Санкт-Петербургский политехнический университет Петра Великого
Физико-Механический институт
«Высшая школа прикладной математики и вычислительной физики»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №9

по дисциплине «Математическая статистика»

Выполнил студент:
Ярмак Дмитрий Юрьевич
группа: 5030102/90101

Проверил:

к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург 2022 г.

Содержание

Список иллюстраций

1 Постановка задачи

Исследование из области солнечной энергетики [1]. На рис 1 показана схема установки для исследования фотоэлектрических характеристик.

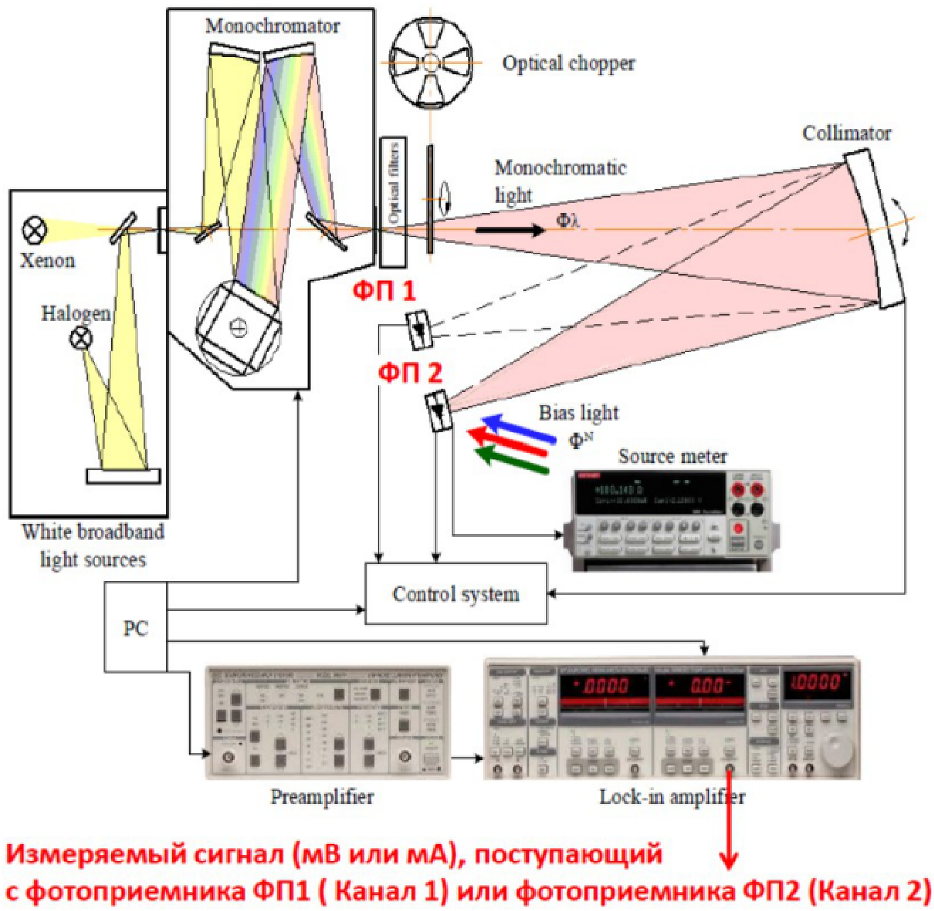

Рис. 1: Схема установки для исследования фотоэлектрических характеристик

Калибровка датчика ФП1 производится по эталону ФП2. Зависимость между квантовыми эффективностями датчиков предполагается одинаковой для каждой пары измерений.

$$QE_{\text{ФП}2} = \frac{I_{\text{ФП}2}}{I_{\text{ФП}1}} * QE_{\text{ФП}1} \tag{1}$$

QE - квантовые эффективности эталонного и исследуемого датчиков, I - измеренные токи.

Исходные данные Имеется 2 выборки данных с интервальной неопределенностью. Одна из них относится к эталонному датчику ФП2, другая - к исследуемому датчику ФП1.
Задача. Треубется определить коэффициент калибровки

$$R_{21} = \frac{I_2}{I_1} \tag{2}$$

при помощи линейной регрессии на множестве интервальных данных и коэффициента Жаккара.

