

УДК 681.511

ПОИСК ПЕРИОДИЧНОСТЕЙ В ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ МЕТОДАМИ АВТОРЕГРЕССИОННОЙ МОДЕЛИ

Б. З. Белашев^а, М. К. Сулейманов^б

^аИнститут геологии Карельского научного центра РАН, Петрозаводск

^бОбъединенный институт ядерных исследований, Дубна

Обработка данных интерференционных экспериментов в физике высоких энергий предполагает привлечение методов спектрального анализа. Изложены методы спектрального анализа, использующие авторегрессионную модель экспериментальных данных: максимума энтропии, Писаренко и Прони. Возможности методов продемонстрированы на примере экспериментальных и симулированных зашумленных данных.

The treatment of data of the interferential experiments in high energy physics needs to attract the spectral analysis methods. The spectral analysis methods are explained: the maximum entropy technique, the Pisarenko and the Prony based on the autoregressive model data. The possibilities of these methods have been demonstrated on the examples of experimental and simulation data with the noise.

Альтернативой повышению энергии сталкивающихся частиц при исследовании свойств вещества на малых расстояниях является наблюдение и изучение тонких эффектов связанных состояний атомов и элементарных частиц [1]. Особый интерес представляют энергии и фазы интерференции состояний, чувствительные к поведению потенциала на малых расстояниях и внешним воздействиям [2]. Информацию о связанных состояниях систем получают в интерференционных экспериментах [3]. Данные такого эксперимента представлены осциллирующими зависимостями, по спектральным параметрам которых судят об энергетических и фазовых характеристиках системы [4]. Достоверность получаемой информации зависит от способности выделять из экспериментальных данных периодичности и определять их параметры.

Математическая теория и алгоритмический аппарат выявления скрытых периодичностей имеют длительную историю и разнообразные практические приложения. Методы выявления периодичностей подразделяют на аппроксимационные, позволяющие аппроксимировать экспериментальные данные функцией, соответствующей априори заданной модели, и фильтрационные, дающие информацию о компонентах и их параметрах при минимальных априорных сведениях. По сравнению с аппроксимационными методами фильтрационные методы дают меньшую точность при определении спектральных характеристик сигнала [5].

В работе речь пойдет о поиске скрытых периодичностей в данных самой различной природы с помощью методов на основе авторегрессионной модели данных (АМД) [6].

Методы на основе АМД занимают промежуточное положение среди обсуждаемых групп методов обработки данных. Аппроксимация в них представлена постулатом о

локальной связи данных, а фильтрация — операциями по получению спектра и выделению компонент. Благодаря промежуточному положению методы АМД используют преимущества каждой из описанных групп и характеризуются более высокой точностью определения спектральных параметров и разрешением по сравнению с фильтрационными методами на основе дискретного преобразования Фурье (ДПФ) [7, 8].

В отличие от алгоритмов ДПФ, в которых наблюдаемый сигнал раскладывают на гармонические компоненты, некоторые из АМД-методов (метод Прони) используют разложение на компоненты в виде затухающих колебаний. При обработке данных интерференционного эксперимента такое представление оказывается более реалистичным по сравнению с традиционным, поскольку учитывает время жизни состояний. Эта тенденция продолжена в вейвлет-анализе, где разложение на солитоноподобные компоненты позволило проследить изменения спектральных характеристик сигналов в пространстве и времени [9].

Метод Прони дает возможность определить фазы колебаний с точностью более высокой, чем при использовании ДПФ. Такая возможность представляется важной с учетом информативности фаз в интерференционных экспериментах.

В работе проведено моделирование АМД спектральных методов максимума энтропии (ММЭ), Писаренко и Прони [10–12] на примере симулированных и экспериментальных данных.

Авторегрессионная модель данных. АМД случайного процесса постулирует локальную связь данных: текущее значение функции $y(l)$, аппроксимирующей исследуемый процесс, определено предшествующими отсчетами данных $x(l-k)$ в виде линейной комбинации:

$$y(l) = \sum_{k=1}^p a(k)x(l-k) \quad (1)$$

с глубиной p и параметрами $a(k)$. Параметры $a(k)$ содержат всю информацию о спектральных характеристиках процесса [13].

