

Программы AggW и AggWR

оценки вейвлет-агрегированного сигнала многомерного временного ряда.

А.А. Любушин, доктор физ.-мат. наук
Институт физики Земли РАН им. О.Ю.Шмидта,
123995, Москва, Большая Грузинская, 10; факс: +007-499-2556040;
e-mail: lyubushin@yandex.ru
<http://AlexeyLyubushin.narod.ru/Index.htm>

Программы предназначены для выделения общих вариаций, имеющих нестационарный характер («всплесков»), присутствующих сразу во всех (число рядов – не менее 3-х) анализируемых временных рядах и, в то же время, для подавления индивидуальных шумов, характерных только для одного, того или иного, временного ряда. Качественно агрегированный сигнал можно описать как такой скалярный сигнал, который в максимальной степени аккумулирует в себе наиболее общие вариации, присутствующие сразу во всех анализируемых процессах. В то же время метод агрегирования подавляет те составляющие, которые характерны только для одного процесса и имеющих, как правило, характер локальных помех, обусловленных спецификой места проведения измерений, техногенными причинами или ошибками измерений.

Отличие между программами **AggW** и **AggWR** заключается в различном способе вычисления так называемых агрегированных вейвлет-коэффициентов, обратное преобразование от которых позволяет получить скалярный агрегированный сигнал.

Описание программ.

Обрабатываемые временные ряды должны представлять собой результаты синхронных наблюдений в виде числовых текстовых файлов, имеющих структуру «одна запись – один отсчет» (или «длинная колонка чисел»). Каждому временному ряду должен соответствовать свой файл. Если исходные данные представляют собой таблицы, то временные метки и прочая служебная информация не должны находиться в первой колонке таблицы. Небольшие пропуски данных должны быть восполнены какими-то «правдоподобными» значениями. Перед запуском программы в рабочей директории должен быть создан вспомогательный файл со стандартным именем "list", содержащий имена анализируемых файлов, перечисленных «в столбик» (см. программу **MakeList**).

Программы считывают из файла "list" имена файлов, содержащих анализируемые временные ряды, открывает их, последовательно считывает из них временные отсчеты и закрывает. Если файл "list" отсутствует, то программа останавливается с соответствующим сообщением. Если исходные файлы содержат разное число отсчетов, то обработка будет производиться по выборке длины, равной минимальной длине временных рядов.

Далее пользователю необходимо ответить на следующие вопросы:

- 1) С какого по номеру отсчета следует начать обрабатывать сигналы. Ввод этого параметра позволяет пропустить от начала определенное число отсчетов и начать обрабатывать лишь тот фрагмент временных рядов, который представляет интерес. Если пропускать начальные фрагменты нет необходимости, то на этот запрос следует ответить «1».
- 2) Следует ввести длину временного фрагмента в числе отсчетов (не менее 64), которую желательно обработать. Реальная длина фрагмента может не совпасть с

желательной, поскольку тот или иной ряд может иметь меньшую длину, чем сумма отсчетов, пропущенных от начала и желательная длина обрабатываемого фрагмента. В этом случае берется минимальная среди анализируемых сигналов длина, оставшаяся после пропуска начальных отсчетов. Если необходимо проанализировать всю допустимую длину выборок, но минимальная длина неизвестна, то можно ввести произвольно большое число, например «1000000000» и тогда программа сама найдет реальную длину обрабатываемого фрагмента.

- 3) Надо ли переходить к рядам в приращениях («дифференцировать ряды»). Если ряды носят существенно низкочастотный характер, то такой переход необходим. В противном случае, если доминируют низкие частоты, всегда будет наблюдаться высокая когерентность на самых старших уровнях детальности вейвлет-разложения.
- 4) Следует ввести длину окна адаптации в числе отсчетов.
- 5) Необходимо ввести идентификатор используемого вейвлета, который для обычных вейвлетов Добеши равен их порядку, то есть числам 2, 4, 6, 8, 10, 12, 14, 16, 18, 20. Для того, чтобы ввести симлет, надо ввести число, равное порядку симлета плюс 100, то есть возможные ответы для симлета следующие: 108, 110, 112, 114, 116, 118, 120.
- 6) Следует ввести значение порога представительности L_{\min} – положительное целое число, смысл которого заключается в том, что оценки вычисляются лишь для тех уровней детальности вейвлет-разложения, для которых число вейвлет-коэффициентов на уровне при длине выборки, определяемой длиной временного ряда, больше или равно этому пороговому значению.
- 7) Следует ввести значение T_{ini} начальной временной метки в выводном файле.
- 8) Следует определить шаг по времени T_{step} во временных метках, введя значения T_{scale} и dT , после чего шаг по времени вычисляется по формуле $T_{step} = T_{scale} / dT$. Это удобно, например, в ситуации, когда временные ряды определены с шагом 1 сутки от начала 1900-го года, а временные метки в выводном файле необходимо задать в годах. Тогда можно ввести $T_{ini} = 1900$, $T_{scale} = 1$, $dT = 365.25$, где параметр dT равен среднему числу суток в году с учетом високосных лет.

