

УДК 517.1:550

ДИСКРЕТНЫЙ МАТЕМАТИЧЕСКИЙ АНАЛИЗ И ГЕОЛОГО-ГЕОФИЗИЧЕСКИЕ ПРИЛОЖЕНИЯ

© 2010 А.Д. Гвишиани, С.М. Агаян, Ш.Р. Богоутдинов, А.А. Соловьев

*Геофизический центр РАН, Москва, 11929;
e-mail: sh.bogoutdinov@gcras.ru*

В статье описывается новый оригинальный подход к анализу данных. Называется он «Дискретный математический анализ» (ДМА). Авторами предпринята попытка моделирования конечного предела нечеткими методами. С его помощью по сценарию классической математики найдены свои ответы на основные вопросы в дискретной ситуации: непрерывность, связность, плотность, тренд и др. ДМА представляет собой серию алгоритмов, нацеленных на решение основных задач анализа данных: кластеризацию, трассирование, морфологический анализ, тренд и т. д. Все алгоритмы ДМА носят универсальный характер и базируются на конечном пределе. Отметим, что этот подход к конечному пределу и построению ДМА авторы считают одним из возможных. В данной статье описаны три алгоритма кластеризации — «Кристалл», «Роден», «Монолит» — и приведены их некоторые геофизические приложения.

Ключевые слова: дискретный математический анализ, алгоритмы, нечеткие методы, кластеризация.

ВВЕДЕНИЕ

Прогресс в развитии компьютерной техники, появление Интернета привели к большим объемам геолого-геофизических данных, геоинформационный анализ которых является важным направлением в современных геологии и геофизике. Причем происходило это одновременно с переходом от аналоговых представлений результатов к цифровым, вследствие чего возросла роль дискретных методов анализа данных и дискретной математики в целом.

Одно из направлений развития дискретной математики связано с умением человека анализировать данные. Действительно, человек прекрасно выделяет аномалии, кластеризует и трассирует их, находит сигналы на записях и делает многое другое. Объясняется это более естественным и устойчивым, по сравнению с математическим, характером восприятия им свойств дискретности и стохастичности. Приведем три примера.

1. Гладкая в математическом смысле функция f на отрезке $[a, b]$ после даже достаточно тщательной дискретизации $[a, b]$ либо под воздействием небольшого стохастического возму-

щения $\varepsilon: f \rightarrow f + \varepsilon$ потеряет это свойство, но по-прежнему останется гладкой для человека;

2. Математическую монотонность f на $[a, b]$ может нарушить любое сингулярное возмущение, в то время как человеческое восприятие тренда более устойчиво к нему. Лишь достаточно «большое» возмущение заставит человека изменить свое решение о монотонности f на $[a, b]$.

3. В многомерном конечном массиве X любой, в частности геолого-геофизической, природы особую роль реперных точек играют наиболее «плотные» из них, сильнее всего концентрирующие X вокруг себя. Они важны для анализа X , например, при кластеризации или трассировании в нем. Нетривиальное формальное выражение плотности в X не может быть построено в рамках классической математики, потому что для нее X — дискретное пространство, все точки которого одинаково изолированы и неинтересны.

Но человек бессилен перед большими объемами и большими размерностями данных. Авторы поставили задачу смоделировать умение человека анализировать данные, тем самым передав его компьютеру для работы в больших размерностях и с большими объемами. В таком моделировании необходимо учитывать то об-

стоятельство, что человек мыслит не числами, а нечеткими понятиями. Еще Норберт Винер (1966) отмечал: «По-видимому, главное преимущество человека перед компьютером — это его способность оперировать с нечетко очерченными понятиями». Именно для этих целей американский математик Лотфи Заде создал теорию нечетких множеств (Zadeh, 1965). Основанные на ней нечеткая математика и нечеткая логика обладают достаточно большими возможностями для моделирования человеческих представлений и рассуждений по сравнению с обычными множествами и булевой логикой, и потому именно они послужили технической основой нашего подхода. Нечеткая математика и нечеткая логика позволяют дать строгое математическое описание в действительности расплывчатых экспертных утверждений и преодолеть лингвистический барьер между человеком, суждения и оценки которого являются приближенными, качественными и нечеткими, и компьютером, который может выполнять только четкие инструкции.

Есть еще одно обстоятельство, делающее нечеткую математику привлекательной для геологии и геофизики: нечеткая математика является формальным аппаратом, способным наиболее адекватно учесть не только мнения экспертов, но и нечеткость данных. Действительно, в геологии и геофизике мы почти всегда имеем дело с приближенными величинами. Это свидетельствует о естественности нечеткого подхода к геологии и геофизике, поскольку информации в них изначально присущ расплывчатый характер.

Авторами разработан оригинальный подход к анализу данных. Называется он «Дискретный математический анализ» (ДМА). Для его понимания необходимо обратиться к стандартному математическому анализу: «Дискретный анализ — область математики, занимающаяся изучением свойств структур конечного характера. Дискретный анализ представляет собой важную направление в математике, имеющее характерные для него предмет исследования, методы и задачи, специфика которых обусловлена, в первую очередь, необходимостью отказа в дискретном анализе от основополагающих понятий классической математики — предела и непрерывности — и (в связи с этим) тем, что для многих задач дискретного анализа сильные средства классической математики оказываются, как правило, малопримемыми» (Математическая..., 1979, с. 207).

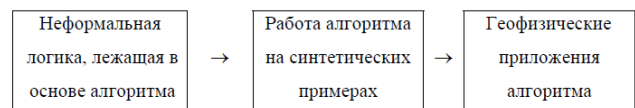
Авторами была предпринята попытка исправить это положение путем моделирования конечного предела нечеткими методами. С его помощью по сценарию классической математики найдены ответы на основные вопросы в дис-

кретной ситуации: непрерывность, связность, плотность, тренд и др. ДМА представляет собой серию алгоритмов, нацеленных на решение основных задач анализа данных: кластеризацию, трассирование, морфологический анализ, тренд и так далее. Все алгоритмы ДМА носят универсальный характер и базируются на конечном пределе. Отметим, что этот предлагаемый подход к конечному пределу и построению ДМА авторы считают одним из возможных.

На рис. 1 представлена схема ДМА. Первые три блока относятся к формальным основам ДМА и завершаются определением конечного предела. С их помощью строго определяются все объекты ДМА, в частности, кластеры, трассы в многомерных массивах, которые ищут алгоритмы, представленные слева. Остальные ветви относятся к конечным временным рядам и реализуют в серии алгоритмов функциональный подход к ним, напоминающий классический математический анализ гладких функций. Именно этим объясняется название представленного подхода. Для рядов строятся дискретные аналоги гладкости, разрывности, монотонности, экстремума, выпуклости, изучается их геометрия.

Таким образом, ДМА — подход к изучению многомерных массивов и временных рядов, базирующийся на моделировании предела в конечной ситуации и реализованный в серии алгоритмов. Основой построения конечного предела послужил более устойчивый, по сравнению с математическим, характер восприятия человеком дискретности и стохастичности. Другими словами, то, что академик Кейлис-Борок (1968, с. 322) называл моделированием «на глаз»: решает не математика, а человек, и его решение нужно формально выразить. Нечеткая математика и нечеткая логика обладают достаточными возможностями для моделирования человеческих представлений и рассуждений, и потому именно они послужили технической основой ДМА.

Алгоритмы ДМА. В статье мы по возможности будем придерживаться следующей схемы:



Еще раз отметим, что все встречающиеся далее понятия имеют в рамках ДМА строгое формальное выражение. Без этого алгоритмы невозможны.

НЕЧЕТКИЕ СРАВНЕНИЯ

Определение. Нечеткое сравнение $n(a,b)$ на действительных числах a и b измеряет в знаковой переменной шкале отрезка $[-1,1]$ степень превосходства b над a :

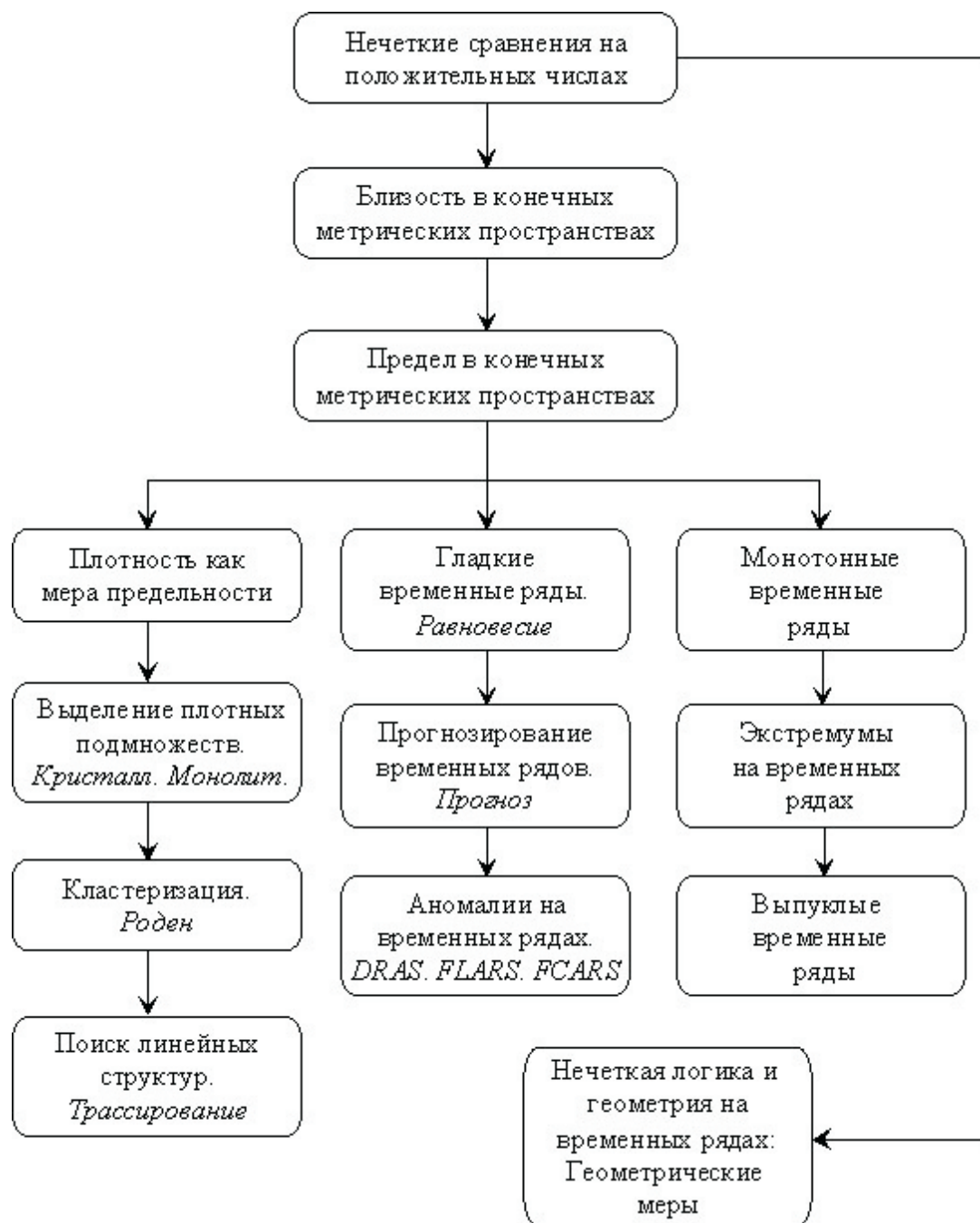


Рис. 1. Схема ДМА

$$n(a, b) = \mu_{es}(a < b) \in [-1, 1]. \quad (1)$$

Широкое применение нашел вариант базового сравнения $n_{\gamma, \nu}(a, b)$, $\nu > 0$, определенный на неотрицательных числах, а также его вариации специального вида $n_{\gamma, \nu}(a, b)$.

