

Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра «Прикладная математика»

**КУРСОВАЯ РАБОТА
ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА»**

Выполнил
студент группы 3630102/70301

Камянский Д.В.

Проверил
к. ф.-м. н., доцент

Баженов А.Н.

Санкт-Петербург
2020

1 Постановка задачи

1.1 Подготовка данных

- Считать данные светимости;
- Выделить временные интервалы с различным типом циклического движения;
- Разделить матрицы состояния плазмы по соответствующим временным интервалам.

1.2 Расчеты

- Рассчитать положение центра масс системы для каждого момента времени;
- Рассчитать периоды прохождения центром масс циклической траектории и соответствующую частоту движения;
- Рассчитать значения характеризующие частоту движения центра масс.

1.3 Анализ

Рассмотреть характер вращения системы и зависимость частоты вращения от времени.

2 Теория

2.1 Центр масс

Центр масс - геометрическая точка, характеризующая движение тела или системы частиц как целого. Положение центра масс системы определяется следующим образом:

$$r_c = \frac{\sum m_i r_i}{\sum m_i} \quad (1)$$

где r_c - радиус-вектор центра масс, r_i - радиус-вектор i -й точки системы, m_i - значение светимости в i -й точке.

2.2 Выборочная медиана и выборочное среднее

Выборочная медиана - это значение, которое разбивает выборку на две равные части. Половина наблюдений лежит ниже медианы, и половина наблюдений лежит выше медианы. Медиана вычисляется следующим образом. Изучаемая выборка упорядочивается в порядке возрастания (N - объем выборки). Получаемая последовательность a_k , где $k = 1, \dots, N$ называется вариационным рядом или порядковыми статистиками. Если число наблюдений N нечетно, то медиана оценивается как $m = a_{\frac{N+1}{2}}$. Если число наблюдений N четно, то медиана оценивается как $m = (a_{\frac{N}{2}} + a_{\frac{N}{2}+1})$

Пусть X_1, \dots, X_n - выборка. Среднее значение выборки оценивается по формуле:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

2.3 Выборочный коэффициент корреляции Пирсона

Пусть по выборке значений $\{x_i, y_i\}_1^n$ двумерной с.в. (X, Y) требуется оценить коэффициент корреляции $\rho = \frac{\text{cov}(X, Y)}{\sqrt{D_X D_Y}}$. Естественной оценкой для ρ служит его статистический аналог в виде выборочного коэффициента корреляции, предложенного К.Пирсоном, —

$$r = \frac{\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2 \frac{1}{n} \sum (y_i - \bar{y})^2}} = \frac{K}{s_X s_Y}, \quad (2)$$

где K, s_X^2, s_Y^2 — выборочные ковариация и дисперсии с.в. X и Y [1, с. 535].

2.4 Автокорреляция уровней временного ряда

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют **автокорреляцией уровней ряда**. Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Формула для расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1) \cdot (y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}} \quad (3)$$

где

$$\bar{y}_1 = \frac{\sum_{t=2}^n y_t}{n-1}; \bar{y}_2 = \frac{\sum_{t=2}^n y_{t-1}}{n-1} \quad (4)$$

Эту величину называют коэффициентом автокорреляции уровней ряда первого порядка, так как он измеряет зависимость между соседними уровнями ряда t и y_{t-1} . Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ , то ряд содержит циклические колебания с периодичностью в τ моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

3 Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки JupyterLab. Используются библиотеки `scipy` для загрузки данных из файла, `pymru` для простоты использования различных статистических функций, `statsmodels` для построения коррелограммы, `matplotlib` и `seaborn` для визуализации результатов. Исходный код лабораторной работы приведён в приложении в виде ссылки на репозиторий GitHub.

4 Результаты и Анализ полученных данных

На видео движения светимости плазмы можно выделить 2 интервала: интервал вращения ($[157.0, 167.0]$), где чётко различимо вращение ядра (область с самыми высокими значениями светимости):

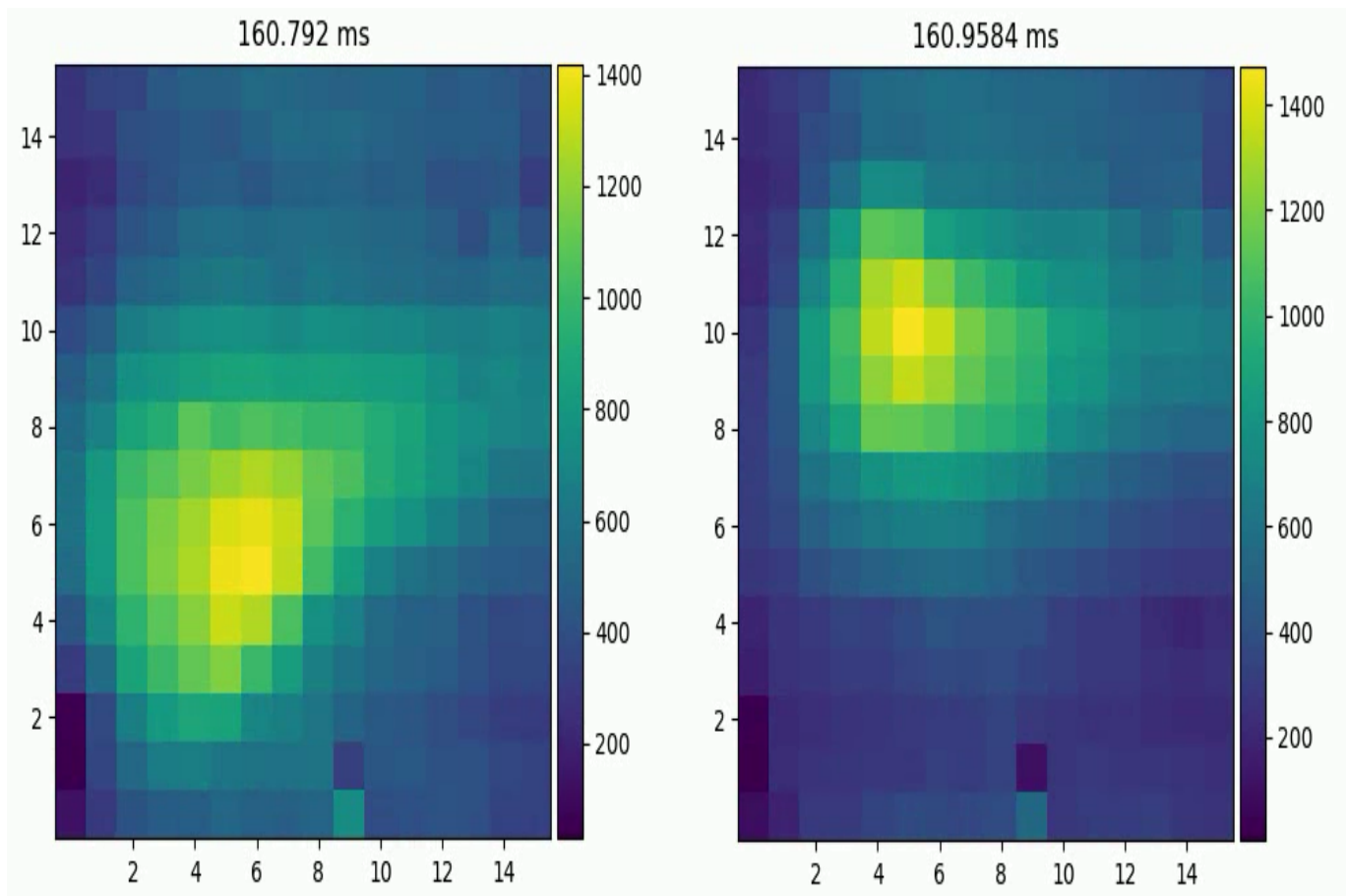


Рис. 1: Интервал вращения $[157.0, 167.0]$

Интервал колебания ($[167.0, 175.0]$), где ядро либо остаётся на месте, либо отклоняется очень незначительно от изначального положения и быстро к нему возвращается:

