

УДК 519.72

МОРФОЛОГИЧЕСКИЕ И ДРУГИЕ МЕТОДЫ ИССЛЕДОВАНИЯ ПОЧТИ ЦИКЛИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ НА ПРИМЕРЕ РЯДОВ КОНЦЕНТРАЦИИ CO_2 ¹⁾

© 2021 г. В. К. Авилон¹, В. С. Алешновский², А. В. Безрукова², В. А. Газарян^{2,4}, Н. А. Зюзина², Ю. А. Курбатова¹, Д. А. Тарбаев², А. И. Чуличков^{2,3,*}, Н. Е. Шапкина^{2,5}

¹ 119071 Москва, Ленинский пр-т, 33, ИПЭЭ им. А.Н. Северцова РАН, Россия

² 119991 Москва, Ленинские горы, 1, стр. 2, МГУ им. М.В. Ломоносова, физический факультет, Россия

³ 119017 Москва, Пыжевский пер., 3, ИФА им. А.М. Обухова РАН, Россия

⁴ 125993 Москва, Ленинградский пр-т, 49, Финансовый университет при правительстве РФ, Россия

⁵ 125412 Москва, ул. Ижорская, 13, ИТПЭ РАН, Россия

*e-mail: achulichkov@gmail.com

Поступила в редакцию 26.11.2020 г.

Переработанный вариант 26.11.2020 г.

Принята к публикации 11.03.2021 г.

На основе методов морфологического анализа, развитых под руководством Ю.П. Пытьева, предложен метод фильтрации временных рядов, позволяющих выделить почти циклическую составляющую, длительность цикла которой не является постоянной, а значения элементов ряда внутри циклов изменчивы. Эффективность подхода иллюстрирована на примере декомпозиции временного ряда концентрации CO_2 в атмосфере Земли. Остаток ряда после фильтрации составляющей ряда, моделирующей суточную изменчивость, становится стационарным, что позволяет для его дальнейшего исследования использовать методы математической статистики и фурье-анализа. Верификация полученных результатов проводилась путем сравнения с результатами фурье-анализа. Для исследования цикличности с периодом, большим суток, используются фурье-разложение и вейвлет-анализ исходного ряда. Библиография: 16. Фиг. 8.

Ключевые слова: цифровая обработка сигналов, квазипериодические сигналы, декомпозиция, форма сигнала, фурье-анализ, вейвлет-анализ.

DOI: 10.31857/S0044466921070048

1. ВВЕДЕНИЕ

В настоящее время обработка и анализ результатов длительных измерений, проводимых через заданные промежутки времени, вызывают интерес в различных сферах, от экономики до метеорологии. Для этого используется достаточно обширный арсенал методов как математически обоснованных (например, метод Фурье для стационарных рядов, вейвлет-преобразования для нестационарных), так и эвристических [1]. В этой ситуации интерес представляет разработка методов анализа временных рядов, основанных на достаточно общих математических моделях, дающих оптимальные решения на широком классе задач.

В данной работе для исследования цикличности временных рядов в первую очередь используется морфологический подход, развиваемый в школе Ю.П. Пытьева [2]–[4]. Предлагается способ морфологической фильтрации, позволяющий выделить в сигнале составляющую, которая представляет собой последовательность фрагментов, схожих по форме, но меняющихся по длительности и амплитуде в некоторых заданных пределах. Такая фильтрация может быть полезна для декомпозиции исходного временного ряда на “почти циклическую” составляющую, связанную с сезонностью, и остаток, который может быть подвергнут дальнейшему исследованию с целью выделения дополнительных циклических составляющих.

Эффективность предложенного подхода исследована на примере анализа данных измерения концентрации CO_2 , представленных в виде временных рядов. В результате морфологической

¹⁾ Работа выполнена при частичной финансовой поддержке РФФИ (код проекта № 19-29-09044).

фильтрации ряда получен стационарный остаток, его дальнейшее исследование проведено методами фурье-анализа и математической статистики. В результате этого анализа были выделены короткопериодные составляющие ряда. Для верификации полученных результатов эти же данные исследовались методами фурье-анализа и вейвлет-анализа, было проведено сравнение результатов.

2. МОРФОЛОГИЧЕСКИЙ ПОДХОД К АНАЛИЗУ СИГНАЛОВ

На практике, в частности, при дистанционном зондировании атмосферы, при изучении временных рядов метеорологических параметров и др., анализируемый сигнал имеет вид суммы почти повторяющихся фрагментов, схожих по форме с сигналами иной формы – со стационарным или нестационарным сигналами, с широкополосной шумовой составляющей и др. Задача состоит в разделении суммы этих сигналов на составляющие и анализе параметров каждой из них.

Широко распространенный метод фурье-анализа не всегда удобен, так как позволяет представить исследуемый сигнал в виде суммы ортогональных слагаемых синусоидальной формы, в то время как при нестационарном характере сигналов каждый из участков сигнала может иметь свои характерные особенности. В этом случае более удобным является вейвлет-анализ, позволяющий исследовать динамическую структуру сигнала [5]. Вейвлет-анализ также дает возможность представить участок сигнала в виде линейной комбинации масштабируемых базисных функций заданной формы, однако, желательно, чтобы форма слагаемых была адаптирована к анализируемому сигналу.

Более гибкую модель сигнала заданной формы дает морфологический подход [2]–[4]. Морфологические методы анализа сигналов, развиваемые в школе Ю.П. Пытьева, предназначены для анализа структуры сигнала, сохраняющейся при его преобразовании, принадлежащем заданному классу. В частности, если каждый из двух сигналов представляет собой набор “пиков” и “провалов” (т.е. локальных максимумов и минимумов), то считать их сигналами одной формы разумно, если положения “пиков” и “провалов” совпадают, а упорядоченность локальных экстремумов по амплитуде сохраняется (самый высокий “пик” у обоих сигналов расположен в одной и той же точке и т.д.). Это свойство сигналов сохранится при любом монотонном преобразовании их амплитуд.

Похожий подход к анализу сигналов и изображений предложен в работах А.Н. Каркищенко и А.В. Гончарова [6] и [7], используемое в них описание изображений и сигналов получило название “знакового описания”. Изучению свойств такого описания посвящена монография [8], исследования продолжены в работе [9]. Методы, в которых используется свойство неотрицательности производных сигнала, развиваются также в работе [10].

В методах морфологического анализа [3] форма сигнала $f(t)$, $t \in T$, понимается как множество V_f сигналов, полученных из $f(t)$ всевозможными монотонными преобразованиями его амплитуды; очевидно, при таком понимании формы указанное выше характерное свойство сигнала сохраняется. Математическую модель сигнала и множество преобразований, сохраняющих форму, обычно выбирают так, чтобы для любого предъявленного для анализа сигнала $g(t)$, $t \in T$, была разрешима задача его наилучшего приближения элементами множества V_f , решение которой называют проекцией g на V_f . Операцию проецирования, в ряде случаев определенную конструктивно, называют формой сигнала (изображения) g [3].

