

Содержание

1 Введение	3
1.1 Зачем нужны промыслово-геофизические исследования	3
1.2 Оборудование для каротажа	3
1.2.1 Стандартные датчики	4
1.2.2 Акустическая шумометрия	4
1.3 Анализ данных ПГИ	5
1.4 Решаемые в исследовании задачи	6
2 Часть 1. Локализация зон притока	8
2.1 Актуальность и обоснование выбора темы	8
2.2 Используемые методы	8
2.2.1 Коррекция увязки по глубине	8
2.2.2 Выделение точек изменения тренда	14
2.2.3 Построение разрывных трендов профиля	17
2.2.4 Примеры результатов построения профилей	18
2.2.5 Модель слабого контроля	20
2.3 Результаты	22
2.3.1 Поиск событийных участков	22
2.3.2 Детекция притока по данным температуры оптического кабеля .	30
2.3.3 Детекция и классификация зон притока на тестовой скважине .	33
3 Часть 2. Определение фазового состава флюидов вдоль ствола скважины по данным ПГИ	37
3.1 Актуальность и обоснование выбора темы	37
3.2 Описание данных	37
3.2.1 Осевой датчик фоновой температуры	38
3.2.2 Азимутально распределенный термоанемометр	38
3.3 Формулировка задачи	39
3.3.1 Физическая постановка задачи	39
3.3.2 Математическая постановка задачи	39
3.4 Используемые методы	40
3.4.1 Коррекция фоновой температуры	40

3.4.2	Анализ данных распределенного термоанемометра	42
3.5	Результаты	51
3.6	Обсуждение	54
4	Заключение	55
5	Список литературы	56
6	Приложение	58

1 Введение

1.1 Зачем нужны промыслово-геофизические исследования

Промыслово-геофизические мероприятия проводятся для оптимизации процесса разработки месторождения с целью увеличения добычи нефти. ПГИ могут проводиться как при работающей скважине, так и при остановленной. Иногда для сбора достаточно-го для анализа количества данных проводится несколько исследований на различных этапах работы скважины.

Одна из основных задач ПГИ - локализация зон притока флюида в скважину. Как правило, при этом решается еще и задача определения типа флюида - вода, нефть, газ или смесь флюидов - и определение характеристик флюида - плотность и скорость.

Одно из наиболее эффективных мероприятий для увеличения добычи углеводородов - проведение многостадийного гидроразрыва пласта [1]. В скважину под большим давлением (до 100МПа) закачивается жидкость разрыва высокой вязкости (на водной или нефтяной основе). Под влиянием давления жидкости разрыва пласт расщепляется. После этого в скважину закачивается жидкость-песконоситель при давлении, позволяющем удержать скважину в открытом состоянии. Жидкость-песконоситель - более вязкая, чем жидкость разрыва, и содержит проппант - песок или другие гранулы диаметром обычно порядка 0.5-1.2мм. Шарики проппанта затекают в образовавшиеся трещины. При откачке жидкости давление снижается, но образовавшиеся трещины не схлопываются полностью из-за наличия в них шариков проппанта, при этом часть проппанта разрушается, а часть вдавливается в породу. Раскрытые трещины позволяют значительно увеличить прирост нефти.

ПГИ могут проводиться при проведении многостадийного гидроразрыва пласта [2]. Анализ данных ПГИ позволяет оценить качество полученных трещин и локализовать зоны притоков в скважине.

Другая часто решаемая задача - поиск неисправностей в структуре скважины, то есть нежелательных трещин, или утечек [3].

1.2 Оборудование для каротажа

В ходе ПГИ в скважину опускается каротажная сборка приборов, которая проводится по всей глубине скважины и непрерывно пишет данные. Приборы могут состоять из одного или нескольких датчиков, измеряющих различные физические величины.

1.2.1 Стандартные датчики

Часто используемый для ПГИ прибор PLT-9.2 [4] содержит в себе датчики температуры, давления, влагомер и расходомер. Для привязки показаний прибора к траектории скважины используется локатор муфт и датчик уровня естественного гаммаизлучения [5].

1.2.2 Акустическая шумометрия

В рамках измерения спектральный шумомер записывает акустические сигналы в скважине в широком диапазоне частот с хорошим разрешением: анализируемые в данной работе данные пассивной акустической шумометрии были получены приборами с диапазоном измерений от 8Гц до 60кГц и от 8Гц до 100кГц, записанными в 512 или 1024 частотных каналов с высокой чувствительностью. Каждой точке траектории, в которой производились измерения, сопоставляется запись акустического шума порядка 0.1сек с частотой дискретизации 200кГц, которая для анализа обычно преобразуется в вектор, компоненты которого соответствуют интенсивности сигнала в определенном диапазоне частот. Запись показаний прибора может проводиться как при движении прибора, так и на остановках (например, каждые 2 метра) - чтобы избежать дорожного шума, который может скрыть низкочастотные сигналы малой интенсивности.

Акустический шум появляется при турбулентном движении флюида по стволу скважины.. Согласно [6], [2], мощность акустического излучения W при таком движении зависит от параметров трубы и флюида:

$$W \sim \frac{\rho U^8 d^2}{a_0^5}, \quad (1)$$

Где ρ – средняя по сечению трубы плотность флюида, a_0 – скорость звука во флюиде, U – скорость потока и d – диаметр трубы, который задает характерный масштаб турбулентности.

Так как мощность зависит от скорости в восьмой степени, шумометрия - очень чувствительный способ измерения скорости потока. Это позволяет достичь хороших результатов в случае однофазного течения; в случае смеси флюидов меняются и другие компоненты формулы (1) - плотность флюида и скорость звука во флюиде, поэтому показания шумометрии нужно рассматривать в комбинации с остальными показаниями датчиков.

Показано [7], [8], что сочетание пассивной акустической шумометрии с классическими приборами геофизических исследований позволяет улучшить результаты анализа

данных ПГИ.

В частности, одно из достоинств спектральной шумометрии заключается в том, что с ее помощью можно выделить и дифференцировать заколонные перетоки - то есть потоки флюида между обсадной колонной и породой, которые невозможно определить с помощью, например, расходомера. Выявление заколонных перетоков важно для проверки технического состояния скважины, определения расположения продуктивных пластов, оценки эффективности гидроразрыва и т.п.

Также чувствительность современных приборов спектральной шумометрии позволяет уловить притоки со скоростями ниже чувствительности обычных расходомеров, которые при таких скоростях просто не раскручиваются. Поэтому акустическая шумометрия особенно важна в горизонтальных и около-горизонтальных малодебитных скважинах.

Последнее время актуальными становятся измерения акустической шумометрии и температуры с использованием оптоволоконных кабелей – технологии DTS (distributed temperature sensing) и DAS (distributed acoustic sensing) [9].

1.3 Анализ данных ПГИ

Анализ полученных в результате каротажных исследований данных может проводиться как в ручном режиме, так и с использованием написанных экспертами программ. Невозможность непосредственно измерить такие важные показатели, как пористость и плотность пластов, делают очень сложной задачу численного моделирования скважины для решения задачи с высокой точностью, поэтому обычно используется ручной анализ и/или численное моделирование [10].

Рассмотрим для примера задачу локализации зон притока с использованием данных спектральной акустической шумометрии.

При наличии данных пассивной акустической шумометрии «ручной» способ анализа данных обычно предполагает построение спектрограммы в смысле изменения спектра сигнала при продвижении вдоль траектории скважины. Изменения спектра сигнала могут указывать на наличие зон притока.

В [8] описаны способы локализации зон притока в малодебитных горизонтальных скважинах с использованием спектральной акустической шумометрии. Описаны некоторые характеристики спектрограмм, позволяющие выделить события: так, поток флюида через трубу большого диаметра генерирует низкочастотный шум, а через трубу меньшего диаметра - высокочастотный [11]; также поток через пористую среду генери-

ирует высокочастотный шум, а течение через естественную или искусственно созданную скважину - низкочастотный. Для избавления от статистически не значимого шума, вызванного движением прибора или фоновым шумом скважины, используются различные техники фильтрации шума, например, на основе вейвлет-преобразования [11]. В целом притоки углеводородов можно выделить по широкому спектру производимого акустического шума.

1.4 Решаемые в исследовании задачи

В данной работе рассмотрены две задачи: локализация зон притока в скважину и определение фазового состава флюида по данным термоанемометрии.

Задача локализации зон притока решается на основе задачи выявления событийных участков в скважине. Событийными участками могут быть притоки воды, нефти или газа, а также изменения показаний приборов, связанные с вариацией траектории скважины. Это могут быть удары каротажной сборки приборов о стенки скважин, изменения показаний из-за стыка труб, или, в случае горизонтальных скважин, индикаторы стратификации жидкости не в местах наличия притоков - в частности, могут существовать места скопления воды на вогнутых по отношению к горизонту участках траектории и газовые облака на выпуклых участках. Так как задача разделения участков на зоны притока и зоны стратификации жидкостей, не связанных с притоком, требует дальнейших кропотливых исследований, в том числе учета мнения экспертов, в данной работе она не рассмотрена.

Тем не менее, как будет показано далее, предложенный для решения задачи выявление событийных участков подход может быть доработан до семейства алгоритмов, решающих больший диапазон задач анализа данных ПГИ, в том числе и задачу локализации зон притока. В частности, предложенный подход может помочь в решении любой проблемы, которую можно свести к задаче классификации - причем как бинарной классификации, так и задачи классификации на несколько классов [12]. В качестве задачи классификации на несколько классов можно привести задачу локализации зон притока с указанием типа притока - вода, нефть или газ; эта же задача может решаться с применением трех моделей бинарной классификации.

Задача определения фазового состава флюида вдоль ствола скважины по данным термоанемометрии решается с помощью анализа геометрической формы циклов термоанемометрического датчика. Форма зависит от типа флюида, где находится датчик во время цикла (12 сек). Показано, что можно использовать два признака геомет-

рии цикла, вычисляемых на основе значений температуры – как фоновой температуры вдоль оси скважины, так и показаний специального распределенного прибора термоанемометрии, – в качестве координат пространства для кластерного анализа. В этих координатах удается разделить точки-цикли на три достаточно отчетливых кластера, соответствующих воде, нефти и газу. Для учета неопределенности разделение проводится с помощью модели смеси гауссовых распределений, позволяющей приписать каждой точке-циклу вероятность принадлежности к тому или иному кластеру. Это дает возможность не только определить для каждой точки наиболее вероятный соответствующий ей тип флюида, но и оценить фазовый состав при предполагаемом наличии смеси флюидов.

2 Часть 1. Локализация зон притока

2.1 Актуальность и обоснование выбора темы

Данных ПГИ и мониторинга скважин встроенными оптоволоконными системами становится все больше, и зачастую на решение поставленных задач не хватает рук экспертов и времени и ресурсов для численного моделирования; ответ нужен «в реальном времени». Требуется разработать надежные системы автоматизированной обработки – как минимум для выполнения основной части рутинной работы, производимой зачастую вручную; и как оптимум – для подготовки вариантов интерпретации данных и рекомендации решений для выработки экспертом окончательных заключений.

Методы машинного обучения могут оказать серьезную помощь в построении таких систем. Каждая скважина уникальна, и по данным невозможно однозначно описать происходящие в скважине процессы. Таким образом, возможность применения методов машинного обучения с учителем сильно ограничена, поскольку недостаточно хорошо размеченных датасетов. Поэтому в данном исследовании мы сконцентрировались на методах обучения без учителя.

В данной части работы обсуждаются три задачи интерпретации данных ПГИ. Задачи связаны между собой возрастающей необходимостью уметь автоматически локализовать и классифицировать зоны притока флюидов по глубине скважины – то есть определять, на каких интервалах глубины скважины наиболее вероятна зона притока флюида и определить тип флюида – вода, нефть, газ или смесь флюидов.

2.2 Используемые методы

2.2.1 Коррекция увязки по глубине

При проведении анализа данных ПГИ, полученных от датчиков в каротажной сборке приборов необходимо, во-первых, привязать показания по длине ствола скважины, и, во-вторых, синхронизировать сигналы от датчиков, поскольку приборы в сборке расположены на некотором расстоянии друг от друга, обычно не более нескольких метров. Пример каротажной сборки приборов показан на рис.1. Для привязки по стволу используются гамма-каротаж совместно с локатором муфт [5] и другие решения. Синхронизация сигналов производится предобработкой «сырых» данных на основе информации о расположении датчиков друг относительно друга и их технических характеристиках. В идеале ожидается, что все события в анализируемых временных рядах точно привязаны

к истинным значениям координаты траектории скважины.

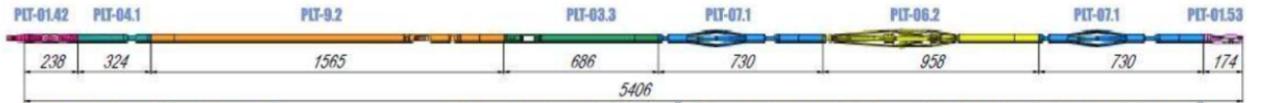


Рис. 1: Пример расположения датчиков в каротажной сборке. Здесь PLT-X - различные приборы; расстояния между ними указаны в миллиметрах.

Тем не менее, на первом этапе работы с полевыми данными (предоставленными нам уже после указанной предобработки) было решено проверить, насколько ситуация близка к идеальной. Корреляционный анализ показал, что рассинхронизация может превышать разумные погрешности, например 0.5м. Поэтому мы разработали и реализовали алгоритм предварительной синхронизации полевых данных. Отметим, что причинами рассинхронизации могут быть как технические сбои приборов и записывающей/считывающей аппаратуры, так и человеческий фактор. Это, вероятно, порождает ошибки в стандартных алгоритмах предобработки. Поэтому дополнительное использование корреляционного анализа позволяет распознать участки траектории с рассинхронизацией и устранить ее.

В предложенном алгоритме синхронизации производится автоматическая коррекция относительных сдвигов датчиков, которые уже не обязательно будут соответствовать реальным расстояниям между ними в приборе.

Постановка задачи Предположим, что имеется N датчиков α_n , $n \in [1, N]$, расположенных на некотором отрезке (сборка приборов) и генерирующих сигналы – последовательности по времени. Для удобства можно представить, что отрезок закреплен, а скважина движется относительно него. Отметим, что координата по времени однозначно пересчитывается в координату по стволу скважины через скорость движения.

Назовем вектором сдвигов набор значений (α_n, s_n) , $n \in [1, N]$, где для каждого датчика α_i , $i \in [2, N]$ указан сдвиг s показаний этого датчика относительно опорного датчика α_1 такой, что $-10\text{м} < s_i < 10\text{м}$ и $s_1 = 0$.

В идеале все сдвиги равны нулю (датчики находятся в одной точке). В реальности они разнесены, отсюда возникает рассинхронизация последовательностей сигналов от датчиков. Нам надо найти оптимальный вектор сдвига с учетом разной физики и зашумленности сигналов. Такой сдвиг с противоположным знаком должен обеспечить синхронизацию последовательностей.

Алгоритм Для разработки алгоритма анализировались данные по горизонтальной скважине Х45. На устойчивость работы алгоритма влияет много факторов, включая разную физику сигналов (температура, акустический шум, скорость флюида, фаза флюида и т.д.), зашумленность и др. Чтобы уменьшить влияние этих факторов, был выбран близкий к горизонтальному участок траектории, так как показания датчиков меньше подвержены искажениям, вызванных физической перестройкой течения в местах изменения наклона скважины. ми изменений траектории событиями. В данном случае длина участка составляла 300 метров (3000 точек). В качестве опорного датчика использовался сигнал показаний глубины – TVD (true vertical depth).

Шаг 1. Назовем оптимальным сдвигом между двумя массивами показаний A_1 и A_2 число $s_{12} = s(A_1, A_2)$ такое, что:

$$s_{12} = \arg \max_{-10\text{м} \leq s \leq 10\text{м}} (\text{corr}(A_1, A_2 - s)), \quad (2)$$

то есть такой сдвиг, применение которого к массивам A_1, A_2 позволит достичь максимального коэффициента корреляции Пирсона между ними.

Оптимальный сдвиг можно оценить перебором возможных значений сдвигов в диапазоне $[-10\text{м}, 10\text{м}]$. В первом шаге алгоритма для каждой пары показаний датчиков вычисляется оптимальный сдвиг, результатом вычисления является антисимметричная матрица 1 размера $N \times N$ (рисX).

Отсутствующие значения в ячейках матрицы 1 означают, что максимум коэффициента корреляции был достигнут на границах рассматриваемого участка $\pm 10\text{м}$. Так как длина колонны обычно не превышает 10м в длину, такие максимумы корреляции нефизичны, то есть, на рассматриваемом интервале траектории данные не имеют заметной корреляции, и алгоритм «поймал» корреляцию уже не связанных друг с другом событий.

Из-за большого количества «пустых значений» в данном случае полученная матрица 1 состоит из одного столбца и одной строки – то есть необходимые взаимные сдвиги можно получить, просто взяв эти столбец или строку. Однако, такой вид матрицы не гарантируется: в общем случае может нарушаться аддитивность оптимальных взаимных сдвигов. Поэтому автоматическое определение оптимального вектора сдвигов предполагает дополнительный шаг.

Шаг 2. Если бы отсутствовали любые сдвиги между датчиками, матрица 1 состояла бы только из нулей. Будем к этому стремиться. Назовем оптимальным вектором сдвигов такой вектор сдвигов $(\alpha_n, s_n)^{opt}, n \in [1, N]$, при применении которого к данным матрица

1 была бы минимальной по евклидовой норме, то есть:

$$S^{opt} = \arg \min_{s \in [-10, 10]^N} \sum_{i,j} (M_{ij})^2 \quad (3)$$

При вычислении нормы матрицы отсутствующие значения полагаются равными нулю.

Для оптимизации L_2 -нормы матрицы M предлагается следующий алгоритм:

$S^0 = [0] \times N$ – изначально нулевой вектор сдвигов,

$i = 1$ – номер строки матрицы M ,

$k = 0$ – номер итерации,

$\varepsilon = 10^{-3}$ – регулируемый критерий останова сходимости,

$M^0 = M$ – полученная на шаге 1 матрица попарных сдвигов,

$F^0 = L_2(M) = \sum_{i,j} (M_{ij})^2$ – начальное значение минимизируемой нормы матрицы.

while $F^k > \varepsilon$:

$$m = \frac{1}{N} \sum_{n=1}^N M_{in}$$

$$S_i^k = m$$

Нулевая матрица T^k размера NxN заполняется следующим образом:

$$\forall i \in [1, \dots, N], \forall j \in [1, \dots, i-1, i+1, \dots, N] \rightarrow T_{ij}^k = m, T_{ji}^k = -m$$

$$M^k = M^{k-1} - T^k$$

$$k = k + 1$$

$$i = i + 1 \text{ если } i < N, \text{ иначе } i = 0$$

$$F^k = L_2(M^k)$$

Таким образом, на каждой итерации из соответствующей строки матрицы вычитается ее среднее; к соответствующему столбцу добавляется это среднее для сохранения антисимметричности; вычисленное среднее добавляется к изначально пустому вектору сдвигов.

На рассматриваемом примере оптимизируемый функционал в результате алгоритма уменьшился с $F^0 = 320.5$ до $F^K = 2.5 \times 10^{-4}$ за $K = 10$ итераций. Результаты на

примере некоторых данных скважины X45 показаны на следующих рисунках: создание и преобразование матрицы сдвигов на рис.2, предложенная схема расположения датчиков на рис.3 и демонстрация сдвига данных на рис.4.

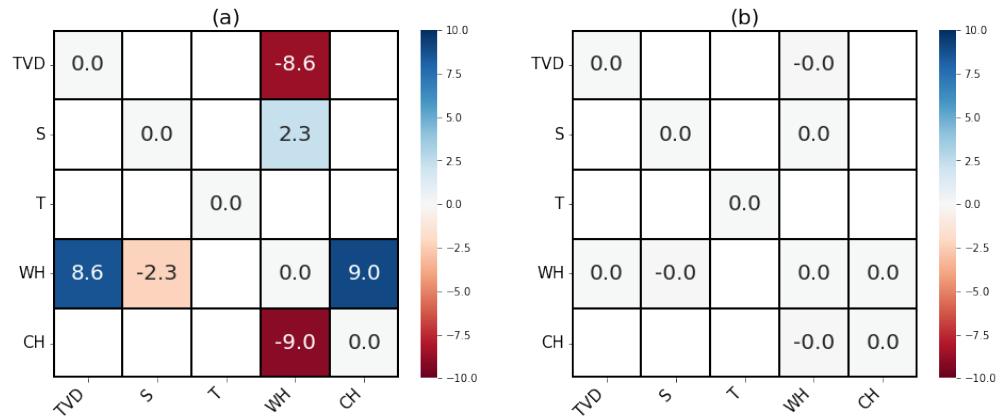


Рис. 2: Матрицы попарных оптимальных сдвигов. (а) - вариант до оптимизации, (б) - результат после минимизации L_2 -нормы. Видно, что удалось достичь идеального случая - с определенной точностью все несоответствия стали нулевыми.



Рис. 3: Примерная предложенная схема сдвига датчиков по полученным значениям. Здесь сдвиги отцентрированы относительно датчика глубины - TVD.