Определение.

1) Если $a, b \in \mathbb{R}^+$, то

$$n_{0, \nu}(a, b) = \frac{b - a}{(a^\nu + b^\nu)^{1/\nu}}.$$

2) Для $\gamma \in (-1, 1)$ положим

$$n_{\gamma, \nu}(a, b) = \psi_\gamma(n_{0, \nu}(a, b)),$$

$$\text{где } \psi_\gamma(t) = \begin{cases} \frac{t - \gamma}{1 - \gamma}, & t \in [\gamma, 1] \\ \frac{t - \gamma}{1 + \gamma}, & t \in [-1, \gamma]. \end{cases}$$

Заметим, что в общем случае сравнение n не-кососимметрично: $n(a, b) \neq -n(b, a)$ и допускает расширение до сравнения неотрицательного числа $b > 0$ и произвольной неотрицательной совокупности $A = \{0 \leq a_1 \leq a_2 \leq \dots \leq a_N\}$. Были построены три конструкции такого сравнения (бинарное, гравитационное и сигма-конструкция).

Пример (бинарная конструкция):

$$n(A, b) = \frac{\sum_i n(a_i, b)}{N} \in [-1, 1]$$

$$n(b, A) = \frac{\sum_i n(b, a_i)}{N} \in [-1, 1]$$

Нечеткие сравнения дают возможность формализовать нечеткие понятия «маленький» и «большой». Совершенно естественно считать, что если некоторое свойство выражено в шкале – $[-1, 1]$, то попадание в отрезок $[\frac{1}{2}, 1]$ $[[0, \frac{1}{2}]$ означает сильно экстремальное (слабо экстремальное) его проявление. Аналогично попадание в отрезок $[-1, -\frac{1}{2}]$ $[[-\frac{1}{2}, 0]$ означает сильное (слабое) отсутствие такого свойства.

МЕРЫ БЛИЗОСТИ В КОНЕЧНЫХ МЕТРИЧЕСКИХ ПРОСТРАНСТВАХ (КМП)

Близость точек. Зная расстояние $d(x, y)$ между точками в КМП (X, d) , за исключением тривиального случая $d(x, y) = 0$, невозможно ответить на вопрос: «В какой степени точка y близка или далека от точки x ?». Для этого нужен глобальный взгляд на X , а именно сравнение $d(x, y)$ с остальными расстояниями $d(\bar{x}, \bar{y})$.

Предложено несколько вариантов ответа на этот вопрос, но везде «близость» y к x реализована как нечеткая мера $\delta_x(y) \in [0, 1]$ (или $[-1, 1]$) с нормировкой $\delta_x(x) = 1$. В общем случае $\delta_x(y)$ не только убывает с ростом $d(x, y)$, но и зависит от топологического распределения X вокруг x . Приведем несколько конструкций $\delta_x(y)$. Пусть n – нечеткое сравнение (1)

1. Если $dX = \{d(\bar{x}, \bar{y}) : \bar{x} \neq \bar{y} \in X\}$ – совокупность всех нетривиальных расстояний в (X, d) , то $\delta_x(y)$ есть мера малости $d(x, y)$ по сравнению с dX

$$\delta_x(y) = \mu_{es}(d(x, y) < dX) = n(d(x, y), dX) = \frac{\sum_{\bar{x} \neq \bar{y}} n(d(x, y), d(\bar{x}, \bar{y}))}{|X|(|X| - 1)}$$

2. Замена dX на $dX(x) = \{d(x, \bar{y}) : \bar{y} \in X - x\}$ – совокупность расстояний от x до любых точек из X приводит ко второй конструкции близости

$$\delta_x(y) = \mu_{es}(d(x, y) < dX(x)) = n(d(x, y), dX(x)) = \frac{\sum_{\bar{y} \neq x} n(d(x, y), d(x, \bar{y}))}{|X| - 1}$$

Эта конструкция в отличие от первой, вообще говоря, несимметрична – $\delta_x(y) \neq \delta_y(x)$.

3. Наиболее тесно связана с исходным расстоянием $d(x, y)$ следующая симметричная конструкция. Пусть φ – убывающая на полуоси $[0, \infty)$ неотрицательная функция «потенциального типа»: $\varphi(0) = 1$, $\varphi(t_1) > \varphi(t_2)$ при $t_1 < t_2$. С ней

$$\delta_x^{\varphi}(y) = \delta_x(y)$$

$$\delta_x(y) = \varphi(d(x, y)) \quad \forall y \in X.$$

Близость точки к множеству лежит в основе поиска плотных областей в КМП. Были построены две конструкции такой близости: нечеткая принадлежность (плотность) P и монолитность топ. Пусть $x \in X$ и A – подмножество в X .

Нечеткая принадлежность $P_A(x)$: пусть δ – мера близости на X . Значение $\delta_y(x)$ можно трактовать как согласованную с метрикой d нечеткую меру принадлежности x к u в X . Пользуясь набором $\{\delta_y(x) : y \in X\}$, построим меру $P_A(x)$ нечеткой принадлежности x к A :

$$P_A(x) = \frac{\sum \delta_y(x) : y \in A}{|A|} \quad (2)$$

Монолитность $\text{mon}_A(x)$: пусть $0 < r_1 < r_2 < \dots < r_{m(x,A)}$ – упорядоченная совокупность расстояний от точек x до отличных от нее элементов множества A . Числовую совокупность $\{r_i | 1 \leq i \leq m(x,A)\}$ обозначим через $\Gamma A(x)$.

Определение.

1) Пусть h – квант, $h > 0$ и $\mathbb{R}_h^i = \{kh, k = 0, 1, \dots\}$.

Назовем квантованием соответствие

$$t \rightarrow h(t) : \mathbb{R}^+ \rightarrow \mathbb{R}_h^+$$

$$h(t) = \begin{cases} t, & \text{если } t = mh \\ (m+1)h, & \text{если } t \in (mh, (m+1)h) \end{cases}$$

2) Если $D_A(x, r) = \{y \in A : d(x, y) \leq r\}$ – шар в A с центром в x радиуса r и $\psi_r(mh) = 1 - \frac{mh}{h(r) + h}$, то выражение

$$\text{mon}_A(x|r) = \frac{\sum \psi_r(h(r_i)) : r_i \in \Gamma A(x), r_i \leq r}{\sum \psi_r(mh) : mh \in [0, h(r)]} \quad (3)$$

назовем локальной r – монолитностью A в x .

3) Если $r = r_{m(x,A)}$, то $\text{mon}_A(x|r_{m(x,A)})$ считается глобальной монолитностью A в x и обозначается через $\text{mon}_A(x)$.

Монолитность можно считать мерой неизолированности x от A . При этом принадлежность x к A не имеет решающего значения: большие монолитности $\text{mon}_A(x|r)$ могут достигаться для $x \notin A$, и наоборот, если $x \in A$, то монолитность $\text{mon}_A(x|r)$ может быть достаточно маленькой.

АЛГОРИТМ «КРИСТАЛЛ»

Алгоритм «Кристалл» (Агаян, Соловьев, 2004) предназначен для поиска областей повышенной плотности в конечных метрических пространствах. Предлагается стратегия объединения сильнейших, наиболее «плотных» точек во множестве X , которая может быть интерпретирована как своего рода «кристаллизация» в X . «Кристалл» имеет две фазы: «Выбор основ»

и «Кристаллизация». В первой из них ищутся «основы» кристаллизаций – наиболее «плотные» точки в исходном массиве, вокруг которых он группируется сильнее всего. Произвольное плотное подмножество не всегда обязано состоять только из основ. Полное его заполнение происходит во время второй фазы, а именно кристаллизации вокруг основ.

Формализация процедуры кристаллизации основана на «оптическом» подходе, суть которого состоит в следующем. Представим себе, что в каждой точке исходного пространства имеется источник света единичной силы. Тогда каждая точка будет освещать остальные точки пространства с различной силой, зависящей от расстояния между соответствующими точками.

Для точки $x \in X$ и произвольного подмножества $A \subset X$ положим

$$A_x \stackrel{\text{def}}{=} \begin{cases} A, & \text{если } x \notin A \\ A - x, & \text{если } x \in A \end{cases}; \quad |A|_x \stackrel{\text{def}}{=} \begin{cases} |A|, & \text{если } x \notin A \\ |A| - 1, & \text{если } x \in A \end{cases}.$$

Определение. Назовем внешним освещением точки x подмножеством A сумму $O_A(x)$:

$$O_A(x) = \sum_{y \in A_x} \delta_y(x).$$

Определение. Назовем плотностью освещения $P_A(x)$ точки $x \in X$ подмножеством $A \subset X$ среднее внешнего освещения точки x по элементам подмножества A :

$$P_A(x) = \frac{O_A(x)}{|A|_x}.$$

При этом $P_x(x) = 0$ (плотность внешнего освещения точки x точкой x равна нулю).

Определение. Пусть $A \subset X$. Обозначим через $P(A)$ и назовем его качеством (плотностью подмножества) единицу, если A – точка и минимум плотностей освещения им своих точек для не тривиального A :

$$P(A) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{если } |A| = 1 \\ \min_{x \in A} P_A(x), & \text{если } |A| > 1 \end{cases}.$$

Определение. Подмножество $A \subset X$ плотно на фоне X , если

$$P_A(x) \geq P_X(x), \quad \forall x \in A.$$

Определение. Подмножество $A \subset X$ – α -плотно на фоне X , если

$$P_A(x) \geq \alpha P_X(x), \quad \forall x \in A, \alpha \geq 1.$$

Целью алгоритма «Кристалл» является поиск α -плотных подмножеств на фоне исходного множества (Агаян, Соловьев, 2004).

