

01;09

О выделении слабой периодической компоненты из нестационарного временного ряда

© Л.М. Мартюшев, Е.Г. Аксельрод, А.П. Сергеев

Институт промышленной экологии УрО РАН, Екатеринбург
E-mail: mlm@ecko.uran.ru

Поступило в Редакцию 25 марта 2003 г.

Рассмотрены различные подходы к обнаружению слабой периодической компоненты в зашумленном нестационарном временном ряде в случае отсутствия априорной информации о полезном сигнале и невозможности многократных измерений. Предложенные методы использованы для изучения зависимости массы от времени растущего дендритного кристалла хлористого аммония из водного раствора. Обнаружено, что при скоростях роста более $20 \mu\text{m/s}$ на кинетических кривых имеются низкочастотные осцилляции с периодом около 3 с.

Несмотря на обилие литературы, проблема обнаружения периодической компоненты в сильно зашумленных нестационарных временных рядах по-прежнему существует [1–4]. Наибольшая сложность возникает при этом в двух случаях: а) если невозможно многократно повторить эксперимент; б) если отсутствует априорная информация о периодической компоненте, а также виде функциональной зависимости тренда (само исследование должно дать эту информацию). В этих случаях существенным становится значение отношения сигнала к шуму. При возможности многократного повторения эксперимента и/или наличии предварительной информации об искомом сигнале выделить и проанализировать интересующую составляющую теоретически всегда можно. В первом случае увеличение отношения сигнала к шуму происходит как квадратный корень из числа наблюдений, а во втором — использование специально подобранных фильтров всегда позволит выделить сигнал. Особый интерес представляет наиболее сложный случай, когда имеется единственный нестационарный временной ряд и отсутствует априорная

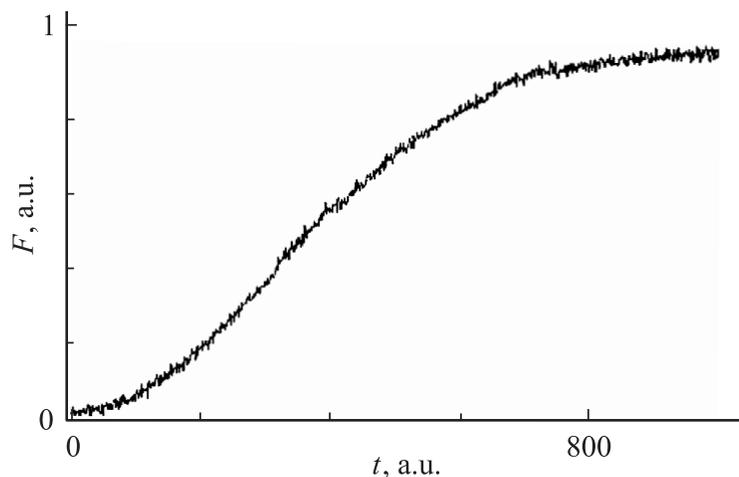


Рис. 1. Модельная кривая F , сгенерированная с помощью наложения на распределение Вейбула гауссова шума (с нулевым средним и дисперсией sd) и прямоугольных импульсов ($n = 3$, $T/N = 1/10$, $N = 1000$, $A/sd = 1.5$). Здесь t — относительное время.

информация о сигнале. Эта ситуация типична при экспериментальных физических исследованиях и ей посвящается работа.

Статья будет состоять из двух частей. В первой части будут исследоваться модельные сгенерированные временные ряды, а во второй — будет проведен анализ результатов конкретных экспериментов по дендритному кристаллическому росту из раствора.

Пусть имеется некоторая зашумленная S-образная кривая (рис. 1) и необходимо выяснить наличие или отсутствие у нее скрытой периодичности. Кривая такой формы выбрана, с одной стороны, для определенности, а с другой — в связи с распространенностью при рассмотрении различных эволюционных и кинетических процессов. Периодическая компонента характеризуется амплитудой A , формой, числом n и длительностью T импульсов. Последний параметр удобно нормировать на общее число точек временного ряда N . Так как нашей задачей является лишь выявление периодичности (и ее периода), а не

ее формы, то для определенности далее использовались импульсы прямоугольной формы. Шум предполагался аддитивный, гауссовый с нулевым средним и со среднеквадратичным отклонением (sd).

