

УСТРАНЕНИЕ ИЗБЫТОЧНОСТИ ДАННЫХ СЕЙСМИЧЕСКОГО МОНИТОРИНГА ГИДРОРАЗРЫВА ПЛАСТА

Е.В. Рабинович

профессор, д.т.н., e-mail: erabinovich1952@gmail.com

П.И. Вайнмастер

аспирант, e-mail: wmaster91@mail.ru

Г.С. Шефель

доцент, к.ф.-м.н., e-mail: ligrlx@yahoo.com

Новосибирский государственный технический университет, Новосибирск, Россия

Аннотация. Важным этапом мониторинга гидроразрыва пластов является построение графической модели трещиноватой зоны. Однако реальные данные, регистрируемые в процессе гидроразрыва, характеризуются большой интенсивностью сейсмической активности в области мониторинга, что затрудняет корректное построение такой модели. Поэтому актуальной является задача устранения избыточности сейсмических данных. В статье рассматривается оригинальный алгоритм, позволяющий устранить избыточность за счёт выделения ограниченного набора локальных гипоцентров сейсмической активности. В основе алгоритма лежат методы кластерного и факторного анализа. Факторный анализ использован для выявления схожих характеристик и сокращения количества переменных, используемых для описания сейсмического источника. Кластерный анализ применён для обнаружения «облаков» сейсмической эмиссии и выделения их локальных гипоцентров. Полученные результаты использованы для построения трёхмерной графической модели трещины.

Ключевые слова: гидравлический разрыв пласта, трещина, кластерный и факторный анализ.

Введение

Гидравлический разрыв пласта (ГРП) — эффективный метод повышения продуктивности нефтяных скважин. Метод заключается в создании искусственной трещины в целевом пласте, которая служит каналом для притока добываемого флюида из продуктивной части пласта в призабойную зону скважины.

Создание трещины осуществляется путём закачки в пласт специальной жидкости с расклинивающим агентом под давлением, значительно превышающим горное давление. Благодаря такому воздействию в пласте происходит образование трещины, которая в процессе проведения ГРП увеличивается в

направлениях пониженной прочности породы. После снятия избыточного давления и прекращения закачки жидкости разрыва трещина остаётся в закреплённом состоянии благодаря расклинивающему агенту (проппанту), создавая высокопроводящий канал для поступления в скважину дополнительного флюида.

Проведение ГРП является дорогостоящей операцией, сопряжённой с риском получения негативных последствий. Так, при превышении критического давления разрыва глинистых экранов возможен гидродинамический контакт добывающей скважины с этими пластами или с нагнетательными скважинами. Такой контакт приводит к резкому обводнению добываемого флюида, вследствие чего скважина становится непригодной к работе. Чтобы избежать риска обводнения, специалистам по работе со скважиной необходимо знать форму и пространственное расположение трещины гидроразрыва.

Существующие математические расчёты не позволяют с достаточной точностью и достоверностью определить параметры раскрывающихся трещин, а бурение исследовательских скважин с отбором кернов является весьма дорогостоящим. Поэтому возникла актуальная необходимость разработки новых средств для определения формы, размеров и динамики образования трещин ГРП.

Одним из таких средств является технология наземного сейсмического мониторинга ГРП, предложенная и описанная в работах [1, 2]. Технология базируется на локации сейсмических сигналов наземной нерегулярной системой приёма и дальнейшей их обработке с целью определения параметров трещины. В данной работе рассматривается один из важных аспектов данной технологии — устранение избыточности микросейсмических данных и определение гипоцентров сейсмической эмиссии.

1. Постановка задачи

Острые трещины, разрываемое при ГРП, имеет размер много меньше расстояния до поверхности приёма. Разрыв породы с высокой степенью достоверности можно считать точечным источником сейсмических импульсов. Это подтверждается регистрацией сейсмоприёмниками возбуждаемых при разрывах пласта волн в виде последовательности сейсмических импульсов.

Регистрация сейсмических импульсов осуществляется наземной нерегулярной системой приёма и представляет собой антенну, состоящую из n одинаковых сейсмоприёмников (регистрирующих датчиков), расположенных в пространстве так, чтобы шумы на их входах были взаимно независимыми. Антенна «сканирует» зону обнаружения, зарегистрированные в ходе гидроразрыва колебания записываются сеймостанцией.

Для определения местоположения сейсмических источников на уровне продуктивного пласта формируется виртуальная трёхмерная сетка с шагом 1 м и с центром, совпадающим с координатами забоя. От каждого узла сетки вычисляются времена распространения сейсмической волны до каждого сейсмоприёмника по формуле:

$$t_k = \frac{\sqrt{h^2 + L_k^2}}{v}, k = 1, \dots, n,$$

где h — расстояние от узла до поверхности, L_k — расстояние между проекцией источника на поверхность земли и k -м сейсмоприёмником, v — средняя скорость распространения сейсмической волны в среде.

