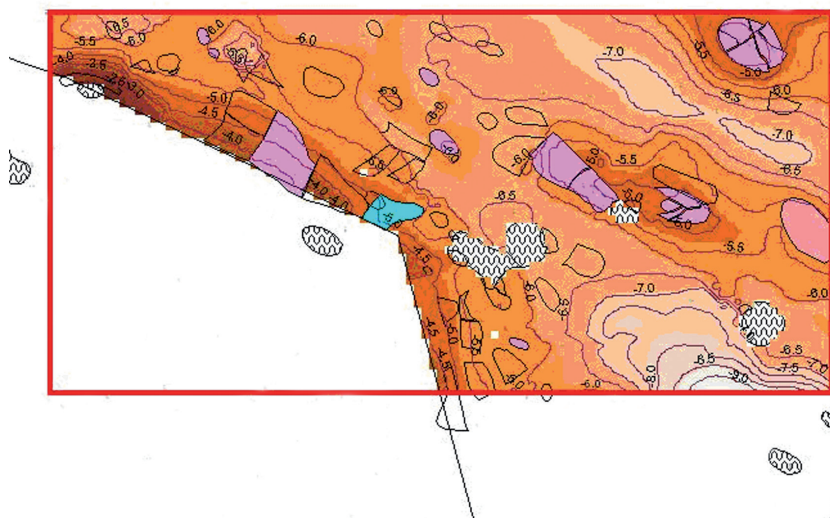
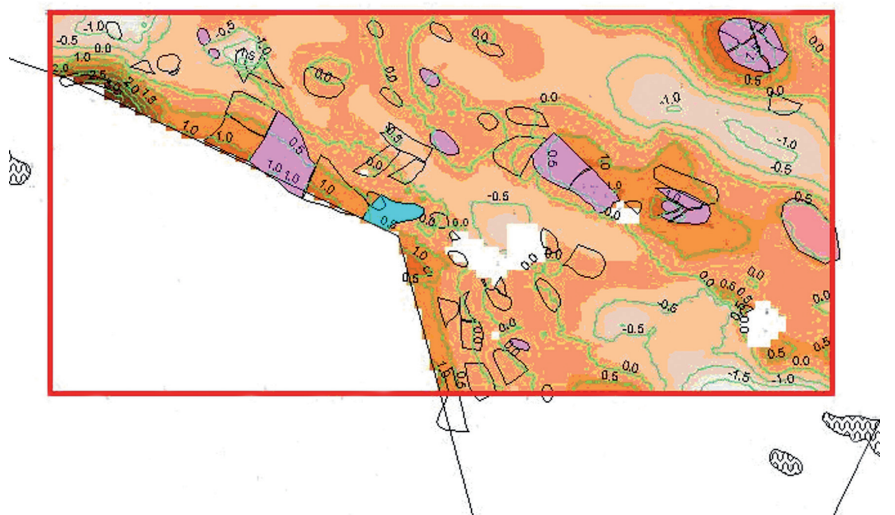