2 Теория

2.1 Анализ данных с интервальной неопределенностью

В первую очередь преставим данные таким образом, чтобы применить понятия статистики данных с интервальной неопределенностью. Один из распространённых способов получения интервальных результатов в первичных измерениях - это "обинтерваливание"точечных значений, когда к точечному базовому зачению x_0, которое считывается по показаниям измерительного прибора, прибавляется *интервал погрешности* ϵ:

$$\mathbf{x} = \dot{x} + \epsilon \tag{3}$$

Интервал погрешности зададим как

$$\epsilon = [-\epsilon; \epsilon]$$

В конкретных измерениях примем $\epsilon = 10^{-4}$ мВ. Согласно терминологии интервального анализа, рассматриваемая выборка - это вектор интервалов, или интервальный вектор $x = (x_1, x_2, ..., x_n)$

2.2 Линейная регрессия

2.2.1 Описание модели

Линейная регрессия - регрессионная модель зависимости одной переменной от другой с линейной функцией зависимости:

$$y_i = X_i b_i + \epsilon_i$$

где X - заданные значения, y - параметры отклика, ϵ - случайная ошибка модели.

В случае, если у нас y_i зависит от одного параметра x_i, то модель выглядит следующим образом:

$$y_i = b_0 + b_1 * x_i + \epsilon_i \tag{4}$$

В данной модели мы пренебрегаем прогрешностью и считаем, что она получается при измерении y_i.

2.2.2 Метод наименьших модулей

Для наиболее точного приближения входных с фотоприемников данных y_i линейной регрессией $f(x_i)$ используется метод наименьших модулей. Этот метод основывается на минимизации нормы разности последовательности:

$$||f(x_i) - y_i||_{l^1} \to min \tag{5}$$

В данном случае ставится задача линейного программирования, решение которой дает нам коэффициенты b_0 и b_1, а также вектор множителей коррекции данных w. По итогу получается следующая задача линейного программирования

$$\sum_{i=1}^{n} |w_i| \to min \tag{6}$$

$$b_0 + b_1 * x_i - w_i * \epsilon \le y_i, \forall i = 1..n \tag{7}$$

$$b_0 + b_1 * x_i + w_i * \epsilon \le y_i, \forall i = 1..n \tag{8}$$

$$1 \le w_i, \forall i = 1..n \tag{9}$$

2.3 Предварительная обработка данных

Для оценки постоянной, как можно будет увидет далее, необходима предварительная обработка данных. Займемся линейной моделью дрейфа.

$$Lin(n) = A + B * n, \forall n = 1..N \tag{10}$$

Поставив и решив задачу линейного программирования, найдем коэффициенты A, B и вектор w множителей коррекции данных для каждого из фотоприемников ФП1 и ФП2. В последствии множитель коррекции данных необходимо применить к погрешностям выборки, чтобы получить данные, которые согласовывались с линейной моделью дрейфа:

$$I^f(n) = x(n) + \epsilon * w(n), \forall n = 1..N \tag{11}$$

В итоге необходимо построить "спрямленные"данные выборки: получить их можно путем вычитания из исходных данных линейной компоненты:

$$I^c(n) = I^f(n) - B * n, \forall n = 1..N \tag{12}$$

2.4 Коэффициент Жаккара

Коэффициент Жаккара - мера сходства множеств. В интервальных данных рассматривается некоторая модификация этого коэффициента: в качестве меры множества (в данном случае интервала) рассматривается его длина, а в качестве пересечения и объединения - взятие минимума и максимума по включению двух величин в интервальной арифметике Каухера соответственно. Можно заметить, что в силу возможности минимума по включению быть неправильным интервалом, коэффициент Жаккара может достигать значения только в интервале [-1; 1].

$$JK(x) = \frac{wid(\wedge x_i)}{wid(\vee x_i)} \tag{13}$$

2.5 Процедура оптимизации

Чтоб найти оптимальный параметр калибровки R_{21} необходимо поставить и решить задачу максимизации коэффициента Жаккара, зависящего от параметра калибровки:

$$JK(I_1^c(n) * R \cup I_2^c(n)) \to max \tag{14}$$

где I_1^c и I_2^c - полученные спрямленные выборки, а R - параметр калибровки. Найденный таким образом R и будет искомым оптимальным R_{21} в силу наибольшего совпадения, оцененного коэффициентом Жаккара.

3 Реализация

Работа была реализована на языке Phyton в среде разработки PyCharm с использованием дополнительных библиотек numpy, scipy, matplotlib.

