

Программа LinShares

оценки меры линейного влияния друг на друга 2-х точечных процессов.

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 оценки степени влияния друг на друга 2-х точечных процессов (последовательностей времен событий). Оценка может быть получена как по всей имеющейся выборке значений моментов времени событий, так и в скользящем временном окне заданной длины и с заданным смещением.

Для 2-х обычных временных рядов для оценки подобного рода может быть использован коэффициент корреляции. В этом случае степень линейной связи оценивается как модуль или квадрат модуля коэффициента корреляции, причем эта мера линейной связи является симметричной. Если интересоваться частотной зависимостью меры линейной связи между 2-мя временными рядами, то можно использовать квадрат модуля спектра когерентности (программы **CrosSp.exe** и **CrosCor.exe**).

При рассмотрении 2-х реализаций точечных процессов (последовательностей времен событий) вопрос оценки степени их взаимной связи является более сложным как с теоретической, так и с методической точки зрения. Эти сложности обусловлены тем, что последовательности времен могут содержать как «затишья» (длительные интервалы отсутствия событий), так и «рой событий», когда промежуток времени между соседними событиями становится малым. Очевидно, что мера связи между 2-мя точечными процессами не может быть симметричной, подобно коэффициенту корреляции или спектру когерентности, поскольку если одна последовательность событий в среднем на некотором временном масштабе τ всегда предвещает другую, то одна из них может быть названа «причиной», а вторая – «следствием». Если их поменять местами, то это должно отражаться и в оценке меры связи между последовательностями событий.

Очевидно также, что степень влияния одной последовательности событий на другую должно зависеть от временного масштаба τ , на котором они рассматриваются. Поэтому программа оценки содержит свободный параметр времени τ («характерное время» сравниваемых точечных процессов), который должен выбираться пользователем из физических или каких-то других соображений. Практически выбор этого параметра осуществляется путем перебора по разумному множеству значений и сравнением результатов для различных τ .

Излагаемый ниже метод является частным случаем общего метода, разработанного для совместного анализа произвольного числа последовательностей событий [Любушин, Писаренко, 1993], см. также [Любушин, 2007], параграф 2.6, стр. 142-148.

Описание программы.

Входные файлы должны представлять собой ASCII числовые таблицы, первая колонка которых является монотонно неубывающей последовательностью моментов времен событий. Запуск программы производится из командной строки с 2-мя аргументами –

именами 2-х входных файлов, которые должны находиться в текущей директории. Обычно это делается с использованием комбинаций клавиш Ctrl-Enter. Например:

```
LinShares.exe North.dat South.dat
```

– здесь 1-й входной файл с именем “North.dat” содержит первую последовательность времен, а файл “South.dat” – вторую. В данном случае входные файлы представляют собой моменты времен землетрясений с магнитудой $M \geq 7.0$ в Северном и Южном полушариях за период 1901-2005 гг. (более подробно – см. ниже, в примере применения).

Ниже приведены 2 примера копий экрана консольного приложения **LinShares.exe**, реализующего диалог с пользователем.

Первый случай применения – это когда оценку требуется получить для всей реализации моментов времени событий:

```

D:\Users\Lbshn\Wrk\LinShares.exe
This is the program for estimating shares of intensity for point processes:
T1<j> & T2<k> - sequences of time moments of events,
within moving time window.
For each process the intensity is split into 3 shares:
 1) Poissonian (pure stochastic);
 2) Self-exciting;
 3) External (due to influence of other process).
The sum of these shares equals 1.
The program contains free parameter Tau which is the time scale
for considering the interacting between 2 point processes.
Input files must be any ASCII tables with the 1-st column presenting monotone
non-decreasing time moments. Output file has a name "Intensity_Shares.txt".

Alexey Lyubushin, IPE RAS, Moscow, Russia
http://AlexeyLyubushin.narod.ru/ e-mail: lyubushin@yandex.ru
-----
Number of T1 time moments =      894
Number of T2 time moments =      799
Time scale value Tau= ? < * > 0 >
1
What is the length of time window = ?
1000000_

```

В этом примере на запрос о значении временного масштаба τ вводится значение 1 (в данном случае – 1 год). Далее следует запрос о длине скользящего временного окна. Для того, чтобы получить оценку для всей реализации, эту длину следует задать произвольно большой (в данном случае – миллион). После ввода этих 2-х параметров программа начинает считать и создает выводной файл со стандартным именем “Intensity_Shares.txt”. Например:

```

1-st time sequence filename=North.dat
2-nd time sequence filename=South.dat
Relaxation time Tau=      1.00000

0.455  0.400  0.146
0.123  0.621  0.256

