

УДК 550(34(012:062):344.2)

АВТОМАТИЧЕСКОЕ УДАЛЕНИЕ НИЗКОЧАСТОТНОГО ШУМА ИЗ СЕЙСМОГРАММ С ПОМОЩЬЮ ДВОИЧНЫХ ОТПЕЧАТКОВ: АЛГОРИТМЫ, СВОЙСТВА, ОГРАНИЧЕНИЯ

© 2024 г. К. Ю. Силкин*

*Единая геофизическая служба РАН,
просп. Ленина, 189, Обнинск, 249035 Россия
e-mail: const.silkin@ya.ru

Поступила в редакцию 30.06.2023 г.

После доработки 24.09.2023 г.

Принята к публикации 20.02.2024 г.

Представляемая статья содержит описание новой методики для автоматической подготовки зашумленной сейсмической записи к дальнейшему анализу с помощью экспертных информационных систем. В основу методики положены двоичные отпечатки сейсмограммы, которые благодаря своему лаконичному, но информативному рисунку позволяют хорошо алгоритмизировать поиск ключевых характеристик низкочастотного шума. Особенно важным нахождение оптимальной частоты низкочастотной фильтрации является в условиях частичного перекрытия спектров сигнала и шума при высокой интенсивности последнего. Именно на такой трудный случай и рассчитано проведенное исследование. В статье проводится анализ разработанной методики на примере нескольких сотен регистраций региональных землетрясений и взрывов. По результатам анализа показано, что достоверные результаты решения могут быть достигнуты более чем в 90% случаев. Отдельно упоминаются слабые стороны и ограничения метода, которые являются продолжением его достоинств. В приложении к статье читатель найдет детально описанный алгоритм, положенный в его основу.

Ключевые слова: землетрясение, взрыв, сейсмограмма, вейвлет-преобразование, двоичные отпечатки, низкочастотный шум, фильтрация

DOI: 10.31857/S0203030624030077, EDN: JPXQUG

ВВЕДЕНИЕ

В работе [Силкин, 2022б] были представлены двоичные отпечатки (ДО) сейсмической записи как полезный инструмент, способный найти несколько различных областей применения при анализе сейсмограмм. Во-первых, двоичные отпечатки позволяют сжимать сейсмическую информацию не менее чем на 4 порядка объема. Во-вторых, в сжатом виде эта информация сохраняет ключевые свойства исходных данных, которые могут быть использованы для анализа волнового поля и распознавания природы сейсмического события. Однако, судя по весьма скромному числу отечественных публикаций на эту тему, данная методика еще плохо знакома российским сейсмологам.

Кроме отмеченных особенностей ДО в работе [Силкин, 2023] было показано, что они способны с успехом быть применены к автоматическому определению параметров фильтра, предназначенного для оптимального устранения низкочастотного шума из сейсмической записи с бережным отношением к полезному сигналу в условиях частичного перекрытия спектров обоих. Очевидно, по отношению к зашумленным данным это действие является необходимой предварительной операцией, которую обязательно надо выполнить, а лишь затем приступать к задачам распознавания. С самого начала в этом деле у нас при разработке методик ДО существовала ориентация на автоматизацию указанной операции.

В работе [Силкин, 2023], кроме того, была представлена подборка нескольких близких аналогов к предлагаемому нами методу среди современных приемов устранения шума сейсмограмм. Поскольку задача данной статьи в развитии алгоритмического обеспечения ДО, то здесь мы только еще раз упомянем публикации, которые легли в основу этого обзора:

- метод двумерной фильтрации [Abma, Claerbout, 1995];
- алгоритм нелокального среднего [Bonar, Sacchi, 2012];
- редактирование частотно-временного изображения с последующим восстановлением по нему сигнала [Herrera, 2014; Mousavi et al., 2016a; Асминг, Баранов, 2006; Асминг и др., 2010; Wang et al., 2010; Mousavi, 2017];
- синхронное сжимающее преобразование [Daubechies et al., 2011; Mousavi et al., 2016b; Tary et al., 2018; Shi et al., 2019; Zhang et al., 2022];
- анализ “атомов” времени-частоты на частотно-временном изображении [Mousavi, Langston, 2016];
- анализ связности смежных пикселей на частотно-временном изображении [Zeng et al., 2022];
- синхронное извлекающее преобразование [Yu et al., 2017];
- прочерчивание хребтовых линий на нечетком частотно-временном изображении [Iatsenko et al., 2016].

Благодаря этому обзору было выполнено сравнение метода ДО с аналогами, выявлены их сходства и различия.

К принципиально общему с аналогами относится следующее: перевод сейсмограммы в двумерное изображение; сильное уменьшение объема информации; выявление малых существенных локусов записи; выделение особо ориентированных сигнатур; использование “мягкого” порога при выборе существенной части изображения; поиск минимума графика спектра при перекрытии спектров шума и сигнала.

К важным отличиям предлагаемого метода, которые отличают его от аналогов, относятся: анализу подвергается изображение энергетического представителя вейвлетограммы,

что позволяет оценивать интенсивность всех волновых компонентов любого масштаба одинаково четко; логарифмический масштаб оси частот помогает в этом; анализ изображения производится по выделенным направлениям (горизонтальное и вертикальное) как наиболее часто встречающимся среди наблюдаемых рисунков; применение оптимального порога даже при многократном превышении интенсивности шума над сигналом сохраняет возможность оценки интенсивности частотных характеристик сигнала; автоматическое применение — как изначальный приоритет; отсутствие принципиального требования предварительного анализа характеристик шума до вступления сигнала.

Двоичные отпечатки как таковые не являются нашим изобретением. В статье [Силкин, 2022б] приведен подробный исторический обзор данного направления, включающий следующие публикации: [Fragoulis et al., 2001; Haitma, Kalker, 2002; Wang, 2003; Cano et al., 2005; Baluja, Covell, 2008; Yoon et al., 2015; Ferrari, 2016; Тур и др., 2017; Попов, Замараев, 2019; Bergen, Beroza, 2019]. Там показано как родилась идея формировать для временных рядов компактные образы, чей лаконичный характер отдаленно напоминает дактилоскопические рисунки. Изначально в роли таких рядов выступали аудиозаписи и с помощью отпечатков производился поиск нужных меломану песен в большой базе музыкальных композиций.

Однако вскоре и к сейсмическим записям был применен метод ДО. Цель в данном случае была примерно та же. Однако в силу уникальности каждой сейсмограммы речь шла, главным образом, о детектировании события как такового, выявлении основных черт структуры волнового поля и классификации событий по их происхождению: землетрясение, подземный взрыв, карьерный взрыв и тому подобное. Масштаб решаемых сейсмологических задач — региональный.

Новое, что было принесено в технологию получения двоичных отпечатков работой [Силкин, 2022б], стало использование (вместо спектрограмм на базе скользящего оконного преобразования Фурье) другого, более адекватного подхода — непрерывного вейвлет-преобразования [Suvichakorn et al., 2011; Силкин, 2020]



Рис. 1. Распределение по эпицентральному расстоянию числа проанализированных событий в зависимости от вида источника.

на основе подходящего материнского вейвлета [Mallat, 1989].

Кроме того, нами предложено не саму вейвлетограмму подавать для получения отпечатка, а ее “энергетическое” представление – энергограмму, введенную в статье [Силкин, 2022а]. Энергограмма демонстрирует не анализируемый волновой процесс непосредственно, а характер перераспределения во времени энергии всех его частотных компонент. Фактически энергограмма – это уже частотно-временной портрет события, однако, чрезвычайно избыточный. Поэтому приложение к ней метода ДО стало оправданным и важным благодаря колоссальному сокращению объема информации с помощью отпечатков.

Изначально у других упомянутых исследователей основная идея применения ДО сейсмограмм заключалась в наработке базы данных эталонных образов большого числа событий, чтобы можно было потом сравнивать отпечаток каждого нового события с этими образцами.

Однако, как впервые было показано в работе [Силкин, 2022б], определенные полезные результаты с помощью ДО можно получить сразу, без длительного накопления объемных баз данных. И, прежде всего, к таким результатам относятся упомянутые уже параметры низкочастотного шума, которые необходимо знать для оптимальной фильтрации записи. Тем более, что нерациональным было бы приниматься за решение задачи распознавания природы событий по зашумленным данным. И, если уже организовывать применение двоичных отпечатков, то было бы правильным использовать их еще и для бережной очистки сейсмограмм от низкочастотных помех.

Непосредственный процесс построения ДО после получения вейвлет-энергограммы записи подробно описан в работе [Силкин, 2022б] и

кратко в статье [Силкин, 2023], поэтому здесь мы не будем повторяться.

Целью же данной статьи является детальный разбор методики автоматической оценки свойств шума на основе отпечатков зашумленной записи. Задачи исследования: представление практического алгоритма получения такой оценки, исследование особенностей предлагаемого решения, а также анализ ограничения в его применении.

Именно эти задачи добавляют новое знание относительно статьи [Силкин, 2023], чье содержание было более теоретическим. Она только представила методику ДО, но без конкретизации применявшихся алгоритмов.