Вычитая из обеих частей равенства (1) значение переменной $x(l)$, ошибку текущего значения $\varepsilon(l) = y(l) - x(l)$ выражают в виде

$$\varepsilon(l) = \sum_{k=0}^p a(k)x(l-k) \quad (2)$$

с параметром $a(0) = -1$. Ошибка текущего значения также представлена линейной комбинацией предшествующих отсчетов данных, что позволяет рассматривать параметры $a(k)$ как коэффициенты линейного фильтра, предсказывающего ошибку.

Ошибку предсказания текущего значения ошибки можно минимизировать, исходя, например, из принципа наименьших квадратов. При независимых сигнале и аддитивном шуме поступающие на вход фильтра предсказания ошибки данные наблюдений на выходе превращаются в белый шум. Такой фильтр называют «отбеливающим».

Умножая обе части соотношения (2) на комплексно-сопряженное значение $x^*(l)$ и усредняя по всем данным наблюдений, для параметров $a(k)$ получаем систему линейных

уравнений (3)

$$\sum_{k=0}^p R(m-k)a(k) = \sigma^2 \delta(m), \quad (3)$$

где σ^2 — дисперсия шума; $\delta(m)$ — дельта-функция; $R(m-k)$ — значения автокорреляционной функции $R(k)$, $m = 0, 1, 2, \dots, p$.

Для нахождения коэффициентов фильтра $a(k)$ обычно применяют рекурсивный алгоритм Дербу–Левинсона [14], состоящий в последовательном уточнении коэффициентов отбеливающего фильтра и оценке мощности предсказанной ошибки, начиная с фильтра первого порядка и кончая фильтром p -го порядка, вычисляющего последовательность: $\{a_{11}, \sigma_1^2\}, \{a_{21}, a_{22}, \sigma_2^2\}, \dots, \{a_{p1}, a_{p2}, \dots, a_{pp}, \sigma_p^2\}$ по формулам:

$$\begin{aligned} a_{11} &= -R(1)/R(0); \quad \sigma_1^2 = (1 - |a_{11}|^2)R(0), \\ a_{kk} &= - \left[R(k) + \sum_{i=1}^{k-1} a_{k-1,i} R(k-i) \right] / \sigma_{k-1}^2, \\ a_{ki} &= a_{k-1,i} + a_{kk} a_{k-1,k-i}^*; \quad \sigma_k^2 = (1 - |a_{kk}|^2) \sigma_{k-1}^2. \end{aligned} \quad (4)$$

Быстродействие и экономичность алгоритма Дербу–Левинсона объясняется тем, что в отличие от алгоритма решения систем линейных уравнений Гаусса для нахождения решения он требует p^2 операций вместо p^3 операций и использует результаты предшествующих вычислений. Для тех же целей разработаны более эффективные алгоритмы [15].

Метод максимума энтропии (ММЭ). В качестве спектрального метода ММЭ был впервые предложен Бургом [10]. Идея состояла в максимизации энтропийного функционала плотности спектральной мощности процесса $P(l)$

$$- \sum_{l=-(n-1)/2}^{(n-1)/2} \ln P(l) - \max \quad (5)$$

при выполнении условий теоремы Винера–Хинчина для $p+1$ известных значений автокорреляционной функции

$$\sum_{l=-(N-1)/2}^{(N-1)/2} P(l) \exp(-j\Delta\omega lk\Delta t) = R(k), \quad (6)$$

где $\Delta\omega$ и Δt — частотные и временные интервалы между отсчетами спектра и данных; j — мнимая единица ($j^2 = -1$), $k = 0, 1, 2, \dots, p$.

Для гауссовского случайного процесса требование максимума энтропийного функционала эквивалентно минимуму функционала предсказываемой ошибки. Решение такой вариационной задачи с коэффициентами отбеливающего фильтра $a(k)$ в качестве множителей Лагранжа дано формулой

$$P(l) = \frac{\sigma^2}{1 + \sum_{k=1}^p a(k) \exp(-j\Delta\omega l \Delta t k)^2}, \quad (7)$$

имеющей простой смысл: спектр мощности сигнала получают делением выходной мощности шума на квадрат модуля спектральной характеристики отбеливающего фильтра. Коэффициенты $a(k)$ находят в АМД путем решения системы уравнений (3).