```

Последние 2 строки выводного файл представляют собой значения частей интенсивности процессов, обусловленных чисто случайным поведением («пуассоновская часть»), самовозбуждающейся компонентой (то есть влиянием процесса самого на себя) и внешним влиянием со стороны второго процесса. Первая строка этой матрицы дает разбиение интенсивности для случая, когда рассматривается влияние 2-го процесса на 1-й,

а вторая строка – наоборот, 1-го процесса на 2-ой. Таким образом, в этом примере интенсивность 1-го процесса раскладывается на такие составляющие: 45.5% обусловлены чисто стохастической компонентой, 40% – самовозбуждением и 14.6% – влиянием 2-го процесса. Вторая строка матрицы дает разложение интенсивности 2-го процесса соответственно: 12.3%, 62.1% и 25.6%. Очевидно, что этот результат зависит от выбора параметра времени τ . Ниже будет рассмотрен пример этих же данных для других значений τ .

Второй случай применения – это когда оценку требуется получить в скользящем временном окне заданной длины и с заданным смещением:

```

D:\Users\Lbshn\Wrk\LinShares.exe
This is the program for estimating shares of intensity for point processes:
T1(j) & T2(k) - sequences of time moments of events,
within moving time window.
For each process the intensity is split into 3 shares:
1) Poissonian (pure stochastic);
2) Self-exciting;
3) External (due to influence of other process).
The sum of these shares equals 1.
The program contains free parameter Tau which is the time scale
for considering the interacting between 2 point processes.
Input files must be any ASCII tables with the 1-st column presenting monotone
non-decreasing time moments. Output file has a name "Intensity_Shares.txt".

Alexey Lyubushin, IPE RAS, Moscow, Russia
http://AlexeyLyubushin.narod.ru/ e-mail: lyubushin@yandex.ru

Number of T1 time moments =      894
Number of T2 time moments =      799
Time scale value Tau= ? ( * > 0 )
1
What is the length of time window = ?
20
What is the shift of moving time window = ?
1

```

В этом примере на запрос о значении временного масштаба τ снова вводится значение 1 (в данном случае – 1 год). Далее следует запрос о длине скользящего временного окна и вводится значение 20 (20 лет). Если длина окна меньше длины выборки, то оценка производится в скользящем временном окне (а не по всей выборке) и следует запрос на смещение временного окна. В данном случае смещение равно 1 (1 год). После ввода этих 3-х параметров программа начинает считать и создает выводной файл с тем же стандартным именем “Intensity_Shares.txt”, но его структура будет иной. Например:

```

1-st time sequence filename=North.dat
2-nd time sequence filename=South.dat
Relaxation time Tau=      1.00000

1921.06      142      87  1.000  0.000  0.000  0.337  0.544  0.119
1922.06      140      92  0.951  0.049  0.000  0.421  0.579  0.000
1923.06      143      94  1.000  0.000  0.000  0.308  0.582  0.109
1924.06      154      98  0.945  0.055  0.000  0.033  0.594  0.373
1925.06      159     102  0.876  0.124  0.000  0.550  0.428  0.022
1926.06      160     107  0.755  0.245  0.000  0.659  0.217  0.124
1927.06      159     115  0.778  0.222  0.000  0.320  0.377  0.303
.....
.....

```

Здесь после информационной «шапки» следует числовая таблица, имеющая 9 следующих колонок:

- колонка 1 – моменты времени, соответствующие правому концу скользящего временного окна;

- колонки 2 и 3 – число событий для 1-й и 2-й последовательности моментов времени в текущем временном окне;
- колонки 4, 5 и 6 – пуассоновская, самовозбуждающаяся и внешняя части средней интенсивности 1-й последовательности моментов времени в текущем окне;
- колонки 7, 8 и 9 – пуассоновская, самовозбуждающаяся и внешняя части средней интенсивности 2-й последовательности моментов времени в текущем окне.

Описание метода.

Пусть $t_j^{(\alpha)}$, $j=1, \dots, N_\alpha$; $\alpha=1, 2$ представляют собой моменты времен 2-х потоков событий. Представим интенсивность какого-нибудь процесса в виде:

$$\lambda^{(\alpha)}(t) = b_0^{(\alpha)} + \sum_{\beta=1}^2 b_\beta^{(\alpha)} \cdot g^{(\beta)}(t) \quad (1)$$

где $b_0^{(\alpha)} \geq 0, b_\beta^{(\alpha)} \geq 0$ - параметры, $g^{(\beta)}(t)$ - функция влияния событий потока с номером β . Представим функцию влияния в виде:

$$g^{(\beta)}(t) = \sum_{t_j^{(\beta)} < t} \exp(-(t - t_j^{(\beta)}) / \tau) \quad (2)$$

