

А.Г. Злобина, И.В. Журбин

# ВОССТАНОВЛЕНИЕ ГРАНИЦЫ ОБЪЕКТА ПО ДАННЫМ МАЛОГЛУБИННОЙ ЭЛЕКТРОРАЗВЕДКИ МЕТОДОМ НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ



Метод электроразведки успешно применяется в малоуглубинных исследованиях при поиске в грунте объектов искусственного происхождения: трубопроводов, котлованов жилищ, фундаментов зданий, пустот и т.д. Эффективной методикой изучения структуры приповерхностных слоев грунта и локализации объектов является площадное электропрофилирование, которое позволяет оценить распределение кажущегося удельного электрического сопротивления грунта на участке исследований.

Измерительная установка состоит из пары питающих электродов, через которую вводится ток в грунт, и пары измерительных электродов, посредством которой снимается падение напряжения на участке исследований. При использовании многоэлектродной электроразведочной аппаратуры электроды располагаются в узлах сетки с постоянным шагом  $b$  [1]. Зная значение падения напряжения, силу тока и взаимное расположение электродов, рассчитывают кажущееся удельное электрическое сопротивление грунта (КС) на текущем участке. Значение относят к условной точке привязки результатов (*точка измерений*) – обычно к геометрическому центру измерительной установки [2, с. 20]. В частности, при использовании дипольной экваториальной установки – к центру квадрата, вершинами которого являются точки расположения электродов, используемых при текущем измерении.

При площадном электропрофилировании перемещение измерительной установки в соответствии с сеткой электродов позволяет сформировать массив значений КС, расположение элементов которого соответствует взаимному расположению точек измерений (рис. 1). Визуализация за счет яркостного представления элементов массива данных позволяет получить графическое изображение – геофизическую карту распределения сопротивления.

В зависимости от структуры и состава грунта объекты поиска могут проявляться как аномалии повышенного или пониженного сопротивления относительно фонового значения – КС вмещающего грунта. Качественная интерпретация данных электроразведки предполагает выявление объекта поиска на фоне вмещающего грунта и определение его границ по аномалии на карте распределения КС. При этом граница аномалии, восстановленная в процессе обработки данных, должна с требуемой точностью соответствовать истинной границе объекта.

## Основные подходы к анализу данных

Обработка и интерпретация данных малоуглубинной электроразведки на территории археологических памятников имеет свою специфику, связанную с процессами формирования и разрушения объектов поиска, особенностями функционирования древних поселений, а также изменением структуры

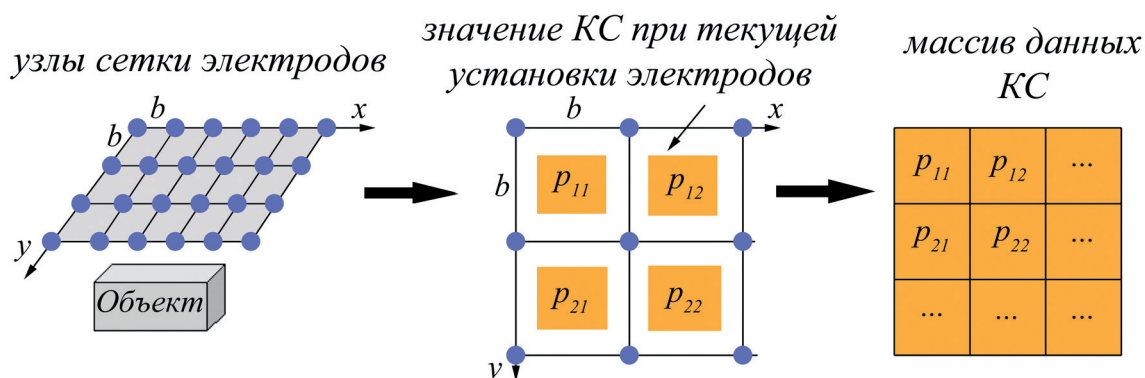


Рис. 1. Формирование массива данных кажущегося сопротивления

приповерхностных слоев грунта (культурного слоя) в результате современной хозяйственной деятельности. Степень влияния природных и антропогенных факторов наглядно демонстрируют данные площадного электропрофилирования на территории средневекового городища Уччакар в Удмуртии (рис. 2а). На участке геофизической съемки проведены археологические раскопки (рис. 2б), что позволяет провести оценку результатов интерпретации и обосновать выявленные особенности распределения КС.

