

*ЗГУРОВСКИЙ М.З.,
БОЛДАК А.А.,
ЕФРЕМОВ К.В.,
СЕРГЕЕВА Н.А.,
ЗАБАРИНСКАЯ Л.П.,
ШЕСТОПАЛОВ И.П.,
НИСИЛЕВИЧ М.В.*

ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ЭМПИРИЧЕСКИХ ИССЛЕДОВАНИЙ ВЗАИМОСВЯЗИ ГЕЛИО- И ГЕОФИЗИЧЕСКИХ ПРОЦЕССОВ

В работе представлены результаты экспериментального исследования взаимосвязи гелио- и геофизических процессов с использованием методов интеллектуального анализа данных, а именно предложенного авторами метода вейвлет-разложения сигналов в неортогональных базисах.

The paper presents the results of an experimental study on the relationship of helio- and geophysical processes using data mining methods, namely the proposed method on the wavelet decomposition of signals in non-orthogonal bases.

Введение

Проблемы исследования сложных природных систем связаны с отсутствием их формальных описаний, сложностью наблюдения за ней и управления такой системой вследствие присутствия большого количества взаимодействующих второстепенных процессов, нестационарности исследуемых процессов и эволюции в целом исследуемых систем, невозможности проведения активных экспериментов [1].

Одной из таких задач является междисциплинарная задача исследования взаимосвязи гелио- и геофизических процессов, в частности активности Солнца и сейсмической активности Земли.

Существует достаточно обширный ряд публикаций, авторы которых устанавливают наличие связи между геоэффективными солнечными событиями и сейсмичностью Земли.

В ряде работ на простом эмпирическом анализе исторических данных или статистическими методами найдено неслучайное соотношение между солнечно-земным взаимодействием и возникновением землетрясений. Так установлено, что в период 11-летнего цикла солнечной активности наиболее высокая сейсмическая активность соответствует эпохам максимума и минимума этого цикла, когда уровень солнечной активности быстро и резко меняется [2-4]. В работах [5-7] сделан вывод о наличии корреляционной связи между вариациями сейсмиче-

ской активности и солнечной активностью. Исследования с применением спектрального анализа временных рядов сильных и слабых землетрясений и чисел Вольфа, характеризующих солнечную активность, методом максимальной энтропии показали, что в периоды повышенных 11-летних циклов солнечной активности повышается активность землетрясений в зонах сжатия Земли, а в периоды пониженной солнечной активности повышается активность землетрясений в зонах растяжения Земли [8].

Однако в работе [9] авторы, используя для статистической проверки гипотезы о существовании связи между солнечной и сейсмической активностью Хи-квадрат критерий и t-критерий Стьюдента, не нашли статистически значимых различий в распределениях глобального числа сильных землетрясений ($M \geq 7,5$) и числа групп солнечных пятен. То есть нельзя уверенно сказать о статистически значимых отношениях между геоэффективными событиями на Солнце и возникновением землетрясений. Но и отвергнуть их существование также нельзя.

Задача данного прикладного исследования состоит в проверке описанной выше гипотезы на основе применения методов интеллектуального анализа к данным, описывающим активность гелио- и геофизических процессов. В этой работе предлагается метод анализа причинно-следственных связей между процессами, основанный на неортогональном дискретном вейвлет-преобразовании.

Постановка задачи эмпирического исследования причинно-следственных связей между гелио- и геофизическими процессами

В качестве характеристики солнечной активности обычно используется ряд чисел Вольфа. Число Вольфа W вычисляется по формуле:

$$W = k(f + 10g),$$

где f – число пятен на солнечном диске; g – число групп, объединяющих эти пятна; k – нормировочный коэффициент, характеризующий условия наблюдения и применяемую аппаратуру.

Для настоящего исследования в качестве данных, характеризующих солнечную активность, был выбран ряд среднегодовых значений чисел Вольфа за период с 1890 г. по 2012 г [10].

