ОБ ОДНОМ АЛГОРИТМЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

X. X. XXXXXXX[1] Y. Y. YYYYY[1]

[1]*Московский государственный университет имени М.В. Ломоносова*

Анализ временных рядов представляет собой мощный инструмент для изучения характеристик некоторой физической величины с течением времени. Временные ряды находят широкое применение в различных областях, включая экономику, климатологию, медицину и физику. Однако в астрономии и астрофизике, где имеется большое количество данных, полученных из наблюдений за астрономическими объектами, анализ временных рядов обретает особое значение. В данном отчете мы рассмотрели один из методов анализа временных рядов в астрономии, использующий быстрое преобразование Фурье.

I. ОПРЕДЕЛЕНИЯ

В дальнейшем при анализе временного ряда мы будем использовать *периодограмму* и *коррелограмму*. *Периодограмма* — это оценка спектральной плотности мощности. В диапазоне частот до частоты Найквиста большинство частот локальных максимумов периодограммы совпадает с частотами локальных максимумов истинной спектральной плотности. Рассмотрим временной ряд $x = x(t)$, который задается дискретным конечным набором вещественных чисел $x_k = x(t_k)$ при $t_k = k\Delta t,\ k = \overline{0,\ N-1}$, где Δt — единичный шаг по времени.
Периодограмма Шустера определяется как

$$D(\nu) = \frac{1}{N^2}\left|\sum_{k=0}^{N-1} x_k e^{-i2\pi\nu t_k}\right|^2 \tag{1}$$

Значения $D(\nu)$ вычисляются по формуле (1), причем только в определенных точках, а именно:

$$\nu_j = \frac{j}{N\Delta t}, \quad j = \overline{0,\ N/2}. \tag{2}$$

Коррелограмма — это оценка автокорреляционной функции, показывающая взаимосвязь двух или более случайных величин, при этом изменения значений одной или нескольких из этих величин сопутствуют систематическому изменению значений другой или других величин. В качестве оценки корреляционной функции c_m берут

$$c_m = \frac{1}{N-m}\sum_{k=0}^{N-m-1} x_k x_{k+m}, \tag{3}$$

$$m = \overline{0,\ N-1}.$$

Периодограмма и коррелограмма связаны соотношением, следующим из теоремы Винера — Хинчина:

$$D(\nu) = \frac{1}{N}\left[2\mathrm{Re}\sum_{m=0}^{N-1}\bar{c}_m e^{-i2\pi\nu m\Delta t} - \bar{c}_0\right], \tag{4}$$

где $\bar{c}_m$ является смещенной оценкой корреляционной функции:

$$\bar{c}_m = \frac{N-m}{N}c_m$$

II. АЛГОРИТМ БЫСТРОГО ПРЕОБРАЗОВАНИЯ ФУРЬЕ

При оценке спектральной плотности мощности и корреляционной функции используется *дискретное преобразование Фурье* (DFT). Возьмем две последовательности

$$\{x_k\}, \quad k = \overline{0,\ N-1},$$
$$\{X_j\}, \quad j = \overline{0,\ N-1},$$

тогда *прямое дискретное преобразование Фурье* (DFT) задается для них как:

$$X_j = \mathrm{DFT}_j\{x_k\}_{k=0}^{N-1} = \sum_{k=0}^{N-1} x_k e^{-i\frac{2\pi}{N}kj}, \tag{5}$$
$$j = \overline{0,\ N-1}.$$

Обратное дискретное преобразование Фурье (DFT^{-1}) определяется в виде:

$$x_k = \mathrm{DFT}_k^{-1}\{X_j\}_{j=0}^{N-1} = \sum_{j=0}^{N-1} X_j e^{i\frac{2\pi}{N}kj}, \tag{6}$$
$$k = \overline{0,\ N-1}.$$

Используя те же частоты: что и в (2) получаем формулу для вычисления периодограмм:

$$D_j = \frac{1}{N^2}\left|\mathrm{DFT}_j\{x_k\}_{k=0}^{N-1}\right|^2, \tag{7}$$
$$j = \overline{0,\ N/2}.$$