4 Результаты

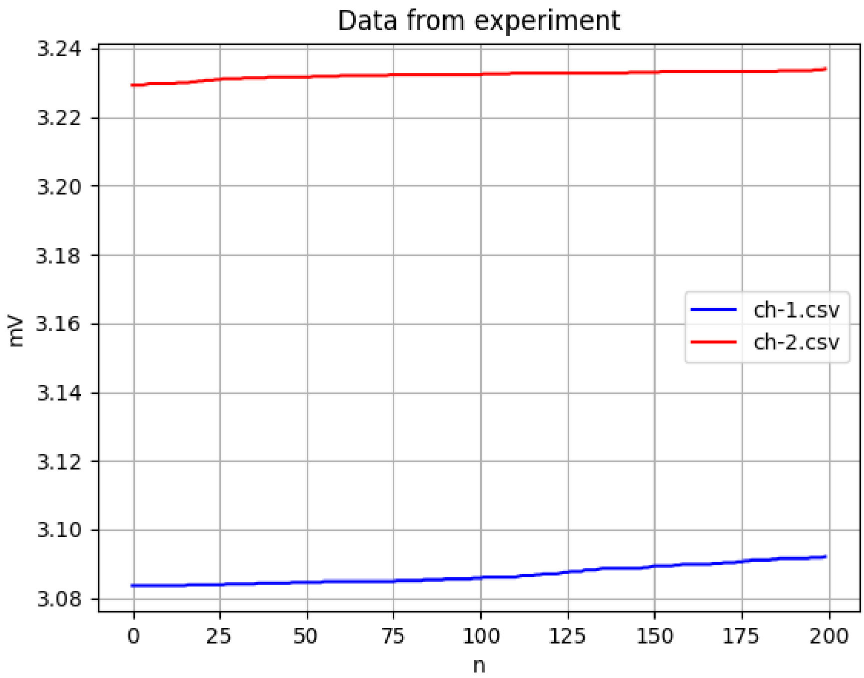

Рис. 2: Исходные данные из эксперимента

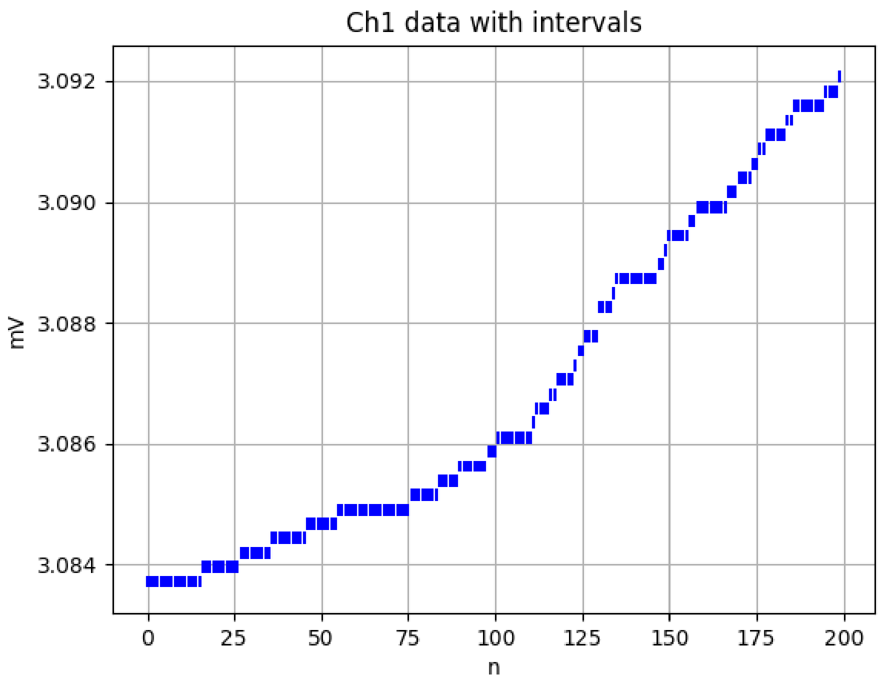

Рис. 3: Интервальное представление исходных данных для ФП1

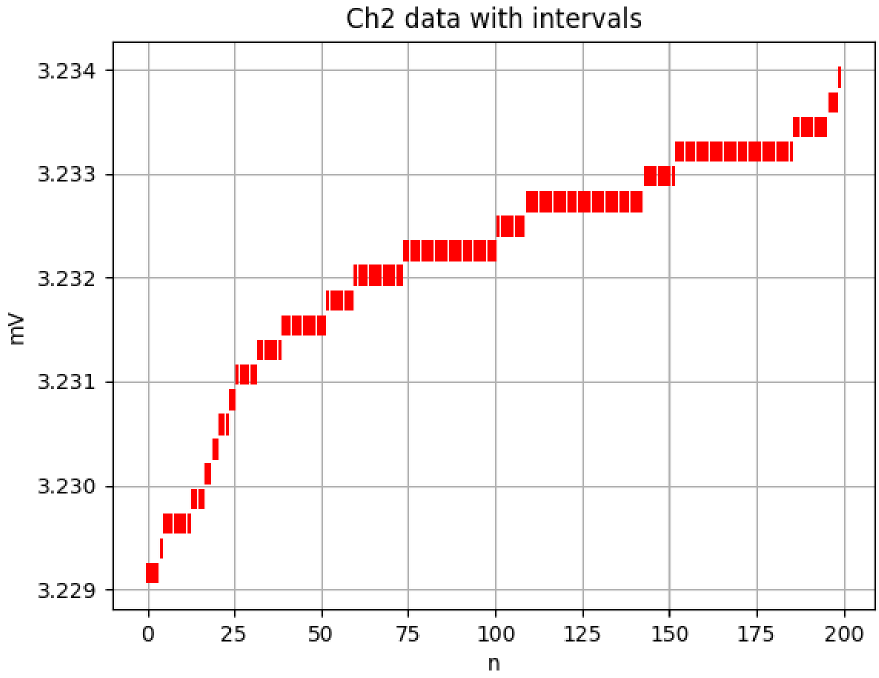