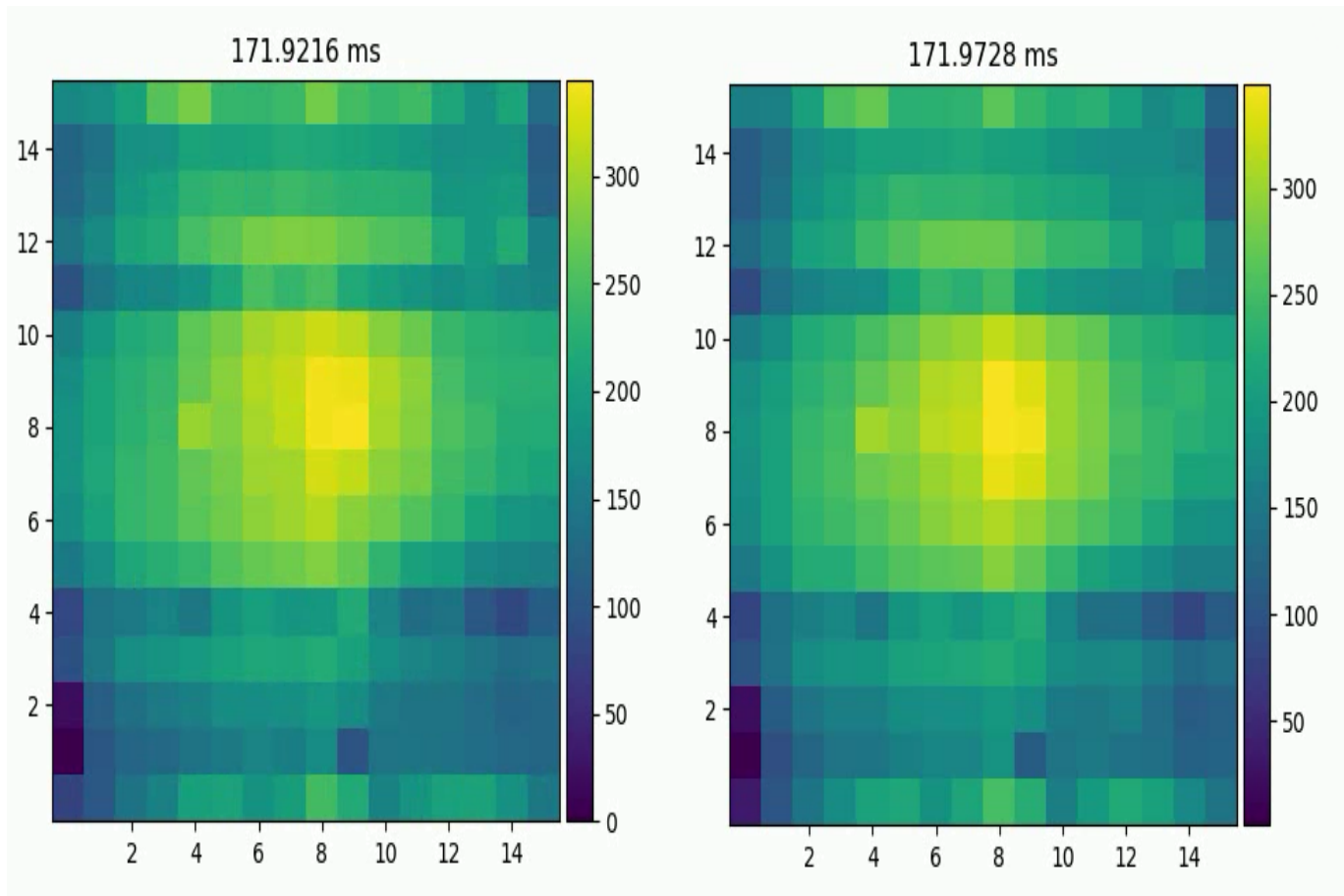


Рис. 2: Интервал колебания [167.0, 175.0]

Для исследования движения (в частности вращения) системы используется исследование движения центра масс. Частота рассчитывается по формуле $\nu = \frac{1}{T}$, где T - период обращения центра масс (время, за которое точка центра масс возвращается в стартовое положение начала периода).

