УДК 51-71

ПРИМЕНЕНИЯ АЛГОРИТМОВ СГЛАЖИВАНИЯ ДЛЯ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ СТАТИСТИЧЕСКИХ СОЛНЕЧНЫХ ДАННЫХ

М.А. КИЯН, Е.В. ФАБРИЧЕВА, В.Н. ПОДЛАДЧИКОВ

Исследованы погрешности метода скользящего среднего и определены условия, при которых применение метода скользящего среднего при анализе солнечной активности искажает значимые особенности этого процесса. Показано, что при аппроксимации колебаний солнечного цикла, период которых существенно больше 13 месяцев, на основе модели 13-месячной скользящей средней возникает лишь незначительная методическая ошибка. Величины таких колебаний несколько уменьшаются, однако процесс в целом не искажается. Однако колебания с периодом в границах от 6 до 12 месяцев 13-месячная скользящая средняя обращает, т.е заменяет выпуклую «волну» на вогнутую и наоборот. В граничных точках, когда период колебаний равен 6 или 12 месяцев, 13-месячное среднее приводит к полной утрате этих колебаний, т.е. уменьшает их величину до нуля. Это имеет принципиальное значения для исследований краткосрочной (месяцы) изменчивости солнечной активности.

ВВЕДЕНИЕ

Сглаживание, очищение имеющихся данных от искажающих шумов, позволяет улучшить воспроизводимость процесса и выявить недоступные без предварительной обработки особенности его динамики, отражающие не только текущий, но и будущий характер процесса.

Исходя из найденных закономерностей в прошлом, можно обосновать и учесть направление изменения динамического процесса, прогнозировать его будущее, предполагая, что основные тенденции прошлого периода сохраняются на период прогноза.

Для сложных процессов построить корректную модель часто не удается. Во многих задачах стоит проблема исследования динамики недостаточно изученного процесса. Так, несмотря на большое количество разнообразных математических моделей поведения солнечной активности, обзор которых дан в работах [1–3], предсказания завершившегося в настоящее время 23-го 11-летнего цикла солнечной активности оказались противоречивыми. Разброс прогнозируемых амплитуд 23-го цикла охватил диапазон изменения амплитуд за весь период наблюдения за солнечной активностью [4]. Систематические наблюдения за солнечной активностью ведутся с 1749 г. и в настоящее время доступна информация только о двадцати четырех 11-летних циклах. Построение модели поведения солнечной активности представляет собой очень сложную задачу, как и из-за ограниченного объема данных, так и недостаточной изученности физических процессов, происходящих на Солнце.

Глубокие и фундаментальные исследования посвящены, например, анализу снижению точности оценивания при нарушении основных предположений метода наименьших квадратов о несмещенности, постоянстве дисперсий, некоррелированности и нормальном распределении помехи [5].

Квазиоптимальные методы оценивания, не требующие точной информации о процессе, такие, как методы скользящего среднего и экспоненциального среднего, также основаны на ряде допущений, которые не могут быть заранее проверены, что приводит к искажению результата.

Однако для практики эти методы являются чаще всего просто надежным инструментом очищения имеющихся данных от искажающих шумов в условиях ограниченной информации об объекте исследования. Поэтому часто при применении методов скользящего среднего не анализируется, при каких допущениях эти методы оптимальны, какие допущения поддаются объективной предварительной проверке и к каким погрешностям приведут недостоверные допущения.

Это может привести к росту погрешности результатов, как будет показано в следующем разделе, вплоть до принятия решения, противоположного правдоподобному.

Традиционно используемое 13-месячное скользящее среднее эффективно применяется для анализа долгосрочных вариаций солнечной активности, изменения амплитуд, формы циклов и их продолжительности, понимания физики процесса пятнообразования, его закономерностей и предсказания будущей солнечной активности. В то же время 13-месячное скользящее среднее может искажать величины максимума, минимума и продолжительности солнечного цикла, как показано в [1], так как не отфильтровывает высокочастотные колебания с периодом меньше одного года.

