

УДК 551.510.532:544.772:517.5

## МЕТОДИЧЕСКИЕ АСПЕКТЫ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ НА ПРИМЕРЕ ОБЩЕГО СОДЕРЖАНИЯ АЭРОЗОЛЯ В СТРАТОСФЕРЕ

С. И. Долгий<sup>1</sup>, А. В. Невзоров<sup>1</sup>, Н. С. Сальникова<sup>1</sup>,  
А. В. Ельников<sup>2</sup>, С. М. Сысоев<sup>2</sup>, В. А. Логинов<sup>2</sup>

<sup>1</sup>Институт оптики атмосферы им. В. Е. Зуева  
Сибирского отделения Российской академии наук, г. Томск

<sup>2</sup>Сургутский государственный университет  
andreyeln@mail.ru

В статье рассмотрены методические аспекты статистического анализа не полностью заполненных временных рядов и интерпретации полученных при этом результатов на примере анализа интегрального коэффициента обратного аэрозольного рассеяния стратосферного аэрозоля.

*Ключевые слова:* лидарное зондирование, стратосферный аэрозоль, временной ход, функция распределения, быстрое преобразование Фурье.

## METHODOLOGICAL ASPECTS OF TIME SERIES ANALYSIS THE CASE OF AEROSOL TOTAL CONTENTS IN THE STRATOSPHERE

S. I. Dolgy<sup>1</sup>, A. V. Nevzorov<sup>1</sup>, N. S. Salnikova<sup>1</sup>,  
A. V. Elnikov<sup>2</sup>, S. M. Sysoev<sup>2</sup>, V. A. Loginov<sup>2</sup>

<sup>1</sup>V. E. Zuev Institute of Atmospheric Optics of Siberian Branch  
of the Russian Academy of Sciences, Tomsk

<sup>2</sup>Surgut State University  
andreyeln@mail.ru

The article deals with the methodological aspects of the statistical analysis of incomplete time series and interpretation of the results obtained by the example of the analysis of the integral aerosol backscattering coefficient of the stratospheric aerosol.

*Keywords:* lidar sounding, stratospheric aerosol, time dependence, distribution function, fast Fourier transform.

**Введение.** Оптически активные компоненты воздуха являются одним из факторов, формирующих погоду и климат Земли, поэтому требуется как можно больше объективных исходных данных о них для создания более точных и адекватных моделей прогноза погоды и изменения климата. Оптической активностью обладает и аэрозоль – твердые частицы, взвешенные в воздухе, что и определяет актуальность представленных в статье результатов наблюдений. Несомненный интерес представляет и то, что в статье анализируются данные, полученные с помощью дистанционного лазерного зондирования атмосферы – метода известного с 60-х годов прошлого столетия [1, 2]. Этот метод позволяет мгновенно получать информацию о вертикальном распределении аэрозоля по всей толще атмосферы, но его возможности сильно ограничены наличием облачности. Оптически плотные облака экранируют посылаемые в атмосферу импульсы оптического диапазона, поэтому обратнорассеянное излучение с высоты 15–30 км получить не удастся. В связи с этим в регулярно сформированном временном ряду интегрального содержания аэрозоля в стратосфере имеются временные интервалы, на которых данные отсутствуют. Другой причиной наличия «пробелов» в рядах – это необходимость регулярного технологического обслуживания сложного лидарного оборудования [3].

Тем не менее результаты натуральных наблюдений, полученные с использованием данного метода, имеют важное геофизическое значение.

Определяемым параметром при лазерном зондировании аэрозоля является коэффициент обратного аэрозольного рассеяния  $\beta_{\pi}^A(H)$  [1, 2, 4], на основе которого и был сформирован анализируемый в статье временной ряд интегрального (суммарного) содержания стратосферного аэрозоля ( $V_{\pi}^A$ ) над Томском в период 2000–2016 годов на основе данных, полученных на Сибирской лидарной станции (далее – СЛС) Института оптики атмосферы им. В. Е. Зуева Сибирского отделения Российской академии наук. Алгоритм формирования анализируемого ряда из вертикальных профилей коэффициентов обратного аэрозольного рассеяния тривиальный и выражается формулой:

$$V_{\pi}^A [\text{ср}^{-1}] = \int_{H_{\text{MIN}}=15 \text{ км}}^{H_{\text{MAX}}=30 \text{ км}} \beta_{\pi}^A(h) \cdot dh; \quad \text{в дискретном виде} \quad V_{\pi}^A [\text{ср}^{-1}] = \sum_{H_{\text{MIN}}=15 \text{ км}}^{H_{\text{MAX}}=30 \text{ км}} \beta_{\pi}^A(H). \quad (1)$$