4.1 Частота вращения системы по исходным данным

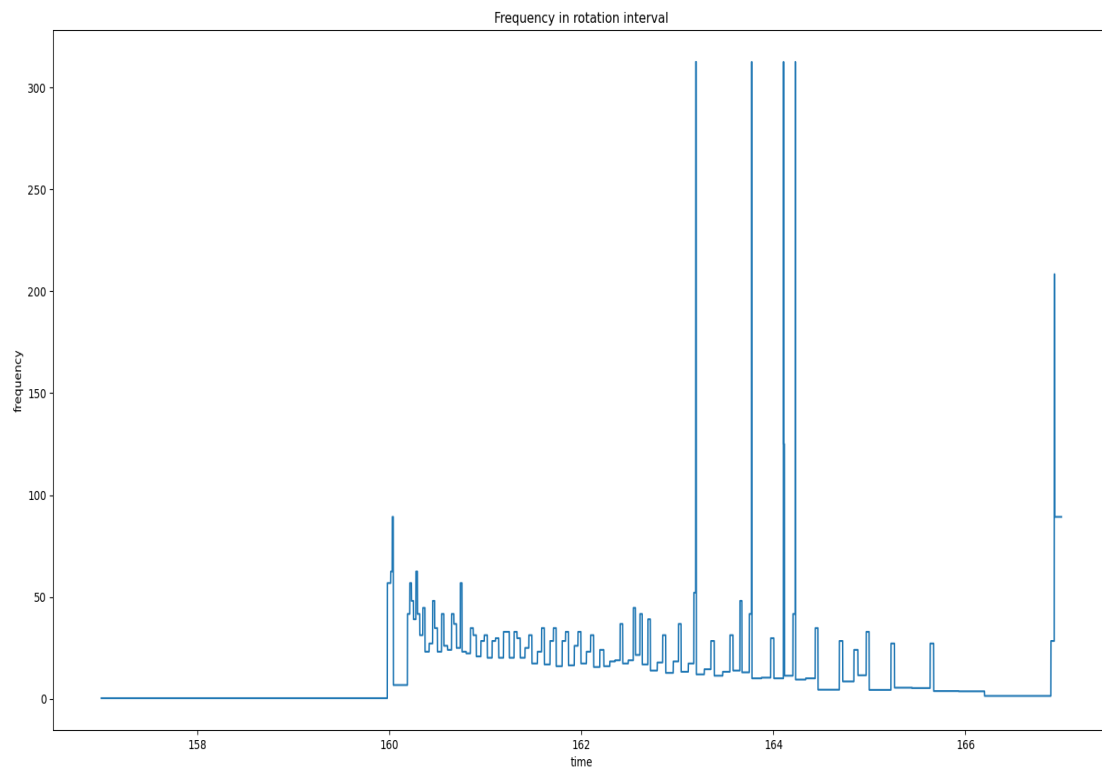


Рис. 3: Частота вращения системы на интервале вращения $[157.0, 167.0]$

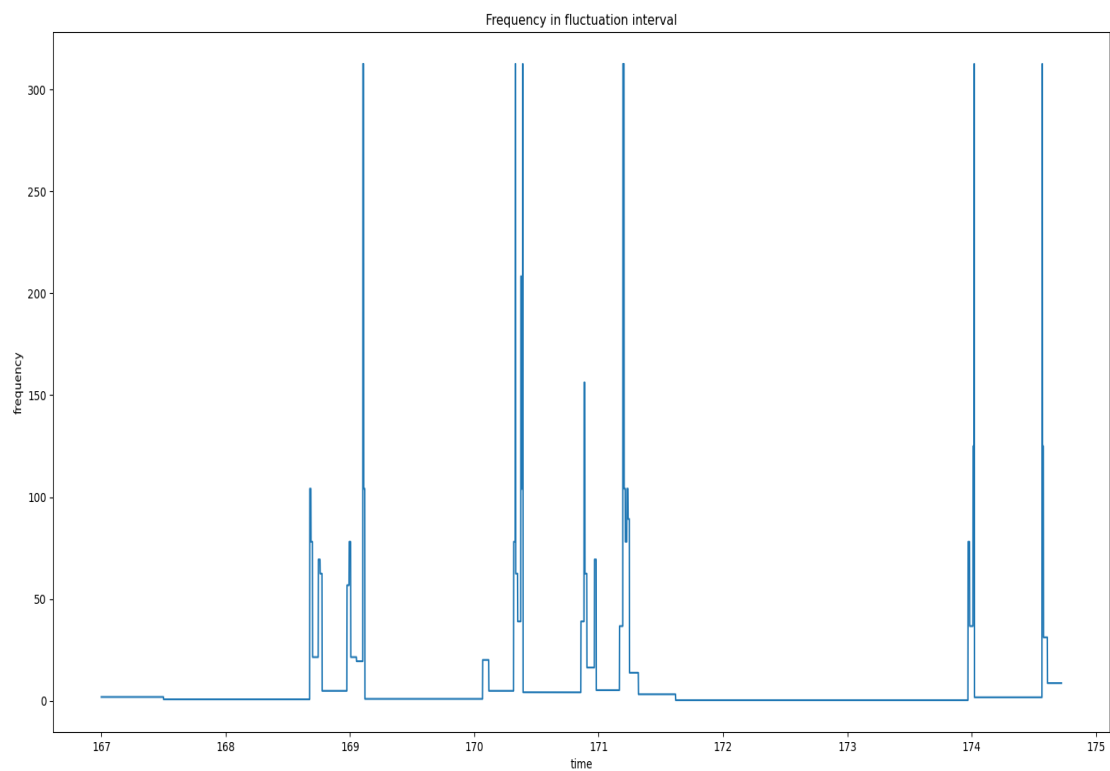


Рис. 4: Частота вращения системы на интервале колебаний [167.0, 175.0]

Таблица 1: Соотношение кол-ва вариаций позиции центра масс с кол-ом временных промежутков

	Вращение	Колебания
Кол-во различных позиций ц.м.	248	106
Кол-во временных промежутков	6250	5000

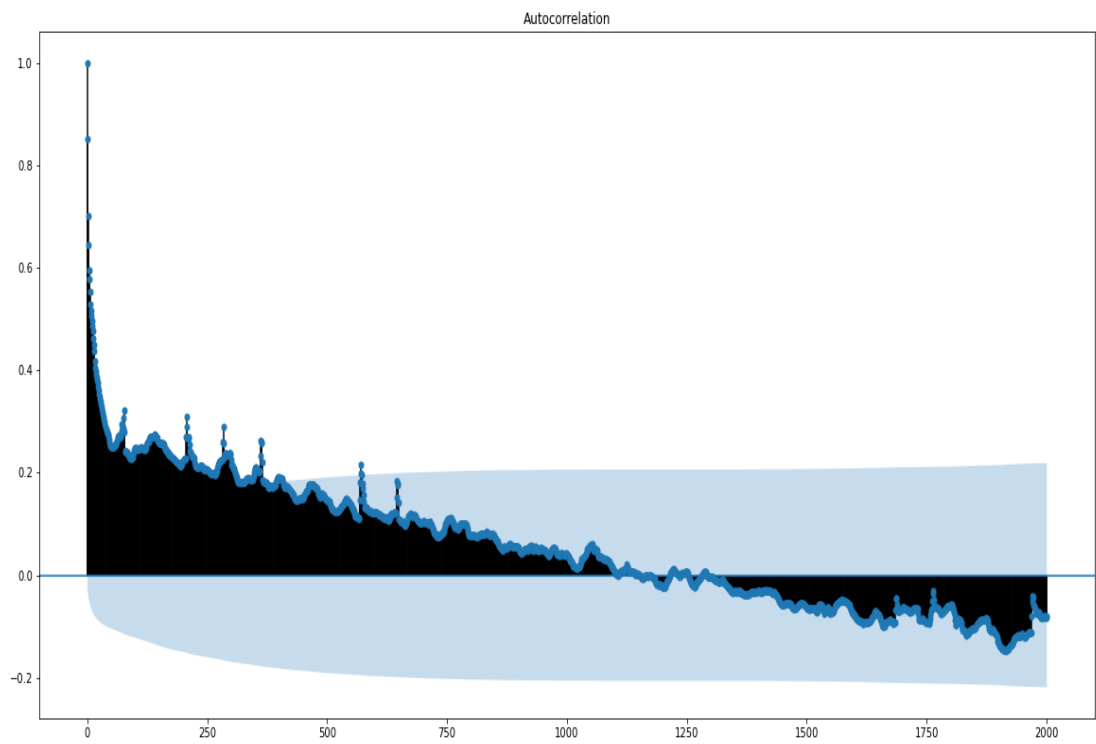


Рис. 5: Коррелограмма частоты вращения на интервале вращения $[157.0, 167.0]$

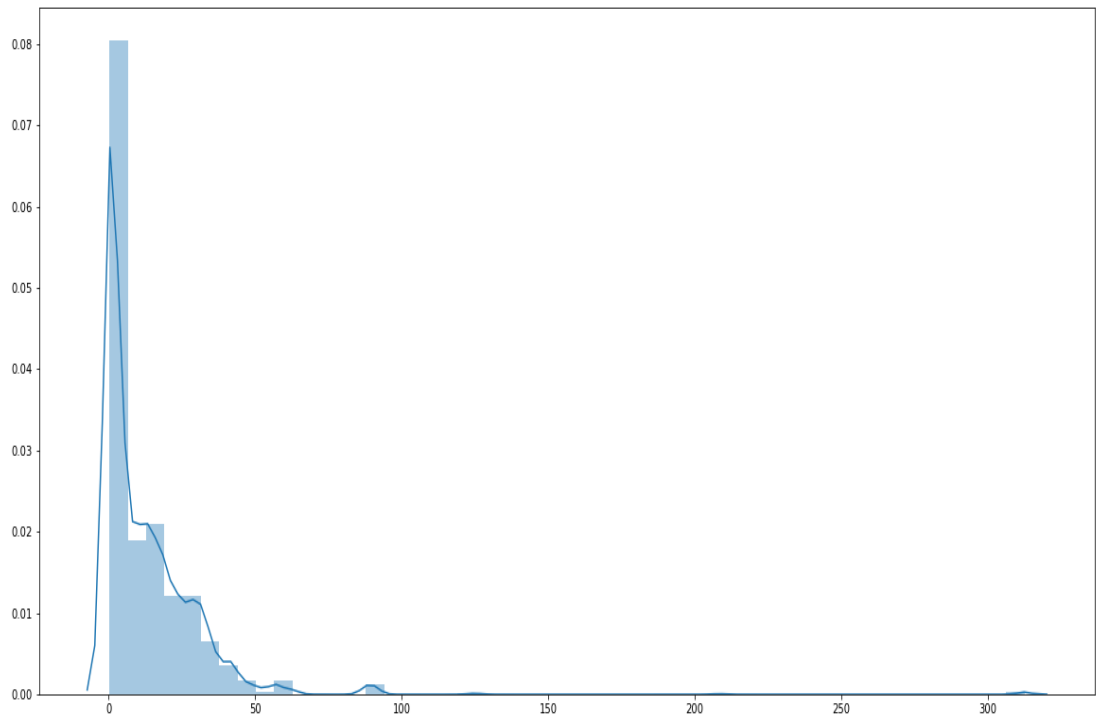


Рис. 6: Гистограмма частоты вращения на интервале вращения $[157.0, 167.0]$

Таблица 2: Выборочные характеристики частоты вращения

	Среднее	Медиана
Вращение	12.93	5.48
Колебания	6.86	1.06

Продолжительные пологие участки графиков - это временные промежутки на которых центр масс длительное время не меняет положения внутри цикла вращения (колебания), что приводит к сильному увеличению периода оборота. Данная проблема хорошо прослеживается при соотношении кол-ва различных позиций центра масс и количества рассматриваемых временных промежутков. Видно, что на большей части времени центр масс не меняет своей позиции.