В настоящей работе методы морфологического анализа, описанные в [2]–[4], адаптируются для анализа временного ряда концентрации CO_2 , полученного на основе данных круглогодичных измерений на эколого-климатической станции “AsiaFlux” во Вьетнаме с 2011 по 2018 г., предоставленных ИПЭЭ им. А.Н. Северцова РАН [11]. Концентрация CO_2 в атмосфере регистрируется ежесекундно с последующим усреднением до получасовых значений и выражена в миллионных долях (ppm). Таким образом, рассматриваемый временной ряд концентрации CO_2 представляет собой упорядоченную последовательность усредненных получасовых значений концентрации CO_2 на определенной высоте над поверхностью Земли. Во временном ряду имеется составляющая, отражающая циклические процессы с периодом в одни сутки, амплитуда этой составляющей существенно больше других составляющих временного ряда. При этом от суток к суткам имеется существенная вариация поведения концентрации CO_2 , что затрудняет применение не только фурье-, но и вейвлет-анализа. В данной работе предлагается математическая модель формы сигнала, отражающая суточную динамику концентрации CO_2 , в виде функции, направление выпуклости которой меняется на противоположное в точках перегиба.

Эти точки являются параметрами формы фрагмента сигнала, моделирующего суточный ход концентрации CO_2 . Считается, что два фрагмента сигнала имеют одинаковую форму, если интервалы их выпуклостей вверх и выпуклостей вниз совпадают.

Для выделения суточной циклической составляющей из исходного временного ряда выбирается составляющая, моделирующая суточную динамику концентрации CO_2 путем решения задачи наилучшего приближения участков ряда сигналами заданной формы. Далее исследуется остаток ряда, представляющий собой разность исходного ряда и его аппроксимации. Остаток удовлетворяет критерию стационарности.

3. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ВРЕМЕННОГО РЯДА

Временной ряд будем рассматривать как конечный набор значений $\xi_i, i = 1, \dots, N$. Здесь каждое значение ξ_i есть результат наблюдения некоторого параметра в i -й момент времени. Зависимость значения параметра от времени является суммой почти циклической составляющей f_i и широкополосной случайной помехи $v_i, i = 1, \dots, N$.

Зададим математическую модель почти циклической составляющей $f_i, i = 1, \dots, N$, считая, что последовательность моментов $\{1, 2, \dots, N\}$ времени измерений можно разбить на участки $\{i_1, \dots, i_2 - 1\}, \{i_2, \dots, i_3 - 1\}, \dots, \{i_{2m-1}, \dots, i_{2m} - 1\}$, длительность которых может изменяться от участка к участку в заданных пределах, и на каждом нечетном участке почти периодическая составляющая имеет выпуклость вверх, а на четном – вниз. В точках перегиба при переходе от выпуклости вниз к выпуклости вверх функция не убывает, а при смене выпуклости вверх на выпуклость вниз – не возрастает. Множество $V_f(i_1, \dots, i_{2m})$ всех таких функций назовем формой почти периодического сигнала, а границы участков $1 \leq i_1 \leq \dots \leq i_{2m} \leq N + 1$ определяют параметры его формы. Итак, форма $V_f(i_1, \dots, i_{2m})$ почти периодического сигнала есть набор значений $f_i, i = i_1, \dots, i_{2m} - 1$, для которых выполнены неравенства

$$\begin{aligned} 2f_j &\geq f_{j-1} + f_{j+1}, & \text{если } i_{2k-1} + 1 \leq j \leq i_{2k} - 2, & \quad k = 1, \dots, m, \\ 2f_j &\leq f_{j-1} + f_{j+1}, & \text{если } i_{2k} + 1 \leq j \leq i_{2k+1} - 2, & \quad k = 1, \dots, m, \\ f_{j-1} &\geq f_j, & \text{если } j = i_{2k}, & \quad k = 1, \dots, m - 1, \\ f_{j-1} &\leq f_j, & \text{если } j = i_{2k+1} & \quad k = 1, \dots, m - 1. \end{aligned} \tag{1}$$

Если рассматривать значения $\xi_i, i = i_1, \dots, i_{2m} - 1$, временного ряда как координаты вектора конечномерного евклидова пространства, то множество $V_f(i_1, \dots, i_{2m})$ является выпуклым замкнутым множеством, и для любого вектора ξ с координатами $\xi_i, i = i_1, \dots, i_{2m} - 1$, однозначно разрешима задача наилучшего приближения

$$P(i_1, \dots, i_{2m})\xi = \arg \inf \{ \|\xi - \mathbf{f}\|^2 \mid \mathbf{f} \in V_f(i_1, \dots, i_{2m}) \} \tag{2}$$

вектора ξ векторами из $V_f(i_1, \dots, i_{2m})$ при фиксированных параметрах формы. Ее решение $P(i_1, \dots, i_{2m})\xi$ называется проекцией ξ на $V_f(i_1, \dots, i_{2m})$, а оператор проецирования $P(i_1, \dots, i_{2m})$ взаимно однозначно определяется множеством $V_f(i_1, \dots, i_{2m})$, и поэтому тоже называется формой сигнала \mathbf{f} [3].

Сформулируем лемму, на основе которой удастся найти решение задачи (2).

Лемма. Пусть в евклидовых пространствах $R^m, R^k, R^{m+k} = R^m \otimes R^k$, где $\mathbf{g} = (\mathbf{t}, \mathbf{s}), \mathbf{g} \in R^{m+k}, \mathbf{t} \in R^m, \mathbf{s} \in R^k$ и $(g_1, \dots, g_m, g_{m+1}, \dots, g_{m+k}) = (t_1, \dots, t_m, s_1, \dots, s_k)$, выпуклые замкнутые конусы V_{m+k}, V_m и V_k заданы системами линейных неравенств: $V_m = \{ \mathbf{t} \in R^m : A\mathbf{t} \geq \mathbf{c} \}, V_k = \{ \mathbf{s} \in R^k : B\mathbf{s} \geq \mathbf{d} \},$

$$V_{m+k} = \{ (\mathbf{t}, \mathbf{s}) \in R^{m+k} : A\mathbf{t} \geq \mathbf{c}, B\mathbf{s} \geq \mathbf{d}, C(\mathbf{t}, \mathbf{s}) \geq \mathbf{g} \} \tag{3}$$

для некоторых матриц A, B, C и векторов $\mathbf{c}, \mathbf{d}, \mathbf{g}$; P_{m+k}, P_m и P_k – проекторы на конусы V_{m+k}, V_m и V_k соответственно. Тогда для некоторого вектора $\tilde{\mathbf{g}} = (\tilde{\mathbf{t}}, \tilde{\mathbf{s}}) \in R^{m+k}$ равенство $P_{m+k}(\tilde{\mathbf{t}}, \tilde{\mathbf{s}}) = (P_m \tilde{\mathbf{t}}, P_k \tilde{\mathbf{s}})$ выполняется тогда и только тогда, когда выполнено неравенство $C(P_m \tilde{\mathbf{t}}, P_k \tilde{\mathbf{s}}) \geq \tilde{\mathbf{g}}$.

Эта лемма обобщает утверждения, приведенные в работах [3] и [4], и доказывается аналогично.

