

УДК 550.334

ПРИМЕНЕНИЕ МЕТОДОВ СЕГМЕНТАЦИИ К ОБРАБОТКЕ ГЕОФИЗИЧЕСКИХ ДАННЫХ

А.Б. Тристанов¹, В.В. Геппенер², П.П. Фирстов¹

¹ *Институт вулканологии и сейсмологии ДВО РАН, г. Петропавловск-Камчатский, e-mail: abt@kcs.iks.ru*

² *Санкт-Петербургский государственный электротехнический университет, г. Санкт-Петербург*

В настоящее время, согласно общепринятой модели геофизической среды, Земля рассматривается как открытая, неравновесная, нелинейная, диссипативная, динамическая система взаимодействующих между собой отдельных частей, имеющих различную физико-химическую природу [6]. В эволюции такой системы отдельностей важное место занимают неустойчивости или отдельные акты неустойчивости (подвижки литосферных плит, геологические катастрофы, извержения вулканов, землетрясения и т. д.).

В отличие от устойчивых равновесных систем сложные неравновесные динамические системы оказываются чувствительными к малейшим изменениям внутренних параметров, флуктуации которых могут быть обусловлены крайне незначительными изменениями характера внешних воздействий. Отсюда следует принципиальная возможность существования триггерных эффектов в геофизических полях [3]. Возможно, именно поэтому прогноз времени, места и силы землетрясений представляет до сих пор проблему, которая требует разработки новых методов обработки геофизических данных.

Один из удачных подходов к обработке геофизических данных развивается в работе [5], где геофизическая информация рассматривается как результат измерений низкочастотных фоновых процессов (НФП). Под НФП понимаются постоянные и носящие случайный характер временные вариации геофизических полей с характерным периодом от нескольких минут до нескольких месяцев и даже лет. НФП имеют место как в сейсмоактивных, так и в асейсмических регионах и, очевидно, являются носителями информации о процессах, происходящих в верхних слоях земной коры, в том числе и связанные с подготовкой очагов будущих землетрясений. Выявление предвестников землетрясений, в рамках такого подхода, основано на построении алгоритмов анализа комплексной геофизической информации, основанных на применении современных методов многомерного анализа временных рядов и поиске предвестников, как увеличение коллективной составляющей в поведении геофизических полей в области подготовки катастрофы.

В данной статье намечен еще один подход к обработке геофизической информации данных, с целью поиска информации о процессах, происходящих в литосфере, по данным, получаемым одномерными наблюдениями.

Традиционный подход к исследованию временных рядов, такой как анализ модели авторегрессивного скользящего среднего (АРСС), имеет ограничение по стационарности исследуемого процесса, т. к. предполагает, что статистические характеристики сигнала не меняются с течением времени. Реальные сигналы редко отвечают этому требованию на

большом временном интервале. Поэтому применение модели АРСС может быть обосновано лишь на небольших временных интервалах, где сигнал условно принимается стационарным.

При исследовании многолетних рядов геофизической информации с целью поиска предвестников землетрясений, сигнал искажается влиянием самых разнообразных шумов (вариации атмосферного давления, температуры, влажности, индустриальная деятельность человека и т. д.). Поэтому к геофизическим сигналам возможно применение современной технологии Data Mining - комплекс процедур и алгоритмов обработки и анализа сигнала, целью которого является обнаружение скрытых структур и закономерностей и, как следствие, получение новых знаний о процессе [1].

Исследование сигналов методом Data Mining является процессом, состоящим из нескольких этапов.

Этап 1. Сбор и хранение исходных данных.

Этот этап включает два синхронных процесса:

1. Регистрацию сигналов и подготовку их цифровых записей.
2. Регистрацию событий и их описание.

С точки зрения проектирования, первый этап заключается в создании базы данных, которая является моделью некоторой предметной области, состоящей из связанных между собой данных об объектах, их свойствах и характеристиках [1] .

Этап 2. Применение методов Data Mining.

Данный этап является ключевым и включает в себя:

2.1. Использование методов выделения особенностей сигналов на основании применения методов сегментации.

2.2. Унификация выделенных сегментов. В связи с тем, что выделенные в первом пункте сегменты являются разнородными как по длине, так и по структуре, следовательно, необходимо выбрать единые для всех сегментов признаки (представление сегментов).