Чтобы преодолеть проблему низкой изменчивости позиции центра масс, были рассмотрены два подхода: слабая фильтрация с последующим ослаблением ядра и сильная фильтрация.

4.2 Частота вращения системы после слабой фильтрации и ослабления ядра

Данный подход опирается на предположение, что причиной малой изменчивости позиции центра масс (далее ц.м.) является малая подвижность ядра системы и чтобы увеличить вес хвостовой части вращения требуется ослабить центральную ядерную часть. Чтобы при подобном ослаблении не увеличивать весовую роль шумовых значений, предварительно отфильтруем систему на значения меньше или равные 25-му перцентиллю.

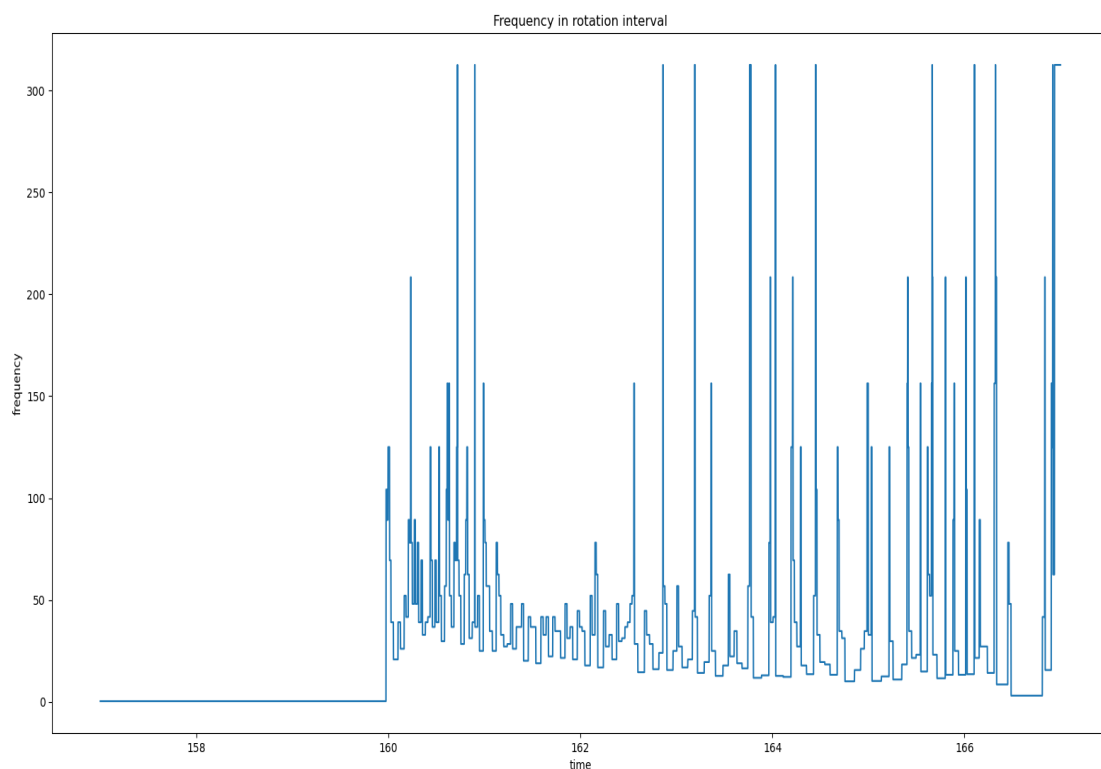


Рис. 7: Частота вращения системы на интервале вращения [157.0, 167.0]