Лемма позволяет последовательно строить проекции фрагментов временного ряда на конус, определяемый неравенствами (1). На первом шаге строится проекция фрагмента ряда, состоящего из трех первых последовательных значений, на конус размерности 3. Далее, проекции фрагментов из 4, 5, ... значений элементов ряда на соответствующие конусы размерности 4, 5 и т.д. строятся путем последовательного присоединения к проецируемому фрагменту по одному значению.

Действительно, пусть для определенности $i_1 = 1$, и построена проекция $P_j(f_1, \dots, f_j)$ на конус V_j , определяемый неравенствами (1), в которые входят только координаты (f_1, \dots, f_j) . Тогда проекция $P_{j+1}(f_1, \dots, f_{j+1})$ на конус V_{j+1} строится последовательным построением пары проекций $P_{j-l+1}(f_1, \dots, f_{j-l+1})$, и $P_l(f_{j-l+2}, \dots, f_{j+1})$ вектора (f_1, \dots, f_{j+1}) на конусы V_{j-l+1} и V_{l+2} ($l = 1, 2, \dots$) соответственно. Здесь первый конус V_{j-l+1} определяется неравенствами (1), в которые входят только координаты (f_1, \dots, f_{j-l+1}) , а второй V_{l+2} неравенствами (1), в которые входят только координаты $(f_{j-l}, \dots, f_{j+1})$, $l = 1, 2, \dots$, до тех пор, пока не будут выполнены неравенства (1), в которые входит хотя бы одна координата как из первого набора, так и из второго. Заметим, что как в первый, так и во второй наборы координат могут входить одна, две и более координат, в первых двух случаях могут отсутствовать неравенства, связывающие только координаты из соответствующего набора; в этой ситуации ограничения на эти координаты отсутствуют, а конус превращается в линейное подпространство.

Рассмотрим подробнее, как строится проекция $P_{j+1}(f_1, \dots, f_{j+1})$ при известной $P_j(f_1, \dots, f_j)$. Пусть имеется два набора координат: (f_1, \dots, f_j) и f_{j+1} . Если выполнены все неравенства из (1), в которые входят f_{j+1} и хотя бы одна координата из (f_{j-1}, f_j) , то искомая проекция построена и равна $P_{j+1}(f_1, \dots, f_{j+1}) = (P_j(f_1, \dots, f_j), f_{j+1})$, так как выполнены условия леммы при $B = 0$, $\mathbf{d} = 0$.

Если эти неравенства не выполняются, следует построить две проекции: $P_{j-1}(f_1, \dots, f_{j-1})$ (она уже вычислена на предыдущих шагах) вектора (f_1, \dots, f_{j-1}) на конус V_{j-1} и $P_2(f_j, f_{j+1})$ вектора (f_j, f_{j+1}) на множество, определяемое неравенствами из (1), в которые входят только координаты (f_j, f_{j+1}) ; если таких неравенств нет, то $P_2(f_j, f_{j+1}) = (f_j, f_{j+1})$. Далее для координат построенных проекций проверяются выполнения неравенств (1), в которые входят как координаты из набора (f_j, f_{j+1}) , так и из набора (f_{j-2}, f_{j-1}) . Если эти неравенства выполняются, то искомая проекция построена, иначе следует продолжить процесс.

Поясним, как построить проекции на конусы на начальных шагах. Пусть, например, как и прежде, $i_1 = 1$, а $i_2 > 4$.

Шаг 1. Выберем первые три координаты f_1, f_2, f_3 , и построим проекцию $P_3(f_1, f_2, f_3) = (q_1, q_2, q_3)$ трехмерного вектора (f_1, f_2, f_3) на трехмерный конус

$$V_3 = \{(t_1, t_2, t_3) : 2t_2 \geq t_1 + t_3\}, \quad (4)$$

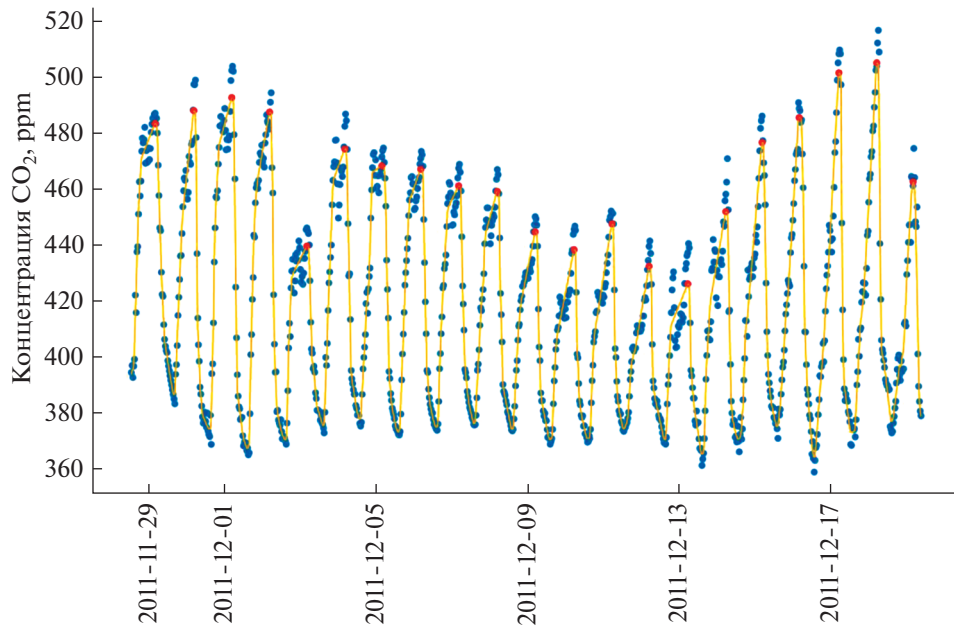
решая задачу выпуклого программирования

$$\inf_t \left\{ \sum_{i=1}^3 (f_i - t_i)^2 \mid 2t_2 \geq t_1 + t_3 \right\}.$$

Эта проекция есть либо сам вектор (f_1, f_2, f_3) , если $2f_2 \geq f_1 + f_3$, либо наилучшая среднеквадратичная аппроксимация вектора (f_1, f_2, f_3) вектором с координатами $(q_1, (q_1 + q_3)/2, q_3)$, т.е. линейно зависящими от времени. Заметим, что эта ситуация соответствует утверждению леммы при $B = 0$, $C = 0$, $\mathbf{d} = 0$, $\mathbf{g} = 0$.

Шаг 2. Добавим к набору координат (f_1, f_2, f_3) вектора \mathbf{f} четвертую координату f_4 . Если выполнены неравенства $2q_3 \geq q_2 + f_4$ для координат проекции $P_3(f_1, f_2, f_3) = (q_1, q_2, q_3)$, вычисленной на первом шаге, и f_4 , то согласно лемме проекция $P_{3+1}(f_1, f_2, f_3, f_4)$ на конус $V_{3+1} = \{(t_1, t_2, t_3, s_1) : 2t_2 \geq t_1 + t_3, 2t_3 \geq t_2 + s_1\}$ в четырехмерном пространстве равна $P_{3+1}(f_1, f_2, f_3, f_4) = (P_3(f_1, f_2, f_3), f_4)$, что соответствует условиям леммы при $\mathbf{d} = 0$. Если это неравенство не выполнено, то дальнейшие действия зависят от результата первого шага.