где $\tau > 0$ является характерным временным масштабом рассмотрения взаимодействия между потоками событий. Таким образом, в соответствии с формулой (2), вес события с номером j становится ненулевым для времен $t > t_j^{(\beta)}$ и экспоненциально затухает с характерным временем τ по мере возрастания текущего времени t . Сумма всех таких затухающих экспонент образует функцию влияния $g^{(\beta)}(t)$ потока с номером β . Параметр $b_\beta^{(\alpha)}$ является масштабирующим множителем и именно он определяет степень влияния потока β на поток α : $\beta \rightarrow \alpha$. Параметр $b_\alpha^{(\alpha)}$ определяет степень влияния потока α на самого себя (самовозбуждение), а параметр $b_0^{(\alpha)}$ отражает число случайную компоненту интенсивности, для которой функция влияния постоянна и тождественно равна 1.

Зафиксируем параметр τ и рассмотрим задачу определения параметров $b_0^{(\alpha)}, b_\beta^{(\alpha)}$. Будем использовать метод максимального правдоподобия. Логарифмическая функция правдоподобия для нестационарного пуассоновского процесса равна [Cox, Lewis, 1966; Cox, 1975]:

$$\ln(L_\alpha) = \sum_{j=1}^{N_\alpha} \ln(\lambda^{(\alpha)}(t_j^{(\alpha)})) - \int_0^T \lambda^{(\alpha)}(s) ds, \quad \alpha = 1, 2 \quad (3)$$

где $[0, T]$ есть интервал наблюдения. Таким образом, необходимо найти максимум функции (3) по отношению к параметрам $b_0^{(\alpha)}, b_\beta^{(\alpha)}$. Принимая во внимание формулу (1) и используя правило дифференцирования сложной функции, нетрудно получить следующее выражение:

$$b_0^{(\alpha)} \frac{\partial \ln(L_\alpha)}{\partial b_0^{(\alpha)}} + \sum_{\beta=1}^2 b_\beta^{(\alpha)} \frac{\partial \ln(L_\alpha)}{\partial b_\beta^{(\alpha)}} = N_\alpha - \int_0^T \lambda^{(\alpha)}(s) ds \quad (4)$$

Поскольку параметры $b_0^{(\alpha)}, b_\beta^{(\alpha)}$ должны быть неотрицательными, то каждый член в левой части формулы (4) равен нулю в точке максимума функции (3) – либо в силу необходимых условий экстремума (если параметры положительны), либо, если максимума достигается на границе, то сами параметры равны нулю. Следовательно, в точке максимума логарифмической функции правдоподобия (3) выполняется равенство:

$$\int_0^T \lambda^{(\alpha)}(s) ds = N_\alpha \quad (5)$$

Подставим выражение (1) в (5) и разделим на длину интервала наблюдения. Тогда получим другой вид формулы (5):

$$b_0^{(\alpha)} + \sum_{\beta=1}^m b_\beta^{(\alpha)} \cdot \bar{g}^{(\beta)} = \lambda_0^{(\alpha)} \equiv \frac{N_\alpha}{T} \quad (6)$$

где $\bar{g}^{(\beta)} = \frac{1}{T} \int_0^T g^{(\beta)}(s) ds$ - среднее значение функции влияния. Подставляя $b_0^{(\alpha)}$ из (6) в (3), получим следующую задачу на максимум, эквивалентную задаче (3):

$$\Phi^{(\alpha)}(b_1^{(\alpha)}, b_2^{(\alpha)}) = \sum_{j=1}^{N_\alpha} \ln(\lambda_0^{(\alpha)} + \sum_{\beta=1}^2 b_\beta^{(\alpha)} \cdot \Delta g^{(\beta)}(t_j^{(\alpha)})) \rightarrow \max \quad (7)$$

где $\Delta g^{(\beta)}(t) = g^{(\beta)}(t) - \bar{g}^{(\beta)}$, при ограничениях:

$$b_1^{(\alpha)} \geq 0, b_2^{(\alpha)} \geq 0, \sum_{\beta=1}^2 b_\beta^{(\alpha)} \bar{g}^{(\beta)} \leq \lambda_0^{(\alpha)} \quad (8)$$

Нетрудно показать, что функция (7) является выпуклой с отрицательно определенным гессианом и, следовательно, задача (7), (8) имеет единственное решение. В программе **LinShares** эта задача решается численно методом проекции градиента.