В частности, проблемы интерпретации данных малоуглубинной электроразведки могут быть связаны с плавным изменением фонового уровня сопротивления на участке исследований. В большинстве случаев наличие *тренда* значений КС соответствует вариации толщины культурного слоя (уменьшение мощности отложений на склонах, результат эрозийных процессов на участках распахки и т.п.) или локальным изменениям рельефа кровли материковых пород. Аналогичная ситуация наблюдается при геоэкологических исследованиях. На приведенной геофизической карте выявлено общее снижение значений КС в южном направлении, что связано с ландшафтом местности – измерения проводились вблизи склона мыса коренной террасы.

Кроме того, на геофизической карте фиксируются *аморфные аномальные области* повышенного сопротивления (рис. 2а). Такие области могут быть связаны с локальными линзами или небольшими прослойками грунтов иных типов во вмещающем слое (слой 4 на рис. 2б). Возникновение таких включений обычно связано с процессами освоения и использования территории поселений (например, выбросы грунта при создании или перестройке сооружений в древности). Они характеризуются достаточной площадью и объемом, при этом отличаются по сопротивлению от фоновых участков. Необходимо отметить, что такие неоднородности не являются объектами поиска. Подобные проблемы возникают и в инженерной геофизике при работе в городских условиях и в зонах строительства. Именно с такими включениями связаны аморфные аномалии повышенного сопротивления (рис. 2а, б).

Сами *объекты поиска* (древние сооружения, ямы различного назначения, печи и т.п.) обычно представляют собой локальные области, существенно отличающиеся по удельному сопротивлению от вмещающих грунтов. На основе априорных сведений возможно прогнозировать форму и уровень удельного сопротивления на участке погребенного объекта.

Процесс выделения аномалий и определения границы объекта поиска, по сути, можно рассматривать как задачу классификации данных, поскольку

необходимо каждый элемент массива данных (каждую точку на изображении – геофизической карте) отнести в один из классов: «Фон», «Объект 1», ..., «Объект N».

Аномалия пониженного сопротивления в центре планшета (рис. 2а) соответствует фундаменту средневекового сооружения. Основания таких сооружений представляют собой линзу уплотненной или прокаленной глины (рис. 2б). Обычно форма оснований близка к прямоугольнику [3, рис. 3, 4]. Глиняная площадка окружена неоднородным гумусированным слоем, который содержит разнообразные включения (песок, глина, уголь, зола, древесный тлен и пр.). Исходя из этого, корректная классификация данных малоуглубинной электроразведки на данном участке измерений должна выявить элементы только двух классов – «Фон» и «Объект».

К задаче моделирования границ объекта применимы методы сегментации, основными из которых являются градиентные методы, пороговые методы, методы кластеризации, методы разрастания регионов [4].

В основе *градиентных методов* лежит предположение о том, что наибольшие значения градиента будут наблюдаться вдоль границ при переходе от объекта к фону. Данный метод не разделяет элементы данных на классы, а только выделяет границы объектов. При наличии плавных переходов между объектами, без выраженных точек перегиба, применение этого метода может привести к ошибкам, которые заключаются в слиянии объектов или выделении ложных границ. При анализе массива данных полевых исследований (рис. 2а) границы самого объекта поиска не выявлены (рис. 2в). Фиксируют лишь фрагменты границы аморфных областей в пределах фоновой области. Границы не замкнуты ввиду влияния выраженного тренда КС на участке исследований.

*Пороговые методы* основаны на сравнении значений элементов данных с заданной пороговой величиной, количество которых, в общем случае, может быть больше одной. В зависимости от числа интервалов методы сегментации по порогу разделяют на двухуровневые и многоуровневые. Пороговые значения определяют границы интервалов, принадлежность к которым выступает в качестве критерия разбиения элементов на классы. При этом наличие тренда КС и неоднородная структура фона приводят к избыточной сегментации данных. В частности, применение формулы Стреджесса при выборе пороговых значений приводит к необоснованной детализации (рис. 2г), что определяет возникновение ложных границ в пределах области фона и избыточное разбиение объекта поиска.