ДАННЫЕ

В качестве исходных данных были использованы более 700 регистраций полутора сотен землетрясений и разнообразных взрывов от Карелии до Шпицбергена, записанных десятью станциями, расположенными в Мурманской области и Норвегии. В статье [Силкин, 2022б] на карте изучаемого арктического региона были показаны эпицентры этих событий, а в работе [Силкин, 2023] – расположение участвовавших в регистрации сейсмостанций. Из соображений экономии здесь мы не станем дублировать эти карты, так как для данного исследования они не требуются.

Среди событий присутствует 54 землетрясения, 70 подземных взрывов и 25 карьерных. Из взрывов 12 (как шахтных, так и поверхностных) были произведены по короткозамедленной схеме. Магнитуда событий (если она определена) составила от 1.5 до 4.0, в среднем 2.4. Эпицентральные расстояния при регистрации определены в диапазоне от нескольких десятков км до нескольких сотен км. Распределение числа событий в зависимости от расстояния источник–приемник отдельно для карьерных и шахтных взрывов, а также землетрясений показано на рис. 1. Зарегистрированные землетрясения происходили в основном в Норвежском и Гренландском морях, поэтому они часто более удалены от сейсмостанций, чем взрывы.

МЕТОДИКА

На рис. 2 в графическом виде представлены все ключевые этапы формирования ДО в последовательности их выполнения. Рис. 2а

демонстрирует исходную запись. Для примера взято землетрясение, произошедшее 1 апреля 2020 г. в 04^ч44^м27^с. Эпицентр толчка находился в Кандалакшском районе Мурманской области. Сейсмограмма зарегистрирована станцией LVZ (канал ВНЗ). Эпицентральное расстояние 176 км. Магнитуда события была определена $ML = 3.0$. Отношение максимальных амплитуд сигнала и шума в интервале частот 0.2–20 Гц оценено как $R_{СШ} \approx 3.5$.

Получение частотно-временного образа записи

Вейвлетограмма, рассчитанная с помощью вейвлета Гаусса 8-го порядка [Mallat, 2008], здесь не показана. Также на данном этапе мы уже не обсуждаем проблему выбора материнского вейвлета, так как она была подробно исследована в работе [Силкин, 2023]. Там было отмечено, что для нашей цели базис вейвлет-преобразования имеет, конечно, значение, но не такое большое как при, например, выделении вступлений фаз. Главное, чтобы он был близок по форме к обычному сейсмическому импульсу. Для этого подходит широкий ассортимент семейств вейвлетов и их разновидностей. Выбор между вейвлетом, полученным с помощью производной 6-го, 8-го или 10-го порядка гауссианы, – в общем-то, дело скорее вкуса, чем критический момент методики, так как различия у получаемых в итоге ДО не существенны.

На рис. 2б сразу приведена полученная на основе вейвлетограммы энергограмма. Причем, чтобы лучше продемонстрировать все, в том числе и слабые волновые составляющие, применен логарифмический масштаб значений этой функции. Стрелка с обозначением “ $f_B - f_H$ ” указывает на середину (f_C) между двумя горизонтальными синими пунктирами. Данные линии прочерчивают сделанный оператором выбор допустимых пределов (верхнего f_B и нижнего f_H) расположения частотной границы шума, которая будет использоваться при задании фильтра. Эти пределы используются для контроля качества работы вводимого здесь алгоритма по выявлению параметров шума. В данном случае $f_B = 0.52$ Гц, $f_H = 0.83$ Гц, $f_C = 0.68$ Гц.

Поверх энергограммы (для сопоставления) красной линией нанесен амплитудный спектр Фурье этой записи.

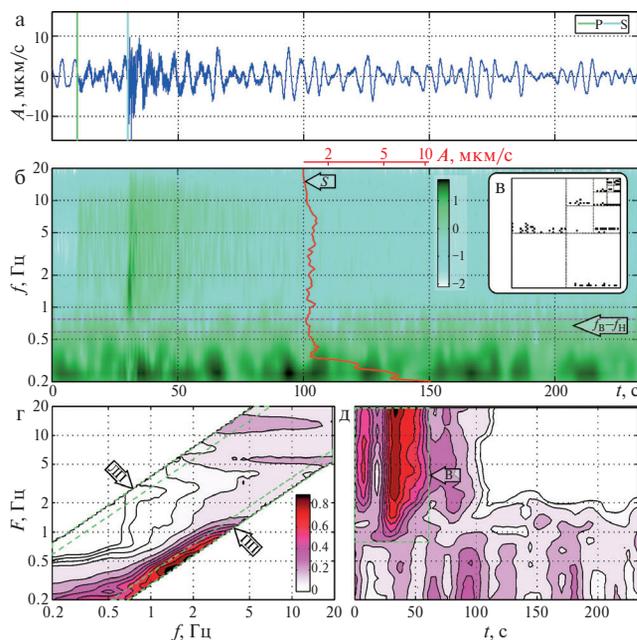


Рис. 2. Основные элементы технологии двоичных отпечатков.

а – запись канала ВНЗ станции LVZ землетрясения 1 апреля 2020 г. в 04^ч44^м27^с: зеленая и бирюзовая линии – отметки вступления Р- и S-волн; б – энергограмма по вейвлетограмме записи, ось времени относительная, цветовая шкала – логарифм значений амплитуды, стрелкой “S” показан наложенный амплитудный спектр Фурье записи, горизонтальная ось значений которого наверху красная, вертикальная ось соответствует оси энергограммы, стрелка с обозначением “ $f_B - f_H$ ” помечает экпертный интервал допустимых пределов границы фона (горизонтальные синие пунктирные линии) и указывает на его середину; в – двоичные отпечатки энергограммы; г, д – избыточные двоичные отпечатки (г – скользящий вейвлет-спектр, горизонтальная ось – частота вейвлет-разложения, вертикальная ось – положение сканирующего частотного окна, цветовая шкала значений – условная плотность отпечатков, стрелки “ЛП” и “ПП” отмечают левую и правую полосы для расчета сечений скользящего вейвлет-спектра (зеленый пунктир), при этом стрелка “ПП” указывает на зону сильного градиента; д – скользящая вейвлет огибающая, вертикальная ось и шкала амплитуд соответствует скользящему вейвлет-спектру, стрелка “В” отмечает область проявления регулярных волн (зеленый пунктир)).

Двоичные отпечатки энергограммы

Как врезка поверх малоинформативной части энергограммы на рис. 2в демонстрируются двоичные отпечатки, рассчитанные по данной энергограмме, в виде централизованного и бинаризованного созвездия максимальных

коэффициентов двумерного дискретного вейвлет-разложения Хаара [Jensen, la Cour-Harbo, 2011; Hussain et al., 2020; Фортуна-Сервантес и др., 2021] децимированной энергограммы. Отпечатки представляют собой фрактальноподобный композит из нескольких последовательных уровней разложения исходной энергограммы на горизонтальные, диагональные и вертикальные составляющие.

Применение в данном случае вейвлет-преобразования с четко определенным материнским вейвлетом (Хаара) преследует 2 задачи. Во-первых, это сжатие информации, а простой, похожий на две разнонаправленные ступеньки вейвлет Хаара, прекрасно с этим справляется [Ranjan et al., 2022]. Во-вторых, благодаря свойствам преобразования – еще и раздельный анализ специфически ориентированных элементов изображения.

На нескольких уровнях горизонтальной составляющей (см. рис. 2в) невооруженным глазом можно видеть четко прочерченные полосы, соответствующие проявлению шума. Остается только должным образом проинтерпретировать их и оценить меру доверия к получаемым результатам.

Избыточные двоичные отпечатки.

Более наглядно интерпретацию отпечатков можно сделать на примере избыточных двоичных отпечатков (ИДО), которые, как показано в статье [Силкин, 2022б], и были изначально задуманы только для этой цели, а потом нашли самостоятельное применение.

Скользящий вейвлет-спектр. Рис. 2г демонстрирует одну из двух ипостасей ИДО – скользящий вейвлет-спектр (СВ-С). Избыточные отпечатки получаются путем расчета простых ДО в скользящем частотном окне, текущее положение середины которого показано на СВ-С с помощью вертикальной оси F . Шаг между окнами и ширина их, а также масштаб оси F – логарифмические. Горизонтальная же ось f отмечает частоту вейвлет-спектра, полученного по результатам осреднения значений *горизонтальной* составляющей нескольких младших уровней разложения Хаара при данном положении скользящего окна. Так как СВ-С рассчитывается в окне с шириной в четверть числа

дискретов частоты исходной энергограммы, то он имеет вид полосы, протянувшейся по диагонали от малых значений f и F к большим.

Замечено в серии опытов, что подобный рисунок (присутствие области высоких значений внизу, часто в виде горизонтальной полосы, нередко с косым отростком, протянувшимся вдоль правой стороны, отмечающим переходный диапазон частот между частично накладывающимися спектрами сигнала и шума) на СВ-С неизменно появляется всякий раз, когда на исходной записи присутствует стационарный низкочастотный шум, сменяемый на более высоких частотах сейсмическими колебаниями иной природы, например, регулярными от землетрясения. Причем правая сторона СВ-С более надежно указывает на частотный рубеж смены волновой картины. Важно, что эта граница тяготеет к участку наиболее сильного градиента от максимума правой стороны по пути к выходу поверхности СВ-С на плато. Именно здесь интенсивность регулярных волн начинает преобладать над низкочастотным шумом. Фильтрация по данной границе будет означать не идеальное устранение помехи, а оптимальное подавление с бережным сохранением полезного сигнала, насколько это возможно.