Решив задачу (7), (8) для заданного τ , можно ввести доли интенсивности $\mu_\beta^{(\alpha)}$, $\alpha = 1, 2$; $\beta = 0, 1, 2$ согласно формулам:

$$\kappa_0^{(\alpha)} = \frac{b_0^{(\alpha)}}{\lambda_0^{(\alpha)}} \geq 0, \quad \kappa_\beta^{(\alpha)} = \frac{b_\beta^{(\alpha)} \cdot \bar{g}^{(\beta)}}{\lambda_0^{(\alpha)}} \geq 0 \quad (9)$$

которые можно назвать элементами матрицы влияния.

Интерпретация этих величин вполне естественна: $\kappa_0^{(\alpha)}$ является частью средней интенсивности процесса с номером α , являющейся чисто стохастической, часть $\kappa_\alpha^{(\alpha)}$

вызвана влиянием самовозбуждения $\alpha \rightarrow \alpha$ и $\kappa_{\beta}^{(\alpha)}$, $\beta \neq \alpha$ обусловлена внешним влиянием $\beta \rightarrow \alpha$. Из формулы (5) вытекает условие нормировки:

$$\kappa_0^{(\alpha)} + \sum_{\beta=1}^2 \kappa_{\beta}^{(\alpha)} = 1 \quad (10)$$

Пример 1: два сейсмических каталога.

В качестве первого примера рассмотрим задачу о гипотетической связи сейсмичности Северного и Южного полушарий. На рис.1 представлены моменты времен сильных землетрясений за период с 1901 по 2005 гг.

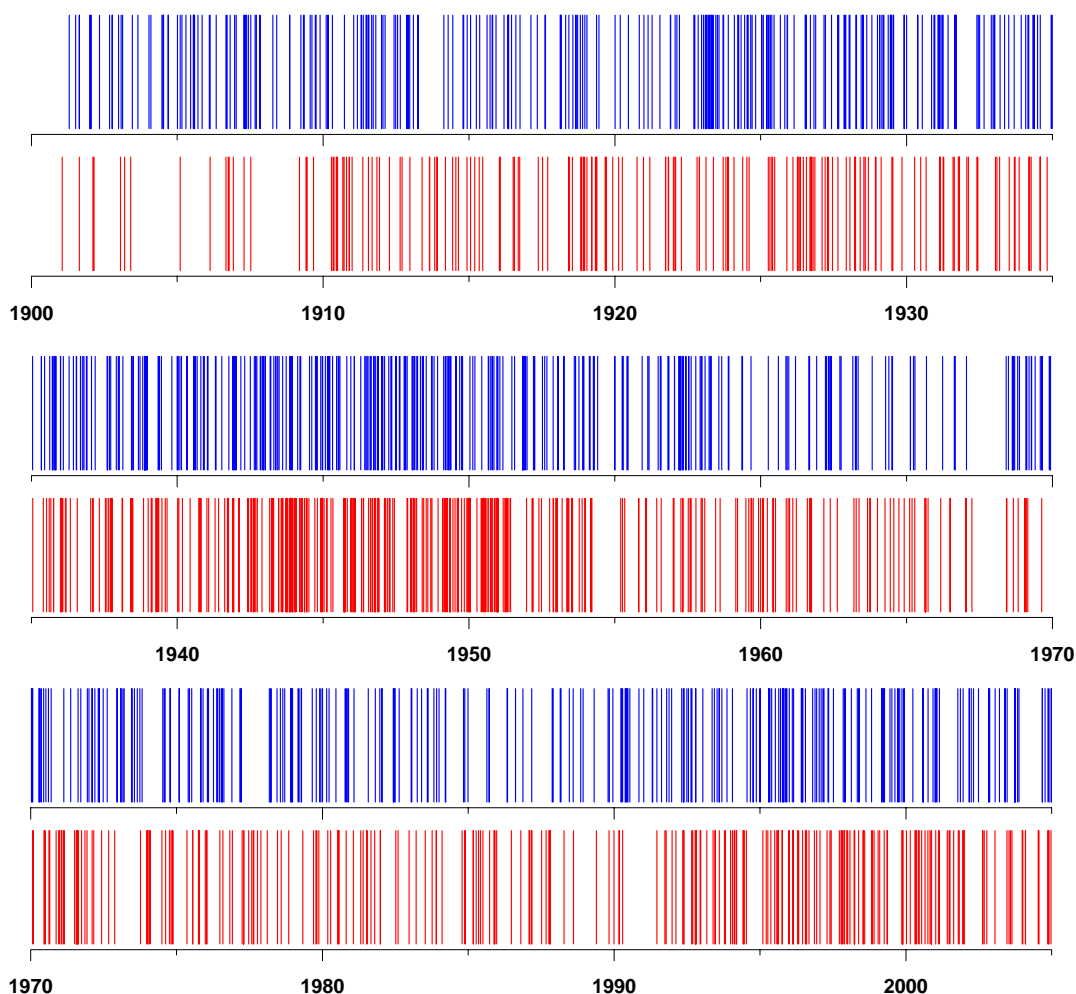


Рис.1. Последовательность моментов времени – землетрясений с магнитудой $M \geq 7.0$ с начала 1901 года по конец 2005 года. Синие линии – землетрясения в Северном полушарии (1-я последовательность времен, 894 события), красные линии – землетрясения в Южном полушарии (2-я последовательность времен, 799 событий).