ММЭ учитывает локальную связь данных, дает оценку спектра мощности шума, не приводит к отрицательным значениям в спектре мощности, имеет лучшее разрешение по сравнению с ДПФ, но уступает по разрешению методам Писаренко и Прони. В силу эквивалентности методу наименьших квадратов ММЭ дает несмещенную оценку спектра. А требование максимальной энтропии или минимальной при заданных ограничениях информации автоматически исключает из оценки спектра все посторонние частотные пики. По этой причине следует ожидать, что в ММЭ спектральные пики, соответствующие ангармонической периодической гистограмме, должны проявляться слабее, чем в ДПФ [16].

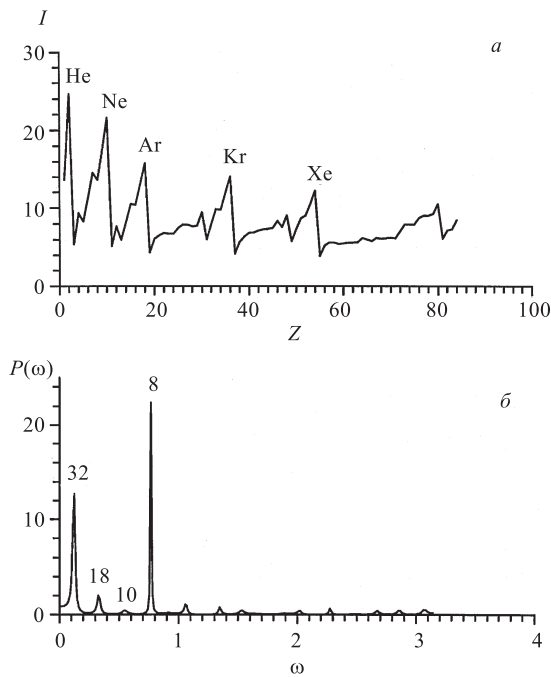


Рис. 1. Зависимость потенциала ионизации атомов в основном состоянии от атомного номера элемента (а) и ММЭ-оценка ее спектра мощности (б). Частотным пикам спектра мощности поставлены в соответствие периоды таблицы Менделеева

Во втором примере характер исследуемого сигнала весьма близок смеси гармоник и шумовой компоненты. В спектре мощности данного сигнала (рис. 2, б) все среднечастотные и высокочастотные пики нашли объяснение. Этим пикам в порядке возрастания периодов были сопоставлены сейшевые колебания водоема, соединенного через пролив с Петрозаводской губой, Петрозаводской губы и Большого Онего. Сопоставление с данными натуральных наблюдений и моделирования дало согласующиеся результаты [18].

ММЭ эффективен при экспрессном определении частотного спектра случайных процессов. Применение ММЭ к выявлению скрытых периодичностей рассмотрим на примере зависимостей потенциала ионизации атома в основном состоянии от атомного номера элемента [17] (рис. 1, а) и скорости течения через Соломенский пролив Петрозаводской губы Онежского озера от времени [18] (рис. 2, а). Для получения частотных спектров в этих примерах была использована программа [19].

Частотный спектр рис. 1, б демонстрирует периоды 8, 10, 18, 32, характерные для таблицы Менделеева. ДПФ-спектр такого ангармонического распределения помимо основных периодов $T_{k,1}$ содержит периоды $T_{k,i} = T_{k,i-1}((i-1)/i)$, $i = 2, 3, 4$, составляющие для наиболее интенсивного и узкого пика периода 8 соответственно 4,0, 2,7, 2,0. В данном спектре слабоинтенсивных пиков оказывается значительно больше, а среди их периодов 6,0, 4,7, 4,1, 3,1, 2,8, 2,3, 2,2, 2,1 только 4,1, 2,8 и 2,1 можно считать близкими к указанным ангармоническим периодам.

