

Модели, алгоритмы и открытый облачный сервис для диагностики состояния природных сейсмических генераторов

Р.Ю. Замаараев, С.Е. Попов

Аннотация. В работе представлены математические модели, алгоритмы и их программная реализация для диагностики состояния природных сейсмических генераторов в ряду высокоэнергетических событий. Модели сейсмологических сигналов на основе информационной энтропии трактуются как строгие характеристики «волновода» из очага к сеймостанции. Наблюдаемые качественные отличия моделей связываются с изменением вида состояния генератора. Представлена программная реализация алгоритмов на основе технологии облачных вычислений Google App Engine в интеграции с веб-сервисами IRIS Data Management Center.

Ключевые слова: сейсмический генератор, виды состояния, сейсмологические сигналы, модели, диагностика, облачный сервис.

Введение

Прогностические задачи сейсмологии являются неизменно важными и актуальными. Общими для них были и остаются проблемы идентификации параметров динамических и статистических моделей очага, которые создает априорная нестационарность сейсмических процессов. Она проявляется как в самом факте появления высокоэнергетического события, так и в структуре сейсмологических сигналов, к этому событию привязанных.

Классификационные методы анализа сейсмической активности, опирающиеся на теорию распознавания образов, не требуют анализа сигналов. Например, в алгоритме КОЗ [1] прогноз реализуется на устойчивых последовательностях сейсмических событий определенных энергий в пространственно-временных ячейках.

Известно также, что возмущение из гипоцентра распространяется в существенно неоднородном литологическом пространстве в виде

продольных и поперечных волн. Регистрируются они измерительной системой в виде трех ортогональных разложений пространственных колебаний.

Однако особенности конкретного события могут содержаться как в малозаметных аномалиях, так и локальных трендах сигналов, разложенных по трем случайным осям. Это соображение используется в работе [2], где состояние генератора анализируется в математических конструкциях на основе вейвлет-преобразования сейсмологических сигналов. Изменение соотношения количеств и амплитуды регулярных и редких элементов структуры сигнала на интервале наблюдения интерпретируется как угасание или активизация генератора.

Настоящая работа опирается на диагностический подход к наблюдениям и экспериментам с акцентом на разработку устойчивых критериев для сравнения видов состояния генерирующих динамических систем. Основная идея работы заключается в выборе модели сейсми-

ческих наблюдений, обеспечивающей использование формы сейсмологического сигнала как классификационного признака.

С одной стороны, это позволило бы решить задачу классификации сейсмических событий на основе яркого, структурированного признака. С другой стороны, сохранило бы привязку наблюдаемых отличий сигналов к временной области, а, значит, возможность получения устойчивых критериев сравнения и интерпретаций событий из теории колебаний.

1. Постановка задачи

В работе предлагается специфический подход к землетрясению как к случайному Вумп-тесту динамических систем, передающих возмущение от гипоцентра к сейсмостанциям. Тогда по записям одного события с нескольких станций можно сравнить условия прохождения сейсмических волн в различных трактах, а по записям с одной станции нескольких событий, относящихся к одному сейсмически активному району, сравнить виды состояния генератора и, по возможности, оценить изменения.

В обеих постановках возникает необходимость разработки алгоритмов для обобщения информации о структуре сигналов по трем ортогональным направлениям. Для устойчивого сравнения сигналов между станциями и событиями алгоритмы должны опираться на некоторую математическую модель наблюдений, содержащую определение видов состояния генерирующей системы.

Необходимость доступа к актуальным и достоверным источникам сейсмической информации и манипуляций большими объемами географически привязанных данных привели к идее создания открытого для научной общест-венности сервиса на одной из развитых облачных платформ.

2. Базовые математические модели

В основу предлагаемых алгоритмов анализа состояния положена группа моделей *информационного* типа [3], связанных с определением вида состояния и структурой системы уникальных объектов-опытов. Теория, обосновываю-

щая данные модели, излагалась ранее также в работе [4].

Базу для построения модели задает энтропийное отображения данных наблюдений и экспериментов:

$$\mathbf{E} = -\mathbf{q} \ln(\mathbf{q}), \quad (1)$$

где $\mathbf{q} = \mathbf{x} / S$ – вектор долей, *эквивалентных вероятности выбора* i -го элемента из вектора данных $\mathbf{x} = \{x_1, \dots, x_m\}$ о значениях некоторого показателя X ; $S = \sum_i x_i$ – сумма элементов вектора \mathbf{x} .

Энтропийное отображение (1) приводит вектор данных к виду модели распределения полной энтропии системы по элементам–опытам в выбранном показателе X и, согласно теории информации [5], взаимно однозначно определяет вид состояния i -го элемента. Элементы с равными и/или упорядоченными по уровню вкладами в полную энтропию определяют информационную структуру системы в виде набора подсистем с особыми свойствами.