Здесь диапазон  $H_{\text{MAX}} \div H_{\text{MIN}} = 30 \div 15 \text{ км}$  – это интервал высот, по которому производится суммирование коэффициента обратного аэрозольного рассеяния. Полученное суммарное значение усредняется подекадно.

Статья является продолжением обсуждения проблемы, поднятой в публикации [5].

**Временной ряд  $V_{\pi}^A$  в период 2000–2016 гг. и его ход по годам.** Временной ход коэффициента  $V_{\pi}^A$  за анализируемый период приведен на рис. 1, отсутствие данных в таком представлении ряда практически незаметно. Тонкой линией на этом же рисунке изображена квадратичная аппроксимирующая кривая, полученная с помощью метода наименьших квадратов:

$$V_{\pi}^A(i) = 1,22 \cdot 10^{-4} + 2,87 \cdot 10^{-7} i - 1,95 \cdot 10^{-10} i^2; \quad (2)$$

где  $i$  – значение ряда.

Ее ход свидетельствует о неуклонном увеличении (правда незначительном) общего содержания аэрозоля в стратосфере. Все имеющиеся значения ряда укладываются в интервал  $3\sigma$  ( $\sigma$  – стандартное отклонение анализируемого ряда). Это позволяет утверждать, что все они, и даже те, которые выделяются из общего хода  $V_{\pi}^A$ , не являются случайными выбросами. Природа их появления обусловлена извержениями вулканов, в результате которых большое количество аэрозоля и аэрозолеобразующих газов достигает атмосферы и переносится по всему земному шару. Извержения вулканов, наблюдаемые в анализируемый период, приведены в статье [5].

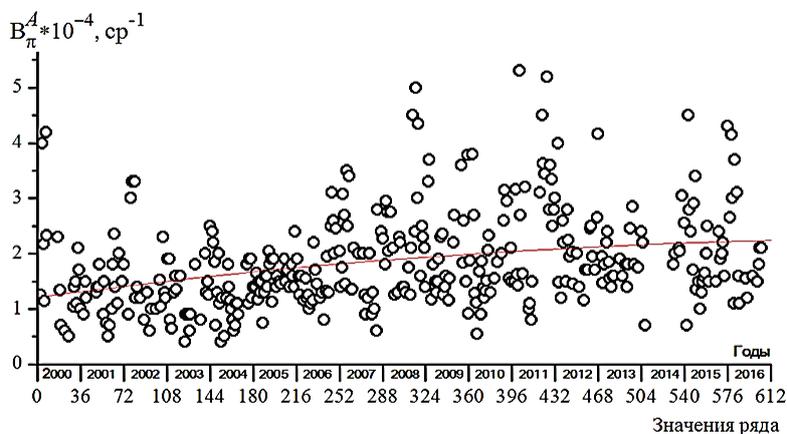


Рис. 1. Временной ряд  $V_{\pi}^A$  в период 2000–2016 гг. и его аппроксимация полиномом 2-го порядка

Проверка анализируемого ряда на нормальность распределения (по модифицированному критерию Шапиро – Уилка (Shapiro–Wilk) [6, 7]) показала, что гипотеза о его нормальности может быть отброшена при уровне значимости 0,05, т. е. ряд хорошо аппроксимируется нормальным распределением.

Наполненность анализируемого ряда данными иллюстрирует рис. 2. На этом рисунке весь ряд представлен годовыми отрезками, здесь отчетливо видны участки ряда, на которых данные отсутствуют. Тем не менее при усреднении по всем 16 годам удалось получить средний подекадный годовой ход  $V_{\pi}^A$  (нижний график с правой стороны на рис. 2), в котором отсутствие данных приводит к меньшей статистической обеспеченности каждого значения.