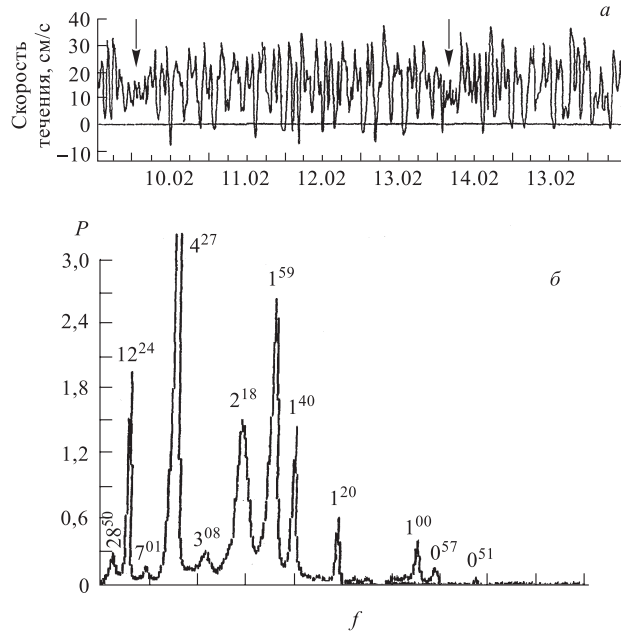


Рис. 2. Временная зависимость скорости течения через Соломенский пролив Петрозаводской губы Онежского озера (а) и ее ММЭ-оценка спектра мощности (б). Частотным пикам соответствуют периоды сейшевых колебаний: Логозера (водоема, соединенного через Соломенский пролив с Петрозаводской губой) — 22–60 мин; Петрозаводской губы — 1 ч 20 мин–2 ч 18 мин; Онежского озера — 3 ч 08 мин–4 ч 27 мин, приливного периода — 12 ч 24 мин

Спектральный анализ выявил ранее неизвестный для озерных водоемов приливной период около 12 часов, а также другие низкочастотные периоды, с трудом поддающиеся интерпретации. Для анализа полученных результатов спектральное разрешение ММЭ оказалось вполне достаточным.

Методы Писаренко и Прони. Метод Писаренко, применяемый для выделения гармонических компонент из их смеси с белым шумом, дает еще более высокое спектральное разрешение, чем ММЭ. В этом методе частоты компонент определяют не визуально, как в ДПФ или ММЭ, по положениям пиков частотного спектра, а вычисляют. Точность определения частоты и спектральное разрешение заданы точностью вычислений.

Алгоритм вычислений частот в методе Писаренко продемонстрируем на примере синусоиды $x(l) = \sin(\Omega l)$. Тригонометрическое тождество

$$\sin(\Omega l) = 2 \cos \Omega \sin(\Omega(l-1)) - \sin(\Omega(l-2)) \quad (8)$$

дает связь между данными:

$$x(l) = 2 \cos \Omega x(l-1) - x(l-2). \quad (9)$$

Фурье-преобразование соотношения (9) приводит к соотношению

$$X(\omega)(1 - 2 \cos \Omega z^{-1} + z^{-2}) = 0, \quad (10)$$

где под z понимают $\exp(-j\Omega)$. Корни уравнения второй степени $z_1 = z_2^*$ определяют частоты Ω и $-\Omega$ в соответствии с выражением

$$\Omega = \arctg(\operatorname{Im} z_i / \operatorname{Re} z_i). \quad (11)$$

В общем случае локальную связь между данными в методе Писаренко задают в виде

$$x(l) = - \sum_{k=1}^{2p} a(k)y(l-k) + \varepsilon(l), \quad (12)$$

а для коэффициентов $a(k)$ получают систему линейных уравнений

$$\sum_{k=0}^{2p} R(m-k)a(k) = \sigma^2 a(m). \quad (13)$$

Коэффициенты $a(k)$ и дисперсию шума σ^2 находят, определяя собственные числа и собственные векторы матрицы автокорреляционных функций $R(m, k)$ [20]. Минимальному собственному числу матрицы $R(m, k)$ соответствует дисперсия σ^2 . При этом удобно использовать рекурсивное выражение $Rc(l+1) = c(l)$, следующее из (13), и определить вектор $c(l+1)$ по вектору $c(l)$, полученному на предыдущей итерации. Вектор $c(0) = [1, 1, 1, \dots, 1]$ удобно принять за начальное приближение. Несколько итераций позволяют получить вектор, близкий к $c(\infty)$. По нему вычисляют $\lambda_{\min} = \sigma^2 = (c^T Rc / c^T c)$, и вектор a находят как $a = c / \lambda_{\min}$.