Требуемым свойством отображения является аддитивность вкладов элементов–опытов как внутри одного показателя, так и между несколькими показателями. Это свойство вытекает из представления о наборе отображений $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$, построенных в показателях X_1, \dots, X_n , как о модели полной энтропии системы опытов, представленной независимыми сообщениями (каналами) $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$.

Для решения диагностических задач по набору показателей X_1, \dots, X_n введена *обобщенная энтропийная модель*

$$\mathbf{E}_{\Sigma, V} = \left\{ \sum_{j \in V} -q_{i,j} \ln(q_{i,j}) \right\}_{i=1, \dots, m}, \quad (2)$$

где V – обобщаемое подмножество показателей, т.е. для решаемых задач имеем алгоритм построения комплексной характеристики из сигналов по требуемым направлениям, событиям и станциям.

Аддитивность элементов в наборе энтропийных моделей открывает принципиально другой важный тип моделей *информативности*. Их построение заключается в процедуре извлечения

вектора значений частных энтропий показателей для избранного элемента системы, т.е. при $i = const$ с последующей стандартизацией:

$$\mathbf{E}_i = \frac{\{E_{i,j=1}, \dots, E_{i,j=n}\} - M[\{E_{i,j=1}, \dots, E_{i,j=n}\}]}{\sigma[\{E_{i,j=1}, \dots, E_{i,j=n}\}]} \quad (3)$$

Интерпретация такой модели очевидна: каждый элемент вектора \mathbf{E}_i есть энтропия j -го показателя при i -ом элементе. Таким образом, получаем модель интенсивности приращения по i -го объекта по каждому показателю и, следовательно, можем оценить и сравнить *информативность* каждого показателя для каждого объекта.

При анализе систем, заданных сигналами на выходе, следует учесть их отличие от выборок: i -ый отсчет показателя по времени не является автономным объектом и не характеризует какое-либо состояние системы. Характеристикой состояния является количество информации, произведенное системой к моменту времени t от начала отсчета [5]. Исходя из физических принципов регистрации сейсмологических сигналов выбрана замена исходной формы на вектор модулей размахов колебаний

$$s_i = |x_{i+1} - x_i| \quad (4)$$

между i -ым и $i+1$ -ым отсчетами. Такое преобразование сигнала приводит к определению аддитивной величины – длины траектории движения материальной точки (центра масс измерительной системы).

Тогда количество информации, приходящееся на i -ый отсчет определяется как кумулятивная функция энтропий размахов

$$\mathbf{H}_C = \left\{ -\sum_{k=1}^i \frac{s_k}{\sum_{i=1}^{m-1} S_i} \ln \left(\frac{s_k}{\sum_{i=1}^{m-1} S_i} \right) \right\}_{i=1, \dots, m-1} \quad (5)$$

Алгоритмы анализа состояния динамических систем на основе полученных моделей не имеют ограничений по размерности и стационарности сигналов, поскольку в базовом отображении никаких условий к этим свойствам данных наблюдений и экспериментов не выдвигается.

3. Программная реализация

Основными требованиями при разработке прикладного программного обеспечения (ПО) были установлены: открытость и масштабируемость; эргономика работы с внешними источниками; актуальность и достоверность источников сейсмической информации.

Рассматривались спецификации трех основных на сегодняшний день облачных технологий: Microsoft Windows Azure, Google App Engine (GAE) и Amazon Web Services (AWS). Существует большое количество ресурсов, например [6, 7], приводящих сравнение достоинств и недостатков этих продуктов. Можно утверждать, что по степени реализации и проработанности сервисов, API, языков программирования и т.п. все они находятся примерно на одном уровне.

Основной упор был сделан на минимизацию вложений в размещение веб-проекта и его полное функционирование в облаке. С технологией Google App Engine (GAE) таких затрат удалось избежать вовсе. Политика Google предоставляет полностью работоспособное бесплатное пользовательское пространство для размещения приложений в облаке, где по основным характеристикам (процессорное время, исходящий/входящий трафик, дисковое пространство и т.п.) предоставляются достаточные квоты [8] для реализации сервиса.

Для сравнения, технология Microsoft Azure предоставляет трехмесячный пробный доступ, который накладывает существенные ограничения по сравнению с аналогичным GAE. По истечении пробного периода необходимо выбрать план оплаты ресурсов [9], что не позволяет откладывать и развивать приложение.

В пользу выбора IRIS DMC (International Research Institutions for Seismology Data Management Center) как источника сейсмологической информации послужил факт наличия разветвленной сети сейсмостанций по всему миру, которые в режиме реального времени поставляют сигналы в базу данных. IRIS предоставляет свободный доступ к API и веб-сервисам [10], позволяющим получать информацию об объектах базы данных (сейсмическая станция, сейсмическое событие, сигнал и т.п.) по средством HTTP-запросов.