Рис. 4. Структурная карта подошвы верхневизейского подъяруса

Рис. 5. Карта локальной составляющей подошвы верхневизейского подъяруса



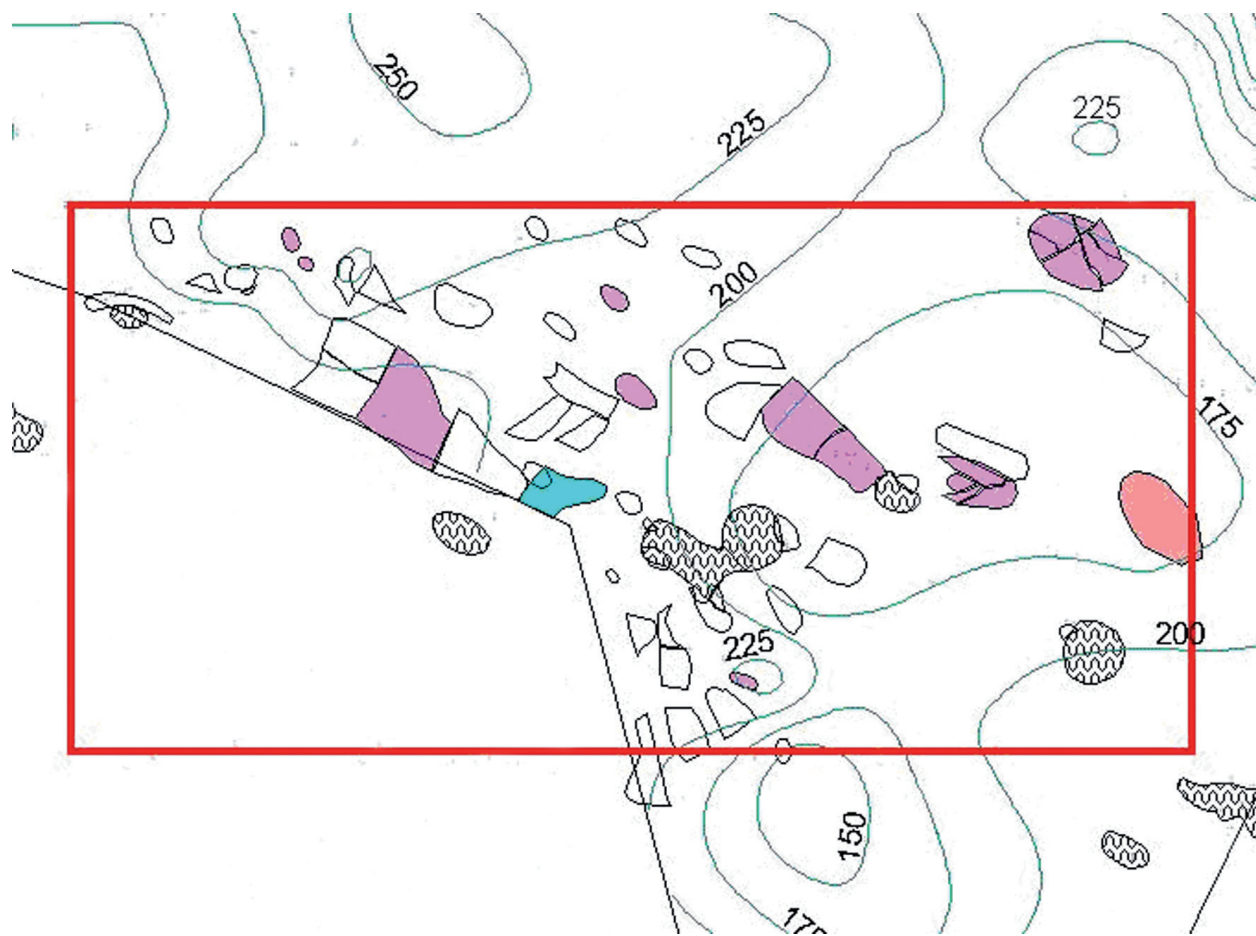


Рис. 6. Максимальные палеотемпературы в толще верхневизейского подъяруса