Ниже приведена копия экрана диалога с программой **AggW**. Аналогичная копия экрана для диалога с программой **AggWR** не приводятся, поскольку для обеих программ диалог идентичен.

Выводной файл программы **AggW** имеет имя “AggW.dat” и представляет собой числовую таблицу из 2-х колонок: 1-я колонка суть значения агрегированного сигнала, а 2-я колонка – это временные метки отсчетов.

Выводной файл программы **AggWR** имеет имя “AggWR.dat” и представляет собой числовую таблицу из 2-х колонок: 1-я колонка суть значения агрегированного сигнала, а 2-я колонка – это временные метки отсчетов.

```

c:\D:\Users\Lbshn\Wrk\AggW.exe
This is the program for computing wavelet-aggregated signal
of multidimensional time series ( 3 <= dim ).

Alexey Lyubushin, IPE RAS, Moscow
http://lyubushin.hotbox.ru/Index.htm, e-mail: lyubushin@yandex.ru
-----
Number of series to process =          10
What is the number of sample to start from = ?
1
What is the desirable number of samples to be processed ?
64 <= Leng
10000000
Real number of samples to be processed <=          2922
Nfour=          4096
From the serie Z01          2922 samples were read.
From the serie Z02          2922 samples were read.
From the serie Z03          2922 samples were read.
From the serie Z04          2922 samples were read.
From the serie Z05          2922 samples were read.
From the serie Z06          2922 samples were read.
From the serie Z07          2922 samples were read.
From the serie Z08          2922 samples were read.
From the serie Z09          2922 samples were read.
From the serie Z10          2922 samples were read.
Real number of samples to be processed =          2922
Do you want to come to incremental (diff.) series (0/1) ?
1
Length of adaptation window (number of samples) = ?
700
Wavelet order= ? (2,4,6,8,10,12,14,16,18,20) - for usual Daubechies wavelets,
(108,110,112,114,116,118,120) - for symlets of order = m-100
(Sym_02 is Haar, Sym_04 and Sym_06 are Daub_04 and Daub_06).
2
What is the value of Lmin - minimal number
of wavelet coefficients on the last MRA-level,
which is sufficient for estimating covariational matrices ?
(          3 <= *          )
10
What is the initial value for time marks in the output files Tini = ?
0
Introduce the step for time marks in the output files: Tstep = Tscale/dT.
What are the values of Tscale and dT ?
1 1

```

Описание метода.

Вейвлет-агрегированный сигнал строится аналогично Фурье-агрегированному (программа **AggF**), но имеет специфику, вытекающую из финитности базисных функций. Отличие состоит в том, что для вейвлет-агрегации опять вводится скользящее временное окно некоторой длины N отсчетов. Раньше, при конструировании мер когерентности в программах **WBRCM** и **WBRCM_mult**, для каждого положения скользящего временного окна, после выполнения предварительных операций, фрагменты временного ряда внутри окна подвергались независимому («короткому») дискретному вейвлет-преобразованию. Для вычисления агрегированного сигнала ряды подвергаются дискретному преобразованию лишь один раз («длинное» преобразование), а смысл временного окна состоит в том, что вычисление агрегированных вейвлет-коэффициентов или робастных агрегированных вейвлет-коэффициентов (см. описание программ **WBRCM** и **WBRCM_mult**) производится лишь для тех канонических коэффициентов, для которых носители исходных коэффициентов лежат внутри текущего временного окна τ длиной N отсчетов.