Методы выделения краткосрочных колебаний чисел Вольфа, которые в основном основаны на спектральном анализе, который также требует принятия ряда допущений, вызывают большой интерес и широко освещены в литературе. Исследование этих краткосрочных вариаций и их обоснованная интерпретация очень важны для понимания характеристик солнечной активности.

Поэтому актуальной задачей является анализ погрешностей метода скользящего среднего, широко используемого в практике исследования солнечной активности.

Метод скользящего среднего не связан с риском расходимости результатов оценивания. Однако он субъективен в отношении выбора параметров сглаживания. Выбор периода сглаживания N при определении скользящей средней, предполагает принятие компромиссного решения между скоростью реагирования на изменение процесса и качеством фильтрации помех.

Оценка методической составляющей ошибки сглаживания затруднена из-за неопределенности модели. В этом основной недостаток метода скользящего среднего, так как невозможно оценить риск, связанный с решением, основанным на полученных оценках.

Цель работы — исследовать методические ошибки метода скользящего среднего и определить условия, при которых применение метода скользящего среднего может привести к ложным выводам.

АНАЛИЗ ОШИБОК СКОЛЬЗЯЩЕГО СРЕДНЕГО

Предположим, что в нашем распоряжении имеется последовательность измерений z_k процесса X_k в присутствии аддитивного шума η_k. Модель процесса можно описать в виде

$$z_k = X_k + \eta_k \,. \qquad (1)$$

Распространенным методом воспроизведения зависимости X от времени в условиях неопределенности изменения этого процесса, является сглаживанием скользящим средним, предложенным Слуцким [6–8].

Скользящее среднее, в котором каждый член ряда заменяется простым или взвешенным средним последних N соседних членов, является одним из самых простых и популярных методов сглаживания N — основной параметр — ширина окна или период скользящего среднего [9]. Полученное значение скользящего среднего $\hat{X}_{k,k+\frac{N-1}{2}}$ относится к середине выбранного периода, то есть дает оценку сглаженного значения с задержкой на $\frac{N-1}{2}$. Здесь первый индекс при $\hat{X}$ указывает номер сглаженной точки, а второй индекс — номер последнего измерения, которое использовалось для сглаживания. Например, при $N = 9$ простое скользящее среднее $\hat{X}_{k,k+4} = \frac{1}{N}\sum_{i=k-4}^{k+4} z_i$ в момент времени $k+4$ дает оценку уровня ряда X_k с задержкой на 4 шага.

Чем меньше параметр N скользящего среднего, тем оно быстрее определяет новую тенденцию, но и одновременно пропускает больше помех, и наоборот, чем больше параметр N, тем медленнее определяется новый тренд, но эффективней фильтруются ложные колебания.

При использовании метода скользящего среднего для сглаживания экспериментальных данных возникает типичная для практики задача выбора такого компромиссного значения ширины окна N, чтобы по возможности точно отразить общую тенденцию зависимости исследуемого процесса, но вместе с тем сгладить незакономерные, случайные уклонения, связанные с неизбежными погрешностями самого наблюдения.

Для решения этой задачи важно выяснить условия, при которых рост N не приводит к значимому росту методической ошибки и когда эти ошибки могут привести к ложным выводам.

С учетом выражения (1) полную ошибку скользящего среднего

$$\tilde{X}_k = X_k - \hat{X}_{k,k+\frac{N-1}{2}}$$

можно представить в виде суммы двух составляющих:

$$\tilde{X}_k = \Delta_k^m + \Delta_k^P, \; \Delta_k^m = X_k - \frac{1}{N}\sum_{i=k-\frac{N-1}{2}}^{k+\frac{N-1}{2}} X_i \,, \; \Delta_k^p - \frac{1}{N}\sum_{i=k-\frac{N-1}{2}}^{k+\frac{N-1}{2}} \eta_i \,,$$

где Δ_k^m — методическая составляющая ошибки, обусловленная изменчивостью самого процесса и возникающего из-за этого отклонения истинного уровня ряда от среднего значения N соседних членов, Δ_k^p — ошибка, возникающая из-за случайных помех.

Дисперсия этой составляющей ошибки с ростом ширины окна уменьшается пропорционально $1/N$.