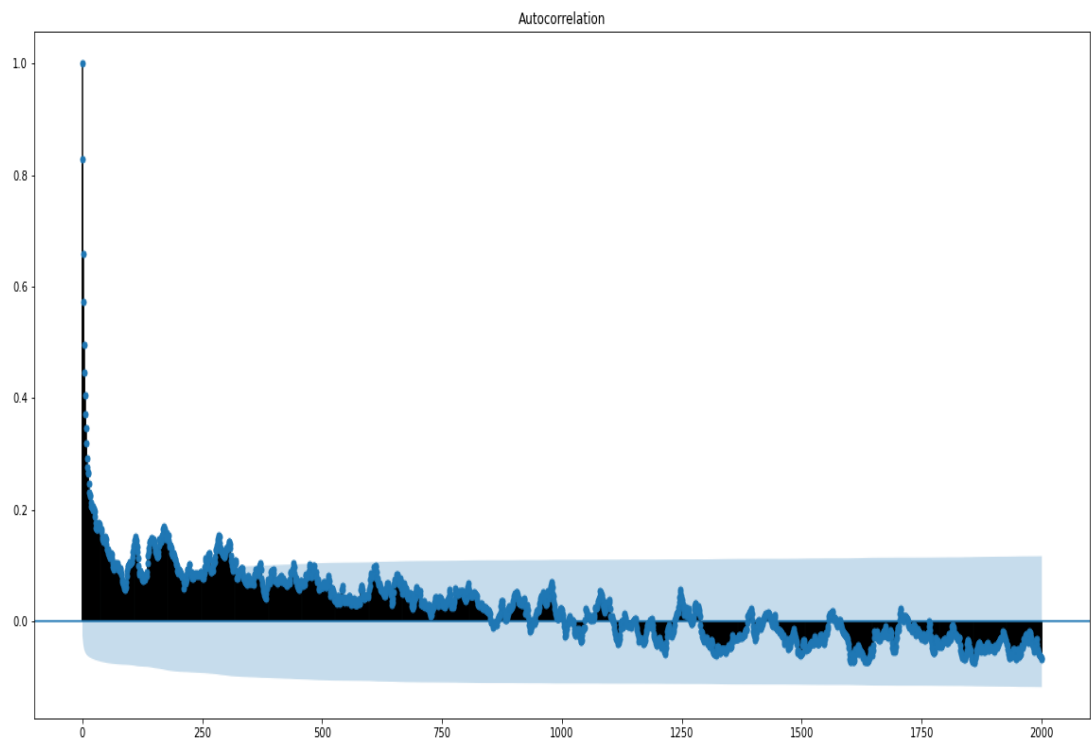


Рис. 8: Коррелограмма частоты вращения на интервале вращения $[157.0, 167.0]$

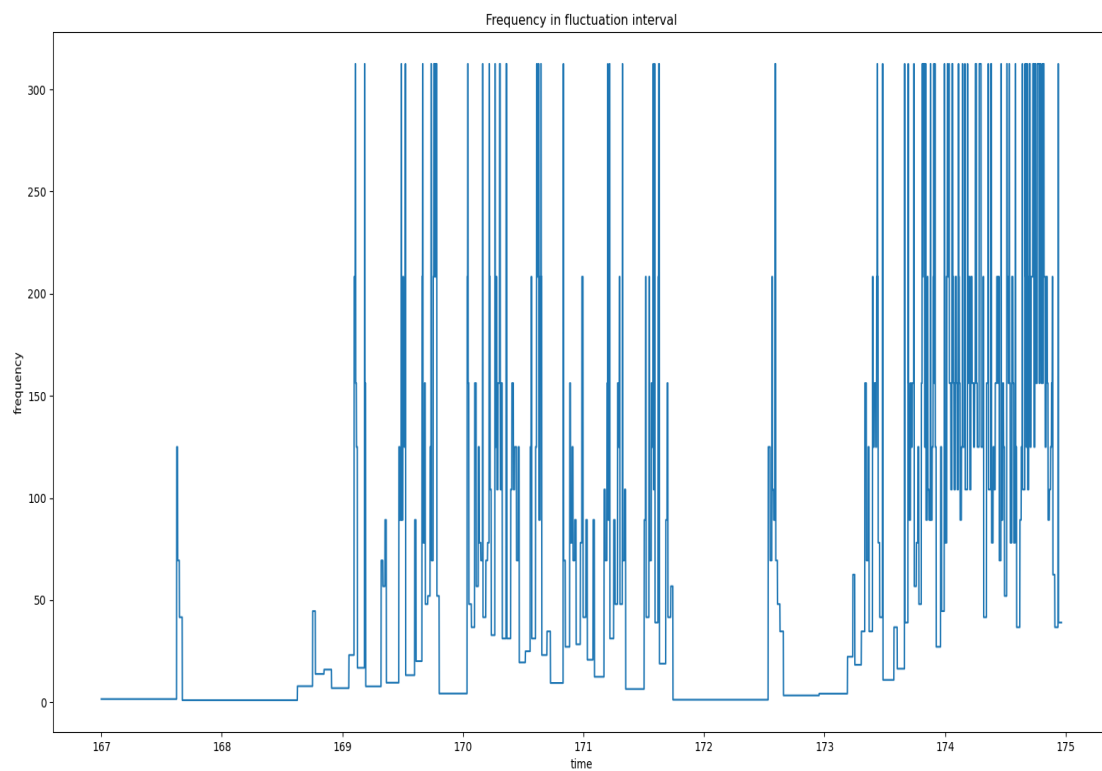


Рис. 9: Частота вращения системы на интервале колебаний $[167.0, 175.0]$

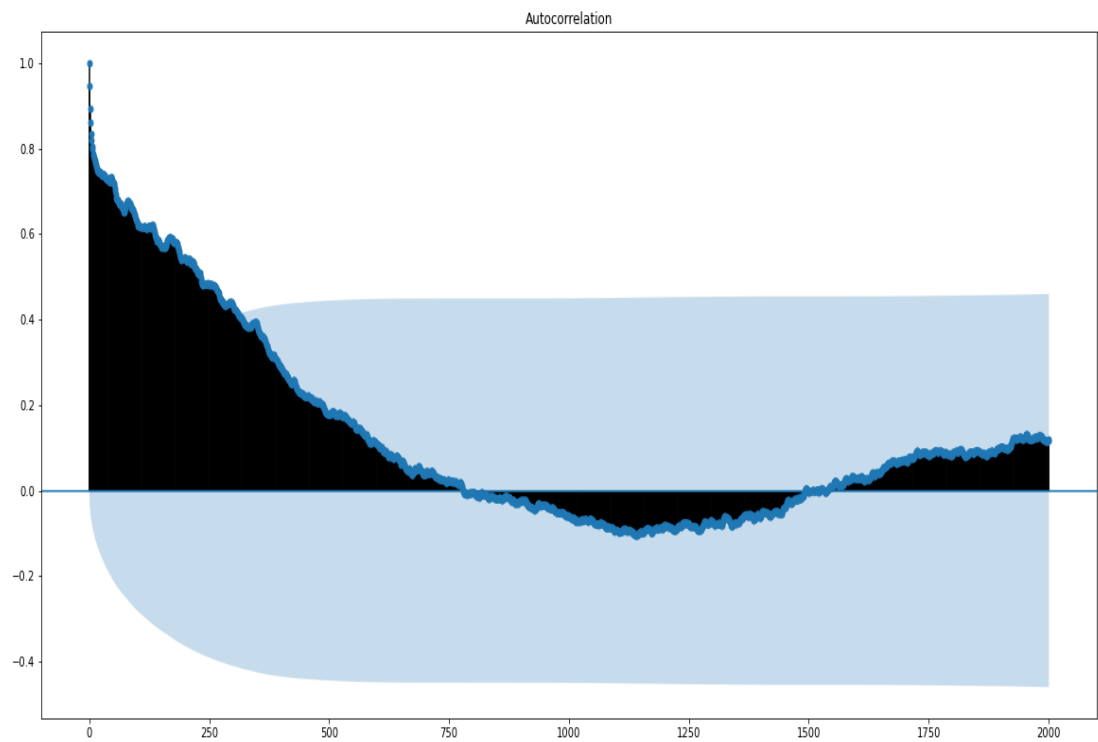


Рис. 10: Коррелограмма частоты вращения на интервале колебаний $[167.0, 175.0]$

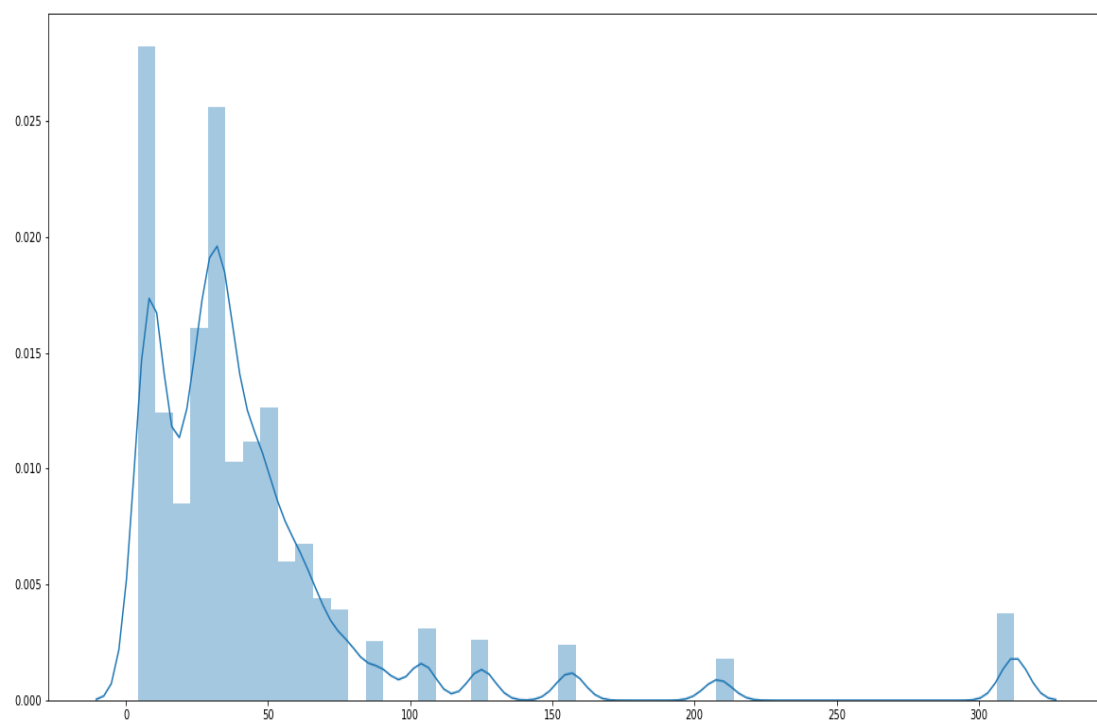


Рис. 11: Гистограмма частоты вращения на интервале вращения $[157.0, 167.0]$

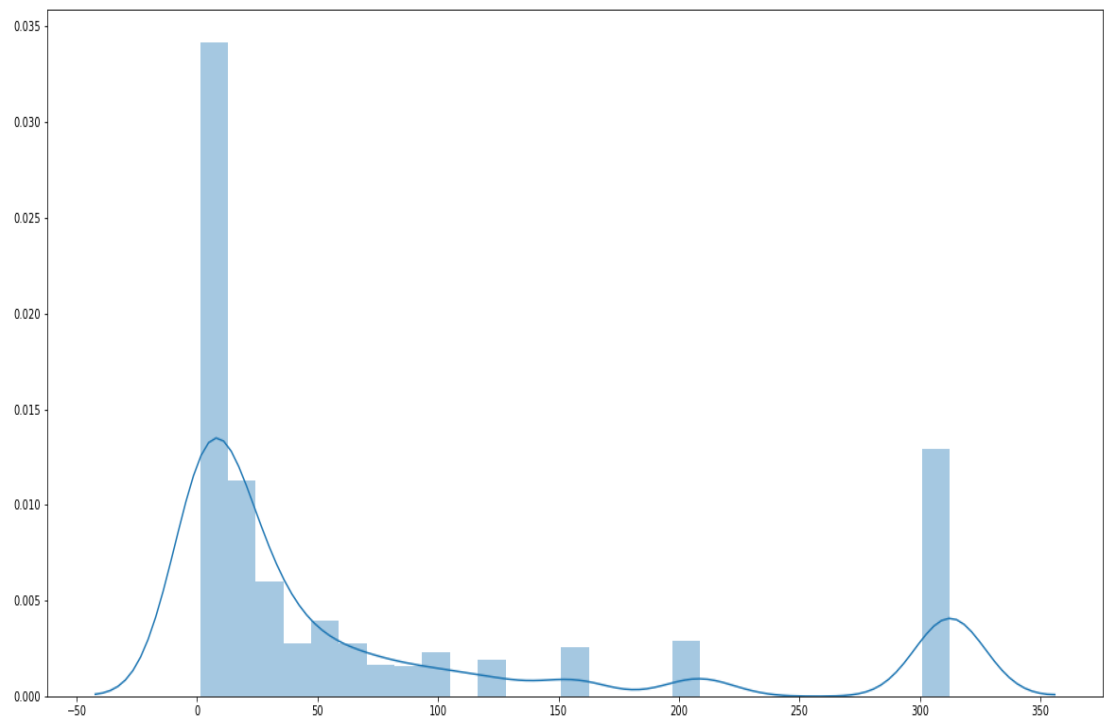


Рис. 12: Гистограмма частоты вращения на интервале вращения $[157.0, 167.0]$

Таблица 3: Выборочные характеристики частоты вращения

	Среднее	Медиана
Вращение	26.35	15.63
Колебания	48.83	10.97

Как видим, среднее и медианное значение частоты сильно возросло, особенно в случае колебаний, что лучше отражает наблюдаемую на видео динамику.

4.3 Частота вращения системы после сильной фильтрации

Второй подход исходит из предположения, что само ядро в достаточной степени подвижно и хорошо выражает общую динамику движения системы, а статичность возникает в результате влияния шумовых значений. Тогда, чтобы усилить влияние ядра производится фильтрация значений ниже 75-го перцентиля.