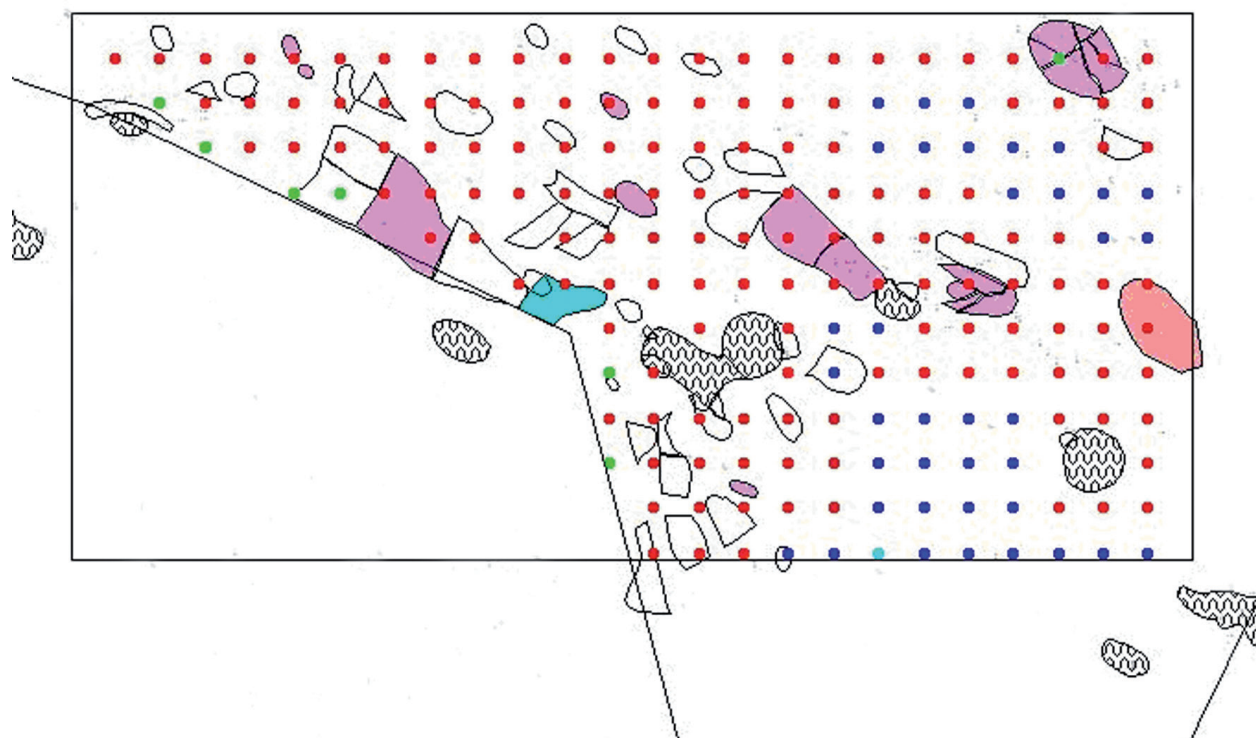


Рис. 7. Результат кластеризации «по углу»

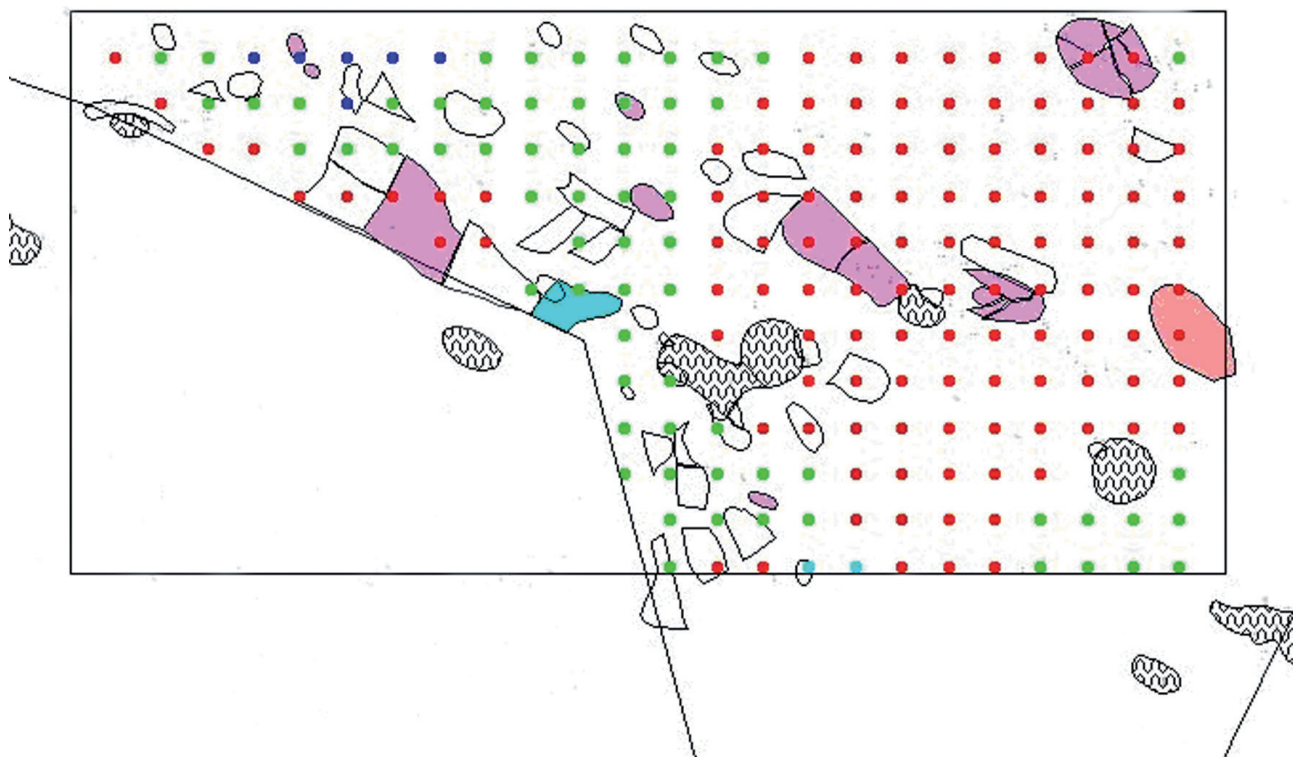


Рис. 8. Результат кластеризации «по расстоянию»



Рис. 9. Результат классификации «по углу»