Каждый вейвлет-коэффициент $c_j^{(\beta)}$ уровня детальности β вычисляется, согласно формуле, по вариациям исходного временного ряда в окрестности отсчета $\tau_j^{(\beta)}$, где $\tau_j^{(\beta)} = j \cdot 2^\beta$, $j = 1, \dots, M_\beta$, где M_β - число вейвлет-коэффициентов на уровне детальности β для всей длительности временных рядов (а не для длины окна N). При использовании

ортогональных вейвлетов Добеши порядка $2p$, радиус окрестности точки $\tau_j^{(\beta)}$, поведение сигнала внутри которой влияет на значение коэффициента $c_j^{(\beta)}$, равен $p \cdot 2^{(\beta-1)}$ отсчетов.

При вычислении агрегированного сигнала, для каждого положения τ правого конца и для каждого уровня детальности β рассматриваются лишь те вейвлет-коэффициенты $c_j^{(\beta)}$, для которых выполняется условие:

$$\tau_j^{(\beta)} + p \cdot 2^{(\beta-1)} \leq \tau, \quad (\tau - N + 1) \leq \tau_j^{(\beta)} - p \cdot 2^{(\beta-1)} \quad (1)$$

где p – число обнуляемых моментов для используемой базисной функции Добеши. Обозначим через $L_\beta^{(A)}(N)$ число коэффициентов, удовлетворяющих условию (1). Верхний индекс «(A)» подчеркивает отличие числа коэффициентов в данном случае от введенного ранее числа $L_\beta(N)$ при вычислении оконных мер когерентности в программах **WBRCM** и **WBRCM_mult**. Как и ранее, рассматриваются лишь те уровни детальности, для которых выполняется условие:

$$L_\beta^{(A)}(N) \geq L_{\min} \quad (2)$$

Заметим, что условие (2), в силу (1), теперь зависит от типа базиса.

Для данной длины окна N и для уровней детальности β , удовлетворяющих (2), в первом окне адаптации, для $\tau = N$, запомним все $L_\beta^{(A)}(N)$ q -мерных векторов агрегированных коэффициентов (формулы (10) или (24) в описании программ **WBRCM** и **WBRCM_mult**). Далее, смещая окно вправо на один отсчет, повторяем независимо в каждом окне адаптации всю процедуру определения вектора агрегированных коэффициентов, однако вычисляем и запоминаем не все облако $L_\beta^{(A)}(N)$ q -мерных векторов для данного окна, а лишь один вектор, соответствующий правому концу окна, моменту времени τ . Тем самым производится адаптация агрегированных вейвлет-коэффициентов к совокупному поведению многомерного сигнала на прошлом временном окне длиной N отсчетов. Полученные значения агрегированных коэффициентов рассматриваются как вейвлет-коэффициенты некоторого скалярного сигнала $W(t)$, который может быть найден путем обратного быстрого вейвлет-преобразования и который назовем *вейвлет-агрегированным сигналом*. Таким образом, вейвлет-агрегированный сигнал – это такой сигнал, вейвлет-коэффициенты которого суть главные значения канонических вейвлет-коэффициентов исходных временных рядов, полученные в скользящем временном окне.

В зависимости от того, какая схема вычисления агрегированных коэффициентов используется – обычная или робастная (формулы (3-10) или (16-24) в описании программ **WBRCM** и **WBRCM_mult**), вейвлет-агрегированный сигнал также является либо просто агрегированным, либо робастным агрегированным. Длину окна N можно взять равной полной длине временных рядов – в этом случае вейвлет-агрегированный сигнал будет неадаптивным, подобно Фурье-агрегированному.

Предварительные операции для временных рядов при вычислении вейвлет-агрегированного сигнала состоят в переходе к приращениям, если это необходимо из-за доминирования низкочастотных вариаций, и масштабировании каждой скалярной

компоненты $Z_k(t)$, $k=1, \dots, q$ многомерного ряда путем замены ее значений на величины $U_k(t)$ согласно следующим правилам:

$$U_k(t) = Z_k(t) / (Z_k^{(\max)}(1, N) - Z_k^{(\min)}(1, N)) \text{ для } 1 \leq t \leq N \quad (3(a))$$

$$U_k(t) = Z_k(t) / (Z_k^{(\max)}(\tau - N + 1, \tau) - Z_k^{(\min)}(\tau - N + 1, \tau)) \text{ для } t > N \quad (3(b))$$

$$Z_k^{(\max)}(r, s) = \max_{r \leq t \leq s} Z_k(t), \quad Z_k^{(\min)}(r, s) = \min_{r \leq t \leq s} Z_k(t) \quad (3(c))$$