Структура матрицы автокорреляционных функций $R(m, k)$ такова, что при наличии в исследуемом процессе гармоник коэффициенты фильтра $a(k)$ действительны и удовлетворяют условию: $a(k) = a(2p-k)$. Составляя и решая алгебраическое уравнение степени $2p$ с действительными коэффициентами $a(k)$ и $a(0) = 1$

$$\sum_{k=0}^{2p} a(k)z^{2p-k} = 0, \quad (14)$$

получают комплексно-сопряженные, равные по модулю единице корни, определяющие частоты гармоник в соответствии с выражением (11).

Положительные частоты гармоник Ω_i , $I = 1, 2, \dots, p$, используют для нахождения плотностей спектральной мощности $P(l)$, решая систему уравнений

$$\sum_{i=1}^p P(i) \cos(\Omega_i l \Delta t) = R(l). \quad (15)$$

Самосогласованность полученных результатов проверяют, определяя дисперсию шума σ^2 , выразив ее соотношением (16)

$$\sigma^2 = R(0) - \sum_{i=1}^p P(i). \quad (16)$$

Метод Писаренко позволяет определить дисперсию аддитивного белого шума. Однако в методе отсутствуют гарантии получения неотрицательных оценок спектральной плотности гармоник. Как и в ММЭ, в методе Писаренко начальную фазу гармоник не определяют.

Этот недостаток преодолен в методе Прони, обобщившем идею Писаренко на случай нестационарных периодических процессов, с использованием в качестве базиса спектрального разложения затухающие колебания.

В методе Прони функция, аппроксимирующая процесс, $y(l)$ выражена через комплексные числа b_i и z_i соотношением

$$\begin{aligned} y(l) &= \sum_{i=1}^{2p} b_i z_i, \quad l = 0, 1, 2, \dots, n-1, \\ b_i &= A_i \exp(j\theta_i), \\ z_i &= \exp(\alpha_i + j\Omega_i \Delta t), \end{aligned} \quad (17)$$

где параметры A_i , α_i , Ω_i , θ_i — соответственно амплитуда, коэффициент затухания, частота и начальная фаза i -й компоненты разложения исходных данных.

Локальную связь между данными постулируют так же, как и в методе Писаренко, так же определяют коэффициенты $a(k)$ и по ним составляют полином (14). Отличие заключается в том, что корнями полинома теперь оказываются комплексные экспоненты z_i .

Исходя из минимума функционала наименьших квадратов

$$\sum_{l=0}^{n-1} (x(l) - y(l))^2 - \min \quad (18)$$

находят вектор $B(b_1, b_2, \dots, b_{2p})$, определяемый соотношением

$$B = (\Phi^{\#} \Phi)^{-1} \Phi^{\#} X, \quad (19)$$

где под X понимают строку данных, под Φ — матрицу Ван-дер-Монда, составленную из степеней корней z_i , под $\Phi^{\#}$ — транспонированную матрицу. А по нему и корням полинома z_i определяют параметры компонент разложения:

$$\begin{aligned} A_i &= |b_i|, & \Omega_i &= \arctg(\operatorname{Im} z_i / \operatorname{Re} z_i), \\ \alpha_i &= (\ln |z_i|) / \Delta t, & \theta_i &= \arctg(\operatorname{Im} b_i / \operatorname{Re} b_i). \end{aligned} \quad (20)$$

Окончательно метод Прони дает:
оценку данных

$$s(t) = \sum_{i=1}^{2p} A_i \exp(\alpha_i t) \exp(j(\omega_i t + \theta_i)), \quad (21)$$