Рис. 4: Интервальное представление исходных данных для ФП2

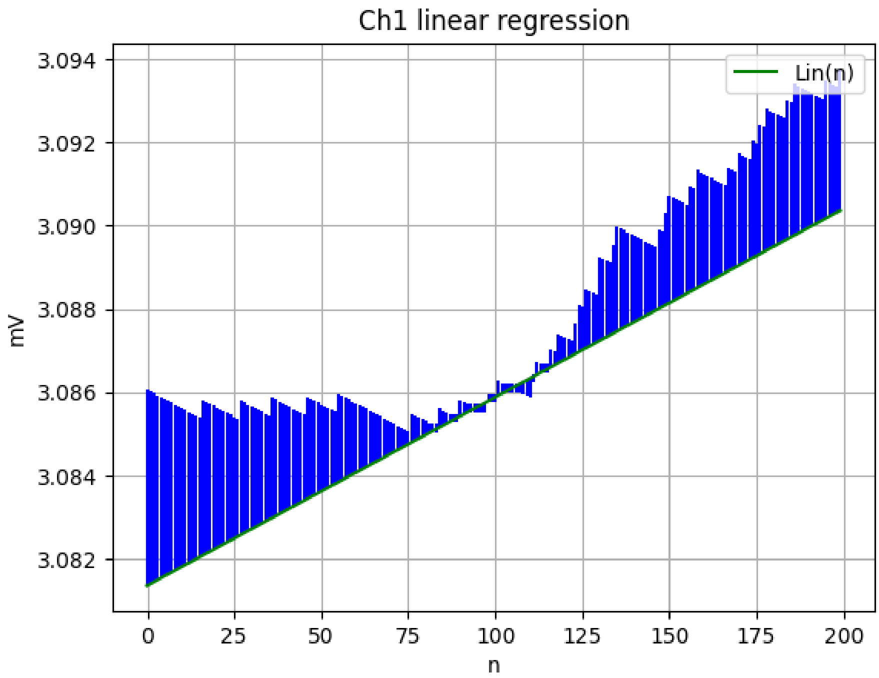

Рис. 5: Линейная модель данных для ФП1

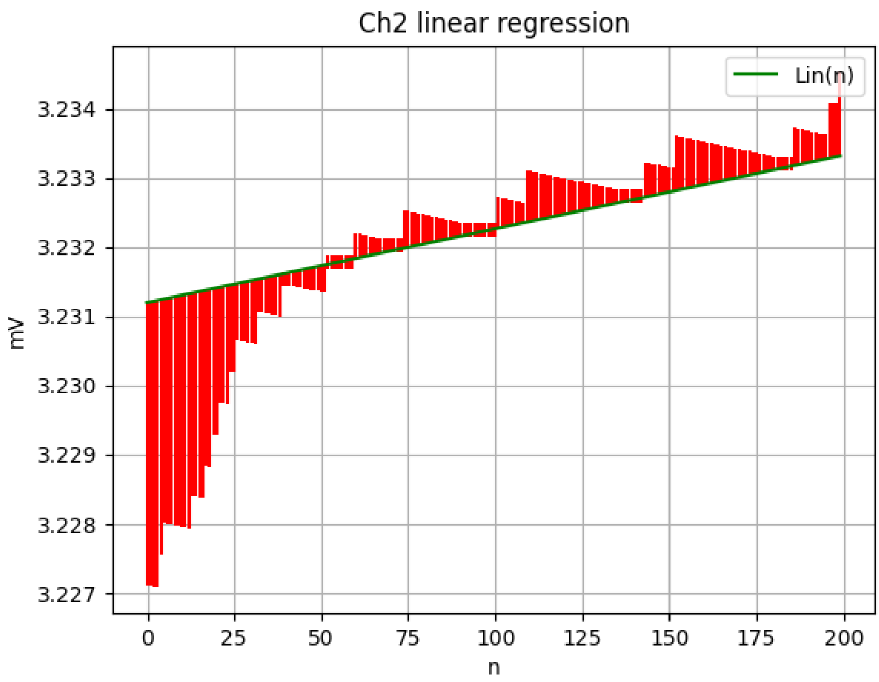

Рис. 6: Линейная модель данных для ФП2