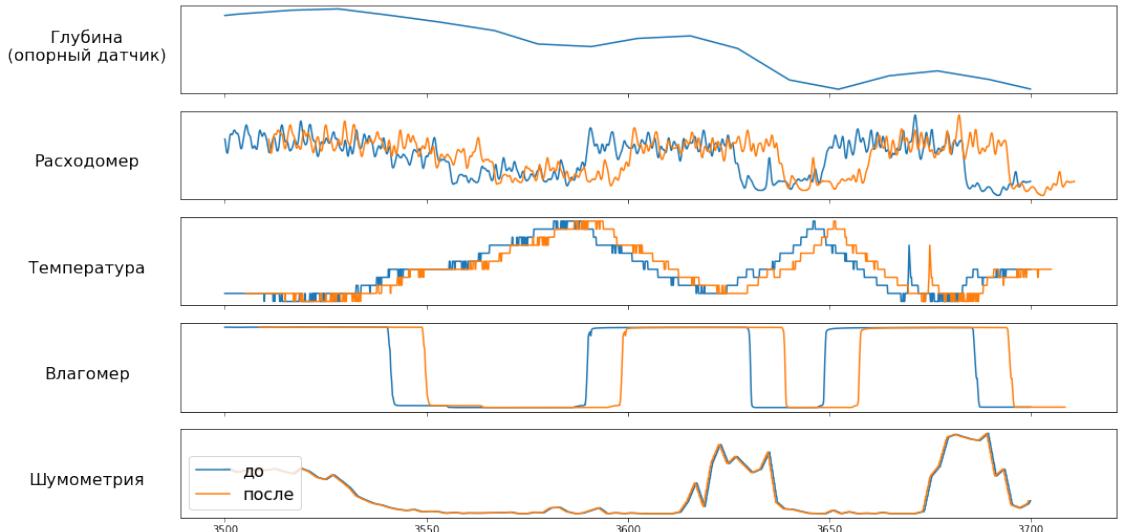


Рис. 4: Демонстрация того, как полученные сдвиги повлияли на визуальное соответствие датчиков.

Можно продемонстрировать необходимость выбора наиболее горизонтального участка траектории для анализа. На рисунках 5,6,7 изображены матрицы получения оптимального вектора сдвигов с использованием всей траектории – порядка километра. На протяжении траектории происходит множество событий, которые вносят шум в данные. Из-за этого не получается достичь удовлетворительных результатов – оптимизация матрицы на шаге 2 мало меняет ситуацию (несмотря на то, что оптимизируемый функционал уменьшился с A до B), и применение полученных сдвигов не соответствует ранее полученным результатам и в принципе тому, что могло бы быть результатом ручной работы.

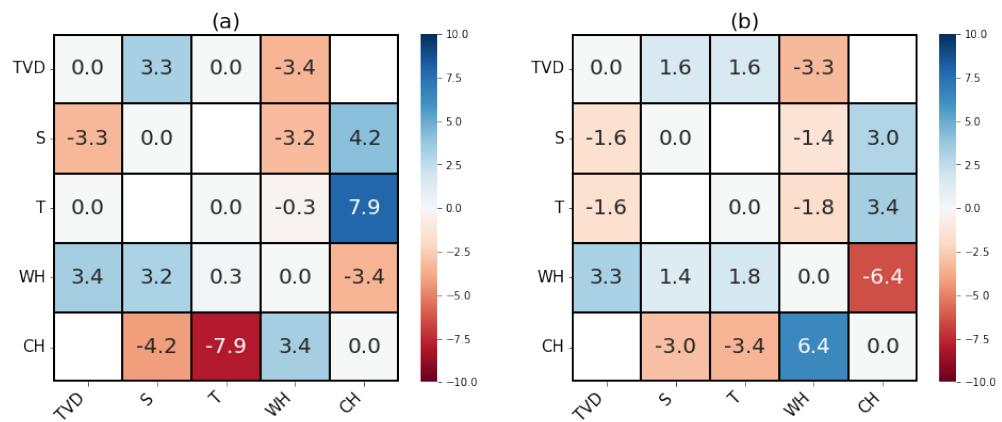


Рис. 5: Матрицы аналогично предыдущему случаю. Видно, что на втором шаге (b) не удалось достичь идеального результата – нулевой матрицы; это показывает, что диапазон траектории для анализа был выбран неверно.

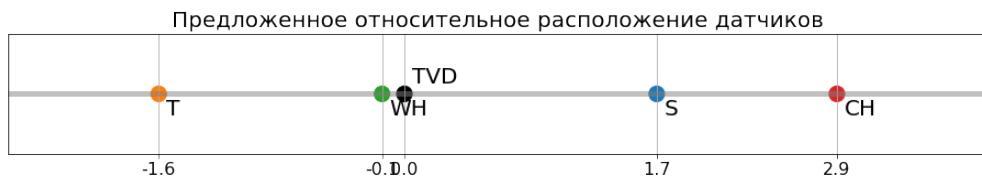


Рис. 6: Полученные сдвиги не совпадают с тем, что мы получили на предыдущем этапе.

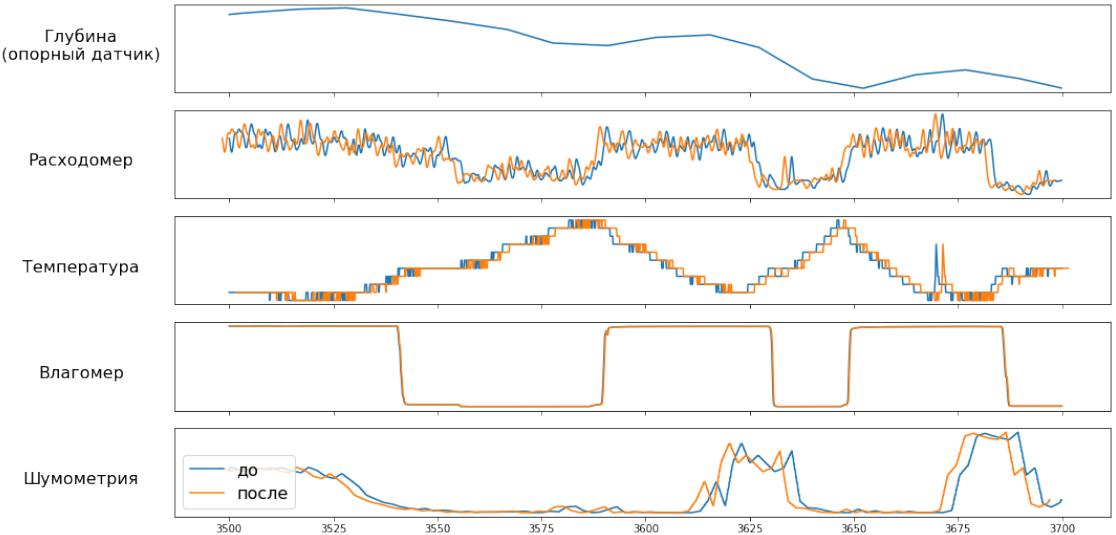


Рис. 7: Видно, что алгоритм почти ничего не поменял в данных - и визуального совмещения пиков сигналов не вышло.

2.2.2 Выделение точек изменения тренда

При анализе показаний временных последовательностей датчиков каротажной сборки приборов зачастую важны не мгновенные значения измеренных физических величин, а их тренды. Тренды могут представлять собой разрывные кусочно-линейные функции, см. пример на рис.9. Для построения таких трендов мы разработали алгоритм, основанный на распознавании точек изменения формы графика сигнала, так называемых точек «перелома» (changepoints).

Отметим, что в каротажном планшете график сигналов построен по координате вдоль ствола скважины, и его принято называть профилем. Поэтому далее в обозначениях будут использоваться сочетания «профиль» или «тренд профиля».

Построение профиля проводится нами в два этапа: вначале ищем точки перелома, затем строим профиль.

Один из распространенных подходов выявления точек перелома заключается в следующем: задается некоторый функционал \mathcal{C} от численного ряда, и после этого подбирается такое разбиение этого ряда, чтобы сумма значений функционалов от отрезков была меньше значения функционала \mathcal{C} от всего ряда. То есть для численного ряда $y = [y_1, \dots, y_N]$ ищем такой набор точек $T_N = 0 = \tau_0 < \tau_1 < \dots < \tau_m < \tau_{m+1} = N; m < N$, чтобы выполнялось

$$\mathcal{C}(T_N) = \sum_{i=1}^{m+1} \mathcal{C}(y_{\tau_{i-1}+1:\tau_i}) < \mathcal{C}(y) \quad (4)$$

Самый простой способ выбрать функцию потерь \mathcal{C} – минимизировать среднеквад-

ратичное отклонение значений ряда на отрезке.

Чтобы исключить тривиальное оптимальное решение в виде разбиения всего ряда на отрезки по одной точке, вводится штраф за количество найденных разбиений, то есть минимизируется не $\mathcal{C}(T_N)$, а $\mathcal{C}(T_N) + \beta f(m)$, где β – параметр регуляризации, m – количество найденных точек перелома профиля и f - обычно линейная функция $f(m) = m$.

Распространенный подход оптимизации функции потерь $F = \mathcal{C}(T_N) + \beta m$ – бинарная сегментация. На каждой итерации ищется одна точка перелома, позволяющая уменьшить функцию потерь F . Если точка находится, каждый из двух получившихся в результате деления отрезков переходит на следующую итерацию. Алгоритм останавливается, когда значение функции потерь перестает уменьшаться. Это вычислительно эффективно, но не гарантирует оптимального деления.

Для выделения оптимальных точек перелома мы будем использовать алгоритм PELT (Pruned Exact Linear Time), описанный в [13].

Метод основан на способе оптимального разбиения, описанного в 2005 году в [14].

Определим $F(s)$ как минимальное значение функции потерь на первых s точках ряда и T_s как набор m точек перелома для первых s точек ряда. Пусть $F(0) = -\beta$. Тогда, разворачивая определение, получаем

$$\begin{aligned} F(s) &= \min_{\tau \in T_s} \sum_{i=1}^{m+1} [\mathcal{C}(y_{\tau_{i-1}+1:\tau_i}) + \beta] = \\ &= \min_t \left\{ \min_{\tau \in T_t} \sum_{i=1}^m [\mathcal{C}(y_{\tau_{i-1}+1:\tau_i}) + \beta] + \mathcal{C}(y_{t+1:\tau_{m+1}} + \beta) \right\} = \\ &= \min_t \{F(t) + \mathcal{C}(y_{t+1:\tau_{m+1}}) + \beta\} \end{aligned} \quad (5)$$

Другими словами, на каждой итерации ищется оптимальное разбиение на укороченной версии ряда, после чего ряд удлиняется на одну точку и проводится проверка: осталось ли разбиение оптимальным или нужно перенести последнюю точку перелома или добавить новую.

PELT является модификацией алгоритма [14], позволяющей убрать некоторые точки в процессе проверки разбиения T на оптимальность, чтобы ускорить вычислительный процесс.

Делается предположение, что существует такая константа K , что уменьшение функционала F при добавлении любой точки будет как минимум больше этой константы, то есть существует минимальное гарантированное значение выигрыша от деления:

$$\exists K \geq 0 : \forall t < s < T \rightarrow \mathcal{C}(y_{t+1:s}) + \mathcal{C}(y_{s+1:T}) + K \leq \mathcal{C}(y_{t+1:T}) \quad (6)$$

Предположение эмпирически доказывается в статье [13] на примере наиболее часто используемых функций потерь \mathcal{C} .

Таким образом, если для кандидата t на новую точку разбиения выполняется условие

$$F(t) + C(y_{t+1:s}) + K \geq F(s), \quad (7)$$

то t не является оптимальной точкой разбиения. Проверка этого условия позволяет сделать вычислительное время алгоритма линейным – отсюда название алгоритма Pruned Exact Linear Time.

Показания датчиков отражают изменения различных физических величин по длине скважины. У этих физических величин есть характеристики поведения, и у датчиков есть инерция измерений – например, плавно меняющиеся показания температуры не могут отражать часто случающиеся события. Поэтому при выделении точек перелома необходимо для каждого датчика вводить свое значение параметра P_1 «минимальное расстояние между событиями». После задания значения P_1 значение штрафа β выбирается в нашем алгоритме автоматически подбором: β уменьшается с определенного стартового значения до тех пор, пока минимальное расстояние между всеми полученными при данном значении β точками не станет меньше P_1 . Результат работы алгоритма PELT на примере сигнала расходомера с различными значениями штрафного параметра β показаны на рис.8. Используется программная реализация алгоритма из библиотеки *ruptures* на языке Python.

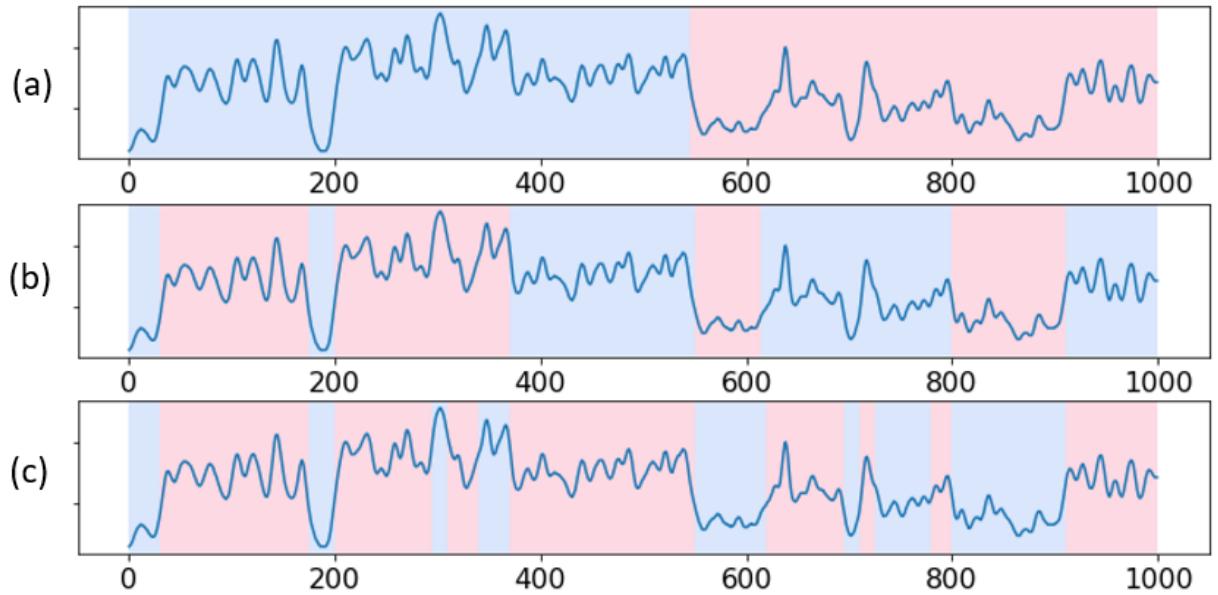


Рис. 8: Выделенные точки перелома при различных значениях штрафного параметра β : $\beta_{(a)} = 50, \beta_{(b)} = 10, \beta_{(c)} = 5$.

2.2.3 Построение разрывных трендов профиля

По полученным с помощью PELT точкам перелома теперь можно построить желаемые типы профилей. Они могут быть как разрывные (расходомер), так и непрерывные (температура). Также можно применять постоянные (влагомер), линейные (расходомер) и квадратичные функции для гладких участков профиля между точками перелома. Эти функции строятся на заданном отрезке с помощью минимизации среднеквадратичного отклонения от исходного профиля. В результате получаются различные типы профиля, которые можно обозначить как D_n или C_n – разрывный (D) или непрерывный (C) со степенью полинома $n = 0, 1, 2$.

Как пример рассмотрим разрывный кусочно-линейный профиль для расходомера (DPL-profile, Discontinuous Piecewise Linear profile). Желаемый результат показан на рис.9.

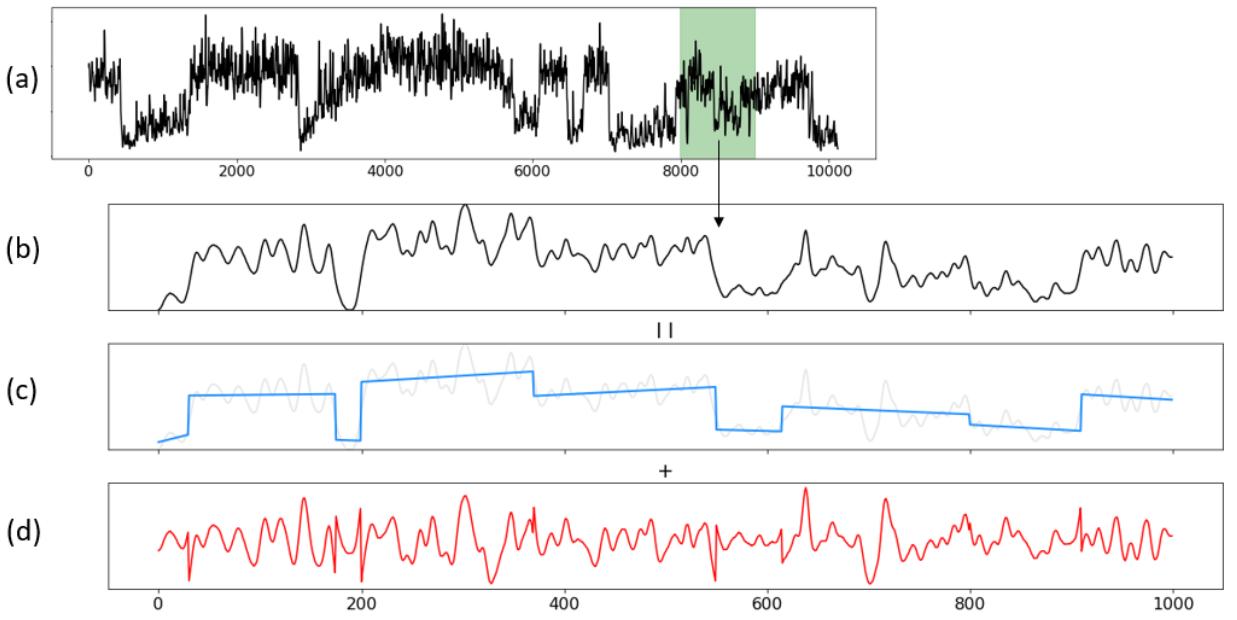


Рис. 9: Пример того, как хочется преобразовать сигнал при выделении профилей. (а) - изначальный сигнал расходомера, рассматриваем выделенный зеленым участок (б). Сигнал разделяется на профиль (с) и остаточный сигнал (д), получаемый вычитанием профиля из начального сигнала.

График остатка, который характеризует «шумовую» компоненту сигнала в отличие от полезной компоненты (профиля), позволяет эксперту оценить визуально адекватность работы алгоритма.

Таким образом, нами предложена и реализована процедура построения профилей, в которой входными параметрами являются типично минимальное расстояние между событиями (точками перелома) и тип профиля – C_n , D_n , $n = 0,1,2$. Назначение этих параметров проводится экспертом – свое для каждого датчика.

2.2.4 Примеры результатов построения профилей

Характер поведения показаний определенного датчика предлагает вид профиля, который можно из него извлечь:

1. Показания влагомера быстро выходят на константу, показывая различные уровни воды в точке измерения – от 0% до 100%. Поэтому каждый участок между двумя точками перелома аппроксимируется горизонтальным отрезком; в итоге получается ступенчатый профиль – D_0
2. Показания расходомера могут как меняться плавно, так и испытывать скачки. Поэтому для каждого участка между двумя точками перелома строится линейная

регрессия; в итоге получается линейный профиль с разрывами – D_1

3. Показания температуры не могут испытывать скачков. Поэтому это C_1 или C_2 профили. При этом возможные небольшие разрывы в точках перелома «сшиваются». Также полезным является построение D_1 профилей для производной температуры, но здесь мы этот вопрос не рассматриваем.

Примеры результатов построения профилей сигналов показаны на рисунках 10,11,12 на выделенных участках траектории скважины.

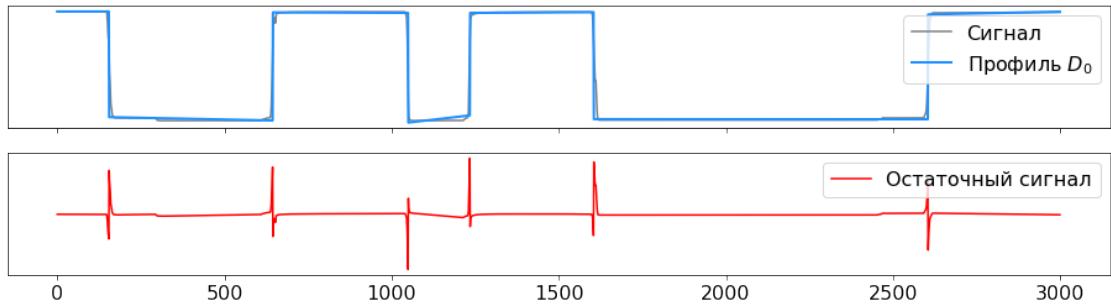


Рис. 10: Ступенчатый разрывной профиль для показаний влагомера.