Одновременно левая сторона СВ-С иногда может оказать поддержку в этом поиске, так как, если на правой стороне имеется такой характерный косой отросток, то сильный линейный градиент СВ-С слева обычно соответствует максимуму справа. В случае слабого проявления переходной зоны такое соответствие не наблюдается, и левая сторона оказывается менее информативной. Тем не менее, в условиях нередкого наложения случайных мешающих факторов и такая подсказка бывает полезной. Хотя, разумеется, помехи могут сказываться и на левой стороне СВ-С, но контрольный источник информации избыточным не станет.

Стрелками “ЛП” и “ПП” на СВ-С отмечены подведенные зеленым пунктиром полосы шириной в 10% от размаха рассчитанного скользящего вейвлет-спектра, соответственно, левая полоса и правая полоса. Эти полосы используются для расчета характерных графиков профилей СВ-С Л и СВ-С П, по которым определяется верхняя частотная граница шума.

Детально это прописано в приложении (подраздел “Избыточные отпечатки”).

Скользящая вейвлет-оггибающая. На рис. 2д представлена скользящая вейвлет-оггибающая (СВ-О), являющаяся второй независимой стороной единой функции ИДО данной записи. При этом СВ-О рассчитывается по *вертикальным* составляющим нескольких уровней разложения Хаара энергограммы. На представленном рисунке стрелка “В” отмечает выделенную зеленым пунктиром область существования в двумерном частотно-временном пространстве исключительно регулярных волн.

Нижняя кромка этой области также используется для оценки верхней границы шума, так как здесь аналогичная картина, которая наблюдается и на СВ-С, только это взгляд с другой стороны. На СВ-О мы видим, что феномены высокочастотных регулярных волн от землетрясения и их энергетические проявления исчезают, когда в области низких частот им на смену приходит колебательный процесс иной природы – стационарный шум. Задача алгоритма в данном случае выявить точку выхода на плато после спуска с вершины максимума в сторону понижения частоты. Этот рубеж не будет означать идеального сохранения полезного сигнала, но даст возможность устранить помеху настолько тщательно, как это только возможно.

Подготовка контрольных параметров

Предварительно, чтобы оценить качество алгоритма для определения свойств низкочастотного шума по каждой записи, оператором с помощью всех доступных ему в такой работе средств были отмечены два уровня частоты (они были продемонстрированы на примере рис. 2б). Эти значения ограничивают снизу и сверху допустимое с точки зрения оператора положение границы между шумом и полезным сигналом. В силу индивидуальности каждой сейсмограммы ширина полосы между этими уровнями могла принимать разные значения. Гистограмма для данного параметра показана на рис. 3.

Можно видеть, что распределение значений интервалов ширины (в логарифмическом масштабе) близко к нормальному с наиболее частотным значением около 0.5 Гц. Это говорит



Рис. 3. Гистограмма распределения ширины экспертного диапазона частот, содержащего допустимое значение верхней границы шума для оптимальной фильтрации.

о том, что при задании частот f_B и f_H оператор старался быть аккуратным и по возможности не придал допустимому интервалу между ними слишком большую ширину.

Принцип, заложенный в выделение верхнего и нижнего пределов допустимого интервала частот, которому может принадлежать оптимальная граница шума, основан на компромиссе между идеальным сохранением полезного сигнала и идеальным устранением шума при фильтрации. Оператор, отмечая эти пределы, полагал, что f_B хотя и обрезает сигнал, но не сильно, а f_H хотя и не совсем исключает шум, но существенно. Разумеется, все это субъективные оценочные суждения, основанные, тем не менее, на экспертном опыте. Ожидается, что алгоритм двоичных отпечатков должен выдать в итоге значение в пределах $f_B - f_H$, желательно вблизи середины f_C между ними.

В качестве меры попадания в цель формулируется гауссиана $g(f)$, построенная так, чтобы она удовлетворяла условиям: $g(f_C) = 1$ и $g(f_B) = g(f_H) = 0.7$. При этих вводных колоколообразная функция качества результата определяется нами следующим образом:

$$g(f) = \exp \left[\frac{4 \ln 0.7}{(f_B - f_H)^2} \left(\frac{f_B + f_H}{2} - f \right)^2 \right].$$

Значение 0.7 при достижении пределов f_B и f_H выбрано исходя из общераспространенного в статистике критерия, согласно которому именно такое значение принимается в качестве

значимого уровня [Баврина, Борисов, 2021]. Однако при выходе за указанные пределы $g(f)$ становится, пусть и мене значимым, но все же не нулевым.

Кстати, для рассмотренного выше примера записи кандалакшского землетрясения были получены следующие значения функции качества: $g_{\text{ДО}} = 0.99$ и $g_{\text{ИДО}} = 0.87$. Первый можно признать идеальным, второй – вполне достоверным.

Алгоритм определения границы шума подробно описан в приложении. Там же даются указания к измерению степени доверия к получаемому результату. Далее мы перейдем к анализу полученных результатов по всем 716 записями, вовлеченным в проведенное исследование.

АНАЛИЗ

Предварительные замечания

Прежде чем приступать к анализу отметим, что даже в таком относительно несложном случае, представленном на рис. 2, где имеется не очень сильный шум, спектр которого довольно скромно перекрывается со спектром сигнала, тем не менее, простой Фурье-спектр совсем не работает для нашей цели. На рис. 2б с сопоставимым наложением приведен амплитудный спектр исходной записи. Он хорошо демонстрирует интервал энергичного низкочастотного шума. Однако график спектра здесь сходит на уровень низкого плато, содержащего и регулярные волны, слишком рано, на уровне 0.36 Гц, где шум еще преобладает над большей частью сигнала. Справедливости ради, следует отметить, что существенный минимум спектра Фурье наблюдается на частоте 0.77 Гц, что близко к f_C (0.68 Гц). Однако он так слабо выступает от линии плато, что нет никаких шансов пытаться обнаружить его в реальных условиях.

Конечно, если бы заранее было известно положение на оси времени основных проявлений энергии регулярных волн от этого землетрясения, то спектр Фурье выдал бы более четко разделяемый результат. Однако предлагаемый алгоритм предназначен работать на этапе, когда такое знание еще не получено или его достоверность подвергается сомнению из-за наложения интенсивного шума. Для выявления фаз регулярных волн хорошо зарекомендовал себя

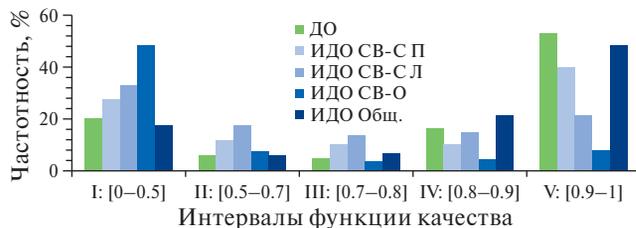


Рис. 4. Гистограммы распределения функции качества результатов алгоритма двоичных отпечатков с разделением на классы.

Условные обозначения содержат виды вычисляемых оценок качества для: ДО – двоичных отпечатков (простых); ИДО СВ С П – правого профиля скользящего вейвлет-спектра; ИДО СВ С Л – аналогично левого профиля; ИДО СВ О – профиля скользящей вейвлет-оггибающей; ИДО Общ. – итогового обобщения результатов разных сторон ИДО.

метод СТА/LTA, однако он начинается с применения к сейсмограмме высокочастотного фильтра, по результатам которого уже работает вычисление отношения средних значений во временных окнах разной ширины [Chen, Stewart, 2006; Асминг, Федоров, 2014]. Как определить нижний предел для такого фильтра – вопрос, требующий отдельного решения, на поиск которого и направлено наше исследование.

Таким образом, только тщательный частотно-временной анализ дает надежду решить поставленную задачу достоверно, тем более, если применять его в автоматическом режиме.

Общий анализ результатов

Итак, разделим область значений функции $g(f)$ на 5 классов, как показывает рис. 4, и определим частотность появления отклика алгоритма в том или ином классе качества. При этом классы III–V считаются достоверным результатом, так как покрывают интервал 0.7–1.0. Ответы алгоритма ИДО показаны по разделениям на его ипостаси и стороны: правый профиль СВ-С, левый профиль СВ-С и профиль СВ-О.

