

Данилов В.И. и др.
Б1-10-8579

+

Ц 840 В

Д-183



1136/75

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ЯДЕРНЫХ ИССЛЕДОВАНИЙ

Б1-10-8579

ДЕПОНИРОВАННАЯ ПУБЛИКАЦИЯ

Дубна 1975₂

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ЯДЕРНЫХ ИССЛЕДОВАНИЙ

Лаборатория ядерных проблем

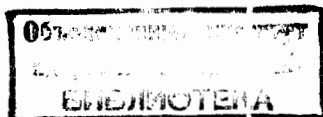
Лаборатория вычислительной техники и автоматизации

В.И.Данилов, К.Н.Данилова, Л.М.Сороко, Т.А.Стриж

Б1-10-8579

АНАЛИЗ БОЛЬШИХ МАССИВОВ ДАННЫХ

С ПОМОЩЬЮ АЛГОРИТМА ФУРЬЕ



НАУЧНО-ТЕХНИЧЕСКАЯ БИБЛИОТЕКА

4 окт 75

г.Дубна, ноябрь 1974 г.

АННОТАЦИЯ

Рассмотрено применение алгоритма Фурье для анализа больших массивов данных. В качестве исходного массива взяты среднемесячные и среднесуточные числа Вольфа. Разработаны схемы определения уровня шумов измерений в частотной области, по величине которого делаются оценки значимости информационных компонент анализируемого явления или процесса.

Приведены затраты времени ЭВМ для вычисления Фурье - образа.

ВВЕДЕНИЕ

Спектральные методы анализа цифровых массивов данных с использованием быстрых алгоритмов счета постоянно начинают использоваться во многих разделах физики и техники. В обзоре^{/1/} приведена общая теория алгоритмов быстрого преобразования Фурье, рассмотрены типичные программы вычислений на цифровых машинах Фурье-образа вещественных и комплексно-значных функций.

Там же подробно проанализированы специфические особенности проведения вычислений на ЭВМ с применением быстрых алгоритмов преобразования Фурье.

Приведенные в работе^{/2/} и других работах^{/3,4,5,6/} примеры использования алгоритма Фурье для анализа спектрометрической информации относятся к сравнительно небольшим массивам данных. Между тем быстрые алгоритмы Фурье имеют особенно большие преимущества при обработке достаточно больших массивов данных.

В настоящей работе даны примеры анализа больших массивов данных с использованием быстрых алгоритмов преобразования Фурье. В качестве исходных данных были взяты массивы чисел Вольфа \bar{W} для среднемесячных и среднесуточных измерений^{/7/}.

Исходные данные записываются в виде функции $f(t)$, $t = 0, 1, 2, \dots, N-1$, где N - полное число точек, превышающее число измерений n на такую величину, чтобы $N = 2^{\bar{m}}$, где \bar{m} - целое число. Дискретный Фурье-образ функции $f(t)$ равен по определению

$$F(\omega) = \sum_{t=0}^{N-1} f(t) e^{-i \frac{2\pi}{N} \omega t}$$
$$\omega = 0, 1, 2, \dots, N-1.$$

Дискретный Фурье-образ $F(\omega)$ является комплекснозначной функцией и его можно характеризовать двумя вещественными функциями. В первом варианте это вещественная

$$\operatorname{Re} F(\omega) = \sum_{t=0}^{N-1} f(t) \cos \frac{2\pi}{N} \omega t$$

и

$$\operatorname{Im} F(\omega) = \sum_{t=0}^{N-1} f(t) \sin \frac{2\pi}{N} \omega t$$

части. Во втором варианте - это абсолютное значение Фурье-образа

$$|F(\omega)| = \sqrt{[\operatorname{Re} F(\omega)]^2 + [\operatorname{Im} F(\omega)]^2}$$

и фазовый угол

$$\varphi(\omega) = \operatorname{arctg} \frac{\operatorname{Im} F(\omega)}{\operatorname{Re} F(\omega)}$$

Преимущества, которые дает алгоритм преобразования Фурье, заключается в том, что расположение частотных пиков практически не меняется, если наблюдаемый процесс связан не с самой исходной функцией $f(t)$, а с её производной. Действительно, при переходе от исходной функции $f(t)$ к её производной $\frac{d^m f(t)}{dt^m}$ Фурье-образ $F(\omega)$ исходной функции $f(t)$ умножается на множитель $(i\omega)^m$, который изменяет лишь форму пиков, а число и расположение пиков практически не изменяет. Аналогичное свойство имеет место для некоторых других линейных операторов. Только нелинейные операции создают новые частотные компоненты, которых не было в исходном сигнале, и которые являются комбинациями исходных частот.

Алгоритмы вычисления Фурье-образа

Дискретный Фурье-образ последовательности чисел Вольфа W вычислялся двумя способами. Первый способ использует вариант Бриджа^{/8/} алгоритма быстрого преобразования Фурье. Число $N = 3072$. Этот вариант быстрого алгоритма используют при малой оперативной памяти ЭВМ. Для этого исходную последовательность разбивают на несколько последовательностей длиной N' , соответствующей возможностям ЭВМ, а $N = N' \cdot n$. Элементам полного массива присваивают два индекса: k и j , так что компонента Фурье-образа $B(k, j)$ соответствует $(k + N'j)$ -ой компоненте Фурье-образа полного массива. Тогда

$$B(k, j) = \sum_{m=0}^{N-1} f(m) W_N^{(k + N'j)m}$$

$$k = 0, 1, 2, \dots, N' - 1$$

$$j = 0, 1, 2, \dots, n - 1$$

$$a \quad W_N^p = \exp\left(\frac{2\pi i}{N} \cdot p\right)$$

Замена $m = (l \cdot n + q)$ дает /8/:

$$B(k, j) = \sum_{q=0}^{n-1} \left[\sum_{l=0}^{N'-1} f(l \cdot n + q) W_{N'}^{kl} \right] W_N^{(k + N'j)q}$$

Выражение в квадратных скобках находится по стандартной программе вычисления БПФ на данной ЭВМ. Алгоритм повторно применяют n раз для каждой подпоследовательности исходного массива, затем вычисляют массив $B(k, j)$, из элементов которого формируют все N компонент Фурье-образа.

Вычисления проводились на ЭВМ СДС 1604-А, полное время счета с учетом времени на трансляцию и печать результатов составляло около 5 минут для $N = 3072$.

Второй способ вычислений основан на усовершенствованной программе БПФ^{/9/}, которая требует меньшего объема оперативной памяти и меньших затрат времени для вычислений, чем в первом способе.

Для этого исходный массив чисел Вольфа дополнялся нулями до $N = 4096$, т.е. до $N = 2^{\bar{m}}$, где \bar{m} - целое число. Такое дополнение, как известно^{/1/}, не влияет на результаты счета. Здесь нет необходимости использовать вариант Бюиджа.

Вычисления по второму способу производились на ЭВМ БЭСМ-6. Счетное время равнялось 2 мин., а коммерческое - около 8 минут.

Результаты

На рис. I приведены абсолютные значения Фурье-образа ($|F(\omega)|$) массива среднемесячных чисел Вольфа для $N = 3072$. Интервал между двумя соседними измерениями равен одному месяцу. Частота ω принимает значение от $\omega = 0$ до $\omega_{\max} = \frac{N}{2}$. Уменьшение числа частотных компонент связан с тем, что исходная функция - вещественна. Выбранному интервалу отсчета среднемесячных чисел Вольфа W и размеру массива данных соответствует связь $\omega T = 256$ между частотой ω и периодом T .