Результаты возвращаются в формате XML (QuakeML) для событий в виде целочисленного массива таймсерий в формате ASCII. Это свойство полностью удовлетворяет используемой модели наблюдений и экспериментов, поэтому для реализации обозначенных выше требований были выбраны платформа облачных вычислений Google App Engine (GAE) и IRIS DMC.

В итоге был сконструирован облачный сервис "Seismatica" с применением инструментария Google Web Toolkit (GWT), технологии GWT-RPC [11], Google Map APIs (GMap), Google Chart API (GChart), Google Users API, классы пакета java.net, а также, с использованием службы IRIS Web Services (IRIS WS) [10]. Сервис доступен по адресу <http://seismatica.appspot.com>.

Пакеты GWT, GMap, GChart применялись для построения графической оболочки сервиса (GUI). GUI представлен веб-формой, предлагающей пользователю возможность сформировать граничные условия для математической модели в виде параметров запросов к веб-службам IRIS WS и списка таймсерий, формирующих расчетную матрицу значений, возможность отображения сейсмообъектов на картах Google Maps, визуализации результатов посредством компонентов GChart.

Технология GWT-RPC позволила организовать взаимодействие клиент-серверной части посредством асинхронных вызовов удаленных методов сервлетов, содержащих программную логику расчета математической модели с передачей матрицы значений таймсерий, а также выполнения http-запросов к IRIS WS без перезагрузки GUI. Данный подход снял проблему сохранения состояния переменных как на стороне сервера, так и на стороне клиента, что позволило инкапсулировать функционал и сделать модель поведения сервиса приближенной к стационарному приложению.

Все изменения интерфейса, расчеты модели, и визуализация результатов происходят в едином контексте, позволяя инициировать различные варианты начальных условий, получая пул результатов в одной сессии, что дает возможность для сравнительного анализа. GUI организован в виде панели вкладок, каждая из которых представляет некий вариант анализа при

различных начальных условиях в математической модели.

Классы пакета java.net совместно с технологией GWT-RPC применялись для взаимодействия серверных компонентов с сервисами IRIS WS. Методы программного интерфейса Google Users использовались для внедрения в сервис механизмов авторизации пользователей на базе действующих учетных записей Google [15].

Последовательность действий пользователя от формирования граничных условий до конечного результата анализа при работе с облачным сервисом "Seismatica" можно разделить на 4 этапа:

1. выбор параметров сейсмических событий (временной интервал, минимальную/максимальную магнитуду, ее тип, глубину и т.п. [12]) для получения списка опорных сейсмических событий;

2. выбор опорного события и параметров сейсмических событий и станций – радиусов расположения от точки привязки опорного события и диапазон магнитуд [12, 13] для получения списка ближайших событий;

3. выбор сейсмической станции, группы ближайших событий, параметров таймсерий (временное окно – количество секунд до и после зафиксированного времени события, маска названия канала сейсмостанции, код расположения [14]) для получения списка таймсерий;

4. отображения исходных данных и построение моделей сигналов.

Графический интерфейс пользователя, отвечающий за каждый из этапов, представлен на Рис. 1, а взаимодействие компонентов облачного сервиса показано на Рис. 2.

На параметры 1-3 этапов накладываются ограничения, вызванные двумя факторами: ограничения служб IRIS WS [10] и косвенные ограничения на исполнение программного кода в среде Google App Engine при бесплатном доступе.

Например, при расчете конечной матрицы значений таймсерий в программном коде объекта *AnalyzeAgentService* принудительно сделаны 10-тисекундные задержки вызова процедуры *CalculateAnalysisScheme* к службе IRIS WS. Они введены для обхода ограничений на время выполнения одного вызова GWT-RPC, которое не должно превышать 30 секунд (квота бесплатного доступа). А время ответа службы IRIS