Итоговый ответ, выдаваемый за результат ИДО, формируется в соответствии со сложной системой измерения доверия к ответу каждой сотрудничающей части ИДО таким образом, чтобы постараться взять от них только самое удачное и по возможности проигнорировать промахи. Можно видеть, что по большей части такой подход оправдался, так как ни одна часть алгоритма для ИДО не равнозначна их итоговому обобщению. Более

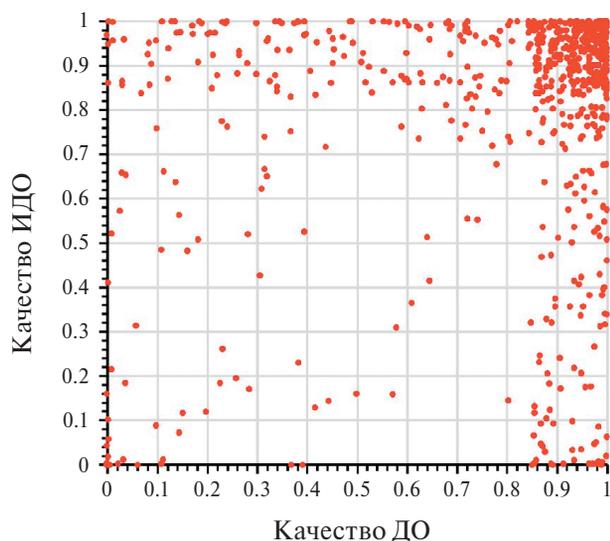


Рис. 5. Диаграмма рассеяния значений качества результатов простых двоичных отпечатков (ДО) и итоговых обобщенных избыточных двоичных отпечатков (ИДО).

того, суммарный итог превосходит по качеству любую свою исходную часть.

Рис. 4 демонстрирует наибольшую предсказуемость у класса качества V, то есть алгоритм двоичных отпечатков (обычных и избыточных) действительно чаще всего дает результат, близкий к f_c . Если посчитать какова суммарная частотность попадания ответов в достоверные классы III–V, то для простых отпечатков это будет 74.1%, а для избыточных — 76.4%. Результат ИДО на первый взгляд может показаться недостаточно высоким. Можно задаться вопросом: “Зачем строить что-то избыточное, если оно не на много лучше простого?” Ответ в том, что алгоритм избыточных отпечатков не полностью дублирует, а временами дополняет алгоритм простых, что хорошо видно на рис. 5.

Диаграмма рассеяния показывает в правом верхнем углу густо заселенное ядро ($g_{ДО} > 0.85$ и $g_{ИДО} > 0.75$), соответствующее одновременно (для каждой сейсмограммы обособленно) высокому качеству решения по обоим видам отпечатков. При этом в две стороны от ядра протянулись рукава с тоже довольно высокой плотностью точек. Они соответствуют записям, для которых лишь простые отпечатки сработали качественно (правый рукав) или, наоборот, лишь избыточные (верхний рукав). Благодаря такой взаимной “поддержке” в достижении цели итоговое качество решения может

превысить достоверный уровень 0.7 для 90.2% сейсмограмм, что конечно нельзя игнорировать.

Упоминание слова “может” в последнем предложении не случайно. Ведь во время автономной работы алгоритма наперед неизвестно какой из видов отпечатков для данной сейсмограммы окажется более точным. Для решения этой проблемы и построена эмпирическая экспертная система оценки доверия к результату работы обоих видов отпечатков.

Данная система представляет собою довольно большое дерево решений относительно величины штрафа, применяемого как понижающий коэффициент к первоначально стопроцентному уровню доверия. Для описания данной системы целиком нам потребовалось бы перечислить все условия назначения штрафа на каждом этапе алгоритма, где оно применимо. Это привело бы к ещё большему увеличению объема текста статьи, чего хотелось бы избежать. В приложении только приведены несколько примеров ситуаций как это работает. В будущем мы вообще не планируем сохранять подобное нагромождение решающих правил, а заменим ее на нейросеть, построенную и обученную для решения данной задачи.

Цель данной статьи заключается лишь в представлении алгоритма для получения величины граничной частоты фильтрации шума и дать меру доверия к ней. Тем не менее, уже начатые работы по разработке нейросетевой экспертной системы, показали, что для понимания принципа ее работы крайне важной была реализация этого дерева решений.

Специальный анализ результатов

Рассмотрим теперь как ДО работают в некоторых особых случаях и наметим ограничения метода.

Короткозамедленные взрывы. Среди проанализированных взрывов имелось 12 короткозамедленных, зарегистрированных на 72 сейсмограммах. Достоверность определения верхней границы шума по ним отличается от средней. Лучше показали себя простые отпечатки. Они выдали результаты с качеством более 0.7 для 84.7% записей такого сорта. В то же время ИДО смогли достичь достоверного результата только в 70.8% случаев. В целом, однако, как показала

практика, работать с записями короткозамедленных взрывов двоичным отпечаткам вполне по силам. Изначально предполагалось, что больша́я по длительности череда интерферирующих регулярных волн от множества быстро взрывааемых зарядов [Адушкин, Спивак, 2013] будет похожа по внешнему виду для ДО на исковую запись стационарного шума. Однако алгоритм нашел между ними отличия и справился хорошо.

Землетрясения. Такой тип источника анализируется на обычном уровне достоверности для ДО (74.1% достоверных ответов) и несколько лучше обычного – для ИДО (78.3%).

Влияние соотношения амплитуд сигнала и шума. Также надо рассмотреть, как влияет отношение энергий сигнала и шума на оценку характеристик последнего. Результаты определения частотности выпадения того или иного класса качества решения для разных относительных уровней шума отдельно для ДО и ИДО показаны на рис. 6. Все значения процентов на этих рисунках рассчитаны относительно полного числа проанализированных записей. Раздельное оценивание каждого диапазона шума было бы неуместным в силу неравнозначной встречаемости сейсмограмм с разными значениями $R_{СШ}$.

Можно видеть определенный диапазон отношения амплитуд сигнала и шума, в пределах которого отпечатки чаще дают качественный результат. Обычно они востребованы при высоком, но не экстремальном шуме. Когда $R_{СШ} \in [0.1, 1]$, то имеется 43.7% достоверных определений границы сигнал/шум для простых отпечатков и 45.3% для избыточных из общего числа записей с любым шумом. На втором месте по встречаемости правильного решения – следующий вариант при $R_{СШ} \in [1, 10]$. На его долю приходится 17.9 и 21.4% достоверных результатов соответственно.

Во всех остальных случаях соотношения сигнала и шума результативные решения встречаются реже. Хотя, надо отдать должное, и сейсмограмм с $R_{СШ} \notin [0.1, 10]$ суммарно не так много. К ним относится лишь пятая доля всего материала. Но если учесть этот факт, то окажется, что преобладающая часть сейсмограмм с $R_{СШ} \in [0.01, 0.1]$ также была обработана верно. Ошибочными среди подобных были лишь 0.1%

определений для простых отпечатков и 0.7% – для избыточных.

ОБСУЖДЕНИЕ

Постараемся сразу ответить на возможную критику предлагаемого метода. Одновременно покажем, что его слабые стороны в действительности могут быть и достоинствами.

Вычислительные затраты. Несомненно, некоторую трудность в практическом применении метода ДО составят его довольно высокие вычислительные затраты. По сравнению с обычным спектральным анализом, особенно алгоритмом быстрого преобразования Фурье, данный метод требует на порядок больше ресурсов. Причиной тому перевод одномерного ряда сейсмограммы в двумерное изображение вейвлетограммы. Но еще больше вычислений требуется для формирования его энергетического представления – энергограммы. Дополнительно накладывается двумерное вейвлет-преобразование Хаара, но в силу своей фрактальной природы оно вычисляется быстрее, чем одномерное. Избыточные двоичные отпечатки требуют многократного вычисления обычных. Однако, поскольку все это применяется к серьезно децимированной энергограмме, затраты времени на данном этапе весьма скромны.

Разумеется, отмеченная сложность может быть оправдана в определенных условиях, к которым правомерно отнести следующие.

Во-первых, никакая предусмотрительность не бывает излишней и, если обработчик сейсмограмм в своей повседневной работе захочет иметь дополнительный независимый инструмент анализа записей, то не стоит его ограничивать.

Во-вторых, во многих сложных случаях обычно применяемые для анализа сейсмограмм методы (спектральный анализ, наборы узкополосных фильтров, спектрограммы и тому подобное) не обеспечивают такой точности как непрерывное вейвлет-преобразование, на котором основан наш метод ДО. О причинах и интенсивности размытия энергетических проявлений частотных компонент на спектрограммах и вейвлетограммах, а также методах повышения четкости частотно-временных

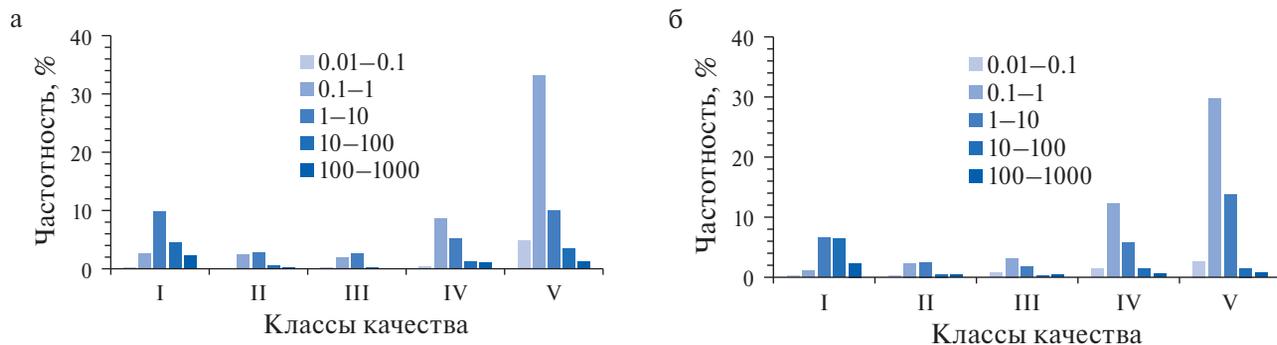


Рис. 6. Гистограммы распределения значений качества результатов решения для простых отпечатков (а) и избыточных (б) при разных соотношениях энергии полезного сигнала и низкочастотного шума (градации в условных обозначениях) с разделением на классы качества. Классы качества соответствуют рис. 3.