Величина методической составляющей ошибки зависит от особенностей динамики процесса, проявляющихся в скорости изменчивости и других характеристиках. Дж.М. Херст в своей книге «The Profit Magic of Stock Transaction Timing» (1970) рассмотрел влияние ширины окна на изменение характеристик точности скользящей средней и показал, что скользящее среднее может значительно снизить краткосрочные колебания, в то время как колебания, период которых значительно превышает N, остаются почти неизменными.

Рассмотрим потенциальную точность скользящей средней и условия, при которых скользящая средняя может привести не только к уничтожению колебания, но и к появлению «обратных» колебаний, когда выпуклая «волна» заменяется на вогнутую или наоборот.

Обозначим изменение процесса на i-м шаге через δ_i. Тогда при $i > k$ его изменение $S_i^+ = X_i - X_k$ на интервале времени между i-м и k-м наблюдениями равно $S_i^+ = \sum_{j=k+1}^{i} \delta_j$, а при $i < k$ изменение процесса $S_i^- = X_k - X_i$ на интервале между k-м и i-м наблюдениями определяется выражением $S_i^- = \sum_{j=i+1}^{k} \delta_j$. Тогда методическую ошибку сглаживания можно записать в виде

$$\Delta_k^m = \frac{1}{N}\left(- \sum_{i=k-\frac{N-1}{2}}^{k-1} S_i^- + \sum_{i=k+1}^{k+\frac{N-1}{2}} S_i^+ \right).$$

Предположим, что скорость процесса не изменяется на интервале сглаживания N:

$$\delta_{k-\frac{N-1}{2}} = \ldots = \delta_k = \ldots = \delta_{k+\frac{N-1}{2}} = \delta .$$

Тогда изменения процесса $S_i^+ = (i-k)\delta$ и $S_i^- = (k-i)\delta$.

Методическая ошибка сглаживания при постоянной скорости равна

$$\Delta_k^m = -\frac{\delta}{N} \sum_{i=k-\frac{N-1}{2}}^{k-1} (k-i) + \frac{\delta}{N} \sum_{i=k+1}^{k+\frac{N-1}{2}} (i-k).$$

Оба слагаемых в полученном выражении равны по величине, но противоположны по знаку и в сумме дают ноль.

Следовательно, при постоянной скорости скользящее среднее не искажает процесс. Точность сглаживания характеризуется дисперсией ошибки $\frac{\sigma_\eta^2}{N}$, обусловленной случайными помехами, которая убывает с увеличением ширины окна N. Поэтому в условиях низкой изменчивости процесса относительно уровня случайных помех целесообразно увеличивать N, чтобы эффективней выделять тренд.

Предположим, что точка, в которой выполняется сглаживание на k-м шаге, совпадает с локальным экстремумом процесса. При этом $\frac{N-1}{2}$ соседних точек, предшествующих k-й точке, и $\frac{N-1}{2}$ следующих после нее точек, симметричны относительно экстремума. В этом случае изменения процесса до достижения экстремума равны по величине и противоположны по знаку его изменениям после этой точки, т.е. $S_i^+ = -S_i^-$.

Тогда методическую ошибку сглаживания можно представить в виде:

$$\Delta_k^m = \frac{2}{N} \sum_{i=k+1}^{k+\frac{N-1}{2}} S_i^+ .$$

Если в k-й точке достигается максимум, то $S_i^+ < 0$ для всех $i = k+1, k+2, \ldots, k + \frac{N-1}{2}$. Следовательно, $\Delta_k^m < 0$ и скользящее среднее уменьшает колебание на величину Δ_k^m и может его обнулить или даже привести к появлению «обратных» колебаний. Если в k-й точке достигается минимум, то скользящее среднее дает положительные отклонения от тренда, так как $S_i^+ > 0$ и методическая ошибка $\Delta_k^m > 0$.

ПРИМЕНЕНИЕ СКОЛЬЗЯЩЕГО СРЕДНЕГО ДЛЯ ВОССТАНОВЛЕНИЯ 11-ЛЕТНЕГО СОЛНЕЧНОГО ЦИКЛА

Одной из наиболее практически значимых задач является выявление закономерностей 11-летнего солнечного цикла по данным среднемесячных значений чисел Вольфа, которые являются основными показателями солнечной активности.