Сейсмическую активность можно характеризовать суммарной сейсмической энергией, высвобожденной при крупных землетрясениях. Энергия землетрясений определяется (в эрг) по формулам [11]:

$$\lg E = 11,8 + 1,5M_s$$

для гипоцентров на глубине меньше 100 км и

$$\lg E = 5,8 + 2,4m_b$$

для гипоцентров на глубине больше 100 км, где E – энергия землетрясения, M_s – магнитуда по поверхностным волнам, m_b – магнитуда по объемным волнам.

Для расчета суммарных годовых значений сейсмической энергии землетрясений за период 1890 – 2012 гг. на основе каталогов Международного сейсмологического центра [12] и Национального центра информации о землетрясениях Геологической службы США [13] был составлен каталог землетрясений с магнитудами $M_s \geq 6$ и $m_b \geq 5,5$.

Для получения однородного ряда данных было проведено их сглаживание в окне шириной 7 лет по трем значениям. Аналогичная процедура предварительной обработки данных была применена к ряду чисел Вольфа. На рис.1 приведены исходные и сглаженные данные о солнечной активности в виде ряда чисел Вольфа R и сейсмической активности в виде выделяющейся энергии землетрясений E .

Ранее при изучении взаимосвязи сейсмической активности Земли и активности Солнца исследователями ГЦ РАН были выявлены веко-

вые циклы солнечной и сейсмической активности, продолжительностью около 100 лет [14-18].

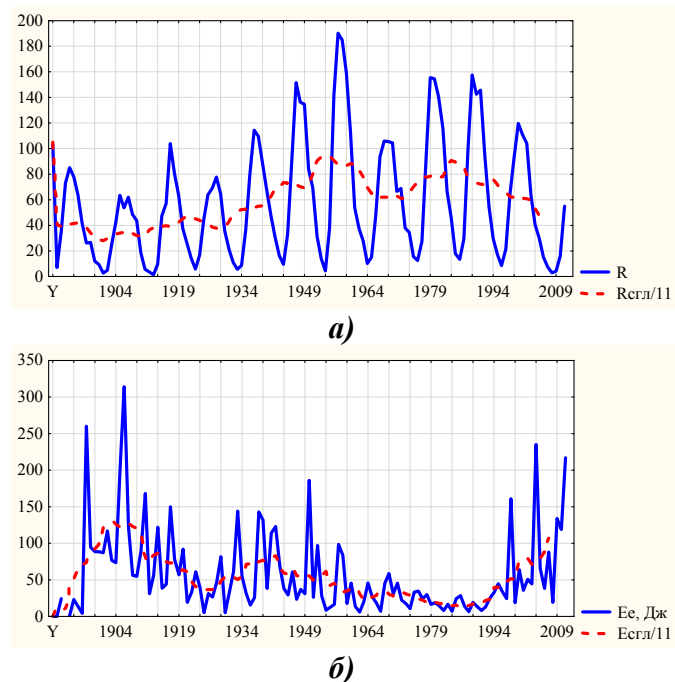


Рис.1. Данные, используемые в исследовании взаимосвязи гелио- и геофизических процессов: а) среднегодовые и сглаженные R_{sgl} значения чисел Вольфа; б) суммарные годовые E_e сглаженные $E_{сгл}$ значения сейсмической энергии.

Цель этого исследования состоит в проверке гипотезы о причинно-следственной связи между гелио- и геофизическими процессами на основе интеллектуального анализа охарактеризованных выше данных.

Метод анализа причинно-следственных связей на основе разложения сигналов в неортогональных базисах

Континуальный многомасштабный вейвлет-анализ [19] представляет собой довольно гибкий и, в то же время, весьма мощный инструмент временного и спектрального анализа, целью которого является построение частотно-временных характеристик сигналов, которые во многом проясняют природу не только сигналов в целом, но также и их составляющих.

Вейвлет-преобразование для непрерывного сигнала $x(t)$ относительно вейвлет-функции ψ определяется следующим образом [20]:

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-b}{a}\right) dt,$$

где ψ^* означает комплексное сопряжение к функции ψ , параметр $b \in R$ соответствует временно́му сдвигу и называется параметром положения, параметр $a > 0$ задает масштабирование и называется параметром растягивания.