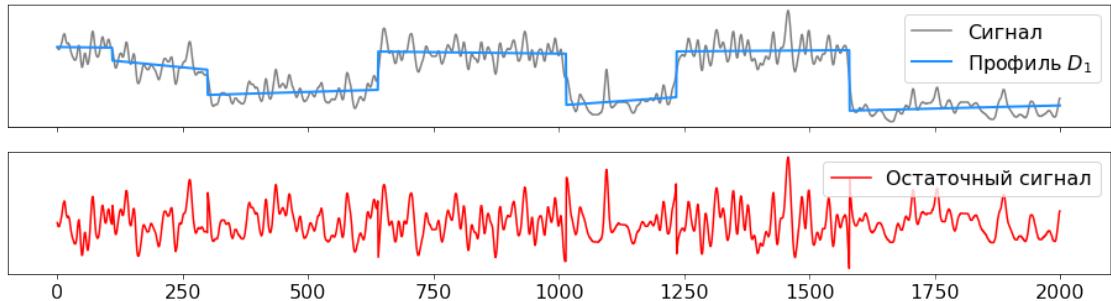


Рис. 11: Кусочно-линейный разрывной профиль для показаний расходомера.

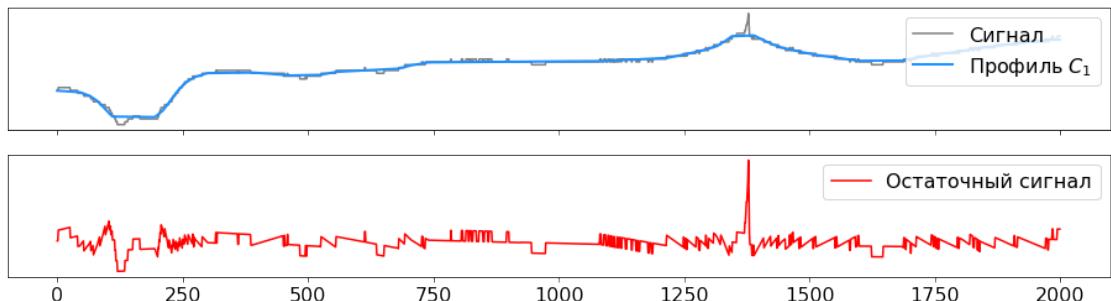


Рис. 12: Кусочно-линейный профиль для показаний температуры без разрывов. В данном случае позволяет убрать ступенчатую структуру температуры, появившуюся из-за слишком маленькой вариации значений.

2.2.5 Модель слабого контроля

При построении алгоритма интерпретации на основе анализа трендов профилей, полученных в параграфе «Построение разрывных трендов профиля» для данных ПГИ и шумометрии, мы предлагаем использовать модель слабого контроля (WS, weak supervision). Это вероятностная модель машинного обучения без учителя, впервые описанная в [15] как способ быстрого создания большого количества датасетов без экспертной разметки.

Суть модели заключается в объединении результатов работы слабых разметочных функций (LFs, labelling functions) в итоговую разметку. Разметочные функции – эмпирические либо любые другие правила, позволяющие решать поставленную задачу классификации – это может быть как классический вычислительный алгоритм, так и модель машинного обучения. Преимущество модели заключается в том, что разметочные функции не обязаны по отдельности быть очень точными, покрывать все множество возможных объектов для разметки или даже не противоречить друг другу; в результате обучения модель объединяет все их предсказания в одну разметку.

Авторы [15] описывают WS как генеративную, или порождающую модель (generative model), то есть как вероятностную модель совместного распределения данных. Разметочные функции описываются как «шумные» - решающие задачу не полностью или неточно, и процесс обучения позволяет избавиться от этого шума, учитывая корреляции разметочных функций.

Опишем вкратце процесс обучения модели на примере решения задачи бинарной классификации. Процесс обучения состоит из минимизации двух функционалов.

Шаг 1. Назовем разметочной функцией $\lambda(x)$ функцию, которая каждому объекту обучающего датасета S ставит в соответствие одну из меток $-1, 0, 1$, где метки -1 и 1 соответствуют двум классам решаемой задачи бинарной классификации, а метка 0 соответствует случаю «недостаточно информации, чтобы принять решение».

Для каждой разметочной функции $\lambda(x)$ введем скалярные вероятности α, β как:

α - вероятность того, что разметочная функция разметит объект не нулем,

β - вероятность того, что разметочная функция разметит объект меткой -1 или 1 правильным образом.

Разметочные функции порождают разметку Λ .

На первом шаге подбираются параметры $(\hat{\alpha}, \hat{\beta})$ такие, что

$$\hat{\alpha}, \hat{\beta} = \arg \max_{\alpha, \beta} \sum_{x \in S} \log \mathbb{P}_{(\Lambda, Y) \sim \mu_{\hat{\alpha}, \hat{\beta}}} (\Lambda = \lambda(x)) = \arg \max_{\alpha, \beta} \sum_{x \in S} \log \left(\sum_{y' \in \{-1, 1\}} \mu_{\alpha, \beta}(\lambda(x), y') \right) \quad (8)$$

Другими словами, полученные оптимизацией функционала векторы параметров $(\hat{\alpha}, \hat{\beta})$ максимизируют правдоподобие полученной разметки.

Здесь $\mu_{\hat{\alpha}, \hat{\beta}}$ - генеративная модель, позволяющая создавать разметку Y для решения задачи бинарной классификации, используя вероятностные метки Λ для классов -1, 0, 1.

Шаг 2. После нахождения наиболее хорошо описывающих обучающий датасет наборов коэффициентов $(\hat{\alpha}, \hat{\beta})$ решается задача логистической регрессии с модифицированной лосс-функцией. Подбирается набор весов \hat{w} такой, что

$$\hat{w} = \arg \min_w L_{\hat{\alpha}, \hat{\beta}}(w, S), \quad (9)$$

где

$$L_{\hat{\alpha}, \hat{\beta}}(w, S) = \frac{1}{|S|} \sum_{x \in S} \mathbb{E}_{(\Lambda, Y) \sim \mu_{\hat{\alpha}, \hat{\beta}}} \left[\log \left(1 + e^{-w^T f(x) Y} \right) | \Lambda = \lambda(x) \right] + \rho \|w\|^2 \quad (10)$$

В этой формуле:

- Логарифм под матожиданием – лосс-функция для классической задачи логистической регрессии для случая бинарной классификации с метками -1, 1:

$$L(Y, w^T f(x)) = \log \left(1 + e^{-w^T f(x) Y} \right) \quad (11)$$

- Слагаемое $\rho \|w\|^2$ – стандартная регуляризация весов.

Поскольку разные комбинации функций разметки дают разметку разного качества, функцию потерь логистической регрессии следует скорректировать, чтобы давать более шумной разметке меньший вес. Эта коррекция выполняется с помощью функции $\mathbb{E}_{(\Lambda, Y) \sim \mu_{\hat{\alpha}, \hat{\beta}}}$.

Другими словами, уровень доверия к каждой разметочной функции характеризуется коэффициентами вероятностной модели, которые подбираются в процессе обучения. Можно предложить аналогию, в которой каждая разметочная функция – это «эксперт». «Эксперты» обладают различной квалификацией: могут с различной частотой отвечать на любой вопрос «не знаю» (метка 0) и с различной частотой выдавать положительные (метка 1) и отрицательные (метка -1) ответы. Эти частоты соответствуют вероятностям $(\hat{\alpha}, \hat{\beta})$, введенным в предыдущем параграфе. Модель слабого контроля собирает

всех «экспертов» в одном месте и для каждого объекта в датасете задает им вопрос – «что думаете?», а после получения меток от каждого «эксперта» на каждый объект принимает решение с учетом того, как показания «экспертов» соотносились между собой.

В данной работе предлагается взять модель слабого контроля как основу для создания модели-советника – автоматизированного алгоритма интерпретации, предлагающего эксперту варианты интерпретаций для выработки окончательного заключения. Предполагается, что функции разметки могут содержать выжимку из экспертных знаний, широко описанных в литературе, чтобы экспертам не пришлось выполнять рутинную работу. Так как итоговое решение строится с оценкой неопределенности, не требуется жесткого контроля над качеством функций разметки – они должны быть просто достаточно адекватны с точки зрения специалистов. При построении и реализации алгоритма используются программные модули рассмотренной модели из библиотеки *snorkel* на языке Python.

2.3 Результаты

2.3.1 Поиск событийных участков

Постановка задачи Рассматриваются данные горизонтальной скважины X45: датчики глубины, расходомера, температуры, влагомера и два канала пассивной акустической шумометрии – низкочастотный (0-2кГц) и высокочастотный (3-32кГц). Имеется порядка 10000 точек, записанных с шагом 0.1м по глубине скважины. Ставится задача по имеющимся данным автоматически локализовать «событийные» участки – то есть те, на которых наблюдается отклик нескольких датчиков одновременно, что может указывать на наличие зоны притока или другие события в скважине.

Используемые для анализа данные показаны на планшете на рис.13.

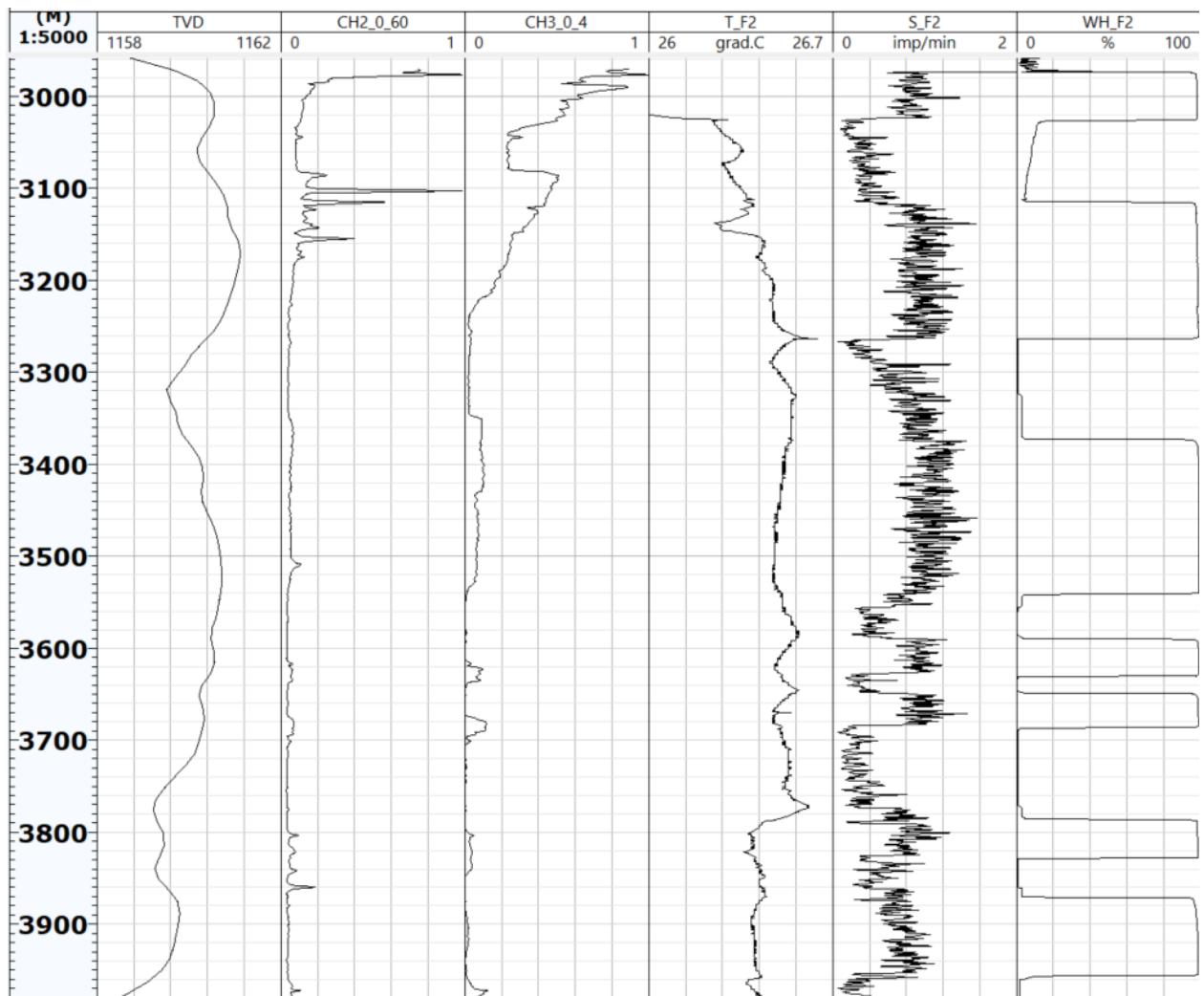


Рис. 13: Данные скважины 0Х, используемые для анализа. Слева направо: показания глубины скважины, высокочастотной шумометрии, низкочастотной шумометрии, температуры, расходометрии и влагомера.

Методы Нужно локализовать отклики во всех доступных сигналах, а затем совместить их в одну разметку. Отклики могут быть несогласованы между собой различными способами:

- В какой-то точке траектории есть отклики всех датчиков, кроме одного – например, в этой точке событие было, но слабое, и одному датчику не хватило чувствительности
- В какой-то точке траектории нет никаких откликов, кроме отклика одного датчика – например, в этой точке нет никакого значимого события, но один датчик чересчур чувствительный, и показания изменились «на пустом месте»
- В какой-то точке траектории часть датчиков показала отклик, а часть «промол-

чала» – и на первый взгляд нет однозначного решения по поводу того, каким показаниям стоит доверять.

Для решения задачи, усложненной такими обстоятельствами, подойдет модель слабого контроля, описанная выше.

В выбранной терминологии, для каждого датчика можно подобрать одну или больше функций разметки – то есть экспертных правил, которые позволяют предположить, произошло на данном участке какое-то событие или нет. Функция разметки в данной постановке задачи должна выдавать один из трех ответов на каждый интервал траектории: «есть событие» (метка 1), «нет события» (метка -1), «недостаточно информации для принятия решений» (метка 0).

Здесь используем по одной функции разметки на каждый датчик. Поскольку тренды профилей для разных датчиков могут иметь разные координаты точек перелома для одного и того же события (да и событие имеет некоторую протяженность), вводится некоторое окно, которому эта метка присваивается. В данном случае окно имеет размер в 50 точек (5м). Т.е., «непрерывная» траектория при разметке заменяется на интервалы с метками через каждые 5 метров.

В качестве индикатора события для расходомера, влагомера, показаний температуры и шумометрии хорошо подходят характеристики кусочно-разрывных профилей, описанных в предыдущей главе. Перед вычислением меток из всех сигналов – расходомера, влагомера, показаний датчика температуры и двух каналов шумометрии – выделяются кусочно-разрывные профили. Таким образом, для каждого сигнала у нас есть набор точек перелома, значения профиля и остаточного сигнала.

Получается:

- Для расходомера: считаем, что произошло событие (метка 1), если в окне находится хотя бы одна точка перелома; иначе – метка 0
- Для влагомера: метка 1, если в окне находится хотя бы одна точка перелома, иначе – метка 0
- Для температуры: метка 1, если в окне находится хотя бы одна точка, модуль значения остаточного сигнала превышает заданное значение, иначе – метка 0. Так как сигнал ввиду низкой вариативности значений практически ступенчатый, перед выделением точек перелома он был аппроксимирован сплайнами третьей степени.

- Для низкочастотного шума: метка 1, если в окне находится хотя бы одна точка перелома, иначе – метка 0
- Для высокочастотного шума: метка 1, если в окне находится хотя бы одна точка, модуль значения остаточного сигнала превышает заданное значение, иначе – метка 0.

Указанные функции выдают только положительные либо нейтральные метки. С другой стороны, показания датчика глубины позволяют присвоить некоторым точкам отрицательные метки – то есть увеличить вероятность того, что данная точка не содержит интересующего нас события. Некоторые изменения показаний датчиков могут являться исключительно следствием влияния траектории – например, вода может накапливаться на вогнутых по отношению к горизонту участках траектории, что вызовет скачок влагомера, который нас не интересует. Поэтому вводится еще одна функция разметки:

- Для траектории: метка -1, если внутри окна максимальное отклонение значений траектории от среднего значения внутри окна превышает 1.5σ стандартных отклонений значений траектории по всей глубине скважины (внутри данного окна наблюдается большая, чем в среднем, осцилляция значений), иначе 0.

Мы реализовали эту процедуру для траектории, но требуются дальнейшие исследования по повышению надежности такой разметки, поэтому здесь эти результаты не приводятся.

Результаты На рисунках 14 и 15 представлены предложенные функции разметки и результат обучения модели. Результаты работы разметочных функций – по одной на каждый датчик – пронумерованы как $\lambda_1, \dots, \lambda_5$ и расположены справа от соответствующих датчиков. На рис.14 серым изображены изначальные сигналы; на рис.15 – извлеченные из них профили для анализа.

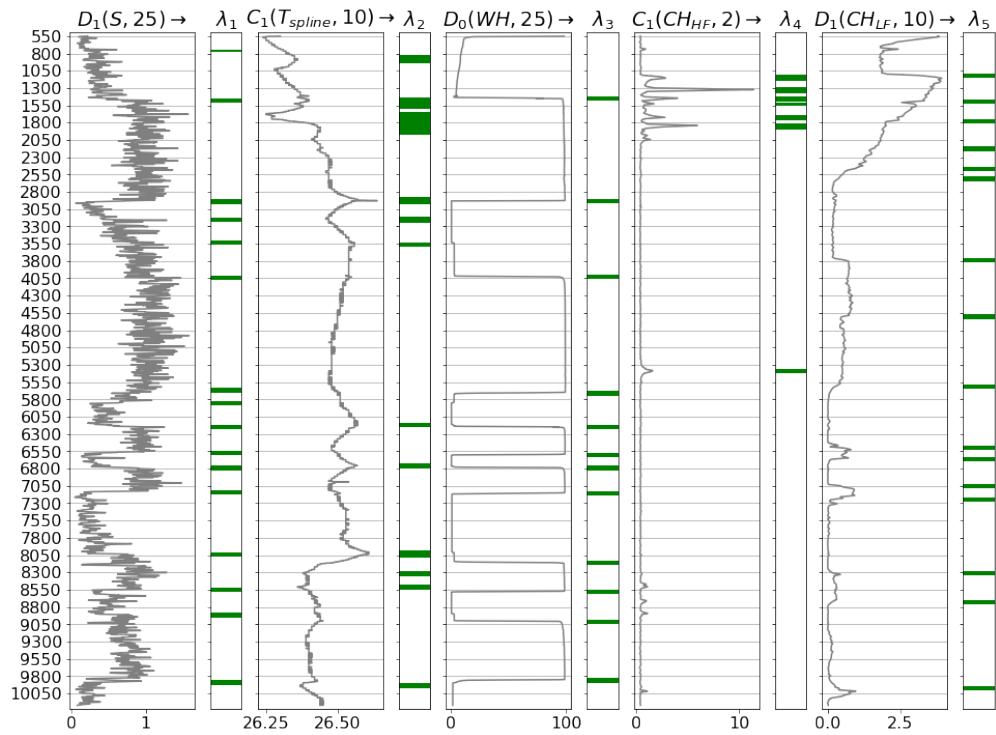


Рис. 14: Разметочные функции, используемые для создания итоговой разметки. Серым показаны сигналы, на основе которых построены разметочные функции, зеленым выделены диапазоны, где соответствующие разметочные функции отмечали вероятность события.

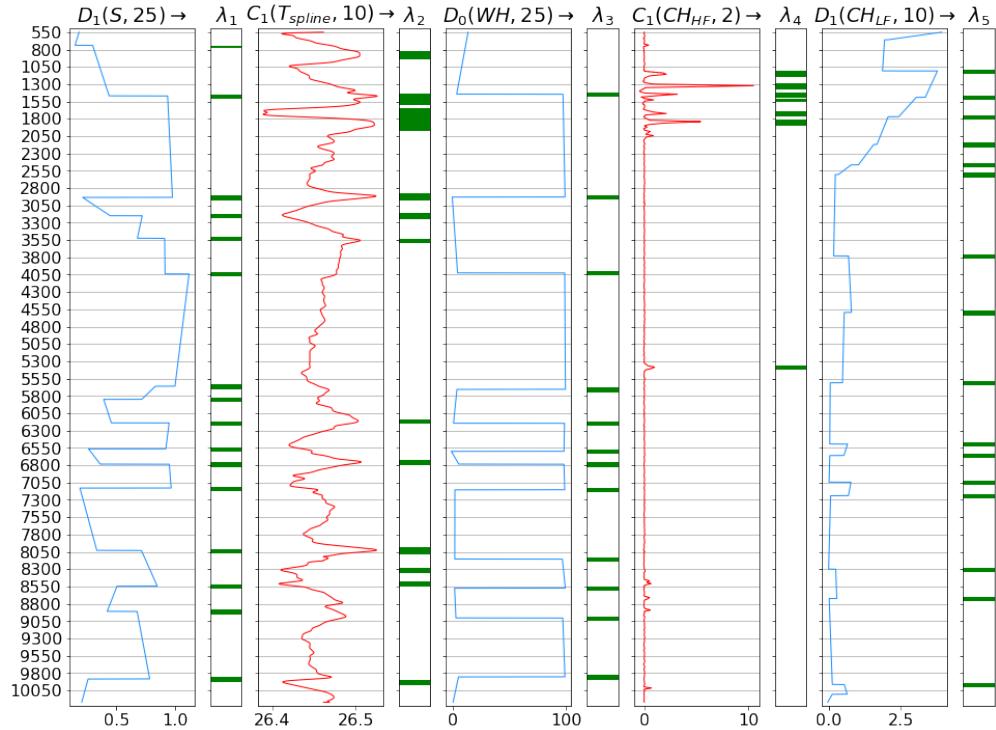


Рис. 15: Разметочные функции совпадают с указанными на предыдущем рисунке, но для демонстрации вместо оригинальных сигналов нарисованы выделенные профили (голубой цвет) или остаточные сигналы (красный цвет).