изображений мы рассуждали в работе [Силкин, 2023].

В-третьих, предлагаемый алгоритм изначально разрабатывался не для интерактивной работы, когда опыт обработчика позволяет увидеть все необходимое без дополнительных преобразований, а для применения в автоматическом режиме, чтобы быть встроенным в информационные системы более общего характера. Поэтому, если есть инструмент, надежно работающий в подобных условиях, заплатить за который требуется дополнительным временем вычислений, то на такую жертву можно пойти.

Потребность в предварительной оценке спектра фона. Другим моментом, способным вызвать критику, является нужда алгоритма в предварительном определении спектральных характеристик фоновой записи. Можно справедливо заметить, что часть сейсмограммы до первого вступления априори содержит тот стационарный низкочастотный шум, от которого мы пытаемся избавиться. Однако эти характеристики фона не могут быть использованы непосредственно в качестве конечного результата алгоритма вместо него самого. Ведь задача предлагаемого метода не отфильтровать шум идеально, а найти компромисс между его подавлением и сохранением сигнала в условиях частичного (а порой и сильного) перекрытия их спектров. Это особенно важно, когда интенсивность шума многократно превышает энергию сигнала.

Однако стоит отметить, что двоичные отпечатки сами по себе в принципе отлично работают

без этих данных, описывающих фон, но некоторые сведения о спектральных свойствах записи до вступления сигнала полезны для измерения степени доверия к получаемым результатам в каждом конкретном случае применения.

Трудности в обнаружении слабого шума. Из сделанных наблюдений о качестве работы метода в условиях разного соотношения сигнал/шум следует очевидный вывод о том, что чем ярче представлен искомый предмет (шум в данном случае), тем лучше он и находится. На практически свободных от низкочастотного шума сейсмограммах его поиск обычно заканчивается провалом, хотя ценность подобного поиска все равно сомнительна. Ведь такие сейсмограммы скорее всего можно в дальнейшем использовать непосредственно, не подвергая их низкочастотной фильтрации.

В то же время ДО продемонстрировали себя как очень полезный инструмент для надежного избавления записей от шума, когда он действительно представляет собою серьезную помеху.

Региональная привязка. Другим узким местом метода, которым, однако, обладают все другие методы распознавания сейсмических событий, и это было детально разобрано в работе [Силкин, 2022а], является его привязка к региону и масштабу исследований. Мы проверяли алгоритм на данных из определенной части Арктики и ограничились региональным масштабом. Ключевым ограничением метода, которое постоянно присутствовало, но не отмечалось явно, был диапазон частот, в котором вейвлет-анализ проводился: 0.2–20 Гц.

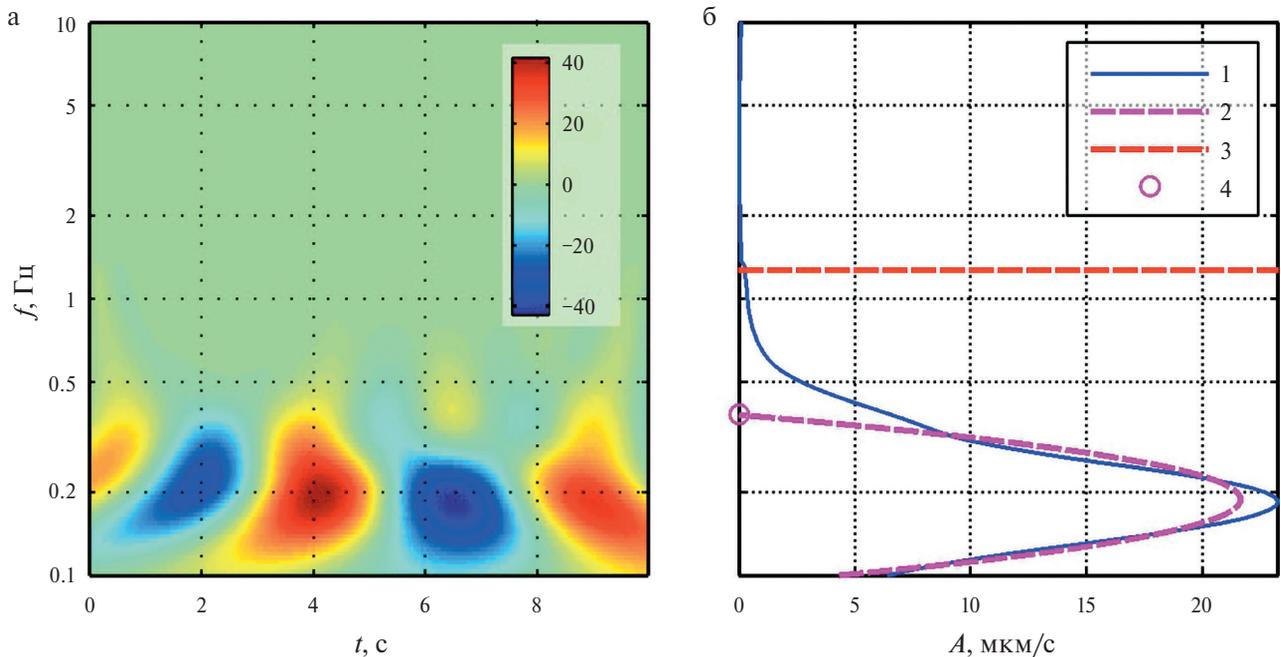


Рис. 7. Иллюстрация методики определения характерных спектральных параметров фоновой записи.

а – вейвлетограмма записи до вступления Р-волны, цветовая шкала – амплитуда в мкм/с; б – средний вейвлет-спектр шума, вертикальная ось совпадает с вейвлетограммой (1 – график спектра, 2 – аппроксимирующий полином для главного максимума спектра, 3 – отметка 1-процентного уровня, 4 – частота начала затухания преобладающей спектральной компоненты фона).

По-видимому, в иных условиях этот диапазон надо будет выбирать особо.

ВЫВОДЫ

В результате проведенного исследования на большом числе реальных примеров было убедительно показано, что технология двоичных отпечатков является полезным надежным инструментом для автоматического детектирования и оценивания свойств низкочастотного шума на сейсмических записях, содержащих сигналы от региональных взрывов и землетрясений. Мы не сравнивали предлагаемый метод по качеству результатов с современными аналогами устранения шума, рассмотренными в обзоре из работы [Силкин, 2023] и упомянутыми во введении данной статьи. Причина в том, что все алгоритмические построения у других авторов, как нам представилось, проходят проверку на единичных событиях или синтетических примерах.

Дальнейшее развитие методики с помощью замены эмпирических решающих правил на

нейросетевую экспертную систему может повысить надежность предлагаемого алгоритма.

ПРИЛОЖЕНИЕ.

АЛГОРИТМЫ ОЦЕНКИ ВЕРХНЕЙ ГРАНИЦЫ НИЗКОЧАСТОТНОГО ШУМА НА ЗАШУМЛЕННОЙ СЕЙСМОГРАММЕ С ПОМОЩЬЮ ДВОИЧНЫХ ОТПЕЧАТКОВ

Подготовительные действия и вводные обозначения

Определение спектральных характеристик фона. При измерении уровня доверия к результатам каждого вида отпечатков используется предварительное определение двух характерных параметров спектра фоновой записи: частота начала затухания преобладающей спектральной компоненты фона f_0^{Φ} и частота 1-процентного уровня спектра фона f_1^{Φ} (частота, на которой график спектра фона впервые станет в 100 раз меньше своего максимума). Оба эти параметра показаны на рис. 7.

Частота начала затухания преобладающей спектральной компоненты определяется как больший из корней квадратного уравнения, задающего полином второй степени, аппроксимирующий по

методу наименьших квадратов главный максимум спектра фона. Во время аппроксимации проверяется, чтобы коэффициент детерминации полинома не был менее 0.7. В противном случае (но это случается крайне редко) аппроксимация принимается неудачной и вместо параметра f_0^Φ используется частота спектрального максимума фона. Чтобы аппроксимация вышла лучше, частотные пределы вейвлетограммы сдвинуты вниз на октаву.