На Рис. I видно, что числа Вольфа W имеют очень сложную периодическую структуру. Наиболее ярко выраженные пики расположены на частотах $\omega \sim 23$, которые соответствуют периоду $T \sim 11$ лет. Кроме этого наблюдается еще много других пиков. Чтобы решить, какие из этих пиков являются информационными, а какие - случайными выбросами, требуется оценить уровень шума и его изменение по частотному диапазону.

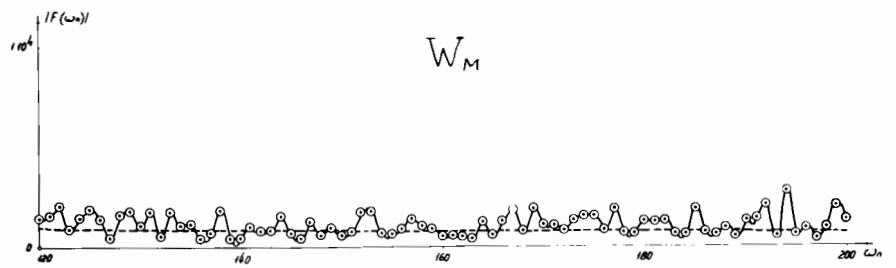
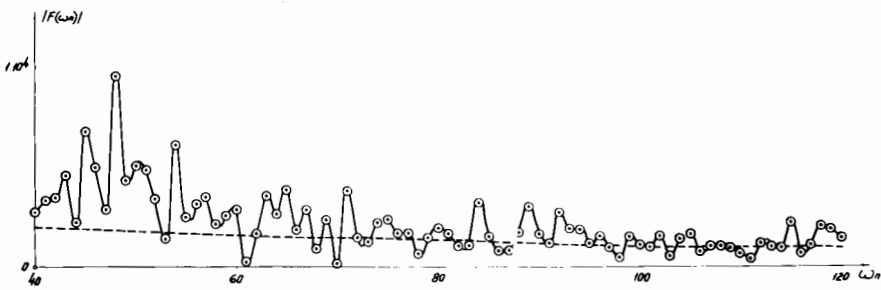
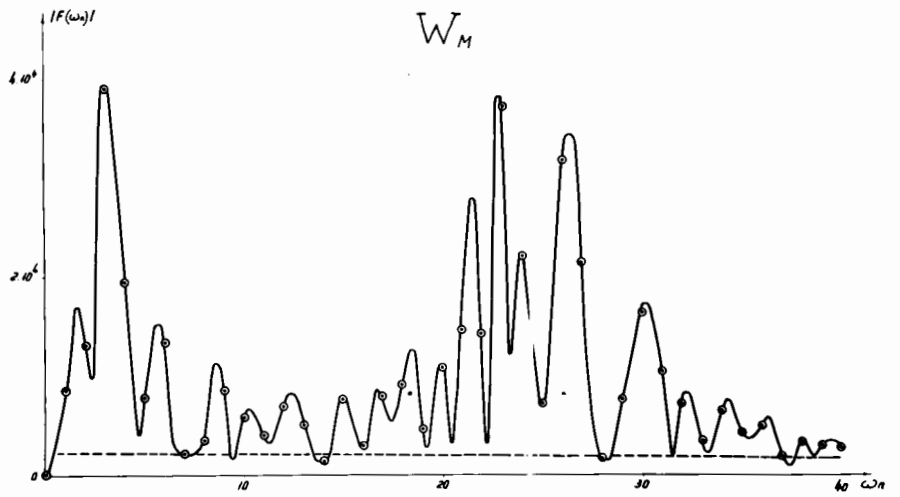


Рис. 1. Абсолютные значения фурье-образа $|F(\omega)|$ массива среднемесячных чисел Вольфа для $N = 3072$.

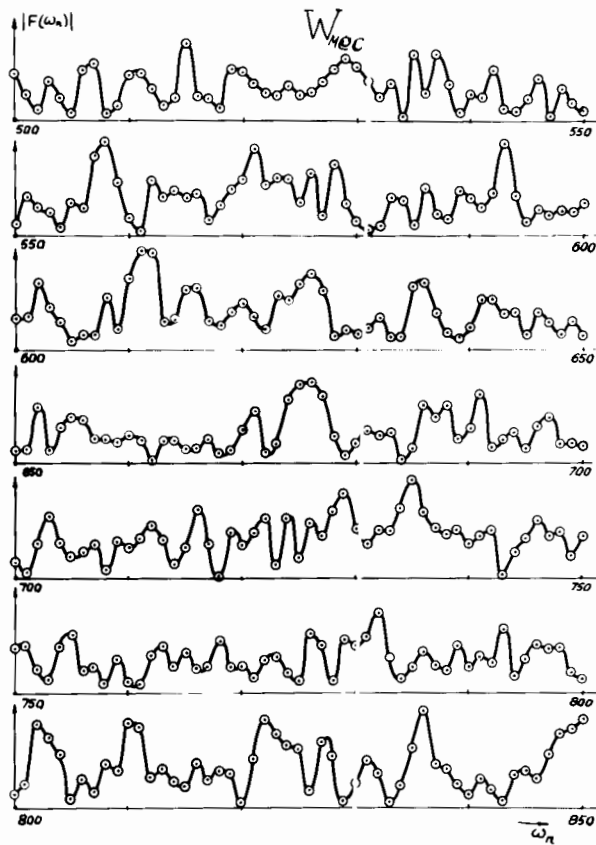
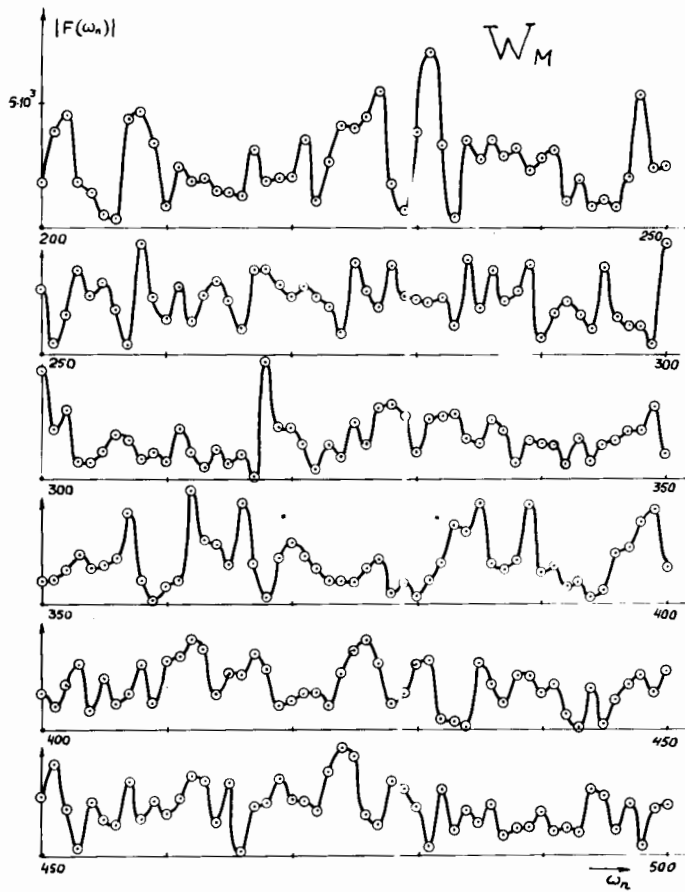


Рис. I. Продолжение.