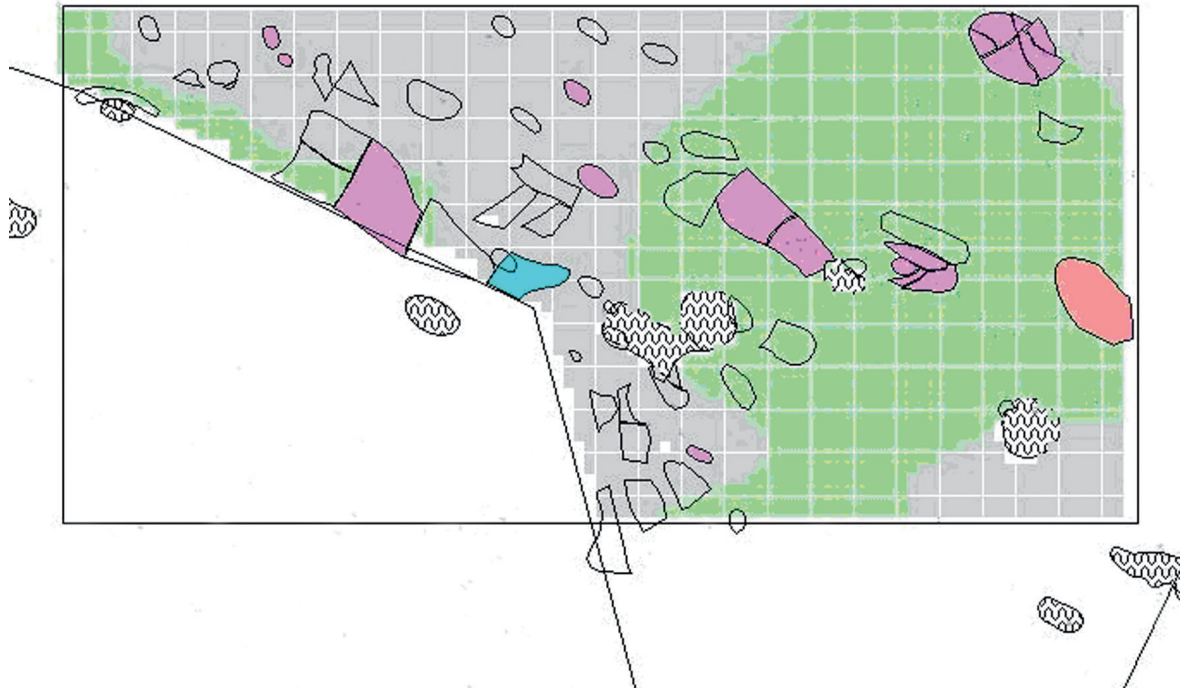


Рис. 10. Результат классификации «по расстоянию»

кластер попадают все месторождения по результатам кластеризации «по углу» и все, кроме одного газового и четырех мелких газоконденсатных, по результатам кластеризации «по расстоянию». Кластеры, обозначенные зеленым и синим цветами на рис. 7, отвечают непродуктивным объектам. На рис. 8 в них попадает несколько мелких месторождений. Районирование территории (рис. 7) можно использовать в первом приближении для поиска продуктивных структур в перспективных (красных) зонах. На рис. 8 в «красную зону» попали более крупные месторождения. Результаты классификации показаны на рис. 9, 10.

По результатам классификации в первом приближении можно сделать два вывода.

1. Образы мелких месторождений в данном признаковом пространстве не отличаются от образов преобладающей части пустых структур.
2. Результаты классификации «по расстоянию» визуальнo оцениваются как более достоверные.

### Выводы

1. Предложен удобный инструментарий для районирования территории исследования и поиска объектов с заданными свойствами на множестве пространственно распределенных параметров. Программный комплекс реализован средствами ГИС-технологий в среде ГИС

ArcView 3.n и может применяться для прогноза нефтегазоносности и решения других задач данного направления.

2. Полученные результаты классификации-кластеризации территории исследования могут стать первым шагом в обосновании необходимости пересмотра геолого-геофизических материалов по структурам, отнесенным к перспективным зонам, которые в настоящий момент считаются бесперспективными. Для повышения достоверности прогноза его нужно выполнить в признаковом пространстве большего объема.

**Ключевые слова:** геоинформационная система, числовые региональные модели, прогноз нефтегазоносности.

### ЛИТЕРАТУРА

1. Гребенников С.Е. Геолого-математическое моделирование и географические информационные системы в задаче мониторинга седиментационных бассейнов / С.Е. Гребенников, А.П. Лобасов // Вестник Киев. нац. ун-та. Геология. – 2001. – Вып. 19. – С. 28-31.
2. Лобасов О.П. Палеотемпературные модели нефтегазоносных горизонтов (на примере нижнекаменноугольных отложений Днепровско-Донецкой впадины) / А.П. Лобасов, П.О. Фенота // Теоретические и прикладные аспекты геоинформатики. – Киев, 2007. – С. 34-38.