Результат работы модели – вероятностная разметка на два класса P_{WS} , которую можно преобразовать в дискретную разметку на два класса (1 или 0) или оставить так для оценки неопределенности.

На рисунках 16, 18 и 19 показаны результаты работы модели при различных параметрах минимального расстояния между событиями (значения указаны в скобках в метрах; точки разнесены на расстояние 0.1м). Первый результат предлагается считать оптимальным, следующие два получены при вариации минимального расстояния между событиями с уменьшением (18) и увеличением (19) всех значений. Видно, что от этого функции разметки становятся более или менее шумными – срабатывает детекция ложных событий или наоборот, важные события оказываются пропущены. Тем не менее, общий вид итоговой разметки примерно сохраняется. Это объясняется тем, что показания температуры и высокочастотной шумометрии анализируются по значениям остатка после выделения профиля, а расходомера, влагомера и низкочастотной шумометрии – по выделенным точкам перелома. Таким образом, при увеличении минимального интервала между событиями уменьшается количество точек перелома, но становятся более шумными остатки; и наоборот – при уменьшении минимального интервала между событиями количество точек перелома увеличивается, но остатки температуры и высокочастотной шумометрии показывают меньше информации.

На рисунках 16, 18 и 19 класс 0 («недостаточно информации») изображен белым цветом, класс 1 («вероятно событие») - зеленым.

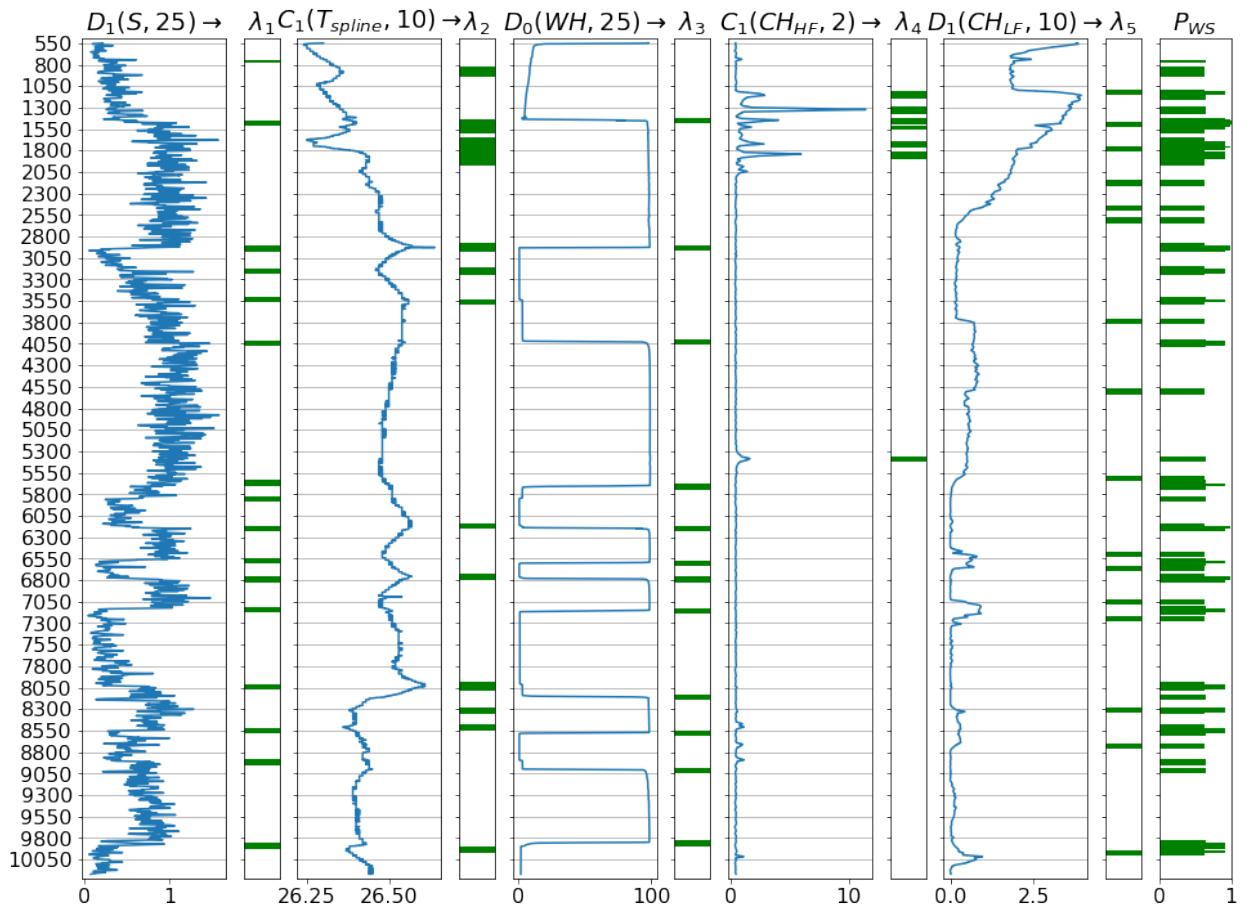


Рис. 16: Из разметочных функций в результате работы модели слабого контроля создается финальная разметка в последней колонке. Длина столбиков отражает вероятность события в данной точке.

Здесь показан результат работы модели (последний столбец P_{WS}) для участка в 1500 точек в приближении.

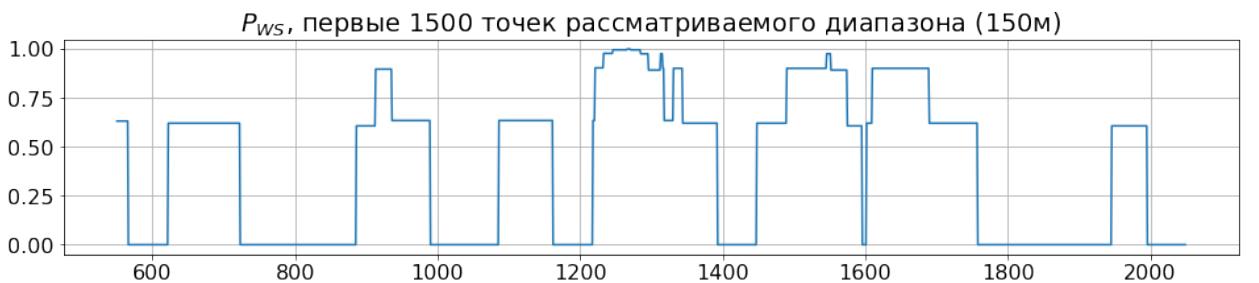


Рис. 17: Увеличенный участок последней колонки с предыдущего рисунка.

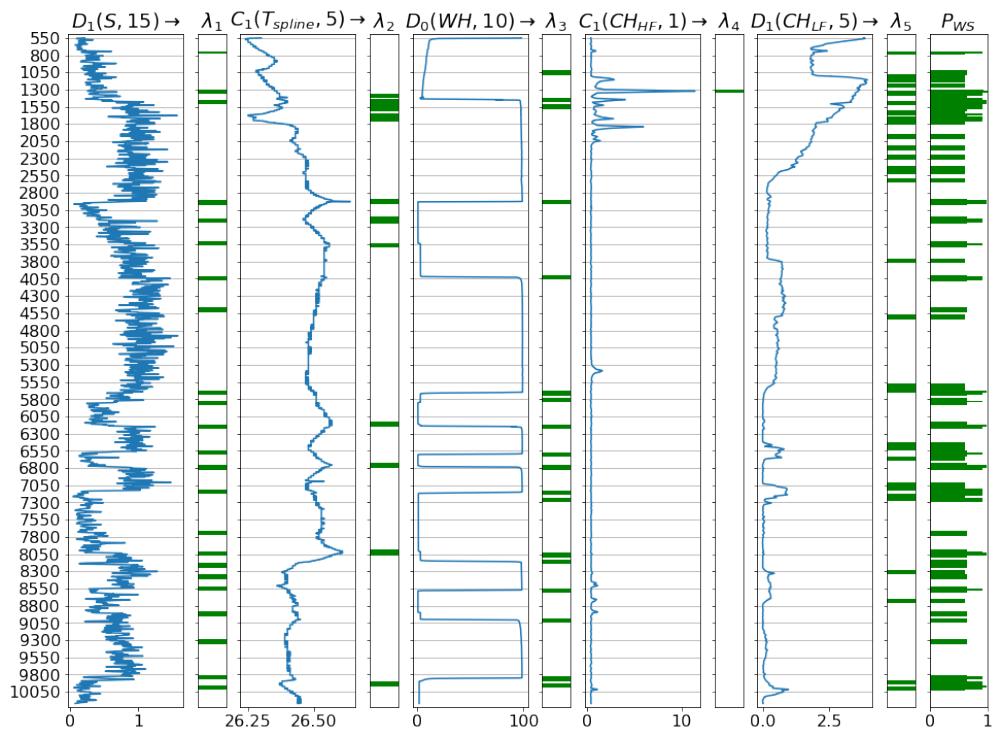


Рис. 18: Проверка устойчивости модели: для каждой разметочной функции минимальная ширина между событиями для выделения профилей была уменьшена. Видно, что финальная разметка изменилась, но не сильно.

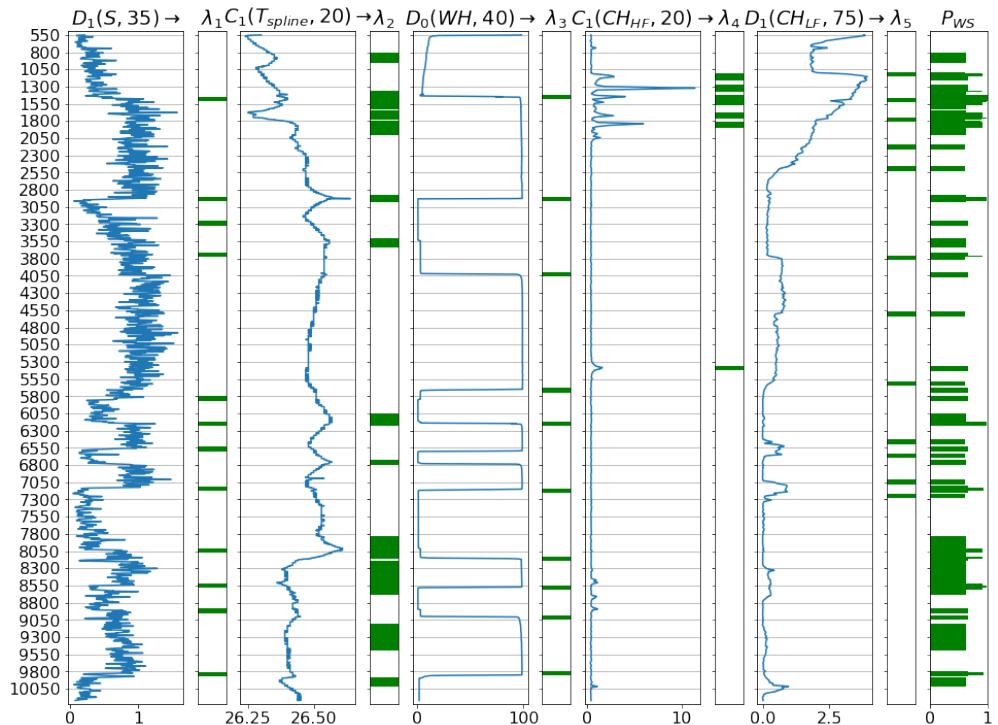


Рис. 19: Аналогично предыдущему рисунку, но значения минимальной ширины между событиями были увеличены относительно оптимальных, а не уменьшены.

2.3.2 Детекция притока по данным температуры оптического кабеля

Постановка задачи Рассматривается вертикальная скважина L1. Для анализа доступны данные температуры, полученные с помощью волоконно-оптического кабеля, протянутого по длине скважины. Показания записывались в диапазоне порядка шести часов с шагом примерно 7 секунд – всего около 3000 измерений для участка траектории порядка 2.5км с шагом 0.5м. Пример данных в начале, в середине и в конце измерений показан на рис.20.

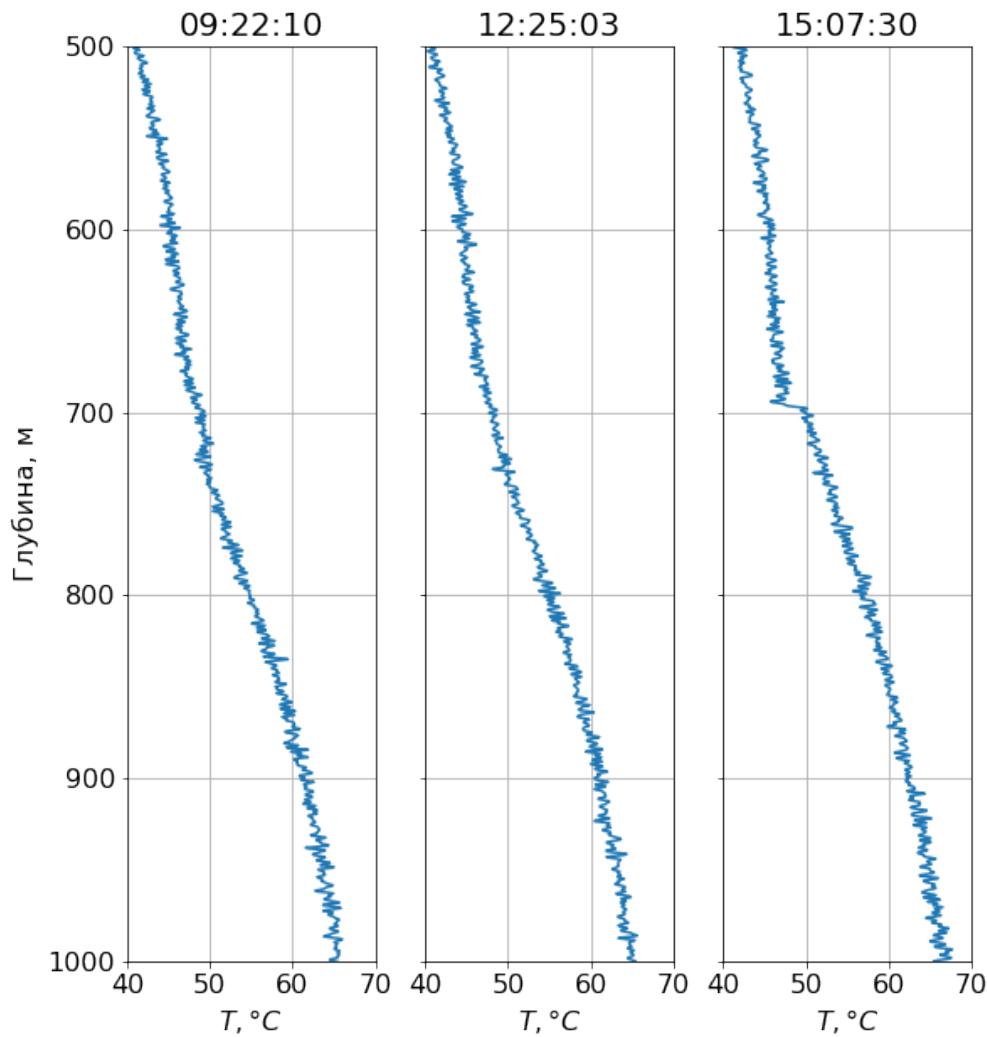


Рис. 20: Пример данных в три различных периода времени. На третьем графике виден пик в районе 700м, на других двух он незаметен. Задача состоит в том, чтобы по всем сигналам выловить возможные пики, которые могут наблюдаться не на всем времени измерений.

Необходимо по всем доступным данным локализовать возможные места притока в скважине. Отклик от притока – температурная аномалия.

Методы и результаты На предыдущем рисунке видно, что по последнему измерению можно заметить скачок температурных измерений на глубине порядка 700 м, в то время как на других двух ничего такого заметить нельзя. Данные шумные, однобразные и их много – поэтому для минимизации шума можно просто их усреднить – результат показан на рис.21(а).

В полученном сигнале видно уже два скачка – на глубине порядка 700 и 1600, но они все еще слабые. Поэтому необходимо вычесть «гладкий» профиль сигнала и рассматривать только остаточный «шум». В качестве метода получения гладкого профиля выбрали кусочно-непрерывную аппроксимацию сплайнами третьей степени. После вычитания аппроксимации сплайнами остается сигнал, показанный на рис.21(б) – здесь скачки уже видны гораздо лучше.

Тем не менее, в сигнале наблюдается низкочастотная осциллирующая структура, которая может быть артефактом измерений. Она автоматически убирается с помощью высокочастотного фильтра на основе преобразования Фурье (зануление в спектре всех частот ниже той, которая имеет максимальную интенсивность, и последующее восстановление сигнала). Спектр и остаточный сигнал показаны на рис.21(с),(д).

Чтобы получить из данного сигнала некоторое подобие вероятности, позволяющее ранжировать только самые интересные участки, можно возвести модуль этого сигнала в какую-нибудь степень α , $1 < \alpha < 5$. Результат возведения в степень с $\alpha = 4$ отнормирован на интервал $[0,1]$ и показан на рис.21(е). Полученные пики для координат около 700 и 1600 локализуют места притока, а их амплитуды – «вероятность» – притока.

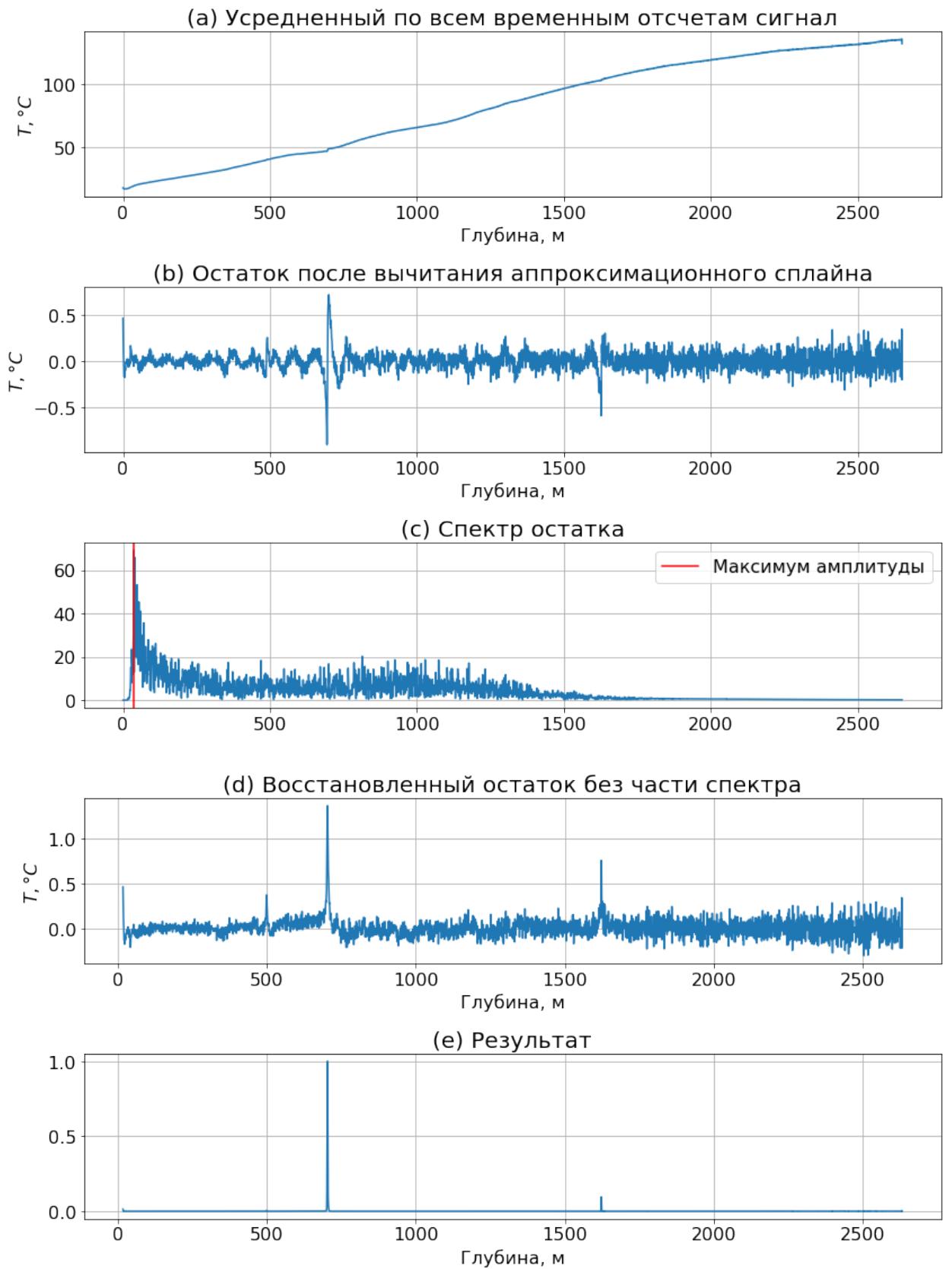


Рис. 21: Последовательные этапы выделения полезной информации из усредненного сигнала.