1. Если $P_3(f_1, f_2, f_3) = (f_1, f_2, f_3)$, то следует построить проекцию $P_3(f_2, f_3, f_4)$ трехмерного вектора (f_2, f_3, f_4) на конус V_3 , определенный в (4), и проверить включение $(f_1, P_3(f_2, f_3, f_4)) \in V_{3+1}$. Если



Фиг. 1. Временной ряд концентрации CO₂; точки – результат регистрации, сплошная линия – результат морфологической фильтрации. Шаг по времени 0.5 ч.

это включение выполнено, то $(f_1, P_3(f_2, f_3, f_4))$ – искомая проекция, так как выполнены условия леммы при $A = 0, \mathbf{c} = 0$ для $V_{1+3} = V_{3+1}$.

2. Если $P_3(f_1, f_2, f_3) \neq (f_1, f_2, f_3)$, то проекция $P_{3+1}(f_1, f_2, f_3, f_4)$ есть результат аппроксимации фрагмента (f_1, f_2, f_3, f_4) временного ряда фрагментом, значения которого есть линейная функция времени, в соответствии с условиями леммы при $B = 0, C = 0, \mathbf{d} = 0, \mathbf{g} = 0$.

Итак, если заданы параметры формы $V_f(i_1, \dots, i_{2m})$ временного ряда, то построить наилучшее приближение любого временного ряда рядами заданной формы можно, основываясь на приведенных здесь результатах. Однако на практике интервалы выпуклости вверх и вниз могут быть непостоянными и меняться в известных пределах: $(i_1, \dots, i_{2m}) \in I, I = \{(i_1, \dots, i_{2m}) : \underline{i}_k \leq i_k \leq \bar{i}_k, k = 1, \dots, 2m\}$. Интерес представляет выбор значений этих параметров, при которых анализируемый временной ряд наиболее близок к соответствующей форме $V_f(i_1, \dots, i_{2m})$.

Задачу наилучшего приближения временного ряда сигналами формы $V_f(i_1, \dots, i_{2m})$ выбором параметров формы назовем задачей морфологической фильтрации временного ряда. Формально это есть задача на поиск минимума

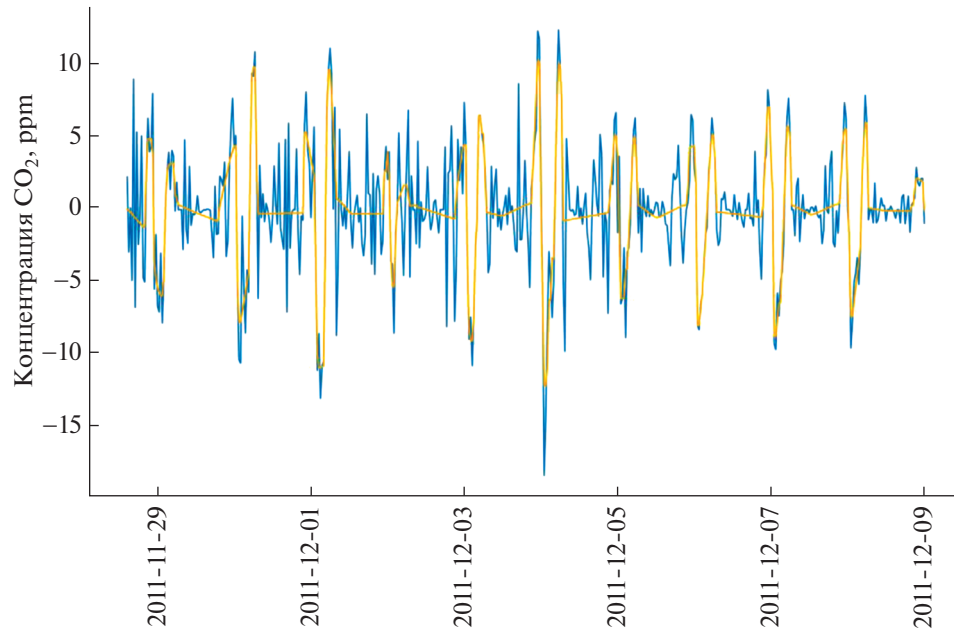
$$\inf_{i \in I} \inf \{ \|\xi - \mathbf{f}\|^2 \mid \mathbf{f} \in V_f(i_1, \dots, i_{2m}) \},$$

где I – множество допустимых значений параметров формы $V_f(i_1, \dots, i_{2m})$.

4. ЗАДАЧА ФИЛЬТРАЦИИ СИГНАЛА ЗАДАННОЙ ФОРМЫ

Рассмотрим пример анализа временного ряда, при котором используется морфологическая фильтрация, описанная в разд. 2 настоящей статьи. На фиг. 1 точками отображена временная зависимость концентрации CO₂, измеренная с шагом 0.5 ч. По вертикали отложена концентрация в миллионных долях (ppm), по горизонтали – время. Видно, что фрагменты временного ряда длительностью примерно 24 ч схожи по форме, однако имеются вариации как в амплитуде, так и в длительности фрагментов. В целом, 24-часовые фрагменты хорошо описываются сигналами, математическая модель формы которых описана в предыдущем разделе настоящей статьи, однако имеются и отличия, которые тоже могут представлять интерес для исследователя.

Для описания суточного хода концентрации CO₂ была применена морфологическая фильтрация сигнала. Морфологическая фильтрация осуществляется следующим образом. Задаются



Фиг. 2. Остаток временного ряда концентрации CO_2 после вычитания результата морфологической фильтрации и результат фильтрации остатка для выделения парных пиков.

ограничения на параметры формы, затем для каждого возможного значения параметров i_1 и i_2 строится проекция фрагмента временного ряда на множество фрагментов, выпуклых вверх, и выбираются такие их значения, которые минимизируют отличие анализируемого фрагмента от заданной формы. Затем для найденного значения i_2 и для каждого возможного значения i_3 строится проекция фрагмента временного ряда на множество фрагментов, выпуклых вниз. Далее проверяется условие сшивания фрагментов, выпуклых вниз и вверх. Условие сшивания задается неравенствами (3). Если они не выполнены, корректируются значение найденной проекции и значение i_2 положения точки перегиба, далее эта процедура повторяется для каждого участка заданного типа выпуклости.

На фиг. 1 приведен результат морфологической фильтрации (сплошная линия). Хорошо видны повторяющиеся участки длительностью примерно 24 ч, отвечающие суточной изменчивости концентрации CO_2 . Средний период, вычисленный по исследованному фрагменту временного ряда, составил 24.0 ч.

В результате применения алгоритма морфологической фильтрации к временному ряду концентрации CO_2 наряду с выделением суточной циклической составляющей из исходного временного ряда проведено исследование остаточной составляющей, представляющей собой разность исходного ряда и его аппроксимации, полученной в результате морфологической фильтрации. График значений остатков показан на фиг. 2; на нем явно угадываются повторяющиеся парные пики. Они были выделены из остаточной составляющей повторной морфологической фильтрацией с периодом, меньшим, чем 24 ч. Результат повторной морфологической фильтрации приведен на фиг. 2 в виде кривой оранжевого цвета.