С целью фокусировки антенны на каждом узле виртуальной сетки и повышения отношения сигнал / шум осуществляется накопление энергии сейсмотрасс, записанных всеми приёмниками, с учётом рассчитанных временных сдвигов (задержек). Каждая суммарная сейсмограмма характеризует сейсмический сигнал, полученный от одного узла виртуальной трёхмерной сетки за некоторый временной отрезок проведения ГРП. На основе полученных данных осуществляется обнаружение искомого сейсмического импульса с помощью квадратурного корреляционного [3] или статистического [4] алгоритмов и, если импульс найден, сопоставление данного узла сетки с точечным сейсмическим источником. Каждый такой источник характеризуется координатами X , Y в горизонтальной плоскости пласта, глубиной расположения источника H , максимальной амплитудой колебания A и моментом времени T , в который была зарегистрирована максимальная амплитуда.

Однако реальные данные, регистрируемые в ходе гидроразрыва, характеризуются большой интенсивностью сейсмической активности во всём объёме области мониторинга. Такая избыточность может быть вызвана резонансными свойствами среды и повторным возникновением сейсмических сигналов в тех областях, где уже зарегистрирована трещина при увеличении её толщины. Избыточность сейсмических данных затрудняет корректное построение графической модели трещины. Поэтому актуальной является задача устранения избыточности микросейсмических данных, в частности — выделение гипоцентров сейсмической активности среди набора импульсных источников.

2. Теория

В процессе распространения трещины, как правило, образуются локальные «облака» сейсмической эмиссии, включающие в себя набор близко расположенных источников со сходными характеристиками. Построение графической модели трещины может обеспечить знание координат гипоцентров таких облаков сейсмичности.

В данной работе предложен оригинальный алгоритм определения координат гипоцентров, основанный на применении методов факторного и кластерного анализа.

Кластерный анализ является способом разбиения многомерных объектов на однородные группы. С его помощью осуществляется выделение «облаков» импульсных сейсмических источников. Центры «масс» кластеров используются в качестве гипоцентров, набор которых является достаточным для дальнейшей обработки. Таким образом, кластерный анализ устраняет избыточность наблюдений, мешающую построению адекватной графической модели.

Факторный анализ, в свою очередь, является средством сокращения количества переменных, используемых для описания объекта. С его помощью осуществляется предварительная подготовка данных к кластеризации, заключающаяся в устранении влияния схожих характеристик, искажающих результат кластерного анализа, и ускорении процедуры обработки данных.

Как показано в [5], схожие характеристики при кластеризации получают больший вес, т. к. расстояния между кластерами вычисляются из разностей по каждой переменной. Вследствие этого результат кластеризации искажается: преимущественно учитываются схожие параметры с большим весом. В исследуемой задаче положение сейсмического источника задаётся координатами (X, Y, H) , влияние которых может подавить другие параметры объекта (A, T) и исказить итоговый результат. Для устранения такого негативного воздействия применён факторный анализ.

Ускорение обработки данных особенно актуально при работе с сейсмическими сигналами, т. к. количество источников сейсмических импульсов составляет десятки тысяч, а их обработка должна осуществляться в режиме «реального времени». Применён критерий «отсева» факторов, который позволяет, с одной стороны, оставить только набор значимых факторов (содержащих полезную информацию), а с другой — отсеять факторы, принципиально не влияющие на результаты исследования.

В рамках данной работы также исследован другой вариант применения факторного анализа — для устранения выбросов на завершающем этапе алгоритма (после кластеризации).

Для того чтобы осуществить предварительную обработку сейсмических данных в соответствии с описанными требованиями, был использован вариант факторного анализа — метод главных компонент (МГК). Данный метод используется во многих прикладных задачах и зачастую применяется в «тандеме» с кластерным анализом [5].

Суть метода заключается в следующем. Совокупность исследуемых объектов представляется в виде точек в n -мерном пространстве, где n — число исследуемых признаков. Осуществляется построение прямой линии, «пронизывающей» облако точек так, что проекция точек на неё имеет максимальное среднеквадратичное отклонение (СКО). Так образуется первая главная компонента. На следующем шаге определяется новая прямая, ортогональная первой и также максимизирующая СКО проекции точек — вторая главная компонента. Процесс продолжается до тех пор, пока количество прямых не достигнет размерности пространства n . Таким образом осуществляется переход к новой системе координат, где переменными являются факторы — линейные комбинации исходных признаков объекта.