На рис. 1 приведены среднемесячные числа Вольфа с 1749 до 2011 г. Динамика чисел Вольфа показывает циклический характер поведения солнечной активности с периодом приблизительно 11 лет и кратковременные вариации, которые могут иметь как случайную природу, так и отражать закономерности динамики солнечной активности.

Для снижения помех, вызванных измерительными ошибками, во многих случаях используют 13-месячное скользящее среднее чисел Вольфа. Если z_k — месячное среднее число Вольфа для k-го месяца, то 13-месячное скользящее среднее с центром в k-м месяце определяется следующим образом:

$$\hat{X}_{k,k+6} = \frac{1}{24}\sum_{i=-6}^{5} z_{k+i} + \frac{1}{24}\sum_{i=-5}^{6} z_{k+i} .$$

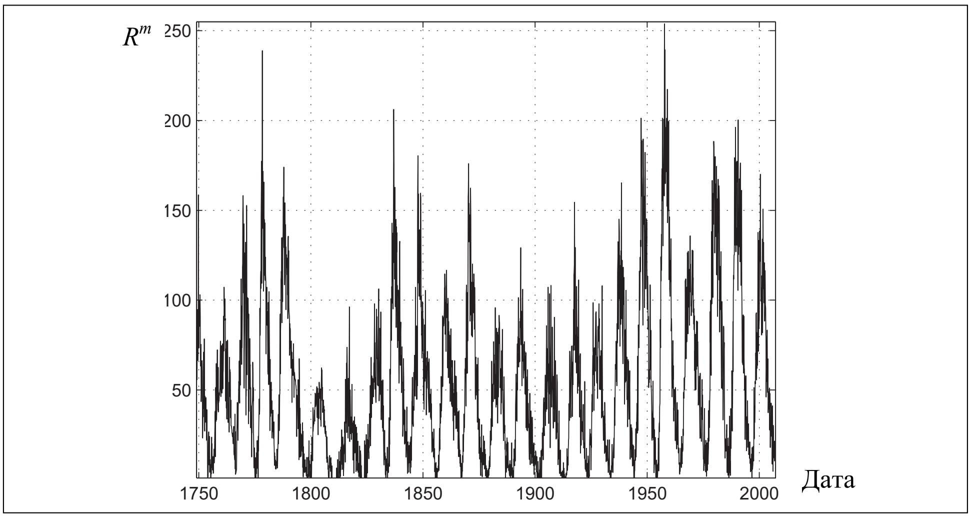

Рис. 1. Среднемесячные числа Вольфа

При осреднении пять последующих среднемесячных значений числа Вольфа z_{k+i} $(i = 1,...,5)$ и пять предыдущих z_{k+i} $(i = -5,...,-1)$ учитываются с одинаковым весом равным $\frac{1}{12}$, а z_{k-6} и z_{k+6} с весом $\frac{1}{24}$. На рис. 2 показаны 13-месячные сглаженные числа Вольфа с 1749 по 2012 г.