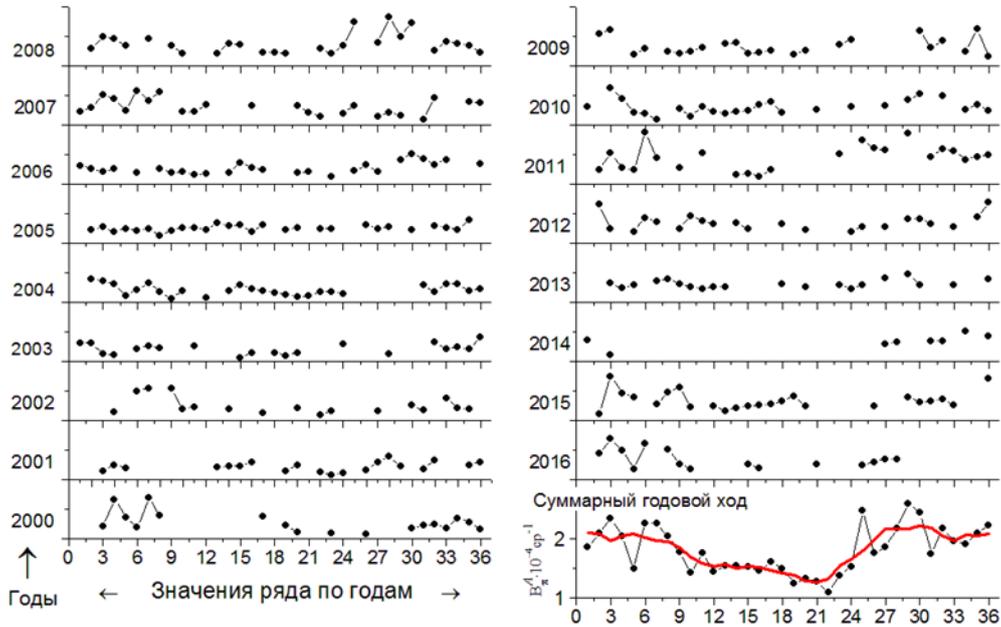


Рис. 2. Временные ходы  $V_{\pi}^A$  по соответствующим годам и их средний годовой ход

**Анализ гистограмм распределения вариаций временного ряда  $V_{\pi}^A$ .** Гистограммы – это наглядное представление функции плотности вероятности случайной величины, в нашем случае вариаций исследуемого ряда  $V_{\pi}^A$  [8]. Они строятся на основе случайных выборок исследуемой величины, и, следовательно, отсутствие данных во временном ряде на получаемые результаты влияния также не оказывает.

Как уже отмечалось в статье [5], вариации временного ряда  $V_{\pi}^A$  обусловлены тремя независимыми процессами: 1) естественными вариациями стратосферного аэрозольного слоя как природного образования; 2) поступлением аэрозоля в стратосферу от вулканических извержений; 3) годовыми вариациями, которые обусловлены сезонными изменениями и являются детерминированными. Следовательно, анализируемый ряд можно представить как результат трех процессов со своими собственными модами.

При анализе временных рядов на нормальность от детерминированности, как правило, стараются избавиться. Но в статье приводится сравнение результатов проверки на нормальность временных рядов  $V_{\pi}^A$  как при наличии годового хода (т. е. детерминированной составляющей), так и без него. Удаление этой составляющей из исходного ряда  $V_{\pi}^A$  было реализовано вычитанием из соответствующих декадных значений  $V_{\pi}^A$  каждого года среднегодовых значений, которые были предварительно сглажены по 5 точкам (на рис. 2 эти значения представлены жирной линией), что и привело исходный ряд к ряду с нулевым средним значением.