2.3. Классификация событий и кластеризация сегментов. После получения подготовленного материала необходимо его систематизировать с выделением классов, которые, в применении к данным геофизических наблюдений с целью прогноза землетрясений, соответствуют трем основным группам:

- сервисные события, связанные с регламентированным воздействием на систему регистрации (пропуски данных);
- внешние антропогенные воздействия на исследуемую систему;
- воздействие на систему, связанное с изменением напряженно деформированного состояния геологической среды.

Поскольку для сегментов не существует априори заданных классов, то к ним требуется применить методику кластеризации. В задаче кластеризации важно, чтобы каждый экземпляр данных выражался четким числовым значением, с этой целью в процесс исследования необходимо ввести этап унификации сегментов.

2.4. Сопоставление событий кластерам сегментов. Сопоставление происходит в хронологическом порядке, основываясь на гипотезе, что в один момент времени происходит одно событие. Данный пункт является, по сути, генератором новых знаний.

Рассмотрим в общем виде подход к выделению особенностей (задача сегментации) в сигнале.

Согласно работе [5], задача сегментации сводится к построению детектора, который по данной реализации сигнала $f(t)$ устанавливает истинность одной из гипотез H_0 и H_1 . Гипотезы определяются следующим образом:

H_0 : исследуемый сигнал $f(t)$ соответствует модели M_θ ;

H_1 : существует момент времени τ , такой, что исследуемый сигнал $f(t)$ соответствует модели M_θ при $t < \tau$ и модели $M_{\theta'}$ при $t \geq \tau$.

В условиях поставленной задачи, авторами предлагается применить два метода сегментации:

1. Адаптация авторегрессионной модели (АМ).

2. Исследование динамики показателя Гельдера.

Идея первого метода заключается в последовательном выполнении следующих шагов алгоритма [5], которые схематически приведены на рис. 1:

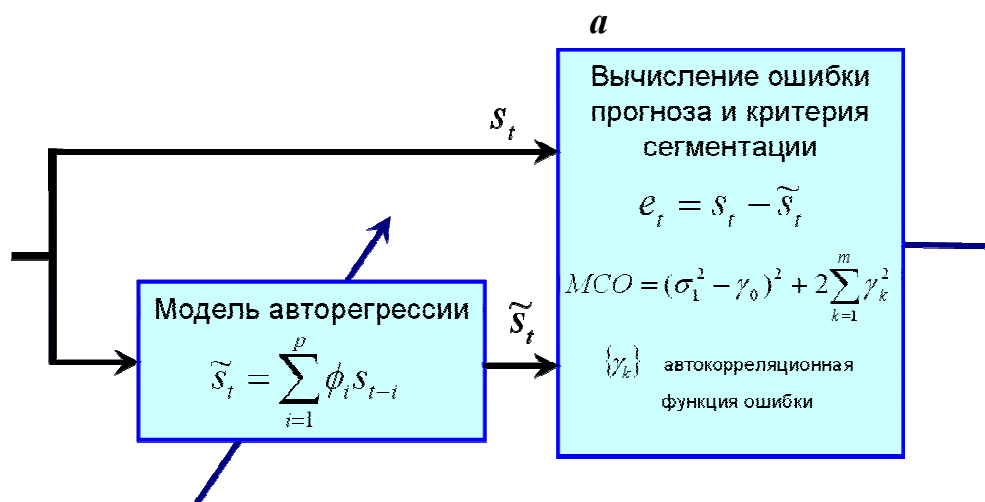


Рис. 1. Структурная схема метода сегментации на основе авторегрессивной модели.

- идентифицирование опорной модели, с этой целью по имеющимся наблюдениям выбирается «типичный» стационарный участок сигнала и строится модель M_θ ;
- после получения текущего значения наблюдаемого сигнала находится его прогнозное значение с использованием модели M_θ ;
- определяется степень различия реального и модельного значений сигналов (ошибка прогноза и критерий сегментации).

В случае, если различие реального и модельного сигналов значительное, то устанавливается метка перехода к новому сегменту и обновляется опорная модель. В качестве опорной модели используется модель авторегрессии, а в качестве критерия сегментации - мера спектральной ошибки (MSO).

В основе второго метода заложена идея регулярности сигнала [2,6,8]. Математической характеристикой регулярности является показатель Гельдера, который является критерием сегментации сигнала.