Операции (3) нормируют каждый временной ряд: в первом окне адаптации на единичный размах (формула (3(a))), а в последующих окнах, смещаемых на один отсчет вправо, нормировка производится только для самого правого отсчета τ , не затрагивая результатов предыдущих нормировок (формула (3(b))). Тем самым достигается адаптация к однородному масштабу вариаций исходных временных рядов в окне длиной N отсчетов только слева от текущей точки.

Еще раз подчеркнем, что если обрабатывать не всю выборку, а лишь начальный кусок произвольной длины, то отбрасывание из обработки последней порции данных никак не скажется на результате анализа первой части данных и агрегированный сигнал на начальном интервале времени будет иметь тот же вид, что и при обработке всей информации

Пример применения.

Рассмотрим применение метода агрегированных сигналов для поиска предвестников сильных землетрясений. На рис.1 представлены графики данных геофизического мониторинга в Северо-Восточном Китае. Данные для анализа были любезно представлены проф. Чанг Чаоченгом, Центр анализа и прогноза землетрясений (Prof. Zhang Zhaocheng, Center for Analysis and Prediction of Earthquakes, State Seismological Bureau, China). Набор исходных данных состоит из 10 временных рядов, представляющих собой результаты синхронной регистрации следующих геофизических параметров: электросопротивление горных пород - 3 ряда, графики 1-3 на рис.3.3; наклоны - 3 ряда, графики 4-6 на рис.1; вариации уровней подземных вод в скважинах - 4 ряда, графики 7-10 на рис.3.3.

Характерный линейный размер наблюдательной сети = 200 км. Обрабатываемый интервал наблюдений составляет 8 лет, с 01 января 1972 года по 31 декабря 1979 года. Интервал взятия отсчетов составляет 1 сутки, что обеспечивает длину рядов по 2922 отсчета в каждом. В течение обработанного интервала наблюдений произошло катастрофическое Тяньшаньское землетрясение - $M=7.8$, 28 июля 1976 года (отмечено вертикальной линией). Момент этого землетрясения соответствует 1671 дню от начала 1972 года и он наиболее ярко отражен в постсейсмической реакции на графике 9 рис.1 вариаций подземных вод в пункте, находящейся непосредственно в эпицентральной зоне.

Для построения вейвлет-агрегированного сигнала был использован вейвлет Хаара, длина временного окна адаптации 700 суток и порог представительности $L_{\min} = 10$. На графике 11 рис.1 представлено поведение агрегированного сигнала первых 1670 отсчетов (т.е. выборки, непосредственно примыкающей к моменту толчка) всех 10 временных рядов. Оно характеризуется амплитудной аномалией, предшествующей землетрясению и начавшейся приблизительно за 100 суток до момента толчка. Кроме того, график 11 содержит другие всплески для моментов времени менее 1000 дней, которые, по всей видимости, являются постсейсмическими реакциями или предвестниками других землетрясений.

