

## Сингулярный спектральный анализ в солнечно-земной физике

Ожередов В.А.<sup>1</sup>, Бреус Т.К.<sup>1</sup>, Обридко В.Н.<sup>2</sup>

<sup>1</sup>Учреждение Российской академии наук Институт космических исследований (ИКИ РАН) 117997 Москва, Профсоюзная 84/32, Россия [ojymail@mail.ru](mailto:ojymail@mail.ru)

<sup>2</sup>Учреждение Российской академии наук Институт земного магнетизма, ионосферы и распространения радиоволн (ИЗМИРАН) имени Н.В. Пушкова, Троицк, Московской области

В работе прогнозируются 23-й и 24-й циклы солнечной активности и найдены аналоги спектральных компонент гелиогеофизических и медико-биологических длинных (порядка 26 лет ежедневных измерений) рядов в высокочастотной области спектра. Использован метод расширенного спектрального анализа, позволяющий получать разреженный спектр с узкими пиками, лишенный искажений, вносимых эпизодически появляющимися нестационарными выбросами.

### Введение

Целью настоящей работы является: 1) демонстрация прогностических способностей расширенного метода сингулярного спектрального анализа SSA на примере предсказания 23-го и 24-го циклов солнечной активности, и 2) иллюстрация применения расширенной версии SSA для сжатия спектральной информации о временном ряде, которое позволяет получать спектры с ограниченным набором разреженных узких пиков. Это обстоятельство предоставляет возможность поиска правдоподобных аналогов спектральных компонент гелиогеофизических и медико-биологических рядов.

### Материалы

Анализировались ежесуточные данные по смертности от инфаркта миокарда, полученные в Миннесоте за 26 лет в период с 1968 по 1996 гг. (10593 отсчета) и любезно предоставленные проф. Ф.Халбергом. Данные о ежесуточных значениях Кр - индекса и чисел Вольфа брались с Интернет сайтов <http://www.sec.noaa.gov/ftpmenu/lists/geomag>.

### Методы

Для создания модели идеального сигнала используется процедура фильтрации исходных временных рядов на основе концепции линейной оболочки, являющейся логическим продолжением и развитием SSA-методики [1], кратко описанной ниже.

Рассмотрим всевозможные последовательности из  $N$  отсчетов временного ряда, называемые в дальнейшем реализациями, полученные в результате последовательных сдвигов временного окна на один интервал дискретизации вперед. К числу  $N$  предъявляется единственное требование – оно должно быть существенно большим, чем максимальный период в получаемом нами спектре. Последовательные реализации образуют траекторию в  $N$ -мерном пространстве. Метод SSA позволяет получить информацию о топологии линейного многообразия (гиперплоскости) некоторой заранее заданной размерности  $Q \ll N$ , на которую оптимальным (в смысле квадратичной нормы) образом проецируется траектория  $N$ -мерных реализаций. Эту  $Q$ -мерную гиперплоскость условимся называть *линейной*

*оболочкой сигнала*. Проекцию траектории на линейную оболочку по известному из теории SSA алгоритму можно снова свернуть во временной ряд, который мы будем называть *сигналом*, порожденным *линейной оболочкой*. Метод получения такого сигнала мы будем называть *фильтрацией в рамках линейной оболочки*.

Следующая доказанная нами теорема о сжатии спектральной информации устанавливает связь размерности линейной оболочки и степени разреженности спектра порождаемого ею сигнала:

Временной ряд любой длины, порождаемый линейной оболочкой размерности  $Q$ , принадлежит семейству ранга  $Q + 1$ . Следовательно, для него существует разложение, при котором количество спектральных пиков не превосходит увеличенную на единицу размерность линейной оболочки.

Концепция линейной оболочки – это «идеализация», позволяющая придать «глобальный» характер найденным у отрезка временного ряда закономерностям. Поскольку какие-либо априорные предположения о модели мешающей шумовой компоненты изначально отсутствуют, то вполне разумно допустить, что расположение  $N$ -мерных реализаций по отношению к линейной оболочке сигнала может изменяться нестационарно (из-за шума). Отсюда следует, что факт возможного наличия нестационарных выбросов следует учитывать при вычислении топологии линейной оболочки.

В связи с вышеизложенным, мы предлагаем процедуру поиска линейной оболочки, сопряженную с нахождением собственного базиса т.н. *самосогласованной ковариационной матрицы*, позволяющей учитывать выбросы и как бы приписывать им малые веса в формировании топологических характеристик линейной оболочки.