Исследовались два случая: $n = 3$, $T/N = 1/10$ либо $1/100$. Необходимо было определить предельное отношение A/sd , при котором возможно обнаружить периодическую компоненту. Очевидно, что увеличение n будет приводить к уменьшению искомого предельного отношения, поэтому, выбрав n столь малым, мы рассчитываем получить оценку для наиболее неблагоприятной ситуации — малого числа возможных периодов на временном ряде.

1. Рассмотрим первый случай ($n = 3$, $T/N = 1/10$, $N = 1000$, частота 0.003). $A/sd \leq 2$ визуально нельзя определить, имеется или нет особенность на модельной кривой (рис. 1). Для выявления осцилляций на подобных кривых использовались различные способы [3,4]:

1 а. Спектральное преобразование от нестационарного ряда и от временного ряда, подвергнутого дифференцированию (рис. 2, а). Как видно, данные методы ничего не выявляют. Необходимо отметить, что, как показали расчеты при $A/sd > 2$ (при едва заметных осцилляциях), эти прямые методы позволяют с хорошей точностью выявить периодичность. При этом наиболее хорошие результаты получаются при использовании спектрального окна типа Blackman, а предварительное дифференцирование часто не улучшает результат.

1 б. Было проведено вычитание трендов различных типов. Подбор их проводился в специализированном пакете TableCurve2D. Предпочтение отдавалось трендам, имеющим достаточно большой коэффициент корреляции r^2 . На рис. 2, б приведены результаты спектральных преобразований после вычитания из нестационарного ряда (рис. 1) двух различных трендов, имеющих одни из максимальных значений r^2 . Как видно, результаты существенно отличаются при использовании различных трендов. Для первого случая максимальный пик соответствует частоте 0.0019, вид второго ближе к шумовому. Отметим, что weibullовский тренд, используемый для построения кривой (рис. 1), по результатам расчетов в TableCurve2D имел $r^2 = 0.9989$. Анализ многочисленных расчетов показал, что, используя в качестве вычитаемого тренда различные регрессионные зависимости, достаточно хорошо описывающие исходный временной ряд, можно получить для кривой типа рис. 1 всевозможные выводы о флуктуирующей компоненте

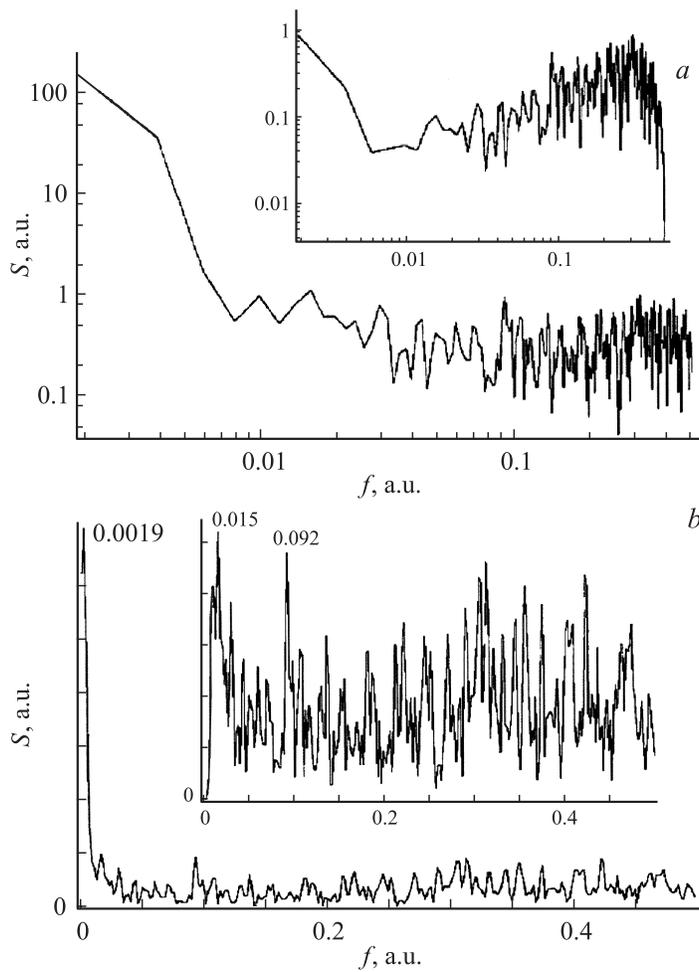


Рис. 2. Результаты спектрального анализа кривой, представленной на рис. 1. Преобразование выполнено по 512 точкам, с использованием спектрального окна Blackman от: *a* — недифференцированного (на основном графике) и дифференцированного (на вставке) временного ряда; *b* — остатков, полученных при вычитании тренда $\ln F = a + bt^{1.5} + c \ln t$ (*a, b, c* — параметры регрессионной кривой), $r^2 = 0.9984$ (на основном графике) и остатков, полученных при вычитании полинома 20-й степени $r^2 = 0.9992$ (на вставке).