Для оценки качества результатов факторного анализа мы руководствовались двумя основными характеристиками. Первая характеристика — это «объяснённая» фактором дисперсия [6] — доля от общей дисперсии данных, которую составляет дисперсия фактора. Объяснённая дисперсия характеризует разброс данных вдоль координатной оси, заданной фактором: чем она выше, тем больше информации об исследуемом объекте несёт фактор. Вторая важная характери-

стика — матрица факторных нагрузок, содержащая корреляции между факторами и исходными переменными. На её основе делается вывод о связи между новыми переменными (факторами) и исходными параметрами объекта.

Как уже упоминалось, каждый сейсмический источник характеризуется пятью параметрами: X , Y , H , A , T . В результате применения МГК были предварительно определены пять новых факторов, среди которых проводилась процедура отсева. Для этого был использован критерий Кайзера: если величина дисперсии фактора меньше дисперсии исходной переменной, то фактор «отсеивается» как помеховая составляющая. Важно отметить, что предварительно данные были стандартизованы так, что среднее значение каждой переменной стало равным 0, а дисперсия — 1. Поэтому, согласно критерию Кайзера, отсеивались факторы с дисперсией меньшей единицы.

Из пяти факторов были отсеяны два, поскольку их дисперсия оказалась 0,74 и 0,75. Три оставшихся фактора, дисперсия которых превысила единицу, были сохранены. Дисперсия этих факторов оказалась выше дисперсии исходных переменных, благодаря тому, что соответствующие оси лучше описывают разброс данных. Иными словами, СКО проекции данных на оси факторов оказалось больше, чем СКО проекции данных на оси исходных переменных.

В результате, был осуществлён переход от множества переменных (X , Y , H , A , T) к множеству факторов (F_1 , F_2 , F_3).

Матрица факторных нагрузок для факторов F_1 , F_2 , F_3 приведена в таблице 1. Жирным шрифтом выделены корреляции (нагрузки), превышающие по модулю величину «порога достоверности» — минимального значения корреляции, при котором корреляция считается значимой (0,7).

Таблица 1. Факторные нагрузки

Переменная/фактор	F_1	F_2	F_3
X	-0,776800	0,164289	0,026227
Y	-0,751429	-0,150475	0,025240
H	0,001522	0,003128	0,999430
A	0,253022	-0,773072	0,004291
T	-0,306739	-0,715682	0,014660
Дисперсия фактора	1,26174	1,159484	1,050418
Доля объяснённой дисперсии (%)	26,5235	23,1897	21,0084

Отрицательные корреляции говорят о том, что ось соответствующей главной компоненты (фактора) и исходная ось направлены в разные стороны. Подобный эффект проявляется при ортогональном вращении факторов в рамках МГК, и никаких выводов на основании знака корреляции делать не следует.

На рис. 1 приведён трёхмерный график факторных нагрузок, построенный на основе таблицы 1. Он отражает расположение исходных переменных в новой системе координат, заданной факторами. Координаты по осям $Factor_1$, $Factor_2$

и $Factor_3$ представляют собой величину корреляции между переменными и соответствующими факторами.

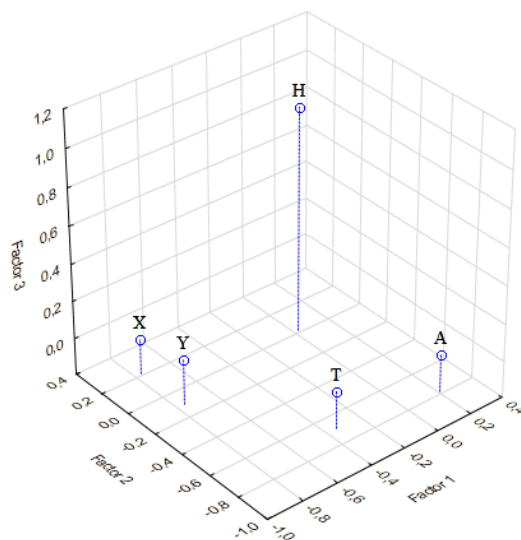


Рис. 1. График факторных нагрузок

Для оценки результатов факторного анализа были проанализированы дисперсии, объяснённые факторами, и факторные нагрузки (табл. 1).

Фактор F_1 имеет наибольшее значение дисперсии — он объясняет 26,5 % от общей дисперсии выборки, т. е. сильнее всего влияет на разброс данных. При этом высокие нагрузки у этого фактора (0,78 и 0,75 соответственно) обнаружены с переменными X и Y , определяющими координаты точечного источника внутри трёхмерной сетки пласта. Таким образом, первый фактор характеризует пространственное положение точечного сейсмического источника.