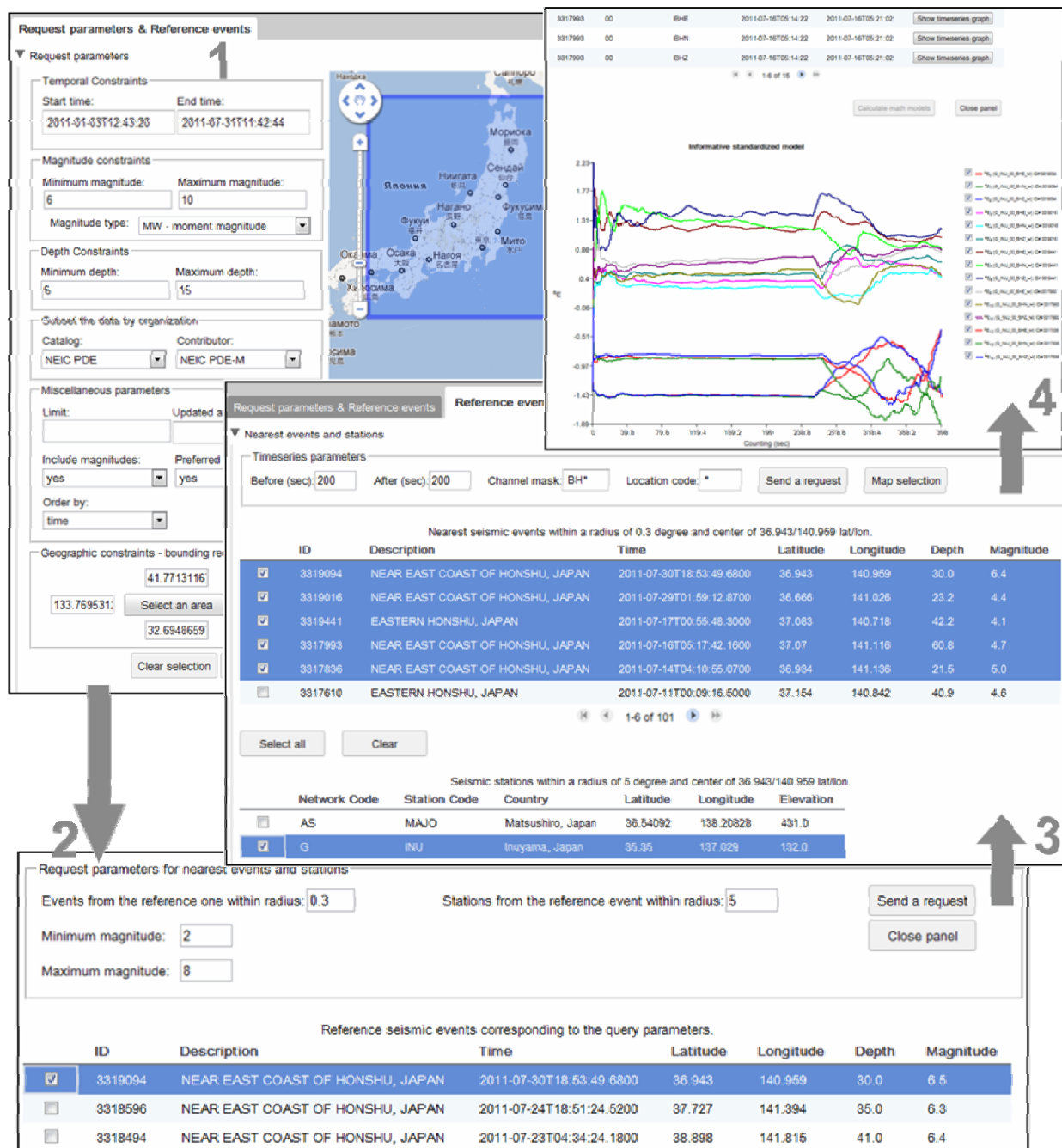


Рис. 1. Графический интерфейс пользователя (фрагменты) на каждом из этапов работы с облачным сервисом "Seismatica"

WS [16] по всем запросам в сумме превосходит этот порог. По той же причине введено ограничение на размер временного окна сигнала.

Сервис обладает системой подсказок и предупреждений в ответ на некорректные действия

пользователя. Все термины, обозначения, единицы измерения применяемые как в GUI, так и при составлении системы подсказок взяты с документации по службам IRIS WS [12-14, 16].