**Ковариация** может быть определена как мера (линейного) сходства между  $M$ -мерными последовательностями [2]. Аналогично, *взвешенная ковариация* может быть определена как мера линейного сходства во взвешенной норме, т.е. если мы имеем два ряда  $\{\xi^i\}_{i=1}^M$  и  $\{\eta^i\}_{i=1}^M$ , то  $\text{cov}_w \{\xi, \eta\}$  определяет ( $w_i$  – веса)

$$F \equiv \min_{a,b} \frac{1}{\sum_{i=1}^M w_i} \sum_{i=1}^M w_i (a\xi^i + b - \eta^i)^2 \quad (1)$$

Последовательное дифференцирование (1) по  $a$  и  $b$  и приравнивание частных производных  $\frac{\partial F}{\partial a}$  и  $\frac{\partial F}{\partial b}$  к нулю дает

$$F = D_w \{ \eta \} \left( 1 - \left( \frac{\text{cov}_w \{ \xi, \eta \}}{\sqrt{D_w \{ \xi \} D_w \{ \eta \}}} \right)^2 \right),$$

где

$$\text{cov}_w \{ \xi, \eta \} \equiv \frac{1}{\sum_{i=1}^M w_i} \sum_{i=1}^M w_i \xi^i \eta^i - \left( \frac{1}{\sum_{i=1}^M w_i} \sum_{i=1}^M w_i \xi^i \right) \left( \frac{1}{\sum_{i=1}^M w_i} \sum_{i=1}^M w_i \eta^i \right) \quad (2)$$

и  $D_w \{ \xi \} \equiv \text{cov}_w \{ \xi, \xi \}$ .

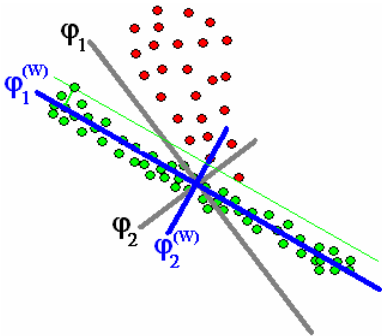


Рис.1. Кластеры регулярных реализаций (зеленые (светлые) точки) и выбросов (красные-темные точки).

Диагональные элементы взвешенной ковариационной матрицы (взвешенные дисперсии) определяют среднеквадратичный разброс проекции реализаций на базисные направления с учетом весов, приданных отдельным реализациям.

Рассмотрим, например, двумерную схему расположения реализаций ( $N = 2$ ), изображенную на Рис.1. Здесь отчетливо видны группы точек – концов векторов реализаций, образующие в пространстве кластеры. Наиболее многочисленная группа точек, обозначенных на рисунке зеленым (серым) цветом, значительно растянута вдоль прямой линии. Представляемые этими точками реализации мы условимся считать *регулярными*, а сами точки – образующими *кластер I*. Кластер красных (темных) точек – *кластер II* (реализации-выбросы) – образует как бы «вырост» из первого кластера. Визуальный анализ Рис.1 показывает, что базисным направлением линейной оболочки, в которую укладывается траектория регулярных реализаций, является длинная синяя (черная)

линия, обозначенная как  $\varphi_1^{(w)}$ , в то время как вычисление собственных векторов тривиальной ковариационной матрицы дает направление, помеченное как  $\varphi_1$  (без верхнего индекса).

Предлагаемый нами итеративный алгоритм поиска самосогласованной матрицы формально записывается как

$$\sigma_{W_{i+1}} = \sigma_W [\sigma_{W_i}], \quad (3)$$

т.е. на каждой последующей итерации вычисляется взвешенная ковариационная матрица, отвечающая весам как функциям расстояния до линейной оболочки, порождаемой предыдущей матрицей. При этом исходно веса всех реализаций равны единице, т.е. в стартовой точке алгоритма ковариационная матрица тривиальна. При наличии кластера регулярных реализаций данная процедура сходится, и предел последовательности

$$\hat{\sigma}_W \equiv \lim_{i \rightarrow \infty} \sigma_{W_i}$$

является самосогласованной ковариационной матрицей, поскольку порождаемая ею весовая последовательность в соответствии с формулой (2) формирует саму эту матрицу.