Фактор F_2 в свою очередь наиболее коррелирован с переменными A и T (соответствующие коэффициенты корреляции равны 0,77 и 0,72). Можно сделать вывод, что данный фактор характеризует параметры динамики импульса, возбуждаемого источником (момент достижения максимума колебания и соответствующую амплитуду). Эти параметры действительно взаимосвязаны: области трещиноватой зоны, сформированные в близкие моменты времени, как правило, имеют схожие значения амплитуды колебаний, что и объясняет высокие коэффициенты корреляции между характеристиками. Доля объяснённой фактором F_2 дисперсии составила 23,19 % от общей дисперсии данных, т. е. он является вторым по значимости влиянием на разброс данных (после фактора F_1).

Наконец, по фактору F_3 значимая корреляция была обнаружена только с переменной H , определяющей глубину, на которой располагается рассматриваемая плоскость внутри пласта (0,99). Т. к. величина корреляции оказалась близка к 1, можно сделать вывод, что выделенный фактор почти полностью повторяет исходную переменную и не подвергся изменениям при обработке. Это подтверждается и тем, что доля дисперсии фактора F_3 от общей дисперсии

данных составила 21 %, или почти $1/5$, — т. е. доля сопоставима с вкладом одной переменной из пяти, присутствующих до обработки.

Суммарная доля дисперсии, объяснённой тремя выделенными факторами F_1 , F_2 и F_3 , достигла 71 % от общей дисперсии исходного набора данных. Применение критерия Кайзера позволило отсеять два «незначимых» фактора, которые объясняли оставшиеся 29 % от общей дисперсии. В рамках МГК такие факторы интерпретируются как помеховая (шумовая) составляющая, не содержащая полезной информации об исследуемом объекте. При этом размерность данных сократилась на 40 % (с пяти признаков до трёх).

Таким образом, в результате применения факторного анализа было достигнуто предварительное устранение избыточности, позволяющее ускорить следующий далее процесс кластеризации.

Учитывая вышесказанное, факторы F_1 , F_2 , F_3 были выбраны в качестве замены исходных переменных при дальнейшей обработке.

Для выделения «облаков» сейсмичности были применены средства кластерного анализа. Очевидным вариантом представлялось выделение кластеров среди всей области локации, однако такое решение имеет серьёзные недостатки. В ходе экспериментов было обнаружено, что при таком подходе центры масс кластеров оказываются сконцентрированными близко к забою скважины, а удалённые области не получают представления. Альтернативным решением является разбиение области локации на множество вспомогательных областей меньшего размера и выделение естественных кластеров внутри каждой из них. Так как вспомогательные области распределены по всему объёму исследуемой области (и, особенно, на границах трещиноватой зоны), упомянутый недостаток удалось устранить.

Вспомогательные области были представлены в виде параллелепипедов, имеющих управляемые параметры — длину, ширину и высоту, и покрывающих всю область локации. Чтобы избежать потери информации, которая может находиться на границе параллелепипедов, параллелепипеды частично перекрывают друг друга. Пример такого разбиения области локации представлен на рис. 2. Чёрные точки внутри представляют точечные источники сейсмических импульсов, а трубка в центре — скважину.

В качестве метода кластеризации был использован агломеративный алгоритм, заключающийся в последовательном объединении точек в кластеры и построении кластерного дерева. Важным этапом является выбор метрики для расчёта парных расстояний между объектами, а также стратегии кластеризации, использование которых окажется наиболее эффективным в рамках рассматриваемой предметной области.

Были исследованы два метода расчёта расстояния между объектами: метод наиболее удалённых соседей (Complete-link) и метод попарного среднего (Average-link).

Метод удалённых соседей определяет межклассовое расстояние как наибольшую дистанцию между объектами двух кластеров. Формула для расчёта дистанции $d(r, s)$ между двумя кластерами r и s имеет следующий вид:

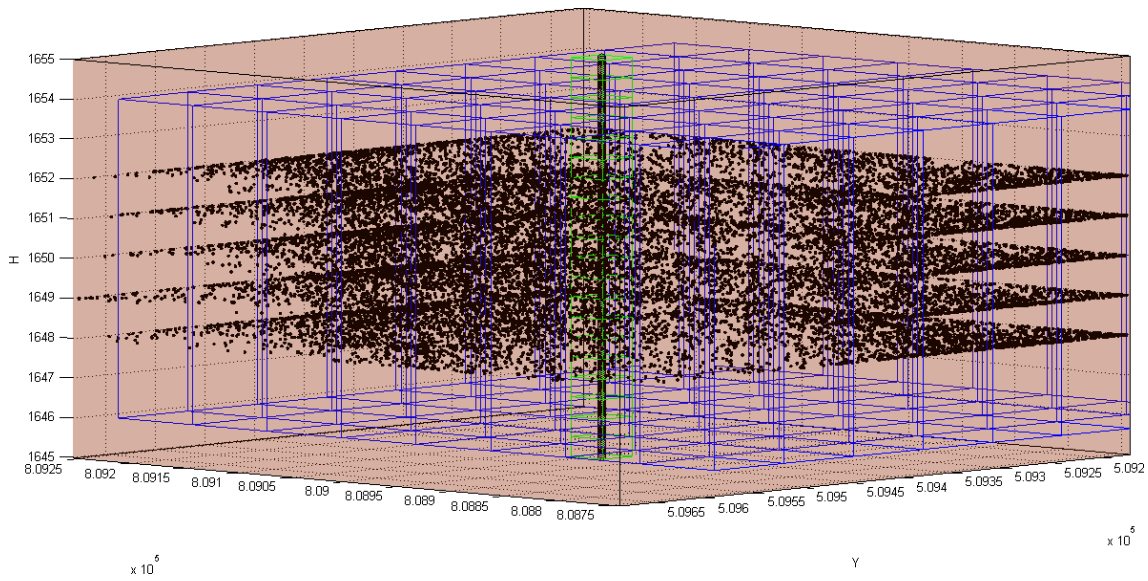


Рис. 2. Разбиение области локации на вспомогательные области (параллелепипеды)

$$d(r, s) = \max[\text{dist}(X_{ri}, X_{sj})],$$

где $i = 1, \dots, n_r$, $j = 1, \dots, n_s$.

Таким образом, при построении кластерного дерева дистанция между двумя кластерами рассчитывается по максимальному расстоянию между их объектами. Такой подход положительно сказывается на охвате всей области локации потому, что способствует формированию кластеров на границах трещиноватой зоны, т. к. расстояние между ними в этом случае максимально.

Метод невзвешенного попарного среднего, в свою очередь, является наиболее универсальным и определяет расстояние $d(r, s)$ между двумя кластерами как среднее расстояние между всеми парами объектов в них:

$$d(r, s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} \text{dist}(x_{ri}, x_{sj}).$$

В отличие от метода удалённых соседей, он не способствует выделению кластеров в приграничных зонах области локации, но, тем не менее, распределяет кластеры наиболее равномерно.

Для сравнения описанных методов была проведена серия экспериментов. Смоделирована тестовая выборка источников сейсмической активности, предварительно разделённых на несколько групп по различным параметрам: в первом опыте — по пространственным характеристикам, представленным факторами F_1 и F_3 , во втором — по характеристикам возбуждаемого сейсмического импульса, представленным фактором F_2 , а в третьем — по совокупности всех параметров. Для каждого из опытов были рассмотрены два случая: когда количество классов равно трём и данные в них расположены «компактно» (так,

что классы хорошо различимы между собой) и когда количество групп равно десяти, а данные в некотором смысле «размыты» между группами.

В качестве меры для расчёта расстояния между точками были использованы Евклидова мера и мера Чебышёва. Мера Евклида представляет расстояние между двумя точками x и y в виде геометрического расстояния в многомерном пространстве:

$$d(X_j, X_i) = \sqrt{\sum_{k=1}^N (x_{ki} - x_{kj})^2}.$$

Мера Чебышёва, в свою очередь, определяет расстояние как максимальную разность между двумя аналогичными параметрами объектов:

$$d(X_j, X_i) = \sum_{k=1}^N |x_{ki} - x_{kj}|.$$

Результаты сравнения сведены в таблицу 2.

Таблица 2. Результаты сравнения методов кластеризации

Метод	Мера расстояния	Среднее кол-во кластеров при «компактном» разбиении	Среднее кол-во кластеров при «размытом» разбиении	Средний % объектов, верно отнесённых к кластерам
Complete-Link	Euclidean	4	16	68%
Complete-Link	Chebyshev	3	12	92%
Average-Link	Euclidean	3	11	94%
Average-Link	Chebyshev	4	14	81%

Из таблицы видно, что результаты, наиболее близкие к исходному разбиению (3 группы при «компактном» разбиении и 10 групп при «размытом»), были получены во втором и третьем опытах. Здесь также наиболее высоким оказался средний процент объектов, верно отнесённых к кластерам — 92 % и 94 % соответственно. Также можно сделать вывод, что метод Completelink лучше работает с мерой Чебышёва, а метод Averagelink — с Евклидовой мерой.

Учитывая, что наилучшие результаты кластеризации были получены в третьем эксперименте, в качестве стратегии кластеризации был выбран метод Averagelink, а в качестве меры расстояния между объектами — мера Евклида.

Стоит заметить, что в некоторых случаях были обнаружены «лишние» (по сравнению с исходным разбиением объектов) кластеры. Чтобы избежать их появления, был введён критерий оптимального разбиения, описываемый формулой:

$$d_{\min}(C_i, C_j) > \max\{h(C_i), h(C_j)\},$$

где $d_{\min}(C_i, C_j)$ — минимальное расстояние между парой объектов из двух кластеров C_i и C_j ; $\max\{h(C_i), h(C_j)\}$ — максимальное из расстояний среди объектов одного кластера. Оптимальным считается то количество кластеров, при котором все кластеры удовлетворяют описанному условию. Подробное рассмотрение критериев оптимального кластерного разбиения является темой отдельной статьи и описано в [7].