Проверка остаточной составляющей на стационарность с помощью расширенного теста Дики–Фуллера [12] показала, что, в отличие от исходного ряда, ряд остатков удовлетворяет условию стационарности, следовательно, для дальнейшего выделения циклических компонент ряда с периодами в пределах суток возможно применение фурье-анализа.

Приближенное представление ряда остатков X_t в виде частичной суммы ряда Фурье

$$X_t \approx a_0 + \sum_{i=1}^I \left[a_i \cos\left(\frac{2\pi t}{T_i}\right) + b_i \sin\left(\frac{2\pi t}{T_i}\right) \right], \quad (5)$$

где $t = 1, 2, \dots, N$, N — число элементов ряда, a_i , b_i — коэффициенты Фурье, T_i — период i -й гармоники, $i = 1, \dots, I$, I — число гармоник, позволяет обнаружить колебания ряда X_t остаточной

составляющей с периодами в пределах суток, вносящие наибольший вклад во временной ряд X_t , и оценить вклад i -й гармоники с помощью периодограммы – значений $\text{Per}_i = \frac{N}{2}(a_i^2 + b_i^2)$, $i = 1, \dots, I$, а ее относительный вклад в суммарную периодограмму ряда $(X_t - a_0)$ оценить значением

$$\frac{\text{Per}_j}{\sum_{i=1}^I \text{Per}_i} = \frac{\frac{N}{2}(a_j^2 + b_j^2)}{\frac{N}{2} \sum_{i=1}^I (a_i^2 + b_i^2)}, \quad i = 1, \dots, I.$$

В результате применения фурье-анализа к ряду остаточной составляющей показано, что наиболее существенный вклад вносят гармонические компоненты разложения Фурье (5) с периодами 11.9, 7.95 и 6 ч.

Алгоритм морфологической фильтрации требует достаточно больших вычислительных затрат по сравнению с линейными разложениями по заданной системе функций, поэтому для исследования рядов большой длительности воспользуемся более быстрыми алгоритмами фурье-анализа и вейвлет-анализа.

5. СТАТИСТИЧЕСКИЙ АНАЛИЗ ВРЕМЕННОГО РЯДА КОНЦЕНТРАЦИИ УГЛЕКИСЛОГО ГАЗА

Статистический анализ временных рядов концентрации углекислого газа также проводится на основе данных эколого-климатической станции “AsiaFlux” во Вьетнаме с 2011 по 2018 г. На фиг. 3 графически представлен временной ряд концентрации CO_2 за период с 2011 по 2018 г. На основе визуального анализа графика концентрации CO_2 выдвинем предположение о том, что представленный ряд содержит регулярную и случайную компоненты. Такое предположение подтверждается также заключениями метеорологов о механизмах формирования уровня концентрации CO_2 в атмосфере под воздействием различных природных и антропогенных факторов. При этом регулярная компонента включает основную тенденцию ряда динамики показателя (тренд), а также его сезонные и циклические составляющие, характеризующиеся различными периодами колебаний концентрации CO_2 .

Визуальный анализ рядов динамики концентрации CO_2 позволяет также высказать предположение о том, что значительных изменений амплитуды колебаний концентрации CO_2 с 2011 г. по 2018 г. как в сторону заметного увеличения, так и в сторону уменьшения, не наблюдается, поэтому в качестве модели декомпозиции изучаемых рядов динамики предлагается аддитивная модель [13]:

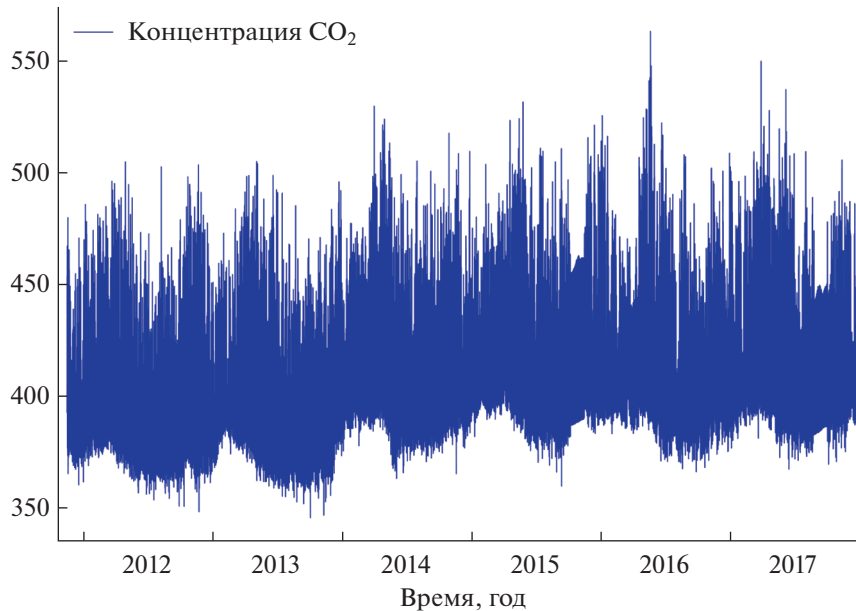
$$X_t = S_t + C_t + u_t + \varepsilon_t, \tag{6}$$

где S_t – постоянная сезонная составляющая, C_t – циклическая составляющая, u_t – тренд, определяющий основную тенденцию временного ряда, ε_t – нерегулярная составляющая, $t = 1, 2, \dots, N$, где N – число элементов ряда.

При изучении основной тенденции изменчивости концентрации CO_2 за рассматриваемый временной период целесообразно на начальном этапе, предшествующем декомпозиции, провести процедуру сглаживания исходного ряда с целью повышения его устойчивости по отношению к выбросам различной природы. В результате сглаживания методом простой скользящей средней исходного ряда динамики концентрации CO_2 X_t , $t = 1, \dots, N$ (6), получаем ряд скользящих средних \hat{X}_t :

$$\hat{X}_t = \frac{\sum_{i=t-p}^{t+p} X_i}{K}, \quad p = \frac{K-1}{2}, \quad t = 1, \dots, N, \quad i = p+1, \dots, N-p, \tag{7}$$

где K – интервал сглаживания, равный периоду колебаний концентрации CO_2 , который будем задавать различным образом, учитывая периоды колебаний, вносящие наибольший вклад в ряд динамики концентрации CO_2 . Если провести усреднение ежесекундных значений концентрации CO_2 не за полчаса, а за сутки, т.е. в качестве X_t , $t = 1, \dots, N$, рассмотреть среднесуточные зна-



