

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

Отчет по лабораторной работе 9  
по дисциплине  
«Математическая статистика»

Выполнил студент:  
Файзрахманов А. Р.  
группа: 5030102/90201

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

Санкт-Петербург, 2022

## Содержание

<b>1</b>	<b>Введение</b>	<b>3</b>
<b>2</b>	<b>Теория</b>	<b>3</b>
2.1	Представление данных . . . . .	3
2.2	Предварительная обработка данных . . . . .	4
2.3	Коэффициент Жаккара . . . . .	4
<b>3</b>	<b>Реализация</b>	<b>4</b>
<b>4</b>	<b>Результаты</b>	<b>4</b>
<b>5</b>	<b>Обсуждение</b>	<b>10</b>
	<b>Литература</b>	<b>10</b>

## Список иллюстраций

1	Схема установки для исследования фотоэлектрических характеристик. . . . .	3
2	Результаты измерений величины токов. . . . .	4
3	Интервальное представление данных для выборки 1. . . . .	5
4	Интервальное представление данных для выборки 2. . . . .	5
5	Линейная модель дрейфа данных из выборки 1. . . . .	5
6	Линейная модель дрейфа данных из выборки 2. . . . .	6
7	Гистограмма значений множителей коррекции $w$ из выборки 1. . . . .	6
8	Гистограмма значений множителей коррекции $w$ из выборки 2. . . . .	6
9	Скорректированная модель данных из выборки 1. . . . .	7
10	Линейная модель дрейфа данных из выборки 1. . . . .	7
11	Линейная модель дрейфа данных из выборки 2. . . . .	7
12	Гистограмма значений множителей коррекции $w$ из выборки 1. . . . .	8
13	Гистограмма значений множителей коррекции $w$ из выборки 2. . . . .	8
14	Скорректированная модель данных из выборки 1. . . . .	8
15	Скорректированная модель данных из выборки 2. . . . .	9
16	Гистограмма скорректированной модели данных из выборки 1. . . . .	9
17	Гистограмма скорректированной модели данных из выборки 2. . . . .	9
18	Зависимость коэффициента Жаккара от $R_{21}$ . . . . .	10
19	Гистограмма объединенной выборки при $R_{21}$ . . . . .	10

# 1 Введение

**Постановка задачи.** Исследование из области солнечной энергетики[1]. На Рис. 11 показана схема установки для исследования фотоэлектрических характеристик.

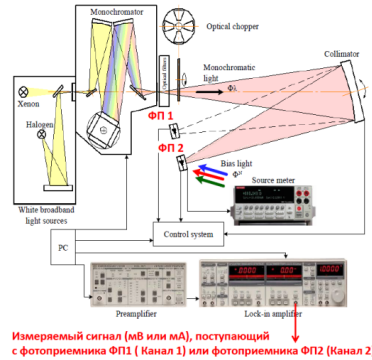


Рис. 1: Схема установки для исследования фотоэлектрических характеристик.

Калибровка датчика ФП1 производится по эталону ФП2. Зависимость между квантовыми эффективностями датчиков предполагается постоянной для каждой пары наборов измерений (1)

$$QE_{ФП2} = \frac{I_{ФП2}}{I_{ФП1}} \cdot QE_{ФП1}. \quad (1)$$

$QE_{ФП1,2}$  - квантовыми эффективностями эталонного и исследуемого датчика,  $I_{ФП1,2}$  - измеренные токи.

**Исходные данные.** Имеется 2 выборки данных с интервальной неопределенностью. Одна из них относится к эталонному датчику ФП2. Другая выборка соответствует исследуемому датчику ФП1. Данные представлены в виде двух текстовых файлов с числом отсчетов 50-200.

Названия файлов имеют формат:

$FN1 = \text{'Канал 1\_700nm0.03.csv'}$ ,

$FN2 = \text{'Канал 2\_700nm0.03.csv'}$ .

Здесь 700 и 0.03 указывают на условия проведения измерений.

## 2 Теория

Некоторые сведения по анализу данных с интервальной неопределенностью [2], [3].

### 2.1 Представление данных

В первую очередь представим данные таким образом, чтобы применить понятия данных с интервальной неопределенностью.