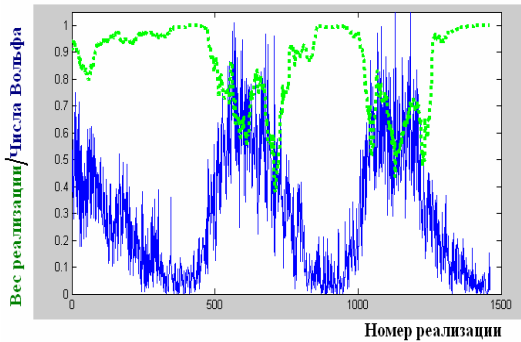
Если учесть, что *характерным расстоянием* в контексте критерия каменистой осыпи Кеттела является  $\sqrt{\lambda_Q}$ , т.е. корень квадратный из  $Q$ -го собственного значения (обычной) ковариационной матрицы, то вполне допустим следующий явный вид зависимости  $w_i$  от расстояния  $r_i$  между реализацией и линейной оболочкой:

$$w_i = \exp \left( - \frac{r_i^4}{\lambda_Q^2} \right) \quad (4)$$

Точка перегиба  $y$  (4) приходится на значение  $\sqrt[4]{\frac{3}{4}} \sqrt{\lambda_Q} \approx 0.93 \sqrt{\lambda_Q} \sim \sqrt{\lambda_Q}$ , т.е. по порядку величины – на характерное по критерию Кеттела расстояние. Описанная процедура приводит к тому, что на каждом шаге собственный базис взвешенной ковариационной матрицы как бы «разворачивается» на некоторый угол, так что обновленная линейная оболочка становится все ближе к регулярным реализациям и все дальше от выбросов, мы назвали этот алгоритм *методом итеративного вращения собственного базиса*. По существу, этот метод является обобщением принятой в математической обработке наблюдений процедуры, в которой точки, отстоящие от среднего значения более, чем на 3 сигмы, выбрасываются и среднее значение и средняя квадратическая ошибка перевычисляется.

В качестве примера временного ряда, к которому применяется алгоритм итеративного разворота базиса, рассмотрим последовательность суточных чисел Вольфа ( $W_n$ -индексов) за период 26 лет (с 1968 по 1996 гг.). На Рис.2 сплошной синей (темной) линией показан сам график временного ряда, а пунктирной зеленой (светлой) – конечные веса реализаций, полученные после разворота базиса.

Из **Рис.2** следует, что выбросы, вероятнее всего, появляются в максимумах солнечной активности. Этого результата следовало ожидать, поскольку количество пятен в группах зависит от фазы солнечного цикла: в максимуме группы значительно больше. Поэтому



**Рис.2.** Веса реализаций  $W_n$ -индекса, соответствующие согласованной с топологией ЛО ковариационной матрице (зеленая пунктирная линия). Синей сплошной линией изображена солнечная активность ( $W_n$ -индекс)

появление одной–двух групп на диске меняет число Вольфа на несколько десятков, а иногда и более сотни, единиц. Этот эффект особенно значим при использовании суточных значений числа пятен. В минимуме этот эффект значительно ниже. В среднемесячных значениях эффект выбросов уменьшается, но все равно остается сильнее в максимуме цикла. Заметим, что по смыслу задачи мы обсуждаем здесь полные величины отклонений от сглаженных кривых. Естественно отношение сигнал-шум в минимуме будет значительно выше.

## Результаты

### 1) Прогноз солнечной активности в 23-м и 24-м циклах.

Используя отфильтрованный в рамках линейной оболочки временной ряд чисел Вольфа за период 1860 – 1996гг. в качестве обучающего материала, мы построили линейно-регрессионный прогноз этого временного ряда на два цикла вперед. Результат показан на **Рис.3**.

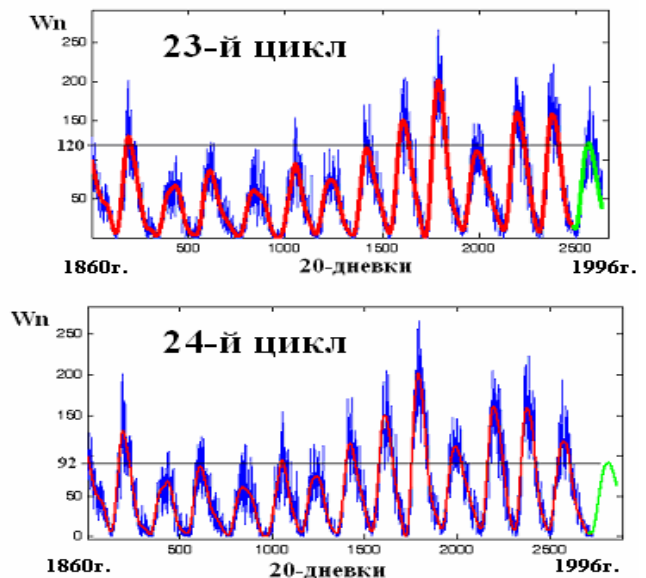
Различные методы прогноза «сглаженного» хода 23-его цикла давали дату наступления пика в районе марта 2000г.  $\pm 15$  месяцев, а величину пика – от 160 до 200. Наш прогноз дает дату наступления максимума в мае 2000г., а высоту пика – 120. Реальные дата наступления максимума 23-его цикла и его величина – апрель 2000г. и 120,8.