Однако, есть возможность использовать более быстрый алгоритм для вычисления данного преобразования. Эта процедура называется *быстрым преобразованием Фурье* (FFT). Чтобы выполнялся данный алгоритм число, точек временного ряда N должно составлять:

$$N = 2^p, \tag{8}$$

где p — целое положительное число. Тем не менее, не всегда N можно представить в такой форме. Поэтому зачастую ряд дополняют нулями, чтобы длина нового ряда составляла $N_1 = 2^p \geq N$. После этого берут длину нового ряд равную $N_2 = 2N_1$.
Помимо формулы (3) для вычисления коррелограмм можно использовать формулу, которая задействует обратное преобразования Фурье (FFT^{-1}):

$$\bar{c}_m = N\ \mathrm{Re}\ \mathrm{FFT}_m^{-1}\{D_j\}_{j=0}^{N_2-1}, \tag{9}$$
$$m = \overline{0,\ N-1}.$$

III. АЛГОРИТМ АНАЛИЗА

В этом разделе приведен алгоритм анализа равномерного временного ряда. Возьмем временной ряд и определим основные его параметры:

$$x_k = \alpha + \beta t_k + A_1 \cos\left(2\pi\nu_1 t_k - \varphi_1\right) + \sigma_x \xi_k, \tag{10}$$

$$t_k = k\Delta t, \quad k = \overline{0,\ N-1},$$

где Δt — постоянный шаг выборки; α и β — параметры линейного тренда; A_1, ν_1, φ_1 — амплитуда, частота и фаза колебательного компонента; $\xi_k, k = \overline{0,\ N-1}$ — шумовой компонент ряда, распределенная по нормальному закону:

$$\xi \sim \mathcal{N}(0,1);$$

σ_x — среднеквадратическое отклонение шумового компонента ξ_k, задаваемое с помощью соотношения

$$\sigma_n = \sqrt{\frac{A_1^2}{2\gamma}}, \tag{11}$$

где γ — отношение «сигнал-шум».
Алгоритм вычисления характеристик ряда и построения периодограмм и коррелограмм приведен ниже.

A. Построение графика временного ряда $x_k = x(t_k)$

При рассмотрении графического изображения данного ряда мы убеждаемся в наличии линейного тренда, а также квазипериодической составляющей.

B. Исключение тренда и центрирование ряда

Аппроксимируем функцию временного ряда с помощью линейной функции и центрируем временной ряд. Соответственно, для этого получим оценки коэффициентов $\hat{\alpha}$ и $\hat{\beta}$ функции:

$$x_k = \hat{\alpha} + \hat{\beta} t_k. \tag{12}$$

Оценим данные коэффициенты, используя метод наименьших квадратов (МНК):

$$\hat{\beta} = \frac{\langle t_k x_k \rangle - \langle t_k \rangle \langle x_k \rangle}{\langle t_k^2 \rangle - \langle t_k \rangle^2}, \tag{13}$$

$$\hat{\alpha} = x_k - \hat{\beta} \langle t_k \rangle. \tag{14}$$

После этого необходимо вычесть линейный тренд из временного ряда. Отсюда мы получаем оценку центрированного ряда:

$$\hat{x}_k^0 = \hat{x}^0(t_k) = x_k - \hat{\alpha} - \hat{\beta} t_k, \quad k = \overline{0,\ N-1}. \tag{15}$$

C. Графическое представление центрированного ряда

Затем представим оценку центрированного ряда в виде графика $\hat{x}^0(t_k)$.

D. Получение периодограммы

Чтобы вычислить периодограмму, необходимо использовать *быстрое преобразование Фурье*, которое описывает формула (5). Для этого необходимо дополнить данный нам временной ряд нулями. Получаем новый ряд длиной $N_2 = 2N_1,\ N_1 = 2^p \geq N$. После применяем FFT:

$$X_j = \mathrm{FFT}_j\{x_k^0\}_{k=0}^{N_2-1} = \sum_{k=0}^{N_2-1} x_k^0 e^{-i\frac{2\pi}{N_2}kj}, \tag{16}$$

$$j = \overline{0,\ N_2-1},$$

после чего рассчитывается *периодограмма Шустера* по формуле (1)

$$D_j = \frac{1}{N^2}\left[(\mathrm{Re}\ X_j)^2 + (\mathrm{Im}\ X_j)^2\right]. \tag{17}$$

$$j = \overline{0,\ N_2-1},$$

Отсчеты периодограммы соответствуют частотам

$$\nu_j = j\Delta\nu, \quad j = \overline{0,\ N_2/2}, \quad \Delta\nu = \frac{1}{N_2\Delta t}.$$

E. Оценка дисперсии ряда

Оценка дисперсии временного ряда производится по формуле (см. Калинина)

$$\sigma_o^2 = \frac{1}{N-1}\sum_{k=0}^{N-1}(x_k^0)^2. \tag{18}$$

F. Построение периодограммы

Теперь представим в виде графика периодограмму, а также отождествим выраженные спектральные линии. Строим график $D(\nu)$. Пики частот, которые превышают критический уровень $D_{\mathrm{crit}} = \sigma_0^2 X_1/N$, являются указанием на существование в зависимости периодической компоненты.