АЛГОРИТМ «РОДЕН»

Несмотря на широкое применение кластер-анализа, общепринятого определения кластера

не существует. Ниже приводится одно из возможных решений этой задачи, а именно дается конструктивное формальное определение кластера на основе конструкции плотности $P(2)$.

Определение. Подмножество A в X назовем кластером относительно плотности $P(P$ -кластером), если

$$P_A(x) \geq P_B(x) \quad \forall x \in A, \forall B: A \subset B \subseteq X.$$

Утверждение. Если $P_A(x)$ – линейная плотность, индуцированная мерой близости δ , то кластерность A эквивалентна следующему свойству:

$$P_A(x) \geq \delta_y(x) \quad \forall x \in A, \forall y \in X - A. \quad (4)$$

Если интерпретировать $\delta_y(x)$ и $P_A(x)$, как притяжение точки x соответственно точкой y и подмножеством A , то приведенное утверждение означает, что кластер притягивает свои точки сильнее, чем это делает любая внешняя по отношению к нему точка. Кластерность есть сочетание плотности и отделимости, поэтому численное выражение плотности и отделимости дадут численное выражение кластерности, необходимое для ее поиска.

Определение. Пусть A – подмножество в X , P – конструкция плотности на X .

1. Плотность A

$$P(A) = \min_{x \in A} P_A(x).$$

2. Отделимость (кластерность) A от своего дополнения \bar{A} в точке $x \in A$

$$Cl_A(x) = \frac{P_A(x)}{E_A(x)}, \quad \text{где } E_A(x) = \max_{y \in \bar{A}} \delta_y(x).$$

3. Отделимость (кластерность) A в X :

$$Cl(A) = \min_{x \in A} Cl_A(x).$$

Поиск кластеров осуществляет алгоритм «Роден» (Гвишиани и др., 2002; Mikhailov et al., 2003). Он базируется на двух вещах – конструктивности определения кластера (4) и принципе высекания. Понимая, что такое кластер заданного качества и отделимости, «Роден» ищет его в исходном пространстве X , отсекая от него все лишнее. Этим объясняется название алгоритма. Остановимся подробнее на простейшем, глобальном варианте алгоритма «Роден».

Сначала задаются характеристики высекаемого кластера: уровень качества α и уровень отделимости β . Обозначим через K_n текущую версию кластера, получившуюся из X после высекания из него n точек $\{x_i\}_1^n$: $K_n = X - \{x_1, \dots, x_n\}$. Начальной версией K_0 считаются все: $K_0 = X$. Высекание из K_n точки x_{n+1} может происходить по двум причинам.

Во-первых, по причине неудовлетворительного качества $P(K_n) < \alpha$, тогда x_{n+1} – слабейшая по плотности точка K_n : $x_{n+1} = \arg \min P_{K_n}(x)$.

Во-вторых, по причине неудовлетворительной отделимости $Cl(K_n) < \beta$, и тогда x_{n+1} – слабая по отделимости точка K_n : $x_{n+1} = \arg \min Cl_{K_n}(x)$

Алгоритм «Роден» завершает работу при встрече первого попавшегося кластера с заданными характеристиками. Вполне возможно, что внутри него может быть еще один кластер более высокого качества $\bar{\alpha} > \alpha$. «Роден» выделит его, если повысить начальный уровень плотности до $\bar{\alpha}$ и скорректировать $\beta \rightarrow \bar{\beta}$. Так, настраивая алгоритм «Роден» на тот или иной уровень характеристик, можно диагностировать КМП(X, d) на предмет наличия в нем сгущений разного «качества».

АЛГОРИТМ «МОНОЛИТ»

Алгоритм «Монолит» (Богоутдинов, 2006) для произвольного подмножества A в КМП X строит оболочки монолитности $A^{(i)}$ на основе конструкции $\text{mon}_A(x|r)$ (3). Определенную во всех точках пространства X функцию $\text{mon}_A(x|r)$ можно трактовать как функцию принадлежности на X для нечеткого свойства «предельности X к A ».

Через $\mathfrak{Mon}_A(X|r)$ обозначается подмножество тех точек из X , в которых это свойство проявлено (выражено) в значительной степени $\mathfrak{Mon}_A(X|r) = \{x \in X : \text{mon}_A(x|r) - \text{немаленькая по модулю } \text{Im } \text{mon}_A(\cdot|r)\}$.

Благодаря определенным выше нечетким сравнениям это равенство имеет точный формальный смысл.

При переходе $A \rightarrow A^{(1)} \stackrel{\text{def}}{=} \mathfrak{Mon}_A(X|r)$ убираются изолированные точки из A , и наоборот – присоединяются к A ей не принадлежащие, но достаточно им окруженные. Другими словами, при переходе $A \rightarrow A^{(1)}$ одновременно происходят два процесса: топологическое «сглаживание» и топологическая «фильтрация» A .

Первая итерация $A^{(1)}$ называется монолитной оболочкой A и является одним из вероятных вариантов дискретного аналога топологического замыкания A в X . Далее возможны последующие (высшие) итерации:

$$A^{(i+1)} = \mathfrak{Mon}_{A^{(i)}}(X|r).$$

Алгоритм «Монолит» начинает свою работу с построения радиуса локализации $r=r(X)$, как решения функционального уравнения

$$n(r, dX) = 0.5, \quad dX = \{d(\bar{x}, \bar{y}) : \bar{x} \neq \bar{y} \in X\}$$

на основе нечеткого сравнения n (1). Радиус $r(X)$ формализует понятие близко в пространстве X . Далее выбираются остальные параметры монолитности: квант h и вес ψ . Алгоритм «Монолит» заключается в построении подпространства $A^{(i)}$ для некоторого заданного i .

Обычно множество A получается как тот

или иной выбор на X относительно какой-либо функции $f : X \rightarrow \mathbb{R}^+$.

В алгоритме «Монолит», ориентированном на поиск сгущений и кластеризацию, функция f является степенью близости (степенью предельности) пространства X к произвольной своей точке: $f(x) = P_X(x)$ или $f(x) = \text{mon}_X(x|r)$.

Множество A в данном случае будет множеством основ пространства X , то есть точек, вокруг которых X концентрируется сильнее всего. Соответственно

$$A = \{x \in X : n(P_X(X), P_X(x)) \geq 0\} \quad \text{и}$$

$$A = \{x \in X : n(\mathfrak{Mon}_X(X|r), \text{mon}_X(x)) \geq 0\}.$$

Массивные сгущения в X обязаны содержать достаточное количество основ, а при переходе $A \rightarrow A^{(1)}$ в данном случае одновременно происходит освобождение в X от изолированных основ и своего рода кристаллизация вокруг массивных их скоплений, что в результате приводит к массивным сгущениям в X .

ГЕОФИЗИЧЕСКИЕ ПРИЛОЖЕНИЯ

Плодотворные геофизические приложения алгоритмов «Кристалл» и «Роден», связаны с методом деконволюции Эйлера (МДЭ). Этот известный в практике метод основан на аппроксимации измеренного аномального гравитационного или магнитного поля (ΔT) в скользящем «окне» полем некоторого элементарного источника однородной плотности или намагниченности, аномалия от которого является однородной по Эйлеру функцией (Hood, 1965; Thompson, 1982; Reid et al., 1990).

В дальнейшем (x, y, z) – декартовы координаты, ось Oz направлена вертикально вниз, ось Ox – на север, ось Oy – на восток. Параметр $N = -n$ принято называть структурным индексом. Известно, что уравнение Эйлера на координаты (x_0, y_0, z_0) элементарного источника имеет вид

$$(x - x_0) \frac{\partial \Delta T}{\partial x} + (y - y_0) \frac{\partial \Delta T}{\partial y} + (z - z_0) \frac{\partial \Delta T}{\partial z} = N(A - \Delta T(x, y, z)),$$

где (x, y, z) – координаты точки наблюдения, в которой задано поле ΔT и вычислены его производные, A – фоновое значение поля.

МДЭ заключается в определении четырех неизвестных параметров (x_0, y_0, z_0) и A в скользящем «окне», содержащем более четырех точек, путем решения системы линейных уравнений. Таким образом, каждому «окну» соответствует одно эйлерово решение (эйлерова точка) и одно фоновое значение A .

Эйлеровы точки для изолированных аномалеобразующих тел (АОТ) располагаются по их контурам и дают оценку глубины их залегания,

но в сложных интерференционных случаях эйлеровы точки образуют размытые облака, которые трудно интерпретировать. Тем не менее оказывается, что в этих облаках имеются плотные области, которые располагаются вблизи контуров тел, поскольку, благодаря небольшому размеру палетки и ее перемещению с перекрытием, одно и то же АОТ будет оказывать доминирующее влияние в поле при нескольких соседних положениях палетки. Следовательно, к этому АОТ будут относиться несколько решений инверсии – они должны быть близки и поэтому образовывать плотные скопления. Таким образом, имеет место эквиваленция: чем четче будут выделены в эйлеровых решениях плотные области, тем лучше будет известна топология АОТ.

Алгоритм «Кристалл» использовался совместно с МДЭ в интерпретации участка аэромагнитной карты аномального поля (ΔT) a в районе массива Ахаггар (Хоггар) вблизи южной границы Алжира (рис. 2). Аэромагнитные измерения в этом районе были проведены на высоте 300–350 м в 1974 г. Для обработки данные были переинтерполированы на равномерную сетку с шагом 1 км. В результате была получена сетка размерностью 401 точка по оси Ox ($\Delta x = 1$ км), и 271 точка по оси Oy ($\Delta y = 1$ км). При этом ряд узлов не обеспечен данными.

МДЭ применялся к данному полю дважды (Соловьев и др., 2005): со структурными индексами $N = 3$ и $N = 0.5$. Рассмотрим каждый из них.

Известно, что при $N = 3$ МДЭ порождает решение, соответствующее эквивалентному точечному диполю, поле которого наилучшим образом аппроксимирует наблюдаемые значения аномалий в скользящем окне. Для локализованных магнитных объектов положение такого диполя приблизительно совпадает с центром магнитных масс.

При $N = 3$ эйлеровы решения определялись с помощью квадратного окна 5×5 точек (5×5 км) со сдвигом в две точки и со структурным индексом $N = 3$ (точечный диполь). В результате получилось 17153 эйлеровых решения. Число их было уменьшено до 2719 путем стандартной фильтрации, принятой в МДЭ. Проекция этих решений на горизонтальную плоскость xOy представлены на рис. 3а.

Несмотря на то, что более 84% точек было отброшено, результат такой фильтрации нельзя было напрямую интерпретировать в терминах геологического строения. Алгоритм «Кристалл» применялся к оставшимся эйлеровым решениям. На рис. 3б приведен результат его работы: из исходного набора в 2719 решений алгоритм отобрал 669 точек, сформировав 38 плотных сгущений.