Один из распространенных способов получения интервальных результатов в первичных измерениях - это «обинтерваливание» точечных значений, когда к точечному базовому значению  $\hat{x}$ , которое считывается по показаниям измерительного прибора, прибавляется интервал погрешности  $\epsilon$ :

$$x = \hat{x} + \epsilon \quad (2)$$

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  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ .

Информационным множеством в случае оценивания единичной физической величины по выборке интервальных данных будет ту же интервал, который называют информационным интервалом. Неформально говоря, это интервал, содержащий значения оцениваемой величины, которые «совместны» с измерениями выборки («согласуются» с данными этих измерений).

## 2.2 Предварительная обработка данных

Зададимся линейной моделью дрейфа.

$$I_{\Phi\P} = A + B \cdot n, n = 1, 2, \dots, N. \quad (3)$$

Поставим и решим задачу линейного программирования по методике и средствами [5], найдем  $A$ ,  $B$  и вектор  $w$  множителей коррекции данных.

Также построим «спрямленные» данные выборки, вычтя из исходных данных «дрейфовую» компоненту.

$$I_{\Phi\P}^c = I_{\Phi\P} - B \cdot n, n = 1, 2, \dots, N. \quad (4)$$

## 2.3 Коэффициент Жаккара

По мере развития интервального анализа, были введены различные определения и конструкции оценки меры совместности интервальных объектов. Вместе с тем, в практике обработки данных часто необходимо оперировать с относительными величинами. Рассмотрим коэффициент Жаккара совместности интервалов.

$$JK(x) = \frac{wid(\bigwedge x_i)}{wid(\bigvee x_i)} \quad (5)$$

В выражении используется ширина интервала, а вместо операций пересечения и объединения множеств — операции взятия минимума и максимума по включению двух величин в полной интервальной арифметике (Каухера).

## 3 Реализация

Лабораторная работа выполнена на языке программирования Python с помощью библиотек numpy, matplotlib, scipy, statsmodels. Отчет написан в среде разработки TexWorks с помощью pdfLaTeX

## 4 Результаты

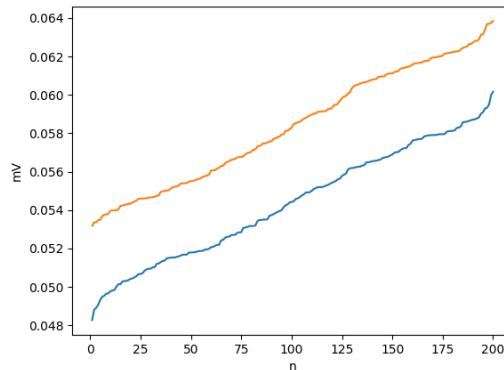


Рис. 2: Результаты измерений величины токов.

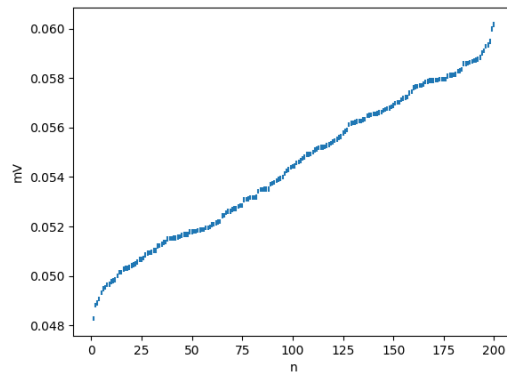


Рис. 3: Интервальное представление данных для выборки 1.

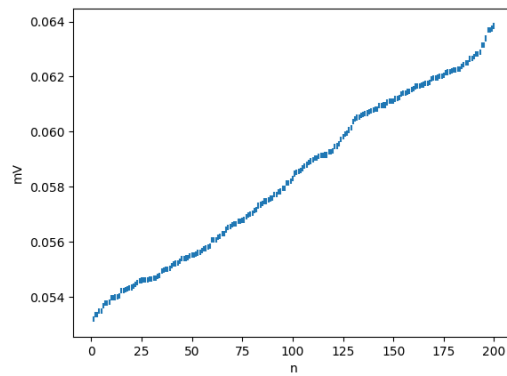


Рис. 4: Интервальное представление данных для выборки 2.