Если определить нормированную функцию:

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right),$$

то вейвлет-преобразование можно записать в виде:

$$T(a,b) = \int_{-\infty}^{\infty} x(t) \psi_{a,b}^* dt.$$

Гибкость вейвлет-анализа достигается за счет применения различных функций ψ , и, следовательно, различных процедур синтеза вейвлет-спектра сигнала. В соответствии с тем, каким образом формируется базис разложения сигнала (система функций ψ), выделяют неортогональные, ортогональные и биортогональные вейвлет-преобразования. В зависимости от способа представления анализируемых сигналов принято выделять непрерывное и дискретное вейвлет-преобразования.

Предлагаемый метод анализа причинно-следственных связей между процессами основан на неортогональном дискретном вейвлет-преобразовании.

Модель сигналов. Пусть задан набор временных последовательностей $X = \{x_{1..n}\}$ и выдвигается гипотеза о существовании подмножества $X_E = \{x_m \mid m \subset 1..n\}$ последовательностей («причин»), формирующих исходную последовательность («следствие»), которая также входит в X . Необходимо разработать метод проверки такой гипотезы, учитывая то, что входящие последовательности могут присутствовать в исходной последовательности с произвольным сдвигом и масштабом. Также необходимо учитывать тот факт, что в причинно-следственной связи «следствие» не может предшествовать «причине» по времени.

Аддитивную модель сигнала, соответствующую проверяемой гипотезе, можно представить следующим образом:

$$y(t) = \sum_{i=1}^n w_i e_i(a,b,t), \quad (1)$$

где w_i – соответствует степени влияния сигнала $e_i(a,b,t)$ на результат, а сигналы $e_i(a,b,t)$ представляют собой масштабированные и смещенные сигналы $x_i \in X_E$:

$$e_i(a_i, b_i, t) = x_i\left(\frac{t-b_i}{a_i}\right).$$

Модель (1) является двухуровневой: внутренняя модель описывает преобразования каждого отдельного сигнала x_i , а внешняя – формирование их аддитивной смеси.

Таким образом, для разложения сигнала «следствия» в базисе сигналов-«причин» необходимо найти такие значения сдвигов b_i и масштабов a_i сигналов x_i , которые максимизируют w_i для всех i .

Определение параметров масштаба и сдвига. Рассмотрим сначала случай, когда сигналы заданы непрерывной моделью.

Пусть задана некоторая функция $f(z)$, которая формирует последовательность значений выборки x , тогда функция $F(a,b,z)$, которая задает масштабирование и сдвиг по z , принимает следующий вид:

$$F(z) = f\left(\frac{z-b}{a}\right)$$

Учитывая, что результатом вейвлет-анализа является поверхность, точки которой указывают на «сходство» между вейвлет-функцией и сигналом в заданном интервале, можно использовать в качестве вейвлет-функции одну из тех, которые описывают входящие последовательности.

Применим формулу вейвлет-преобразования, заменив дискретные последовательности соответствующими им сигналами с помощью одного из видов интерполяции:

$$\begin{aligned} f(a_k, b_k) &= \sum_{i=1}^N \int_{-\infty}^{\infty} w_i x_i\left(\frac{t-b_i}{a_i}\right) x_k\left(\frac{t-b_k}{a_k}\right) dt = \\ &= w_k \int_{-\infty}^{\infty} w_i x_k\left(\frac{t-\bar{b}_k}{a_k}\right) x_k\left(\frac{t-b_k}{a_k}\right) dt + \\ &+ w_i \sum_{i=1, i \neq k}^N \int_{-\infty}^{\infty} x_i\left(\frac{t-b_i}{a_i}\right) x_k\left(\frac{t-b_k}{a_k}\right) dt \end{aligned} \quad (2)$$

В формуле (2) значения \bar{a}, \bar{b} соответствуют тем значениям сдвига и масштаба, с которыми последовательность x_k входит в модель (1).