Пусть q — вероятность того, что на N точек периодограммы находится одна со значением $D > D_{\mathrm{crit}}$ при отсутствии периодического сигнала. Тогда вероятность обнаружения периодического сигнала при наличии на периодограмме хотя бы одной такой точки меньше либо равна $1-q$. Значения q и X_1 связаны (см. Витязева). При анализе они задаются заранее.

G. Вычисление коррелограммы

Используя формулы (3) и (9), получаем выражение для коррелограммы

$$\bar{c}_m = \frac{1}{N}\mathrm{Re}\ \mathrm{FFT}_m^{-1}[|X_j|^2]_{j=0}^{N_2-1}, \tag{19}$$

$$m = \overline{0,\ N^*-1}.$$

H. График коррелограммы

На данном этапе мы строим коррелограмму, т. е. график $\bar{c}_m(t_k)$.

I. Взвешенная коррелограмма

Дальше можно получить сглаженную периодограмму. Для этого необходимо коррелограмму умножить на функцию Тьюки:

$$W_m = (1-a) + 2a\cos\frac{\pi m}{N^*}. \tag{20}$$

Взвешенная коррелограмма вычисляется таким образом

$$\tilde{c}_m = \bar{c}_m W_m, \quad m = \overline{0,\ N^*-1}. \tag{21}$$

J. Сглаженная периодограмма

Ряд $\tilde{c}_m$ дополняют нулями до заданной длины ряда N_2. После вычисляется сглаженная периодограмма так

$$\tilde{D}_j = \frac{1}{N^*}\left[2\mathrm{Re}\ \mathrm{FFT}_j\{\tilde{c}_m\}_{m=0}^{N_2-1} - \tilde{c}_0\right], \quad j = \overline{0,\ N_2/2}. \tag{22}$$

K. Построение графика сглаженной периодограммы

Строим графики $\tilde{D}_j(\nu_j)$. Отсчеты по частотам совпадают у D_j и $\tilde{D}_j$.

IV. ПРИМЕНЕНИЕ МЕТОДА

Возьмем модельный временной ряд, заданный формулой (10). Сам ряд представлен в файле time_series.dat. Считаем, что нам заданы $\Delta t = 1$ с, $N = 230,\ q = 0.01,\ X_1 = 9.0,\ A_1 = 1,\ \nu_1 = 1$ Гц, $\varphi_1 = 0,\ \gamma = 0.50$.
Следуя нашем алгоритму, строим график временного ряда $x_k = x(t_k)$ (см. рис 1). Используя формулы (13) и (14), получаем оценку коэффициентов аппроксимирующей линейной функции:

$$\hat{\alpha} \approx -0.11 \pm 0.16,$$
$$\hat{\beta} \approx 0.05185 \pm 0.0013.$$

После этого проводим пересчет x_k в x_k^0. И строим график центрированного ряда $x_k^0(t_k)$ (см. рис 2).
Далее, пользуясь формулами (16), (17), получаем значения D_j и ν_j для построения периодограммы. Затем получаем по формуле (18) оценку дисперсии ряда

$$\sigma_0^2 \approx 1.661745.$$

Теперь строим периодограмму временного ряда $D(\nu)$ (см. рис 3). Проводим критический уровень $D_{\mathrm{crit}} = \sigma_0^2 X_1/N \approx 0.065$ и находим яркую спектральную линию на частоте 0.1 Гц.

Мы взяли $a = 0.25$, $N^* = 0.5N$, $N^{**} = 0.2N$. По формуле (19) рассчитываем выражения $\bar{c}_m$ для построения коррелограммы. Строим график коррелограммы $\bar{c}_m(t_k)$ (см. рис 4). Используя формулы (20) и (21), рассчитываем значения $\tilde{c}_m$ взвешенной коррелограммы.
Последнее, что остается это рассчитать по формуле (22) взвешенную периодограмму $\tilde{D}_j$ и ее построить. График взвешенной периодограммы представлен на рис. 5 и 6.