Подобная проблема возникает при классификации *методом кластерного анализа*. Точки со схожими характеристиками в пространстве признаков концентрируются вокруг некоторого центра, образуя кластер. Задача классификации данных сводится к разбиению пространства признаков на кластеры с последующим присвоением каждой точки изображения к тому или иному кластеру в зависимости от меры схожести. Учитывая, что при классификации данных малоуглубинной электроразведки может быть использован только один признак (значение удельного сопротивления), данный подход имеет те же ограничения, что и пороговый метод.

На первом этапе *метода разрастания регионов* выбираются центры регионов (классов), к которым последовательно присоединяются точки из ближайшей окрестности, сходные по своим свойствам. Недостатком данного метода также является избыточная сегментация данных и, соответственно, выделение ложных границ. В рассмотренном примере метод разрастания регионов приводит

к избыточной классификации данных (рис. 2д), фоновая область разбита на 4 разных класса, а объект поиска – на 5 классов.

Таким образом, рассмотренные методы не эффективны при выделении границы объекта в условиях тренда значений КС и выраженных аморфных аномальных областей на фоновых участках. В связи с этим при восстановлении границы объекта более адекватны методы нечеткой кластеризации, позволяющие разбить все элементы на классы в зависимости от их степени принадлежности.

### Алгоритм Fuzzy c-means

Традиционным методом нечеткой кластеризации является алгоритм Fuzzy c-means. Согласно этому алгоритму вводится матрица нечеткого разбиения  $F$ , значения элементов которой соответствуют степени принадлежности элементов массива данных ко всем классам. Основная идея метода заключается в следующем: при инициализации задается три параметра – количество классов  $c$ ; экспоненциальный

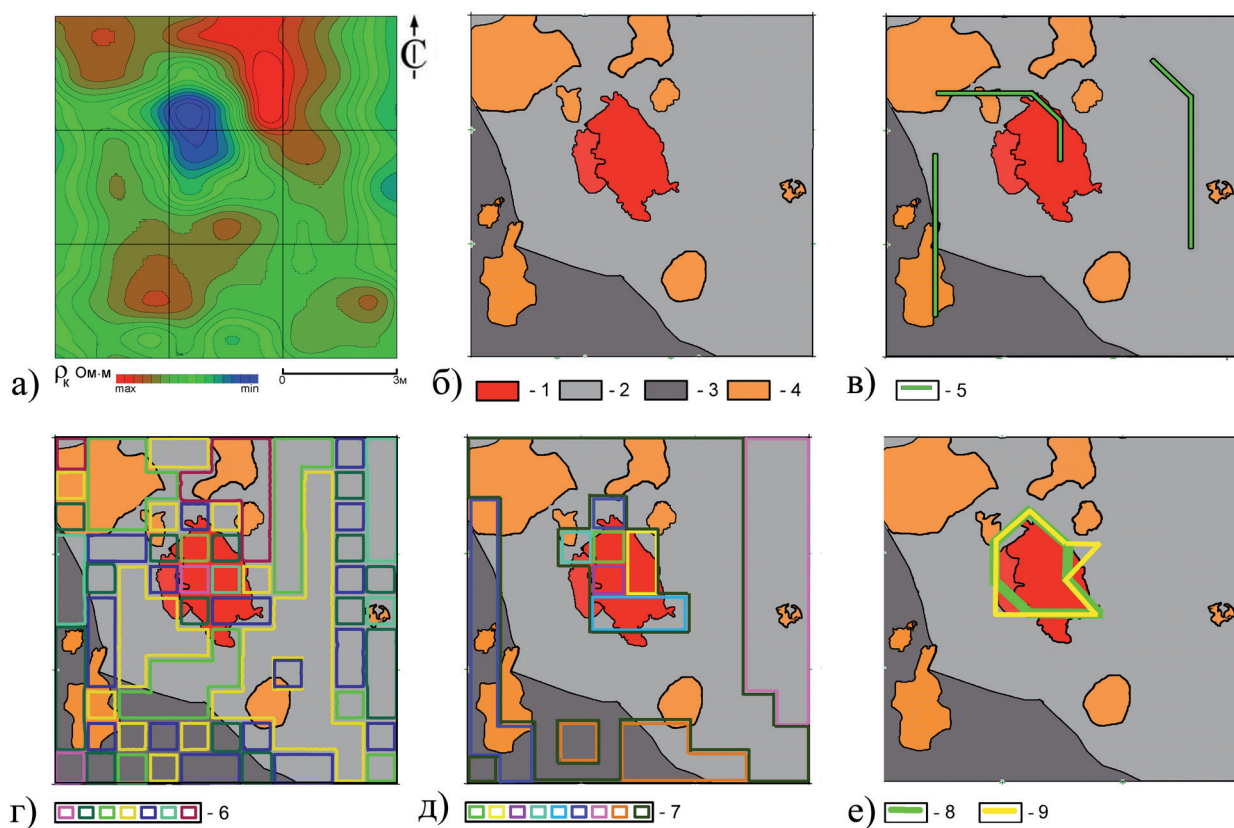


Рис. 2. Результаты комплексных исследований городища Учкакар:

- а) карта распределения сопротивления; б) обобщенная прорисовка планиграфического разреза раскопа; в) результат применения градиентного метода; г) классификация пороговым методом; д) классификация методом разрастания регионов; е) классификация методом нечеткой кластеризации

Условные обозначения: 1 – прокаленная глина, глина различных оттенков; 2 – перемешанный слой (суглинок, гумус, зола, уголь); 3 – бурый гумус; 4 – золистый суглинок; 5 – восстановленная граница; 6 – границы классов; 7 – границы классов; 8 – восстановленная граница объекта при  $\alpha = 0,5$ ; 9 – восстановленная граница объекта при  $\alpha = 0,9$

вес  $m$ , определяющий нечеткость классов; параметр останова алгоритма  $\epsilon$ . На каждом шаге итерации рассчитываются центры классов, вычисляются расстояния между всеми точками и центрами классов, обновляется матрица разбиения до тех пор, пока норма разницы матриц нечеткого разбиения на текущей и предыдущей итерации не станет меньше  $\epsilon$  [5, с. 2-3].

Существенным ограничением применения этого алгоритма является необходимость априорного задания количества классов. Неоднозначность выбора определяет недостаточность предварительной информации о структуре участка грунта. Одним из вариантов решений этой проблемы является задание изначально большего числа классов с последующим объединением близкорасположенных кластеров. Процесс агломерации классов прекращается, когда все расстояния между оставшимися классами превышают заданный порог.

Вторым параметром алгоритма является экспоненциальный вес  $m$ . Также не существует обоснованного правила его выбора, но при  $m \rightarrow \infty$  все элементы принадлежат ко всем классам с одинаковой степенью  $1/c$ . В этом случае применение нечеткой кластеризации нерезультативно. Как правило, этому параметру присваивают значение, равное двум [6, с. 220]. Экспериментальные исследования показали, что при оценке структуры геологического разреза изменение значения параметра  $m$  практически не влияет на результаты классификации [7]. На основании этого при использовании алгоритма Fuzzy c-means для кластеризации данных малоуглубинной электроразведки принято  $m = 2$ .

Результатом работы алгоритма Fuzzy c-means является конечная матрица нечеткого разбиения  $F$ . Основанием включения текущего элемента массива данных в конкретный класс является сравнение степени принадлежности с заданным порогом –  $\alpha$ -сечением. Под  $\alpha$ -сечением для двух классов «Фон» и «Объект» будем понимать следующее: если степень принадлежности элемента массива данных к классу «Фон» больше заданного порогового значения  $\alpha$ , то этот элемент принадлежит к классу «Фон», иначе – к классу «Объект».

Конечную матрицу нечеткого разбиения можно представить двумерным массивом, в котором значение каждого элемента отражает степень принадлежности соответствующего значения КС к классу «Фон» (рис. 3). Серым цветом выделена область расположения объекта, красным контуром – его граница. Степень принадлежности элементов на «фоновом» участке в целом близка к единице, а элементов внутренней области объекта близка к нулю.

1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	0,99	1,00	1,00	0,99
1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00
1,00	1,00	0,99	0,98	0,94	0,94	0,96	0,95	0,96	1,00	1,00	1,00
1,00	0,99	0,78	0,20	0,03	0,02	0,01	0,02	0,15	0,70	1,00	1,00
1,00	1,00	0,30	0,00	0,01	0,03	0,03	0,02	0,00	0,28	0,99	0,99
1,00	1,00	0,26	0,01	0,04	0,03	0,05	0,02	0,01	0,20	0,99	1,00
1,00	0,99	0,13	0,00	0,04	0,03	0,05	0,03	0,00	0,28	1,00	1,00
1,00	0,99	0,19	0,01	0,02	0,04	0,02	0,03	0,02	0,35	1,00	1,00
1,00	1,00	0,41	0,00	0,00	0,01	0,00	0,00	0,00	0,38	0,99	1,00
1,00	1,00	0,84	0,53	0,33	0,18	0,16	0,19	0,45	0,90	1,00	1,00
1,00	1,00	1,00	1,00	1,00	0,99	0,99	0,99	1,00	0,99	0,99	1,00
1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	1,00	1,00	1,00