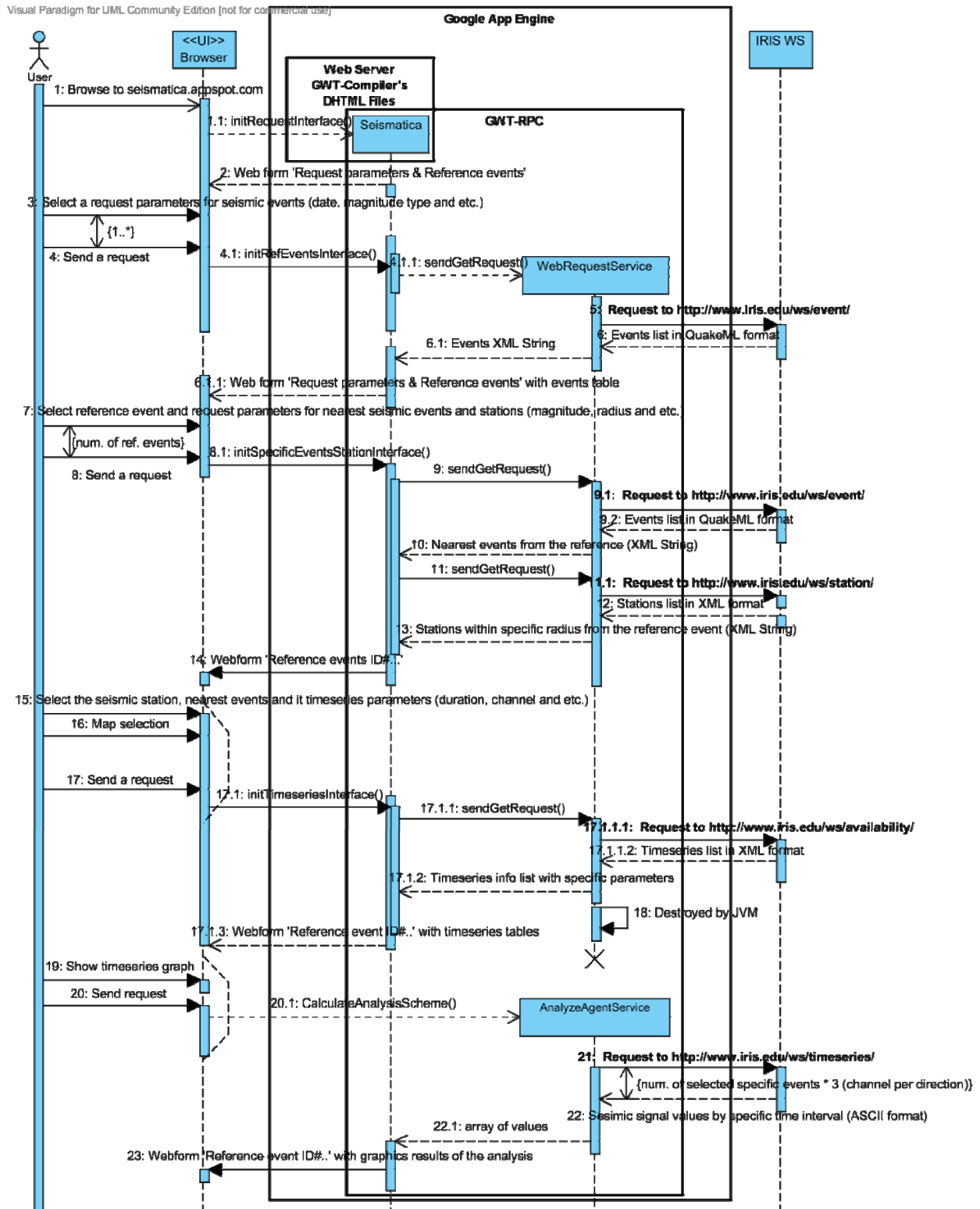


Рис. 2. UML-диаграмма последовательности взаимодействий программных компонентов облачного сервиса "Seismatica"

4. Пример использования алгоритмов и сервиса

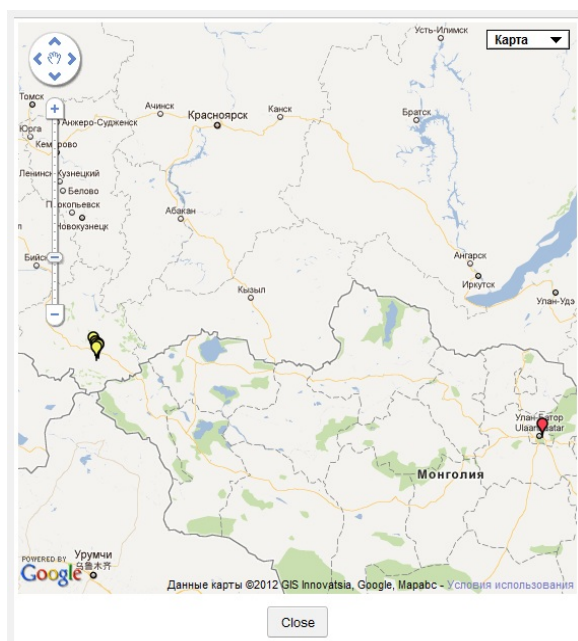
Для примера в качестве опорного события было выбрано известное Алтай–Саянское землетрясение от 27.09.2003г. Его отличительной особенностью является отсутствие значимой предыстории, т.е. зарегистрированных в IRIS DMS и сопоставимых по энергетическому классу землетрясений в том же районе в 5-летней ретроспективе. Поэтому диагностика, в этом случае, заключалась в сравнении основного возмущения с последующими крупными автошоками.

В Табл. 1 собраны сведения по заметным сейсмическим событиям, произошедшим после опорного (выделено жирным шрифтом). При их отборе использовались формальные критерии: установленная магнитуда свыше или равная 5 и отсутствие других заметных автошоков во временном окне.