Первое слагаемое в формуле (2) является автокорреляционной функцией [21], значение которой будет максимальным при условии максимальной «похожести» сигналов $e_k(a_k, b_k, t)$ и $e_k(\bar{a}_k, \bar{b}_k, t)$:

$$\max_k(f(a_k, b_k)) = f(\bar{a}_k, \bar{b}_k).$$

Значения функции (2) можно приближенно вычислить с помощью одной из формул численного интегрирования [22]. В случае применения формулы прямоугольников имеем:

$$f(a_k, b_k) = \sum_{i=1}^N \int_{-\infty}^{\infty} w_i x_i \left(\frac{t-b_i}{a_i}\right) x_k \left(\frac{t-b_k}{a_k}\right) dt \approx \sum_{i=1}^N \sum_{t=t_{\min}}^{t_{\max}} w_i x_i \left(\frac{t-b_i}{a_i}\right) x_k \left(\frac{t-b_k}{a_k}\right) \Delta t \quad (3)$$

где $(t_{\min}; t_{\max})$ – пределы интегрирования, соответствующие области, на которой заданы значения последовательности x_k ; Δt – шаг интегрирования.

Для приведения значений функции f к диапазону $(-1; 1)$ необходимо нормировать значения, вычисленные по формуле (3), разделив их на произведение энергий функций y_i и x_k :

$$f(a_k, b_k)_{norm} = \frac{\sum_{t=t_{\min}}^{t_{\max}} (y_i x_k \left(\frac{t-b_k}{a_k}\right) \Delta t)}{\sum_{t=t_{\min}}^{t_{\max}} (y_i^2 \Delta t) \sum_{t=t_{\min}}^{t_{\max}} ((x_k \left(\frac{t-b_k}{a_k}\right))^2 \Delta t)} \quad (4)$$

Задача вычисления значений w_k , определяющих степень влияния сигналов x_k на результат y , может быть решена с помощью реализации последовательной схемы поиска компонент.

На первом шаге по формуле (4) для каждого сигнала $x_k \in X_E$ осуществляется поиск параметров сдвига и масштаба, максимизирующих w_k . На втором шаге выполняется проверка достоверности значений сдвига и масштаба, которая заключается в нахождении коэффициента w_k с помощью линейной регрессии. На третьем шаге осуществляется вычитание сигнала $e_k(a_k, b_k, t)$ из y . При этом x_k остается в списке исходных последовательностей-«причин», благодаря чему можно обнаружить такие эффекты влияния как «эхо».

Применение предложенного метода для анализа периодических сигналов требует уточнения, связанного с выбором и согласованием размеров временных окон, в пределах которых исследуются сигналы. Окном будем называть отрезок входного сигнала, который включает в себя ровно один период. Теперь задача сравнения последовательностей будет заключаться в сравнении окна входного сигнала с применением к нему соответствующих преобразований с окном выходного сигнала.

Для определения размера окна могут быть применены средства спектрального анализа на основе преобразования Фурье [23]. Однако, в том случае, когда сигнал плохо аппроксимируется гармоническими функциями, такой способ определения размера окна может давать ошибочный результат. Решение задачи определения размера окна для таких сигналов может основываться на анализе их вейвлет-разложения в ортонормированном базисе вейвлет-функций.

Результаты экспериментальных исследований

Предложенный метод реализован в системе обеспечения жизненного цикла данных для научных исследований, разрабатываемой Мировым центром данных «Геоинформатика и устойчивое развитие» [24]. На рис.2 представлено описание этого сценария на языке [25], используемом в этой системе. Такой сценарий представляет собой совокупность взаимосвязанных процессов обработки данных, связь между которыми осуществляется с помощью портов. В представленном сценарии используются компоненты, предназначенные для импорта исходных данных, для экспорта результатов анализа причинно-следственных связей, для выполнения вейвлет-преобразований и преобразований Фурье, для интерполяции и сглаживания данных, а также компонента для анализа взаимосвязи двух исходных сигналов.