Различные методы прогноза «сглаженного» хода 24-его цикла дают дату наступления пика в районе второй половины 2011г. – первого квартала 2012г., а величину пика – от 80 до 120. Наш прогноз дает дату наступления максимума в апреле 2011г., а высоту пика – 92. В работах [3,4] проанализировано большое количество (несколько десятков) различных прогнозов 24 цикла. Несмотря на очень большой разброс результатов авторы [4] пришли к выводу о том, что наиболее вероятное значение высоты максимума 24 цикла составляет 90-100. Более того, в [5] проанализированы некоторые аномалии различных индексов солнечной активности и показано, что они

возможно указывают на ожидаемое наступление цепочки из нескольких невысоких циклов.

### 2) Выявление аналогов спектральных компонент параметров гелиогеомагнитной активности в медикобиологических рядах смертности от инфаркта миокарда в Миннесоте

Схема спектрального анализа в рамках концепции линейной оболочки выглядит следующим образом. Временной ряд делится на две части – обучающую (две трети ряда) и экзаменационную (оставшаяся одна треть). Длина реализации  $N$  выбирается равной удесятеренному максимальному из желаемых периодов в искомом спектре. Исходная размерность линейной оболочки выбирается равной единице, и отфильтрованный в рамках линейной оболочки сигнал линейно-авторегрессионным способом прогнозируется на экзаменационную часть временного ряда. Затем делается экспертная оценка удовлетворительности прогноза. Если прогноз сочтен удовлетворительным, делается FFT-анализ всего ряда, если нет – размерность линейной оболочки увеличивается на единицу, и приведенная цепочка действий повторяется.



**Рис.3.** Исходный временной ряд чисел Вольфа (темная синяя кривая), отфильтрованный ряд в рамках концепции линейной оболочки (светлая красная кривая), и его линейно-авторегрессионное продолжение вперед на 23-й и 24-й циклы солнечной активности (светлая зеленая кривая)

Нами анализировались суточные данные по смертности от инфаркта миокарда, полученные в Миннесоте с 1968 по 1996 гг. (10593 отчета), на наличие аналогов спектральных компонент временного ряда суммарного Кр-индекса. Далее исследовались динамические спектры и искались сходные периоды и их вариациях в циклах солнечной активности. Подобие вариаций ритмов в цикле СА – дополнительный аргумент в пользу гипотезы, о том, что гелиогеомагнитная активность играла роль

«биологических часов» в процессе эволюции биологических объектов.

### 3) Синхронизация ритмов Кр-индекса и смертности от инфаркта миокарда (ИМ) в Миннесоте.

Были взяты ежесуточные ряды Кр-индекса и смертности от ИМ за 26 лет (10593 отсчета). Затем в скользящем временном окне длиной 2000 отсчетов методом итеративного разворота базиса, строилась

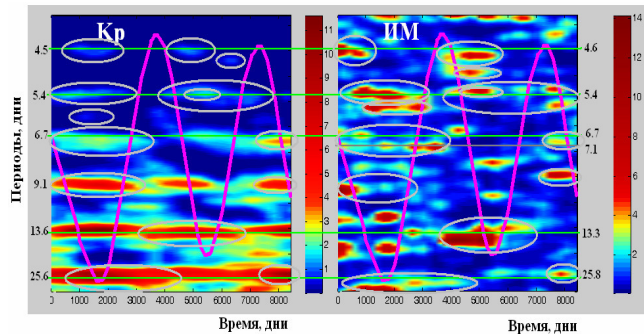


Рис.4 Динамические спектры Кр-индекса и ИМ в области высоких частот. Сиреневая (серая синусоидальная) кривая обозначает солнечную активность [6]. Серыми овалами выделены проявления анализируемых ритмов.

45-мерная линейная оболочка, и вычислялся обычный Фурье-спектр порождаемого ею ряда. Эти спектры накапливались в матрицу, которая затем усреднялась при помощи гауссовой маски, имеющей характерные размеры 0,02 циклов/день по оси частот и 5 временных отсчетов по оси времени. Полученные в результате изображения представлены на Рис.4.