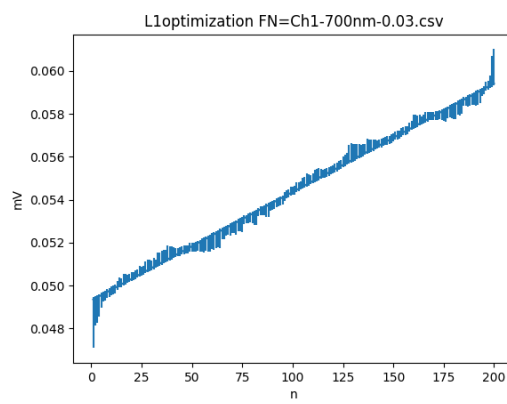


Рис. 5: Линейная модель дрейфа данных из выборки 1.

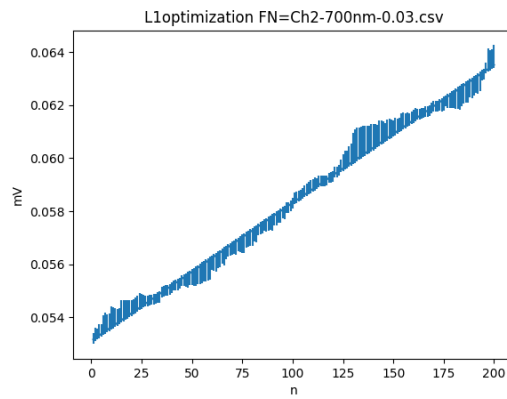


Рис. 6: Линейная модель дрейфа данных из выборки 2.

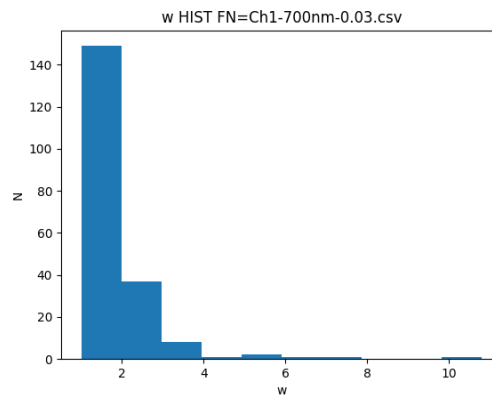


Рис. 7: Гистограмма значений множителей коррекции  $w$  из выборки 1.

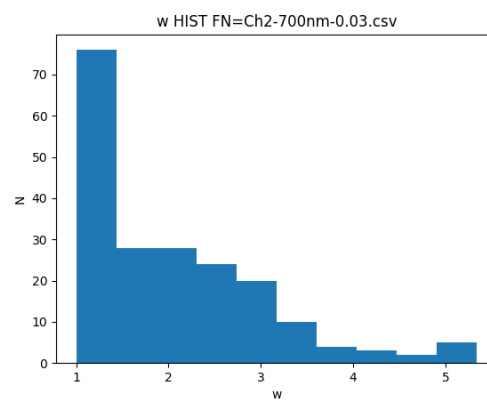


Рис. 8: Гистограмма значений множителей коррекции  $w$  из выборки 2.

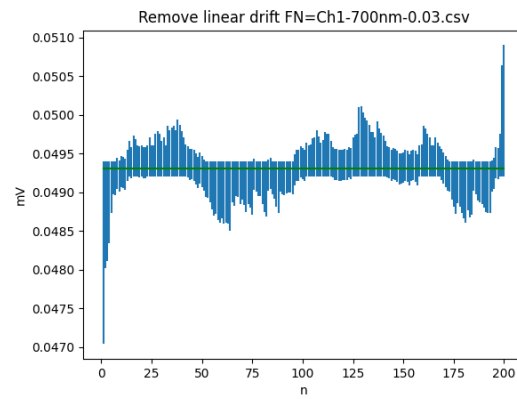


Рис. 9: Скорректированная модель данных из выборки 1.

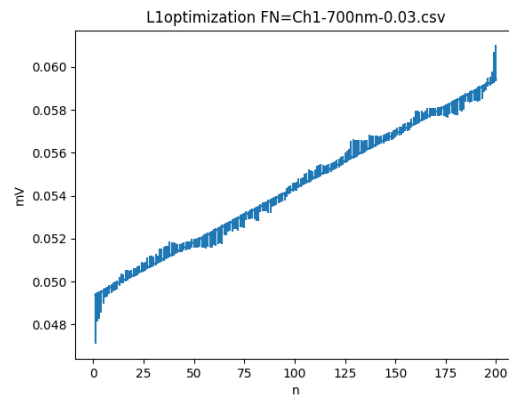


Рис. 10: Линейная модель дрейфа данных из выборки 1.