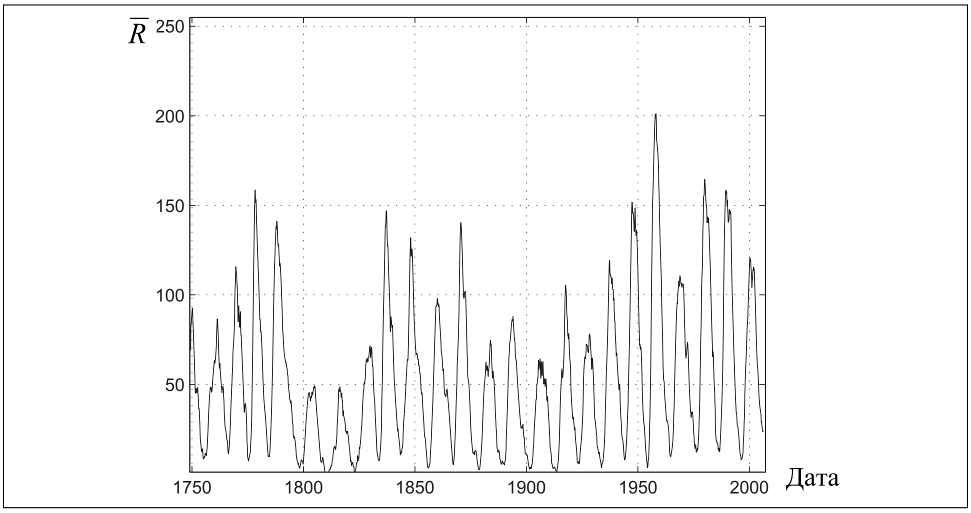

Рис. 2. 13-месячные сглаженные числа Вольфа

Полученные плавные кривые на рис. 2 используются для анализа краткосрочных и долгосрочных вариаций солнечной активности, изменения амплитуд, формы циклов и их продолжительности, понимания физики процесса пятнообразования, его закономерностей и предсказания будущей солнечной активности.

Вариации чисел Вольфа отражают влияние сил, которые являются следствием физических процессов на Солнце. Их анализ имеет принципиальное значение для понимания природы солнечной активности. Поэтому важно оценить, при каких условиях методическая ошибка 13-месячной скользящей средней может привести к потере колебаний или к появлению обратных колебаний.

ПОГРЕШНОСТИ 13-МЕСЯЧНОГО СРЕДНЕГО

Для выявления условий, при которых методическая ошибка 13-месячного скользящего среднего может привести к потере колебаний или к появлению обратных колебаний, проводилось сравнение методической ошибки 13-месячного сглаживания синусоидальных гармонических колебаний с различными периодами.

На рис. 3 показаны синусоидальные колебания с периодом около 9 месяцев (пунктирная кривая) и результаты сглаживания на основе 13-месячной средней (сплошная кривая), которые, как видно из рисунка, привели к появлению обратных колебаний.

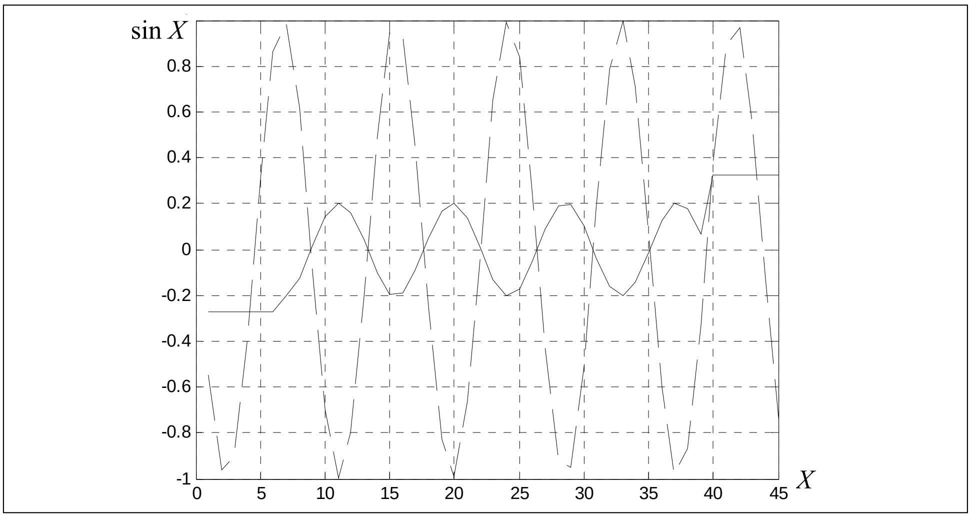

Рис. 3. Синусоидальные колебания с периодом 9 месяцев (пунктирная кривая) и результаты сглаживания на основе 13-месячной средней (сплошная кривая)

Период колебаний 9 месяцев приблизительно в 1,5 раза меньше, чем ширина окна. При сглаживании максимальных точек 13-месячное среднее рассчитывается путем усреднения значений всех точек, принадлежащих одной выпуклой волне с максимумом в усредняемых точках и двум вогнутым волнам. Из 13-ти суммируемых точек значения синусоиды положительные только в тех, которые принадлежат выпуклой волне, а в $2/3$ суммируемых точках, принадлежащим вогнутым волнам, эти значения отрицательные. Поэтому в точке максимума 13-месячное скользящее среднее приобретает отрицательное значение и приводит к замене выпуклой волны на вогнутую.