Результаты проверки на нормальность распределений вариаций исходного временного ряда, как с сезонным ходом, так и после его вычитания, в предположении наличия в нем одной моды представлены на рис. 3. На рис. 3а его вариации изображены в виде гистограмм, а на рис. 3б в виде сплошных ломаных кривых. Плавные кривые на рис. 3б представляют их аппроксимацию нормальным (Гауссовым) распределением:

$$y(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right]. \quad (3)$$

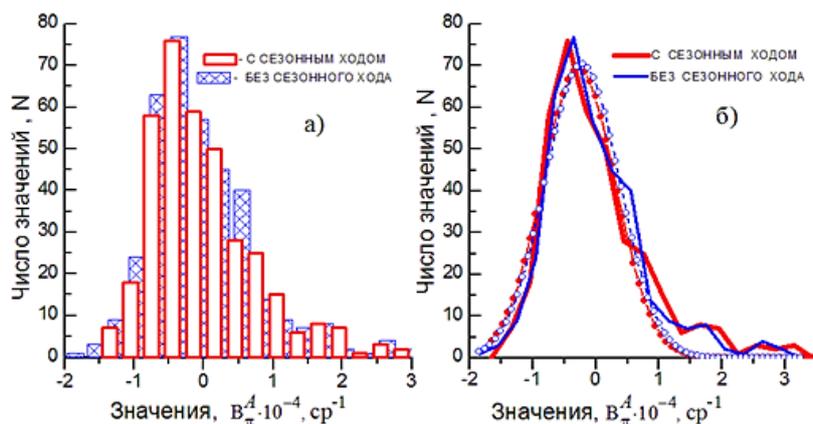


Рис. 3. Гистограммы распределения вариаций рядов  $V_{\pi}^A$ :

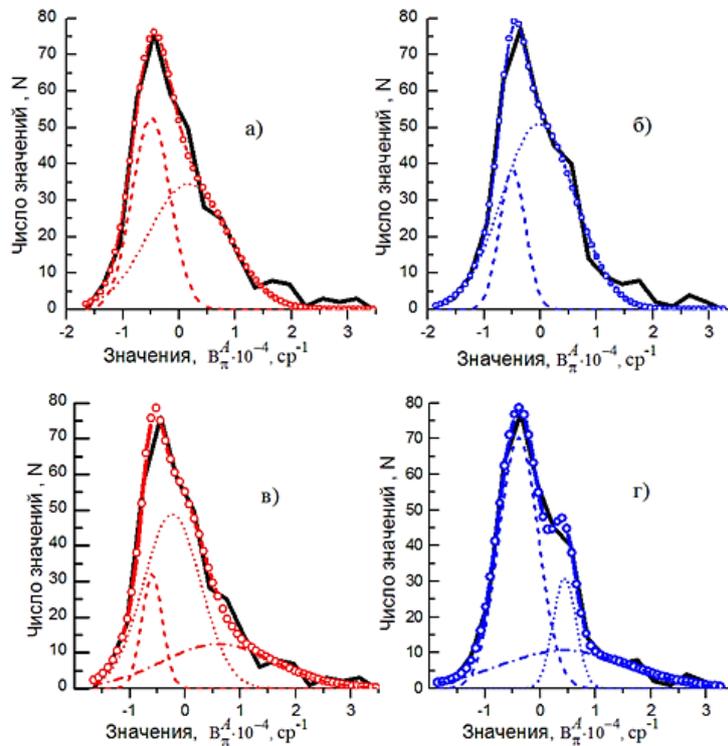
- а) эмпирические гистограммы распределения вариаций временных рядов  $V_{\pi}^A$  с годовым ходом и после его вычитания;  
б) эти же гистограммы в виде ломаных кривых и аппроксимирующие их нормальные распределения (красная кривая со сплошными точками – при наличии в ряде годового хода, синие со светлыми точками – после вычитания годового хода).

Видно, что гистограммы временных рядов (с наличием сезонного хода и после его вычитания) отличаются, но эти отличия вносят несущественные изменения как в форму аппроксимирующих кривых (рис. 3б), так и в результаты количественного сравнения эмпирических и теоретических распределений, проверка которых на подобие проводилась по коэффициенту корреляции и критерию  $\chi^2$  (табл. 1). Это свидетельствует о том, что детерминированные изменения из года в год (годовые изменения) вносят значительно меньший вклад во временной ход коэффициента  $V_{\pi}^A$ , по сравнению с естественными вариациями и вариациями, обусловленными извержениями вулканов.