Рис. 3. Массив степени принадлежности к классу «Фон»

### Восстановление границы объекта: параметры и оценки

Расположение объекта в грунте априорно не известно, следовательно, сетка электродов, установленная на участке исследований, может быть ориентирована произвольным образом относительно границ объекта. При этом линейные участки *истинной границы* объекта могут проходить либо вдоль линий сетки электродов (рис. 4а), либо между ними (рис. 4б). В первом случае изменение значения КС, вызванное влиянием объекта, наблюдается на геофизической карте в точках измерений, отстоящих не более чем на  $b/2$  относительно истинной границы. Во втором случае точки измерений на геофизической карте принадлежат истинной границе. Ведем допущение, что граница определена корректно, если восстановленная линия границы смещена от истинной не более чем на  $b/2$ . Область допустимых значений *корректной границы* выделена красным контуром (рис. 4). Следовательно, задача восстановления границы объекта в грунте сводится к моделированию линии границы и оценке ее корректности – принадлежности к области допустимых значений.

Для анализа точности восстановленной границы введены оценки относительного отклонения и коэффициента разброса. *Относительное отклонение* восстановленной границы от истинной границы объекта  $\delta$  вычисляется по формуле  $\delta = M/N$ , где  $M$  – количество точек восстановленной границы, не совпавшие с точками истинной границы объекта;  $N$  – общее количество точек, принадлежащих истинной границе объекта. По сути, относительное отклонение является качественной оценкой и характеризует степень соответствия *конфигурации* восстановленной и истинной границ. Данная оценка принимает значения в диапазоне  $[0; 1]$ , возрастание численного значения  $\delta$  соответствует большей степени искажения формы восстановленной границы.

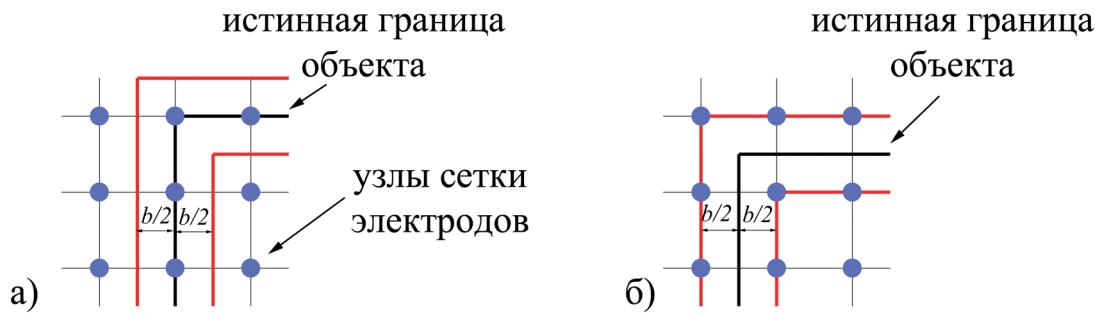


Рис. 4. Взаимное расположение линии границы объекта и сетки электродов

Для численной оценки степени «смещения» восстановленной линии границы за область допустимых значений корректной границы введена оценка – коэффициент разброса  $k$ , который рассчитывается по формуле  $k = DX_1 / DX_2$ , где  $X_1$  – множество элементов массива данных, соответствующих восстановленной границе;  $X_2$  – множество элементов, соответствующих точкам класса «Фон», смежных с граничными точками;  $DX_1$  и  $DX_2$  – дисперсии множеств  $X_1$  и  $X_2$ , соответственно (рис. 5а). Область допустимых значений корректной границы выделена красным контуром. Сдвиг восстановленной границы за пределы области допустимых значений (рис. 5б) определяет увеличение значения коэффициента  $k$ . Это объясняется тем, что при «смещении» восстановленной границы за пределы объекта поиска часть элементов класса «Фон» входит в качестве новых элементов в множество  $X_1$ . С учетом однородности фона это определяет возрастание разброса значений множества  $X_1$  (соответственно – возрастание  $DX_1$ ) при практически неизменном значении  $DX_2$ .