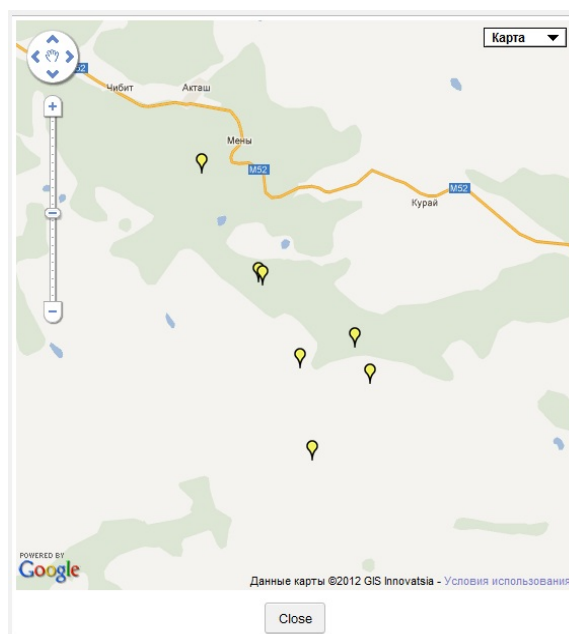
Точкой наблюдения была выбрана сейсмическая станция в Улан-Баторе, Монголия (код IU-ULN, Ulaanbaatar, Mongolia, ш. 47,8651, д. 107,0532). На Рис.3 показано географическое положение станции и отобранных событий.

Табл. 1. Сведения о выбранных для анализа сейсмических событиях

№	ID IRIS DMS	Дата и время события	Долгота	Широта	Магнитуда
0	1646230	2003-11-11 22:42:31.5300	50.1050	87.8364	5.0
1	1667932	2003-10-17 05:30:20.6800	50.1579	87.7116	5.0
2	1654214	2003-10-13 05:26:38.1500	50.2522	87.6298	5.3
3	1640547	2003-10-09 16:06:03.7700	50.0524	87.8312	5.1
4	1594940	2003-10-01 01:03:26.3100	50.1599	87.7071	6.3
5	1647584	2003-09-27 18:52:46.0600	50.0871	87.7632	6.6
6	1646446	2003-09-27 15:31:22.0700	50.0738	87.8570	5.0
7	1645208	2003-09-27 11:33:25.1300	50.0089	87.7788	7.4



а)



б)

Рис. 3. География сейсмических событий и станции наблюдения

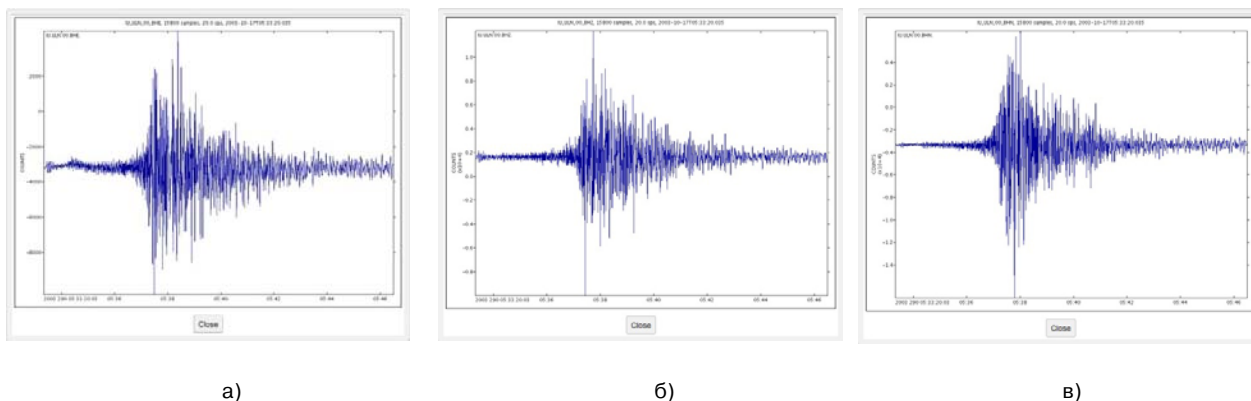


Рис. 4. Пример сейсмологических сигналов: станция IU-ULN, событие №1, а) восточная, б) северная и в) вертикальная составляющие

По схеме работы сервиса (Рис. 2) для каждого выбранного события были извлечены запросом к IRIS DMS сейсмозаписи по трем ортогональным осям измерительных систем станции. Пример сигналов с характерными образами ударной волны, прошедшей через множество передающих систем со своими резонансами, модуляторами и демпферами, показан на Рис. 4.

Каждый сигнал претерпел последовательно отображения (4), (1) и (5), что привело к частным кумулятивным моделям полной энтропии сигналов. Из алгоритмов (1) и (5) следует, что эти модели приближены к s -образной форме и отличаются соотношением длины верхней и нижней ветвей и положением точки перегиба.