Считается, что функция $f(t)$ удовлетворяет условию Гельдера порядка $\alpha: 0 \leq \alpha \leq 1$ в точке t_0 , если существует постоянная A , такая, что для всех x из окрестности t_0 справедливо неравенство

$$|f(t) - f(t_0)| \leq A|t - t_0|^\alpha.$$

Одним из эффективных методов оценки локальной регулярности сигнала является вейвлет - преобразование. Если вейвлет имеет компактный носитель, то значения вейвлет-коэффициентов зависят от значения исследуемой функции в окрестности некоторой точки и пропорциональны масштабу. На малых масштабах они характеризуют локальное поведение функции [2,6,8].

Если возможно оценить показатель Гельдера в каждой точке, то существует возможность выделить участки его снижения, а, следовательно, сегментировать сигнал. Для реализации алгоритма в реальном масштабе времени может быть использован быстрый алгоритм вычисления непрерывного вейвлет-преобразования (НВП), предложенный в работе [9], который основан на свойствах преобразования Меллина (ПМ).

В качестве примера рассмотрим сегментацию многолетнего временного ряда объемной активности радона (ОА Rn) в почвенном воздухе с 1 ноября 1999 г. по 1 января 2005 г. (частота дискретизации 0.5 сутки^{-1}), полученного в опорном пункте Паратунка в

зоне полного влагонасыщения [9]. На верхней кривой рис. 2 приведена кривая динамики OA_{Rn} и ее сегментация, а на нижней кривой показана мера спектральной ошибки.

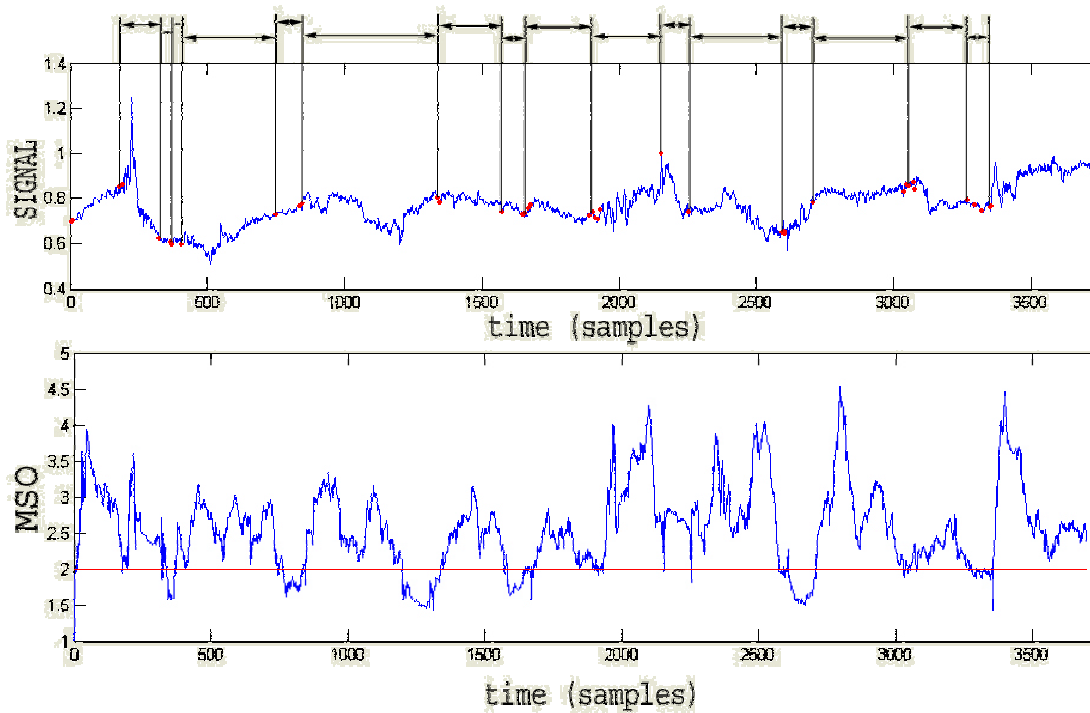


Рис. 2. Пример сегментации сигнала объемной активности радона в подпочвенном воздухе (верхняя кривая) и мера спектральной ошибки (нижняя кривая).

На рис. 3 приведен пример сегментации сигнала сейсмического шума с помощью двух алгоритмов. Точками со стрелками на рисунке обозначены границы сегмента, обнаруженные алгоритмом адаптации AP - модели, вертикальными пунктирными линиями – границы, обнаруженные алгоритмом, основанным на оценке показателя Гельдера. Видно, что при возникновении «быстрых» изменений в сигнале происходит излишняя сегментация по методу адаптации AP-модели, а второй алгоритм работает более корректно, но он не выявил других сегментов.