Моменты событий заданы в «дробных» годах. В следующих таблицах представлены элементы матриц взаимного влияния для разных значений параметра τ .

$\tau = 0.25$ года	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Северное полушарие	0.619	0.258	0.123
Южное полушарие	0.389	0.481	0.130

$\tau = 0.5$ года	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Северное полушарие	0.523	0.333	0.144
Южное полушарие	0.238	0.576	0.187

$\tau = 1$ год	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Северное полушарие	0.455	0.400	0.146
Южное полушарие	0.123	0.621	0.256

$\tau = 2$ года	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Северное полушарие	0.439	0.447	0.114
Южное полушарие	0.077	0.584	0.339

Из этих таблиц видно, что с ростом τ в интенсивности сейсмического процесса Южного полушария все большую часть оказывает влияние Северное полушарие. В то же время обратное влияние Южного полушария на Северное зависит от времени τ мало и остается примерно на уровне 10-15%.

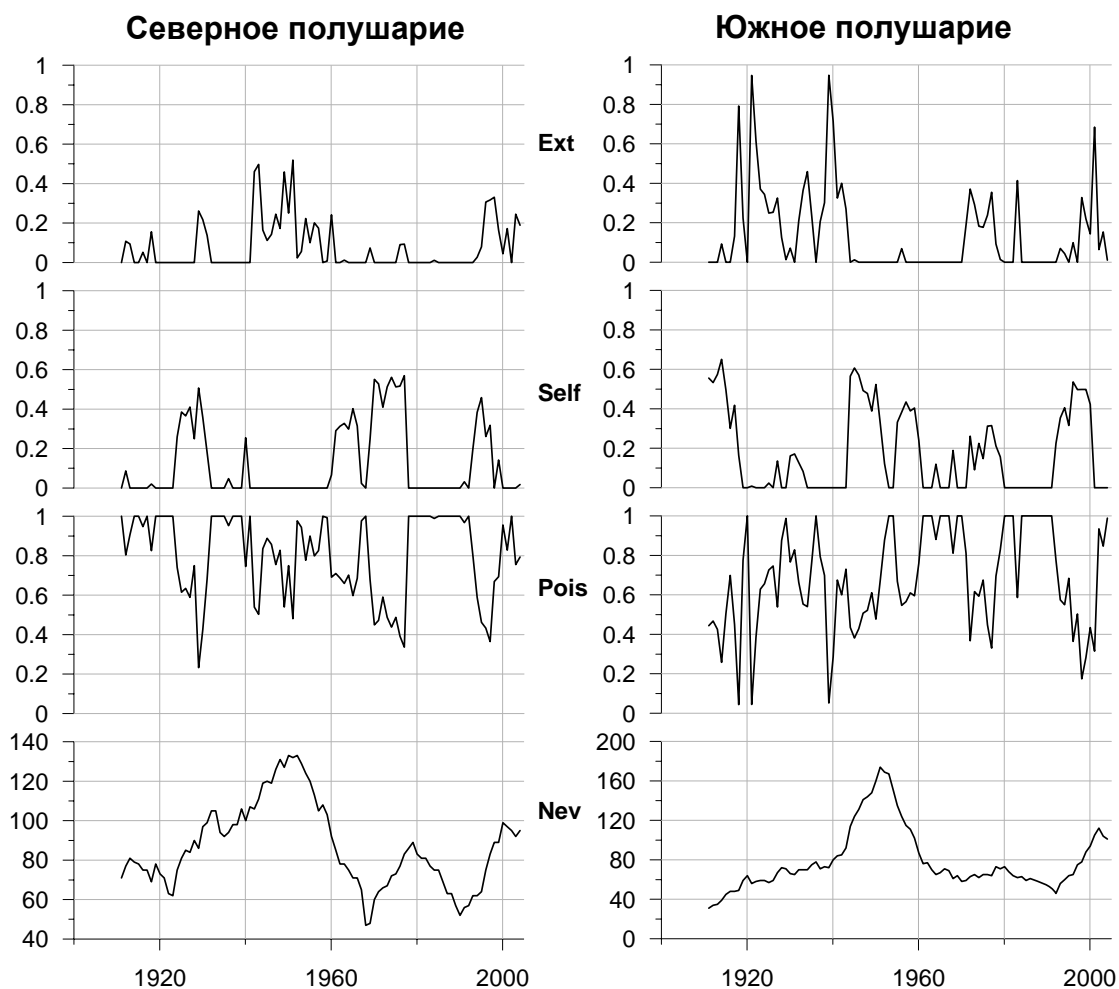


Рис.2. Оценка изменения долей интенсивности сейсмического процесса в скользящем временном окне длиной 10 лет со смещением 1 год для $\tau = 1$ год. Сверху вниз: доля внешнего влияния, доля самовозбуждения, пуассоновская часть и число событий в окне. Зависимости даны от правого конца скользящего временного окна.

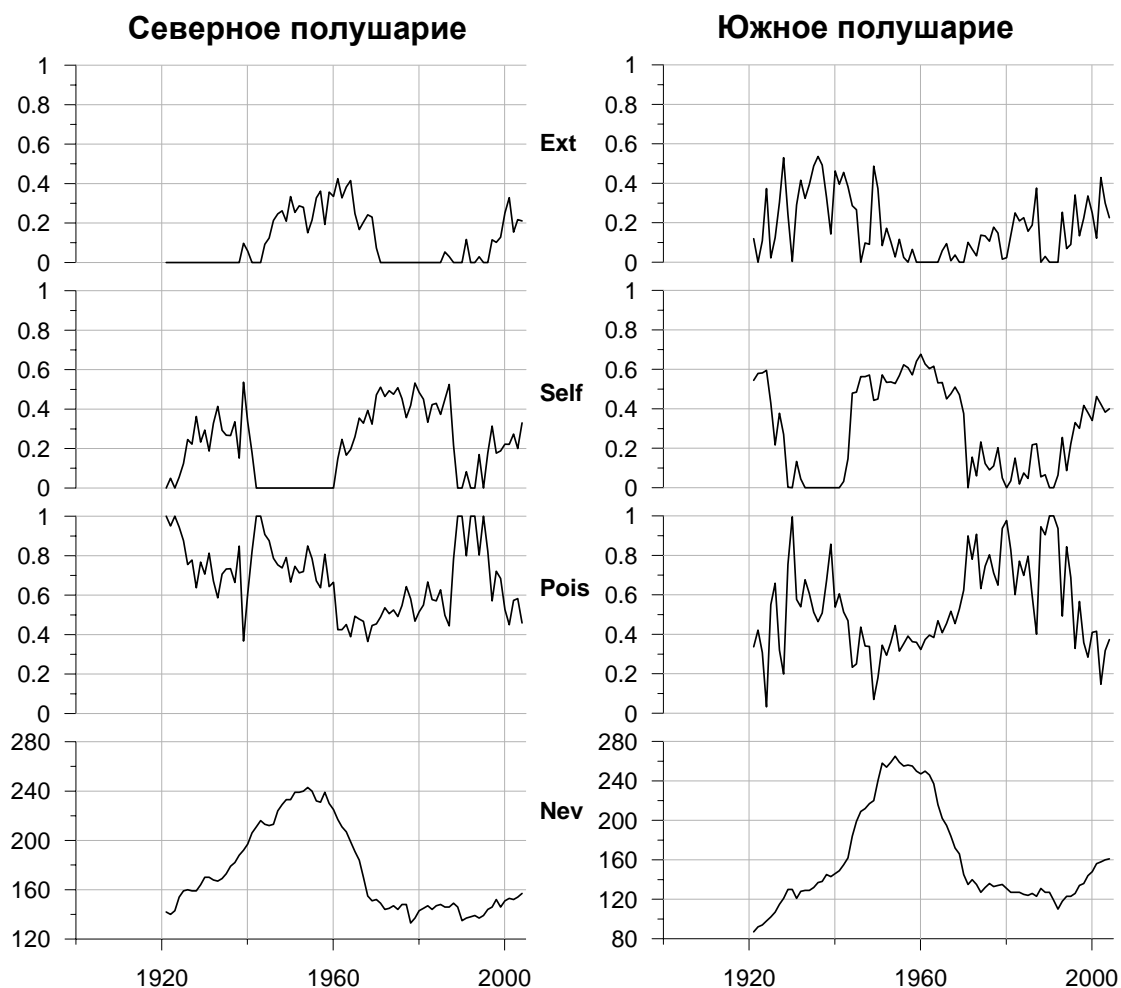


Рис.3. Оценка изменения долей интенсивности сейсмического процесса в скользящем временном окне длиной 20 лет со смещением 1 год для $\tau = 1$ год. Сверху вниз: доля внешнего влияния, доля самовозбуждения, пуассоновская часть и число событий в окне. Зависимости даны от правого конца скользящего временного окна.