Затем по алгоритму (2) построены обобщенные модели событий из трех частных моделей сигналов по направлениям измерения. Как и частные, обобщенные модели имеют s -образную форму и специфические особенности, отражающие вид состояния генератора в каждом из событий. Однако масштаб этих отличий невелик по отношению к размаху модели, что затрудняет анализ. Построения графиков обобщенных моделей событий в облачном сервисе предусмотрены, но здесь не приводятся.

Для сравнения видов состояния генератора в серии из 8 событий по алгоритму (3) построен упорядоченный набор моделей информативности. На соответствующих графиках (Рис. 5) видны этапы различной длительности, отражающие особенности накопления полной энтропии каждого события *относительно других в серии*.

При сопоставлении во временной области сигналов и моделей, для последних можно вы-

делить несколько характерных зон (Рис. 5): зона краевых эффектов – до маркера 1; зона переднего фронта ударной волны – между маркерами 1 и 2; зона заднего фронта ударной волны – между маркерами 2 и 3; зона остаточных колебаний (шумов) в литологических структурах в окрестности сейсмической станции – после маркера 3.

Обращаясь к поведению моделей информативности, прежде всего, выделим очевидную группировку событий №№4, 5 и 7. Поскольку событие №7 является опорным, можем заключить, что после основного землетрясения от 2003-09-27 в 11:33:25, в очаге еще дважды 2003-09-27 в 18:52:46 и 2003-10-01 в 01:03:26 происходили аналогичные по структуре процессы, но с меньшей энергией.

После такого заключения, остальные события могут быть отнесены к сильным автошокам со своими особенностями. Так события №№0, 3 и 6 демонстрируют самостоятельные поведения и должны быть классифицированы как особые процессы.

События №№1 и 2, произошедшие с разницей в 3 дня, имеют стабильное и почти зеркальное к №№7 и 4 поведение. Следовательно, модели полной энтропии этих событий связаны квазилинейно и могут быть отнесены к надклассу процессов в очаге.

Сделанные выше заключения позволяют выбрать группы относительно однородных событий и корректно провести статистическую обработку наблюдений и идентификацию параметров сейсмического генератора.

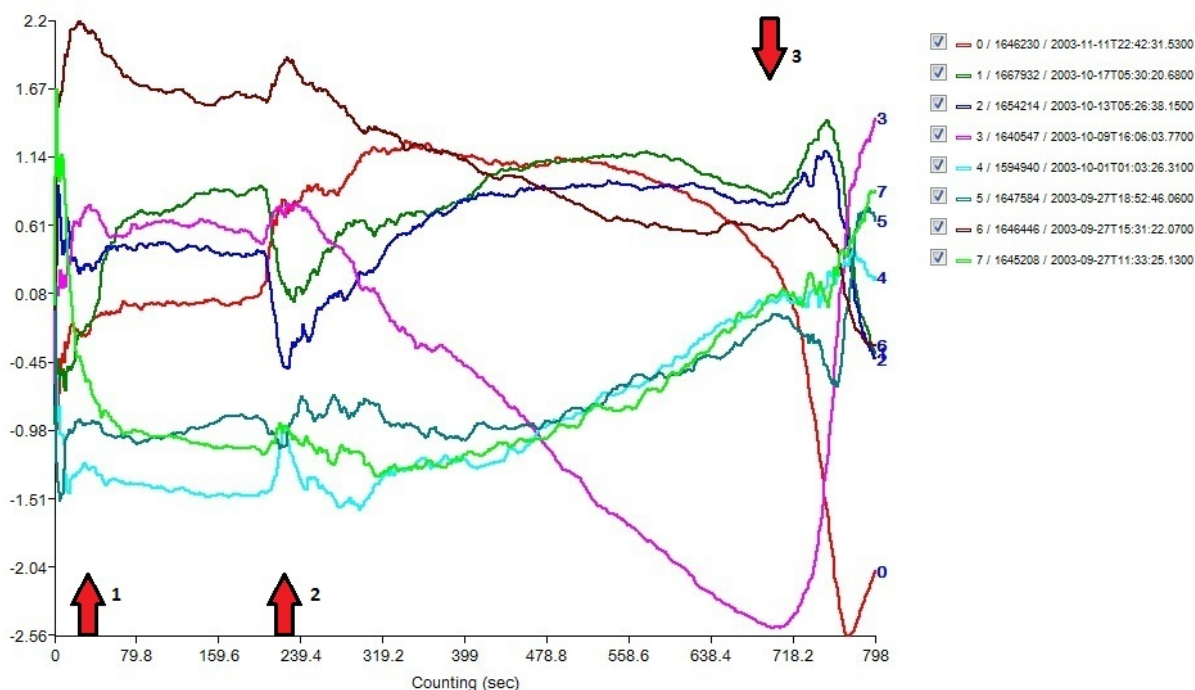


Рис. 5. Модели информативности серии сейсмических событий, привязанных к Алтай-Саянскому землетрясению от 27.09.2003 г.