Далее для каждого K -го сгущения ($K=1...38$) оценивался его центр тяжести $gr(K)$ и средний

магнитный момент $\tilde{m}(K)$, а также их разбросы.

Оказалось, что полученные результаты по направлению вектора суммарной намагниченности хорошо согласуются с представлениями о тектонике региона (рис. 4). Так, было показано, что сходные в тектоническом отношении участки, входящие в состав массива Ахаггар, характеризуются общим для каждого из них направлением намагниченности пород.

При структурном индексе $N = 0.5$ МДЭ порождает решения, соответствующие положению ближайших к поверхности особых точек аномального поля, связанных с положением изломов верхней кромки. В этом случае эйлеровы решения определялись с помощью квадратного окна размером 9×9 точек (9×9 км) со сдвигом 3 точки. Выбранная сетка породила 7319 эйлеровых решений. Число их было уменьшено до 7000 путем фильтрации полученных решений. Проекция этих решений на горизонтальную плоскость (xOy) показаны на рис. 5а. «Кристалл» отобрал из них 708 точек, сформировав 58 сгущений (рис. 5б).

В условиях древних платформ, к которым относится изучаемый район, аномалеобразующие тела залегают в кристаллическом фундаменте, а перекрывающая фундамент осадочная толща практически немагнитна. Верхние кромки намагниченных объектов, как правило, выходят на поверхность кристаллического фундамента, поэтому полученные оценки глубин сгущений могут рассматриваться как оценки глубин кристаллического фундамента. На рис. 6 изображена карта прогнозных изоглубин кристаллического основания, полученная путем линейной интерполяции средних глубин сгущений. Построенная таким образом карта хорошо согласуется с имеющимися ранее данными о глубинах фундамента и позволила детализировать представление о них.

Алгоритм «Роден» был применен при интерпретации участка (рис. 7, 8) аэромагнитной карты аномального поля (ΔT) a Армориканского массива в Бретани, Франция (Гвишиани и др., 2002; Mikhailov et al., 2003). Выбранный участок (рис. 7) включает восточную часть залива Сен-Мало и часть суши к югу от залива. Важной особенностью строения данного района является система долеритовых даек (рис. 7), которые внедрились в докембрийский фундамент и были позднее срезаны и метаморфизованы герцинскими гранитными интрузиями. В прибрежной зоне дайки имеют небольшую мощность (в среднем около 5 м), их простирание меняется от меридианального до ССВ-ЮЮЗ. Линейные магнитные аномалии в этом районе имеют такое же простирание, но значительно большую ширину, что, возможно, указывает на то, что дайки на глубине имеют большую мощность.

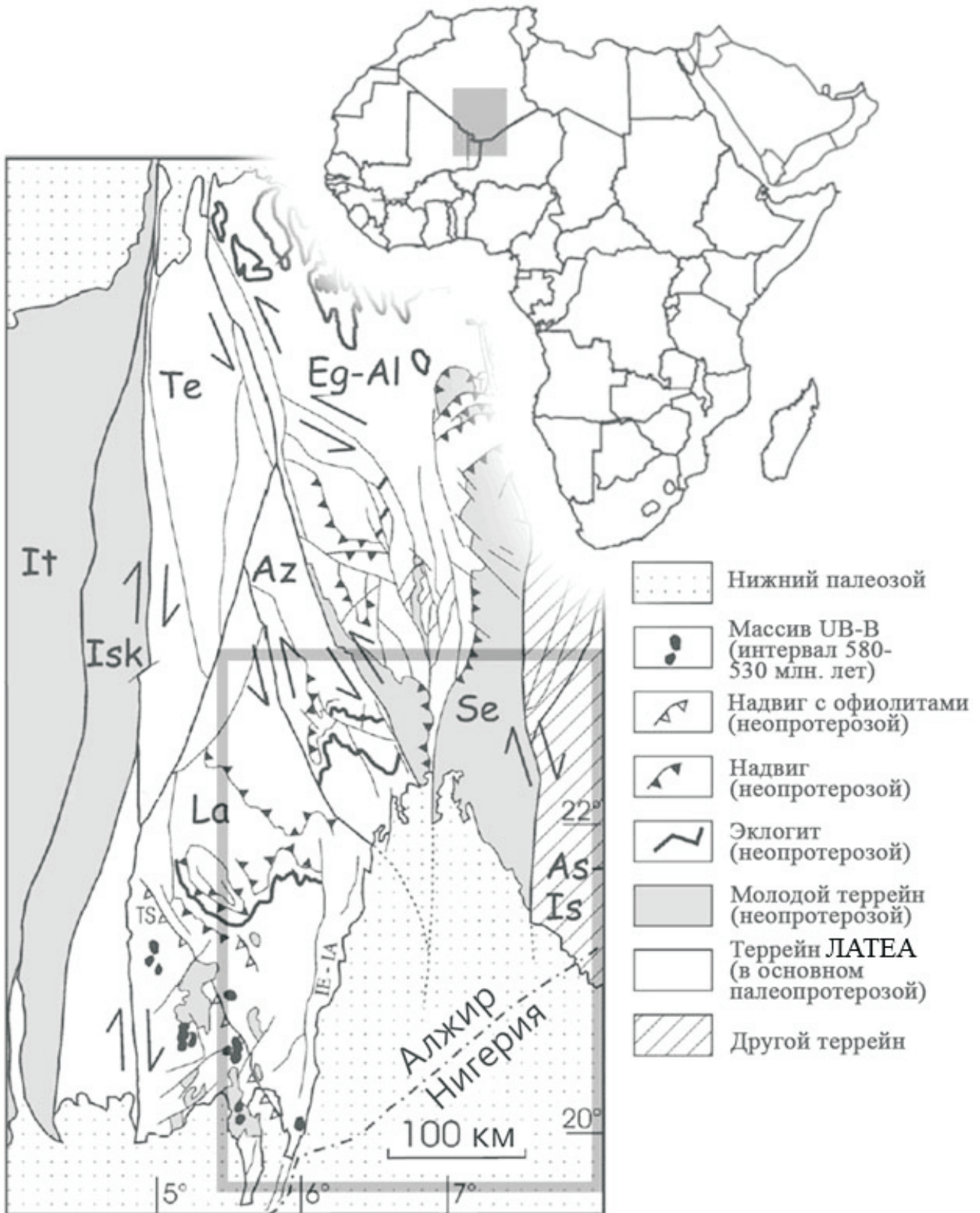


Рис. 2. Массив Ахаггар (Хоггар) вблизи южной границы Алжира. Геологическая карта соответствует прямоугольнику на карте Африки сверху. Исследуемая область соответствует прямоугольнику на геологической карте. Стрелками обозначены разломы сдвигового типа, буквами – отдельные террейны массива Хоггар: Eg-Al – террейн Эгере-Алексод; Te – блок Тефедест; Az – террейн Азроу-н’Фад; Isk – террейн Искел; It – террейн Ин-Тейдини.

Аэромагнитные измерения были проведены на высоте 350 м с малым расстоянием между профилями, что позволило провести интерполяцию на равномерную сетку с шагом 250 м. На полученной карте (рис. 8) видно, что крупные изометричные аномалии располагаются вдоль кадомского пояса, а квазилинейные узкие аномалии протягиваются параллельно дайкам на побережье и уходят далее на акваторию залива. На побережье южнее береговой линии (рис. 8) аномалии имеют простирание С-Ю (СВ-ЮЗ), далее в заливе они меняют свое простирание на ССЗ-ЮЮВ. Структура этих квазилинейных аномалий на пересечении с метаморфическим поясом становится весьма сложной, что затрудняет интерпретацию поля в этом районе. С помощью алгоритма «Роден» была изучена структура источников квазилинейных магнитных аномалий на суше и в заливе, а также их соотношение с метаморфическим поясом.

Для сравнения различных подходов и выбора параметров кластеризации был выбран небольшой участок карты в районе интенсивной изометричной аномалии 1 (рис. 8). На рис. 9а

дана структура аномального поля (ΔT)а (изолинии и объемное изображение) и эйлеровы точки, рассчитанные с размером окна 9×9 точек (2×2 км). Несмотря на то, что более 40 % точек было отброшено, поскольку они имели весьма малые сингулярные значения, полученное распределение трудно поддается интерпретации. Далее были применены «стандартные» критерии, а именно: были отброшены решения, которые имели низкий уровень толерантности, или глубину более 2 км, или располагались на расстоянии, которое в пять раз больше размера окна. Хотя в целом картина несколько прояснилась (рис. 9б), полученное распределение вряд ли можно использовать для трассирования контуров магнитоактивных тел.

На рис. 9в представлены результаты, полученные после применения алгоритма «Роден». Полученные плотные кластеры позволяют протрассировать положение возможных источников магнитных аномалий. В частности, хорошо оконтурено изометричное тело в северной части карты и серия линейных тел с СВ-простиранием параллельно кадомскому метаморфическому поясу.