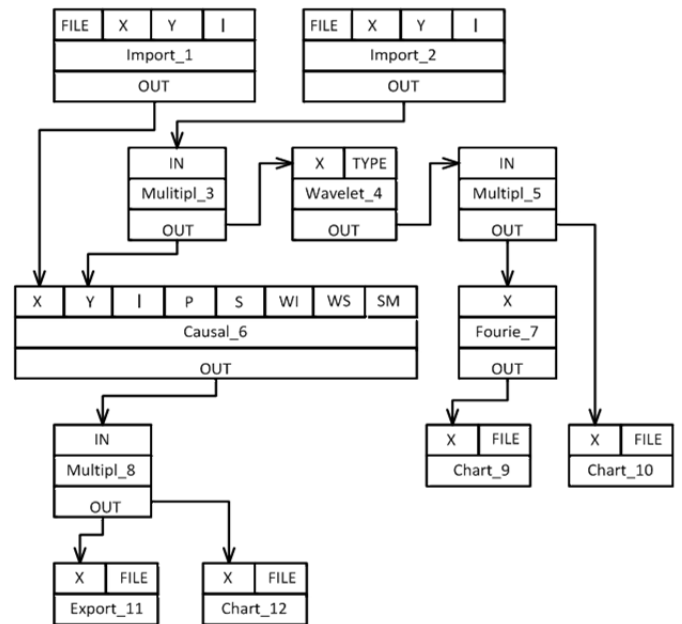
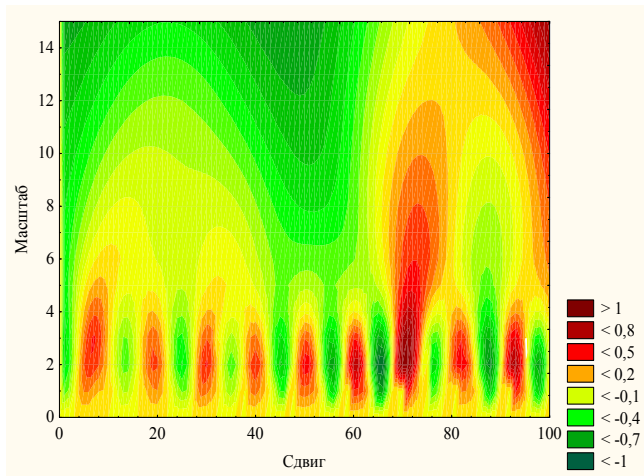


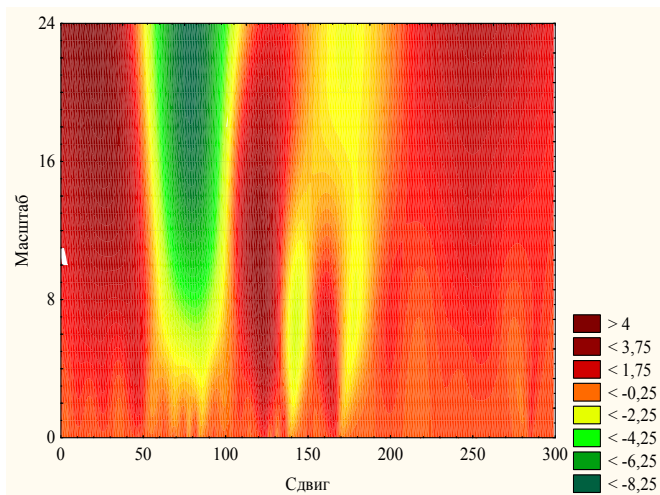
Рис.2. Описание сценария обработки данных для решения задачи анализа взаимосвязи процессов.

Результаты, полученные с помощью разработанных программных средств, представлены ниже.

Как видно из рис.3,а, на котором отображены результаты вейвлет-анализа для достоверного ряда чисел Вольфа, отчетливо прослеживаются известные 11-летние солнечные циклы.



а)



б)

Рис.3. Анализ ряда солнечной активности на периодичность: а) достоверный ряд чисел Вольфа за период 1890-2013 гг.; б) ряд чисел Вольфа за период 1705-2011 гг.

На рис.3,б можно также наблюдать периодичность и на больших масштабах, в частности, циклы длительностью 70-100 лет, что соответствует вековому «циклу Гляйсберга».