2.3.3 Детекция и классификация зон притока на тестовой скважине

Постановка задачи В данной задаче рассматриваются данные тестовой скважины М.

Скважина имеет заранее предопределенные характеристики: в ней имеется четыре набора отверстий на известной глубине. Сверху находится клапан для входа газа, ниже – три группы отверстий для входа воды в группах по 144, 144 и 216 штук соответственно. Задача заключается в том, чтобы локализовать зоны этих перфораций и определить, какой флюид втекает – газ или вода. Схема скважины показана на рис.22.

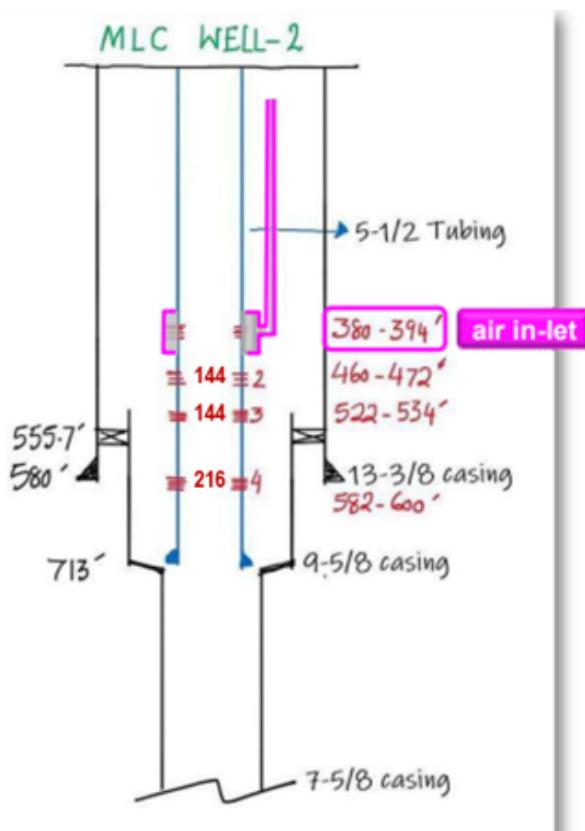


Рис. 22: Схема скважины.

Для анализа используются данные прибора HFND – он содержит в себе температурный датчик и датчик пассивной акустической шумометрии. Здесь мы концентрируемся на анализе данных шумометрии. Данные записывались в диапазоне 0-100кГц с шагом 0.167м.

Методы и результаты Предварительное сравнение спектрограммы со схемой перфораций показало, что отклик от притоков воды наблюдается в низких частотах, а

газовый приток издает высокочастотный шум.

Поэтому из всего измеряемого шумометром диапазона частот 0-100кГц для анализа мы выбрали четыре частотных диапазона:

- Низкие частоты: 160-320Гц
- Низкие частоты, расширенные: 160-1280Гц
- Высокие частоты: 5-10кГц
- Высокие частоты, расширенные: 5-40кГц

По каждому частотному диапазону в каждой точке траектории суммируются интенсивности сигналов по всем полосам, в которых производилась запись; результат показан на рис.23.

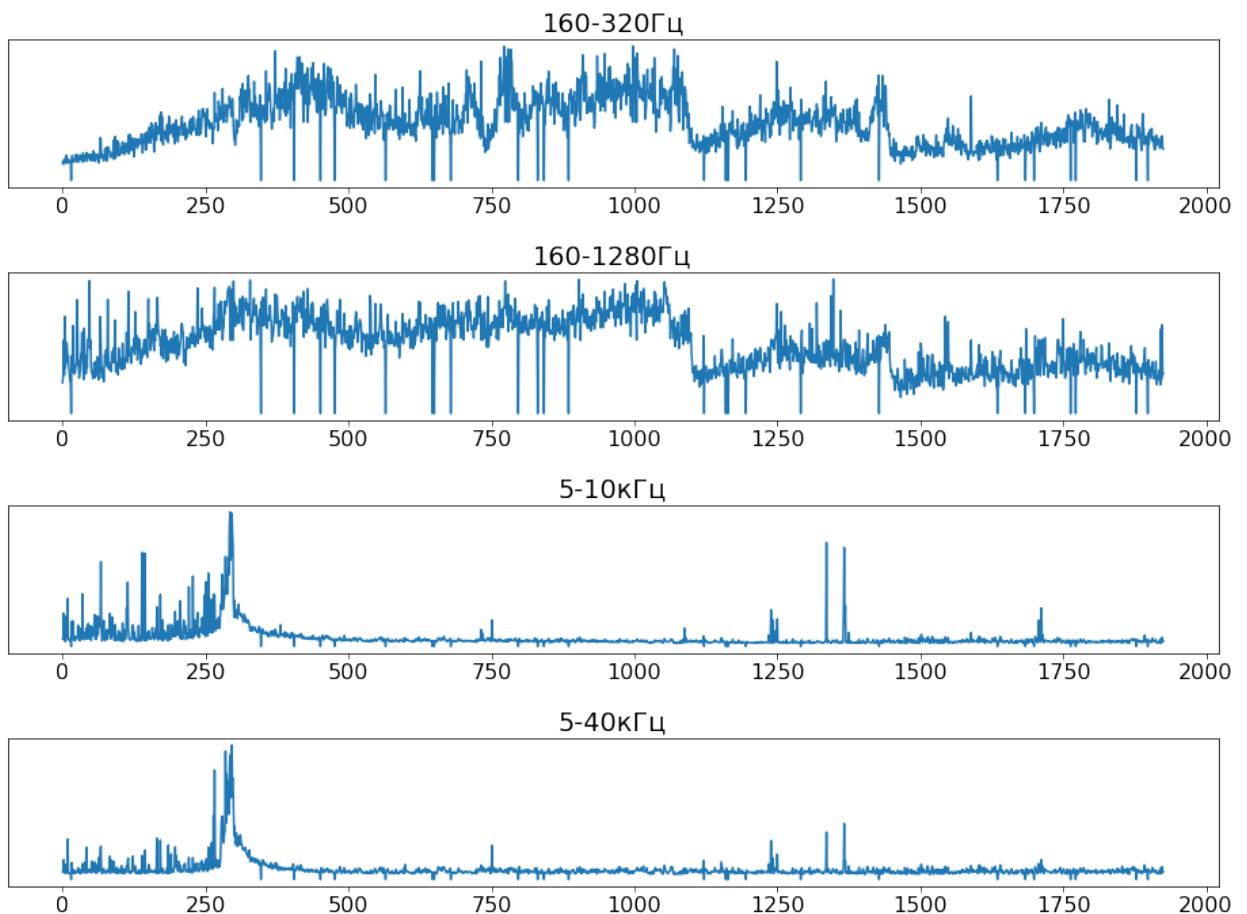


Рис. 23: Выбранные для анализа частотные диапазоны. Суммированием по всем полосам было получено усреднение интенсивности сигнала в каждой точке, которое потом используется для анализа.

Видно, что притоки воды можно локализовать по точкам изменения профиля интенсивности низкочастотных сигналов. Это можно объяснить следующим образом: прибор

при измерении двигался снизу вверх; когда он проходил вдоль перфораций с водными притоками, притоки издавали низкочастотный шум – поэтому общая интенсивность звука увеличивалась ступенчато.

Чтобы отследить это, для первых двух интервалов мы извлекаем описанные в методах разрывные кусочно-линейные профили D_1 и автоматически определяем участки, где скачок профиля был наибольший. Результат показан на рис.24, скачки профиля подсвечены зеленым, искомые зоны – желтым.

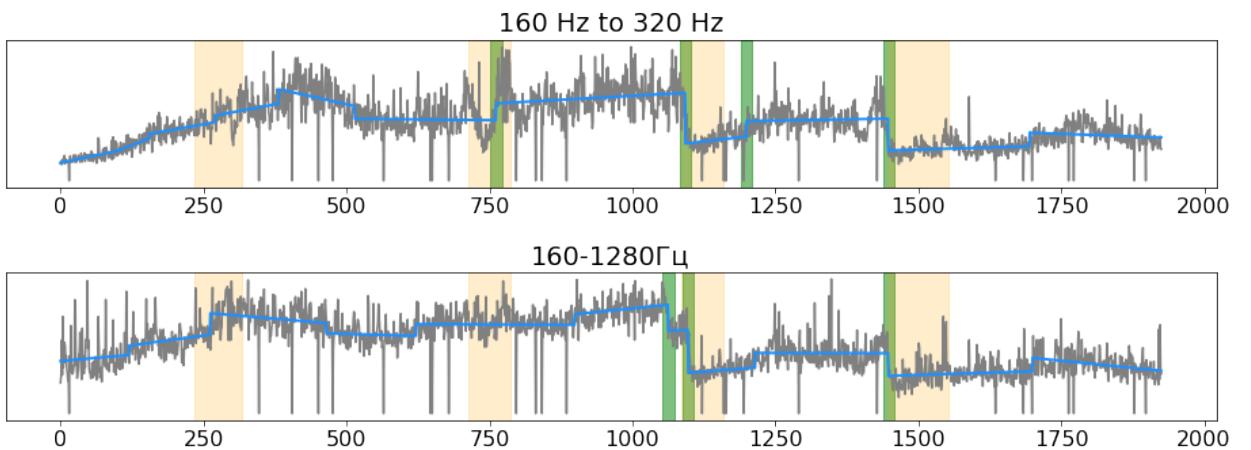


Рис. 24: Результат анализа низкочастотных полос. Желтым указаны интервалы, которые хочется выделить; зеленым – полученные индикаторы. Первая полоса отвечает притоку газа, который не виден на низких частотах; другие три – водные притоки.

Приток газа характеризуется пиком в высокочастотных диапазонах, после которого наблюдается хаотичное изменение интенсивности. Так как событие определяет пик сигнала, то мы смотрим на остаточный сигнал после выделения профиля, а не на сам профиль.

Мы хотим получить характеристику, которая служила бы индикатором хаотичности интенсивности сигнала, которая появляется после прохода зоны газового притока. Для этого сглаживаем модуль остаточного сигнала по окну и снова выделяем профиль – и ищем максимальные скачки аналогично случаю с низкими частотами. В тот момент, когда хаотичность сигнала начинает возрастать, наблюдается скачок полученной кривой. Результат показан на рис.25.

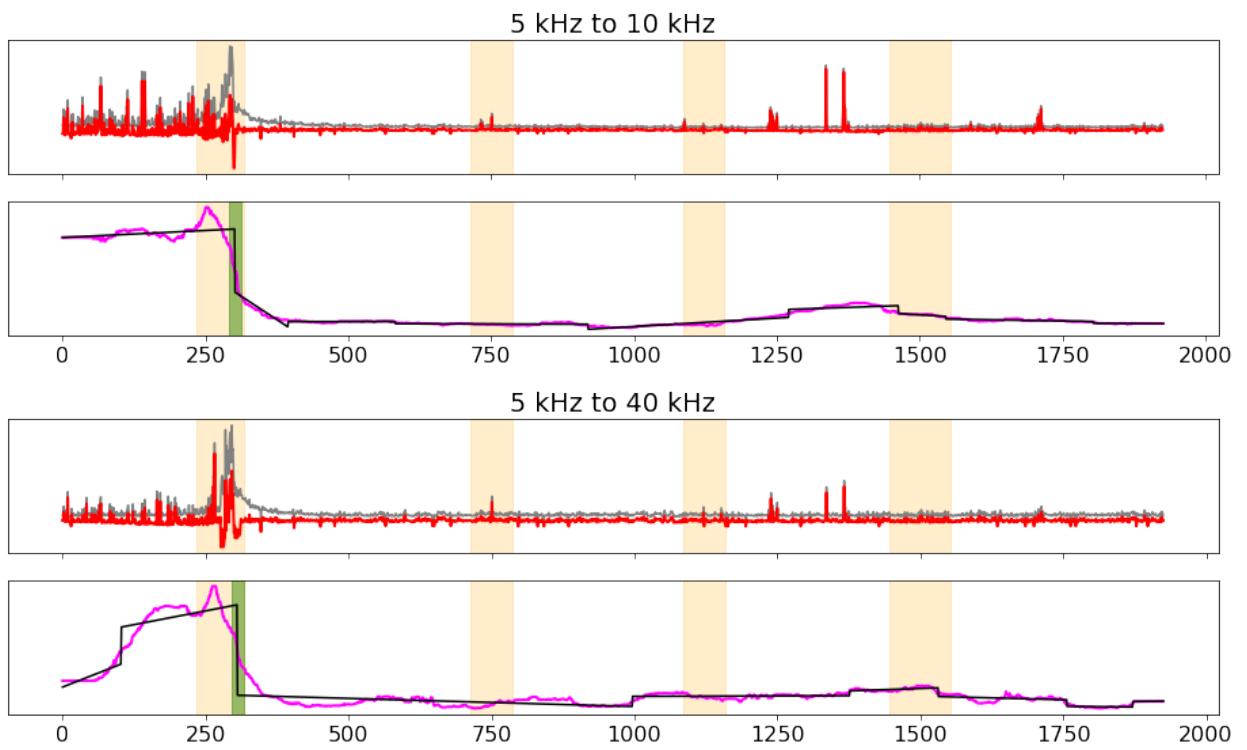


Рис. 25: Результат анализа высокочастотных сигналов. Из изначального сигнала (серый) выделяется профиль и остаточный сигнал (красный). Осцилляции красного сигнала усредняются в розовый сигнал, из которого потом выделяется профиль (черный) и находится точка максимального прыжка - зеленая полоса.

Предварительно выделенные таким образом зоны притока по низкочастотным и высокочастотным сигналам не попадают в точности в интервалы перфораций, указанные на схеме скважины. Тем не менее, их положение примерно соответствует желаемому результату; это показывает, что прибор HFND можно успешно использовать для анализа в комбинации с другими каротажными приборами.

3 Часть 2. Определение фазового состава флюидов вдоль ствола скважины по данным ПГИ

3.1 Актуальность и обоснование выбора темы

Общепринятые методы ПГИ позволяют получать высококачественные данные для вертикальных и околовертикальных скважин [16].

Специально для низкодебитных скважин в Шлюмберже разработан еще один прибор PLT-0.6.3, работа которого основана на явлении термоанемометрии [16], [17]. Интерпретация получаемых им данных проводится с целью определить фазовый состав, распределение и скорости флюидов по длине скважины. Разработаны процессы для калибровки и интерпретации данных термоанемометрии, результаты которых хорошо согласуются с результатами совместной интерпретации данных расходомера, влагометра и других ПГИ данных. Добавление PLT-0.6.3 в сборку приборов повышает качество интерпретации и увеличивает диапазон применимости сборки.

Невозможность получения исчерпывающей информации о потоке флюида в скважине приводит к неоднозначности результатов интерпретации. Для построения наиболее правдоподобной интерпретации, причем с оценкой неопределенности, необходимо иметь различные варианты предсказаний фазового состава флюида на основе имеющейся информации. В данной работе предлагается дополняющий методы [16], [17] новый подход определения фазового состава флюида в скважине на основе измерений датчика термоанемометра.

Рассматривается горизонтальная газо-нефтяная скважина с наличием воды. Используемые для исследования данные получены с помощью каротажной сборки приборов, которая включает в себя базовый модуль PLT-9.2 (измеряемые величины: температура, давление, гамма-каротаж, влагомер, резистивиметр, акустический шумомер) и экспериментальный прототип распределенного термоанемометра PLT-0.6.3. Из временных рядов всех данных ПГИ для анализа мы используем данные температуры и термоанемометрии. При этом данные температуры предварительно корректируем для устранения тепловой инерции датчика.

3.2 Описание данных

Рассматриваемый набор данных (скважина Т1) содержит показания двух типов датчиков: распределенного термоанемометра – шесть одинаковых ТА датчиков, равномерно расположенных по азимуту на окружности в сечении скважины, и вспомогательного

датчика, измеряющего фоновую температуру в центре окружности на оси прибора.

3.2.1 Осевой датчик фоновой температуры

Фоновая температура в центре скважины измерена скважинным прибором PLT-9.2, который включает в себя датчик температуры со следующими характеристиками [4]:

Диапазон измерения температуры, $^{\circ}C$	от 0 до +120(150)
Основная погрешность измерения температуры, $^{\circ}C$	1
Разрешающая способность измерения температуры, $^{\circ}C$	0.001
Тепловая инерция измерения температуры, сек.	1

В исследуемом наборе данных температура записана с нерегулярно меняющимся шагом по времени от $\Delta t_{min} = 0.7$ сек. до $\Delta t_{max} = 4.1$ сек со средним значением $\Delta t_{avg} = 1.8 \pm 0.6$ сек и фиксированным шагом по пространству $\Delta x = 0.1m$.

Разброс значений температуры лежит в интервале от $T_{min} = 27.6^{\circ}C$ до $T_{max} = 28.5^{\circ}C$ со средним значением $T_{avg} = 28.3 \pm 0.2^{\circ}C$.

3.2.2 Азимутально распределенный термоанемометр

Прибор PLT-0.6.3 включает в себя шесть азимутально расположенных термоанемометров. Каждый из этих датчиков работает в цикле нагревания-охлаждения: нагревается 4 секунды и охлаждается 8 секунды. Датчики нагреваются и охлаждаются попарно – пара составляется противоположными по азимуту датчиками. Другими словами, если пронумеровать датчики по кругу номерами $i = 1,..,6$, то общий циклический процесс начиная с произвольного момента времени $t = t_0$ может выглядеть так:

1. $t_0 \leq t < t_0 + 4c$: датчики 1 и 4 нагреваются; датчики 2, 3, 5, 6 охлаждаются.
2. $t_0 + 4c \leq t < t_0 + 8c$: датчики 2 и 5 нагреваются; датчики 1, 3, 4, 6 охлаждаются.
3. $t_0 + 8c \leq t < t_0 + 12c$: датчики 3 и 6 нагреваются; датчики 1, 2, 4, 5 охлаждаются.
4. $t_0 + 12c \leq t < t_0 + 16c$: датчики 1 и 4 нагреваются; датчики 2, 3, 5, 6 охлаждаются,

И так далее.

Запись измерений температуры происходит каждую 0.1 секунды, что дает 120 точек на каждый цикл нагревания-охлаждения. Пример пяти последовательно измеренных циклов нагревания-охлаждения показан на рис.26.

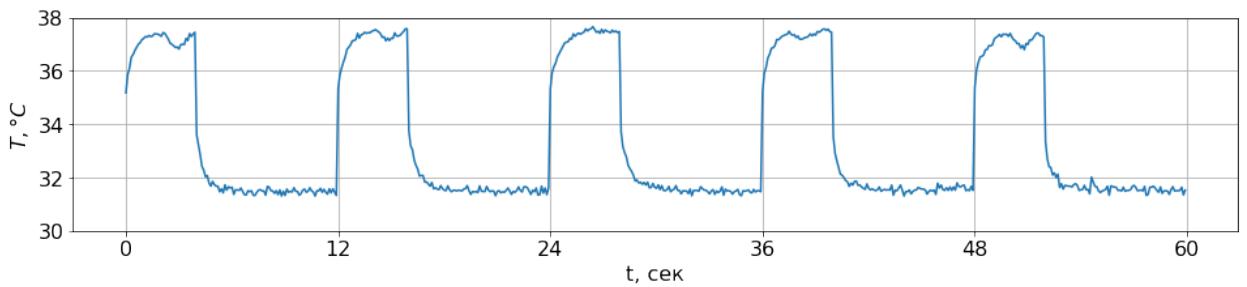


Рис. 26: Пример пять последовательно расположенных циклов нагревания-охлаждения.

3.3 Формулировка задачи

3.3.1 Физическая постановка задачи

Прибор движется по скважине; погруженные во флюид датчики нагреваются и охлаждаются. Характер поведения циклов нагревания-охлаждения зависит от типа флюида и его скорости относительно датчика.

Датчики термоанемометрии достаточно точные, чтобы можно было пренебречь тепловой инерцией и считать все показания измеренными с хорошей точностью; это неприменимо к осевой температуре, которая измеряется гораздо менее точным датчиком. Поэтому осевую температуру необходимо скорректировать с учетом характеристик прибора.

По данным обоих приборов необходимо в общей постановке задачи определить скорость и фазовый состав протекающего через каждый датчик флюида. В данной работе опускается задача вычисления скорости флюидов, в существующих исследованиях выполняющаяся на основе корреляции признаков циклов нагревания-охлаждения с лабораторными данными.