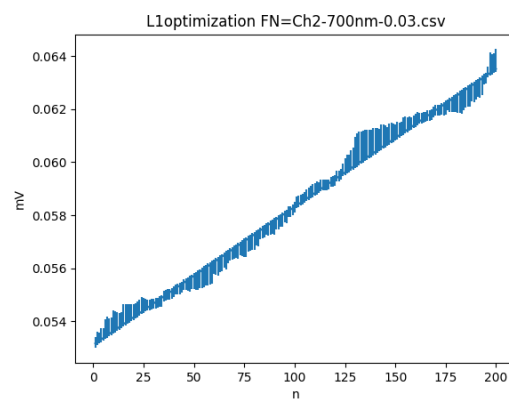


Рис. 11: Линейная модель дрейфа данных из выборки 2.



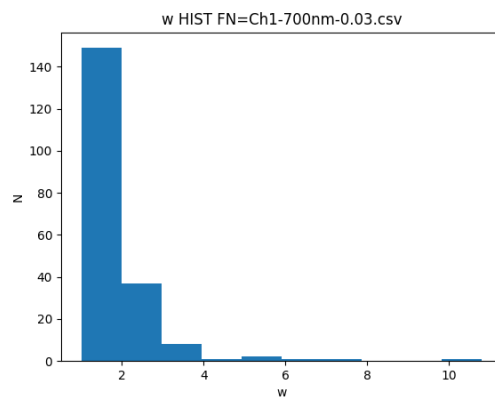


Рис. 12: Гистограмма значений множителей коррекции  $w$  из выборки 1.

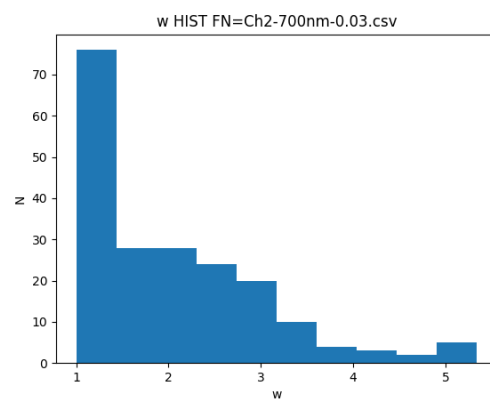


Рис. 13: Гистограмма значений множителей коррекции  $w$  из выборки 2.

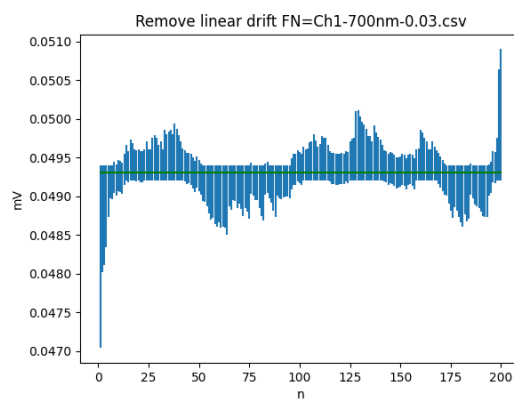


Рис. 14: Скорректированная модель данных из выборки 1.

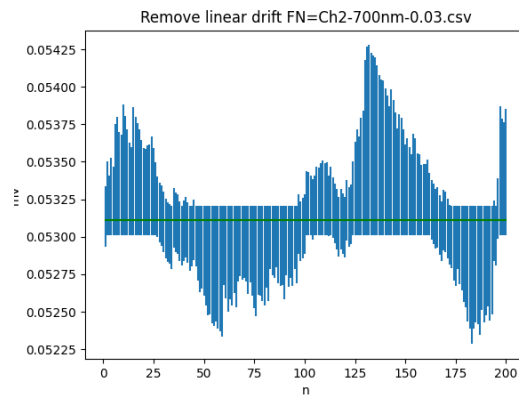


Рис. 15: Скорректированная модель данных из выборки 2.