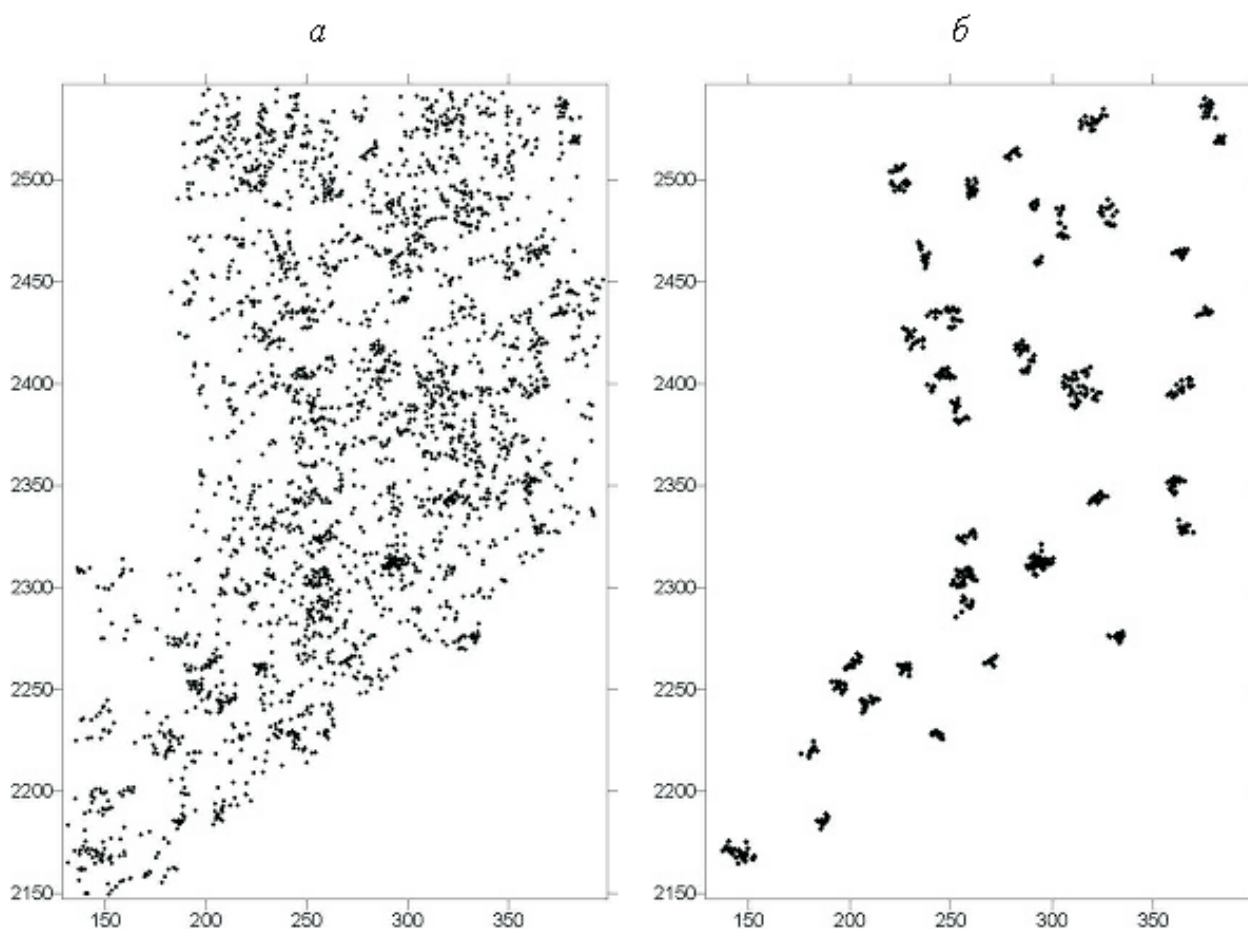


Рис. 3. Проекция эйлеровых решений, полученных при структурном индексе $N = 3$, на горизонтальную плоскость (xOy): а – результат стандартной фильтрации; б – результат применения к ним алгоритма «Кристалл». По осям отложены километры.

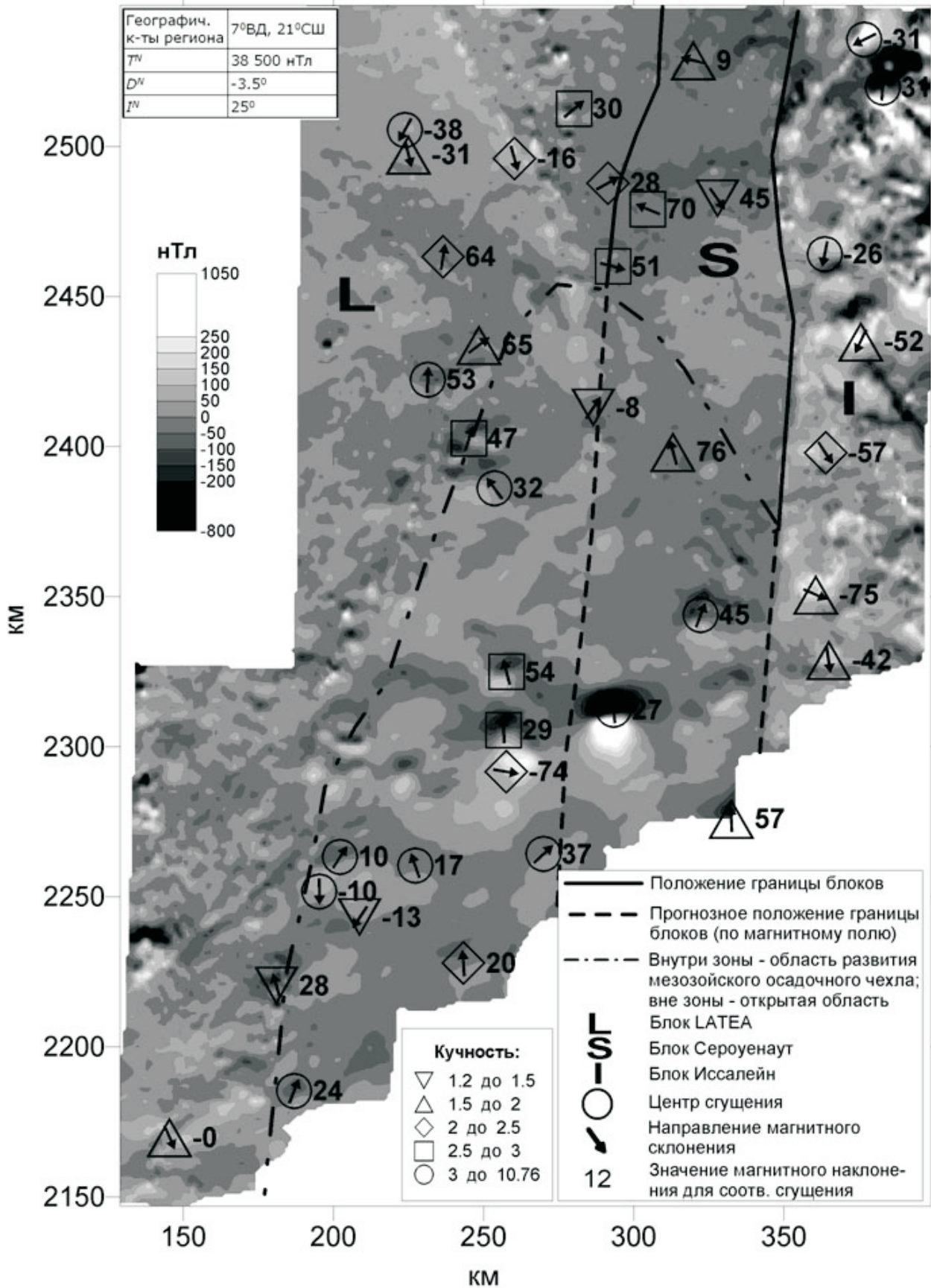


Рис. 4. Результаты определения средних углов намагниченности для каждого из выделенных алгоритмом «Кристалл» сгущений.

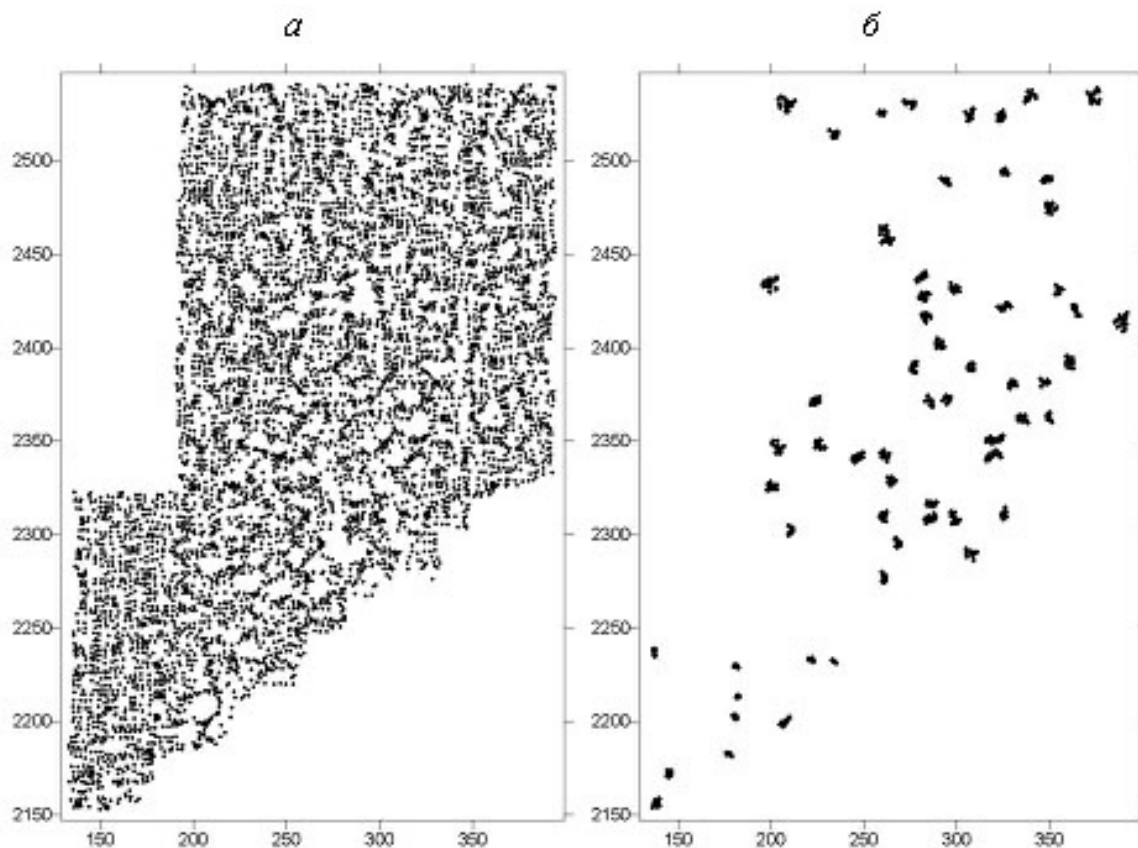


Рис. 5. Проекция эйлеровых решений, полученных при структурном индексе $N = 0.5$, на горизонтальную плоскость (xOy): *а* – результат стандартной фильтрации; *б* – результат применения к ним алгоритма «Кристалл». По осям отложены километры.

На рис. 10 приведены результаты применения алгоритма «Роден» к участку карты, отмеченному пунктирной линией на рис. 7. Расчеты проведены при тех же значениях параметров, что и на рис. 9*в*. Из начального набора в 34500 решений алгоритм «Роден» отбросил 7000 точек, сформировав плотные кластеры, которые оконтуривают изометричные и линейные тела. В частности, четко обозначились контуры трех крупных изометричных тел (рис. 8). Линейные кластеры с СВ-простираем южнее изометричного тела 2 и далее к телу 1 могут рассматриваться как продолжение на восток главного кадомского шва (рис. 7, ГКШ), в то время как гипотеза о продолжении на восток разлома Канкаль-Плуэ (рис. 7, РКП) не находит своего подтверждения.