На рис. 4 представлена аппроксимация исследуемых рядов суммой двух (рис. 4а, 4б) и трех (рис. 4в, 4г) нормальных распределений, что означает следующее: исходный ряд состоит из двух или трех независимых процессов, которые формируют, соответственно, две или три различные моды. Ряды с наличием годовых вариаций представлены на рис. 4а и 4в, а на рис. 4б и 4г – после их удаления. Верхняя и нижняя части рисунка содержат кривые распределения экспериментальных значений гистограммы (сплошная линия), две моды (рис. 4а и 4б) или три моды (рис. 4в и 4г), суммы их значений.

Явное отличие в распределении вариаций анализируемых рядов при аппроксимации их двумя модами следующее: при наличии в ряде годового хода в распределении вариаций доминирует мода № 1 (мелкомасштабная) с максимумом, приходящимся на значение  $V_{\pi}^A \approx -5 \cdot 10^{-5} \text{ ср}^{-1}$ . Максимальное значение моды № 2 (крупномасштабной) при значении  $V_{\pi}^A \approx 3 \cdot 10^{-5} \text{ ср}^{-1}$  примерно в полтора раза меньше. При удалении сезонных вариаций в суммарном распределении доминантной становится крупномасштабная мода № 2.

Параметры мод выбирались из условия наибольшего совпадения экспериментальных кривых и суммарного распределения. Результаты сравнения эмпирических и теоретических кривых в цифровом виде также представлены в табл. 1.



**Рис. 4.** Кривые гистограмм распределения вариаций временных рядов  $V_{\pi}^A$  (15 ÷ 30 км) и их аппроксимация в предположении 2 мод (верхняя часть) и 3 мод (нижняя часть): сплошные черные линии – кривая распределения вариаций исходных рядов; линии со светлыми кружками – результирующая кривая сложения мод; пунктирные линии – 1-ая мода; точки – 2-ая мода; штрих-пунктир – 3-я мода.

Результаты 3-модовой аппроксимации исходных временных рядов представлены на рис. 4в и 4г, а результаты количественного сравнения вариаций эмпирических рядов и теоретических модельных также представлены в табл. 1.

Таблица 1

**Параметры мод и результаты сравнения вариаций для ряда**

	С годовыми вариациями		После вычитания годовых вариаций	
	Параметры мод [ср-1]	Результаты сравнения	Параметры мод [ср-1]	Результаты сравнения
одна мода	<p><b>Мода :</b> Максимум = <math>-2,72 \cdot 10^{-5}</math>;                      Полуширина = <math>1,1 \cdot 10^{-4}</math>;                      Значение в максимуме 69,4.</p>	<p><math>\chi^2 = 52,0</math>  <math>R^2 = 0,921</math></p>	<p><b>Мода :</b> Максимум = <math>-2,14 \cdot 10^{-5}</math>;                      Полуширина = <math>1,18 \cdot 10^{-4}</math>;                      Значение в максимуме 70,4.</p>	<p><math>\chi^2 = 43,8</math>  <math>R^2 = 0,937</math></p>
две моды	<p><b>Мода № 1 :</b> Максимум = <math>-4,99 \cdot 10^{-5}</math>;                      Полуширина = <math>6,77 \cdot 10^{-5}</math>;                      Значение в максимуме 52,9.</p> <p><b>Мода № 2 :</b> Максимум = <math>1,63 \cdot 10^{-5}</math>;                      Полуширина = <math>1,41 \cdot 10^{-4}</math>;                      Значение в максимуме 34,5.</p>	<p><math>\chi^2 = 15,8</math>  <math>R^2 = 0,981</math></p>	<p><b>Мода № 1 :</b> Максимум = <math>-4,93 \cdot 10^{-5}</math>;                      Полуширина = <math>4,54 \cdot 10^{-5}</math>;                      Значение в максимуме 38,7.</p> <p><b>Мода № 2 :</b> Максимум = <math>-2,14 \cdot 10^{-5}</math>;                      Полуширина = <math>1,18 \cdot 10^{-4}</math>;                      Значение в максимуме 70,4.</p>	<p><math>\chi^2 = 13,1</math>  <math>R^2 = 0,985</math></p>