(начиная от шумового поведения до периодического с существенно различающимися частотами).

1 в. Основываясь на данных, приведенных в [5], делались попытки применить модное в настоящее время вейвлетное преобразование. Для этого использовался специальный модуль пакета MATLAB. Было протестировано несколько типов наиболее распространенных вейвлетов: Морлета, Мексиканская шляпа и другие, а также использована различная степень детализации. Однако, как показали модельные расчеты, для рассматриваемого типа задач вейвлетные методы выделения периодичности оказываются неэффективными.

1 г. Использование метода скользящего среднего для приведения временного ряда к стационарному виду наталкивается на те же трудности, что и в методе 1 б. Выбрав слишком мало точек сглаживания, мы рискуем вычистить вместе с нестационарностью и возможную периодичность, а увеличив число точек сверх некоторого предела, напротив, можем искусственно привнести определенные частоты.

Однако с помощью этого метода оказалось возможным построить простой и удобный для программной реализации метод (Σ -метод), позволяющий обнаружить периодическую компоненту в ряде, представленном на рис. 1. Проводится многократный расчет остатков после вычитания из исходного временного ряда тренда, найденного с помощью метода скользящего среднего при последовательном изменении числа точек усреднения. Затем для каждого такого ряда остатков рассчитывается амплитудный спектр (или автокорреляционная функция), и далее, отнормированные на максимальное значение, они суммируются. В результате суммарная функция представляет собой своего рода вероятность нахождения той или иной частоты в исследуемом временном ряде. Очевидно, что именно отсутствие предварительной информации о периоде осцилляционной компоненты вынуждает использовать такой перебор с накоплением. Для ряда, приведенного на рис. 1, соответствующая суммарная автокорреляционная функция приведена на рис. 3. Для расчета этой функции использовались скользящие средние, рассчитанные по 50 . . . 200 точкам. Разница между двумя ближайшими максимальными пиками составляет 331, и соответственно относительная ошибка в определении частоты составляет менее 1%. Таким образом, данный метод позволяет обнаружить скрытую периодичность нестационарного сигнала.

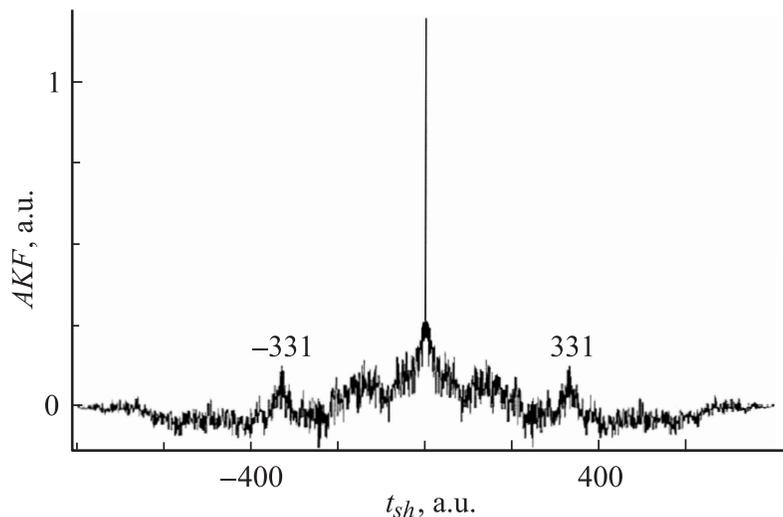


Рис. 3. Суммарная автокорреляционная функция (AKF) в зависимости от времени сдвига, рассчитанная для кривой, приведенной на рис. 1.

