Санкт-Петербургский политехнический университет Петра Великого

Физико-механический институт

Высшая школа прикладной математики и вычислительной физики

Интервальный анализ
Отчёт по лабораторной работе №1

Выполнил:

Студент: Дамаскинский Константин
Группа: 5040102/10201

Принял:

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

2022 г.

Содержание

Список иллюстраций

1. Постановка задачи

Проводится исследование из области солнечной энергетики. На рис. 1 показана схема установки для исследования фотоэлектрических характеристик.

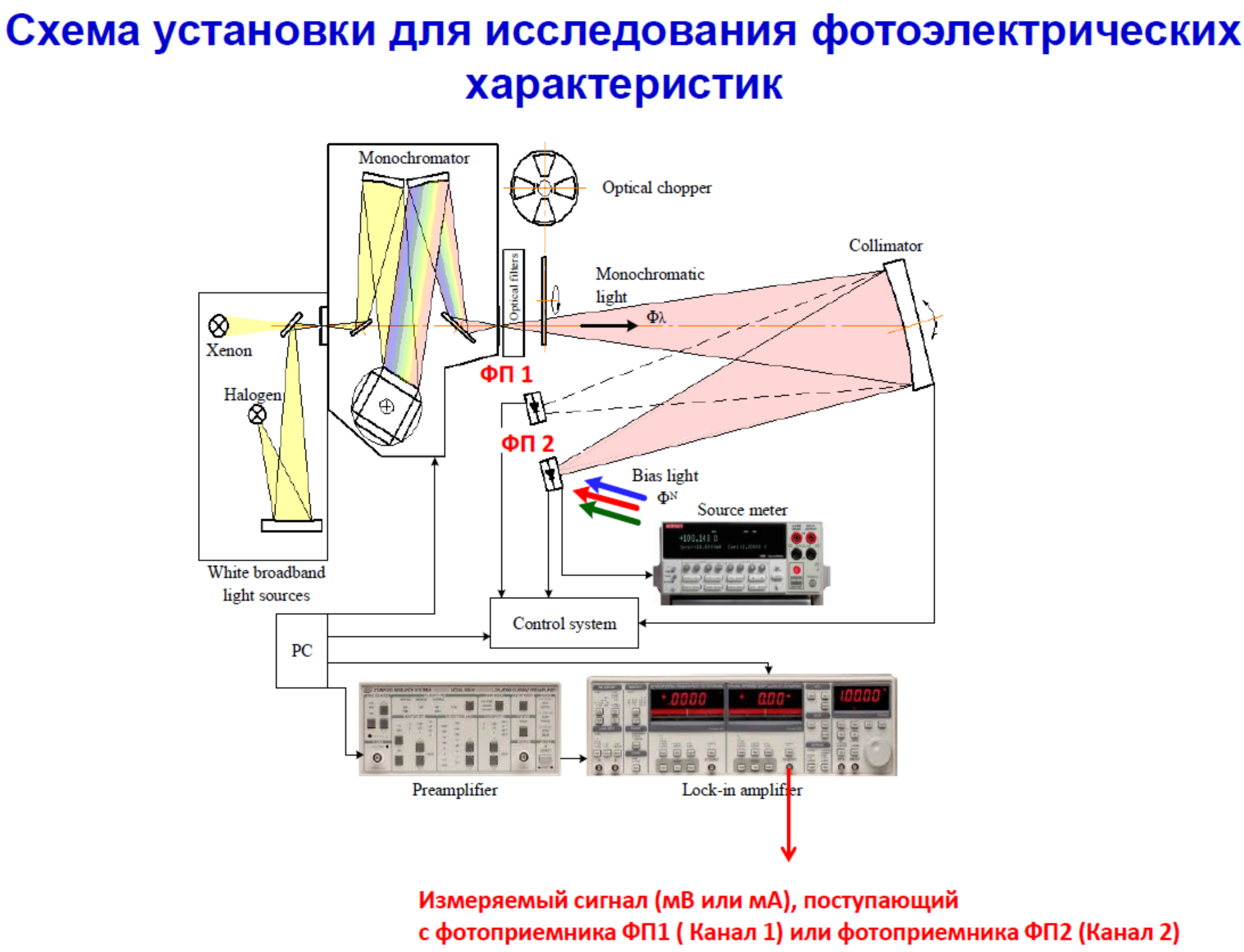

Рис. 1. Схема установки для исследования фотоэлектрических характеристик

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

$$QE_2 = \frac{I_2}{I_1} \cdot QE_1 \tag{1}$$

где QE_2, QE_1 – эталонная эффективность эталонного и исследуемого датчика, I_2, I_1 – измеренные токи. Данные с датчиков находятся в файлах **Канал2_800nm_0.2.csv**, **Канал1_800nm_0.2.csv** и полагаются линейными.

Требуется определить коэффициент калибровки:

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

на основе линейной регрессии на множестве интервальных данных и коэффициента Жаккара.

2. Теория

2.1. Простая линейная регрессия для вещественных данных

Пусть заданы две последовательности $X = \{x_i\}_{i=1}^n, Y = \{y_i\}_{i=1}^n,\ x_i, y_i \in \mathbb{R}\ \forall i = \overline{1,n}$. **Простой линейной регрессией** для этих последовательностей называется функция:

$$f(x) = \beta_0 + \beta_1 \cdot x \tag{3}$$

подобранная так, чтобы вектор $F = \{f(x_i)\}_{i=1}^n$ был в каком-то смысле максимально близок к вектору Y.

Таким образом, для решения задачи простой линейной регрессии необходимо найти коэффициенты β_0, β_1. В зависимости от выбираемого метода поиска коэффициентов будет меняться и мера близости подобранной линейной функции к вектору Y.

В данной работе будет использоваться метод наименьших квадратов (МНК). Данный метод позволяет решить задачу простой линейной регрессии, поставив задачу минимизации второй (евклидовой) нормы разности векторов F и Y:

$$\sum_{i=1}^{n} \|\beta_0 + \beta_1 x_i - y_i\|_2 \xrightarrow[\beta_0, \beta_1]{} \min \tag{4}$$

2.2. Обынтерваливание данных для интервальной регрессии

Поскольку показания датчиков обладают погрешностью, полученные данные на самом деле следует рассматривать как интервалы, центр которых совпадает со считанными показаниями, а радиус равен некоторой базовой погрешности ε, умноженной на вес w_i. ε является константой.

Для каждого из наборов данных $X^{(1)}$ и $X^{(2)}$, прочитанных из соответствующих файлов, построим простую линейную регрессию на вещественных числах в результате чего получим аппроксимацию:

$$Lin_k(i) = a_i^{(k)} \cdot i + b_i^{(k)},\ k \in \{1, 2\},\ i = \overline{1,n} \tag{5}$$

Определим для каждой из выборки вектор весов W_k простым способом: если значение аппроксимирующей прямой Lin_k в точке i не попадает в интервал $x_i^{(k)} \pm \varepsilon$, то увеличим радиус интервала в $w_i^{(k)}$ раз так, чтобы $Lin_k(i)$ оказалось на одной из границ интервала.

После того, как мы получили два интервальных вектора из $\mathbb{IR}^n$, вычтем из $x_i^{(k)}$ "наклонную" составляющую $a_i^{(k)} \cdot i$, получив таким образом "горизонтальные" векторы, для которых будем находить искомый коэффициент пропорциональности R_{21}.

2.3. Коэффициента Жаккара. Поиск R_{21}

Коэффициент Жаккара позволяет оценить, насколько хорошо совмещаются друг с другом заданные интервалы $x_1, \ldots, x_n$. Вычисляется путём деления длины интервала-пересечения на длину интервала объединения по формуле:

$$JK(x_1, \ldots, x_n) = \frac{wid\left(\bigcap\limits_{i=\overline{1,n}} x_i\right)}{wid\left(\bigcup\limits_{i=\overline{1,n}} x_i\right)} \tag{6}$$

Используя данный коэффициент, мы можем подобрать такой $R_{21} \in \mathbb{R}$, чтобы полученные интервалы X_2 и $R_{21} \cdot X_1$ были максимально совместны. Для этого необходимо вычислять коэффициент Жаккара для совокупности компонент этих векторов.