Окончание табл. 1

	С годовыми вариациями		После вычитания годовых вариаций	
	Параметры мод [ср-1]	Результаты сравнения	Параметры мод [ср-1]	Результаты сравнения
ТРИ МОДЫ	<b>Мода № 1</b> : Максимум = $-5,97 \cdot 10^{-5}$ ; Полуширина = $3,71 \cdot 10^{-5}$ ; Значение в максимуме 32, 2.	$\chi^2 = 9,74$ $R^2 = 0,991$	<b>Мода № 1</b> : Максимум = $-3,91 \cdot 10^{-5}$ ; Полуширина = $7,16 \cdot 10^{-5}$ ; Значение в максимуме 70, 4.	$\chi^2 = 3,56$ $R^2 = 0,997$
	<b>Мода № 2</b> : Максимум = $-2,36 \cdot 10^{-5}$ ; Полуширина = $1,03 \cdot 10^{-4}$ ; Значение в максимуме 49, 0.		<b>Мода № 2</b> : Максимум = $4,38 \cdot 10^{-5}$ ; Полуширина = $4,09 \cdot 10^{-5}$ ; Значение в максимуме 31, 4.	
	<b>Мода № 3</b> : Максимум = $6,036 \cdot 10^{-5}$ ; Полуширина = $2,04 \cdot 10^{-4}$ ; Значение в максимуме 12, 4.		<b>Мода № 3</b> : Максимум = $4,44 \cdot 10^{-5}$ ; Полуширина = $2,34 \cdot 10^{-4}$ ; Значение в максимуме 10, 7.	

Анализ результатов, представленных в табл. 1, показывает, что удаление из временного ряда детерминированной составляющей (сезонных колебаний) приводит к тому, что экспериментальные кривые распределения находятся в более удовлетворительном согласии с их модельными представлениями. Кроме того, увеличение числа мод при аппроксимации экспериментальных данных также имеет тенденцию к лучшей аппроксимации экспериментальной кривой и ее модельным представлением.

**Частотные составляющие временных рядов  $V_{\pi}^A$ .** Для выявления частотных составляющих временных ходов  $V_{\pi}^A$  обоих рядов (с наличием годового хода и при его отсутствии) было использовано быстрое преобразование Фурье (далее – БПФ). Но прежде ряды с помощью полинома 2-го порядка были приведены к нулевому среднему и им придана форма окна Хемминга [9, 10]:

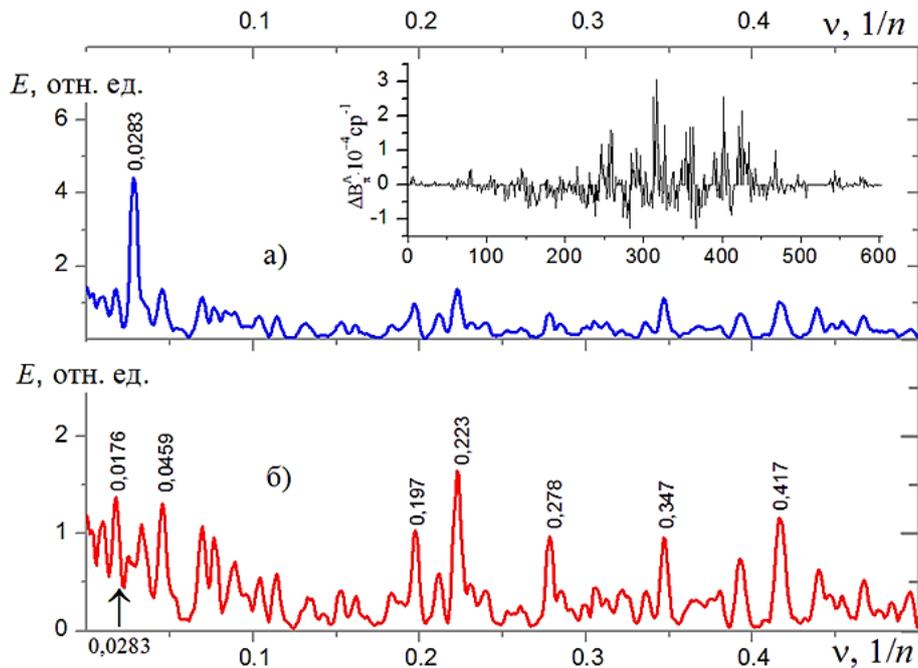
$$w(n) = 0,53836 - 0,46164 \cdot \cos\left(\frac{2\pi n}{N-1}\right), \quad (4)$$

где  $n$  – номер соответствующего значения ряда;  
 $w(n)$  – вес соответствующего значения ряда;  
 $N$  – длина ряда.