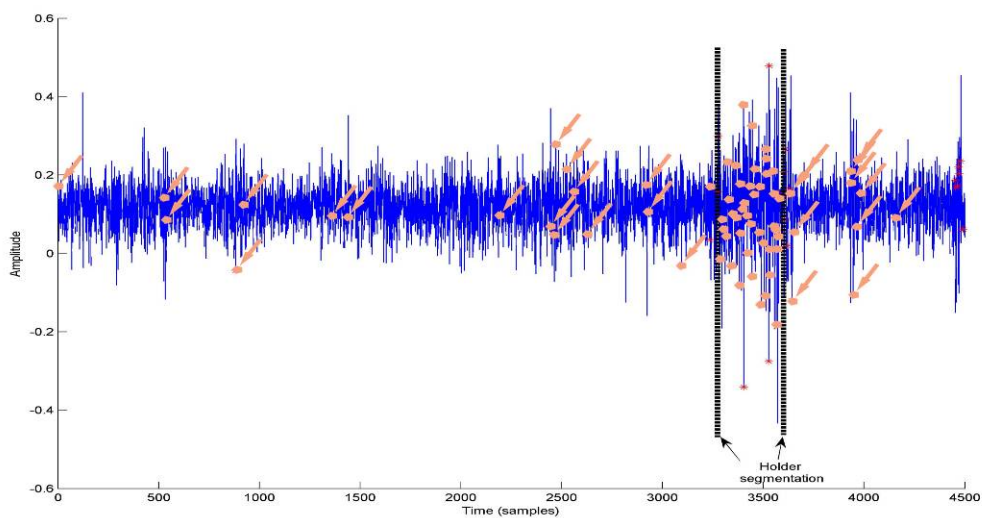


Рис. 3. Пример сегментации сигнала длительностью 36 с с помощью обоих методов AP и ДПГ.

Из проведенных экспериментов видно, что большая информативность этапа сегментации достигается при совместном использовании двух алгоритмов, один из которых

направлен на выявление быстрых изменений, а второй - ориентирован на медленноменяющиеся процессы.

Следует отметить, что при решении задач анализа данных нельзя ограничиваться применением какого-то одного метода. Применение методов Data Mining позволяет создавать эффективные автоматизированные системы обработки геофизических сигналов.

Работа выполнена при финансовой поддержке гранта ДВО РАН 05-III-Г-08-140.

Список литературы

1. *Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И.* Методы и модели анализа данных: OLAP и Data Mining. СПб.: БХВ-Петербург, 2004. 336 с.
2. *Бассвиль М., Вилски А., Банвенист А. и др.* Обнаружение изменений свойств сигналов и динамических систем. М.: Мир, 1989. 278 с.
3. *Децеровский А.В., Лукк А.А., Сидорин А.Я.* Природа временных вариаций геофизических полей // Геофизика на рубеже веков. М.: Миннаука РФ, 1999. С.194-215.
4. *Добеши И.* Десять лекций по вейвлетам. Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. 464 с.
5. *Любушин А.А.(мл.)* Геофизический мониторинг: шумы, сигналы, предвестники // Проблемы геофизики XX века. М.: Наука, 2003. С. 70-94.
6. *Садовский М.А.* О естественной кусковатости горных пород // Докл. АН СССР. 1979. Т. 247. № 4. С. 829-831.
7. *Тристанов А.Б.* Обнаружение изменений в сигнале методом вейвлет-анализа // Труды Второй Всероссийской научной конференции «Проектирование инженерных и научных приложений в среде MATLAB». М.: ИПУ РАН, 2004. С. 1798-1821.
8. *Тристанов А.Б., Руленко О.П., Фирстов П.П.* Некоторые особенности сейсмического шума в скважине НИС-1 в июле-августе 2003 года // Материалы конференции, посвященной дню вулканолога. Петропавловск-Камчатский, 2004. С. 82-89.
9. *Фирстов П.П., Рудаков В.П.* Результаты регистрации подпочвенного радона в 1997-2000 гг. на Петропавловск-Камчатском геодинамическом полигоне // Вулканология и сейсмология. 2003. № 1. С. 26-41.
10. *Mallat S.A.* Wavelet tour of signal processing. New York. Academic Press. 1998. 670 p.
11. *Zhang Tong et al.* A fast algorithm of continuous wavelet transform based on Mellin transform. Chinese. J. of Signal Processing, 1996. 12(4). P. 128-131