Как видно из рис.4, аналогично солнечной активности, для сейсмической активности характерна цикличность. Расстояние между двумя первыми максимумами равно около 100 лет. Третий максимум менее достоверный, потому, что третий период еще не закончился и вейвлет-функция не полностью пересекается с этим периодом, вследствие чего возникает краевой эффект.

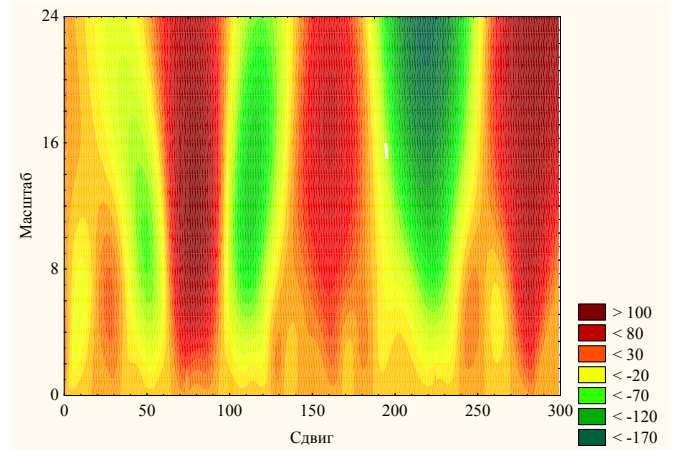


Рис.4. Результат вейвлет-анализа ряда сейсмической активности 1705-2013 гг. на периодичность.

В таблице 1 приведены результаты анализа взаимосвязи гелио- и геофизических процессов. Проанализирована связь между 11-летними циклами солнечной активности и сейсмической активности для достоверных рядов (1890-2013 гг.). Размер окна равен 11 годам, диапазон смещений принимает значение в пределах от 0 до 20 лет с шагом в один год, диапазон масштабов – от 1 до 3 с шагом 0,1.

Таблица 1.

Результаты анализа взаимосвязи процессов

Максимальное сходство	Масштаб	Сдвиг
0,85	1	5
0,804	1,1	0
0,792	1	0
0,813	2,4	4
0,888	1,3	1
0,779	1	3
0,874	1	0
0,926	1,7	0
0,85	1,9	19
0,764	1,3	8

Используя данные таблицы 1, можно дать графическую интерпретацию полученных результатов, изображенную на рис.6.

Как видно из рис.5, во многих случаях максимумам солнечной активности соответствуют максимумы сейсмической активности. Однако также существует много несоответствий.

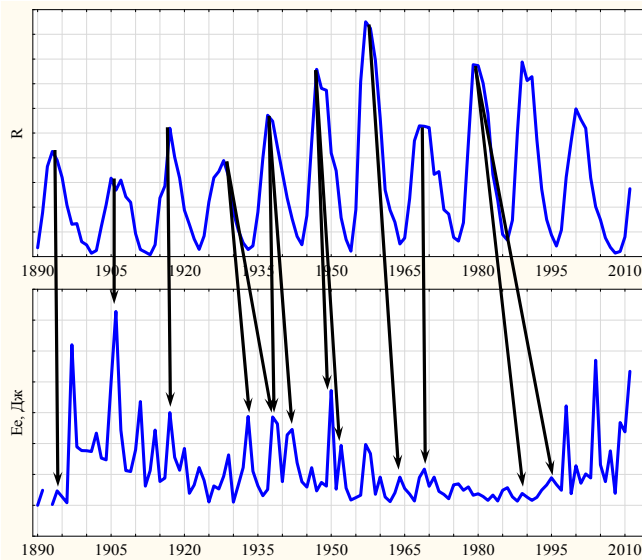


Рис.5. Графическая интерпретация результатов анализа.