Последняя операция необходима для уменьшения процесса «перекачки» энергии, приходящейся на нулевую гармонику, в соседние гармоники более высокой частоты [9, 10]. Вид сформированного для Фурье-анализа временного ряда с наличием годового хода представлен в верхней правой части рис. 5а, основой же рисунка является Фурье-спектр самого временного ряда, сглаженный по пяти точкам. На рисунке видно, что в спектре доминирует гармоника, приходящаяся на частоту  $\nu = 0,02832$  или 35,3 декады. Учитывая число дней в году: февраль – 28 или 29 дней, 7 месяцев – по 31 дню, а 4 месяца – по 30 дней, получаем, что среднегодовая длительность декады – 10,34 дня и максимум по временной шкале соответствует периоду года  $35,3 \times 10,34 = 365$  дней, т. е. как раз гармоника, представляющая годовые вариации. Все другие гармоники, наблюдаемые в спектрах, имеют существенно меньшие значения энергий.

В более контрастном виде эти гармоники изображены на рис. 5б. На нем приведен Фурье-спектр временного ряда  $V_{\pi}^A$  после удаления из него детерминированной компоненты, т. е. вариаций годового хода, также сглаженный по пяти точкам. Местоположение гармоники, появление которой на предыдущем графике (рис. 5а) было обусловлено сезонными колебаниями, на этом графике (рис. 5б) отмечено стрелкой, и оно приходится на локальное минимальное значение. Это указывает на эффективность примененной процедуры удаления сезонных колеба-

ний и приводит к более контрастному выделению гармоник существенно меньших амплитуд. На рис. 5б сверху этих гармоник приведены их частоты, значения которых сведены в табл. 2 и переведены во временной масштаб. Анализ временных параметров этих гармоник не позволяет соотнести их с каким-либо геофизическим процессом, который мог бы являться причиной возникновения этих гармоник.



**Рис. 5. Фурье-спектры энергий временного ряда  $V_{\pi}^A$  :**  
 а) при наличии в нем годового хода (в верхней части рисунка исходный ряд имеющий форму окна Хемминга); б) после его вычитания

Таблица 2

**Значения гармонических составляющих спектра (по рис. 5б)**

№	Частота	Декады	Сутки
1	0,0175	56,9	588
2	0,0459	21,8	225
3	0,197	5,1	53
4	0,223	4,5	47
5	0,278	3,6	37
6	0,347	2,9	30
7	0,417	2,4	25

Наиболее вероятная причина появления их в спектре носит методический характер – это не полностью заполненный временной ряд. На границах (в начале и конце) временного ряда использование окна Хемминга привело к минимизации негативных эффектов применения БПФ к резким изменениям анализируемых данных. Внутри временного ряда резкие изменения значений анализируемых данных в результате отсутствия данных привело к формированию в Фурье-спектре мнимых гармоник.

Однако в физике атмосферы существует понятие квазидвухлетнего цикла, которое формируется изменением глобального направления движения воздушных масс. Термин «квази» означает, что колебания не имеют такой регулярной структуры, которая наблюдается, например, в годовых вариациях, и к тому же имеют неодинаковую амплитуду. Квазидвухлетние колебания проявляются в общем содержании озона и должны бы присутствовать

и общем содержании аэрозоля. Однако в Фурье-образе временного хода  $V_{\pi}^A$  эти колебания не проявились.