На рис.2 и 3 приведены графики изменения частот интенсивности сейсмических процессов при их оценке в скользящем временном окне длиной 10 и 20 лет со смещением 1 год для параметра $\tau = 1$ год.

Пример 2: два климатических временных ряда.

В качестве второго примера рассмотрим задачу о связи моментов времени, соответствующих локальным максимумам двух климатических временных рядов: годовых колец роста арктической сосны в Швеции, 500-1980 [Briffa et al., 1992] и реконструкции средних годовых зимних температур по содержанию изотопов кислорода в ледовых кернах в Гренландии, 553-1973 [Dansgaard et al., 1975].

Оба временных ряда ниже рассматриваются на общем отрезке времени 553-1973. Для уменьшения влияния шума на определение точек локальных максимумов оба ряда предварительно сглаживались гауссовским ядром с радиусом сглаживания 2 года – см. программу **Spectra_Analyzer**. Далее в каждом временном ряду бралось по 50 моментов времени, соответствующих наибольшим локальным максимумам сглаженных кривых. Таким образом, получалась последовательность из одинакового числа «событий», отвечающих характерным точкам поведения исследуемых временных рядов. Исходные

данные, их сглаженные кривые и получившиеся две последовательности времен представлены на рис.4.

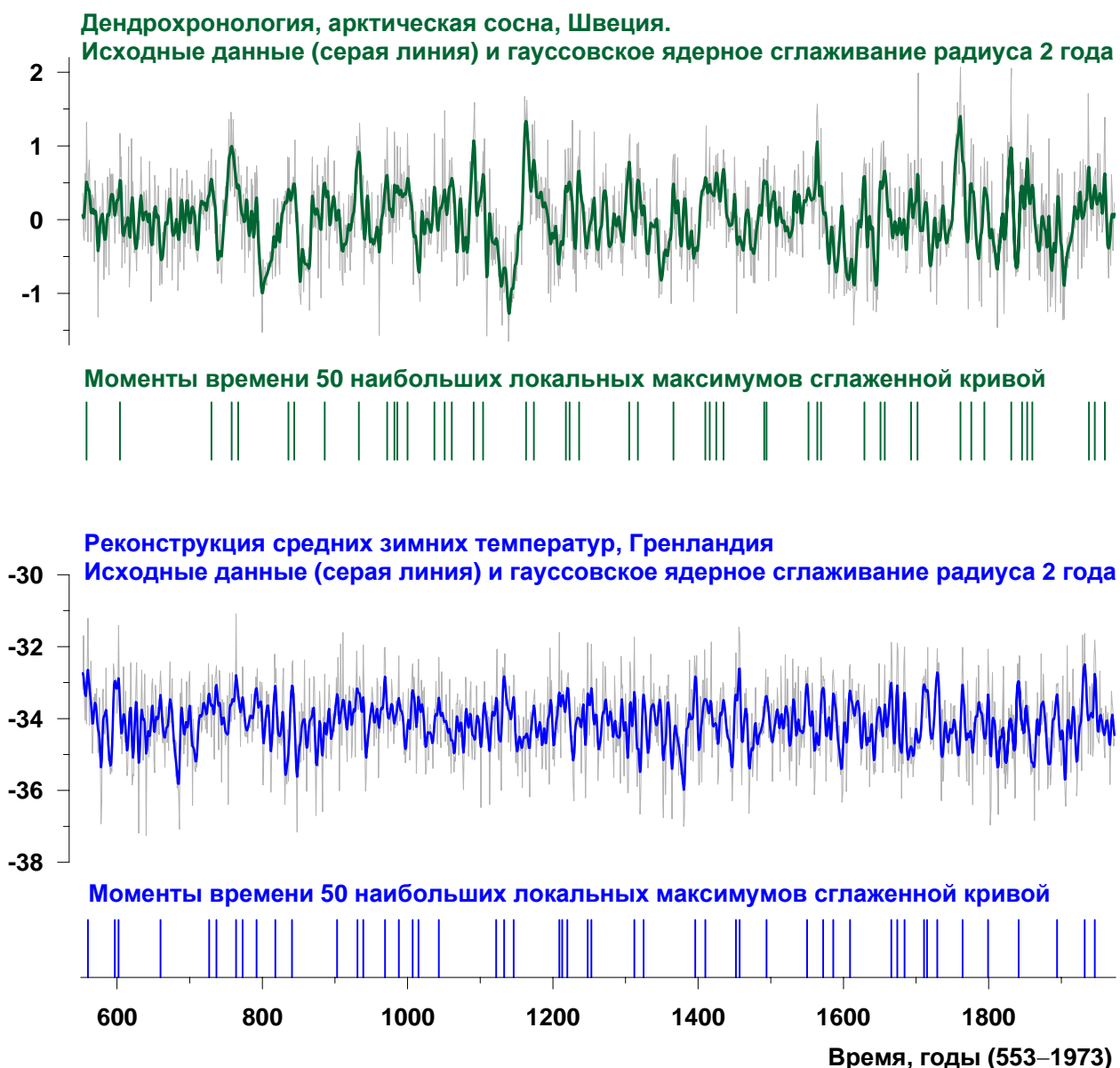


Рис.4. Графики 2-х климатических временных рядов для периода наблюдений 553-1973: дендрохронологии (кольца роста) арктической сосны в Швеции и реконструкции средних зимних температур по содержанию изотопов кислорода в ледовых ядрах в Гренландии. Вертикальными линиями обозначены последовательности «событий» - моментов времени, соответствующих 50 наибольшим локальным максимумам сглаженных кривых. Сглаживание производилось гауссовским ядром с радиусом 2 года.

Ниже представлены таблицы разложения средних интенсивностей последовательностей времени на 3 составляющих: случайную (пуассоновскую), самовозбуждающуюся и обусловленную влиянием другой последовательности для увеличивающихся значений характерного времени τ .

$\tau = 10$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	1.000	0.000	0.000
Гренландия	0.970	0.000	0.030

$\tau = 20$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	1.000	0.000	0.000
Гренландия	0.918	0.000	0.082

$\tau = 30$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	1.000	0.000	0.000
Гренландия	0.899	0.000	0.101

$\tau = 40$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	1.000	0.000	0.000
Гренландия	0.901	0.000	0.099

$\tau = 50$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	0.990	0.000	0.010