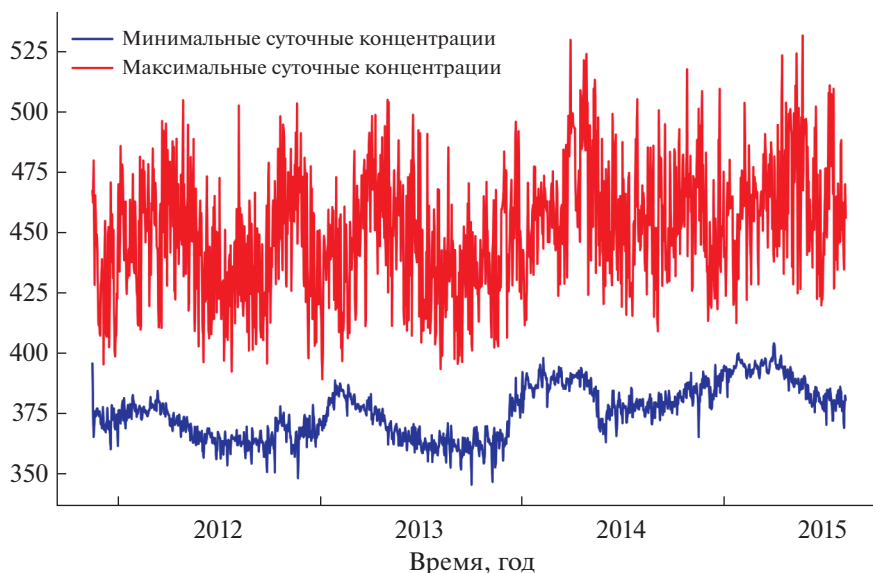
Фиг. 3. Графическое представление временного ряда концентрации CO₂ (2011–2018 гг.)

чения уровней ряда, а интервал сглаживания задать равным 365 дней, в результате сглаживания получится ряд $\hat{X}_t, t = 1, \dots, N$, не содержащий влияния сезонных изменений. Тогда для определения постоянной сезонной компоненты S_t следует усреднить разность значений X_t и \hat{X}_t для каждого дня $t = k + Kj, t = 1, 2, \dots, N$, на протяжении восьмилетнего периода наблюдений за концентрацией CO₂, где k — день года, j — год, $k = 1, \dots, K, j = 0, \dots, J - 1$:

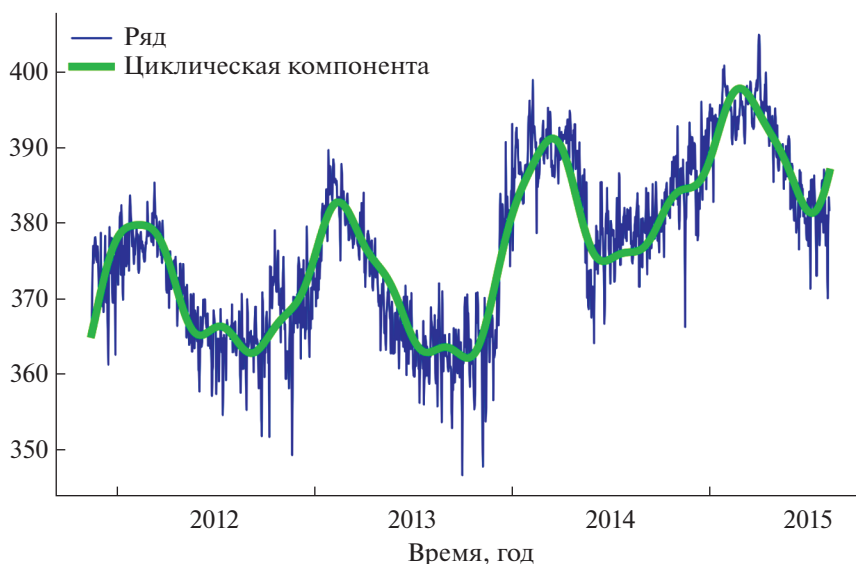
$$S_t = \frac{\sum_{j=0}^{J-1} [X_{k+Kj} - \hat{X}_{k+Kj}]}{J},$$

где $J = 8$ лет, $K = 365$ дней. Для дальнейшего изучения основной тенденции ряда (6), не зависящей от сезонных колебаний, следует вычесть из исходного ряда $X_t, t = 1, \dots, N$, ряд сезонной компоненты $S_t, t = 1, \dots, N$, и снова провести усреднение ряда $X_t - S_t, t = 1, \dots, N$, методом скользящих средних (7). Поскольку в этом случае характер основной тенденции может зависеть от выбора периода усреднения, проведем анализ циклических колебаний ряда динамики концентрации CO₂.

Наряду с временными рядами концентрации CO₂ на различных высотах над поверхностью Земли, важное значение для анализа климатических изменений имеют ряды динамики экстремальных (минимальных и максимальных) суточных значений концентрации CO₂ (фиг. 4). Проверка временных рядов концентрации CO₂ на стационарность с помощью расширенного теста Дики–Фуллера [12] показала, что в то время как ряды динамики нельзя считать стационарными, временные ряды экстремальных суточных значений концентрации CO₂ являются стационарными, поэтому для анализа их циклических колебаний применен фурье-анализ. В результате применения фурье-анализа к ряду минимальных суточных значений концентрации CO₂ выделены следующие колебания концентрации CO₂, вносящие наибольший вклад во временной ряд X_t : 11.37 мес, 3.73 г, 15.1, 6.5, 9, 7.6 и 3.5 мес (перечислены в порядке убывания значений периодограммы). Графическое представление циклической составляющей C_t временного ряда (6) показано на фиг. 5. При анализе максимальной суточной концентрации были выделены совпадающие по значениям, но отличающиеся по величине периодограммы периоды, приведенные далее также в порядке убывания значений периодограммы: 6.5, 15.1, 11.37, 7.6, 5.7 мес. Проверка остаточных составляющих ε_t (5) рядов максимальных и минимальных суточных концентраций с помощью критерия Колмогорова–Смирнова показала, что ряды остатков подчиняются нормальному закону распределения.



Фиг. 4. Графики минимальных и максимальных суточных значений концентрации CO₂.



Фиг. 5. Зависимость минимальных суточных значений концентрации CO₂ и его циклической компоненты от времени.

Таким образом, с помощью морфологического анализа исходного временного ряда концентрации CO₂ и остаточной составляющей, представляющей собой разность исходного ряда и его аппроксимации, полученной в результате морфологической фильтрации, были выявлены суточные колебания концентрации CO₂, а также циклические составляющие с периодами в пределах суток: 12, 8 и 6 ч. Для определения циклических компонент, имеющих более длительные периоды, был применен фурье-анализ стационарных рядов минимальных и максимальных суточных значений концентрации CO₂. В результате фурье-анализа показано, что наибольший вклад в исследуемые за период с 2011 по 2018 г. временные ряды вносят гармоники около 1 г., 3.7 года, 6.5 и 9 мес.

6. ВЕЙВЛЕТ-АНАЛИЗ ВРЕМЕННОГО РЯДА КОНЦЕНТРАЦИИ УГЛЕКИСЛОГО ГАЗА

В случае, когда ряд не является стационарным и преобразовать его в стационарный не удается, оправдано применение вейвлет-анализа. Вейвлеты – это семейство функций, которые полу-

чаются из одной функции посредством ее сдвигов и растяжений по оси времени. С помощью вейвлет-преобразования функция рассматривается в виде разложения на колебания, локализованные по времени и частоте. Таким образом, вейвлет-анализ применяется для обработки нестационарных сигналов, так как отвечает специфике временных рядов, которые демонстрируют эволюцию во времени своих основных характеристик — среднего значения, дисперсии, периодов, амплитуду и т.д. [14] и [15].