Вводные обозначения. Пусть имеется энергограмма $E(t_i, f_j)$ – функция, рассчитанная при дискретных значениях времени t_i и частоты f_j , определяемых следующим образом:

$$t_i = t_{\min} + i\Delta t, \quad (1)$$

$$f_j = \text{ant} \lg \left(\lg f_{\min} + \frac{j}{N_f} \lg \frac{f_{\max}}{f_{\min}} \right), \quad (2)$$

где $i = (1, 2, \dots, N_t) - 1, j = (1, 2, \dots, N_f) - 1$. Здесь N_f – число дискретов частоты (равно в данном случае 128), N_t – число дискретов времени, определяемое следующим образом:

$$N_t = \frac{t_{\max} - t_{\min}}{\Delta t} + 1,$$

где t_{\min} – время вступления продольной волны, $t_{\max} = t_{\min} + 10\Delta_{\text{SP}}$, Δ_{SP} – разница времен прихода S - и P -волн, Δt – дискретность по времени (в данном случае 25 мс).

В процессе вычислений используется понятие уровня доверия к качеству получаемого результата. Он получается следующим образом. В начале каждого алгоритма (ДО, СВ-С Л, СВ-С П и СВ-О) выдается кредит доверия равный 1 (или 100%). В определенных “трудных” местах (некоторые из них как пример упоминаются в описании алгоритмов) к нему применяется пенальти в виде эмпирически подобранного понижающего коэффициента. Как правило, этот коэффициент равен 0.9, но возможны многократные пенальти, поэтому кредит доверия к качеству результата алгоритма способен серьезно пасть.

К случаям понижения уровня доверия в отношении результатов ДО относятся: определение верхней частотной границы сигнала и нижней шума лишь по одному уровню отпечатков;

полное отсутствие проявлений сигнала или шума; сигнатуры шума во временном измерении не охватывают большую часть анализируемого интервала; имеет место взаимное перекрытие спектров шума и сигнала и т.п., а также разнообразные комбинации этих факторов.

Применительно к ИДО понижение уровня доверия вызывают (для каждого варианта одинаково): малое отличие по размаху главного максимума графика профиля от остальных его максимумов; значительное удаление абсцисс середины между главными экстремумами и точки сильнейшего градиента; большое расхождение результата с частотными характеристиками фона и т.п.

Вероятно, можно придумать и какие-то еще дополнительные критерии, сделав эту систему куда более громоздкой. Однако, как мы уже заметили выше, в наших планах создать независимую нейросетевую оценку, а от этой эмпирической отказаться.

Простые отпечатки

Пусть по энергограмме $E(t_i, f_j)$ рассчитаны двоичные отпечатки $\varphi(t_L, f_L)$ как двумерный композит, ассоциированный с аргументами времени и частоты (общее число дискретов композита $n_t \times n_f$). Вейвлет-разложение проводится до уровня $L, L = 1, 2, \dots, N_L$. В представленном примере $N_L = 6, n_t = n_f = 64$.

Отпечатки φ разбираются на составляющие и уровни, из которых формируются средние сечения $\Phi_f^H(L), \Phi_f^V(L), \Phi_t^V(L)$. Здесь надстрочный индекс при Φ указывает на составляющую разложения, по которой берутся значения (H – горизонтальная, V – вертикальная); подстрочный индекс отмечает какое измерение матрицы φ сохраняется (t – временное по строкам, f – частотное по столбцам) после осреднения по другому измерению. Диагональные составляющие не применяются. Число уровней, используемых с этого момента, ограничено несколькими первыми, например, четырьмя: $N_L^{\max} = 4$. Область значений средних сечений принадлежат интервалу $[0, 1]$.

По каждому сечению осуществляется построение вектора поуровневого индикатора $Y_f^H(L), Y_f^V(L), Y_t^V(L)$ в соответствии со следующей схемой:

$$Y(L) = \begin{cases} 0, \text{ если } \Phi_{x_k}(L) < C, \\ k, \text{ при котором } \Phi_{x(k)}(L) = \max \Phi(L), \\ \text{если } \overline{\Phi(L)} \geq C \text{ и } \Phi_{x(k)}(L) \geq C/2 \end{cases}$$

где $k = 1, 2, \dots, N_L^{\max} \cdot 2^{-L-1}$; $L = 1, 2, \dots, N_L^{\max}$. Здесь x – один из аргументов Φ (t или f) соответствующего сечения, $x(k)$ – соответствует t_i или f_j в зависимости от ситуации; $\overline{\Phi(L)}$ – среднее сечение по L -му уровню; C – минимальная значимость сечения отпечатка, задаваемая равной 0.75 для горизонтальной составляющей φ и 0.5 – для вертикальной, эмпирически настроенная константа. Природа этих констант вытекает из того, как формируются исходные данные: длина анализируемой записи равна $10\Delta_{\text{сп}}$ (из которых лишь примерно $1/4$ занята проявлениями регулярных волн), а шум встречается только в нижней половине диапазона частот (в логарифмическом масштабе).

Пусть $v(L)$ – вектор признака обоснованного присутствия искомой особенности записи на соответствующем сечении ДО и L -ом уровне разложения, а σ – число информативных уровней у каждого сечения, определяемые по следующим правилам:

$$v(L) = \begin{cases} 1, \text{ если } Y(L) > 0, \\ 0, \text{ если } Y(L) = 0, \end{cases}$$

$$\sigma = \sum_{L=1}^{N_L^{\max}} v(L).$$

Если $\sigma_f^v > 0$, то есть не менее одного уровня $\Phi_f^v(L)$ несут в себе достаточно четкие проявления сигнала, то определяется нижняя частотная граница его распространения $f_s^\downarrow = f_j$, где f_j берется по формуле (2). Здесь стрелка \downarrow указывает, что определяется нижняя граница. Номер частотного дискрета j (в котором L_0 подразумевает использование только уровней с $v(L) > 0$) задается следующим выражением:

$$j = \left\langle \sum_{L_0}^{2^L} \frac{2^L}{n_f} v_f^v(L) \right\rangle,$$

где угловые скобки означают округление до целого.

Если $\sigma_f^v = 0$, то частота f_s^\downarrow считается невыясненной и в дальнейшем не используется, что снижает доверие к качеству результата.

Исследуется сечение $\Phi_f^h(L)$. Если $\sigma_f^h < 2$, то есть оценка верхней частотной границы шума f_N^\uparrow выполнена недостаточно надежно или вообще не выполнена, то предпринимается попытка определить временной интервал проявления сигнала по $\Phi_f^v(L)$.

Анализируется сечение $\Phi_t^v(L)$. Если $\sigma_t^v = 0$, то сигнал на ДО не виден. Тогда эта информация не используется, а доверие к качеству результата снижается.

В случае $\sigma_t^v > 0$ определяется фактор Q доли длительности ДО, занятой в конце его шумом: $Q = 1 - t_i/n_i$, где t_i берется по формуле (1). Возможные значения у этого фактора лежат в диапазоне $[0, 1]$ и пропорционально ему сокращается минимальная значимость C для следующего этапа. Номер временного дискрета i задается следующим выражением:

$$i = \left\langle \sum_{L_0}^{2^{L+1}} \frac{2^{L+1}}{n_t} [v_t^v(L) + 1] \right\rangle.$$

Повторно вычисляется $Y_f^h(L)$ с уменьшенным C . Если $\sigma_f^h > 0$, то определяется верхняя частотная граница распространения шума $f_N^\uparrow = f_j$. Здесь стрелка \uparrow указывает, что определяется верхняя граница. Номер частотного дискрета j задается следующим выражением:

$$j = \left\langle \sum_{L_0}^{2^{L+1}} \frac{2^{L+1}}{n_f} [v_f^h(L) + 1] \right\rangle.$$

Далее анализируются полученные сведения по f_N^\uparrow и f_s^\downarrow . Если f_s^\downarrow вообще определена и $f_s^\downarrow - f_N^\uparrow < n_f/5$, то за результат алгоритма выдается среднее из этих частот. Иначе f_N^\uparrow не изменяется.

Избыточные отпечатки

Пусть для энергограммы $E(t_i, f_j)$ имеются готовые ИДО: СВ-С $\Xi(f_j, F_j)$ и СВ-О $\Psi(t_i, F_j)$. Здесь t и f имеют смысл дискретных аргументов энергограммы в соответствии с уравнениями (1) и (2), F – сканирующая частота, отмечающая

середины скользящего окна, в рамках которого по энергограмме многократно рассчитываются ДО. Для экономии времени это делается лишь для каждого второго дискрета частоты. Различие между $\Xi(f, F)$ и $\Psi(t, F)$ не только в одном из аргументов, но и в том, что СВ-С рассчитываются по горизонтальной составляющей ДО, а СВ-О – по вертикальной. Чтобы упростить построения индексы i и j далее опустим.

Пусть на $\Xi(f, F)$ взяты средние профили по узким полосам вдоль наклонных боковых сторон: СВ-С Л и СВ-С П. Полосы имеют ширину в одну десятую ширины сканирующего окна, которое в свою очередь имеет ширину $N_f/4$. Все значения поперек полосы усредняются, формируя одномерный график профиля. Влиянием частоты вейвлет-разложения f на них пренебрежем, то есть рассмотрим проекцию профилей на ось F . Профиль с левой стороны обозначим $\overleftarrow{\xi}(F)$, а с правой – $\overrightarrow{\xi}(F)$.