ее фурье-образ

$$S(\omega) = \sum_{i=1}^{2p} A_i \exp(j\theta_i) \frac{1}{(\omega - \omega_i) - j\alpha_i} \quad (22)$$

и спектр энергии

$$E(\omega) = |S(\omega)|^2. \quad (23)$$

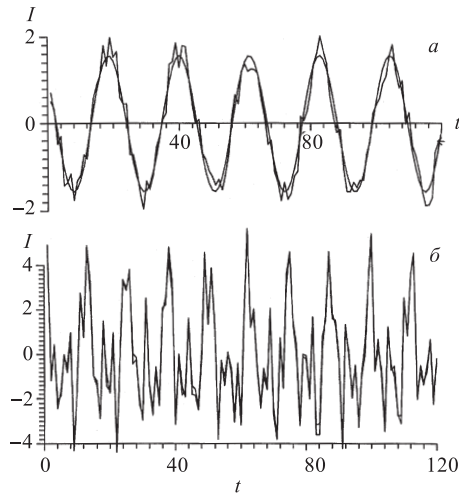


Рис. 3. Результаты моделирования метода Прони на моделях: *а*) простой гармонический сигнал $1,600 \cos(2,021t + 1,0798)$ и его оценка $1,594 \cos(2,021t + 1,0930)$; *б*) смесь гармоник $2,000 \cos(0,5054t + 0,2670) + 2,000 \cos(1,015t + 0,5422) + 1,000 \cos(1,844t + 0,9848) + 1,6 \cos(2,465t + 1,3167)$ и ее оценка $1,932 \cos(0,5054t + 0,2821) + 1,957 \cos(1,015t + 0,5887) + 0,9216 \cos(1,844t + 0,9716) + 1,636 \cos(2,465t + 1,2530)$

с шумом 6% — оценка $1,932 \cos(0,5054t + 0,2821) + 1,957 \cos(1,015t + 0,5887) + 0,9216 \cos(1,844t + 0,9716) + 1,636 \cos(2,465t + 1,2530)$.

Приведенные результаты демонстрируют хорошее соответствие исходным сигналам. В этих и других вычислительных экспериментах показано, что наиболее точно метод Прони определяет частоты, наименее точно — фазы компонент и коэффициенты затухания компонент, а спектральные параметры высокочастотных компонент имеют лучшую точность по сравнению с низкочастотными.

В методах АД априорная информация о процессе, заданная в виде локальной связи данных, носит обобщенный характер, позволяет сформировать модель осциллирующего сигнала, определить его состав и спектральные параметры компонент без каких-либо специальных априорных предположений. Уточнение полученных параметров в рамках сформированной модели возможно, а в ряде случаев и целесообразно с применением других методов, например, таких мощных аппроксимационных методов, как метод Монте-Карло [21].

Выводы. 1. Методы спектрального анализа информативны при обработке данных интерференционных экспериментов, позволяют получить сведения о характеристиках связанных состояний систем: атомов, ядер, частиц.

Как уже отмечалось, преимуществом метода Прони является расширение класса функций, используемых в качестве базиса спектрального разложения. Формализм метода допускает включение в этот класс экспоненциально спадающих и нарастающих сигналов, что позволяет распространить методы спектрального анализа на монотонно меняющиеся и переходные случайные процессы. Возможность вычислять амплитуды, коэффициенты затухания, частоты и особенно фазы колебаний с высокой точностью может оказаться полезной при интерполяции и экстраполяции сигналов, а в некоторых случаях способствовать решению весьма трудной проблемы прогнозирования процессов.

Моделирование метода Прони проведено с помощью авторской программы, написанной на Фортране. На рис.3 представлены исходные сигналы в виде одиночной синусоиды (*а*) и смеси четырех различных гармоник с аддитивным белым шумом (*б*) и результаты их реконструкции по полученным при разложении параметрам. Для синусоиды, заданной формулой $1,600 \cos(2,021t + 1,0798)$ с аддитивным шумом 15% получена оценка $1,594 \cos(2,021t + 1,0930)$, а для смеси гармоник $2,000 \cos(0,5054t + 0,2670) + 2,000 \cos(1,015t + 0,5422) + 1,000 \cos(1,844t + 0,9848) + 1,6 \cos(2,465t + 1,3167)$ с шумом 6% — оценка $1,932 \cos(0,5054t + 0,2821) + 1,957 \cos(1,015t + 0,5887) + 0,9216 \cos(1,844t + 0,9716) + 1,636 \cos(2,465t + 1,2530)$.