Рассматривается задача определения фазового состава для каждого цикла. Так как скважина горизонтальная, в произвольный момент времени некоторые из шести датчиков могут находиться в разных фазах, поэтому анализ каждого датчика необходимо проводить отдельно. Отметим, что прибор может непроизвольно вращаться вокруг своей оси, продвигаясь вдоль скважины.

3.3.2 Математическая постановка задачи

В качестве датасета рассматривается множество последовательно измеренных одним датчиком циклов нагревания-охлаждения. Под циклом понимается 120 значений температуры, измеренных с шагом 0.1 сек на некотором промежутке глубины вдоль скважи-

ны, и соответствующие этому интервалу глубины одно или несколько значений фоновой осевой температуры. Рассматривается шесть датасетов, соответствующие показаниям шести датчиков; в каждом датасете содержится около 1500 циклов.

Используя доступные данные, для каждого цикла в каждом датасете необходимо предсказать метку: либо одна определенная фаза (вода, нефть, газ), либо смешанная фаза с оценкой фазового состава (например, 70% воды и 30% нефти).

3.4 Используемые методы

3.4.1 Коррекция фоновой температуры

Для использования данных фоновой температуры необходимо учитывать два обстоятельства. Во-первых, по сравнению со временем цикла термоанемометра 12 сек температурный датчик имеет довольно значительную тепловую инерцию, равную 1 сек. (более того, мы используем только 2 сек. из 12 в предлагаемом методе анализа данных, как будет видно далее). Во-вторых, датчики термоанемометрии и измерения фоновой температуры находятся в разных приборах каротажной сборки, что приводит к возникновению естественного сдвига во временных рядах при движении этой сборки в потоке флюидов.

Поэтому необходима коррекция данных фоновой температуры как по учету тепловой инерции, так и по проведению увязки по времени данных термоанемометрии и фоновой температуры.

Тепловая инерция измерения температуры определяется [18] величиной показателя тепловой инерции – времени, необходимого для того, чтобы при внесении датчика в среду с постоянной температурой разность температур среды и датчика составила $e^{-1} \approx 0.37$ того значения, которое будет после наступления теплового равновесия. Таким образом, за время измерения с шагом $\Delta t = 1$ сек датчик фоновой температуры T_{bg} с показателем тепловой инерции 1 сек в каждой выбранной точке успевает достичь значения $\alpha(\Delta t) = 0.63$ от реальной разницы между текущей и предыдущей точкой.

Поскольку в исследуемом наборе данных фоновая температура записана с неравномерным шагом $\Delta t \in [\Delta t_{min}, \Delta t_{max}]$, необходимо вначале построить передаточную функцию $\alpha(\Delta t)$. Характерный вид кривой разгона термометра показан на рис.27(а) [19]. Эту кривую приближенно можно представить как

$$T(t) = T_0 \left(1 - e^{-\frac{t}{\tau}}\right), \quad (12)$$

где ε - коэффициент тепловой инерции [20], рис.27(б). Так как для выбранного датчика $\varepsilon = 1\text{с}$, то для рассматриваемого датчика получаем

$$\alpha(\Delta t) = 1 - e^{-\Delta t} \quad (13)$$

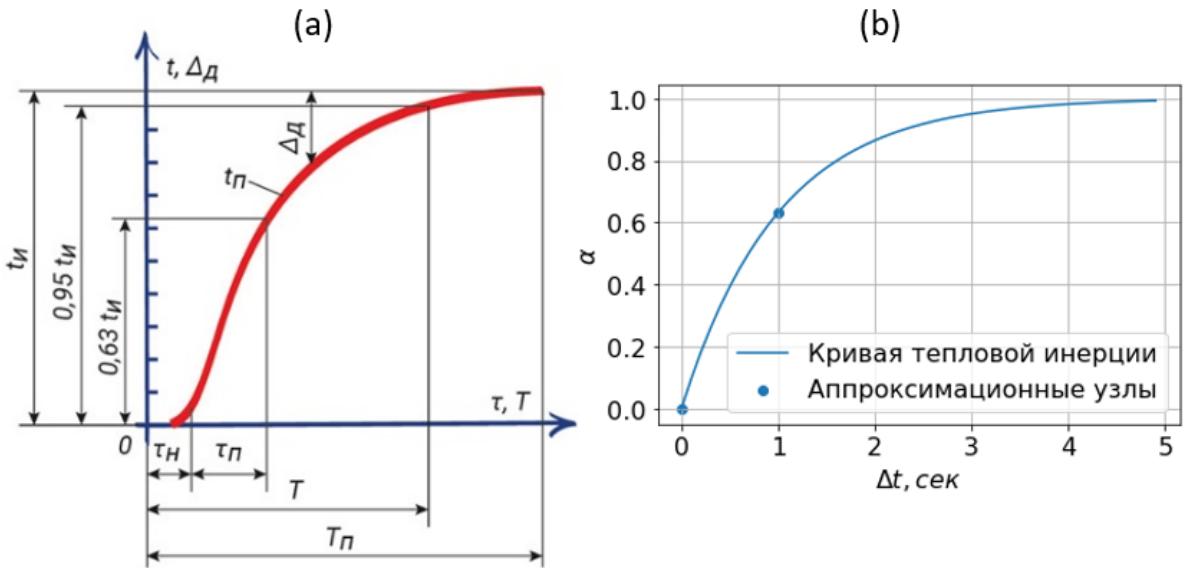


Рис. 27: Характерный (а) и аналитический (б) виды кривой разгона термометра (передаточной функции). Для используемого датчика $\tau_H + \tau_\Pi = 1$.

Для вычисления «истинной» фоновой температуры $T_{bg}(t^i)$ обозначаем данные показаний датчика фоновой температуры через $T_d(t^i)$ и составляем следующее уравнение:

$$T_{bg}(t^{i+1}) - T_d(t^{i+1}) = (1 - \alpha) (T_{bg}(t^{i+1}) - T_{bg}(t^i)) \quad (14)$$

Из него получаем формулу вычисления фоновой температуры:

$$T_{bg}(t^{i+1}) = T_{bg}(t^i) + \frac{T_d(t^{i+1}) - T_{bg}(t^i)}{\alpha} \quad (15)$$

Результат работы алгоритма на некотором участке временного ряда для 50 точек показан на рис.28.

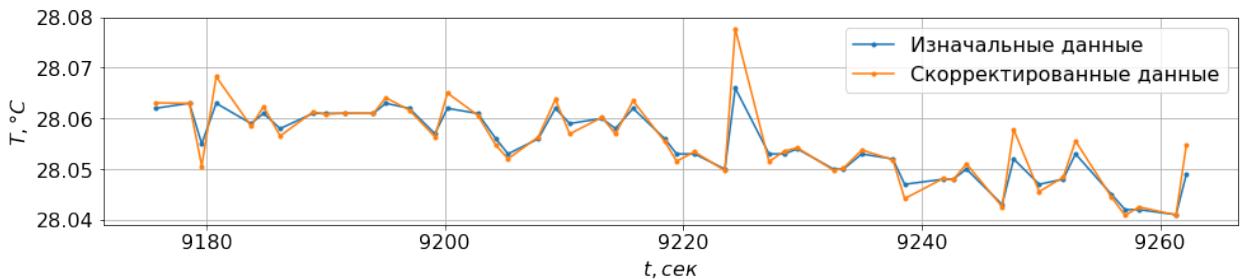


Рис. 28: Результат работы алгоритма коррекции тепловой инерции на 50 точках.

Для коррекции естественного сдвига во временных рядах данных термоанемометрии и фоновой температуры при движении сборки приборов в потоке флюидов мы вводим расстояние L по стволу скважины между датчиками и относительную скорость V приборов и потока. Тогда, очевидно, искомый сдвиг по времени δt оценивается как

$$\delta t = \frac{L}{V}, \quad (16)$$

а вычисление значений $T_{bg}(t^i - \delta t)$ проводится линейной интерполяцией.

Полученные значения фоновой температуры используются для решения задачи разделения флюидов совместно с данными термоанемометра, рассматриваемой в следующем параграфе.

3.4.2 Анализ данных распределенного термоанемометра

Рассматривается задача определения фазы флюида для каждого цикла. Экспериментальные исследования показывают, что форма циклов зависит от скорости флюида и его физических характеристик, см. рис.29.

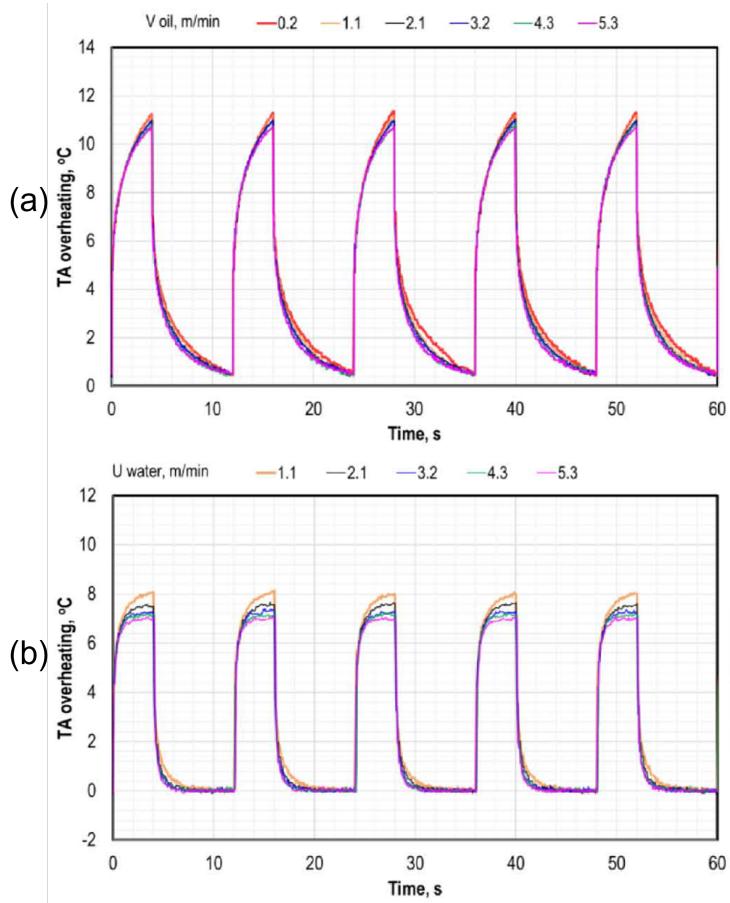


Рис. 29: Характерные циклы нагревания-охлаждения для нефти (a) и газа (b) при различных скоростях потока на протяжении пяти последовательных циклов. Лабораторные данные, результаты статьи [16].

Изменения скорости и характеристик флюидов в определенных пределах могут вызывать изменения численных характеристик кривой графика также в определенных пределах, присущих каждому из трех типов циклов. Назовем такие численные характеристики индикаторами. Например, используются следующие индикаторы:

I_1 - разность максимальной и минимальной температур цикла

I_2 - производная кривой цикла в конце фазы охлаждения

I_3 - производная кривой цикла в конце фазы нагревания

I_4 - интеграл кривой цикла на протяжении фазы охлаждения

I_5 - интеграл кривой цикла на протяжении фазы нагревания

и другие.

На данный момент процесс интерпретации данных термоанемометрии на основе и некоторых других (в том числе, лабораторных) измерений включен в алгоритм обработки ПГИ данных [16], [17].

Из рис.29 видно, что практически все указанные индикаторы хорошо различают две фазы (нефть, вода) независимо от рассмотренных значений скорости потока. Однако, если скорости и параметры флюида претерпевают значительные изменения, то и интервалы изменений индикаторов могут пересекаться для различных фаз. Можно также отметить и другие проблемы обработки измерений, усугубляющие неоднозначность соответствия какого-либо индикатора определенной фазе, например:

- Все датчики термоанемометрии немного отличаются, требуется калибровка
- Использование лабораторных данных требует знания характеристик флюидов и корректировки на скважинные условия
- Распознавание газовой фазы вызывает затруднение из-за близости значений индикаторов для нефти и газа во многих случаях
- За время цикла 12 сек может произойти смена фаз, что приводит к неоднозначности интерпретации индикаторов

Цель исследования - уменьшение влияния указанных проблем на результаты распознавания фаз воды, нефти и газа (в идеале – исключения влияния). Важным условием является также и то, что разрабатываемый метод не должен использовать экспертную

разметку данных, поскольку она недоступна в необходимом объеме, например, для алгоритмов обучения с учителем.

В результате исследований поведения различных циклов из используемого датасета мы разработали метод обработки измерений, в котором подобраны индикаторы и составлены из них комбинации, т.е. признаки, позволившие использовать методы кластеризации для достаточно четкого распознавания всех трех фаз.

Создание признакового пространства. Для построения алгоритма анализировались скважинные данные примерно 1500 циклов одного из датчиков. Обозначим температуру одного цикла как $T(t_i)$, $i = 1, \dots, 120$. Для упрощения обозначений мы опускаем индексы номера цикла и номера датчика из датасета и полагаем, что время соответствует абсолютному значению в датасете. В течение интервала записи данных цикла (12 секунд) датчик может перейти из одной фазы в другую. Чтобы уменьшить вероятность влияния этого события на результат, мы используем только небольшую часть цикла для построения индикаторов, и рассматриваем точки из интервала $i = 40, \dots, 59$, т.е. 2 секунды. Это включает в себя последнюю точку фазы нагревания $i = 40$ и начало фазы охлаждения. Поскольку замеры искажены случайным шумом, мы аппроксимируем точки фазы охлаждения параболой для вычисления значений температуры и ее производной в первой точке фазы охлаждения. Обозначим эту параболу через $T^C(t)$.

После анализа, в том числе и путем перебора, в качестве индикаторов были выбраны

$$T(t_{40}), \quad T^C(t_{41}), \quad \frac{dT^C}{dt}(t_{41}), \quad T_{bg}(t_{41}), \quad (17)$$

где T_{bg} - фоновая температура, вычисляемая согласно алгоритму из предыдущего параграфа.

Чтобы уменьшить влияние скважинных условий и отличий характеристик датчиков, мы проводим обезразмеривание по температуре и формируем два признака из предложенных четырех индикаторов:

$$F_1 = \frac{\frac{dT^C}{dt}(t_{41})}{T(t_{40}) - T^C(t_{41})}, \quad F_2 = \frac{T(t_{40}) - T_{bg}(t_{41})}{T(t_{40}) - T^C(t_{41})} \quad (18)$$

Признаки имеют простой физический смысл: F_1 с размерностью $\left[\frac{1}{\text{сек}}\right]$ характеризует скорость начала охлаждения, а безразмерный F_2 характеризует масштаб перепада температур после окончания нагрева. Отметим, что в обезразмеривании по времени (т.е., исключении $\left[\frac{1}{\text{сек}}\right]$) нет необходимости, так как шаг по времени 0.1 сек в цикле измерений фиксирован для всех датчиков. На рис.30(а) в координатах (F_1, F_2) построены

точки для измерений одного из датчиков. Отчетливо видны три облака с приведенными условными границами (линиями), которые могут соответствовать трем типам флюидов – воде, нефти и газу. Отметим достаточно изотропную форму каждого из облаков. На карте плотности точек, рис.30(б), хорошо выделяются «водяной» и «нефтяной» кластеры («газовый» кластер практически не выделяется из-за существенно меньшего количества точек в нем).

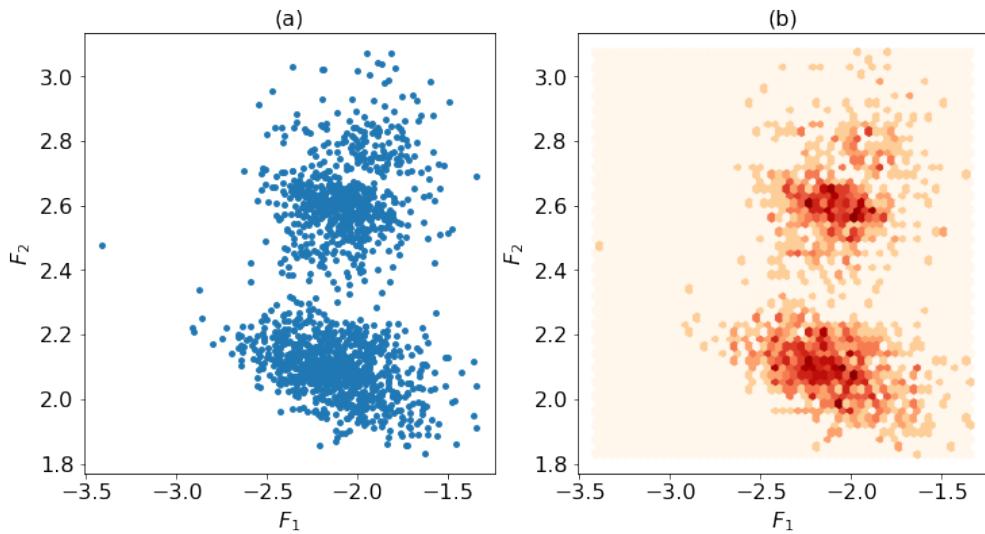


Рис. 30: Скважинные данные термоанемометра для одного из датчиков (около 1500 точек) в координатах предложенных признаков (а) и их раскраска в соответствии с плотностью точек (б)

Кластеризация признакового пространства Разделение данных в полученному пространстве на три кластера делается с помощью модели смеси гауссовых распределений. Так как верхний кластер, который мы будем обозначать как «газ», содержит существенно меньше точек, и он гораздо ближе к среднему кластеру, который мы будем обозначать как «нефть», то кластеризация проводится в два этапа:

1. Модель 1 классификации на 2 гауссовых распределения отделяет нижний кластер («вода») от двух верхних.
2. Оставшиеся после удаления нижнего кластера данных разделяются моделью 2 классификации на 2 гауссовых распределения на два кластера – «нефть» и «газ».

Процесс создания итоговой разметки продемонстрирован на рис.31.

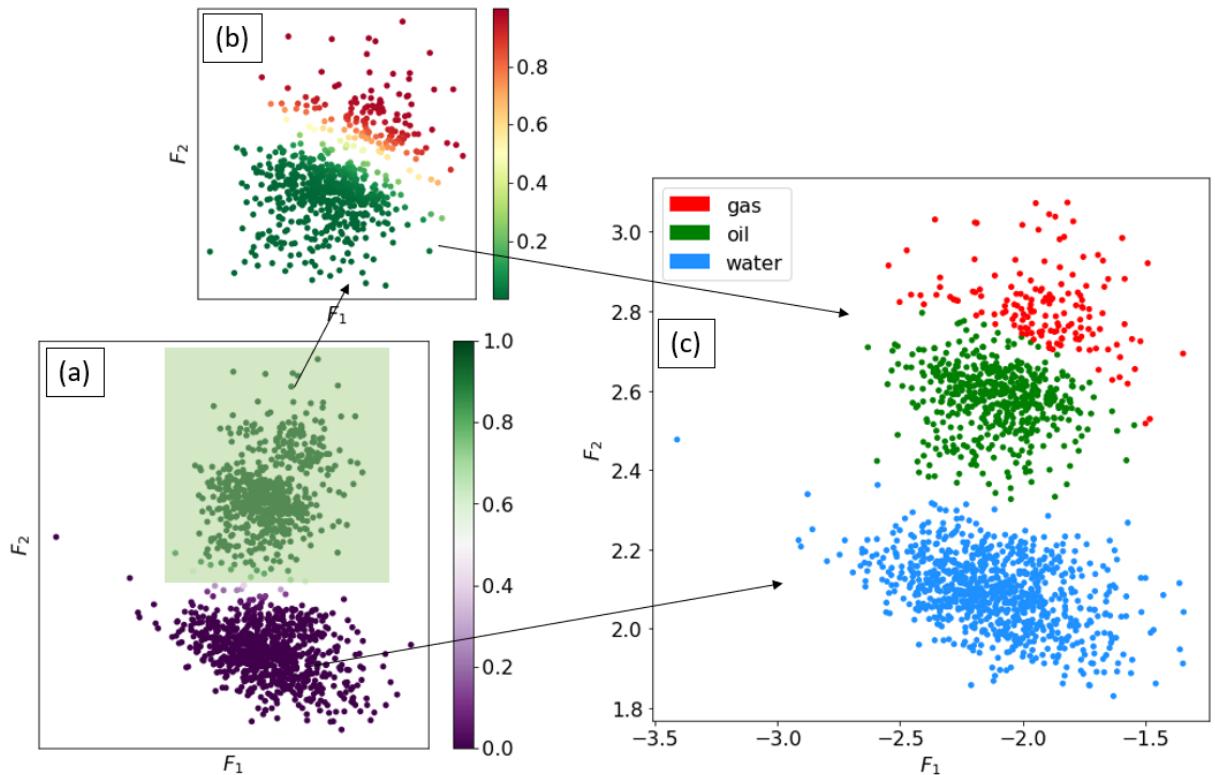


Рис. 31: Процесс создания итоговой разметки. Сначала (а) нижний водяной кластер отделяется от верхнего; верхний кластер разделяется (б) на нефть и газ. Полученные в результате работы двух GMM-моделей метки обединяются в итоговую (с).