Аналогичный график строится и по СВ-О в виде обобщенного спектра $\psi(F)$. Он формируется из $\Psi(t, F)$ как результат вычисления квантиля с 95-процентной вероятностью при всех значениях t и фиксированном значении F . Таким образом выявляются самые высокие значения на СВ-О на данной сканирующей частоте, но свободные от случайных ураганых выбросов. В дальнейшем анализируется график $\psi(F)$, полученный зеркальным переворотом $\psi(F)$ в обратном порядке F , что позволяет все три профиля обрабатывать единым образом.

Все виды профилей – $\overleftarrow{\xi}(F)$, $\overrightarrow{\xi}(F)$ и $\tilde{\psi}(F)$ – проходят через практически одинаковый алгоритм анализа ключевых особенностей своих графиков. Исключение заключается только в том, что для контроля качества анализа первых двух используются сведения о спектральных характеристиках фона, а последнего – нет.

Пусть $y(F)$ – один из этих графиков (рис. 8). По нему отыскиваются следующие параметры:

- 1) абсцисса F_R середины (в логарифмическом смысле или среднее геометрическое) между точками минимума и максимума $y(F)$: $F_R = \sqrt{F_{\min} F_{\max}}$;
- 2) абсцисса F_D минимума производной $y'(F)$;
- 3) абсцисса F_M следующего минимума $y(F)$ после F_D или точки выхода на плато.

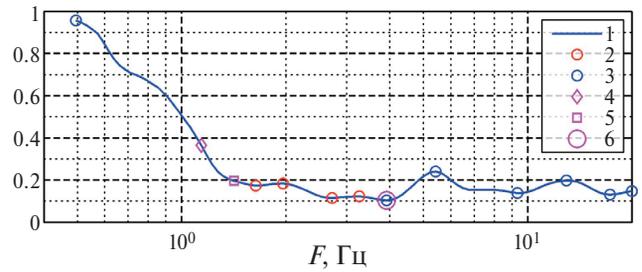


Рис. 8. Пример графика профиля для избыточных двоичных отпечатков. Вертикальная ось – условная плотность отпечатков. 1 – график профиля; 2 – экстремумы графика, которые были удалены на этапе устранения мелких особенностей; 3 – оставленные экстремумы; 4 – точка сильнейшего отрицательного градиента; 5 – точка середины между сильнейшими экстремумами; 6 – точка первого минимума (из оставленных) после сильнейшего градиента.

Для точности определения следует заметить, что y' – это не настоящая производная, а ее численный аналог – разделенная разность 1-го порядка.

Поскольку реальный график любого профиля обычно содержит некоторого рода помехи, то, прежде чем приступить к определению координат упомянутых точек, проводится работа по устранению или смягчению этих мешающих явлений. Если устранить помехи не получается, то оценивается мера доверия к тому, что есть, но с пенальти. Данная работа заключается в следующем.

Отыскиваются все экстремумы на $y(F)$. Устраняется пара смежных экстремумов, если расстояние между ними в пространстве номеров частотных дискретов j , задающих значение частоты в соответствии с формулой (2), менее 3.

Устраняются “мелкие” экстремумы на $y(F)$, выступание которых над (под) соседними относительно максимального размаха значений $y(F)$ попадает в 25-процентный квартиль.

Исследуется степень изрезанности графика $y(F)$. Если график имеет много крупных осцилляций и отношение размаха наибольшей из них к среднему размаху недостаточно велико (менее 3), то это снижает доверие к результатам анализа.

Если найденные значения F_R , F_D и F_M не принадлежат интервалу частот $[f_0^\Phi, f_1^\Phi]$, то это снижает доверие к результатам анализа $\overleftarrow{\xi}(F)$ и $\overrightarrow{\xi}(F)$.

Осуществляется выборка наиболее информативного из показателей F_R и F_D или формируется их обобщение. Мерой для этого служит минимально значимый отрезок дискретов оси F , определяемый через $\Delta j = \langle (j_{\max} - j_{\min}) / 10 \rangle$.

Выявляется наиболее достоверное положение сильнейшего отрицательного градиента F_G . Причем, чтобы сохранить привязку к логарифмическому масштабу измерения производятся не в пространстве частот, а в пространстве номеров дискретов. Если $|j_R - j_D| < \Delta j$, то $F_G = \min(F_R, F_D)$. Если $\Delta j < |j_R - j_D| < 2\Delta j$, то $F_G = \sqrt{F_R F_D}$, но доверие к F_G в данном случае снижается. При выходе за рубеж $2\Delta j$ доверие снижается еще сильнее.

Аналогично в зависимости от соотношения F_G и F_M определяется и оценивается более обоснованное значение искомой верхней частотной границы шума для данного профиля $u(F)$, однако роль F_M в этом случае является вспомогательной.

В итоге по трем профилям $\bar{\xi}(F)$, $\vec{\xi}(F)$ и $\tilde{\psi}(F)$ получается три оценки границы шума: $F_{\bar{\xi}}$, $F_{\vec{\xi}}$ и $F_{\tilde{\psi}}$ соответственно. Между ними устраивается “конкурс” на величину полученной при их отыскании оценки доверия к качеству. Если найден явный фаворит, то побеждает наилучший. Если найден явный аутсайдер, то исключается худший. В случае близких результатов за итог выдается средняя среди них величина. Таким образом, создается ансамбль решающих деревьев, качество работы которого выходит больше, чем любого из них в отдельности.

БЛАГОДАРНОСТИ

Автор выражает благодарность кандидату физ.-мат. наук, ведущему научному сотруднику Кольского филиала ФИЦ ЕГС РАН В.Э. Асмингу за любезную помощь в подборе данных.

КОНФЛИКТ ИНТЕРЕСОВ

Автор данной работы заявляет, что у него нет конфликта интересов.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при поддержке Минобрнауки России (в рамках государственного задания № 075-01471-22) и с использованием данных, полученных

на уникальной научной установке “Сейсмоинфразвуковой комплекс мониторинга арктической криолитозоны и комплекс непрерывного сейсмического мониторинга Российской Федерации, сопредельных территорий и мира”.

СПИСОК ЛИТЕРАТУРЫ

- Адушкин В.В., Спивак А.А.* Влияние трассы на затухание сейсмического сигнала от короткозамедленных карьерных взрывов // *Динамические процессы в геосферах*. 2013. № 4. С. 118–126.
- Асминг В.Э., Баранов С.В.* Использование спектрально- и частотно-временных вариаций сейсмических сигналов для различия землетрясений и взрывов в Евро-арктическом регионе // *Современные методы обработки и интерпретации сейсмологических данных*. Материалы Международной сейсмологической школы. Обнинск: ГС РАН, 2006. С. 23–27.
- Асминг В.Э., Кременецкая Е.О., Виноградов Ю.А. Евтюгина З.А.* Использование критериев идентификации взрывов и землетрясений для уточнения оценки сейсмической опасности региона // *Вестник МГТУ. Труды Мурманского государственного технического университета*. 2010. Т. 13. № 4–2. С. 998–1007. EDN NQZHFN
- Асминг В.Э., Федоров А.В.* Возможности применения автоматического детектора-локатора сейсмических событий по одиночной станции для детальных сейсмологических наблюдений // *Сейсмические приборы*. 2014. Т. 50. № 3. С. 19–29. EDN SLRRSN
- Баврина А.П., Борисов И.Б.* Современные правила применения корреляционного анализа // *Медицинский альманах*. 2021. № 3(68). С. 70–79. EDN TPSSIX
- Попов С.Е., Замараев Р.Ю.* Веб-сервис классификации сейсмических событий на базе системы распределенных вычислений Apache Spark // *Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов (SDM-2019)* // *Сборник трудов Всероссийской конференции с международным участием, Бердск, 26–30 августа 2019 года*. Бердск: ИВТ СО РАН, 2019. С. 440–450. EDN ESWFQV
- Силкин К.Ю.* Использование вейвлет-анализа как дополнительного критерия при идентификации природы сейсмического события // *Структура, вещественный состав, свойства, современная геодинамика и сейсмичность платформенных территорий и сопредельных регионов: материалы XXII Всероссийской с международным участием научно-практической Шукинской конференции (Воронеж, 22–25 сентября 2020 г.)* / Под ред. Л.И. Надежка, Т.Б. Силкиной. Воронеж: ВГУ, 2020. С. 328–334. EDN VHKKWM