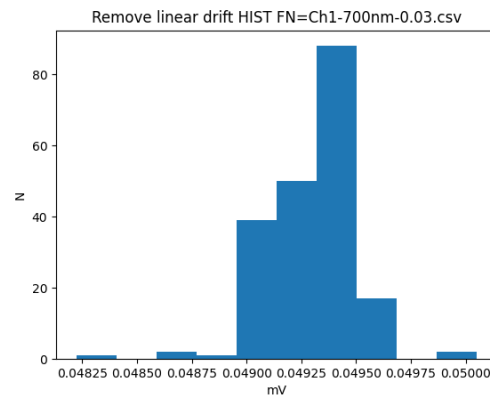


Рис. 16: Гистограмма скорректированной модели данных из выборки 1.

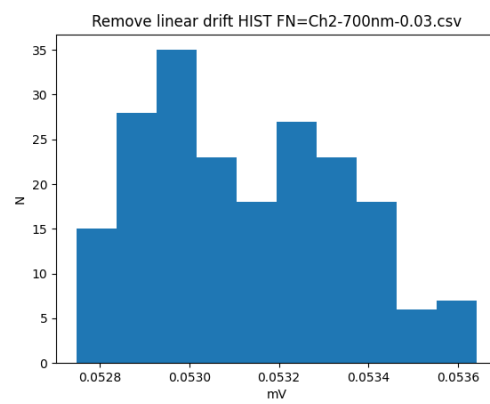


Рис. 17: Гистограмма скорректированной модели данных из выборки 2.

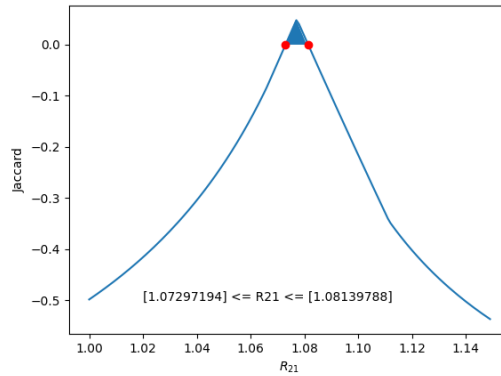


Рис. 18: Зависимость коэффициента Жаккара от  $R_{21}$

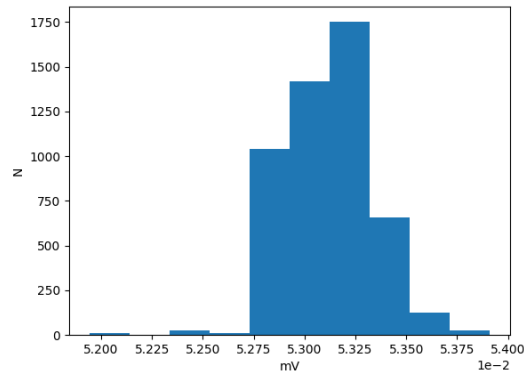


Рис. 19: Гистограмма объединенной выборки при  $R_{21}$

## 5 Обсуждение

Как видно из гистограмм значений множителей коррекции  $w$ , половина массива данных не требует коррекции. Этот факт свидетельствует в пользу того, что линейная модель дрейфа данных является разумным приближением. Из гистограммы объединенной модели видно, что она имеет характерные черты, как гистограммы скорректированной модели данных выборки 1, так и гистограммы скорректированной модели данных выборки 2.

## Литература

- [1] М.З. Шварц. Данные технологических испытаний оборудования для калибровки фотоприемников солнечного излучения. 2022
- [2] А.Н. Баженов, С.И. Жилин, С.И. Кумков, С.П. Шарый. Обработка и анализ данных с интервальной неопределенностью. 2022
- [3] А.Н. Баженов. Введение в анализ данных с интервальной неопределенностью. 2022
- [4] Коэффициент Жаккара [https://en.wikipedia.org/wiki/Jaccard\\_index](https://en.wikipedia.org/wiki/Jaccard_index) Jaccard P. Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines // Bull. Soc. Vaudoise sci. Natur. 1901. V. 37. Bd. 140. S. 241-272.

- [5] С.И. Жилин. Примеры анализа интервальных данных в Octave <https://github.com/szhilin/octave-interval-examples>