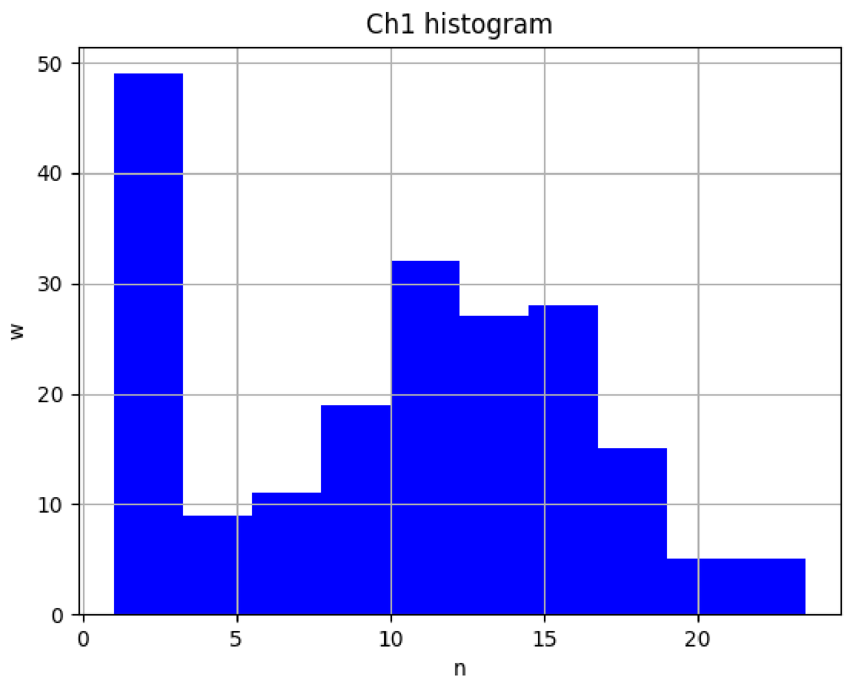

Рис. 7: Гистограмма значений множителей коррекции w для ФП1

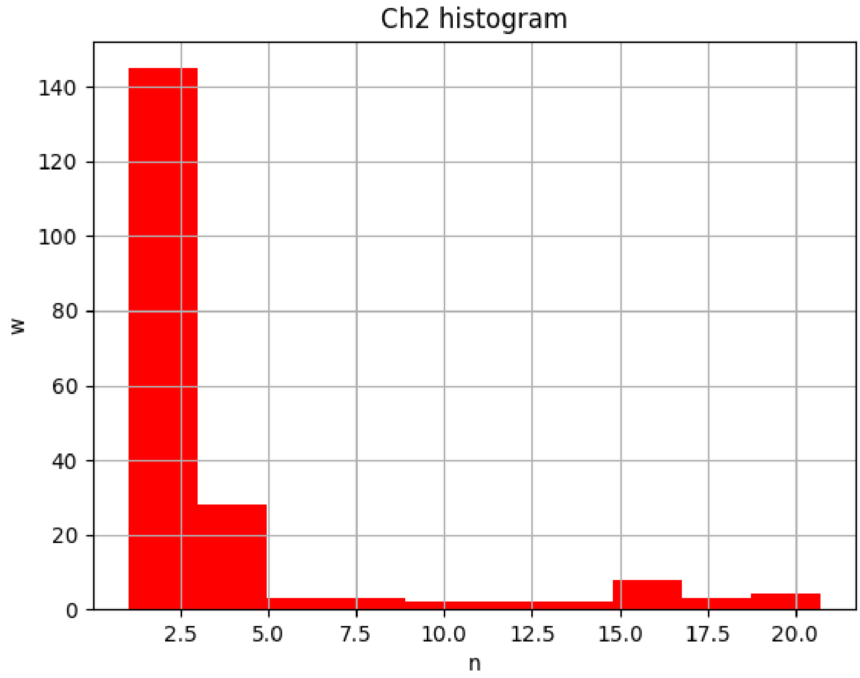

Рис. 8: Гистограмма значений множителей коррекции w для ФП2

Результаты линейного приближения токов:

- Для первого фотоприемника:

$$A_1 = 3.081369, B_1 = 4.520557 * 10^{-5}$$

- Для второго фотоприемника:

$$A_2 = 3.231197, B_2 = 1.064261 * 10^{-5}$$

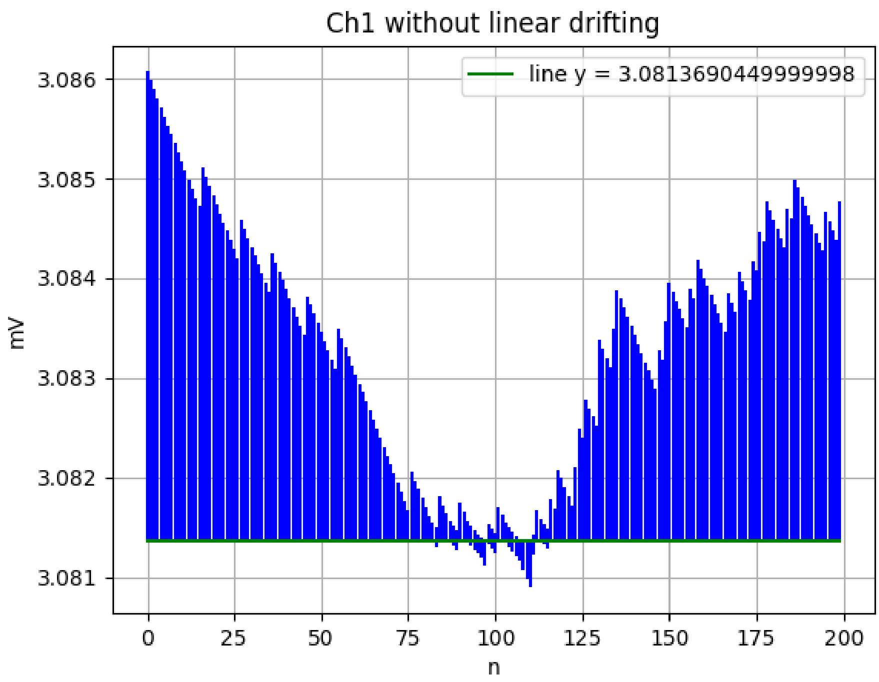

Рис. 9: Скорректированные модели данных для ФП1

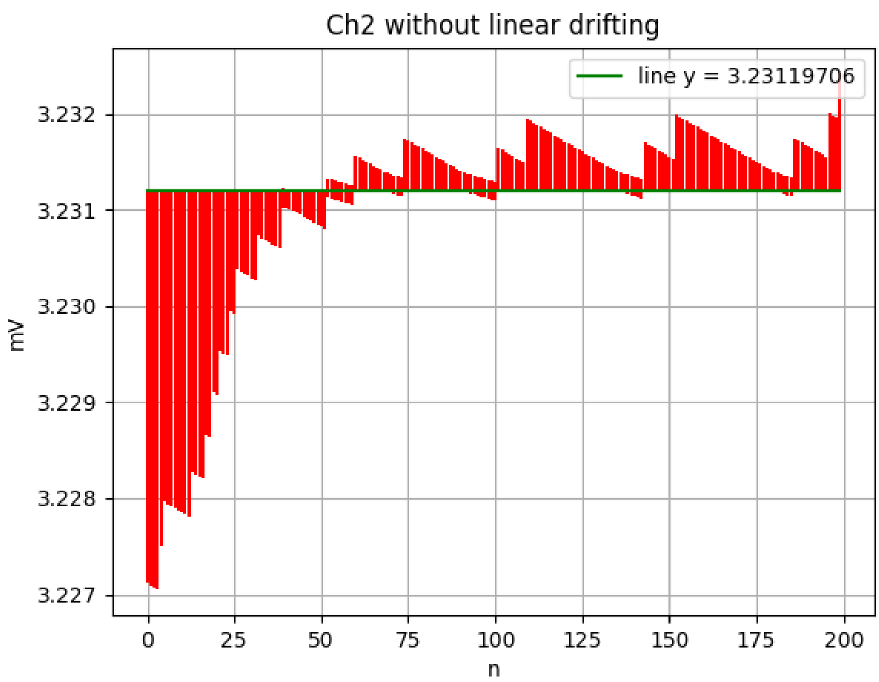

Рис. 10: Скорректированные модели данных для ФП2

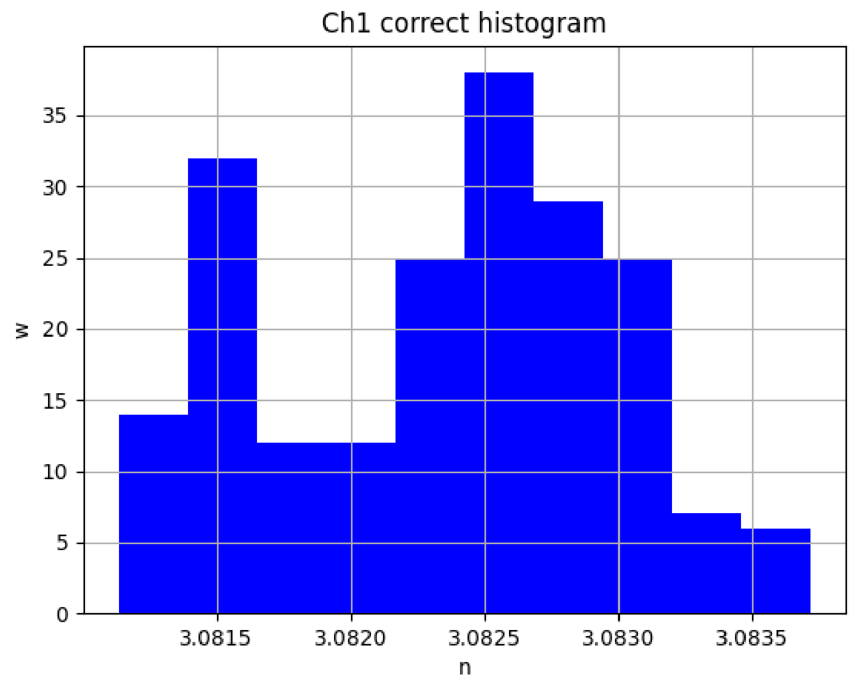