В результате кластеризации был получен набор локальных «облаков» сейсмичности. Для каждого из них был найден гипоцентр — центр масс соответствующего кластера (в пространстве исходных характеристик объектов — X, Y, H, A, T). В результате удалось превратить исходное количество узлов виртуальной сетки локации, представляющих источники сейсмических импульсов (приблизительно 20000), в набор гипоцентров (приблизительно 700), используемых для графического построения модели зоны трещиноватости. При этом каждый узел характеризуется пятью параметрами, а каждый гипоцентр — тремя, что в совокупности приводит к сокращению объёма обрабатываемых данных почти в 50 раз.

На завершающем этапе был вновь применён факторный анализ — теперь к полученному набору гипоцентров, чтобы проанализировать изменение факторных нагрузок после кластеризации, а также выявить и устранить выбросы в наборе гипоцентров.

Таблица 3. Факторные нагрузки после обнаружения гипоцентров

Переменная/фактор	F_1	F_2	F_3
X	0,781574	0,186051	0,206770
Y	0,750792	-0,185946	-0,003375
H	0,052999	0,016990	-0,973803
A	0,205652	-0,797268	0,161902
T	0,224919	-0,737341	-0,044297
Дисперсия фактора	1,273637	1,240787	1,019232
Доля объяснённой дисперсии (%)	25,4727	24,8157	20,3846

Сравнение табл. 3 с табл. 1 показывает, что количество факторов не изменилось. Не изменились и корреляции факторов с исходными переменными. Так, у фактора F_1 высокие нагрузки обнаружены с переменными X и Y (пространственные характеристики); у фактора F_2 — с переменными A и T (параметры импульса), у фактора F_3 — с H , близкая к единице (глубина трещины). При этом нагрузки между фактором F_2 и переменными A и T незначительно выросли (на 0,02), а нагрузка между фактором F_3 и переменной H уменьшилась (также на 0,02). Процент дисперсии, объяснённой новыми факторами, остался тем же — 71 %.

После обработки, заменившей избыточные данные, представляющие «облака» сейсмичности нечётких очертаний, набором ограниченного числа гипоцентров, сохранились все существенные зависимости между переменными, характеризующими источники сейсмических импульсов.

Для устранения выбросов по каждому из факторов был применён критерий Шовене. Всего было найдено 6 выбросов, связанных с фактором F_2 , что составляет 0,01 % от общего количества данных. Проанализировав характеристики соответствующих гипоцентров, было обнаружено, что выбросы связаны с высокими значениями амплитуды в этих узлах сетки локации.

3. Результаты экспериментов

Трёхмерная модель трещины, построенная с помощью триангуляции Делоне на исходных данных, полученных в результате применения описанного алгоритма, приведена на рис. 3. Область локации представляет параллелепипед с квадратным основанием со сторонами по 512 м и высотой 2 м, трещина ориентирована в нем наклонно.

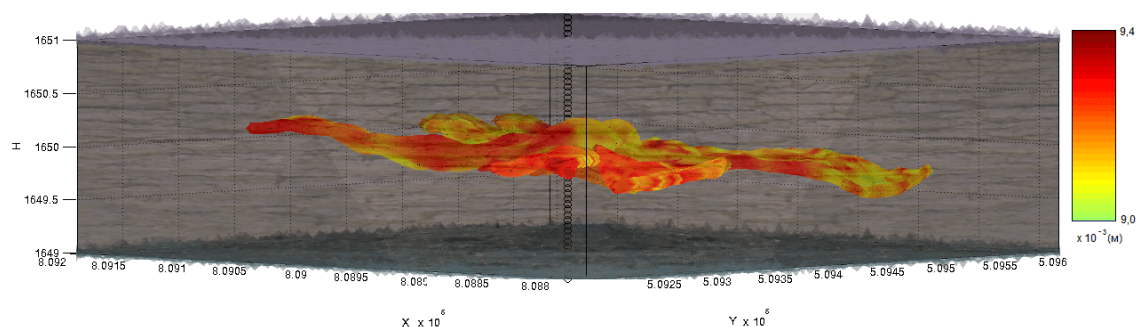


Рис. 3. Трёхмерная модель зоны трещиноватости целевого пласта

Фрагменты трещины имеют раскраску, отображающую амплитуду сейсмических импульсов, которая даёт косвенную оценку толщины закреплённой трещины в целевом пласте.