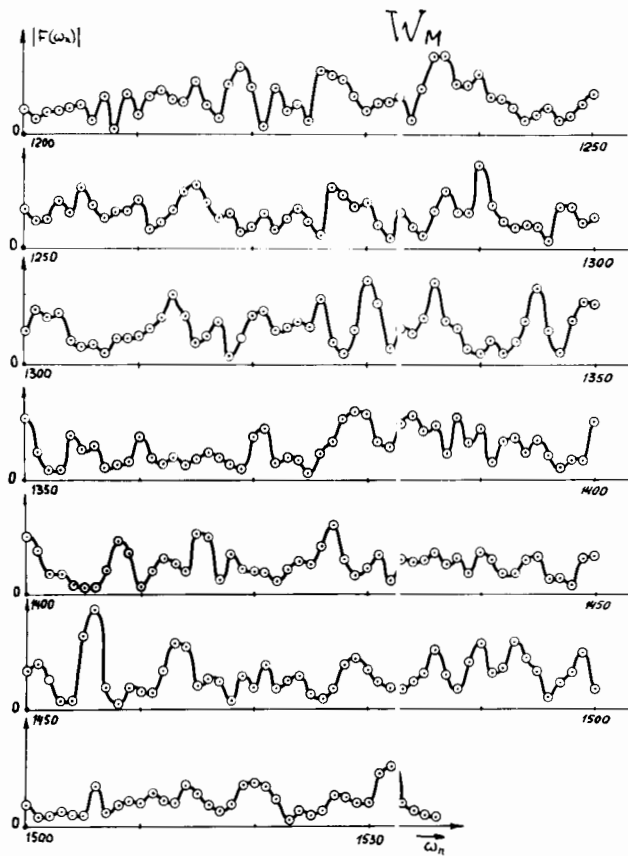
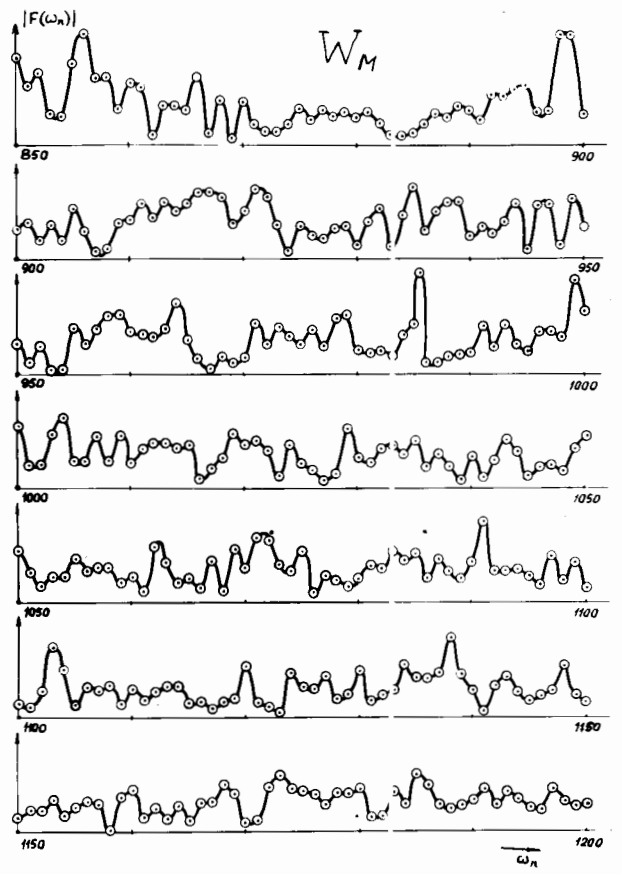


Рис. I. Окончание.

Определение уровня шума

В данной работе уровень шума F_{ω} был оценен двумя методами.

I. Если предположить, что каждое измерение $f(t)$ является независимым от соседних, то функция взаимной корреляции ошибок измерения равна дельта-функции. Это в свою очередь означает, что в частотной области ω ошибки измерения или шум измерений равен постоянной величине при всех значениях ω . Таким образом уровень шума можно определить в области больших ω , а затем экстраполировать в область низких частот.

На Рис.2 дана функция распределения абсолютного значения Фурье-образа $|F(\omega)|$ в области $N = 1100 + 1536$ для среднемесячных чисел Вольфа. На рис.3 дано положение компонент в плоскости $Re F(\omega), Im F(\omega)$. Из этих двух рисунков видно, что радиусы-векторы и фазовые углы распределены случайно, а шумовая компонента имеет среднеквадратичное значение $\sqrt{|F|^2_{\text{средн.}}} \approx 0,8 \cdot 10^3$. Этот уровень шума экстраполировался на информационную область низких частот.

II. Структура шума измерений была проанализирована также для серии среднесуточных измерений чисел Вольфа, для которой, в силу самого характера измерений могла возникнуть некоторая корреляция между соседними измерениями, производимыми с промежутком в одни сутки.

На Рис.4 приведены абсолютные значения Фурье-образа $|F(\omega)|$ среднесуточных чисел Вольфа. При анализе шумов весь диапазон частот ω был разбит на 6 интервалов:

- | | |
|-------------------------------|---------------------------------|
| 1. $0 \leq \omega \leq 100$ | 4. $600 \leq \omega \leq 900$ |
| 2. $100 \leq \omega \leq 300$ | 5. $900 \leq \omega \leq 1200$ |
| 3. $300 \leq \omega \leq 600$ | 6. $1200 \leq \omega \leq 1500$ |

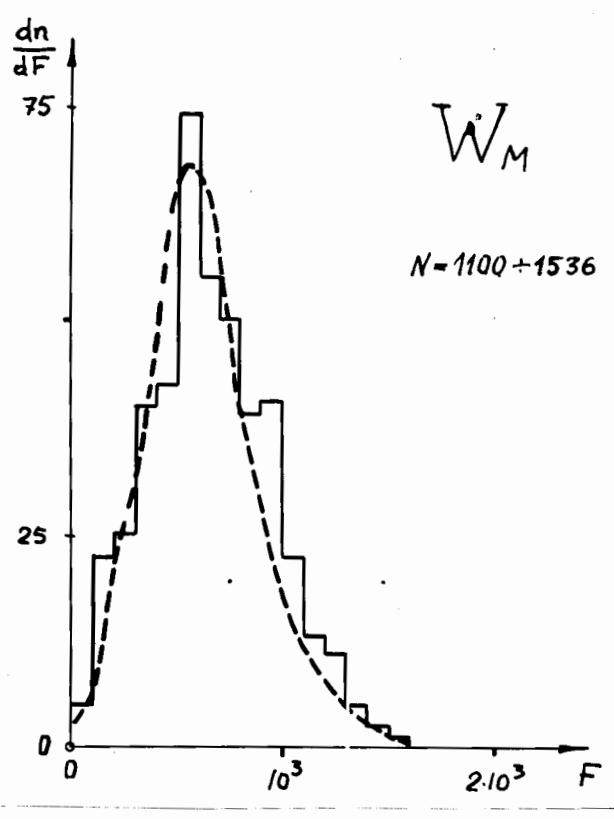


Рис. 2. Распределение абсолютных значений фурье-образа $|F(\omega)|$ средне-месячных чисел Вольфа в области частот $1100 \leq \omega \leq 1536$.