- Силкин К.Ю.* Новые эвристики на основе вейвлет-анализа записи одиночного датчика для распознавания землетрясений и взрывов // Сейсмические приборы. 2022а. Т. 58. № 3. С. 5–24.
<https://doi.org/10.21455/si2022.3-1> EDN HFBDZH
- Силкин К.Ю.* Двоичные отпечатки результатов вейвлет-анализа записи сейсмического события как инструмент формирования его компактного образа для целей нейросетевого распознавания // Российский сейсмологический журнал. 2022б. Т. 4. № 4. С. 42–55.
<https://doi.org/10.35540/2686-7907.2022.4.03> EDN EMMMDС
- Силкин К.Ю.* Двоичные отпечатки сейсмической записи как средство для автоматической фильтрации низкочастотного шума // Сейсмические приборы. 2023. Т. 59. № 2. С. 26–45.
<https://doi.org/10.21455/si2023.2-3> EDN BYNTAQ
- Тур А.И., Кокоулин А.Н., Южаков А.А.* Возможности применения алгоритмов нечеткого поиска и нейронных сетей в технологии Fingerprint // Нейрокомпьютеры: разработка, применение. 2017. № 6. С. 45–48. EDN ZICARH
- Фортуна-Сервантес Х.М., Рамирес-Торрес М.Т., Мартинес-Карранса Х., Мургуа-Ибарра Х.С., Ме-хиа-Карлос М.* Обнаружение объектов в аэронавигации с использованием вейвлет-преобразования и сверточных нейронных сетей: первый подход // Труды Института системного программирования РАН. 2021. Т. 33. № 2. С. 149–162.
[https://doi.org/10.15514/ISPRAS-2020-33\(2\)-9](https://doi.org/10.15514/ISPRAS-2020-33(2)-9) EDN MYCXZQ
- Abma R., Claerbout J.* Lateral prediction for noise attenuation by t - x and f - x techniques // Geophysics. 1995. V. 60. № 6. P. 1887–1896.
<https://doi.org/10.1190/1.1443920>
- Baluja Sh., Covell M.* Waveprint: Efficient wavelet-based audio fingerprinting // Pattern Recognition. 2008. V. 41. № 11. P. 3467–3480.
<https://doi.org/10.1016/j.patcog.2008.05.006>
- Bonar D., Sacchi M.* Denoising seismic data using the nonlocal means algorithm // Geophysics. 2012. V. 77. № 1. P. A5–A8.
<https://doi.org/10.1190/geo2011-0235.1>
- Bergen K.J., Beroza G.C.* Earthquake fingerprints: Extracting waveform features for similarity-based earthquake detection // Pure and Applied Geophysics. 2019. V. 176. P. 1037–1059.
<https://doi.org/10.1007/s00024-018-1995-6>
- Cano P., Battle E., Kalker T., Haitsma J.* A Review of Audio Fingerprinting // Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology. 2005. V. 41. № 3. P. 271–284.
- Chen Z., Stewart R.* A multi-window algorithm for real-time automatic detection and picking of P-phases of microseismic events // CREWES Research Report. 2006. V. 18. P. 1–9.
- Daubechies I., Lu J., Wu H.-T.* Synchrosqueezed Wavelet Transforms: An Empirical Mode Decomposition-like Tool // Appl. Comput. Harmon. Analysis. 2011. V. 30. № 2. P. 243–261.
<https://doi.org/10.1016/j.acha.2010.08.002>
- Ferrari G.M.* Un saber “sudamericano”. La dactiloscopia en el Congreso Científico Latinoamericano, 1901–1909 // Historia Crítica. 2016. № 60. P. 81–101.
<https://doi.org/10.7440/histcrit60.2016.05>
- Fragoulis D., Rousopoulos G., Panagopoulos T., Alexiou C., Papaodysseus C.* On the automated recognition of seriously distorted musical recordings // IEEE Transactions on Signal Processing. 2001. V. 49. № 4. P. 898–908.
- Iatsenko D., McClintock P.V.E., Stefanovska A.* Extraction of instantaneous frequencies from ridges in time-frequency representations of signals // Signal Processing. 2016. V. 125. P. 290–303.
<https://doi.org/10.1016/j.sigpro.2016.01.024>
- Jensen A., la Cour-Harbo A.* Ripples in mathematics: The Discrete wavelet transform. Berlin, Germany: Springer Science & Business Media Publ., 2011. 246 p.
<https://doi.org/10.1007/978-3-642-56702-5>
- Haitsma J., Kalker T.* A Highly robust audio fingerprinting system: Proceedings of ISMIR 2002, 3rd International Conference on Music Information Retrieval. Paris, France, 2002. P. 107–115.
- Hussain I., Fadhil R., Abdulsatar N.M.* An image compression using block truncation coding with two types of wavelet transform (HAAR & DWT) // Test Engineering and Management. 2020. № 83. P. 16707–16713.
- Herrera R., Tary J.B., Baan M., Eaton D.* Body Wave Separation in the Time-Frequency Domain // IEEE Geoscience and Remote Sensing Letters. 2014. № 12. P. 364–368. <https://doi.org/10.1109/LGRS.2014.2342033>
- Mallat S.* A wavelet tour of signal processing: the sparse way. Cambridge, UK: Academic Press, 2008. 109 p.
- Mallat S.G.* A Theory of multiresolution signal decomposition: The wavelet representation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989. № 11. P. 674–693.
<https://doi.org/10.1109/34.192463>
- Mousavi S.M.* Microseismic monitoring and denoising / Diss. ... Doctor. of Philosophy. Memphis, 2017. 188 p.
<https://doi.org/10.13140/RG.2.2.16996.27529>
- Mousavi S.M., Langston C.A.* Hybrid Seismic Denoising Using Higher-Order Statistics and Improved Wavelet Block Thresholding // Bull. Seismol. Soc. Am. 2016. V. 106. P. 1380–1393.
<https://doi.org/10.1785/0120150345>

- Mousavi S.M., Langston Ch., Horton S.* Automatic denoising and detection of microseismic events using the synchrosqueezing // SEG Technical Program Expanded Abstracts. 2016a. P. 825–829.
<https://doi.org/10.1190/segam2016-13262052.1>
- Mousavi S.M., Langston Ch., Horton S.* Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform // Geophysics. 2016b. V. 81. № 4. P. V341–V355.
<https://doi.org/10.1190/geo2015-0598.1>
- Ranjan R., Kumar P., Naik K., Singh V.K.* The HAAR-the JPEG based image compression technique using singular values decomposition // 2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET). Patna, India. 2022. P. 1–6.
<https://doi.org/10.1109/ICEFEET51821.2022.9848400>
- Shi Y., Zhang D., Ji H., Dai R.* Application of Synchrosqueezed Wavelet Transform in Microseismic Monitoring of Mines // IOP Conference Series: Earth and Environmental Science. 2019. № 384. Art. ID 012075.
<https://doi.org/10.1088/1755-1315/384/1/012075>
- Tary J.B., Herrera R.H., van der Baan M.* Analysis of time-varying signals using continuous wavelet and synchrosqueezed transforms // Philos. Trans. R. Soc. A: Math., Phys. Eng. Sci. 2018. V. 376. № 2126. Art. ID 20170254.
<https://doi.org/10.1098/rsta.2017.0254>
- Wang D.-w., Li Y.-j., Zhang K., Xu H.-m.* An Adaptive Time-Frequency Filtering Method for Nonstationary Signals Based on the Generalized S-Transform // Optoelectron. Lett. 2010. V. 6. № 2. P. 133–136.
<https://doi.org/10.1007/s11801-010-9250-0>
- Wang A.L.-Ch.* An industrial strength audio search algorithm // Proceedings of ISMIR 2003, 4th International Conference on Music Information Retrieval. Baltimore, Maryland, USA, 2003. P. 7–13.
- Yoon C.E., O'Reilly O., Bergen K.J., Beroza G.C.* Earthquake detection through computationally efficient similarity search // Science Advances. 2015. V. 1. № 11. P. 1–13.
<https://doi.org/10.1126/sciadv.1501057>
- Yu G., Yu M., Xu Ch.* Synchroextracting Transform // IEEE Transactions on Industrial Electronics. 2017. V. 64. № 10. P. 8042–8054. <https://doi.org/10.1109/TIE.2017.2696503>
- Zeng Z., Lu T., Han P., Zhang D., Yang X.-H., Shi Y., Chang Y., Zhang J., Dai R., Ji H.* Microseismic data denoising in the synchrosqueezed domain by integrating the wavelet coefficient thresholding and pixel connectivity // Geophysical Journal International. 2022. V. 232. № 2. P. 1113–1128. <https://doi.org/10.1093/gji/ggac378>
- Zhang D., Zeng Z., Shi Y., Chang Y., Dai R., Ji H., Han P.* An Effective Denoising Method Based on Cumulative Distribution Function Thresholding and its Application in the Microseismic Signal of a Metal Mine with High Sampling Rate (6 kHz) // Frontiers in Earth Science, 2022. V. 10. № 7. Art. ID 933284.
<https://doi.org/10.3389/feart.2022.933284>

Automatic Denoising of Seismograms using Fingerprints: Algorithms, Properties, Limitations

K. Yu. Silkin*

Geophysical Survey of the RAS, Lenina ave., 189, Obninsk, 249035 Russia

**e-mail: const.silkin@ya.ru*

This article describes the new technique for automatically preparing a noisy seismic record for further analysis using expert information systems. The technique is based on seismogram fingerprints, which, due to their concise but informative pattern, allow the development of a reliable algorithm for finding important noise characteristics. The search for the optimal cutoff frequency for a high-pass filter is especially important under conditions of partial overlap of the signal and noise spectra at a high intensity of the latter. It is precisely this difficult case that this study aims to address. The article analyzes the developed methodology on the example of several hundred registrations of regional earthquakes and explosions. The analysis showed that reliable results can be achieved in more than 90% of cases. In addition to all the problems and limitations of the method, which are an extension of its capabilities, are mentioned. Appendix to the article contains detailed description of the algorithm underlying the method.

Keywords: earthquake, explosion, seismogram, wavelet transform, fingerprints, low-frequency noise, filtering