Исследуемый временной ряд показателей концентрации CO_2 за полтора года, усредненных с периодом в 30 мин, в ходе работы был проверен на стационарность с помощью расширенного теста Дикки—Фуллера. Результат оказался предсказуем — ряды нестационарны, тем самым обосновано применение вейвлет-анализа. Временные ряды минимумов и максимумов на различных высотах, напротив, оказались стационарными, что, вообще говоря, не исключает возможность использования вейвлет-преобразования в качестве одной из методик исследования, в связи с этим вышеупомянутые ряды динамики также использовались в качестве анализируемых временных рядов.

Сущность непрерывного вейвлет-преобразования заключается в следующем. С помощью подходящего материнского вейвлета $\Psi(t)$ вычисляются вейвлет-функции $\Psi_{a,b}(t)$:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right),$$

где параметр a , называемый параметром масштаба вейвлет-преобразования, принимает строго положительные значения и отвечает за ширину вейвлета; величина b есть параметр сдвига, который определяет положение вейвлета на оси t .

Далее определяются вейвлет-коэффициенты $W(a, b)$:

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \Psi_{a,b}^*(t) dt,$$

где $f(t)$ — анализируемый ряд динамики, * обозначает комплексное сопряжение [5].

После этого проводится качественный анализ картины вейвлет-коэффициентов и построение интегрального спектра:

$$S(a) = \int_{a_1}^{a_2} |W(a, b)|^2 db,$$

который показывает наличие циклов в исходном временном ряде. Масштаб a связан с координатами оси времени t как $t = \frac{a}{F_c}$, где F_c — центральная частота вейвлета.

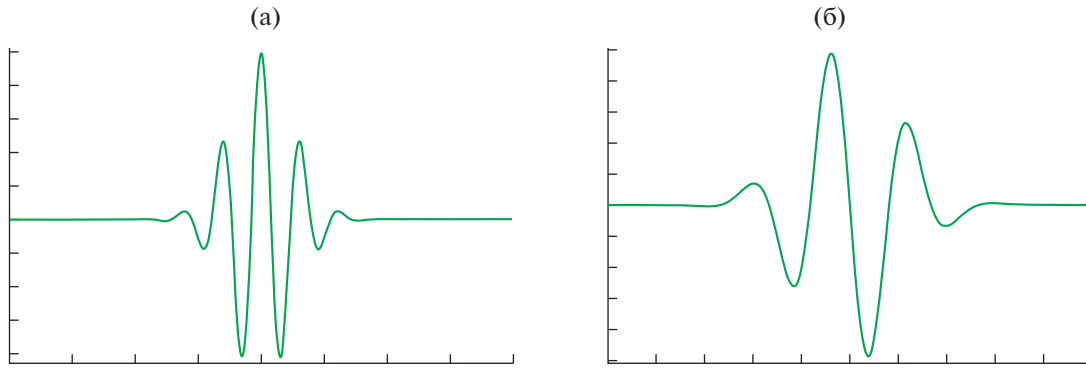
Для исследуемых временных рядов в качестве материнских вейвлетов были выбраны вейвлет Морле и вейвлет Гаусса 7, дающие наиболее информативные и наглядные результаты [16]. Центральная частота данных материнских вейвлетов — 0.8125 и 0.6 соответственно (фиг. 6). Применение иных материнских вейвлетов давали результаты, принципиально не отличающиеся от приведенных в настоящей статье, но с менее выраженными частотно-временными особенностями.

Картины вейвлет-коэффициентов позволяют подробно исследовать структуру периодичностей, содержащихся в исследуемом ряде динамики (в отличие, например, от метода фурье-анализа). Масштаб, коэффициенты которого максимальны (т.е. вносят максимальный вклад во временной ряд) характеризует определенный период цикличности. Иными словами, светлые горизонтальные линии спектра и соответствуют искомым циклам (фиг. 7а и 8а). Картина вейвлет-коэффициентов позволяет качественно оценить величину цикличностей и время, когда они наблюдаются.

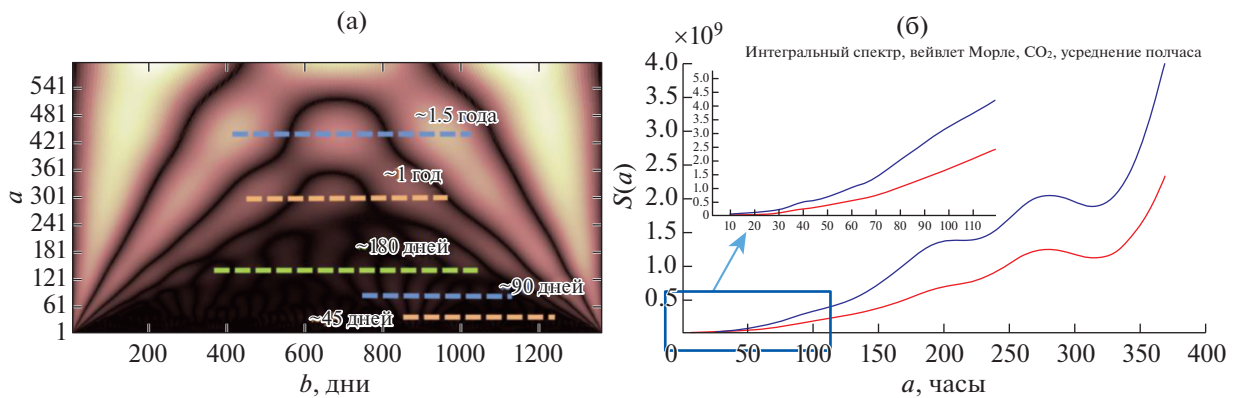
Интегральный спектр, в отличие от картины вейвлет-коэффициентов, позволяет рассчитать величину периода цикла (фиг. 7б и 8б). Локальные экстремумы графика соответствуют максимальным коэффициентам на данном масштабе, показывая значения периода циклов.

Также были получены картины вейвлет-коэффициентов и интегральные спектры для различных показателей концентрации углекислого газа в атмосфере.

