Программа AggF

оценки Фурье-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

Программа предназначена для выделения общих стационарных гармонических колебаний, 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.

Описание программы.

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 исходное данные представляют собой таблицы, то временные метки и прочая служебная информация не должны находится 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**).

Программа считывает 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

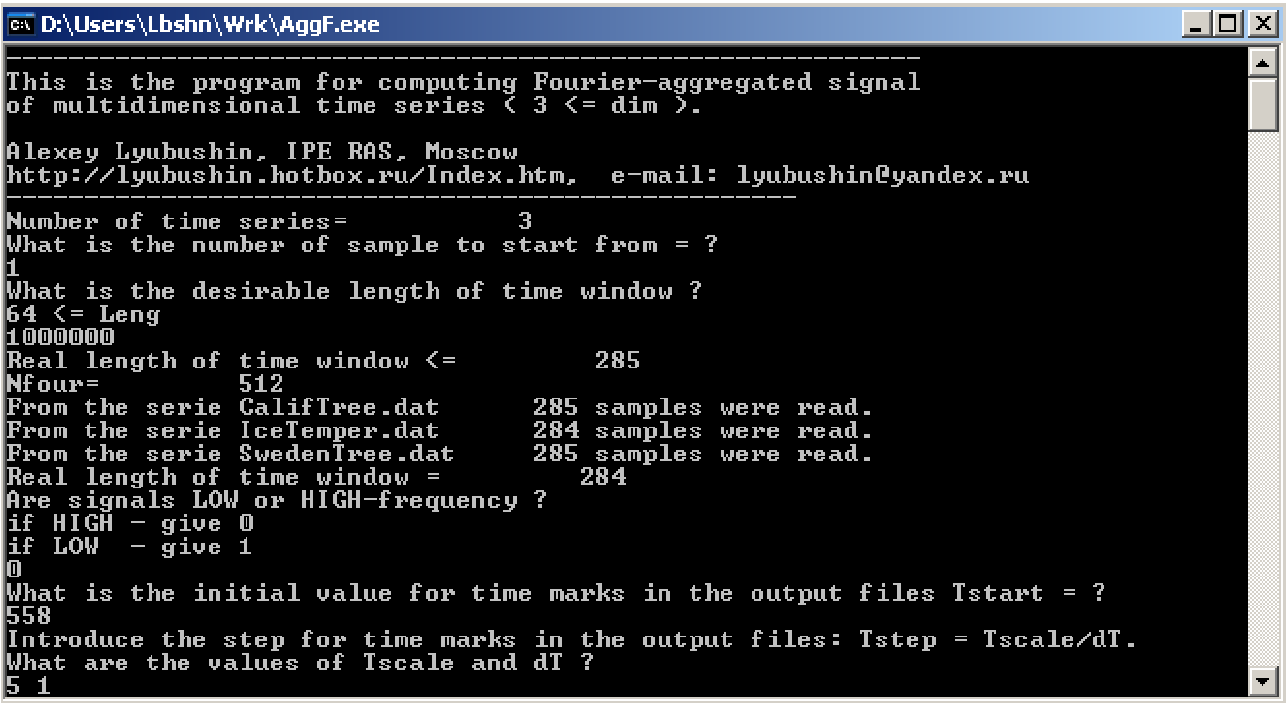
проанализировать всю допустимую длину выборок, но минимальная длина неизвестна, то можно ввести произвольно большое число, например «1000000000» и тогда программа сама найдет реальную длину обрабатываемого фрагмента.

3) Следует ответить на вопрос, каков характер временных рядов: высокочастотный (HIGH) или низкочастотный (LOW). Различие в обработке заключатся в том, что для низкочастотных временных рядов осуществляется предварительный переход к приращениям для всех обрабатываемых рядов, агрегированный сигнал строится для приращений, а потом агрегированные приращения «интегрируются», то есть вычисляется их кумулятивная сумма от нулевого начального значения. Переход к приращениям необходим в случае сильно низкочастотных временных рядов для обеспечения большей стационарности выборок.

4) Следует ввести значение T_{ini} начальной временной метки в выводном файле.