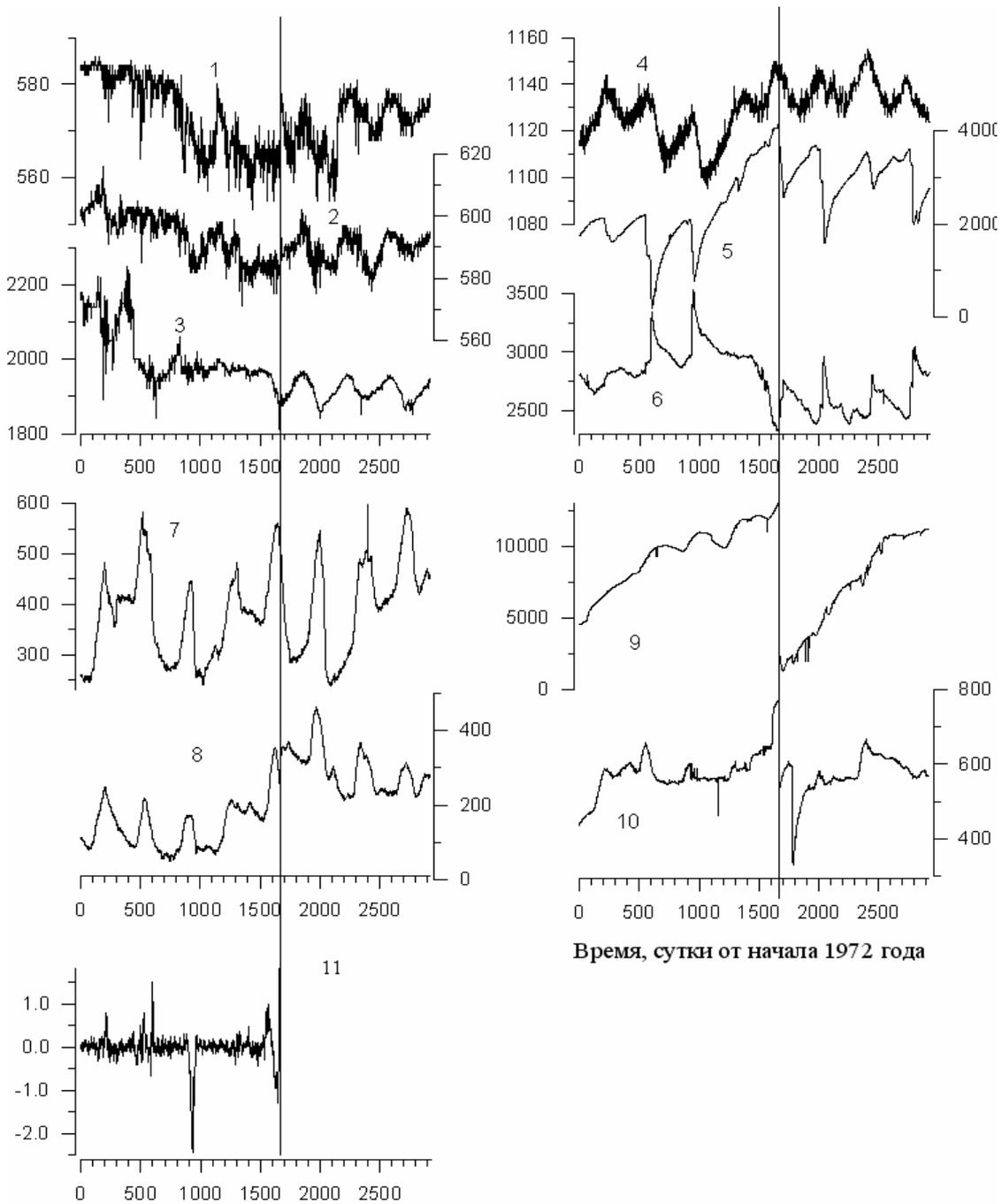


Рис.1.

Рис.2 отражает изменение формы агрегированного сигнала на последнем участке, после 1000-го дня, по мере увеличения длины N обрабатываемой выборки (от 1500 до 1670 отсчетов) и ее приближения к моменту толчка. Видно, что наиболее контрастный предвестник сформировался за 5 суток до землетрясения.