Кроме 11-летних циклов также были проанализированы сглаженные по 7 значениям ряды солнечной и сейсмической активности за 1705-2011 гг. Обнаружены похожие фрагменты при значениях сдвига в 20 лет и со значением масштаба 2.2. Эти фрагменты приведены на рис.7:

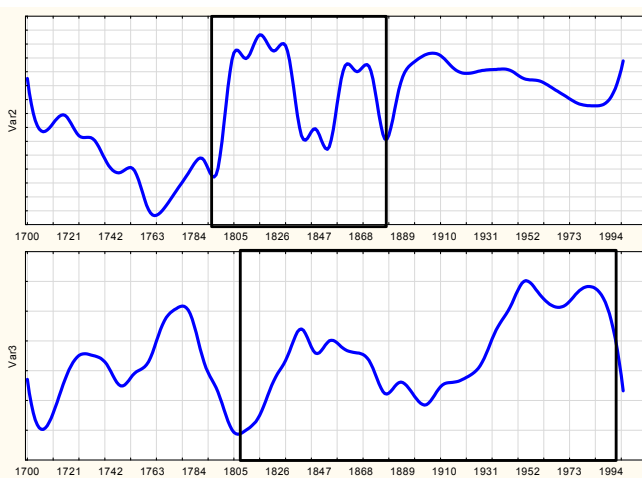


Рис.7. Динамика выделения сейсмической энергии и развития солнечной активности.

Несмотря на то, что замеченное сходство достаточно интересно, оно не позволяет делать однозначные выводы о наличии причинно-

следственной связи. В то же время это может свидетельствовать о каких-то общих чертах процессов, которые лежат в основе исследуемых явлений.

Выводы

В результате применения разработанного авторами метода анализа причинно-следственных связей между процессами, основанного на неортогональном дискретном вейвлет-преобразовании, для проверки гипотезы о наличии связи между геоэффективными солнечными событиями и сейсмичностью Земли можно сделать следующие заключения:

Предложенный метод позволил выделить 11-летние циклы солнечной активности на достоверном ряде чисел Вольфа за период 1890-2013 гг. и на ряде чисел Вольфа за период 1705-2011 гг. На ряде чисел Вольфа за период 1705-2011 гг. выделены периоды длительностью 70-100 лет, что соответствует вековому «циклу Гляйсберга».

Анализ показывает, что нельзя говорить об 11-летней периодичности сейсмичности Земли на ряде значений выделившейся сейсмической энергии за период 1890-2013 гг. Установлено, что для сейсмической активности за период 1705-2011 гг. также характерна цикличность с периодом около 100 лет и по времени не совпадающая с цикличностью активности Солнца.

При анализе сглаженных по 7 значениям рядов солнечной и сейсмической активности за 1705-2011 гг. обнаружены похожие фрагменты при значениях сдвига в 20 лет и со значением масштаба 2.2. Но выявленное сходство не позволяет делать однозначные выводы о наличии причинно-следственной связи между геоэффективными солнечными событиями и сейсмической активностью Земли.

Работа выполнена при финансовой поддержке Национальной академии наук Украины (2268/12, 2268/13) и Российского фонда фундаментальных исследований (грант № 12-01-90418-Укр_a).

Список литературы

1. Растринг Л. А. Адаптация сложных систем. — Рига: Зинатне, 1981. — 375 с.
2. Сытинский А.Д. Зависимость сейсмичности Земли от процессов на Солнце, в межпланетной среде и в атмосфере. В кн. Атлас временных вариаций природных антропогенных и социальных процессов. Том 2. М.: Научный Мир, 1998, с.70-72.
3. Одинцов С.Д., Иванов-Холодный Г.С., Георгиева К. Солнечная активность и глобальная сейсмичность Земли // Известия РАН. Серия физическая, 2007. Т. 71. № 4. С. 608-610.