В целом одновременное возрастание обеих оценок соответствует снижению точности при моделировании границы объекта поиска. При использовании алгоритма нечеткой кластеризации Fuzzy c-means к массиву данных КС ошибки могут быть вызваны некорректным заданием порога ( $\alpha$ -сечения) при оценке степени принадлежности текущего элемента массива классу «Фон».

Исследование характера изменения введенных оценок и результатов применения алгоритма Fuzzy c-means проводилось на основе решения прямой задачи электроразведки для однородной среды, содержащей псевдотрехмерные объекты конечной высоты (программа Res3Dmod, Geotomo Software). Компьютерное моделирование проводилось для случая площадного электропрофилирования дипольной экваториальной установкой (шаг измерений и расстояние между питающим и приемным диполями установки совпадают с шагом сетки электродов).

**Модели объектов поиска**

В качестве моделей объектов поиска выбраны прямоугольные призмы с высотой  $4,5b$ , глубиной погружения  $b/2$ , отношением сопротивления объекта и фона 10 ( $b$  – расстояние между электродами сетки):

1. Прямая треугольная призма (далее – треугольная призма), основанием которой является прямоугольный треугольник с катетами  $8b$  и  $5,5b$  (рис. 6а). Примером реального объекта, аналогичного данной модели, является фрагмент фундамента.
2. Цилиндр с диаметром основания  $3b$  (рис. 6б). Например, колодец или яма, заполненная грунтом.
3. Прямоугольный параллелепипед, стороны основания которого повернуты на  $45^\circ$  относительно ориентации линий сетки электродов

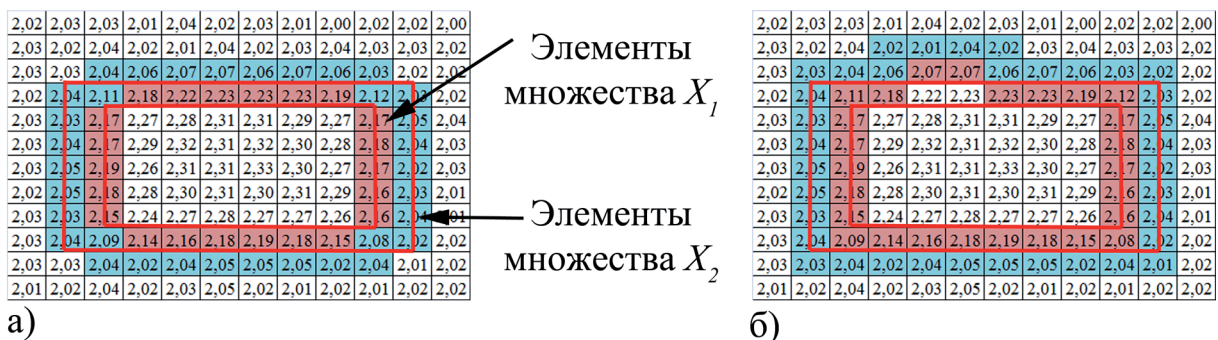


Рис. 5. Расположение элементов массива данных, принадлежащих множествам  $X_1$  и  $X_2$ , для модели прямоугольного параллелепипеда: а)  $\alpha = 0,7$ ; б)  $\alpha = 0,95$

(далее – параллелепипед 1), основание – квадрат со стороной  $6,7b$  (рис. 6в). Данная модель может соответствовать заполненному котловану или фундаменту здания.

4. Прямоугольный параллелепипед, стороны основания которого ориентированы вдоль линий сетки (далее – параллелепипед 2). Основание – прямоугольник со сторонами  $6b$  и  $7b$  (рис. 6г). Данная модель аналогична параллелепипеду 1 при соответствующем повороте сетки электродов.

Линейные участки границ моделей имеют различную ориентацию относительно сетки электродов: параллельные, перпендикулярные и наклонные ( $30^\circ/45^\circ/60^\circ$ ), а также различную конфигурацию: прямая и окружность. В модель добавлен белый шум (5%), так как для полевых наблюдений методом электроразведки максимальная допустимая погрешность измерений составляет 5% [8, с. 8].