Глубокие предметные интерпретации моделей информативности могут быть реализованы при включении в серию модельных или эмпирических сигналов на выходе динамических систем с известными свойствами. Однако даже на этой стадии развития сервиса может быть решена прогностическая задача для очагов с перемежающимся (циклическим) характером изменения вида состояния.

Заключение

Представленные модели, алгоритмы и облачный сервис рассматриваются как новый инструмент для классификации и диагностики сейсмических явлений. В настоящее время ведется разработка феноменологических моделей информативности сигналов и создание на их основе решающих правил и шаблонов диагностических заключений. Естественно, предполагается программная реализация правил и шаблонов в сервисе, а, также, оптимизации интерфейса и расчетного ядра сервиса.

Литература

1. Завьялов А.Д. Среднесрочный прогноз землетрясений: основы, методика, реализация. – М.: Наука. 2006. – 254 с.
2. Богданов В.В., Геппнер В.В., Мандрикова О.В. Моделирование нестационарных временных рядов геофизических параметров со сложной структурой. – С.-Петербург: ЛЭТИ. 2006. – 108 с.
3. Логов А.Б., Замараев Р.Ю., Логов А.А. Анализ состояния уникальных объектов. – М.: Машиностроение, 2010. – 336 с.; ил.
4. Замараев Р.Ю., Попов С.Е. Программный комплекс для интеллектуального анализа данных в горном производстве // Информационные технологии и вычислительные системы. 2010. № 3. С. 86-101.
5. Колмогоров А.Н. Теория информации и теория алгоритмов. – М.: Наука. 1987. – 304 с.
6. Google App Engine (GAE) против Amazon Web Services (AWS) [Электронный ресурс] // HABRAHABR.RU: Новостной сайт, коллективные блоги. URL: http://habrahabr.ru/blogs/cloud_computing/117146/ (дата обращения: 03.02.2012).
7. Google App Engine (GAE) vs Amazon Elastic Computing (EC2) vs Microsoft Azure [Электронный ресурс] // Danny Tuppeny/DanTup. URL: <http://blog.dantup.com/2010/10/google-app-engine-gae->

- vs-amazon-elastic-computing-ec2-vs-microsoft-azure (дата обращения: 03.02.2012).
8. Google App Engine. Quotas [Электронный ресурс] // GOOGLE.COM. URL: <http://code.google.com/intl/ru-RU/appengine/docs/quotas.html> (дата обращения: 02.02.2012)
 9. Windows Azure. Pricing Overview [Электронный ресурс] // MICROSOFT Corp. URL: <https://www.windowsazure.com/ru-ru/pricing/details/#compute> (дата обращения: 02.02.2012)
 10. IRIS DMC Web Services [Электронный ресурс] // IRIS - Incorporated Research Institutions for Seismology. URL: <http://www.iris.edu/ws/> (дата обращения: 02.02.2012)
 11. Making Remote Procedure Calls [Электронный ресурс] // GOOGLE.COM. URL: <http://code.google.com/intl/ru-RU/webtoolkit/doc/latest/tutorial/RPC.html> (дата обращения: 02.02.2012)
 12. IRIS Event Webservice [Электронный ресурс] // IRIS - Incorporated Research Institutions for Seismology. URL: <http://www.iris.edu/ws/event/> (дата обращения: 02.02.2012)
 13. IRIS Station Webservice [Электронный ресурс] // IRIS - Incorporated Research Institutions for Seismology. URL: <http://www.iris.edu/ws/station/> (дата обращения: 02.02.2012)
 14. IRIS Availability Webservice [Электронный ресурс] // IRIS - Incorporated Research Institutions for Seismology. URL: <http://www.iris.edu/ws/availability/> (дата обращения: 02.02.2012)
 15. Users API (интерфейс авторизации пользователей) [Электронный ресурс] // GOOGLE.COM. URL: <http://googleappengine.ru/docs/users/> (дата обращения: 02.02.2012)
 16. IRIS Timeseries Webservice [Электронный ресурс] // IRIS - Incorporated Research Institutions for Seismology. URL: <http://www.iris.edu/ws/timeseries/> (дата обращения: 02.02.2012).

Замараев Роман Юрьевич. Старший научный сотрудник ИВТ СО РАН. Окончил Кузбасский государственный технический университет в 1996 году. Кандидат технических наук. Автор 34 печатных работы, в том числе 4 монографий. Область научных интересов: математическое моделирование, анализ данных. E-mail: zrukem@gmail.com

Попов Семен Евгеньевич. Старший научный сотрудник ИВТ СО РАН. В 2002 г. Окончил Кемеровский государственный университет в 2002 году. Кандидат технических наук. Автор 18 печатных работ. Область научных интересов: математическое моделирование, распределенные вычислительные системы и облачные технологии. E-mail: ogidog@mail.ru