4. Simpson I.F. Solar activity as a triggering mechanism for earthquakes // *Earth and Planet. Sci. letter*, 1968. V. 3. № 5. P. 417-425.
5. Ананьин И.В., Фаддеев А.О. О возможных причинах корреляций между изменениями величин сейсмической активности и средними годовыми температурами на поверхности Земли. В кн. Атлас Временных вариаций природных, антропогенных и социальных процессов. Том 3. М., Янус-К, 2002, с.222-224.
6. Шестопапов И.П., Рогожин Ю.А. Корреляция между микробиологической и сейсмической активностью с учетом взаимосвязей «Солнце-Земля» и генерации нейтронных потоков // *Авиакосмическая и экологическая медицина*. 2005. Т. 39. № 3. С. 20-26.
7. Gousheva, M.N., Georgiva, K.Y., Kirov, B.B., Atanssov, D. On the relation between solar activity and seismicity // *Proceedings of International Conference on Recent Advances in Space Technologies*, 20-22 November 2003.
8. Хаин В.Е., Халилов Э.Н. Пространственно-временные закономерности сейсмической
9. и вулканической активности. Bargas, SWB, 2008. 304 с. ISBN 978-9952-451-00-9.
10. SIDC – Solar Influences Data Analysis Center [Электронный ресурс] – 2013 –Режим доступа: <http://sidc.oma.be/>.
11. Касахара К. Механика землетрясений. Пер. с англ. Москва. Мир. 1985. 264 с.
12. International Seismological Centre [Электронный ресурс] – 2013 –Режим доступа: <http://www.isc.uk>.
13. National Earthquake Information Center – NEIC [Электронный ресурс] – 2013 –Режим доступа: <http://neic.usgs.gov>.
14. Соболев Г.А., Шестопапов И.П., Харин Е.П. Геоэффективные солнечные вспышки и сейсмическая активность Земли // *Физика Земли*. 1998. № 7. С. 85-89.
15. Шестопапов И.П., Харин Е.П. Изменчивость во времени связей сейсмичности Земли с циклами солнечной активности различной длительности // *Геофизический журнал*. 2006. Т. 28. № 4. С. 59-70.
16. Шестопапов И.П., Харин Е.П. Сейсмическая активность Земли и процессы в межпланетной среде // IV Международная конференция “Солнечно-земные связи и предвестники землетрясений”: с. Паратунка Камч. обл. 14-17 авг. 2007 г., сборник докладов. Петропавловск-Камч.: ИКИР ДВО РАН, 2007. С.476-481.
17. Белов С.В., Шестопапов И.П., Харин Е.П. О взаимосвязях эндогенной активности Земли с солнечной и геомагнитной активностью // *Доклады Академии наук*. 2009. Т. 428. № 1. С. 104-108.
18. Харин Е.П., Белов С.В., Шестопапов И.П. Пространственно-временные изменения сейсмичности Земли и солнечная активность // V Международная конференция “Солнечно-земные связи и физика предвестников землетрясений”, с. Паратунка, Камчатский край. 2-7 авг. 2010 г., сборник докладов. Петропавловск-Камчатский: ИКИР ДВО РАН, 2010. С. 470-473.
19. B. Boashash. *Time-Frequency Signal Analysis and Processing – A Comprehensive Reference [Text]* / B. Boashash. Oxford, UK: Elsevier Science, 2003.
20. Raghuvеer M. *Wavelet Transforms: Introduction to Theory & Applications [Text]* / Raghuvеer M. Rao, Ajit S. Ворардикар. Upper Saddle River, New Jersey, US: Prentice Hall PTR, 2008. – 336 p.
21. Орлов А.И. *Прикладная статистика [Текст]* / Орлов А.И. – М.: Издательство «Экзамен», 2011. – 425 с.
22. Каханер Д. *Численные методы и программное обеспечение [Text]* / Каханер Д., Моулер К., Нэш С. – М.: Мир, 2011. – 575 с. – ISBN 5-03-003392-0.
23. М. А. Павлейно. *Спектральные преобразования в MatLab. [Text]* / М. А. Павлейно, В. М. Ромаданов. – СПб., 2007. – 160 с. – ISBN 978-5-983-40121-1.
24. World Data Center for Geoinformatics and Sustainable Development [Электронный ресурс] – 2013 –Режим доступа: <http://wdc.org.ua>.
25. Болдак А.О. , Єфремов К.В. Предметно-орієнтована мова аналітичної обробки даних[Текст] // Вісник НТУУ "КПІ". Інформатика, управління та обчислювальна техніка: Зб. наук. пр. - К.: Век+, - 2012. - № 55. - с.46-54.