2. Методы на основе АМД сочетают сильные стороны аппроксимационных и фильтрационных методов при определении спектральных характеристик сигналов.

3. ММЭ-оценки спектров мощности реальных сигналов отличаются информативностью и применимы для анализа гармонических сигналов и ангармонических распределений. Спектральные параметры осциллирующих сигналов, определенные методом Прони, имеют высокую точность и позволяют представить экспериментальные данные суперпозицией базисных функций.

СПИСОК ЛИТЕРАТУРЫ

1. Хрипович И. Б. Несохранение четности в атомных явлениях. М.: Наука, 1981.
2. Соколов Ю. Л. Интерференционный метод измерения параметров атомных состояний // УФН. 1999. Т. 69. С. 559–583.
3. Подгорецкий М. И., Хрусталева О. А. О некоторых интерференционных явлениях в квантовых переходах // УФН. 1963. Т. 81. С. 217–247.
4. Александров Е. Б., Хвостенко Г. И., Чайка М. П. Интерференция атомных состояний. М.: Наука, 1991.
5. Василенко Г. И. Теория восстановления сигналов. М.: Сов. радио, 1979.
6. Kay S. K., Marple S. L. Spectrum Analysis — A Modern Perspective // Proc. of the IEEE. 1981. V. 69, No. 11. P. 1380–1418.
7. Babic H., Temes G. C. Optimum Low Order Windows for Discrete Fourier Transform System. 1976. V. ASSP-24. P. 512–517.
8. Серебрянников М. Г., Первозванский А. А. Выявление скрытых периодичностей. М.: Наука, 1965.
9. Астафьева Н. М. Вейвлет-анализ: основы теории и примеры применения // УФН. 1996. Т. 166. С. 1145–1170.
10. Burg J. R. Maximum Entropy Spectral Analysis // Proc. of the 37th Meeting Society of Exploration Geophysicist, Oklahoma City, OK, 1967. P. 31.
11. Pisarenko V. F. On the Estimation of Spectra by Means of Nonlinear Functions of Covariance Matrix // Geophys. J. Royal Astronom. Soc. 1972. V. 28. P. 511–513.
12. Prony G. R. B. Essai Experimental and Analytique, etc. Paris, 1975. V. 1. Cahier 2. P. 24–76.
13. Akaike H. Power Spectrum Estimation through Autoregressive Model Fitting // Ann. Inst. Statist. Math. 1969. V. 21. P. 243–247.
14. Gybenko B. G. Round of Error Propagation in Durbin's Levinson's and Trench's Algorithms // Rec. 1979 IEEE Conf. Acoustic, Speech, and Signal Proc. P. 498–501.
15. Hildenbrand F. B. Introduction to Numerical Analysis. Mc.Graw Hill, 1956. Ch. 9.
16. Злоказов В. Б. Анализ скрытых ангармонических периодичностей. Препринт ОИЯИ Р11-89-378. Дубна, 1989.

17. *Протодьяконов М. М.* Свойства породообразующих минералов и их электронное строение. М.: Наука, 1965.
18. *Руднев С. Ф.* Взаимодействие сейш и стокового течения в Онежском озере. 19 с. Деп. в ВИНТИ 29.03.89, № 2060-B89.
19. *Barrodale I., Erickson R. E.* Algorithms for Least-Square Linear Prediction and Maximum Entropy Spectral Analysis // *Geophysics*. 1980. V. 45, No. 3. P. 420.
20. *Farden D. S.* Solution of a Toeplitz Set of Linear Equations // *IEEE Trans. Antennas Propagat.* 1976. V. AP-24. P. 906–908.
21. *Соболь И. М.* Числовые методы Монте-Карло. М.: Наука, 1973.

Получено 6 апреля 2001 г.