5) 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, иллюстрирующая диалог с программой:



```
D:\Users\Lbshn\Wrk\AggF.exe
------------------------------------------------------------
This is the program for computing Fourier-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 time series=          3
What is the number of sample to start from = ?
1
What is the desirable length of time window ?
64 <= Leng
1000000
Real length of time window <=         285
Nfour=          512
From the serie CalifTree.dat      285 samples were read.
From the serie IceTemper.dat      284 samples were read.
From the serie SwedenTree.dat     285 samples were read.
Real length of time window =        284
Are signals LOW or HIGH-frequency ?
if HIGH - give 0
if LOW  - give 1
0
What is the initial value for time marks in the output files Tstart = ?
558
Introduce the step for time marks in the output files: Tstep = Tscale/dT.
What are the values of Tscale and dT ?
5 1
```

Затем программа начинает работать и создает в рабочей директории выводной файл (текстовую числовую таблицу) со стандартным именем “AggF.dat”, состоящий из 2-х колонок: 1-я колонка суть значения агрегированного сигнала, а 2-я колонка – это временные метки отсчетов.

Описание метода.

Агрегированный сигнал строится в два этапа. На первом этапе исходный многомерный ряд заменяется на многомерный же ряд т.н. канонических компонент, которые сохраняют общие сигналы и освобождены от локальных. На втором этапе общие сигналы дополнительно усиливаются путем построения одного скалярного ряда – их первой главной компоненты, который и назван агрегированным сигналом исходного многомерного временного ряда. В зависимости от того, используются для разложения сигнала классические Фурье-методы или вейвлет-разложения, агрегированный сигнал будет называться Фурье-агрегированным или вейвлет-агрегированным. Фурье-агрегированный сигнал был предложен и применен для анализа геофизических временных рядов в [*Любушин, 1998*]. В дальнейшем он применялся не только в физике твердой Земли, но и в гидрологии, метеорологии и океанографии [*Любушин и др., 2004; Любушин, 2007*]

Определим Фурье-агрегированный сигнал, который предназначен для выделения общих стационарных гармонических колебаний и основан на анализе не в скользящем временном окне сравнительно небольшой длины, а по всей имеющейся выборке, либо по ее заданной части. Рассмотрим q-мерный временной ряда $Z(t)$. Выделим из многомерного ряда наблюдений $Z(t)$ i-ую скалярную компоненту $Z_i(t)$ и постараемся так профильтровать $(q-1)$-мерный ряд $Z^{(i)}(t)$, составленный из оставшихся компонент, чтобы получившийся на выходе фильтра скалярный сигнал $C_i^{(Z)}(t)$ имел бы с выделенным рядом $Z_i(t)$ максимальную когерентность на каждой частоте. Для этого необходимо в качестве многомерного частотного фильтра использовать компоненты собственного вектора матрицы:

$$U(\omega) = S_{xx}^{-1/2} S_{xy} S_{yy}^{-1} S_{yx} S_{xx}^{-1/2} \tag{1}$$

в которой в качестве $Y(t)$ фигурирует $Z_i(t)$, а в качестве $X(t)$ – $Z^{(i)}(t)$. В формуле (1) ω - частота, $S_{xx}(\omega)$ - спектральная матрица временного ряда $X(t)$, $S_{xy}(\omega)$ - кросс-спектральная матрица, $S_{yx}(\omega) = S_{xy}^H(\omega)$, “$H$” – знак эрмитова сопряжения. Фильтром служат компоненты собственного вектора, соответствующие максимальному собственному числу этой матрицы, которое равно $\nu_i^2(\omega)$ - покомпонентной квадратичной канонической когерентности [*Brillinger, 1975; Hanan, 1970*] (см. описание программы **SpectCohMes**). Смысл подобной операции состоит в том, что если в компоненте $Z_i(t)$ присутствуют помехи, характерные только для этого ряда и отсутствующие в других компонентах ряда $Z(t)$, то в сигнале $C_i^{(Z)}(t)$ они отсутствуют просто в силу построения. В то же время в ряду $C_i^{(Z)}(t)$ сохраняются все составляющие компоненты $Z_i(t)$, общие для остальных составляющих ряда $Z(t)$ – сигнала $Z^{(i)}(t)$. Назовем ряд $C_i^{(Z)}(t)$ канонической компонентой скалярного ряда $Z_i(t)$.

Определим агрегированный сигнал $A_Z(t)$ многомерного временного ряда $Z(t)$ как первую главную спектральную компоненту многомерного ряда $C^{(Z)}(t)$, составленного из канонических компонент $C_i^{(Z)}(t)$ каждого скалярного временного ряда, образующего исходный ряд $Z(t)$. Напомним, что главной спектральной компонентой в частотной области [*Brillinger, 1975; Hanan, 1970*] называется проекция вектора преобразований Фурье от исходных сигналов на собственный вектор спектральной матрицы, соответствующий максимальному собственному числу. Главная компонента во временном представлении получается обратным преобразованием Фурье от главной компоненты в частотном представлении. Подчеркнем отличие ряда $A_Z(t)$ от простой первой главной компоненты. И в том и в другом случае ряды определяются путем многомерных фильтраций, в которых в качестве многомерных частотных фильтров берутся собственные вектора спектральных матриц, соответствующих их максимальным собственным числам. Однако для обычной первой главной компоненты такой спектральной матрицей является матрица исходного временного ряда $Z(t)$, а для $A_Z(t)$ - спектральная матрица ряда $C^{(Z)}(t)$. Хотя в ходе и той и другой фильтрации происходит выделение общих составляющих, агрегированный сигнал $A_Z(t)$ обладает преимуществом, поскольку в ходе его построения уничтожаются индивидуальные помехи.