V. ЗАКЛЮЧЕНИЕ

В данной работе приведен алгоритм, позволяющий анализировать временные ряды путем построения периодограмм и коррелограмм, а также показано применение данного метода на примере модельного ряда. Периодограммы позволили выявить периодические составляющие и частоты, описывающие ряд. Коррелограммы выявили автокорреляцию в данных, что помогло построить сглаженные периодограммы.

ИСПОЛЬЗУЕМЫЕ МАТЕРИАЛЫ

В. Н. Калинина, В. Ф. Панкин *«Математическая статистика»*
В. В. Витязев *«Спектрально-корреляционный анализ равномерных временных рядов. Учебное пособие»*
Eric Feigelson (Penn State University) *«Time Series Analysis»*
James Holland Jones *«Time Series and Spectral Analysis»*
R. Vio et al. (2005) *«Time series analysis in astronomy: Limits and potentialities»*
Mark Fiecas et al. (2018) *«Spectral Analysis of High-dimensional Time Series»*

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

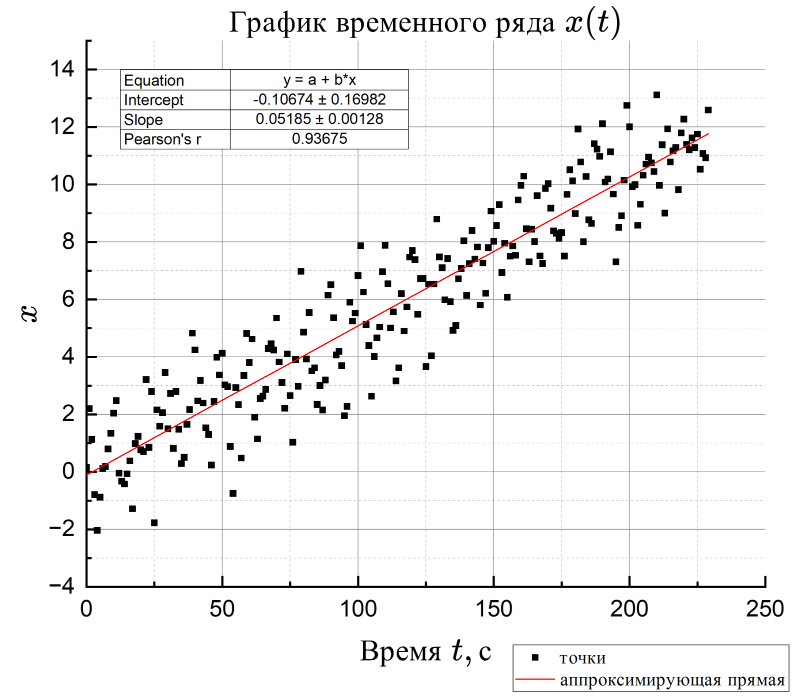

Рис. 1. График модельного временного ряда $x_k = x(t_k)$.

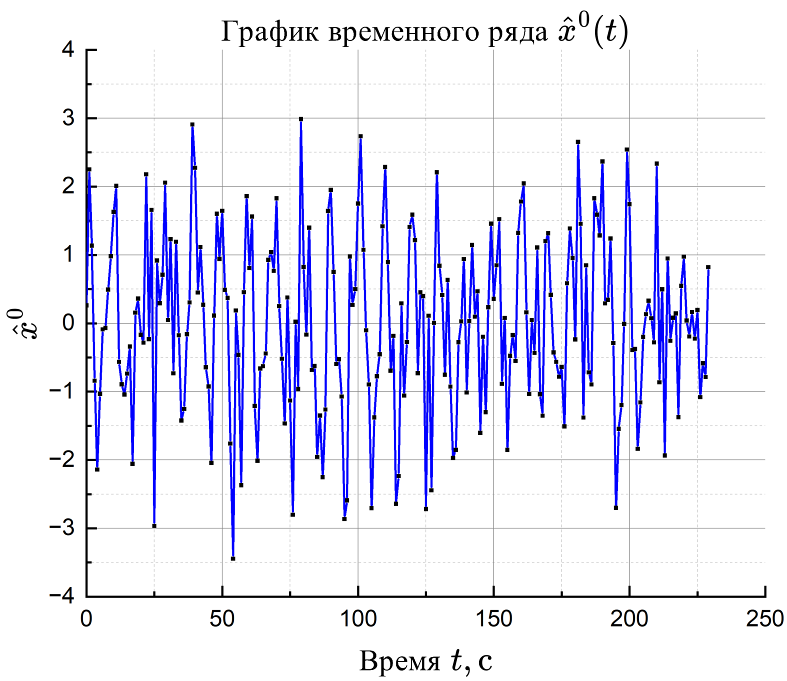

Рис. 2. График центрированного временного ряда $\hat{x}^0 = \hat{x}^0(t_k)$.