13-месячное скользящее среднее обращает все колебания с периодом в границах от 6 до 12 месяцев. В граничных точках, когда период колебаний равен 6 или 12 месяцев, 13-месячное среднее уменьшает их величину до нуля.

Синусоидальные колебания с периодом 6 и 9 месяцев показаны пунктирной кривой на рис. 4 соответственно. Из рисунков видно, что 13 месячное среднее, показанное сплошной кривой, обнуляет эти колебания.

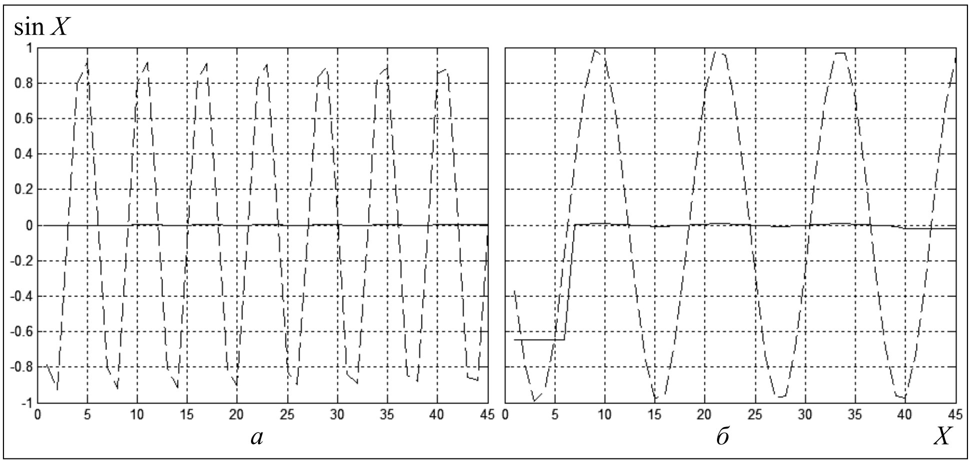

Рис. 4. Синусоидальные колебания (пунктирная кривая) и результаты сглаживания на основе 13-месячной средней (сплошная кривая): *а* — период колебаний 6 месяцев; *б* — период колебаний 12 месяцев

Если период колебаний больше 12, то 13-месячное среднее выявляет их с разной степенью искажения. С ростом периода колебаний погрешности сглаживания уменьшаются.

На рис. 5 показаны синусоидальные колебания с периодом 26 месяцев (пунктирная кривая) и результаты сглаживания на основе 13-месячной средней (сплошная кривая). Из рисунка видно, что 13-месячное среднее в целом воспроизводит колебание, однако несколько занижает его амплитуду.