Гренландия	0.911	0.000	0.089

$\tau = 60$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	0.935	0.000	0.065
Гренландия	0.923	0.000	0.077

$\tau = 70$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	0.867	0.000	0.133
Гренландия	0.935	0.000	0.065

$\tau = 80$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	0.794	0.000	0.206
Гренландия	0.945	0.000	0.055

$\tau = 90$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	0.726	0.000	0.274
Гренландия	0.954	0.000	0.046

$\tau = 100$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	0.667	0.000	0.333
Гренландия	0.962	0.000	0.038

$\tau = 150$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	0.508	0.000	0.492
Гренландия	0.985	0.000	0.015

$\tau = 200$ лет	Пуассон. часть	Самовозбужд. часть	Внешняя часть
Швеция	0.463	0.000	0.537
Гренландия	0.997	0.000	0.003

Из этих таблиц следует, что последовательность характерных точек временного ряда в Гренландии является практически полностью пуассоновской для всех τ . Влияние последовательности характерных точек временного ряда в Швеции пренебрежимо мало. То же самое можно сказать и об обратном влиянии для малых значений τ от 10 до 60 лет. Однако при дальнейшем увеличении τ часть средней интенсивности «климатического процесса в Швеции» становится все более и более быть связанной с последовательностью характерных точек временного ряда в Гренландии – вплоть до его преобладания для $\tau = 200$ лет. Таким образом, на больших временных масштабах изменения климата в Гренландии происходят в среднем раньше, чем в Швеции и могут быть названы «причиной».

ЛИТЕРАТУРА.

- Любушин А.А., В.Ф.Писаренко.* (1993) Исследование сейсмического режима с помощью линейной модели интенсивности взаимодействующих точечных процессов. - Физика Земли. 1993. N12. С.81-87.
- Любушин А.А.* (2007) «Анализ данных систем геофизического и экологического мониторинга». М.: Наука, 2007, 228с.
- Briffa, K. R.; Jones, P. D.; Bartholin, T. S.; Eckstein, D.; Schweingruber, F. H.; Karlén, W.; Zetterberg, P.; Eronen, M.* (1992) Fennoscandian summers from ad 500: temperature changes on short and long timescales – Climate Dynamics, Volume 7, Issue 3, pp.111-119
- Cox D.R., Lewis P.A.W.* (1966) The statistical analysis of series of events. London, Methuen (Русский перевод: *Кокс Д., П.Льюис* (1969) Статистический анализ последовательностей событий. М., Мир. 312с.)
- Cox D.R.* (1975) Partial likelihood – Biometrika. Vol.62, pp.269-276.
- Dansgaard, W., Johnsen, S. J., Reeh, N., Gundestrup, N., Clausen, H. B., Hammer, C. U.* (1975) Climatic changes, Norsemen and modern man. Nature 255: 24-28.