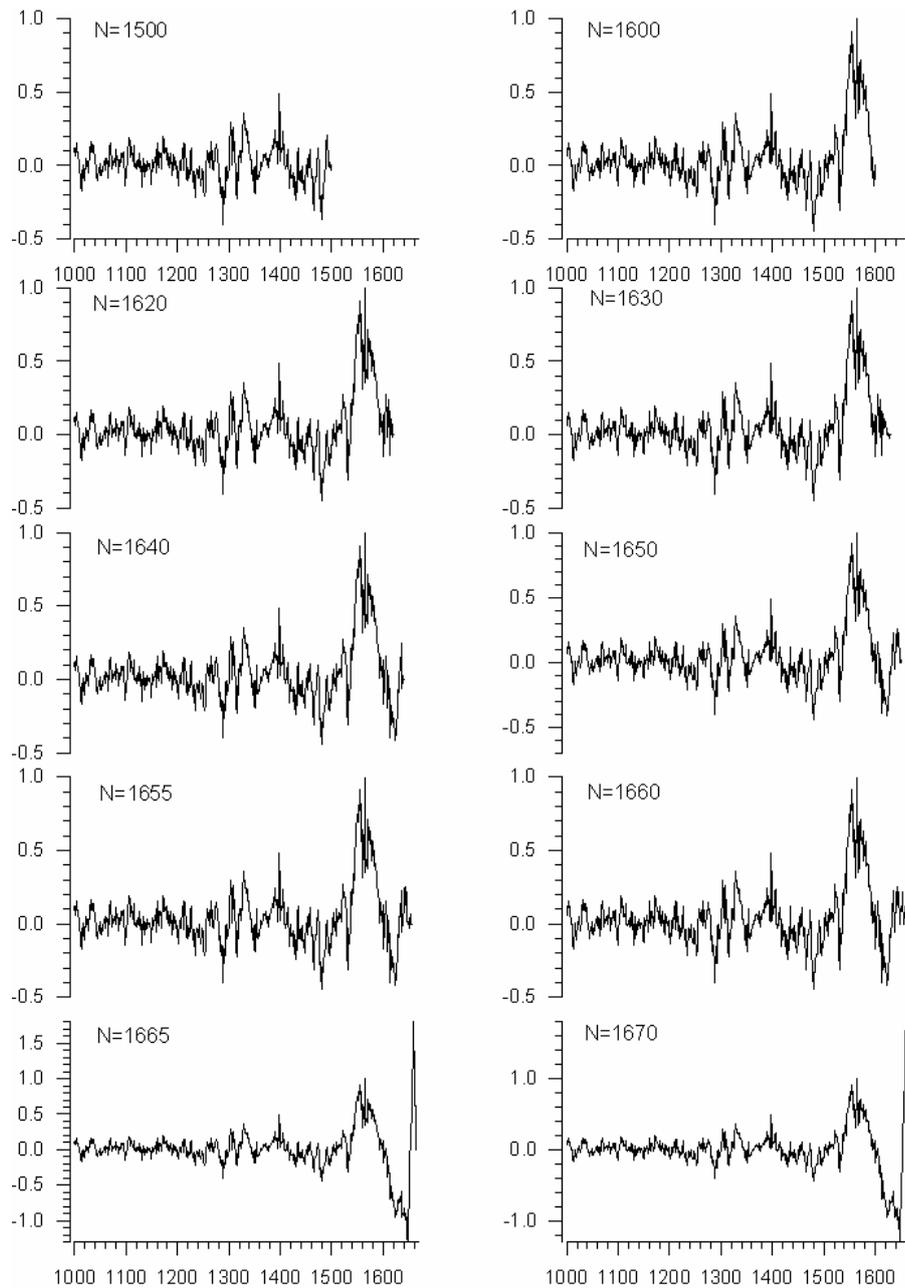


Рис.2.

На рис.3(а) изображен график вейвлет-агрегированного сигнала, вычисленного с теми же параметрами уже для всей выборки, на котором видна аномалия, предшествующая Тяньшанскому землетрясению. Сам момент землетрясения выделен второй вертикальной линией. На рис.3(б) представлен график робастной модификации агрегированного сигнала, которая также надежно выделяет предвестник Тяньшанского землетрясения. Еще раз подчеркнем, что алгоритм агрегации временных рядов является лево-ориентированным. Последнее означает, что если обрабатывать не всю выборку, а лишь начальный кусок произвольной длины, то отбрасывание из обработки последней порции данных никак не скажется на результате анализа первой части данных и агрегированный сигнал на начальном интервале времени будет иметь тот же вид, что и при обработке всей информации.