Определение положения и фазового состава точек смены фаз Переход между фазами, то есть изменение характеристик цикла, соответствует переходу между кластерами в координатах (F_1, F_2) . Пример перехода между фазами продемонстрирован на рис.32. Раскраска точек соответствует полученной ранее дискретной разметке. Справа показаны восемь последовательно измеренных циклов, на основном графике стрелками тех же цветов соединяются точки в координатах (F_1, F_2) , соответствующие этим циклам. Прибор двигался от меньшего номера цикла к большему, то есть от голубого цвета к розовому. Видно, что изменения формы циклов нагревания-охлаждения соответствуют перемещениям в выбранном пространстве признаков.

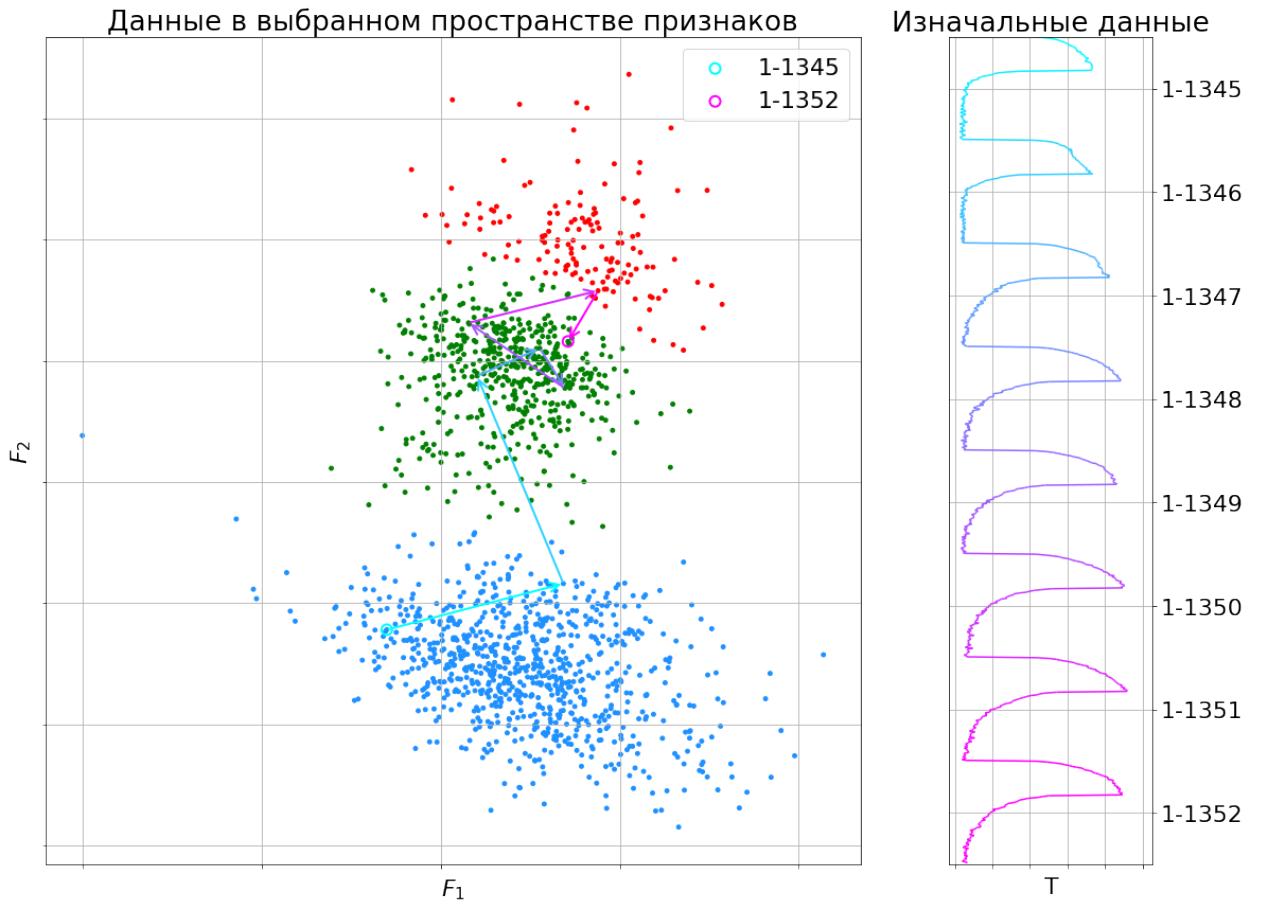


Рис. 32: Демонстрация того, как точки переходят между фазами. На правой колонке видно, как на протяжении 8 циклов меняется форма цикла - и это отражается движением точек, соответствующих этим циклам в выбранном признаковом пространстве. Движение начинается с голубого цикла, находящегося в водяном кластере, и заканчивается в розовом цикле, находящемся на границе нефтяного и газового кластера.

Необходимо автоматически определять точки смены фаз и присваивать им оценку фазового состава флюидов. Так как разделение на кластеры производится вероятностной моделью, то дискретные метки вода-нефть-газ получаются из вероятностных оценок нахождения выбранной точки датасета в каждой из фаз. Так как разделение делается дважды: сначала отделяется вода, а потом разделяются нефть и газ, то на самом деле «сырой» результат работы модели – это два набора вероятностей.

Поэтому:

1. Для каждой точки из класса «вода» доступны вероятности p_{water}^1 и p_{oil}^1 такие, что $p_{water}^1 + p_{oil}^1 = 1$.
2. Для каждой точки из класса «нефть» доступны вероятности p_{water}^1 и p_{oil}^1 такие, что $p_{water}^1 + p_{oil}^1 = 1$; а также вероятности p_{oil}^2 и p_{gas}^2 такие, что $p_{oil}^2 + p_{gas}^2 = 1$.

3. Для каждой точки из класса «газ» доступны вероятности p_{oil}^2 и p_{gas}^2 такие, что $p_{oil}^2 + p_{gas}^2 = 1$.

Форма и процесс получения выделяемых кластеров в координатах ($F_1, F2$) предполагает только два варианта точек смены фаз: «вода-нефть» и «нефть-газ»; вариант «вода-газ» не выделяется, так как соответствующие кластеры получаются двумя различными смесями распределений.

Предложен и реализован следующий алгоритм выделения точек смены фаз и оценки фазового состава. Для модели 1 (отделение воды от нефти и газа) задаются пороги thr_{water}^1 и thr_{oil}^1 такие, что $0 \leq thr_{water}^1 \leq 1$ и $0 \leq thr_{oil}^1 \leq 1$, при этом для существования точек смены фаз необходимо выполнение условия $thr_{water}^1 + thr_{oil}^1 \geq 1$. Для каждого набора вероятностей (p_{water}^1, p_{oil}^1) в зависимости от значений выбранных порогов создается метка М:

Алгоритм для модели 1

1. Если $thr_{water}^1 \leq p_{water}^1$, то М = «вода».
2. Если $1 - thr_{oil}^1 < p_{water}^1 < thr_{water}^1$, то М = « $p_{water}^1\%$ вода+ $p_{oil}^1\%$ нефть».
3. Если $p_{water}^1 \leq 1 - thr_{oil}^1$, то ТМ=«нефть».

Гистограмма распределения значений p_{oil}^1 для датчика 1 показана на рис.33. Для примера указаны значения $thr_{water}^1 = thr_{oil}^1 = 0.9$. Видно, что по результату работы описанного выше алгоритма большая часть точек приобретет метку «вода» (синяя зона) или «нефть» (зеленая зона), в то время как небольшое количество точек в голубой зоне будут классифицированы как смесь флюидов.

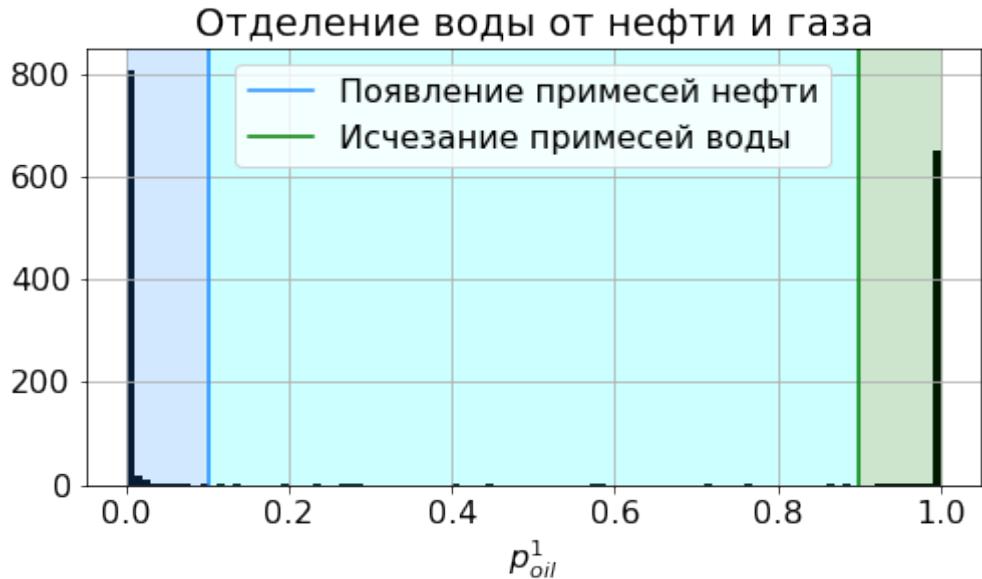


Рис. 33: Распределение значений вероятностей точек не попасть в водяной кластер. Видно, что разделение сильно бимодальное - про большинство точек мы можем с уверенностью сказать, в какой фазе они находятся; но присутствуют и точки между прямыми линиями - им мы присваиваем оценку фазового состава вода+нефть.

Аналогичная процедура проводится для модели 2 (отделение нефти от газа). Задаются пороги thr_{oil}^2 и thr_{gas}^2 такие, что $0 \leq thr_{oil}^2 \leq 1$ и $0 \leq thr_{gas}^2 \leq 1$, при этом для существования точек смены фаз необходимо выполнение условия $thr_{oil}^2 + thr_{gas}^2 \geq 1$.

Алгоритм для модели 2

1. Если $thr_{gas}^2 \leq p_{gas}^2$, то $M = \text{«газ»}$.
2. Если $thr_{oil}^2 < p_{gas}^2 < thr_{gas}^2$, то $M = \langle p_{oil}^2 \% \text{ нефть} + p_{gas}^2 \% \text{ газ} \rangle$.
3. Если $p_{gas}^2 \leq 1 - thr_{oil}^2$, то $M = \text{«нефть»}$.

Гистограмма распределения значений p_{gas}^2 для датчика 1 показана на рис.34. Для примера указаны $thr_{oil}^2 = thr_{gas}^2 = 0.9$. Видно, что по результату работы описанного выше алгоритма большая часть точек приобретет метку «нефть» (зеленая зона) или «газ» (красная зона), в то время как небольшое количество точек в желтой зоне будут классифицированы как смесь флюидов.

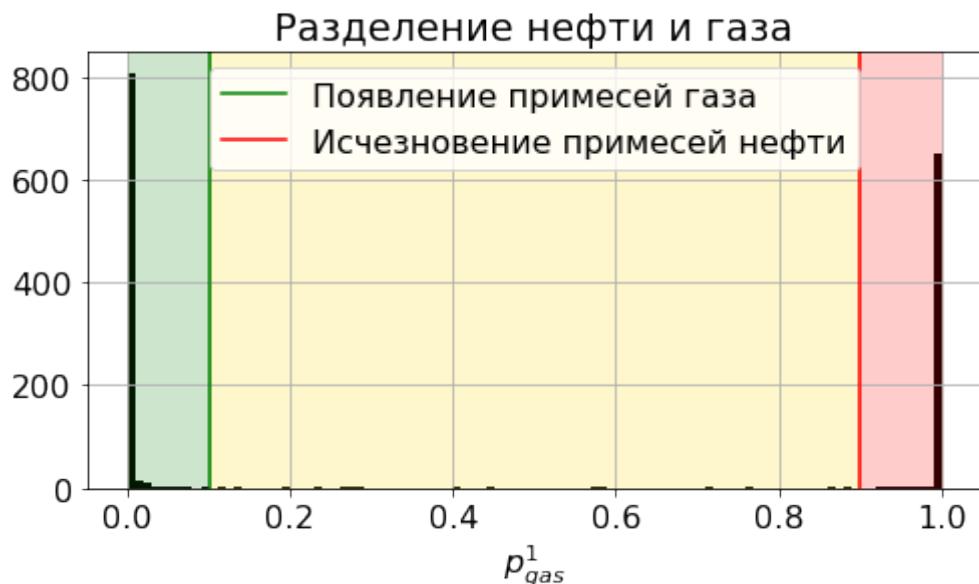


Рис. 34: Разделение нефти и газа происходит аналогично работе модели 1. Здесь распределение опять бимодальное - у большинства точек с уверенностью определяется фаза. Здесь, как и на предыдущей гистограмме, задаются отсечки между дискретной фазой и смесью флюидов.

Так как каждая точка, классифицированная как нефть, имеет как вероятность p_{water}^1 содержать какую-то часть воды, так и вероятность p_{gas}^2 содержания газа, то выбор алгоритма для модели 1 или модели 2 (то есть классификация точки как смеси «вода+нефть» либо «нефть+газ») определяется из сравнения этих вероятностей: если $p_{oil}^1 > p_{oil}^2$, то используется алгоритм для модели 1, и наоборот.

Полученные точки смены фаз показаны на рис.35(а) желтым и голубым цветами; раскраска соответствует гистограммам.

На рис.35(б) представлена финальная разметка с оценкой фазового состава в точках смены фаз.

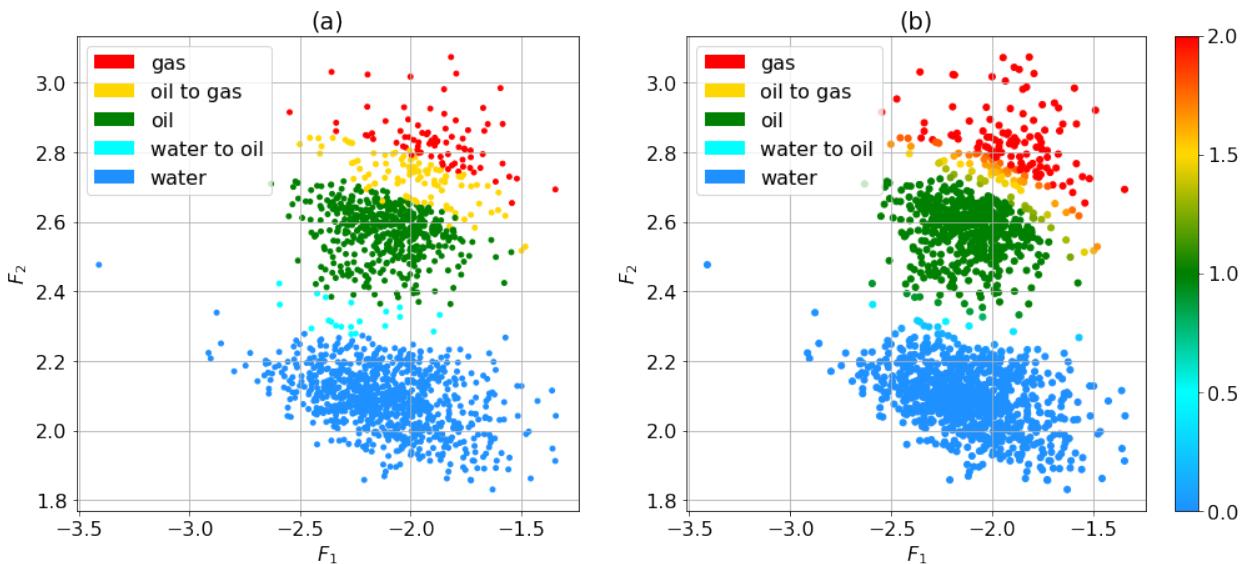


Рис. 35: На (а) - отделение точек смены фаз от точек, содержащих одну фазу, в соответствии с определенными ранее порогами. На (б) - то же разделение, но точкам смены фаз присвоены пропорции фазового состава.

3.5 Результаты

Финальный результат работы алгоритма – разметка с выделением точек смены фаз и оценкой фазового состава – была представлена на рис.35. Так как экспертная разметка была доступна только для случая однозначной классификации каждой точки как принадлежащей одному из трех классов «вода», «нефть» или «газ», то валидацию будем проводить только для первого этапа кластеризации.

После проведения разделения датасета на три кластера по построенным признакам моделью гауссовой смеси распределений получаем разметку на рис.36(а). На рис.36(б). показано сравнение полученной разметки с экспертной разметкой. На экспертной разметке видно, что данные разделяются на кластеры по этим признакам не так явно, как на разметке согласно предложенному методу. При этом основное несовпадение приходится на «нефтяной» кластер: зеленые точки попадают в облака синих и красных точек. Это подтверждается и матрицами ошибок, рис.37: согласно экспертной разметке 23% точек «нефтяного» кластера перешли в воду, а 8% - в газ. Результаты для остальных пяти датчиков показаны в Приложении 1.

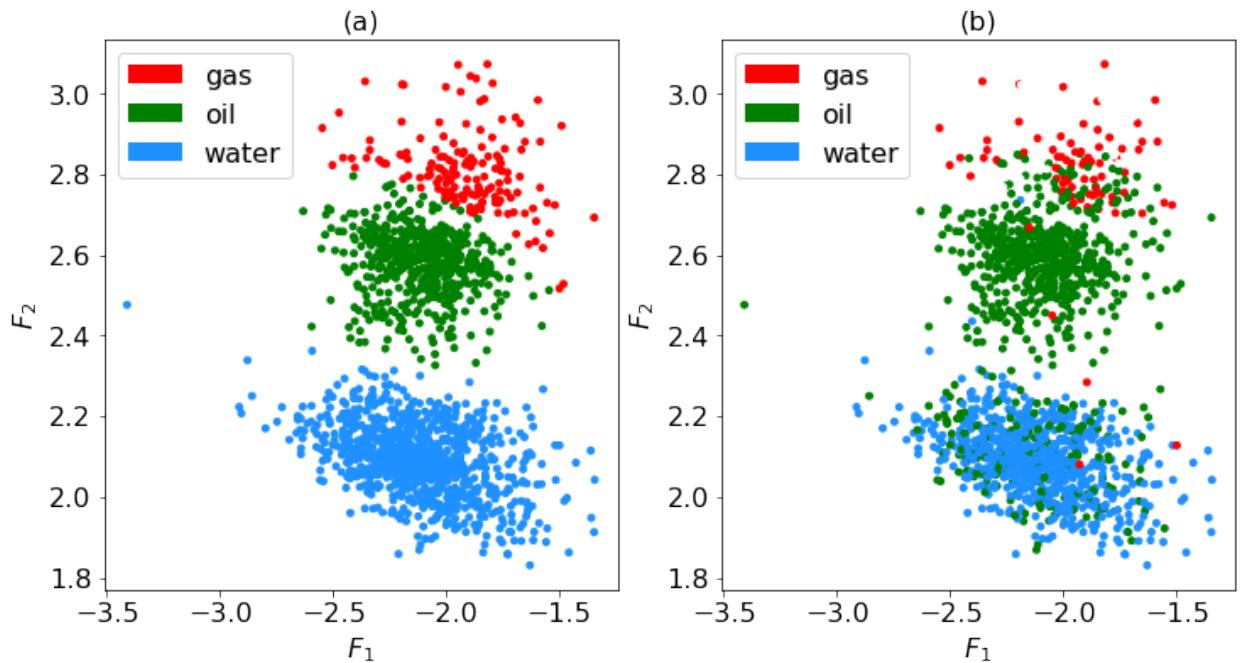


Рис. 36: Сравнение полученной разметки (а) с экспертной (б). Синий цвет – вода, зеленый – нефть, красный – газ.

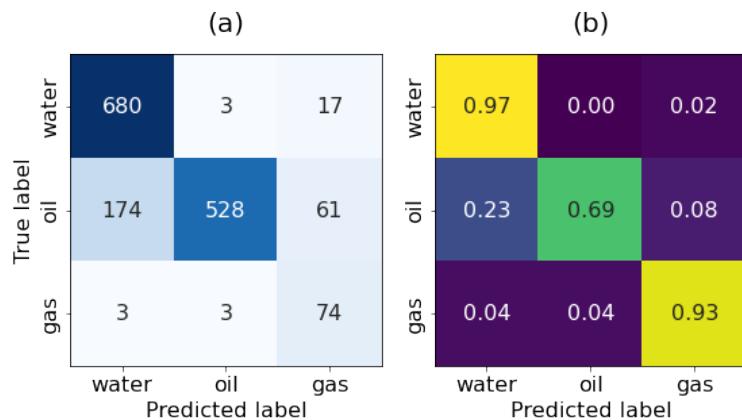


Рис. 37: Матрицы ошибок: (а) количество и (б) доля предсказанных значений.

Одной из причин такого большого расхождения (особенно для фаз нефти и воды, которые обычно очень хорошо распознаются) является смена фаз за интервал времени цикла. В качестве примера на рис.38(d) изображен цикл, попавший в «водяной» кластер по предложенной разметке и в «нефтяной» по экспертной. Видно, что происходит смена фаз, и в начале стадии нагревания цикл действительно больше соответствует нефти, чем воде. Эти же четыре цикла показаны как точки в пространстве выбранных признаков на рис.39.