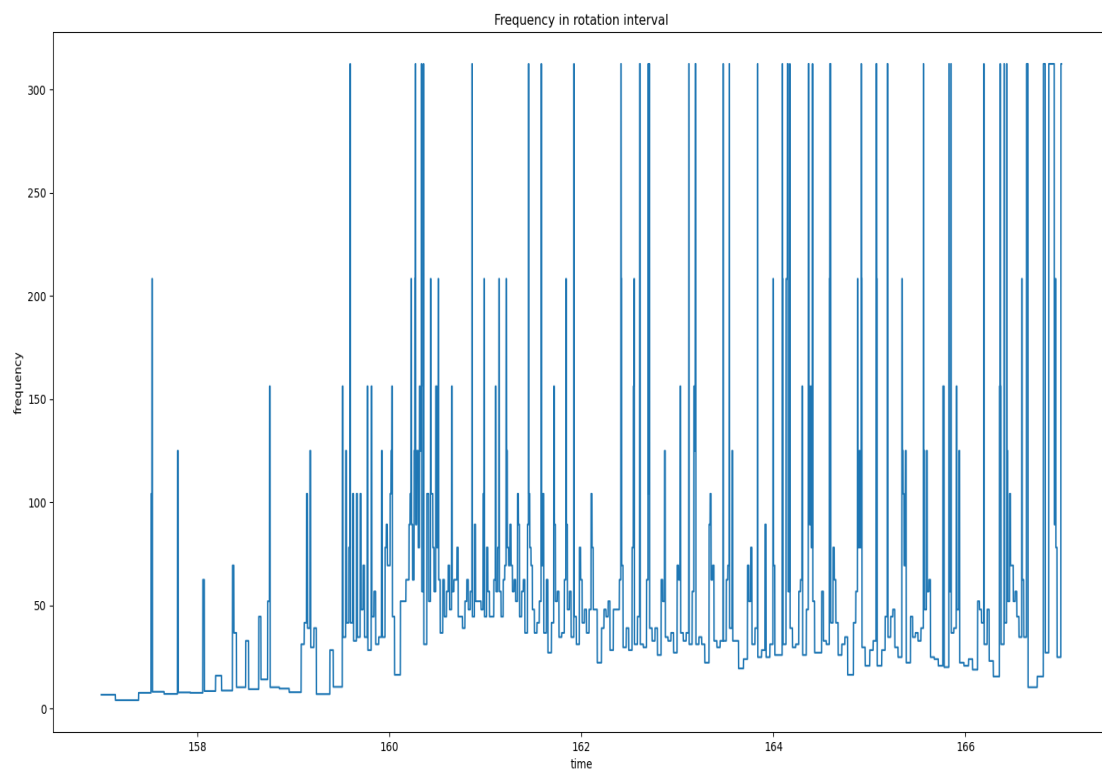


Рис. 13: Частота вращения системы на интервале вращения $[157.0, 167.0]$

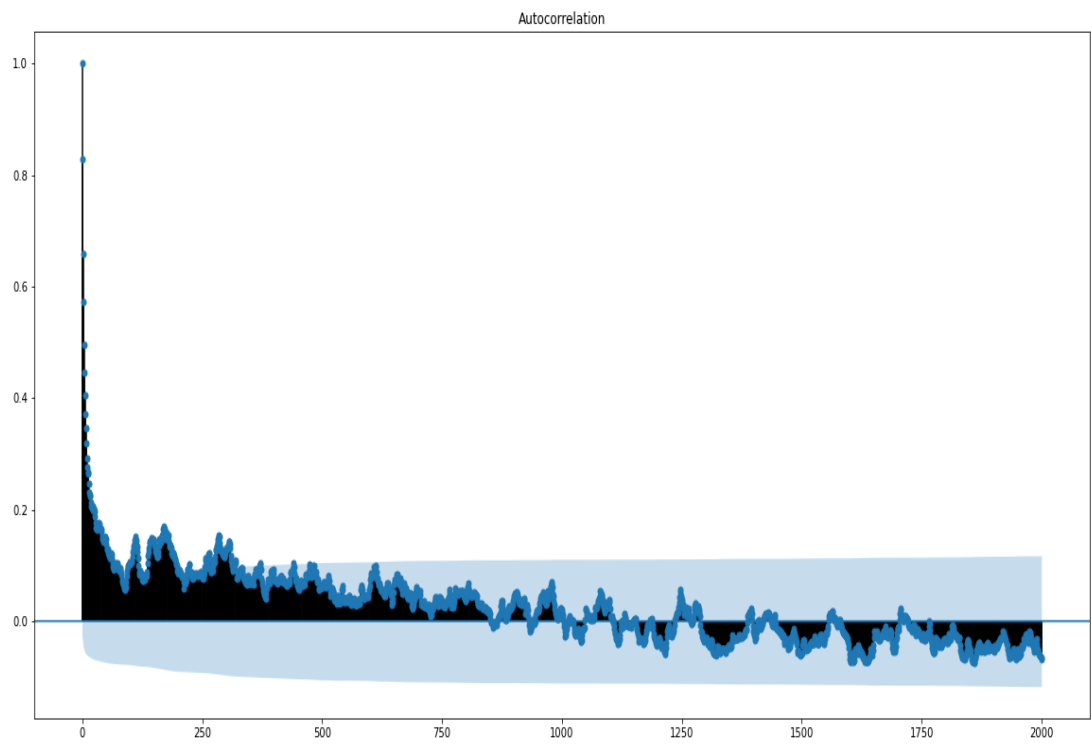


Рис. 14: Коррелограмма частоты вращения на интервале вращения $[157.0, 167.0]$

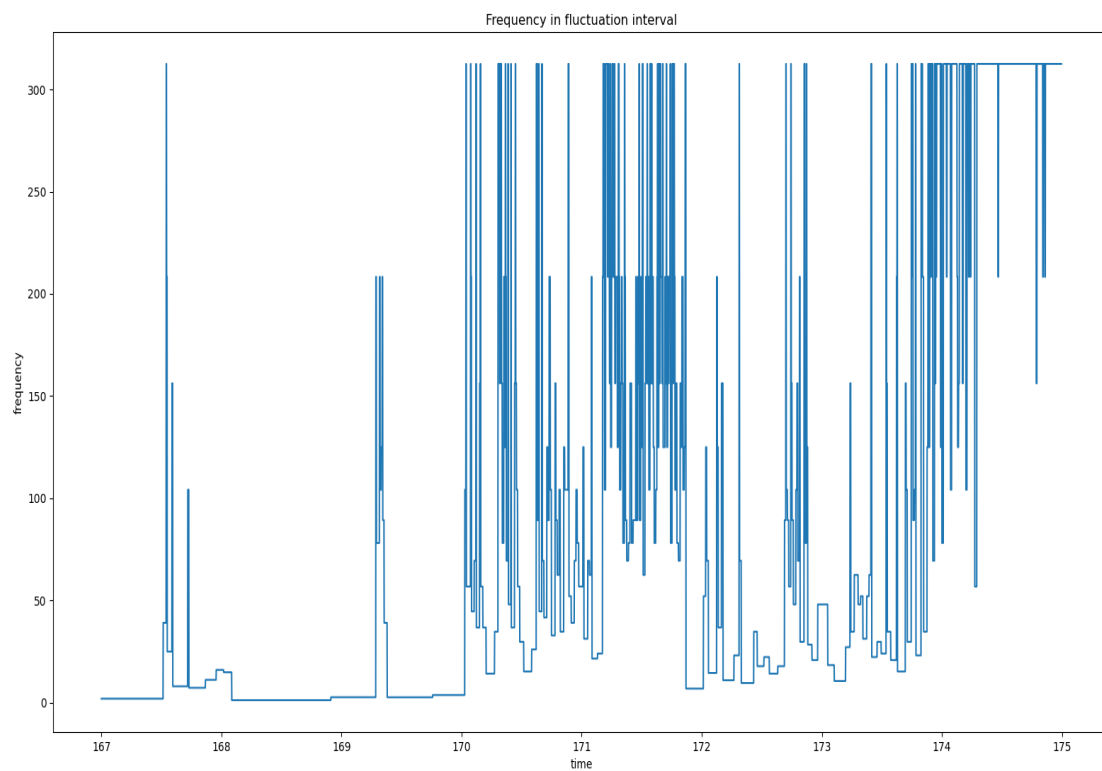


Рис. 15: Частота вращения системы на интервале колебаний [167.0, 175.0]