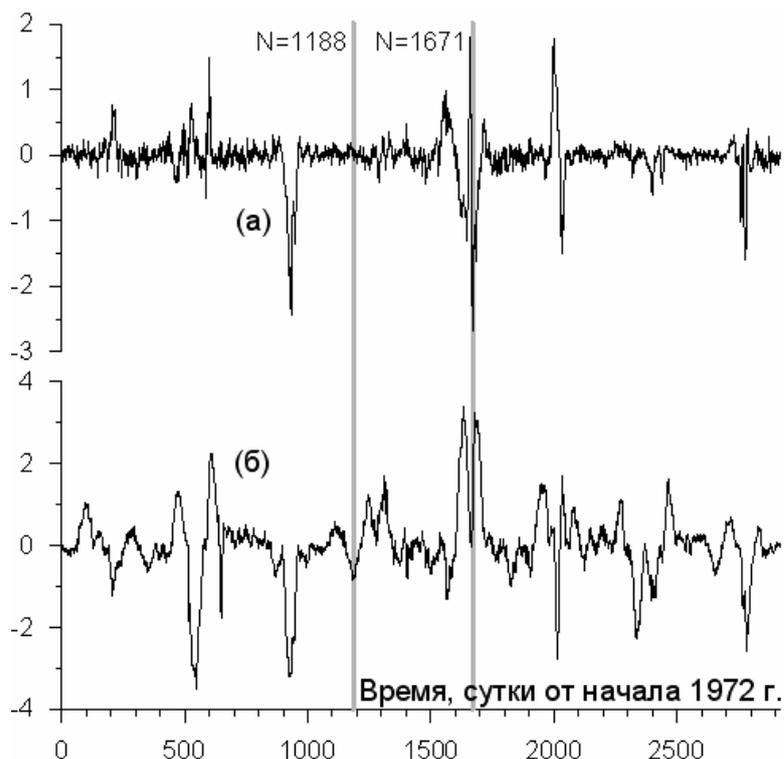


Рис.3.

На первый взгляд рис.3(а) производит большее впечатление, поскольку те аномалии, которые на нем видны, больше выделяются над общим фоном статистических флуктуаций. Однако известно, что перед Тяньшанским землетрясением в том же регионе Китая имело место Хайчэнгское (02.04.1975, $M=7.4$, 1188-й день от начала 1972 года), пожалуй, единственное в практике прогноза, успешно предсказанное (по другим данным). На рис.3 оно также отмечено вертикальной штриховой линией. Удручает то, что на рис.3(а) это событие не отмечено никаким статистически значимым эффектом синхронизации, ни до события, ни после него. В то же время на рис.3(б) заметен как предвестниковый, так и постсейсмический эффекты. Кроме того, все аномалии, заметные на рис.3(а), также присутствуют и на рис.3(б), но в более контрастном виде.

Таким образом, робастный вейвлет-агрегированный сигнал в данном частном случае оказался более чувствителен, чем его предыдущий вариант, хотя делать вывод, что для анализа временных рядов низкочастотного геофизического мониторинга всегда следует использовать робастную агрегацию, преждевременно. Здесь дело не только в том, что робастный вариант метода требует времени счета, на порядок больше, чем старый вариант. Излишняя чувствительность метода может привести к тому, что он будет выделять не только сильные эффекты синхронизации, предшествующие сильным землетрясениям (а именно они и представляют интерес), но менее сильные общие вариации, которые, тем самым, увеличат амплитуду общего фона статистических флуктуаций. Сравнение рис.3(а) и рис.3(б) подтверждают это предположение, хотя реакция робастного сигнала на Хайчэнгское землетрясение является, безусловно, сильным доводом в пользу применения робастной схемы агрегации временных рядов.

Каждая из выделенных выше значимых аномалий, помимо заведомо связанной с Тяньшанской катастрофой и рассмотренной выше, является индикатором увеличения коллективного поведения регистрируемых процессов. Здесь мы не будем пытаться каждую аномалию интерпретировать как предвестник того или иного сильного события в каталоге землетрясений Северо-Восточного Китая. Ведь причиной увеличения коллективного поведения может быть и рой землетрясений слабой силы, и увеличение криповых движений

типа т.н. “медленных событий”. Целью рассмотренного примера анализа данных была именно попытка выделения краткосрочных предвестников Тяньшанского землетрясения, и предложенная методика показала свою перспективность.

ЛИТЕРАТУРА.

- Любушин А.А.* (2000) Вейвлет-агрегированный сигнал и синхронные всплески в задачах геофизического мониторинга и прогноза землетрясений. – Физика Земли, 2000, N3. С.20-30.
- Любушин А.А.* (2002) Робастный вейвлет-агрегированный сигнал для задач геофизического мониторинга – Физика Земли. 2002, N9. С.37-48.
- Любушин А.А.* , *Копылова Г.Н.* (2004) Многомерный вейвлет-анализ временных рядов электротеллурических наблюдений на Камчатке – Физика Земли, 2004, N2, с.82-96.
- Любушин А.А.* (2007) «Анализ данных систем геофизического и экологического мониторинга». М.: Наука, 2007, 228с.