В отличие от оценки в скользящем временном окне, для оценки спектральной матрицы, необходимой для построения агрегированного сигнал, использовалась непараметрическая оценка [*Brillinger, 1975; Hanan, 1970*], путем частотного усреднения периодограмм и кросс-периодограмм. Такой выбор связан с большей структурной устойчивостью классических периодограммных оценок спектров мощности при наличии длинных временных рядов по сравнению с параметрическими авторегрессионными оценками спектральных матриц, которые обладают преимуществами для коротких выборок. По умолчанию используется глубокое усреднение (сглаживание) периодограмм в частотном окне длиной 1/32 часть от общего числа частотных дискретов. Если временные ряды обладали доминирующими низкими частотами, то осуществлялся предварительный переход к приращениям и нормировка на единичную выборочную дисперсию. Перед вычислением спектральных матриц данные подвергались процедурам винзоризации (итерационного подавления выбросов) и косинусного сглаживания на концах с помощью весовой функции:

$$g(u)=\begin{cases}(1-\cos(\pi u/\alpha))/2, & 0\le u<\alpha\\ 1, \quad \alpha\le u\le 1-\alpha, & 0<\alpha<0.5\\ (1-\cos(\pi(1-u)/\alpha))/2, & 1-\alpha<u\le 1\end{cases} \tag{2}$$

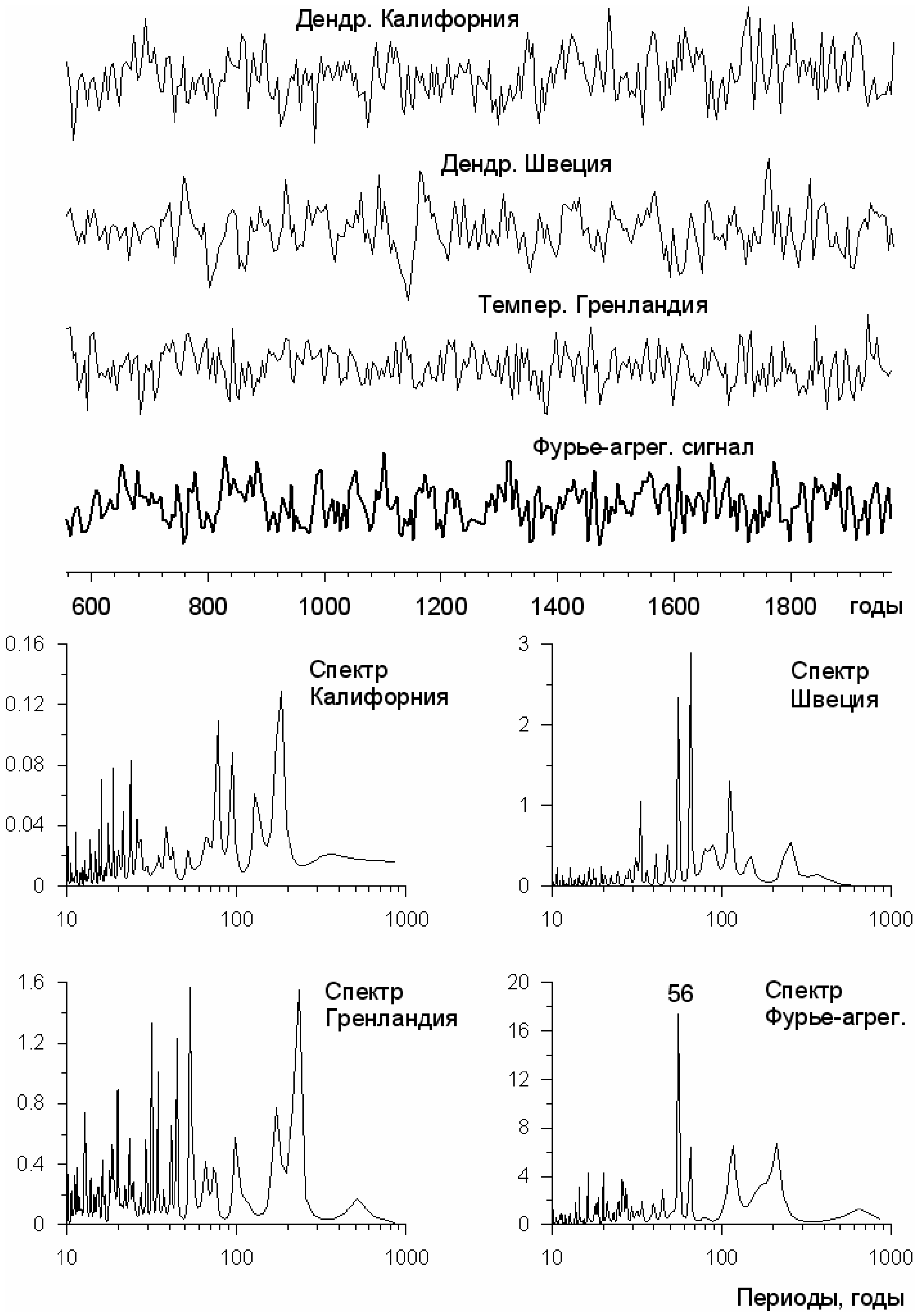
где α – часть длины интервала наблюдений или временного окна, которая «жертвуется» на операцию сглаживания выборки с целью уменьшения смещения оценки, всюду ниже бралось значение $\alpha = 0.125$.