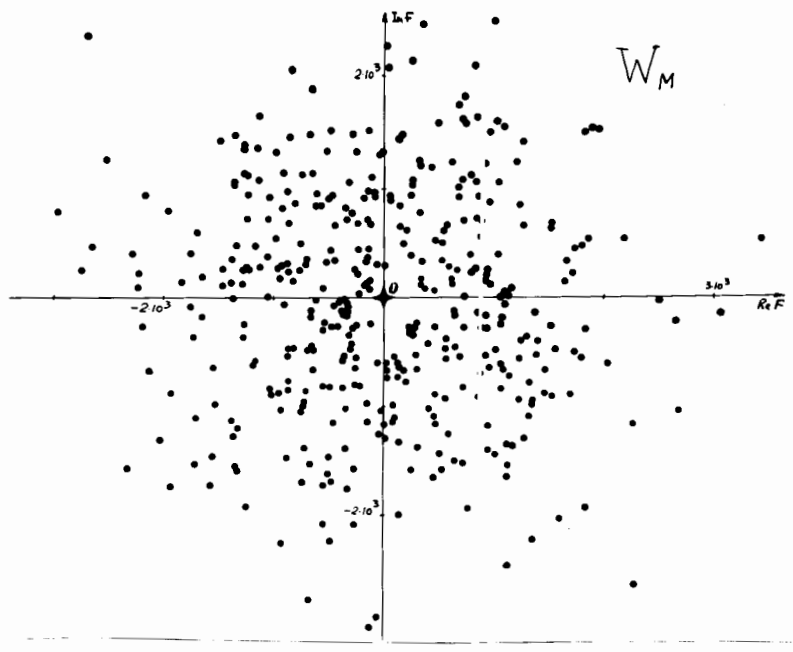


Рис. 3. Расположение компонент фурье-образа $|F(\omega)|$ средне-месячных чисел Вольфа в плоскости (Re, Im) для частот $1100 \leq \omega \leq 1536$.

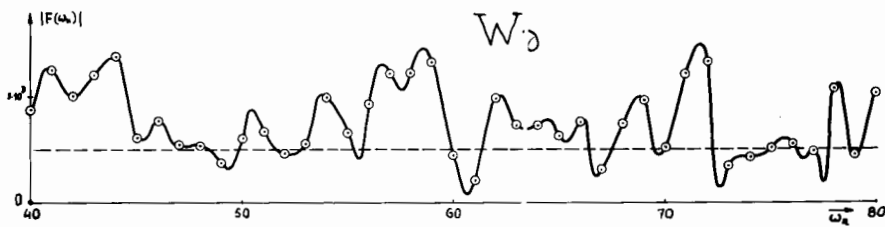
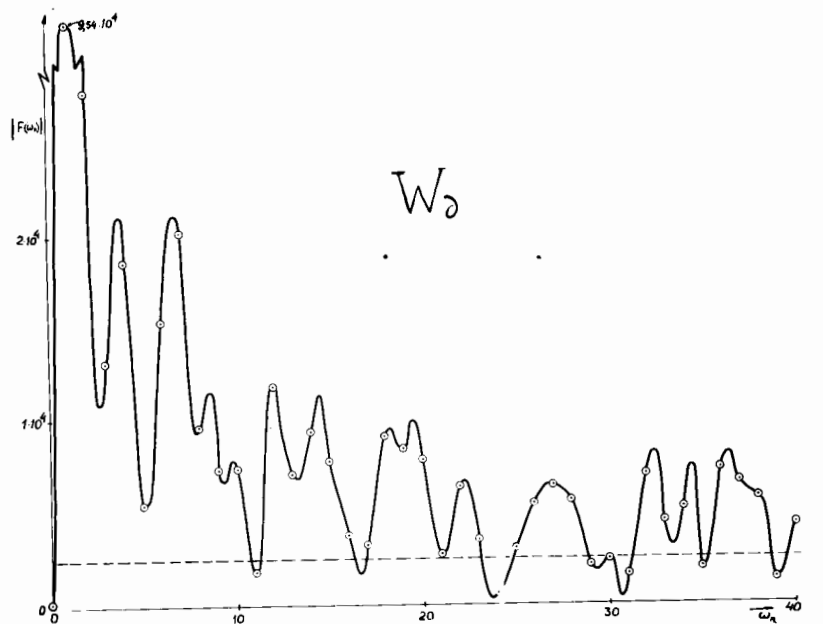


Рис. 4. Абсолютные значения Фурье-образа $|F(\omega)|$ массива среднесуточных чисел Вольфа.

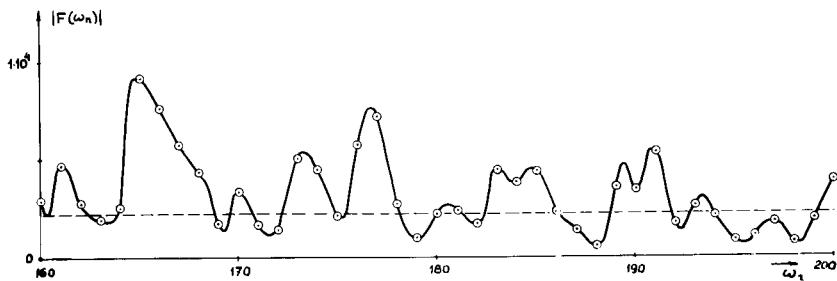
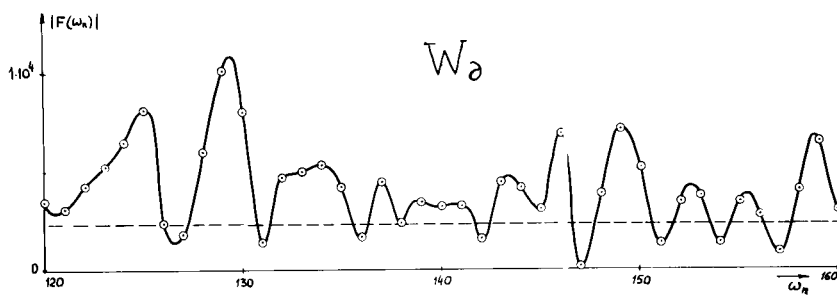
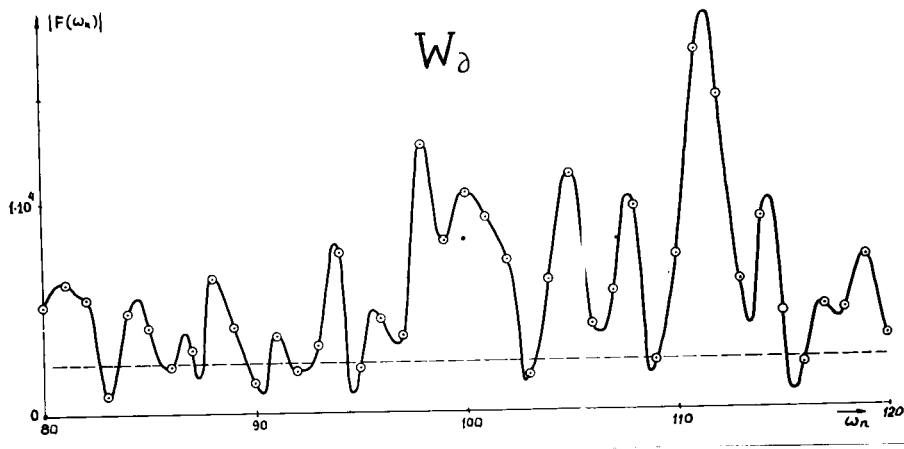
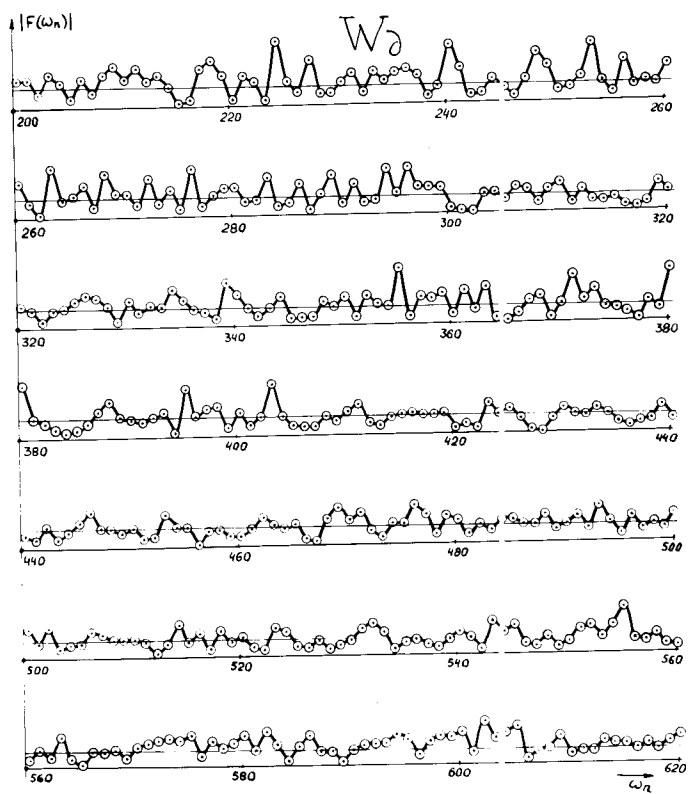