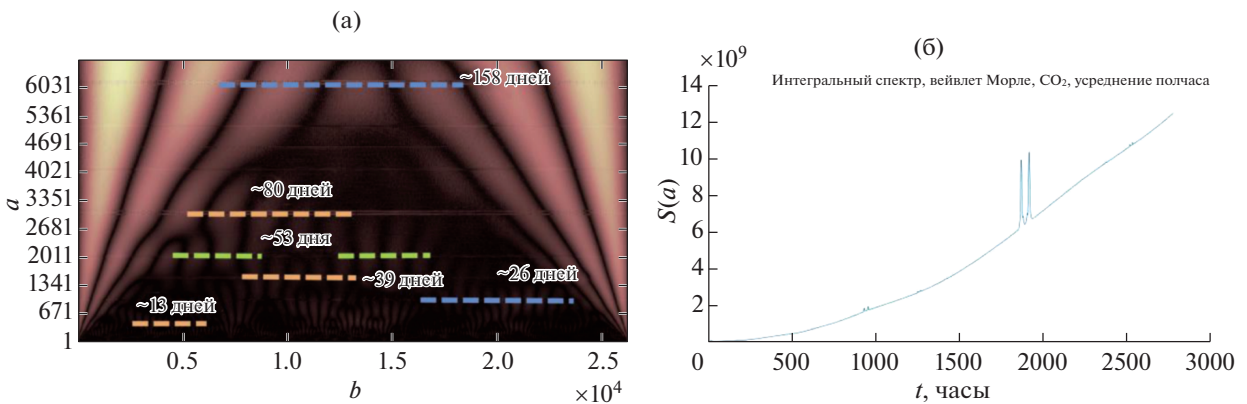
На фиг. 7а представлена картина вейвлет-коэффициентов минимумов и максимумов концентраций углекислого газа, на которой выделены цикличности, что позволяет качественно наблюдать нестационарную структуру временного ряда, показывающего наличие множественных не-продолжительных цикличностей.



Фиг. 6. Вейвлет Морле (а) и Гаусса 7 (б).



Фиг. 7. Коэффициенты вейвлет-разложения временного ряда минимальных суточных концентраций CO_2 с базовым вейвлетом Морле (а), интегральный спектр (б).



Фиг. 8. Коэффициенты вейвлет-разложения временного ряда концентраций CO_2 , усредненного по получасовым интервалам, с базовым вейвлетом Морле (а), интегральный спектр (б).

Соответствующий интегральный спектр (фиг. 7б) также является достаточно размытым. Можно наблюдать пики в 23, 45, около 80 дней для минимума концентрации, а также общие и для показателей минимумов и максимумов пики в 1.04 и 1.5 года.

Данные временные ряды были также проанализированы с помощью вейвлета Гаусса. Картины вейвлет-коэффициентов для данного материнского вейвлета аналогичны, а интегральные спектры размыты еще сильнее, но некоторые слабые пики совпадают с результатами, полученными с помощью вейвлета Морле.

При анализе данных концентрации углекислого газа за полтора года с усреднением в полчаса также выявлено множество непродолжительных (вносящих вклад в ряд динамики на протяжении только части области исследования) цикличностей (фиг. 8а). В связи с чем для получения более информативных результатов необходимо анализировать отдельные участки временного ряда, которые следует выделить в согласовании с климатическими условиями местности.

7. ЗАКЛЮЧЕНИЕ

Рассмотрены математические методы анализа рядов динамики с целью изучения их цикличности — метод морфологической фильтрации, предложенный в данной статье, метод, основанный на анализе фурье-разложения временного ряда, и метод вейвлет-анализа. Метод морфологической фильтрации позволил выделить суточную циклическую составляющую и анализировать остаток ряда для оценивания составляющих с периодом, меньшим, чем 24 ч. Метод, основанный на фурье-преобразовании, позволил выделить циклические составляющие анализируемого ряда с периодом, как большим, так и меньшим, чем 24 ч. Вейвлет-анализ позволил выделить и локализовать во времени составляющие с периодом, большим, чем 24 ч.

Метод морфологической фильтрации дает возможность преобразовать нестационарный ряд в стационарный, что позволяет провести дальнейшее исследование ряда методами фурье-анализа и статистики. В случае нестационарных рядов выделить циклы также позволяет применение вейвлет-анализа. В нашем случае — это 24, 12, 8 и 6 ч для периодов, не превосходящих суточные. Метод вейвлет-анализа позволил локализовать циклические участки с периодом в несколько дней: 23, 45, около 80 дней для минимума концентрации, а также общие и для показателей минимумов и максимумов на всех высотах пики в 1.04 и 1.5 года.

Периодичность в 1.04 года соответствует в целом годовому циклу изменения концентрации CO_2 и обусловлен сезонной динамикой осадков и фенологическими изменениями в развитии растительности, а в 24 ч соответствует суточному циклу. Объяснение природы циклов с другими периодами требует дальнейших исследований.

Таким образом, предложен набор методов, позволяющих выявить цикличность рядов в широком временном диапазоне.

СПИСОК ЛИТЕРАТУРЫ

1. *Кричевский М.Л.* Временные ряды в менеджменте. Ч. 1, 2. М.: РУСФИНС, 2016.
2. *Пытьев Ю.П.* Морфологические понятия в задачах анализа изображений // Докл. АН СССР. 1975. Т. 224. № 6. С. 1283–1286.
3. *Пытьев Ю.П., Чуличков А.И.* Методы морфологического анализа изображений // М.: Физматлит, 2010. 336 с.
4. *Demin D.S., Chulichkov A.I.* Filtering of monotonic convex noise-distorted signals and estimates of positions of special points // J. of Math. Sci. 2011. V. 172. № 6. P. 770–781.
5. *Astafeyeva N.M.* Wavelet analysis: basic theory and some applications // Phys. Usp. 1996. V. 39. P. 1085–1108.
6. *Гончаров А.В.* Исследование свойств знакового представления изображений в задачах распознавания образов // Известия ЮФУ. Технические науки. 2009. Тематический выпуск. С. 178–188.
7. *Каркищенко А.Н.* Исследование устойчивости знакового представления изображений // Автоматика и телемехан. 2010. Т. 9. С. 57–69.
8. *Броневиц А.Г., Каркищенко А.Н., Лепский А.Е.* Анализ неопределенности выделения информативных признаков и представлений изображений. М.: Физматлит, 2013. 320 с.
9. *Мясников В.В.* Локальное порядковое преобразование цифровых изображений // Компьютерная оптика. 2015. Т. 39. № 3. С. 397–405.
10. *Terentiev E.N., Farshakova I.I., Prikhodko I.N.* Problems of accurate localization objects in images // AIP Conference Proc. 2019. V. 2171. P. 110009-1–110009-4.
11. *Дещеревская О.А., Авиллов В.К., Ба Зуй Динь, Конг Хуан Чан, Курбатова Ю.А.* Современный климат национального парка Кат Тьен (Южный Вьетнам): использование климатических данных для экологических исследований // Геофизические процессы и биосфера. 2013. Т. 12. № 2. С. 5–33.
12. *Dickey D.A., Fuller W.A.* Distribution of the Estimators for Autoregressive Time Series with a Unit Root // J. of the American Statistical Association. 1979. V. 74. P. 427–431.
13. *Вуколов Э.А.* Основы статистического анализа. М.: Форум, 2008. 464 с.
14. *Sallie Baliunas, Peter Frick, Dmitry Sokoloff, Willie Soon.* Time scales and trends in the Central England Temperature data (1659–1990): A wavelet analysis // Geophysical Research Letters. 1997. V. 24. № 11. P. 1351–1354.
15. *Витязев В.В.* Вейвлет-анализ временных рядов. СПб.: Изд-во СПбГУ, 2001.
16. *Асташов Р.А., Голубинский А.Н.* Обоснование выбора материнского вейвлета непрерывного вейвлет-преобразования для анализа речевых сигналов // Вестник Воронежского института МВД России. 2014. № 1. С. 7–15.