На рисунках 4 и 5 приведены примеры результатов кластеризации в рамках одной из «вспомогательных» зон (параллелепипедов). Рисунок 4 демонстрирует разбиение на кластеры до применения факторного анализа (т. е. кластеризацию по переменным X , Y , H , A и T , предварительно нормированным), а рисунок 5 — после применения факторного анализа (кластеризацию по факторам F_1 , F_2 и F_3). Оба графика приведены в системе координат (X, Y, H) для визуального сравнения результатов кластеризации.

Из рисунка 4 видно, что в первом случае получились «размытые» кластеры, точки которых рассредоточены по всей области параллелепипеда. Этот результат может быть объяснён тем, что из-за ограниченной разрешающей способности сейсмической антенны, кластеризации подвергались гипоцентры с неточно определёнными характеристиками. Кроме того, имеет место влияние случайных шумов и помех на параметры исходных точечных источников.

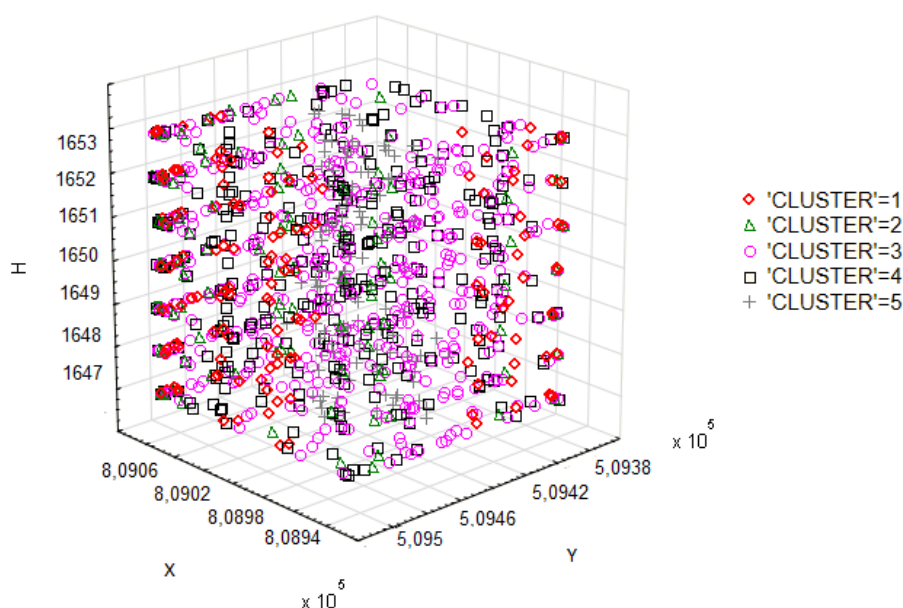


Рис. 4. Результат кластеризации по исходным признакам X , Y , H , A , T

После предварительного выделения факторов и кластеризации на их основе (рис. 5) результат стал более компактным и имеет лучшее разрешение. Можно сделать вывод, что два фактора, «отсеянные» по критерию Кайзера на этапе факторного анализа, в большой мере описывали шумовые характеристики, которые влияли на результат кластеризации в предыдущем эксперименте. Устранив их, удалось получить кластеры, гипоцентры которых могут быть более чётко интерпретированы как источники сейсмических импульсов.

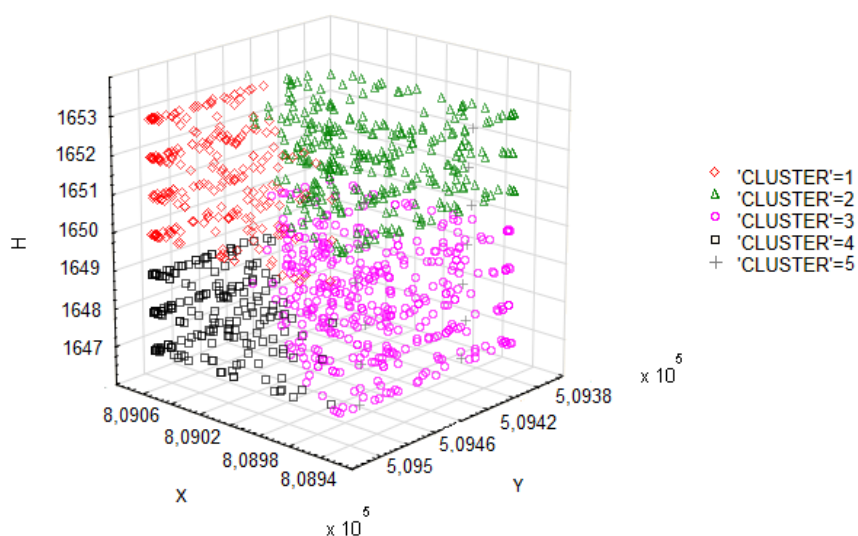
Стоит также отметить, что предварительная фильтрация данных с помощью факторного анализа позволила сократить время кластеризации на 25–30 % (за счёт сокращения количества признаков точечных источников, использованных в обработке).