Для всех рассматриваемых моделей наблюдается одинаковый характер изменения значения  $\delta$  в зависимости от  $\alpha$ -сечения: начиная с некоторого порогового значения  $\alpha$  наблюдается постепенное (или скачкообразное) возрастание значения  $\delta$  (табл. 1). Сравнение конфигурации истинной и восстановленной границ модельного объекта показало, что увеличение значений  $\delta$  происходит при смещении восстановленной границы за пределы области допустимых значений корректной границы (рис. 4). Соответствующие ячейки в табл. 1 отмечены серой заливкой.

Аналогичная зависимость получена для коэффициента разброса  $k$  (табл. 2, рис. 7): увеличение порогового значения  $\alpha$  при нечеткой кластеризации не обеспечивает восстановление корректной границы. В большинстве случаев при  $\alpha \geq 0,9$  абсолютная погрешность превышает  $b/2$ , что определяет скачкообразный рост коэффициента  $k$ .

Зависимости изменения оценки конфигурации (относительное отклонение  $\delta$ ) и степени «смещения» (коэффициент разброса  $k$ ) восстановленной границы позволяют предположить, что при нечеткой кластеризации данных электроразведки абсолютная

погрешность восстановления границ объекта поиска не превышает величину  $b/2$  при  $\alpha$ -сечении равном 0,5-0,8. Данная тенденция характерна для всех рассмотренных моделей возмущающих тел, отличающихся формой и ориентацией относительно сети электродов.

Применение алгоритма Fuzzy c-means при анализе полевых данных электропрофилирования на археологическом памятнике также показало корректность оценок, полученных в результате компьютерного моделирования. При изменении  $\alpha$ -сечения в диапазоне 0,5-0,8 абсолютная погрешность восстановления границы объекта не превышает половины шага сетки электродов (на рис. 2е восстановленная граница показана зеленым), относительное отклонение  $\delta$  от истинной границы и коэффициент разброса  $k$  равны соответственно 0,5 и 0,98. При  $\alpha \geq 0,9$  восстановленная граница существенно меняет конфигурацию, на некоторых участках «смещается» за пределы области допустимых значений корректной границы (на рис. 2е граница показана желтым), относительное отклонение  $\delta$  составляет 0,75 и коэффициент разброса  $k$  – 3,90. Следовательно, метод нечеткой кластеризации при выборе значения  $\alpha \leq 0,8$  обеспечивает моделирование корректной границы объекта поиска по данным площадного электропрофилирования.

### Заключение

Компьютерное моделирование и результаты обработки данных малоглубинного площадного электропрофилирования, полученных в рамках натурального эксперимента, доказали эффективность разработанного метода определения границы объекта в грунте по результатам нечеткой кластеризации. Обоснованный выбор оптимального  $\alpha$ -сечения обеспечивает заданную точность моделирования границы объекта поиска – максимальная абсолютная погрешность отклонения контура восстановленной границы не превышает половины шага измерений.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 15-06-04239а.

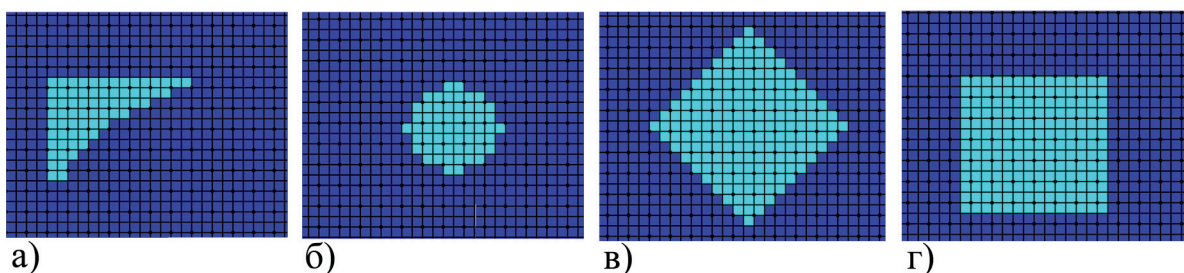


Рис. 6. Горизонтальные сечения моделей: а) треугольная призма; б) цилиндр; в) параллелепипед 1; г) параллелепипед 2

Таблица 1

**Зависимость относительного отклонения  $\delta$  от результатов классификации**

$\alpha$ -сечение	0,5	0,6	0,7	0,8	0,9	0,95
треугольная призма	0,25	0,25	0,25	0,28	0,31	0,40
цилиндр	0,32	0,27	0,27	0,27	0,68	0,73
параллелепипед 1	0,44	0,44	0,44	0,44	0,44	0,50
параллелепипед 2	0,50	0,47	0,47	0,47	0,50	0,53