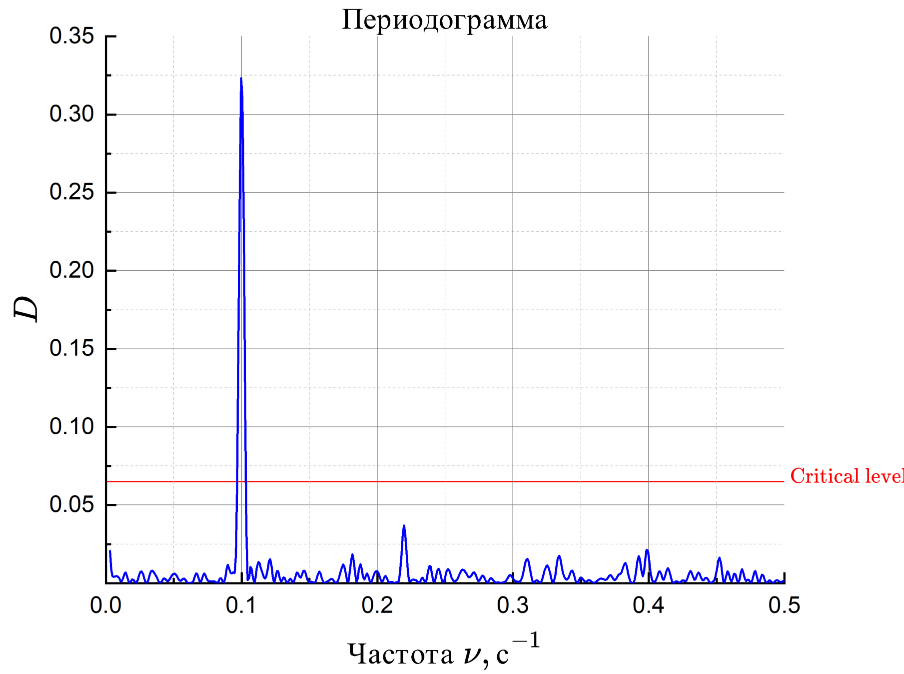

Рис. 3. *Синяя кривая* — периодограмма Шустера временного ряда в зависимости от частоты $D(\nu)$. *Красная линия* — критический уровень детектирования периодического сигнала, $D_{\mathrm{crit}} \approx 0.065$

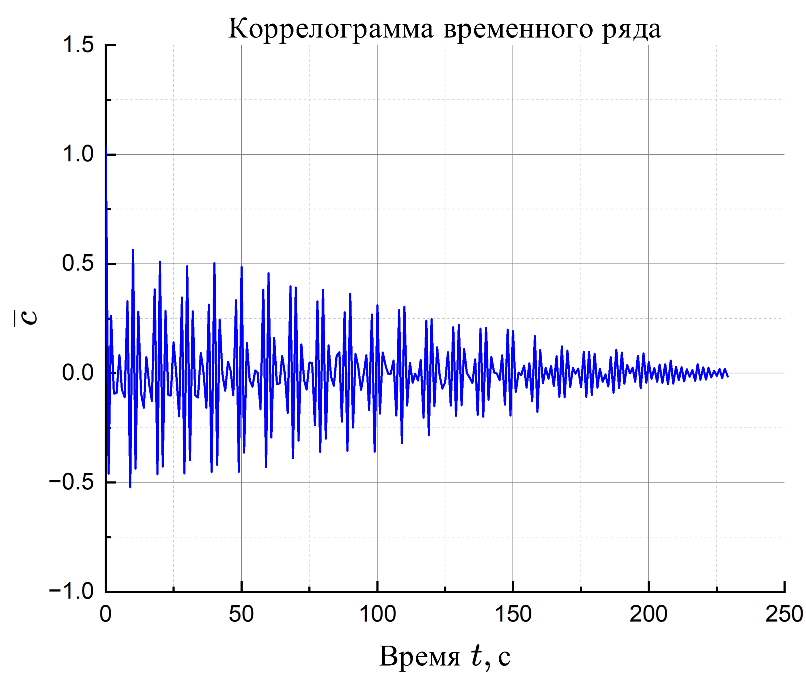

Рис. 4. Коррелограмма временного ряда $\bar{c}(t)$.