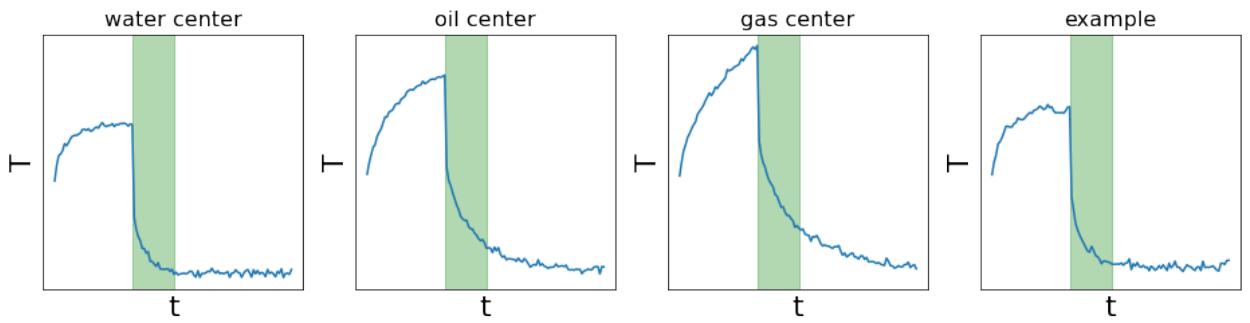


Рис. 38: Примеры циклов. Для первых трех циклов - центров кластеров - предложенная разметка совпала с экспертной. Последний цикл предложенный алгоритмом определил как вода, экспертная разметка - как нефть.

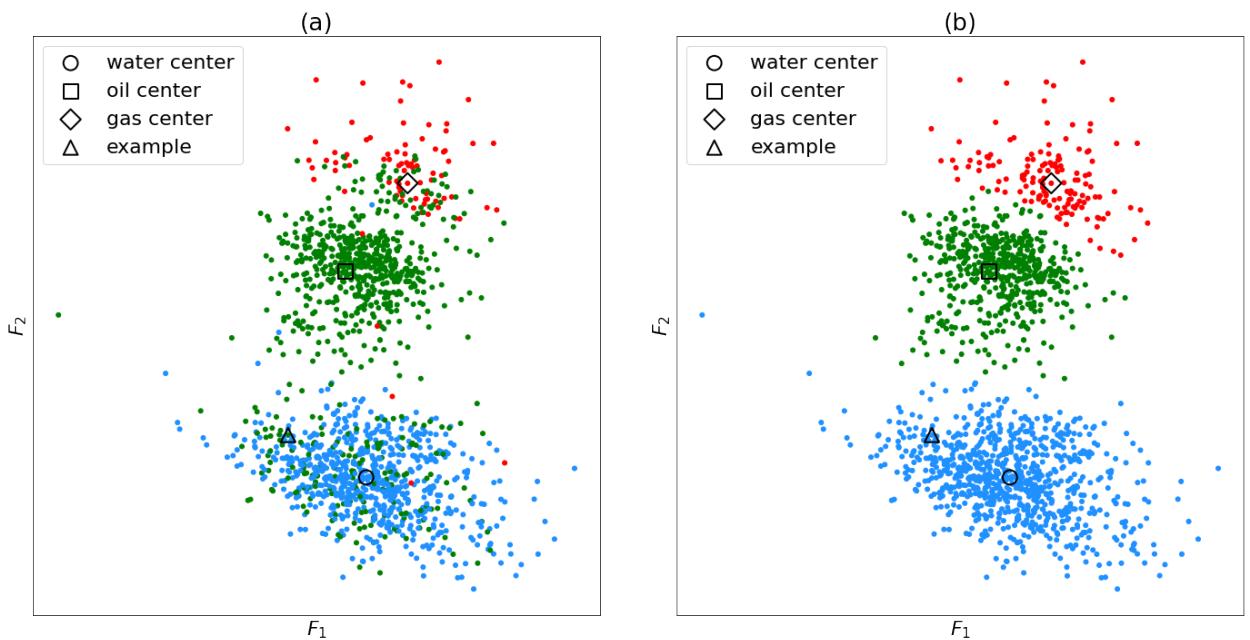


Рис. 39: Циклы с предыдущего рисунка в пространстве выбранных координат.

Следует отметить, что экспертная разметка проводилась на основе единственного индикатора, вычисленного на основе данных каждого цикла нагревания-охлаждения; она не является полностью ручной работой эксперта. Также без 3D-симуляции нужного уровня невозможно получить единственное решение, и в данной главе предлагается лишь один способ снизить неопределенность в интерпретации данных термоанемометрии. Полный цикл анализа данных должен включать в себя другие датчики и в идеале задействовать несколько различных подходов для оценки неопределенности решения; поэтому в данном случае неполное соответствие с экспертной разметкой не является проблемой.

3.6 Обсуждение

В основе экспертной разметки, рассмотренной нами для сравнения, также лежало использование индикатора, вычисленного на основе данных каждого цикла нагревания-охлаждения (в частной беседе экспертом было указано на возможность наличия неточностей из-за неоднозначности распознавания некоторых циклов); поэтому в данном случае неполное соответствие с экспертной разметкой не является проблемой. Предложенный в данной главе новый способ нацелен на снижение неопределенности и ее оценку при интерпретации данных термоанемометрии, поскольку появляется возможность задействовать несколько различных подходов для распознавания фаз, причем в автоматическом режиме.

Оценка фазового состава должна сопровождаться и определением скоростей и масштабного расхода флюидов вдоль скважины. Предложенный здесь способ планируется интегрировать в эту полную задачу путем рассмотрения во взаимосвязи данных от всех шести ТА датчиков в сечении скважины, учета физики течения (например, привлечение уравнения неразрывности), траектории скважины и показаний других датчиков в сборке приборов ПГИ. Автоматизация решения этой полной задачи интерпретации на наш взгляд возможна, но потребует дальнейшего развития предложенного подхода машинного обучения без учителя. Здесь можно будет привлечь методы, разработанные в первой части работы, для выявления надежных и устойчивых индикаторов, характеризующих поток в скважине, а также учета особенностей рассматриваемых типов течений, известных из общепринятых физических моделей, полевых и лабораторных данных, опыта экспертов.

4 Заключение

В работе описаны алгоритмы решения двух типов задач промыслово-геофизических исследований: первый охватывает поиск событийных участков и локализацию зон при тока; второй – оценку распределения фазового состава флюидов по длине скважины. Все исследования проводились на полевых данных реальных скважин.

Использованы алгоритмы машинного обучения без учителя – модель слабого контроля и кластеризация смеси гауссовых распределений; разработаны прототипы программ на языке Python. Во всех задачах в той или иной мере применялись методы цифровой обработки сигналов. При этом уже на этапе цифровой обработки сигналов при подборе параметров фильтрации и т.п. может возникнуть необходимость в машинном обучении без учителя; для этого также возможно использование модели слабого контроля, описанной в первой части.

В целом хочется отметить широкий потенциал модели слабого контроля как модели-советника по принятию решений для эксперта. Представляется, что широкий класс задач анализа данных ПГИ можно формулировать таким образом, что в качестве основы алгоритма будет выступать классификация – на два класса, на несколько или даже с присвоением нескольких меток одному объекту. При наличии достаточного количества экспертных правил (разметочных функций), позволяющих решить задачу в такой формулировке, модель слабого контроля должна выдавать надежные результаты, обоснованные теорией вероятностного машинного обучения.

Важно также отметить следующее. Каждая скважина уникальна, поэтому проверка разработанных алгоритмов требует длительной и кропотливой работы по анализу их применения для всех новых и новых датасетов (скважин). Более того, эта проверка неизбежно влечет за собой улучшение первоначальных алгоритмов, т.е. дальнейшую разработку и модификацию предложенных подходов.

Работа выполнена при финансовой поддержке Московского научно-исследовательского центра «Шлумберже».

5 Список литературы

- [1] Арбузов В. Н. *Эксплуатация нефтяных и газовых скважин*. Томский политехнический университет, 2011.
- [2] *Multistage Hydrofracturing Efficiency Analysis and Horizontal Well Inflow Profiling Using Spectral Noise Logging Technique*, volume Day 3 Wed, October 17, 2018 of *SPE Russian Petroleum Technology Conference*, 10 2018. D033S019R013.
- [3] *Well Integrity Leak Diagnostic Using Fiber-Optic Distributed Temperature Sensing and Production Logging*, volume Day 3 Tue, November 30, 2021 of *SPE Middle East Oil and Gas Show and Conference*, 11 2021. D031S021R001.
- [4] Прибор скважинный plt-9.2. <http://neftegaz.com.ua/skvazhinnaya-apparatura-i-oborudovanie/plt-9-2.html>. Дата обращения: 10.06.2022.
- [5] *Improving Unconventional Completions by Pinpointing Multistage Fracturing with Coiled Tubing: Cases Histories from Burgos Basin*, volume Day 3 Fri, May 23, 2014 of *SPE Latin America and Caribbean Petroleum Engineering Conference*, 05 2014. D031S030R002.
- [6] Lighthill Michael James. On sound generated aerodynamically. i. general theory. *Royal Society*, 1952.
- [7] *Complementing Production Logging with Spectral Noise Analysis to Improve Reservoir Characterisation and Surveillance*, volume All Days of *SPWLA Annual Logging Symposium*, 06 2013. SPWLA-2013-TTT.
- [8] *Identification of Inflow Zones in Low-Rate Horizontal Wells by Spectral Noise Loggin*, volume Day 3 Wed, October 18, 2017 of *SPE Russian Petroleum Technology Conference*, 10 2017. D033S041R002.
- [9] *Innovative Noise and High-Precision Temperature Logging Tool for Diagnosing Complex Well Problems*, volume All Days of *Abu Dhabi International Petroleum Exhibition and Conference*, 11 2012. SPE-161712-MS.
- [10] *Experience of Thermo-Hydrodynamic Studies of Wells in Combination with Noise Logging and Quantitative Interpretation of Data Based on the Simulator*, volume Day 1 Tue, October 22, 2019 of *SPE Russian Petroleum Technology Conference*, 10 2019. D013S003R002.

- [11] *Spectral Noise Logging Data Processing Technology*, volume All Days of SPE Russian Petroleum Technology Conference, 10 2012. SPE-162081-MS.
- [12] Alessio Mazzetto, Cyrus Cousins, Dylan Sam, Stephen H Bach, and Eli Upfal. Adversarial multiclass learning under weak supervision with performance guarantees. *Proceedings of the 38th International Conference on Machine Learning, PMLR 139*, pages 7534–7543, 2021.
- [13] Eckley IA Killick R, Fearnhead P. Pelt algorithm: Optimal detection of changepoints with a linear computational cost. *The Journal of the Acoustical Society of America*, 2012.
- [14] B. Jackson, J.D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and Tun Tao Tsai. An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, 12(2):105–108, 2005.
- [15] A. Ratner, C. D. Sa, S. Wu, D. Selsam, and C. Re. Data programming: Creating large training sets, quickly. *Advances in Neural Information Processing Systems 29*, pages 3567–3575, 2016.
- [16] *New Production Logging Tool for Inflow Profiling of Low-Rate Oil and Water Horizontal Wells: Case Studies of Field Testing an Experimental Prototype*, volume Day 3 Wed, October 18, 2017 of SPE Russian Petroleum Technology Conference, 10 2017. D033S018R002.
- [17] *Field Experience of Integrating Distributed Thermal Anemometer Data Analysis into Production Log Interpretation Workflows*, volume Day 1 Tue, October 22, 2019 of SPE Russian Petroleum Technology Conference, 10 2019. D013S003R003.
- [18] ГОСТ 6616-94 Преобразователи термоэлектрические. Общие технические условия, 1999.
- [19] Н. Ковальчук, И. Пытель, and Л. Тищенко. Способ определения показателя тепловой инерции термопреобразователя. Федеральный институт промышленной собственности, патент SU 1 030 670 A1, МПК G01K 15/00;, 1983.
- [20] А. Разина. Тепловая инерция датчиков температуры. *Журнал «Информатизация и системы управления в промышленности» (ИСУП) № 2(86)*, 2020.

6 Приложение

Здесь приведены результаты работы алгоритма определения фазового состава на шести датчиках одного прибора скважины Т1, описанной в части 2. Далее изображения пронумерованы следующим образом:

- (a) Изначальные данные, преобразованные в выбранных координатах. Изображения (a)-(e) демонстрируют те же данные с различной раскраской.
- (b) Тепловая карта данных, сделанная для демонстрации плотности точек.
- (c) Полученная дискретная кластеризация данных на три фазы: «вода» (синий цвет), нефть (зеленый цвет) и газ (красный цвет).
- (d) Полученная кластеризация с оценкой фазового состава в точках смены фаз. Желтый цвет указывает зоны смены фаз, остальные цвета соответствуют раскраске на (c). Для демонстрации везде выбраны значения порогов

$$thr_{water}^1 = thr_{oil}^1 = 0.99, \quad thr_{oil}^2 = thr_{gas}^2 = 0.8$$

- (e) Экспертная разметка для данного датчика. Цвета соответствуют (c).
- (f) Матрица ошибок для оценки качества кластеризации (c) в сравнении с экспертной разметкой (e). В ячейках – количество точек, соответствующих произвольной комбинации «предсказанная метка – истинная метка».
- (g) Матрица ошибок, построенная аналогично матрице (f), но значения отнормированы по строке.

Видно, что для датчика 3 не удалось достичь хороших результатов – в выбранных координатах не видно четко выраженных кластеров.

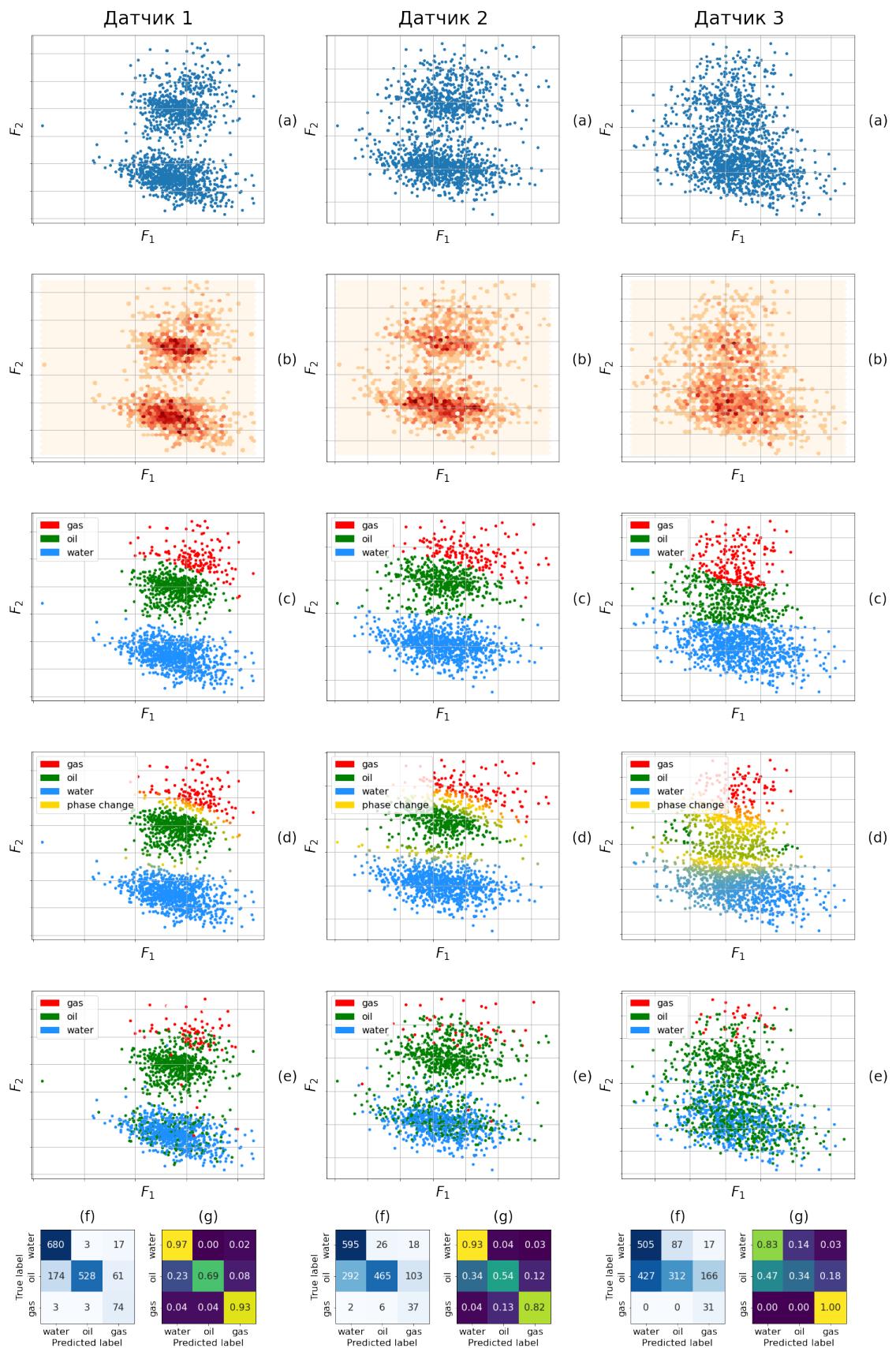


Рис. 40: Датчики 1-3

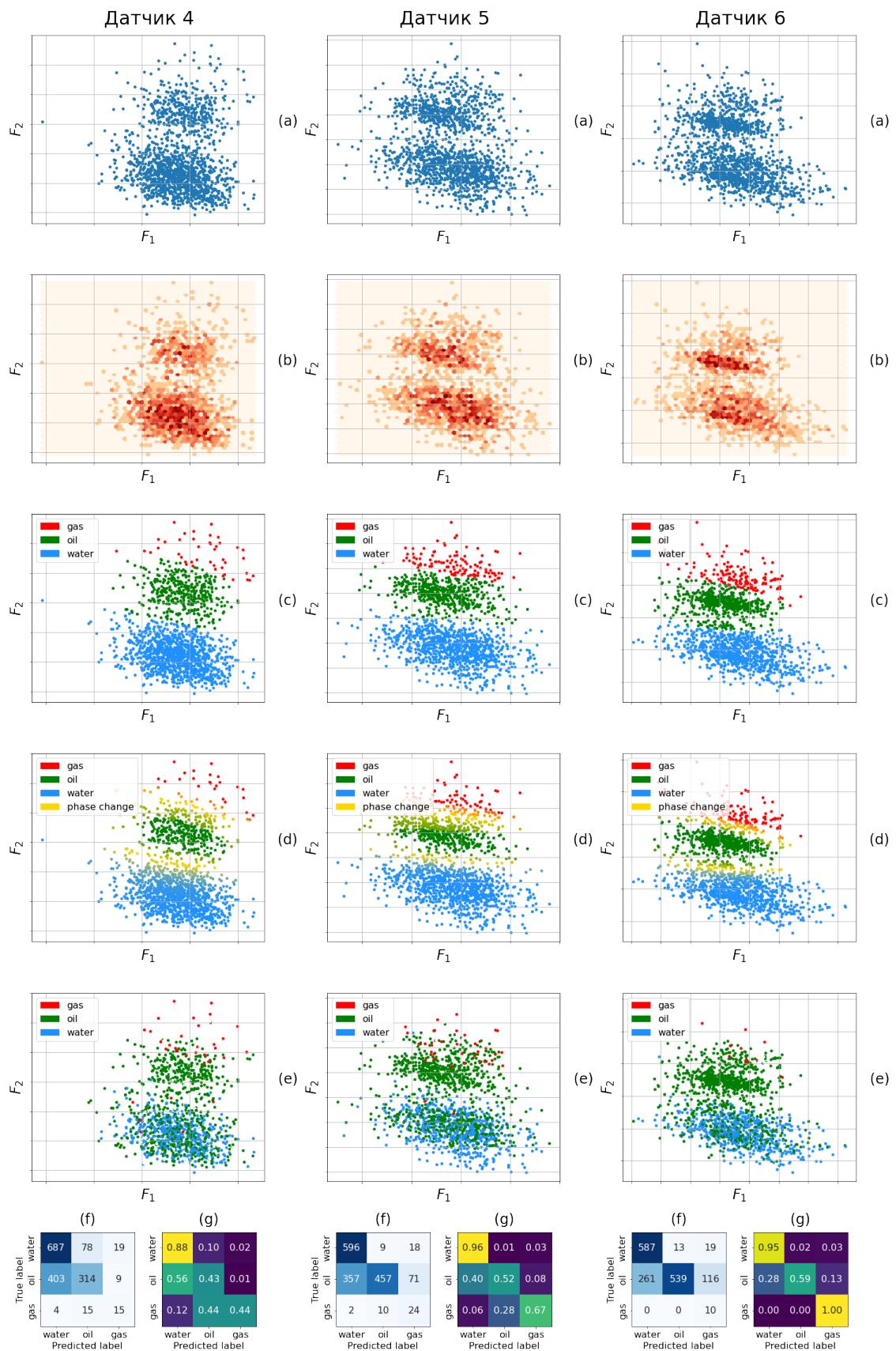


Рис. 41: Датчики 4-5