Рис. 11: Гистограмма скорректированных данных для ФП1

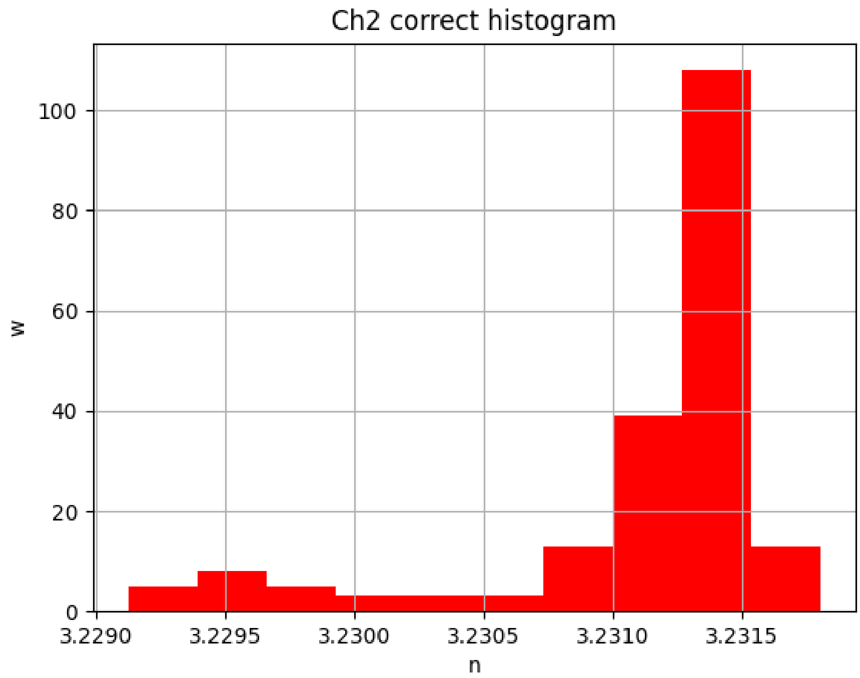

Рис. 12: Гистограмма скорректированных данных для ФП2

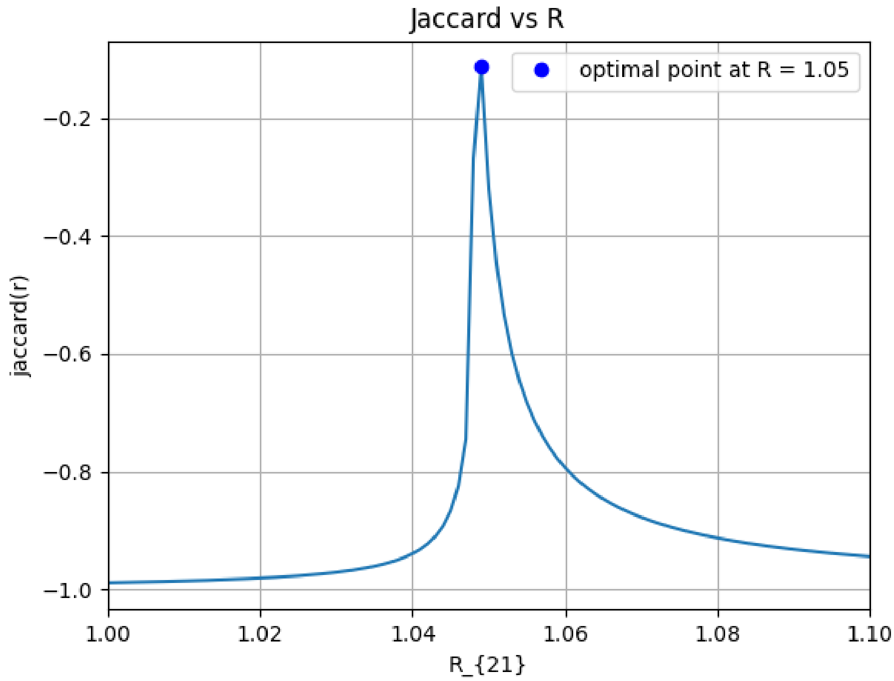

Рис. 13: Значение коэффициента Жаккара от калибровочного множителя R_{21}

Результаты исследований:

$$R_{opt} = 1.049, jaccard(R_{opt}) = -0.113196$$

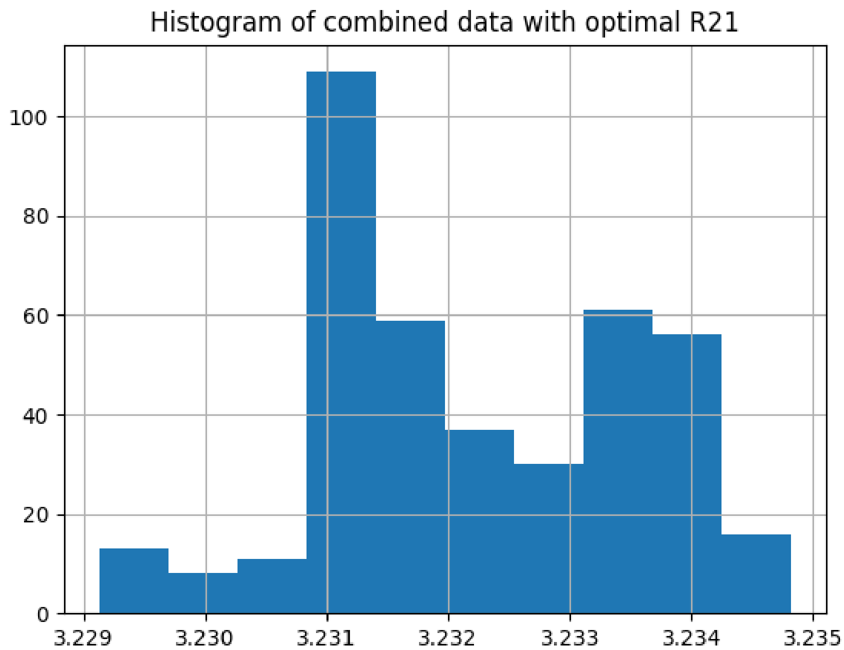

Рис. 14: Гистограмма объединенных данных при оптимальном значении R_{21}

5 Список литературы

[1] М.З.Шварц. Данные технологических испытаний оборудования для калибровки фотоприемников солнечного излучения. 2022.

[2] А.Н. Баженов. Обобщение мер совместности для анализа данных с интервальной неопределённостью. 2022.

6 Приложение

Ссылка на репозиторий с исходным кодом проекта:
https://github.com/AvitusCode/AvitusStatistics/Lab9