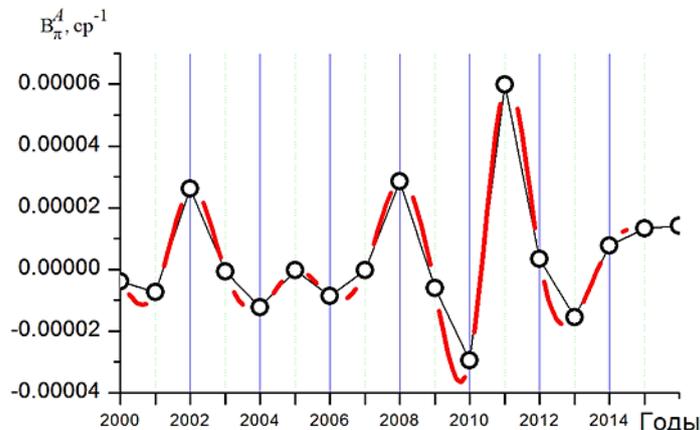


Рис. 6. Ряд годовых значений интегрального коэффициента обратного аэрозольного рассеяния

Наличие квазидвухлетних колебаний во временном ходе интегрального коэффициента обратного рассеяния удалось выявить после вычисления его средних годовых значений и представления их в виде ряда. Полученный временной ряд содержит 17 точек, на рис. 6 он изображен после приведения к нулевому среднему. Здесь же приведена и кривая сплайн-аппроксимации этого ряда. Видно, что годовые суммарные значения подвержены случайным по амплитуде колебаниям, период которых близок к двум годам.

**Заключение.** В статье продемонстрировано, что все вариации временного ряда интегрального содержания аэрозоля в стратосфере не являются артефактными выбросами, это подтвердил и дальнейший анализ временного ряда  $V_{\pi}^A$  на нормальность. Анализ проводился в предположении наличия в ряде нескольких независимых случайных процессов. Было продемонстрировано, что увеличение количества мод, отражающего наличие независимых случайных процессов во временном ряде, приводит к большему соответствию эмпирических значений распределения вариаций ряда теоретическим.

Результаты Фурье-преобразования временного ряда выявили (что и следовало ожидать) доминирующий вклад в его ход сезонных вариаций, при удалении которых в Фурье-спектре проявляются гармоники, обусловленные пропусками экспериментальных данных. Показана эффективность и тривиальных методов анализа эмпирических данных, которые позволили выявить наличие квазидвухлетних колебаний во временном ходе интегрального коэффициента обратного аэрозольного рассеяния.

Работа выполнена при финансовой поддержке гранта РФФИ 18-47-860014.

### Литература

1. Лазерный контроль атмосферы / под ред. Э. Д. Хинкли. М. : Мир, 1979. 416 с.
2. Межерис Р. Лазерное дистанционное зондирование. М. : Мир, 1987. 550 с.
3. Ельников А. В., Кавкьянов С. И., Креков Г. М., Маричев В. Н. Процедура обработки сигналов лазерного зондирования стратосферы // Оптика атмосферы. 1989. Т. 2. № 5. С. 537–540.
4. Матвиенко Г. Г., Балин Ю. С., Бобровников С. М., Романовский О. А., Коханенко Г. П., Самойлова С. В., Пеннер И. Э., Горлов Е. В., Жарков В. И., Садовников С. А., Харченко О. В., Яковлев С. В., Баженов О. Е., Бурлаков В. Д., Долгий С. И., Макеев А. П., Невзоров А. А., Невзоров А. В. Сибирская лидарная станция: аппаратура и результаты / под ред. Г. Г. Матвиенко. Томск : Изд-во ИОА СО РАН. 2016. 414 с.

5. Долгий С. И., Невзоров А. В., Ельников А. В., Сысоев С. М. Статистический анализ временного ряда общего содержания аэрозоля в стратосфере над Западной Сибирью // Вестник кибернетики. 2017. № 3. С. 40–46.
6. Кобзарь А. И. Прикладная математическая статистика. М. : Физматлит, 2006. 816 с.
7. ГОСТ Р ИСО 5479-2002. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения. М. : Изд-во стандартов. 2002. 30 с.
8. Кремер Н. Ш. Теория вероятностей и математическая статистика. М. : ЮНИТИ, 2004. 573 с.
9. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. М. : Книга, 2012. 541 с.
10. Подкорытова О. А., Соколов М. В. Анализ временных рядов / 2-е изд., перераб. и доп. М. : Юрайт, 2018. 267 с.