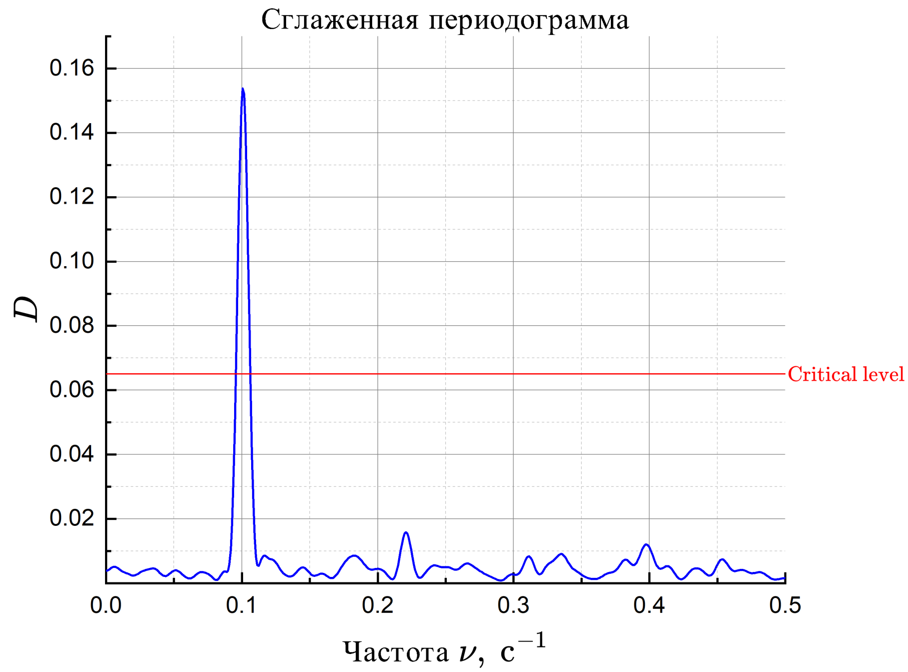

Рис. 5. *Синяя кривая* — сглаженная периодограмма $D(\nu)$ при параметре $N^* = 0.2N$. *Красная кривая* — критический уровень детектирования периодического сигнала, $D_{\rm crit} \approx 0.065$

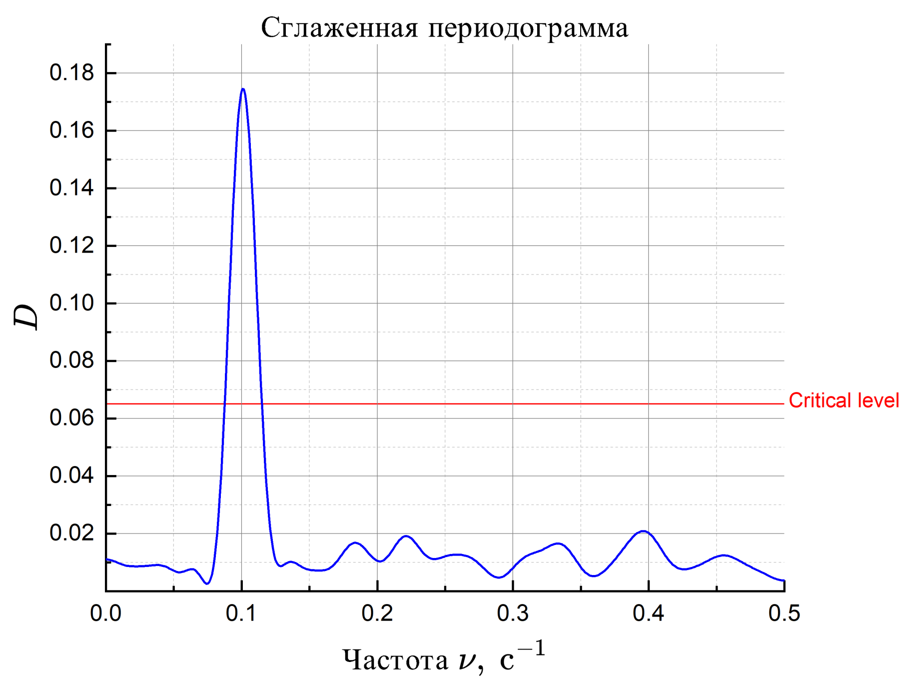

Рис. 6. *Синяя кривая* — сглаженная периодограмма $D(\nu)$ при параметре $N^* = 0.5N$. *Красная кривая* — критический уровень детектирования периодического сигнала, $D_{\rm crit} \approx 0.065$