Важной особенностью полученного решения (рис. 10) являются линейные кластеры с меридиональным простираем в южной части рисунка и СЗ-ЮВ в его северной части. Южнее береговой линии (рис. 10) они имеют то же простираем, что и выходящие на поверхность долеритовые дайки. При этом полученные кластеры показывают, что кроме системы даек меридионального простираем, возможно, имеется и вторая система даек, простираем СЗ-ЮВ. Эйлеровы точки позволяют протрассировать вторую систему кластеров до

их пересечения с кадомским метаморфическим поясом, севернее которого имеется система линейных кластеров, также простирающихся с СЗ на ЮВ. Предварительный анализ показывает, что магнитные тела, соответствующие линейным кластерам простираем СЗ-ЮВ, по-видимому, моложе самого пояса, потеря четкой линейности на их пересечении с поясом, возможно, является признаком более поздних слабых деформаций, произошедших после формирования линейных магнитных тел.

Применение метода кластеризации «Роден» позволило получить решения, более четко трассирующие контуры аномалеобразующих тел, чем это было сделано с применением стандартных приемов отбраковки решений.

Одно из применений алгоритма «Монолит» связано с обработкой спутниковых интерферограмм с целью изучения стабильных участков в районе вулкана Этна. В период с 1992 по 1999 гг. с помощью спутников ERS-1 и ERS-2 было сделано 38 снимков, на основе которых было вычислено 238 интерферограмм (Massonnet et al., 1995; Briole et al., 1997). Каждая из этих интерферограмм представляет собой область 55×55 км, разбитую на более чем 4 млн. узлов (рис. 11*а*). В каждом узле находится масштабированное от 0 до 255 значение,

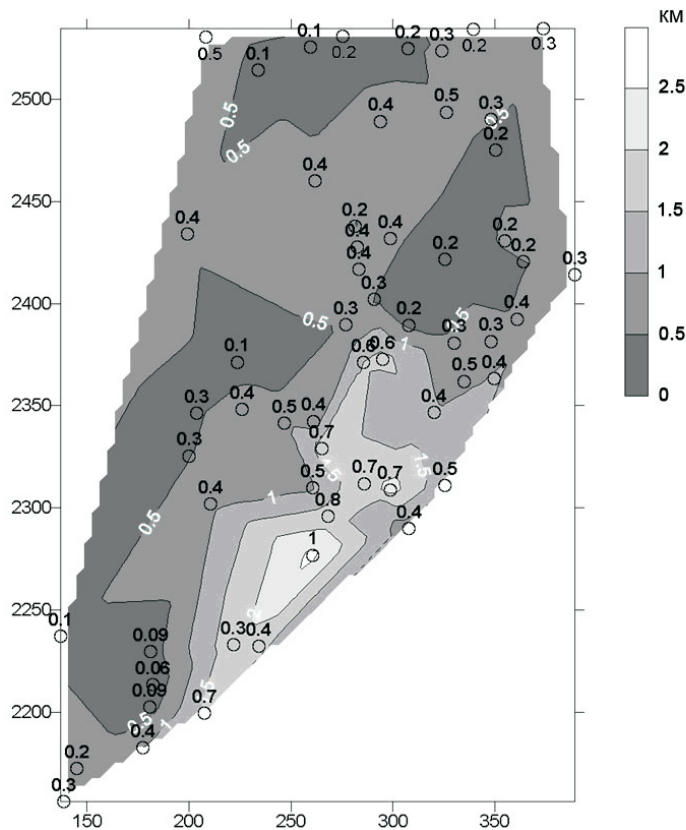
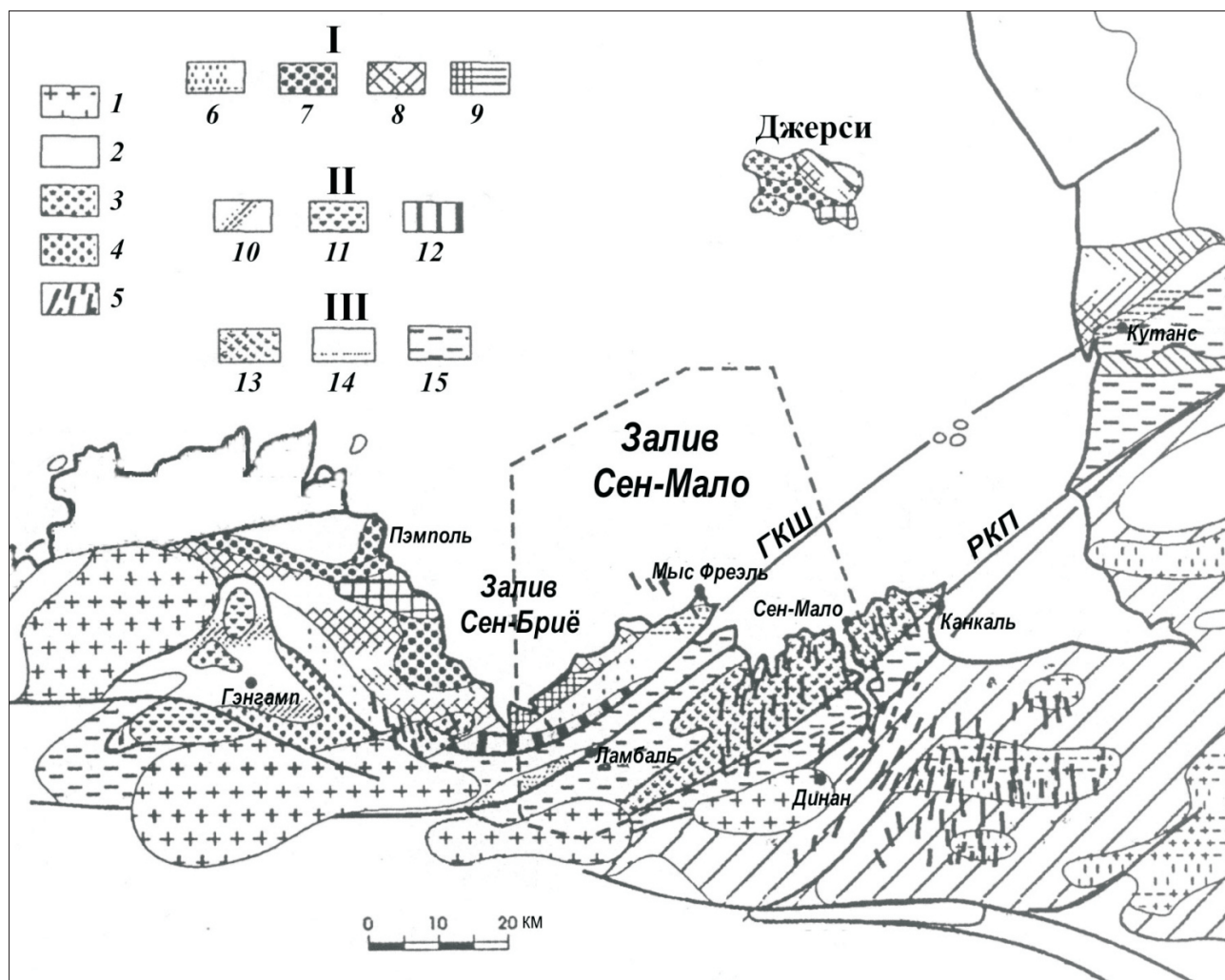


Рис. 6. Карта прогнозных изоглубин поверхности кристаллического фундамента, полученная по результатам интерпретации. Черными кругами обозначены центры сгущений, выделенных алгоритмом «Кристалл». Цифры, расположенные над кругами, – значения средних глубин соответствующих сгущений. По осям отложены километры.

Рис. 7. Тектоническая карта района залива Сен-Мало (Chantraine et al., 2001):

ГКШ – главный кадомский шов; РКП – разлом Канкаль-Плуэ; САЗС – Северо-Армориканская зона сдвига; 1 – вариссийские гранитные интрузии; 2 – Pz и Mz осадочные породы; 3 – пост-кадомские интрузии; 4 – гранитоиды, 5 – долеритовые дайки. Тектонические зоны: I – Сен-Бриё; 6 – габбро-диоритовые интрузии; 7 – осадочные породы; 8 – метаморфизованные вулканические породы; 9 – ортогнейсы; II – Гинган: 10 – граниты; 11 – мигматиты; 12 – метагаббро; III – Сен-Мало: 13 – мигматиты; 14 – граниты; 15 – осадочные породы. Пунктиром отмечен проинтерпретированный алгоритмом «Роден» участок.