Рис. 4. Продолжение.



Рмс. 4. Продолжение.

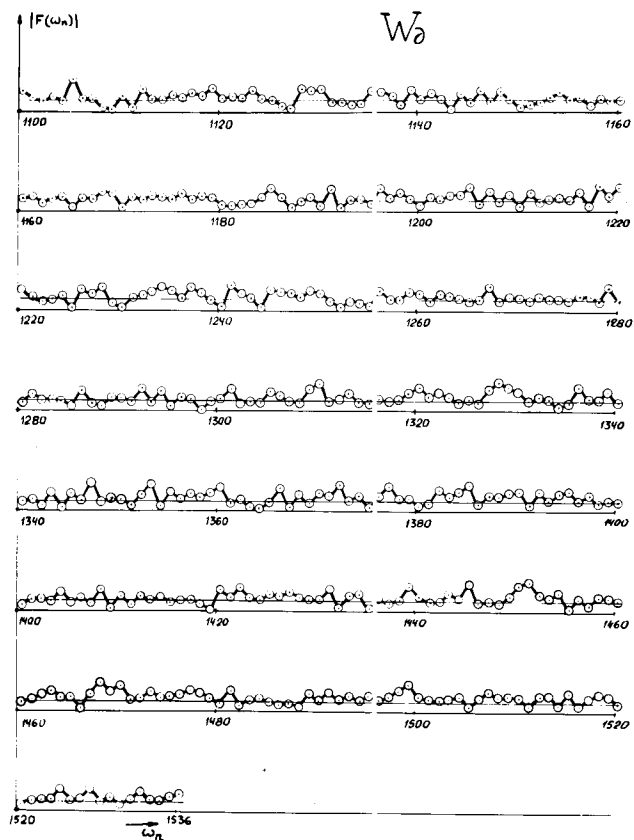
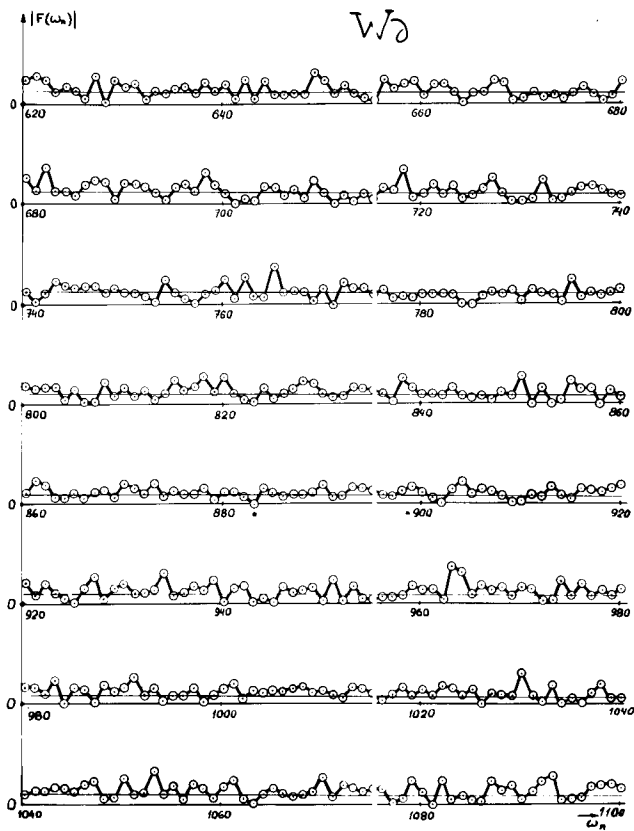


Рис. 4. Окончание.

Для каждого интервала было построено экспериментально наблюдаемое распределение абсолютного значения Фурье-образа $|F(\omega)|$ (Рис.5,6,7,8,9,10). Видно, что для трех последних интервалов это распределение является случайным со среднеквадратичными значениями шума:

$$F_{\omega} = 1,25 \cdot 10^3 \quad (1200 \leq \omega \leq 1500)$$

$$F_{\omega} = 1,3 \cdot 10^3 \quad (900 \leq \omega \leq 1200)$$

$$F_{\omega} = 1,35 \cdot 10^3 \quad (600 \leq \omega \leq 900)$$

В остальных трех диапазонах (Рис.8,9,10) четко проявляются информационные компоненты, значение которых существенно превышает уровень шума. На рисунках информационные компоненты помечены заштрихованными квадратами. Средне-квадратичные шумы в этих интервалах равны

$$F_{\omega} = 1,5 \cdot 10^3 \quad (300 \leq \omega \leq 600)$$

$$F_{\omega} = 2 \cdot 10^3 \quad (100 \leq \omega \leq 300)$$

$$F_{\omega} = 2,5 \cdot 10^3 \quad (0 \leq \omega \leq 100)$$

На Рис.11 значение шума F_{ω} нанесены для всех шести интервалов. Видно, что независимые оценки F_{ω} по интервалам изменяются незначительно.

На Рис.12 приведены двумерные диаграммы Фурье-образа $F(\omega)$ в координатах $\text{Re } F(\omega)$, $\text{Im } F(\omega)$ для среднемесячных чисел Вольфа в диапазоне частот ^{от} $0 \leq \omega \leq 27$, а на Рис.13 в диапазоне $28 \leq \omega \leq 51$. Кривые на плоскости проведены от руки и они соответствуют некоторому "ручному сплайну". Масштабы оси ординат на Рис.12 и Рис.13 не совпадают. Указанные для точек $\omega = 26$ и $\omega = 31$ ошибки соответствуют предположению, что шум F_{ω} является векторно аддитивной величиной.