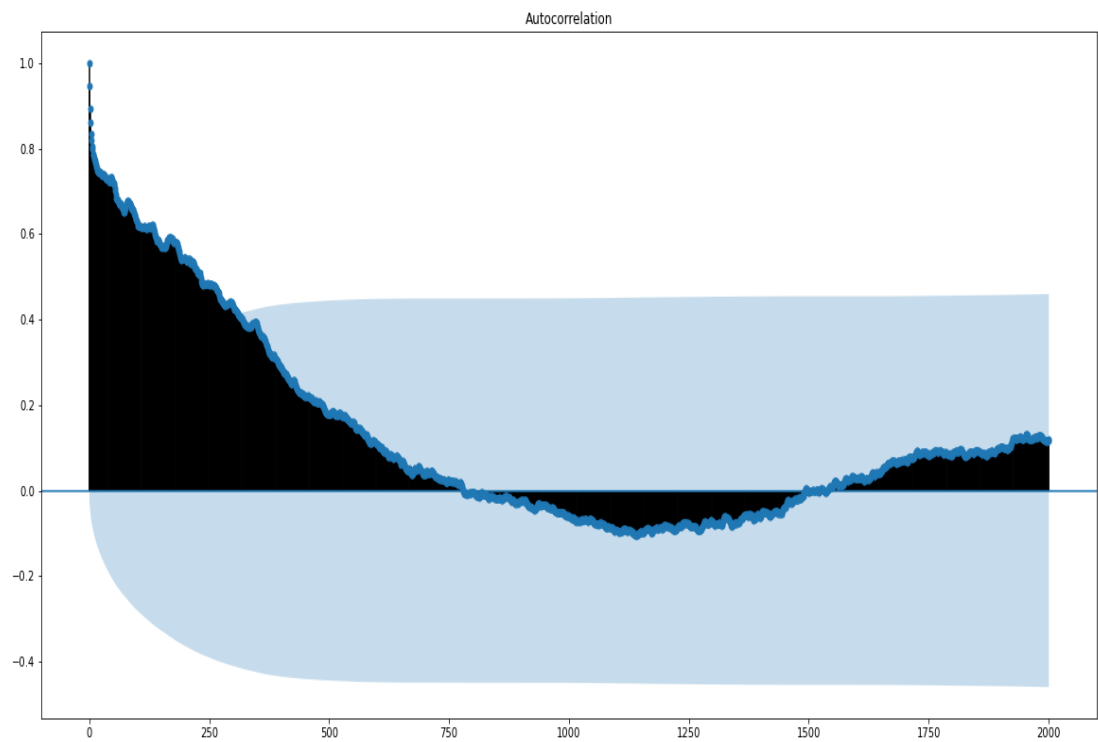


Рис. 16: Коррелограмма частоты вращения на интервале колебаний $[167.0, 175.0]$

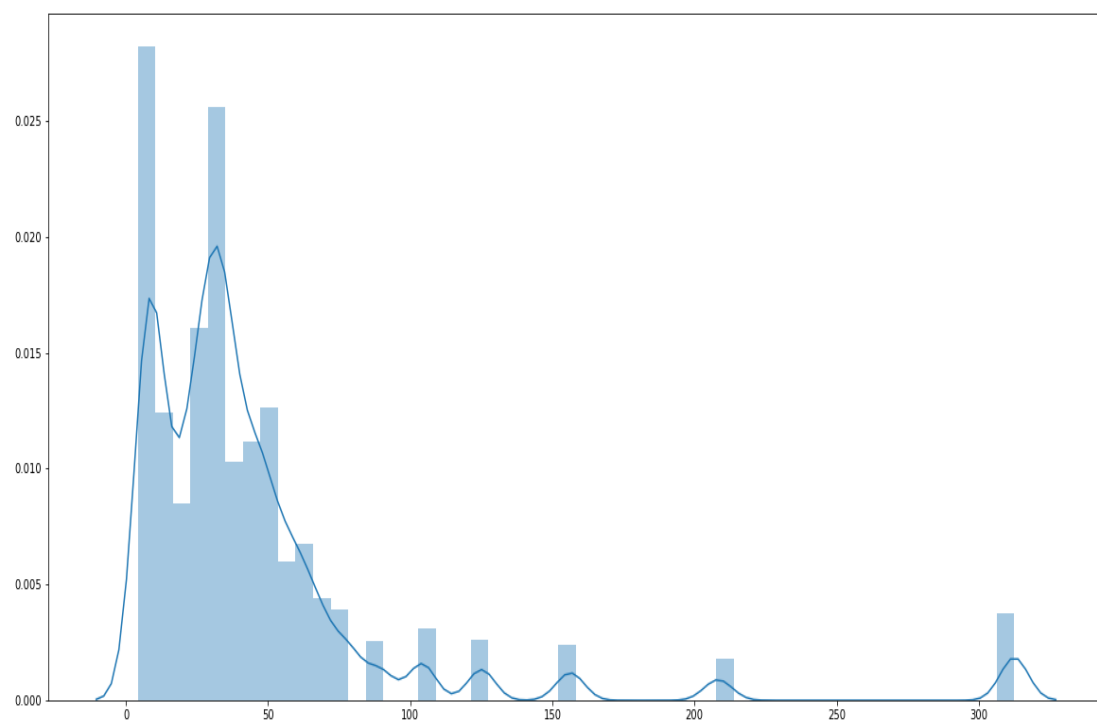


Рис. 17: Гистограмма частоты вращения на интервале вращения $[157.0, 167.0]$

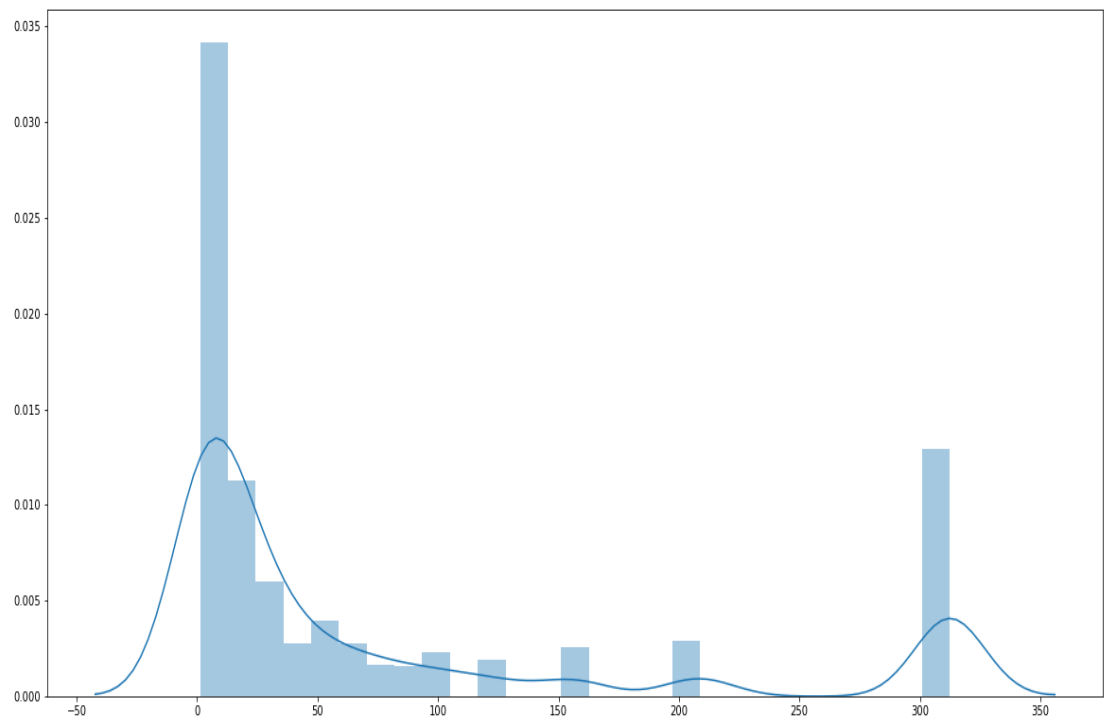


Рис. 18: Гистограмма частоты вращения на интервале вращения $[157.0, 167.0]$

Таблица 4: Выборочные характеристики частоты вращения

	Среднее	Медиана
Вращение	46.1	32.89
Колебания	78.66	22.32

После сильной фильтрации видно, что графики стали значительно лучше отражать наблюдаемую на видео динамику. С графика интервала вращения практически исчезли частотные плато, они присутствуют только в начале, когда система ещё не перешла в фазу активного вращения. На графике колебаний плато попрежнему присутствуют, при чем основная их часть приходится на промежуток времени после окончания интервала вращения и начала интервала колебаний, когда на видео мы наблюдаем стабилизацию системы с последующим возникновением слабых колебательных движений.

Также заметно и улучшение выборочных характеристик: на интервале вращения относительная разность между средним и медианным значениями уменьшилась, а в абсолютном значении характеристики выросли, на интервале колебаний высокая разница между медианой и средним сохраняется, но она объективно обусловлена исходной динамикой.

Из полученных коррелограмм можно сделать выводы: на интервале вращения линейной динамики частоты не наблюдается, так как на большинстве рассмотренных лагов значение коэффициент автокорреляции не превышает значение 0.2, если динамика и есть, то она имеет нелинейный характер, что требует дополнительных исследований; на интервале колебаний коэффициент принимает значимые значения (> 0.5) на довольно широком диапазоне начальных лагов, из чего можно сделать вывод, что на коротких временных промежутках наблюдается цикличность частоты.

5 Обсуждение

Исходя из полученных результатов можно сделать вывод о том, что подход, основанный на сильной фильтрации является наиболее предпочтительным при исследовании движения светимости.

Динамика вращения системы на интервалах вращения и колебания сильно отличается: на интервале вращения [157,167] мы наблюдаем достаточно равномерные циклические колебания частоты со средним значением 46.1 1/мс, после чего система переходит [167,170] в устойчивое состояние, в котором центр масс практически не меняет своего положения, из которого система переходит в состояние быстрых коротких колебаний [170,175] с возрастающей динамикой частоты со средним значением частоты 78.66 1/мс.

6 Приложения

Код программы и полученные видео на GitHub, URL: <https://github.com/dkamianskii/MatStatLabs/tree/master/Course%20Project>

Полученные видео:

1. Изменение проекции светимости по времени - `projections_in_area`
2. Проекция светимости после удаления элементов ядра - `projections_dropped_rot`, `projections_dropped_fluct`
3. Проекция светимости после сильной фильтрации - `projections_filtered`
4. Изменение центра масс системы по времени - `centers_mass`

Список литературы

- [1] Вероятностные разделы математики. Учебник для бакалавров технических направлений. //Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.
- [2] Баженов А.Н., Затылкин П.А. Малоракурсная реконструкция светимости плазмы для сферического токамака. Вычислительные технологии. 2020; 25(1):5–38.