Однако эти операции выполнялись лишь для оценки спектральной матрицы. Затем, для вычисления канонических компонент $C_i^{(Z)}(t)$ данные еще раз преобразовывались в частотное представление с помощью быстрого дискретного преобразования Фурье, но уже без сглаживания и итеративного устранения выбросов, а получившиеся многомерные преобразования Фурье проецировались на собственные вектора. Точно такая же последовательность операций повторялась для вычисления агрегированного сигнала: сначала вычисление спектральной матрицы канонических компонент, нахождение для каждого значения частоты собственного вектора, соответствующего максимальному собственному числу, проекция Фурье-представлений канонических компонент на этот вектор и, наконец, обратное преобразование Фурье от результата проецирования с целью получения временной реализации агрегированного сигнала $A_Z(t)$.

Если осуществлялся переход от исходных временных рядов к рядам в приращениях, то агрегированный сигнал также будет построен для приращений. Поэтому финальная операция в этом случае заключается в интегрировании ряда, то есть вычислении в каждый момент времени суммы текущего и всех предыдущих значений. Отметим, что агрегированный сигнал не имеет физической размерности, поскольку он строится после последовательности операций нормировки исходных данных, его смысл состоит лишь формальном выделении наиболее общих гармонических вариаций.

Пример применения.

Ниже приведен пример использования Фурье-агрегированного сигнала для выделения общей слабой гармонической компоненты, присутствующей в 3-х климатических временных рядах [*Любушин, 2007*].



Сверху вниз представлены графики трех климатических временных рядов: двух дендрохронологий из Калифорнии (остистая сосна) и Швеции (шотландская сосна) и реконструкции зимних температур в Гренландии по ледовым кернам, относящихся к интервалу времени 553-1973 гг. Временные ряды приведены (путем усреднения и прореживания) к интервалу дискретизации 5 лет. Далее идут графики их Фурье-агрегированного сигнала (в «высокочастотной моде») и оценок спектров мощности как исходных рядов, так и агрегированного сигнала. Из сравнения спектров видно, что общей периодической компонентой рядов является лишь 56-годовая периодичность.

Литература.

Любушин А.А. (1998) Агрегированный сигнал систем низкочастотного геофизического мониторинга. – Физика Земли. N3. С.69-74.

Любушин А.А., В.Ф.Писаренко, М.В.Болгов, М.В.Родкин, Т.А.Рукавишникова (2004) Синхронные вариации уровня Каспийского моря по береговым наблюдениям, 1977-1991 гг. – Известия РАН. Физика атмосферы и океана, 2004, N6, том 40, с.821-831.

Любушин А.А. (2007) «Анализ данных систем геофизического и экологического мониторинга». М.: Наука, 2007, 228с.

Brillinger D.R. (1975) Time series. Data analysis and theory. Holt, Rinehart and Winston, Inc., N.Y., Chicago, San Francisco (Русский перевод: *Бриллинджер Д.* (1980) Временные ряды. Обработка данных и теория. М., Мир. 536с.)

Hannan E.J. (1970) Multiple time series. John Wiley and Sons, Inc., N.Y., London, Sydney, Toronto (Русский перевод: *Хеннан Э.* (1974) Многомерные временные ряды. М., Мир, 575с.)