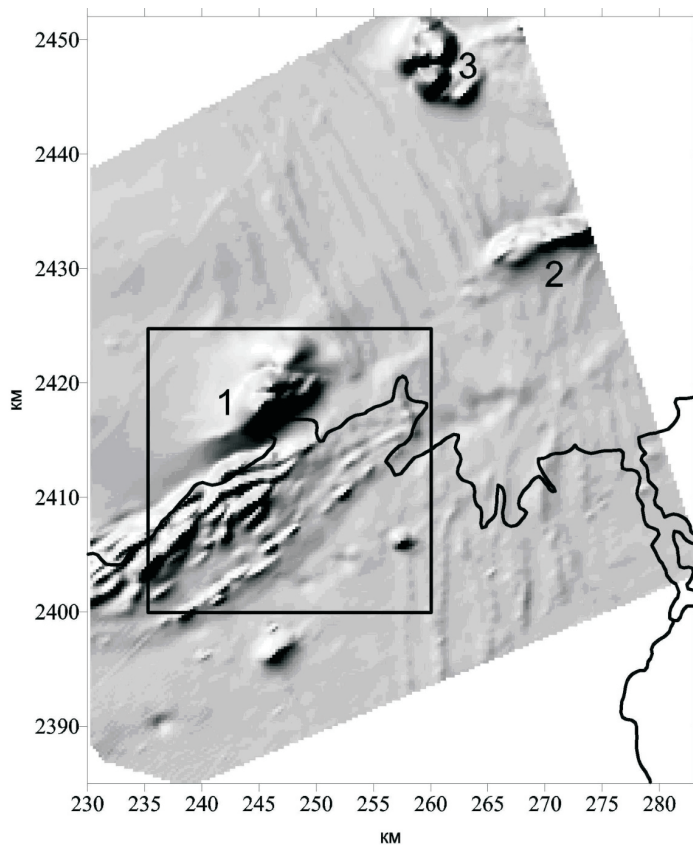
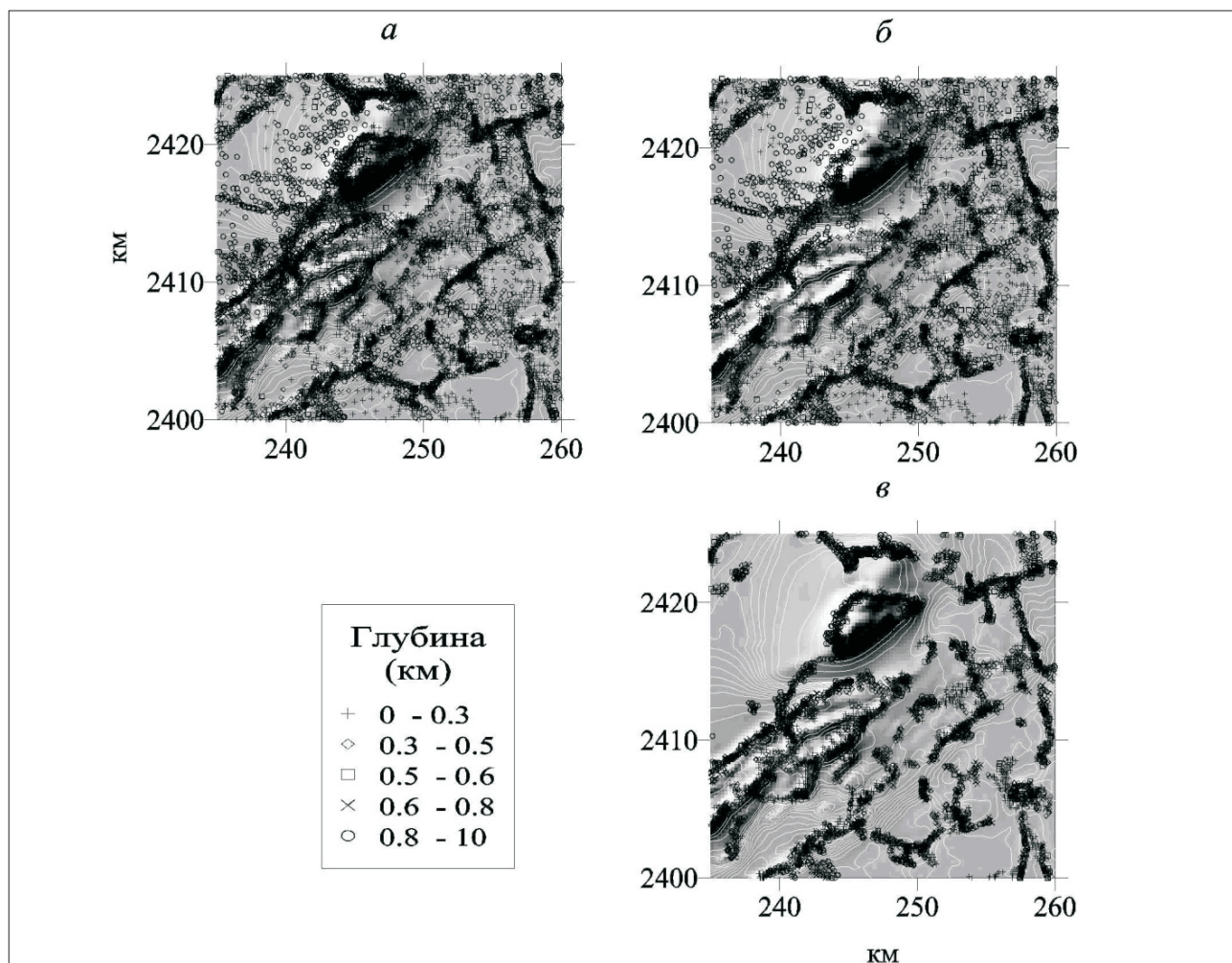


Рис. 8. Аномальное поле (ΔT) участка в районе залива, отмеченного на рис. 7 пунктиром; 1-3 – контуры трех крупных изометричных тел. Квадратом обозначен проинтерпретированный участок, на котором сравнивались различные методы очистки решения (рис. 9).

Рис. 9. Аномальное поле (ΔT) *a*. Результаты Эйлеровой деконволюции для части области, отмеченной на рис.8 квадратом: *a* – все решения; *б* – очистка стандартными методами; *в* – результат кластеризации алгоритмом «Роден»



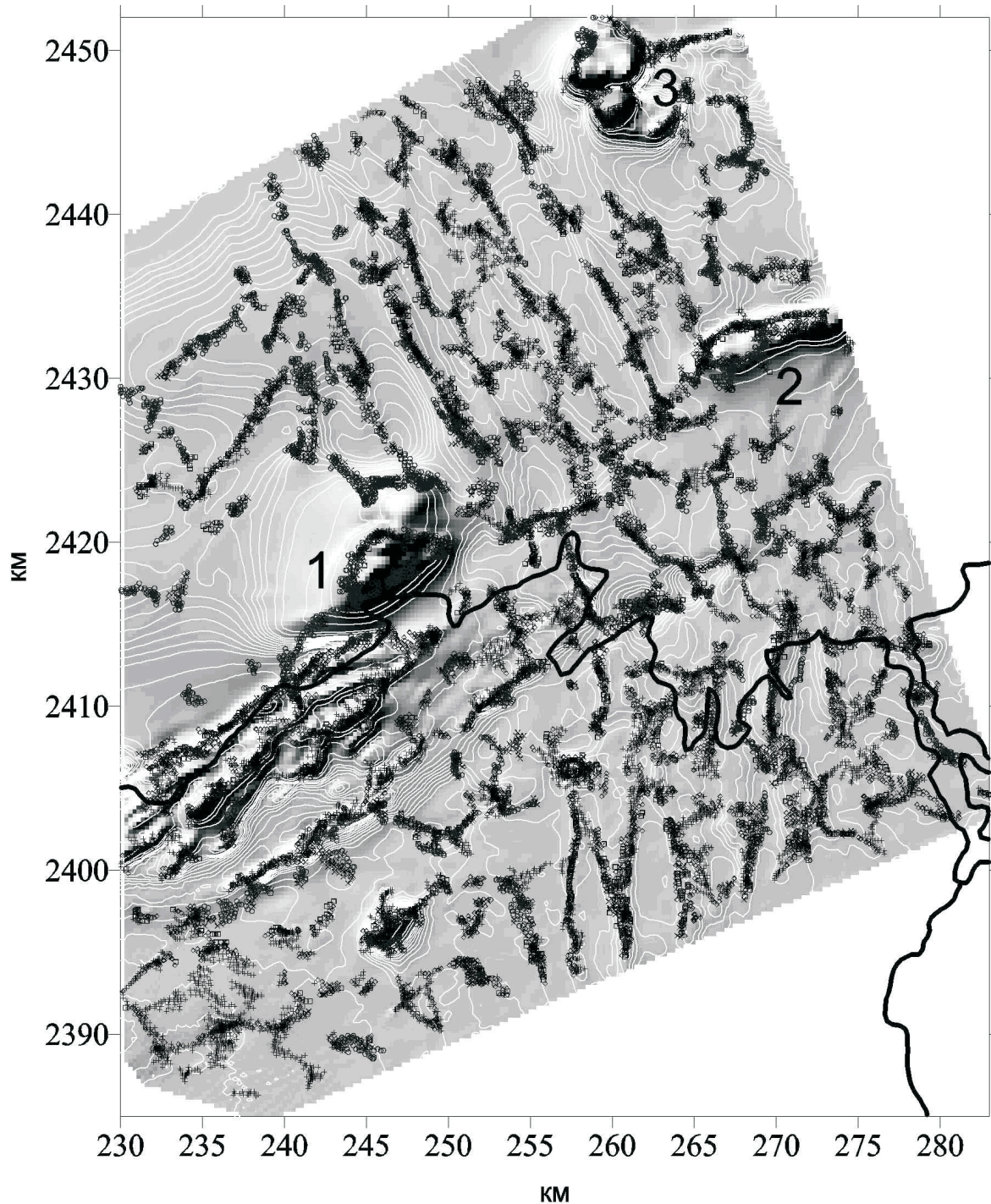


Рис. 10 Результаты кластеризации Эйлеровых решений в районе залива Сен-Мало алгоритмом «Роден». Местоположение проинтерпретированного участка представлено на рис. 7.

представляющее собой изменение фазового расстояния. Особый интерес представляли области, в точках которых значения изменялись не резко (гладкие точки). Этим областям соответствуют лавовые области, не покрытые деревьями. Визуально это – одинаковые по тону участки.

В работе использовалось самое простое и естественное выражение гладкости рельефа в узле в виде разброса значения в нем от значе-

ний в соседних узлах. Гладкие узлы – это узлы с маленьким разбросом. Строгое выражение этого получено при помощи нечетких сравнений. Гладкие узлы изображены на рис. 11б и именно к ним применялся «Монолит».

В результате применения алгоритма 4 раза получилась картина, изображенная на рис. 11в. Изолированные гладкие узлы исчезли, небольшие дыры в массивных скоплениях заполнились.