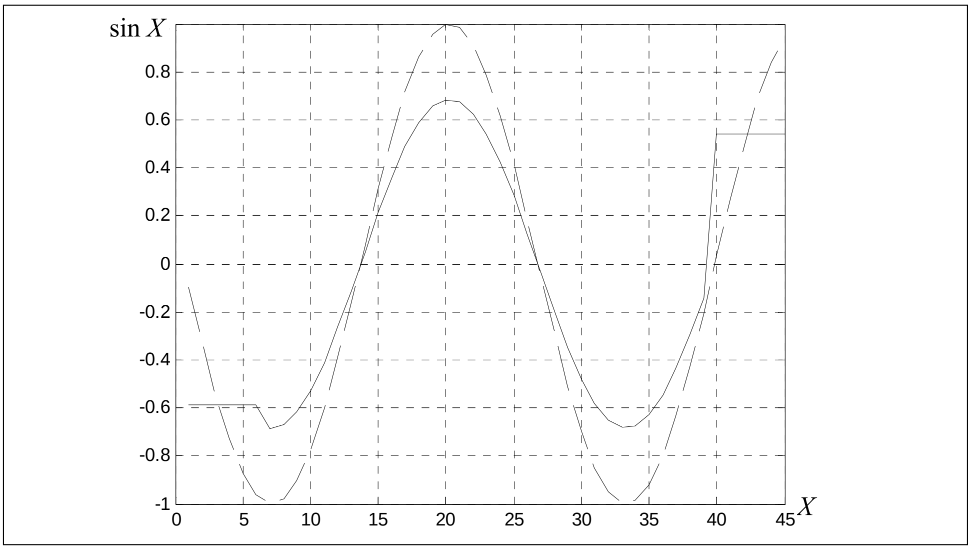

Рис. 5. Синусоидальные колебания с периодом 26 месяцев (пунктирная кривая) и результаты сглаживания на основе 13-месячной средней (сплошная кривая)

Таким образом, метод 13-месячного скользящего среднего хорошо подходит для воспроизведения среднесрочных и долгосрочных вариаций солнечной активности, но его возможности ограничены при выделении краткосрочных вариаций. По-видимому, задача исследования краткосрочных колебаний солнечной активности нуждается в таком подходе к выявлению закономерности динамики солнечного цикла, который учитывает особенности сглаживания краткосрочных колебаний

ЗАКЛЮЧЕНИЕ

В данной работе выполнен анализ условий, при которых традиционно используемое для восстановления динамики 11-летнего солнечного цикла 13-месячное скользящее среднее искажает значимые особенности исследуемого процесса.

Показано, что при аппроксимации колебаний солнечного цикла, период которых существенно больше 13 месяцев, на основе модели 13-месячнного скользящего среднего возникает лишь незначительная методическая ошибка. Величины таких колебаний несколько уменьшаются, однако процесс в целом не искажается.

Однако колебания с периодом в границах от 6 до 12 месяцев, 13-месячное скользящее среднее обращает, т.е выпуклую «волну» заменяет на вогнутую и наоборот. В граничных точках, когда период колебаний равен 6 или 12 месяцев, 13-месячное среднее приводит к полной утрате этих колебаний, т.е. уменьшает их величину до нуля. Это имеет принципиальное значение для исследований краткосрочной (месяцы) изменчивости солнечной активности.

Такие ложные вариации могут быть основной причиной того, что обычно используемое 13-месячное скользящее среднее может привести к росту погрешности, вплоть до принятия решения, противоположного правдоподобному.

Методы и принципиально новые пути аппроксимации солнечных данных, позволяющие избежать ошибок выделения краткосрочных вариаций солнечной активности будут рассмотрены в будущих публикациях

ЛИТЕРАТУРА

1. *Hathaway H., Robert R., Wilson R., Reichmann E.* A synthesis of solar cycle prediction techniques // Journal of Geophysical Research. — 1999. — **104**. — № A10. — P. 22375–22388.
2. *Petrukovich A.A., Klimov S.I.* The use of solar wind measurements for the analysis and prediction of geomagnetic activity // Cosmic Research. — 2000. — **38**, № 5. — P.
3. *Conway A.J.* Time series, neural networks and the future of the Sun // New Astronomy Reviews. — 1998. — № 42. — P. 343–394.
4. *Храмова М.Н., Красоткин С.А., Кононович Э.В.* Прогнозирование солнечной активности методом фазовых средних. — http://zhurnal.ape.relarn.ru/arcticles /2001/107/pdf.
5. *Себер Дж.* Линейный регрессионный анализ. — М.: Мир, 1980. — 456 с.
6. *Слуцкий Е.Е.* Избранные труды. — М.: Изд-во АН СССР, 1960. — 291 с.
7. *Редкозубов С.А.* Статистические методы прогнозирования в АСУ. — М.: Энергоиздат, 1981. — 152 с.
8. *Андерсон Т.* Статистический анализ временных рядов. — М.: Мир, 1976. — 755 с.
9. *Бокс Дж., Дженкинс Г.* Анализ временных рядов, прогноз и управление. Пер. с англ. 1974. — Вып. 1–2. — 604 с.

Поступила 08.07.2013