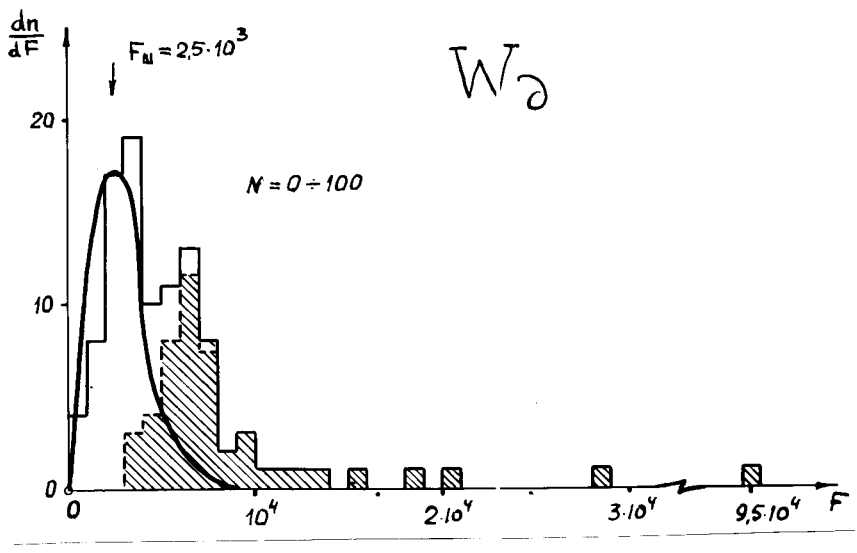


Рис. 5. Распределение абсолютных значений фурье-образа $|F(\omega)|$ среднесуточных чисел Вольфа в области частот $0 \leq \omega \leq 100$.

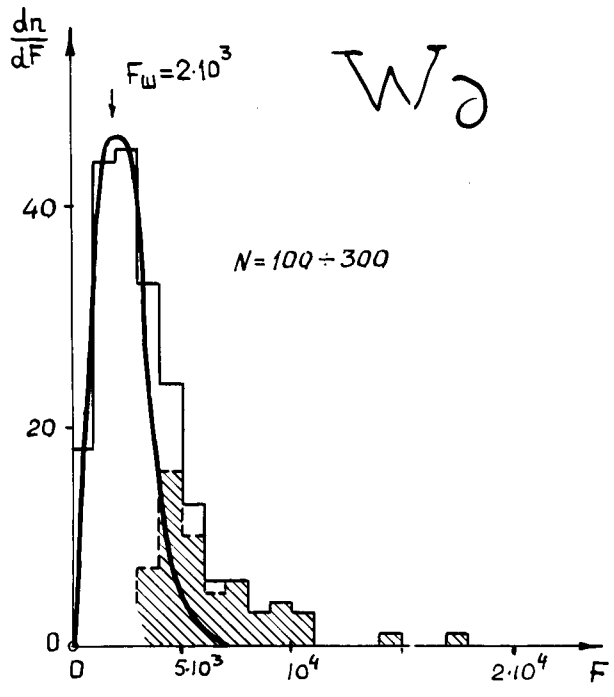


Рис. 6. Распределение абсолютных значений фурье-образа $|F(\omega)|$ среднесуточных чисел Вольфа в области частот $100 \leq \omega \leq 300$.

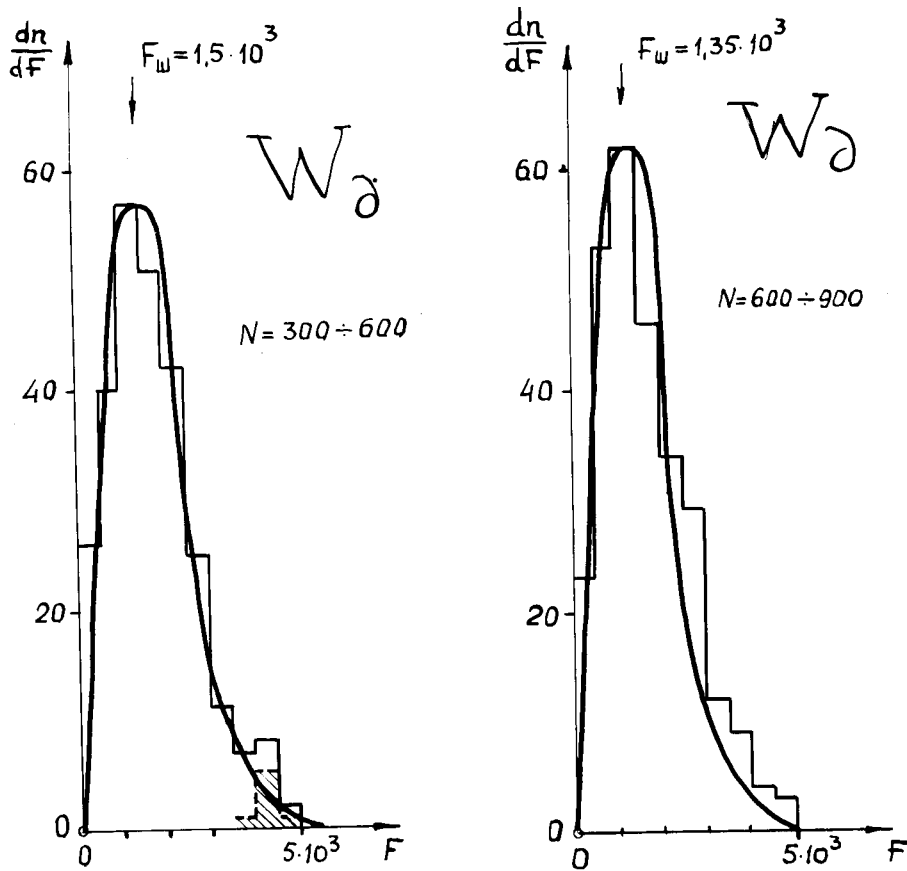


Рис. 7 и 8. Распределение абсолютных значений фурье-образа $|F(\omega)|$ среднесуточных чисел Вольфа в области частот:

$$300 \leq \omega \leq 600 \text{ /слева/}$$

$$600 \leq \omega \leq 900 \text{ /справа/}$$

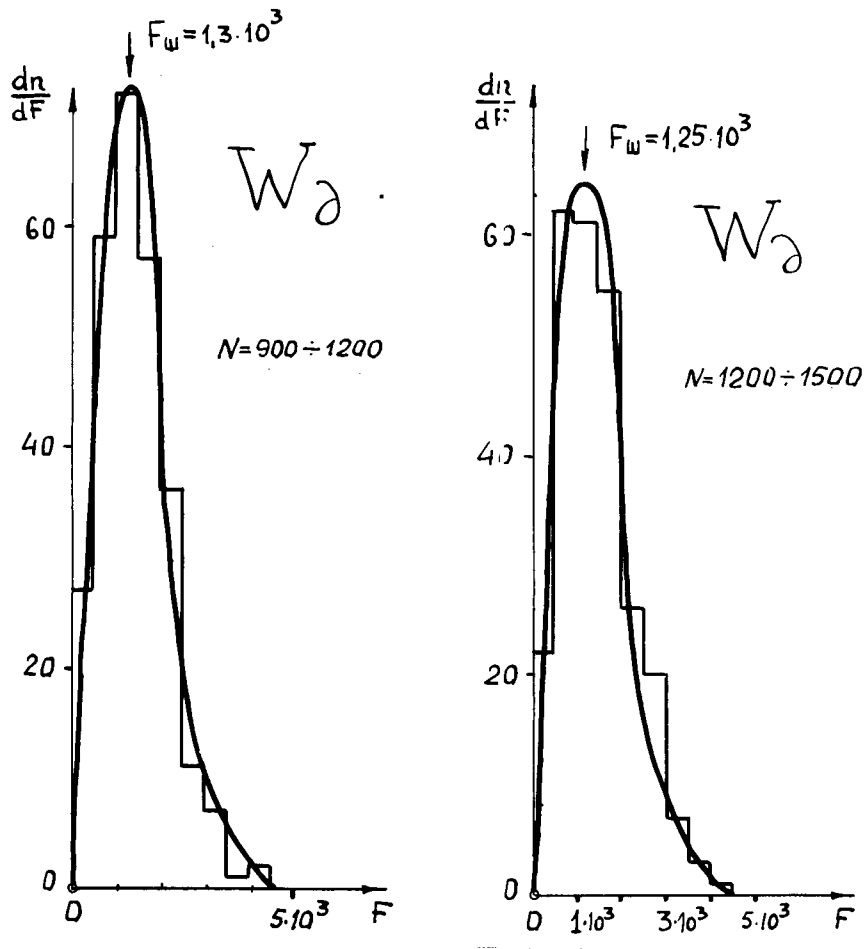


Рис. 9 и 10. Распределение абсолютных значений фурье-образа
среденсutoчных чисел Вольфа в области частот:

$$900 \leq \omega \leq 1200 \quad / \text{слева} /$$

$$1200 \leq \omega \leq 1500 \quad / \text{справа} /$$

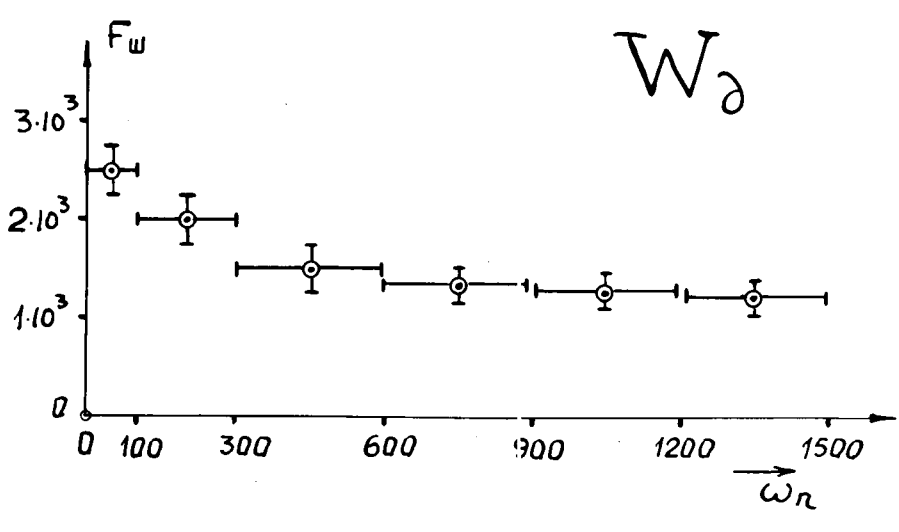


Рис. II. Значения шума F_w для шести интервалов частот ω_n .

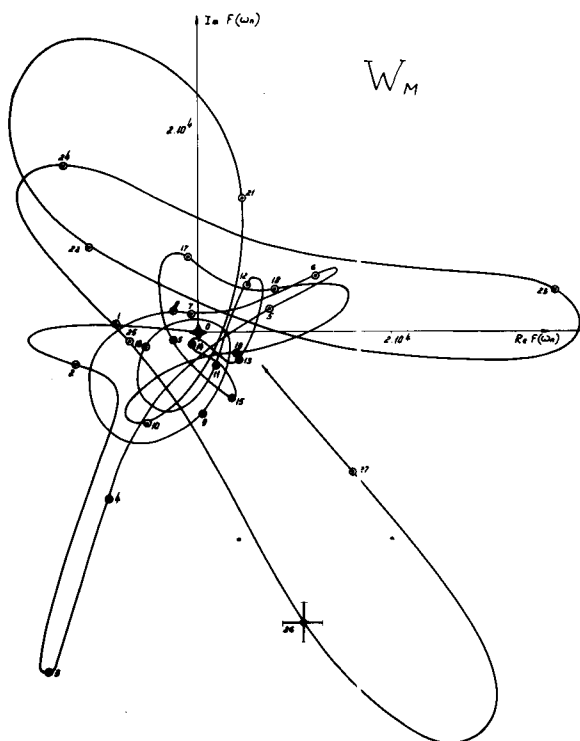


Рис. 12. Двухмерные диаграммы значений фурье-образа
 среденмесячных чисел Вольфа в диапазоне частот $0 \leq \omega \leq 27$.

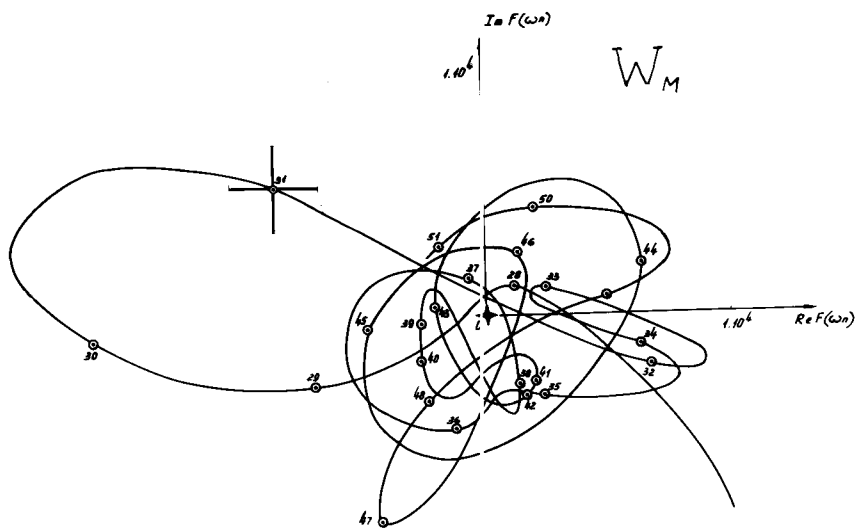


Рис. 13. Двухмерные диаграммы значений фурье-образа
 среденмесячных чисел Вольфа в диапазоне частот $28 \leq \omega \leq 51$.

Информационные компоненты

В табл. I приведены наиболее ярко выраженные информационные компоненты Фурье-образа среднемесячных чисел Вольфа. В качестве естественных реперов взяты частоты ω , соответствующие периоду обращения планет вокруг Солнца.

В третьей графе даны отношения высоты пика к среднеквадратичному шуму.

В табл. II приведены значения информационных компонент, реперами для которых взяты суммарные и разностные частоты ^{от} обращения планет вокруг Солнца.

В табл. III реперами являются обертоны частоты ω для планеты Юпитер, а в табл. IV - для комбинации трех и четырех частот.

Аналогичные данные приведены для среднесуточных измерений чисел Вольфа в табл. V + XI.

ВЫВОДЫ

1. Показано, что алгоритм преобразования Фурье является весьма эффективным при анализе больших массивов данных, имеющих сложную периодическую структуру.

2. Алгоритм Фурье позволяет оценить эффекты шумов измерений в условиях, когда о дисперсии измеряемых значений нет никаких данных.

3. Время, затраченное на обработку массива данных с $N = 4096$ составляет на ЭВМ БЭСМ-6 около 2 минут счетного времени и около 8 минут коммерческого времени.

Таблица I.

Наиболее интенсивные компоненты фурье-образа среднемесячных чисел Вольфа. Реперные частоты ω_n обратно пропорциональны периодам обращения планет вокруг Солнца.