4. Заключение

В работе предложен оригинальный алгоритм устранения избыточности сейсмических данных при мониторинге ГРП, обеспечивающий определение координат гипоцентров сейсмической активности. В рамках алгоритма применены, исследованы и адаптированы к решаемой задаче методы факторного и кластерного анализа.

Использование факторного анализа (в качестве предварительной обработки данных перед кластеризацией) позволило сократить количество анализируемых признаков объекта с пяти до трёх. В результате были устранены помеховые составляющие, искажающие результат кластеризации, и достигнуто значительное ускорение процесса выделения кластеров (на 25–30 %).

Применение кластерного анализа, в свою очередь, позволило сократить из-

Рис. 5. Результат кластеризации по факторам F_1 , F_2 , F_3

быточность исходных данных путём выделения компактных «облаков» сейсмичности и определения их центров «масс» в качестве единичных гипоцентров микроземлетрясений. Применение кластерного анализа позволило сократить избыточность представления данных почти в 50 раз.

Применение факторного анализа на завершающем этапе показало, что в результате обработки структурные характеристики сейсмических данных не изменились, и все значимые корреляции между переменными были сохранены.

Наконец, использование критерия Шовене позволило устранить выбросы среди гипоцентров, связанные с аномальным значением амплитуды импульса (кол-во выбросов составило 0,01 % выборки).

На основе выделенных гипоцентров с помощью метода Делоне была построена трёхмерная графическая модель трещиноватой зоны пласта.

ЛИТЕРАТУРА

1. Новаковский А.Ю., Рабинович Е.В. Сейсмическая измерительная система для локализации трещин при гидроразрыве нефтяного пласта // Идентификация, измерение характеристик и имитация случайных сигналов (состояние, перспективы развития): сборник материалов конференции. 2009. С. 171–174.
2. Рабинович Е.В., Туркин А.С. Наземная локация микросейсмических сигналов для мониторинга гидравлического разрыва пласта // Доклады ТУСУР. 2012. № 1(25), Ч. 1. С. 104–112.
3. Correlation detection of seismic pulses for hydraulic fracturing monitoring / E.V. Rabinovich [and others] // 16 International conference of young specialists on micro/nanotechnologies and electron devices (EDM). Altai, Erlagol, 2015. P. 162–165.
4. Statistical detection of seismic pulses for hydraulic fracturing monitoring / E.V. Rabinovich [and others] // 11 International forum on strategic technology (IFOST 2016). Novosibirsk, 2016. Pt. 1. P. 375–378.

5. Юдин Г.Б. Кластеры на факторах: proetcontra // Социологические методы в современной исследовательской практике. М. : ГУ-ВШЭ, 2007. С. 296–302.
6. Иберла К. Факторный анализ. М. : Статистика, 1980. 398 с.
7. Пространственная фильтрация сейсмических событий, возникающих при гидравлическом разрыве пласта / Е.В. Рабинович [и др.] // Автоматика и программная инженерия. 2013. № 1. С. 53–60. URL: <http://www.jurnal.nips.ru/sites/default/files/Paper-2013-1-7.pdf> (дата обращения: 28.05.2017).

ELIMINATION OF DATA REDUNDANCY IN SEISMIC MONITORING OF HYDRAULIC FRACTURING

E.V. Rabinovich

Dr.Sc. (Eng.), Professor, e-mail: erabinovich1952@gmail.com

P.I. Vaynmaster

Postgraduate Student, e-mail: wmaster91@mail.ru

G.S. Shefel

Ph.D. (Phys.-Math.), Associate Professor, e-mail: ligr1x@yahoo.com

Novosibirsk State Technical University, Novosibirsk, Russia

Abstract. The important stage of hydraulic fracturing monitoring is constructing a graphical model of fracturing area. However, the real data recorded in the process of hydraulic fracturing are characterized by high intensity of seismic activity in the monitoring area, which hinders correct constructing of such model. Therefore, the problem of eliminating the redundancy of seismic data is important. The original algorithm is considered in this article, which allows eliminating redundancy by detection of local hypocenters of seismic activity. The algorithm is based on the methods of cluster and factor analysis. Factor analysis is used to identify similar characteristics and reduce the number of variables used to describe a seismic source. Cluster analysis is used to detect clouds of seismic emission and find their local hypocenters. The results obtained are used to construct a three-dimensional graphic fracture model.

Keywords: hydraulic fracturing, fracture, factor and cluster analysis.

Дата поступления в редакцию: 21.06.2017