Таблица 2

**Зависимость коэффициента разброса  $k$  от результатов классификации**

$\alpha$ -сечение	0,5	0,6	0,7	0,8	0,9	0,95
треугольная призма	2,99	2,99	3,67	6,65	10,89	9,91
Цилиндр	1,43	2,44	2,44	2,44	9,17	10,65
параллелепипед 1	0,56	1,39	2,09	4,30	6,14	11,27
параллелепипед 2	1,20	1,25	1,25	2,16	6,15	8,77

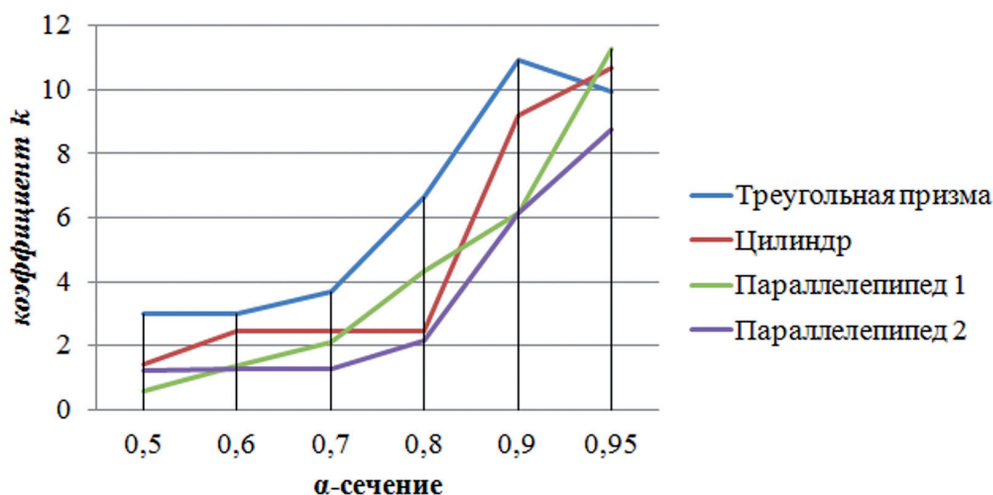


Рис. 7. Зависимость коэффициента  $k$  от  $\alpha$ -сечения для рассматриваемых моделей

**Ключевые слова:** нечеткая кластеризация, алгоритм Fuzzy c-means,  $\alpha$ -сечение, электроразведка, аномалия, граница объекта, моделирование.

**ЛИТЕРАТУРА**

1. Zhurbin I.V. Malyugin D.V. On the method of visualization of electrometric data // Archaeological prospection. – 1998. – Vol. 5. – N 2. – P. 73-79.  
 2. Loke M.H. Tutorial : 2-D and 3-D electrical imaging surveys. – Malaysia : Penang : Geotomo Software, 2012. – 161 p.  
 3. Иванова М.Г., Журбин И.В. Археологические и геофизические исследования средневековых поселений бассейна р. Чепцы // Российская археология. – 2014. – № 1. – С. 40-53.  
 4. Young I.T. Gerbrands J.J., van Vliet L.J. Fundamentals of Image Processing [Электронный ресурс]. – URL: [http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL\\_COPIES/TUDELFT/FIP2\\_3.pdf](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/TUDELFT/FIP2_3.pdf) (дата обращения: 26.04.2015).

5. Miller D., Nelson C., Cannon M., Cannon K. Comparison of Fuzzy Clustering Methods and Their Applications to Geophysics Data // Applied Computational Intelligence and Soft Computing. – 2009. – Vol. 2009. – P. 1-16.  
 6. Тараскина А.С. Нечеткая кластеризация по модифицированному методу c-средних // Проблемы интеллектуализации и качества систем информатики. – Новосибирск : Ин-т систем информатики имени А.П. Ершова СО РАН, 2006. – С. 217-228.  
 7. Markovic M. Joint inversion of magnetotelluric and seismic data for crustal characterization. – Lisboa : Universidade de Lisboa, 2013. – 136 p.  
 8. Марченко М.Н., Станкевич В.И., Терещенко А.Ю. и др. Некоторые вопросы метрологического обеспечения инженерно-геофизических изысканий. Электроразведка методами сопротивления и ВП. – М. : МГУ им. Ломоносова, 2013. – 28 с.