	Нептун	Уран	Сатурн	Юпитер	Марс	Земля
	ψ	♁	♄	♃	♂	♁
ω_n	1,55	3,05	8,7	21,6	136,1	256
F/F_ω	5,9	17,8	4,6	10,0	0,6	2,5

Таблица II.

Наиболее интенсивные компоненты фурье-образа среднемесячных чисел Вольфа. Реперные частоты ω_n равны сумме и разности частот обращения планет вокруг Солнца.

	$\text{♁} \pm \psi$	$\text{♄} \pm \psi$	$\text{♄} \pm \text{♁}$	$\text{♃} \pm \psi$	$\text{♃} \pm \text{♁}$	$\text{♃} \pm \text{♄}$
ω_n +	4,6	10,2	11,7	23,1	24,6	30,3
ω_n -	1,5	7,1	5,6	20,0	18,5	12,9
F/F_ω +	4,6	2,7	3,1	17,2	5,4	7,8
F/F_ω -	5,6	1,0	6,3	5,0	4,2	2,7

	$\text{♂} \pm \psi$	$\text{♂} \pm \text{♁}$	$\text{♂} \pm \text{♄}$	$\text{♂} \pm \text{♃}$
ω_n +	137,6	139,2	144,8	157,7
ω_n -	134,6	133,1	127,4	114,5
F/F_ω +	2,0	0,6	1,7	2,3
F/F_ω -	1,2	2,0	0,8	2,5

Таблица III.

Значения F/F_{ω} компонент фурье-образа среднемесячных чисел Вольфа для серии частот ω_n , являющихся обертонами частоты обращения Юпитера вокруг Солнца.

	2	2 ^{×2}	2 ^{×3}	2 ^{×4}
ω_n	21,6	43,2	64,8	86,3
F/F_{ω}	10,0	4,7	2,5	2,5

Таблица IV.

Значения F/F_{ω} компонент фурье-образа среднемесячных чисел Вольфа для реперных частот, образованных комбинацией трёх и четырёх частот обращения планет вокруг Солнца.

	2- η - δ	2- η + ψ	2- η + δ	2- δ - ψ	2+ ψ - δ	2+ η - δ
ω_n	9,84	14,4	15,9	17,0	20,1	27,2
F/F_{ω}	3,0	2,0	1,5	4,0	5,3	8,0

	+2- η - δ - ψ	2+ η - ψ	+2+ η + δ - ψ	2+ η + δ	+2+ η + δ + ψ
ω_n	8,3	28,7	31,8	33,3	34,9
F/F_{ω}	4,9	4,0	3,0	1,5	2,9

Таблица У.

Значения F/F_{ω} компонент фурье-образа среднесуточных чисел Вольфа. Реперными являются частоты обращения планет вокруг Солнца, а также сидерическая и синодическая частоты обращения Солнца вокруг своей оси и синодическая и сидерическая частоты вращения Луны вокруг Земли.

	♀	♀	♁	♂	♃
ω_n	34,9	13,7	8,8	4,5	0,7
F/F_{ω}	3,5	4,0	4,0	7,5	40,0

	☉ _{СИДЕР}	☉ _{СИНОД}	☾ _{СИДЕР}	☾ _{СИНОД}
ω_n	121	112,8	112,6	104
F/F_{ω}	1,5	7,4	7,4?	4,8

Таблица УІ.

Значения F/F_{ω} компонент фурье-образа среднесуточных чисел Вольфа. Реперы - синодическая частота обращения Солнца вокруг своей оси и их кратные значения.

	☉ _{СИНОД}	☉ _{СИД} × 2	☉ _{СИД} × 3	☉ _{СИД} × $\frac{1}{2}$	☉ _{СИД} × $\frac{3}{2}$	☉ _{СИД} × $\frac{5}{2}$
ω_n	112,8	226	339	56,4	170	283
F/F_{ω}	7,4	3,5	1,5	2,5	4,0	2,5

Таблица УП.

Значения F/F_{ω} компонент фурье-образа среднесуточных чисел Вольфа. Реперы - сидорические частоты обращения Солнца вокруг своей оси и их кратные значения.

	$\odot_{\text{СИДЕР}}$	$\odot_{\text{СИД}} \times 2$	$\odot_{\text{СИД}} \times 3$	$\odot_{\text{СИД}} \times \frac{1}{2}$	$\odot_{\text{СИД}} \times \frac{3}{2}$	$\odot_{\text{СИД}} \times \frac{5}{2}$
ω_n	121	242	365	61	182	303
F/F_{ω}	1,5	2,0	2,0	2,0	2,2	2,3

Таблица УШ.

Значения F/F_{ω} компонент фурье-образа среднесуточных чисел Вольфа. Реперные частоты равны сумме и разности частот обращения планет вокруг Солнца.

	$\text{♀} \pm \text{♀}$	$\text{♀} \pm \text{♂}$	$\text{♀} \pm \text{♂}$	$\text{♀} \pm \text{♂}$	$\text{♀} \pm \text{♂}$	$\text{♂} \pm \text{♂}$
ω_n +	48,6	43,6	39,4	22,4	18,2	13,2
ω_n -	21,2	26,2	30,4	4,9	9,2	4,3
F/F_{ω} +	1,0	2,9	1,0	2,6	3,7	3,2
F/F_{ω} -	2,6	2,6	1,0	2,4	3,0	6,0

Таблица ІХ.

Значения F/F_{ω} компонент фурье-образа среднесуточных чисел Вольфа. Реперные частоты равны сумме и разности синодической частоты обращения Солнца вокруг своей оси и частот обращения планет вокруг Солнца.

	$\odot \pm \text{♂}$ сик.	$\odot \pm \text{♂}$ сик.	$\odot \pm \text{♀}$ сик.	$\odot \pm \text{♀}$ сик.
ω_n +	117,8	121,6	126,5	147,7
ω_n -	108,3	104,1	99,1	77,9
F/F_{ω} +	2,0	2,0	3,5	3,2
F/F_{ω} -	4,2	4,8	4,4	2,2

Дань
Дань

Волф
Волф

ЛИТЕРАТУРА

1. Л.М.Сороко, Т.А.Стриж " Спектральные преобразования на цифровых вычислительных машинах", ОИЯИ, Р10-6702, Дубна, 1972 г.
2. Л.М.Сороко, сообщение ОИЯИ, I-5030, Дубна, 1970 г.
3. В.М.Дубовик, Б.Л.Марковский, Л.М.Сороко, Т.А.Стриж, сообщение ОИЯИ Р1-5340, Дубна, 1970 г.
4. В.М.Дубовик, Б.Л.Марковский, Л.М.Сороко, Т.А.Стриж, сообщение ОИЯИ, Р1-6071, Дубна, 1971 г.
5. V.M.Dubovik, B.L.Markovsky, L.M.Soroko, T.A.Strizh.
JINR, E2-5656, Dubna, 1971.
6. Л.М.Сороко, Т.А.Стриж, сообщение ОИЯИ, Р10-6390, Дубна, 1972 г.
7. Сборник "Солнечная активность и жизнь", изд. "Зинятне", Рига, 1967.
8. H.L.Buijs. Fast Fourier Transformation of Large Arrays of Data.
Applied Optics, 8, No.1, 211, 1969.
9. Т.А.Стриж. Быстрое преобразование Фурье. Сообщения ОИЯИ:
Б1-11-7198 и Б1-11-7200, Дубна, 1972 г.