Как показали многочисленные расчеты с сигналами типа приведенного на рис. 1, выводы о наличии периодичности с помощью указанного выше способа удается получать лишь с некоторой вероятностью. Это связано с тем, что в модельных расчетах используется случайный гауссовский шум с дисперсией, практически равной амплитуде сигнала. Выводы, основанные на расчете нескольких десятков модельных сигналов, оказались следующие: при $A/sd = 2; 1.7; 1.5; 1.2$ соответствующие вероятности правильного определения частоты скрытой периодичности составляют 1; 0.80; 0.78; 0.50. Таким образом, при отношениях A/sd менее 1.5 предсказания с помощью Σ -метода резко ухудшаются. Исследования показали, что для увеличения вероятности обнаружения скрытого периодического сигнала можно использовать два метода:

а) аппаратное увеличение дискретизации сигнала (при условии, что это не приводит к ухудшению отношения сигнала к шуму);

б) составить несколько выборок из имеющегося единичного сигнала и для анализа каждой использовать Σ -метод, а затем все полученные

суммарные автокорреляционные функции (или спектры) просуммировать. Оказывается полезным также расчет всевозможных корреляционных функций между остатками, относящихся к различным выборкам и полученных при усреднении по различному числу точек.

В качестве примера приведем результат, полученный при одновременном использовании методов a и b . В работе по-прежнему использовался сигнал $n = 3$, $T/N = 1/10$, однако число точек дискретизации было увеличено в 10 раз ($N = 10\,000$). Из данного ряда формировались выборки двух типов (первый: 5 временных рядов по 1000 точек и второй: 50 рядов по 100 точек), при этом интервалы между точками выбирались равными. Итоговый результат оказался следующим, при A/sd , равном 1.5 и 1, вероятность правильного определения составляла 0.91 и 0.62 соответственно. При уменьшении A/sd менее единицы вероятность предсказания для используемого временного ряда резко уменьшается (0.30 при $A/sd = 0.8$).

2. Рассмотрим второй случай ($n = 3$, $T/N = 1/100$, $N = 1000$). При $A/sd \leq 2.5$ визуально практически нельзя определить, имеется ли особенность на модельной кривой. Использование всех перечисленных выше методов, включая Σ -метод, не приводят к обнадеживающим результатам. Из результатов расчетов модельных кривых можно сделать вывод, что лишь увеличение дискретизации сигнала способно выявлять скрытые периодичности (так если T/N по-прежнему 1/100, но N увеличить до 10000, то сигнал с $A/sd = 2.5$ выделяется с вероятностью 1.00). Таким образом, обнаружение сигналов малой амплитуды и длительности сопряжено с большими сложностями.

Полученные выше результаты были использованы для анализа экспериментальных данных по кинетике роста одиночного дендрита хлористого аммония из водного раствора. Цель исследования состояла в обнаружении на нестационарной кривой (масса растущего дендрита — время) неслучайной периодической компоненты. Технические особенности эксперимента подробно изложены в [6,7]. Проведенный ранее в этих работах статистический анализ кинетических кривых, использующий методы 1, a – z выявил наличие неслучайной низкочастотной компоненты. Однако, как уже говорилось, точность данных методов оказывается не всегда высокой. Поэтому для количественного уточнения полученного ранее результата было проведено дополнительное исследование. В технику эксперимента по сравнению с [6,7] были внесены два изменения: использовалась более помехоустойчивая электронная схема