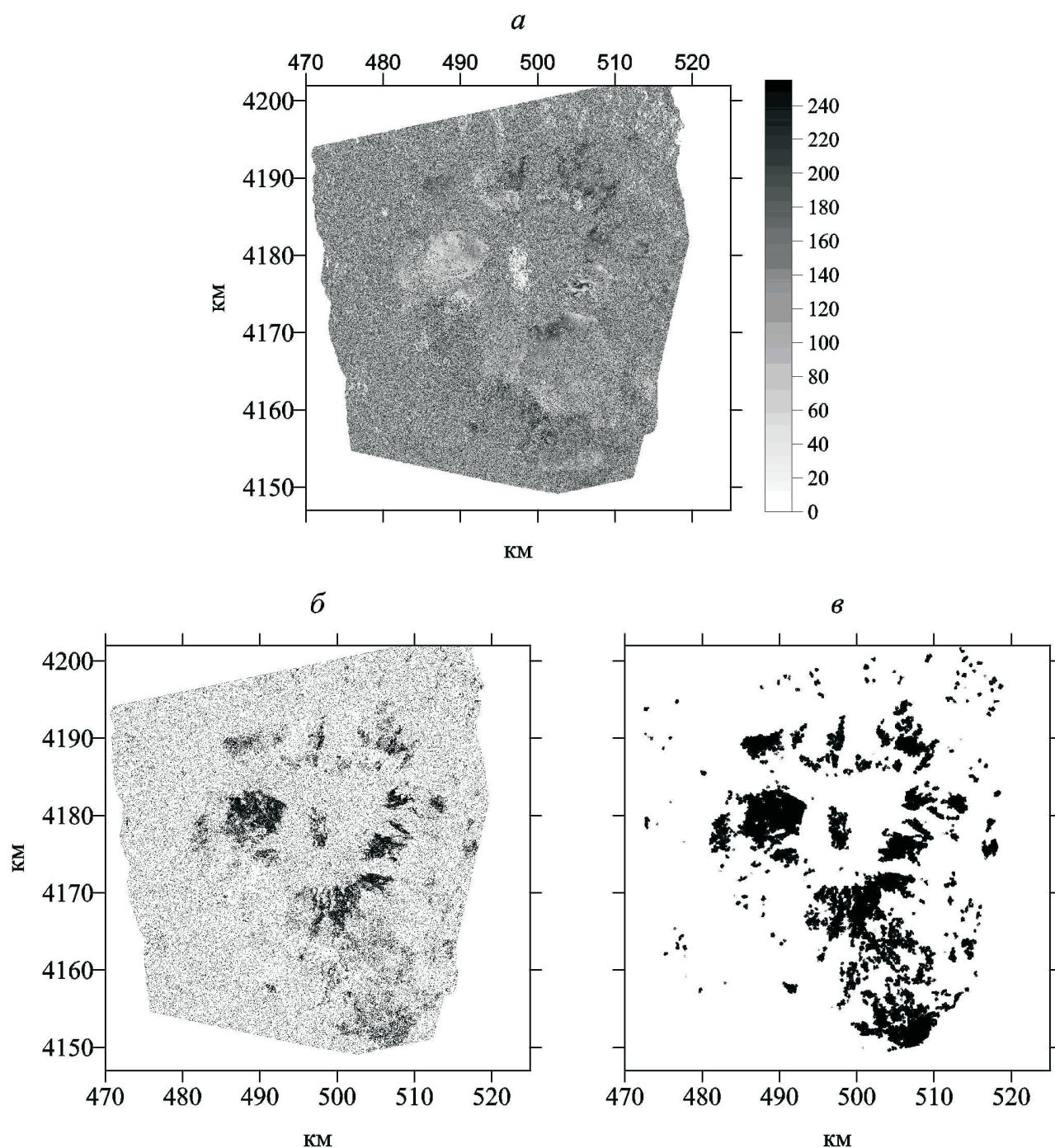


Рис. 11. Обработка спутниковых интерферограмм: *a* — интерферограмма (оттенками серого цвета обозначены снормированные от 0 до 255 изменения фазового расстояния); *б* — «гладкие» точки; *в* — результат применения «Монолита» (4 итерации).

СРАВНЕНИЕ С КЛАССИЧЕСКИМИ МЕТОДАМИ КЛАСТЕРИЗАЦИИ

На примере алгоритма «Монолит» (рис. 12) показано преимущество построенных алгоритмов по сравнению с классическими алгоритмами кластерного анализа (Kaufman, Rousseeuw, 1990). В простой ситуации наверху алгоритмы «Монолит», «иерархический метод» и «k-средние» работают одинаково. В сложных ситуациях

появляется шум, с которым справляется (оказывается к нему устойчивым) только алгоритм «Монолит».

ЗАКЛЮЧЕНИЕ

Алгоритмы Кристалл», «Роден» и «Монолит» скорее относятся к расширенному пониманию кластер-анализа как описательного инструмента для изучения взаимного расположения объектов в метрическом пространстве. В частности,

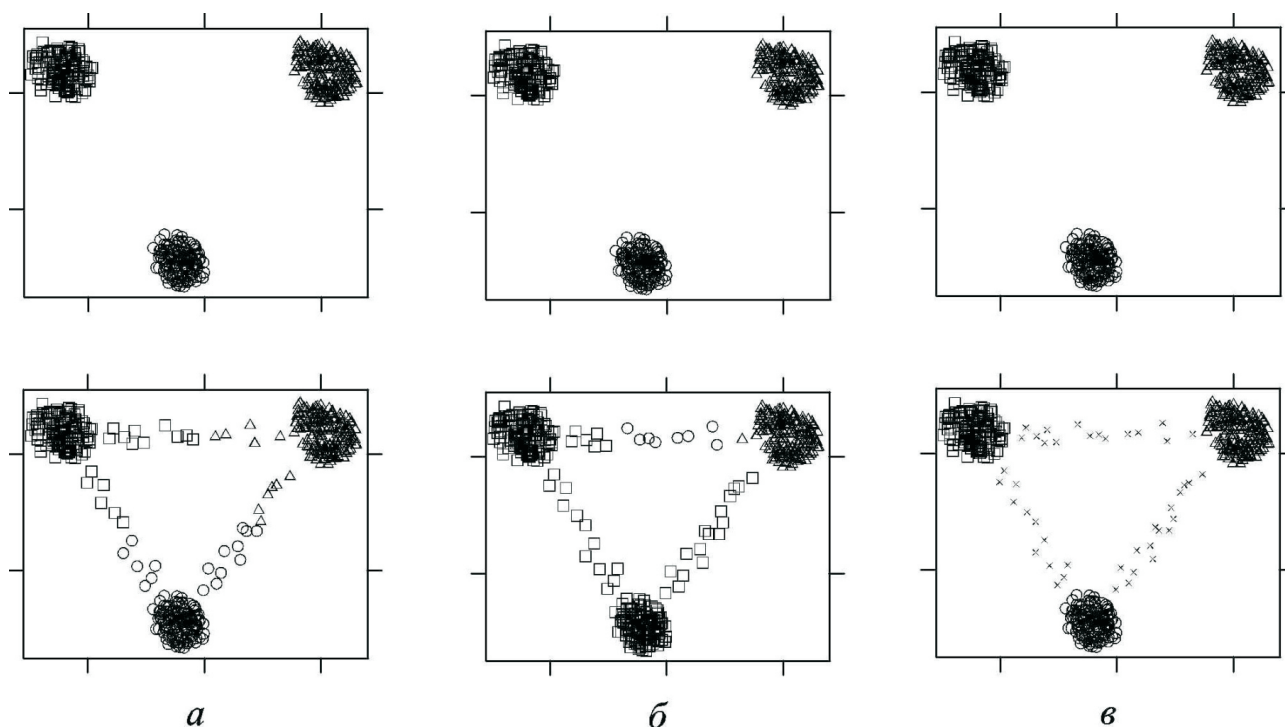


Рис. 12. Сравнение различных методов кластеризации на синтетических примерах: *а* – нечеткие k -средние; *б* – иерархический метод; *в* – алгоритм «Монолит» (треугольниками, квадратами и кружками обозначены кластеры, выделенные различными алгоритмами; Крестиками на последнем рисунке отмечен фон).

под кластер-анализом подразумевается метод изучения списка всех возможных кластеров при использовании некоторого заданного определения кластера. При этом любые два кластера могут пересекаться по любому числу элементов. Такой подход отличается от традиционного понимания кластер-анализа как классификации без учителя и требующего дизъюнктивности кластеров либо устанавливающего верхнюю границу числа общих элементов для любых двух кластеров.

Техническая основа этих алгоритмов в виде нечеткой математики дает возможность выделять более тонкие объекты, чем кластеры, а именно сгущения. Эти алгоритмы нацелены на топологическую фильтрацию пространства, позволяющую отсеивать случайные (изолированные, несущественные) части. В то же время если пространство допускает разбиение на однородные непересекающиеся части, то разработанные алгоритмы позволяют выделить их, и в этом смысле они являются алгоритмами кластеризации.

Список литературы

- Агаян С.М., Соловьев А.А.* Выделение плотных областей в метрических пространствах на основе кристаллизации // *System Research & Information Technologies*. 2004. № 2. Р. 7-23.
- Богоутдинов Ш.Р.* Применение методов нечеткой логики (алгоритм «Монолит») для интерпретации геомагнитных данных // *Материалы 33-й сессии Международного семинара им. Д.Г. Успенского «Вопросы теории и практики геологической интерпретации гравитационных, магнитных и электрических полей»*. Екатеринбург: ИГ УрО РАН, 2006. С. 41-48.
- Винер Н.* Творец и робот. М.: Прогресс, 1966. 104 с.
- Гвишиани А.Д., Диаман М., Михайлов В.О. и др.* Алгоритмы искусственного интеллекта для кластеризации магнитных аномалий // *Физика Земли*. 2002. № 7. С. 13-28.
- Кейлис-Борок В.И.* Сейсмология и логика // *Вычислительная сейсмология*. 1968. Вып. 4, с.317-350.
- Математическая энциклопедия*. М.: Советская энциклопедия, 1979. Т. 2. 552 с.
- Соловьев А.А., Шур Д.Ю., Гвишиани А.Д. и др.* Определение вектора магнитного момента при помощи кластерного анализа результатов локальной линейной псевдоинверсии аномалий ΔT // *ДАН*. 2005. № 1. С. 1-4.
- Briole P., Massonnet D., Delacourt C.* Post-eruptive deformation associated with the 1986-87 and 1989 lava flows of Etna detected by radar interferometry // *Geophys. Res. Lett.* 1997. V. 24. P. 37-40.
- Hood P.* Gradient measurements in aeromagnetic surveying // *Geophysics*. 1965. V. XXX. P. 891-902.

- Kaufman L., Rousseeuw P.J.* Finding Groups in Data: an Introduction to Cluster Analysis. Wiley-Interscience, 1990. 368 p.
- Massonnet D., Briole P., Arnaud A.* Deflation of Mount Etna monitored by spaceborne radar interferometry // *Nature*. 1995. V. 375. P. 567-570.
- Mikhailov V., Galdeano A., Diamant M. et al.* Application of artificial intelligence for Euler solutions clustering // *Geophysics*. 2003. V. 68. № 1. P. 168-180.
- Reid A.B., Allsop J.M., Granger H. et al.* Magnetic interpretation in three dimensions using Euler deconvolution // *Geophysics*. 1990. V. 55. P. 80-91.
- Thompson D.T.* EULDPH: A new technique for making computer-assisted depth estimates for magnetic data. // *Geophysics*. 1982. V. 47. P. 31-37.
- Zadeh L. A.* Fuzzy sets // *Inf. Control*. 1965. № 8. P. 338-353.
- Chantraine J., Egal E., Thieblemont D. et al.* The Cadomian active margin (North Armorican Massif, France): a segment of the North Atlantic Panafrican belt. // *Tectonophysics*. 2001. V. 331. Is. 1-2. P. 1-18.

DISCRETE MATHEMATICAL ANALYSIS AND APPLICATIONS GEOLOGY AND GEOPHYSICS

A.D. Gvishiani, S.M. Agayan, Sh.R. Bogoutdinov, A.A. Solovyov

Institution of the Russian Academy of Sciences Geophysical center of RAS

This article describes a new original approach to the analysis of data which is called the Discrete Mathematical Analysis (DMA). The authors have made an attempt to create a model of a finite limit using fuzzy techniques. The model helped to find the key in terms of classical mathematics to the basic problems of a discrete case – continuity, coherence, closeness, trend etc. DMA is a series of algorithms used to solve the basic tasks of data analysis: clustering, tracing, morphological analysis, trend and so on. All the DMA algorithms are universal and are based on a finite limit. The authors consider this is one of possible approaches to a finite limit and the DMA construction. The article suggests three algorithms of clustering: «Crystal», «Rodin» and «Monolith» and some of their applications in geophysics.

Keywords: Discrete Mathematical Analysis, algorithms, fuzzy techniques, clustering.