Из Рис.4 можно увидеть синхронное усиление в минимумах и ослабление в максимумах солнечной активности спектральных компонент в области периодов около 4.5 дня и 5.4 дня. Так же наблюдается совпадение в появлении ритмов 6.7 дней в и 26.6 дней в Кр-индексе с ритмами 6.7-7.1 дней и с дрейфующим ритмом около 25.8 дней в смертности от ИМ. В области максимумов СА, где, как было показано методом итеративных вращений базиса наблюдается большое количество выбросов, наступают синхронные бифуркации спектральных компонент Кр и смертности от ИМ, Это, по-видимому, связано с тем, что выбросы сбивают фазы регулярных ритмических колебаний..

### Выводы

1) Разработаны алгоритмы построения идеализированной модели временных рядов, адекватной исходному ряду как на интервале наблюдения, так и в будущем, позволяющие избавиться от шумовых компонент и отфильтровать нестационарные выбросы, чтобы получить четкие спектры с малым количеством пиков. Показано, что прогноз временных рядов на базе сигналов, порождаемых линейными оболочками предельно малых размерностей, для солнечной активности (Wp-индекса) дает убедительное соответствие хода кривых

прогноза и реальности в 23 цикле СА. Модель опирается на концепцию линейной оболочки, а метод фильтрации нестационарных выбросов использует алгоритм итеративного разворота базиса.

2). Впервые удалось достоверно проследить синхронные вариации показателей геомагнитной и медико-биологической активности в высокочастотной области динамических спектров. Получено изменение характера этой синхронизации в зависимости от цикла солнечной активности: в максимумах СА, где наблюдается большое количество выбросов, наступают синхронные бифуркации спектральных компонент Кр-индекса и смертности от инфаркта. Это, очевидно, связано с тем, что выбросы сбивают фазы регулярных ритмических колебаний..

### ЛИТЕРАТУРА

- [1] A.Yu., Loskutov, Isrtomin I.A., Kuzanyan K.M., and Kotlyarov O.L. Testing and forecasting the time series of the solar activity by singular spectrum analysis// Nonlin. Phenomena in Complex Systems. 2001.V.4. .N1.P.47-57
- [2] Ю.П. Пытьев, И.А. Шишмарев. Курс теории вероятностей и математической статистики для физиков. Издательство Московского университета, 1983г.
- [3] А.А. Петрукович., А.В. Белов, В.Н.Обридко, Прогноз геоингеофизической обстановки //«Плазменная геофизика» под ред. Л.М.Зеленного и И.С.Веселовского, Физматлит., 2008.. Т. 2.. С. 235-252.
- [4] V.N Obridko. and B.D. Shelting On Prediction of the Strength of the 11-Year Solar Cycle No. 24 /Solar Phys 2008, Solar Physics, Volume 248, Issue 1, pp.191-202
- [5] В.Н.Обридко., Б.Д. Шельтинг Некоторые аномалии эволюции глобальных и крупномасштабных магнитных полей на Солнце как предвестники нескольких предстоящих невысоких циклов (ПИСЬМА В АСТРОНОМИЧЕСКИЙ ЖУРНАЛ, 2009, том 35,№3, с. 38-44 )
- [6] В.А. Ожередов и Бреус Т.К.Новые подходы к статистическому анализу рядов длительных наблюдений гелиогеомагнитной активности и медикобиологических показателей, реагирующих на нее // Геофизические процессы и биосфера. 2008.. Т.7. N1. С.7-32.

### Singular spectral analysis in solar-terrestrial physics

Ozheredov V.A.<sup>1</sup>, Breus T.K.<sup>1</sup>, Obridko V.N.<sup>2</sup>

<sup>1</sup> Space Research Institute (IKI RAS), Profsoyznaya 84/32, 117997 Moscow, Russia [ojymail@mail.ru](mailto:ojymail@mail.ru)

<sup>2</sup>Institute of the Earth Magnetism, Ionosphere and Radio Wave Propagation (IZMIRAN), Troitsk, Moscow District.

*In this work it is predicted 23th and 24th cycles of solar activity and spectral components heliogeophysical and medical long time-series analogues are found in high-frequency part of a spectrum. The expended method of spectral analysis (SSA) is used, allowing receiving the rarefied spectrum with the narrow peaks, deprived of the distortions brought by incidentally appearing non-stationary emissions.*