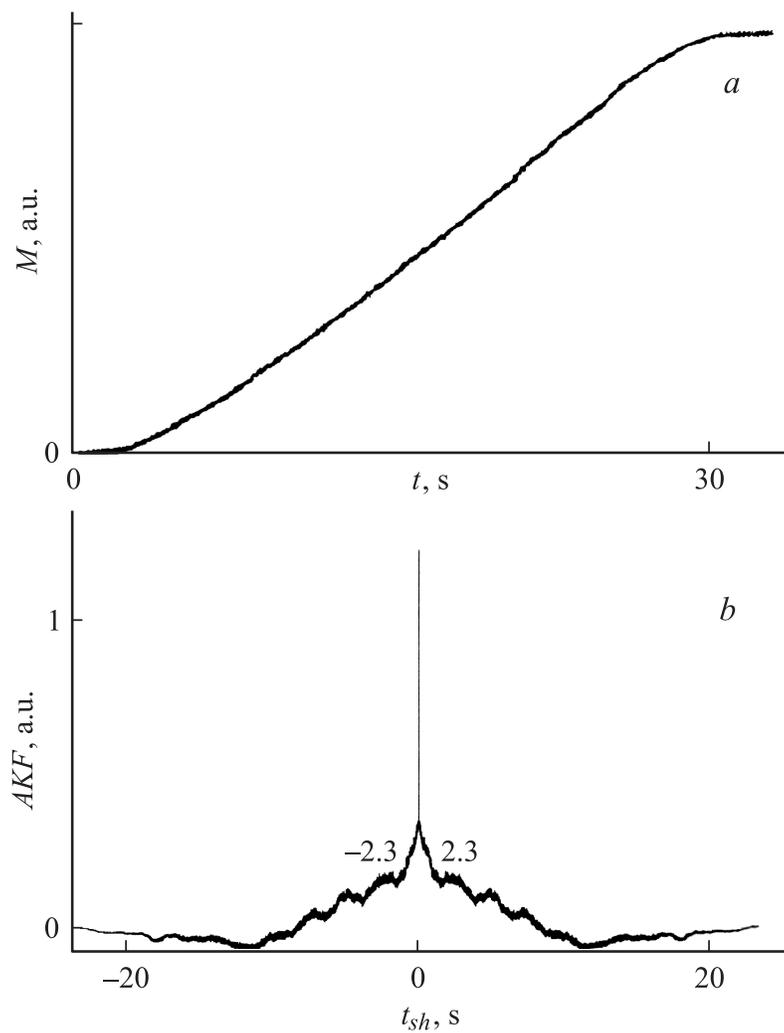


Рис. 4. Результаты эксперимента: a — масса M растущего дендрита со временем t, s ; b — суммарная автокорреляционная функция (AKF) в зависимости от времени сдвига, s .

приема и усиления сигнала, а также применена новая плата сбора данных (АЦП: Lcard-1250). Все это позволило на несколько порядков повысить степень дискретизации сигнала без изменения в погрешности измерения. Было получено несколько кинетических кривых типа приведенной на рис. 4, *a*. Это типичные *S*-образные кривые длительностью не более одной минуты (дискретизированной по 66 152 точкам). Важной особенностью изучаемой системы является невозможность многократного повторения эксперимента при неизменных параметрах, поэтому все полученные временные ряды относились к различным дендритам и, как следствие, их статистический анализ проводился независимо. Для выявления периодичности использовался Σ -метод. Одна из типичных суммарных автокорреляционных функций приведена на рис. 4, *b*. Вид этой кривой указывает на наличие неслучайной компоненты во временном ряде. Обработка данных с помощью Σ -метода всех полученных кривых позволяет говорить о наличии на кинетической кривой (масса дендрита — время) неслучайной периодической компоненты с периодом 2.8 ± 0.8 секунд. Этот результат примерно в три раза меньше полученного ранее и значительно лучше согласуется с имеющимися теоретическими оценками [6,7]. Другой интересный результат, который удалось получить при статистической обработке, заключается в том, что осцилляционная компонента на кинетических кривых зависит от скорости роста дендрита. Наиболее надежно о наличии неслучайных осцилляций можно говорить при скоростях дендрита более $20-30 \mu\text{m/s}$. При меньших скоростях роста дендрита максимумы на кривых, подобных рис. 4, *b*, либо не столь ярко выражены, либо отсутствуют. Поэтому вопрос о наличии осцилляций для таких медленно растущих дендритов должен исследоваться дополнительно.

Список литературы

- [1] Вайнштейн Л.А., Зубаков В.Д. Выделение сигналов на фоне случайных помех. М.: Сов. радио, 1960. 447 с.
- [2] Левин Б.Р. Теоретические основы статистической радиотехники. Т. 2. М.: Сов. радио, 1975. 392 с.
- [3] Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. М.: ИНФРА, 1998. 528 с.
- [4] Обработка экспериментальных данных с использованием компьютера / Под ред. С. Минами. М.: Радио и связь, 1999. 256 с.

- [5] Дьяконов В.П. Вейвлеты. От теории к практике. М.: СОЛОН-Р, 2002. 448 с.
- [6] Аксельрод Е.Г., Мартюшев Л.М., Левкина Е.В. // Письма в ЖТФ. 1999. Т. 25. В. 20. С. 64–70.
- [7] Axelrod E.G., Martiouchev L.M., Lyovkina Y.V. // Phys. Stat. Sol. (a). 2000. V. 182. P. 687–696.