Таким образом, для того, чтобы найти R_{21}, необходимо задать нижнюю и верхнюю границы поиска $\underline{R}, \overline{R}$, а затем при помощи бинарного поиска найти точку максимума коэффициента Жакккара в зависимости от выбранного R_{21}.

Числа $\underline{R}, \overline{R}$ можно найти тривиально, поделив наименьшую верхнюю границу среди интервалов вектора $R_{21} \cdot X_1$ на наибольшую нижнюю границу среди интервалов вектора X_2 и, соответственно, наибольшую на наименбшую соответствующие границы.

3. Реализация

Данная работа реализована на языке программирования Python 3.10 с использованием пакетов numpy и scikitlearn. Код данного отчёта подготовлен с использованием редактора TeXstudio и компилятора pdflatex.

4. Результаты

Ниже приведены графики, полученные в результате работы реализованной программы.

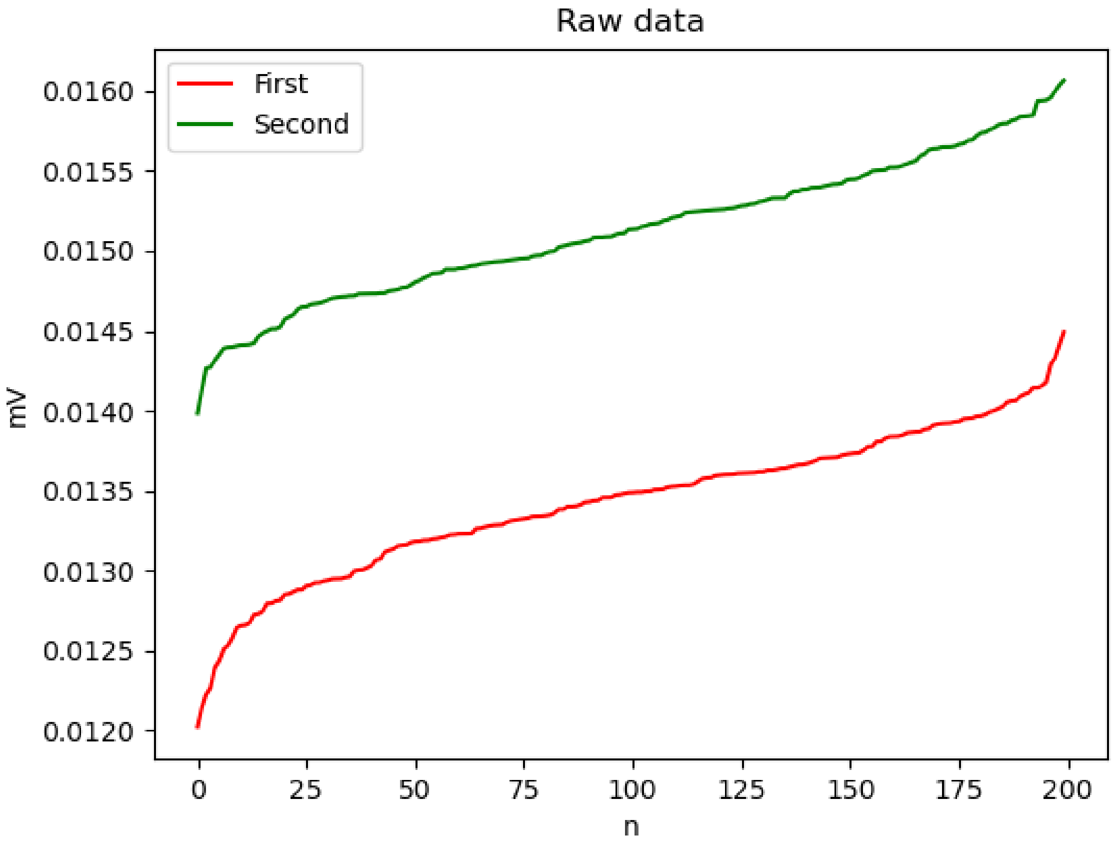

Рис. 2. Загруженные данные

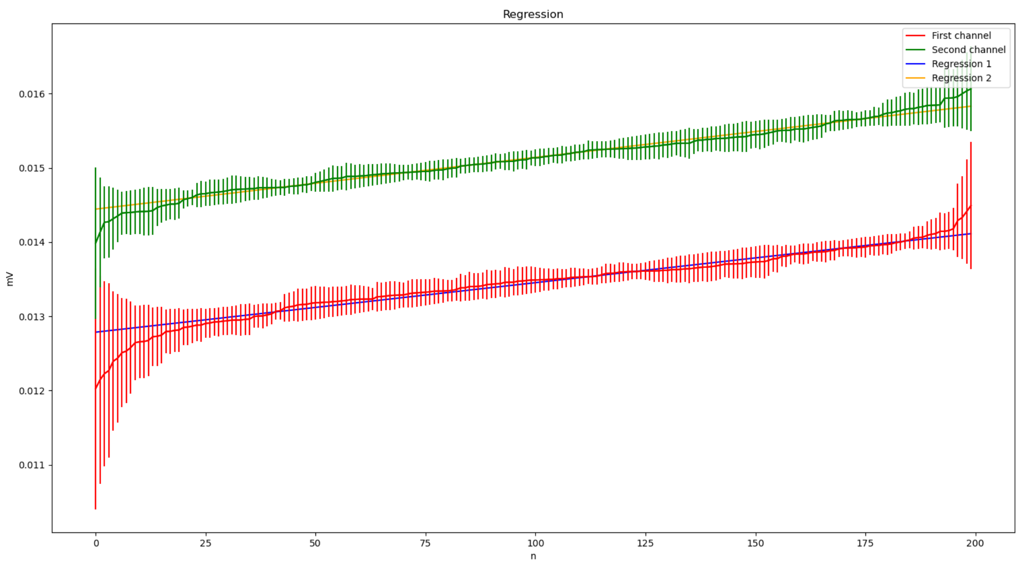

Рис. 3. Линейная регрессия для вещественных данных и результаты обынтерваливания

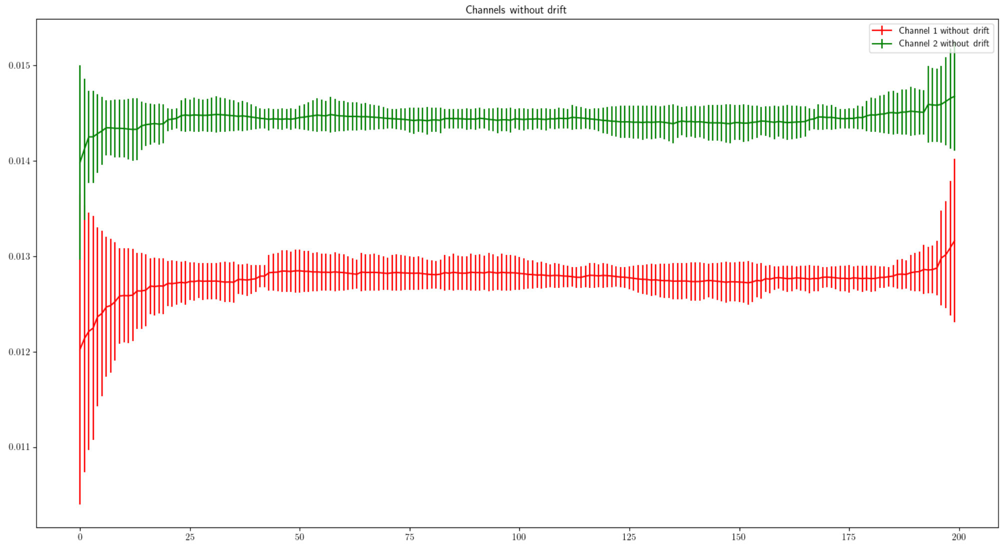

Рис. 4. Данные после вычитания “наклонной” составляющей

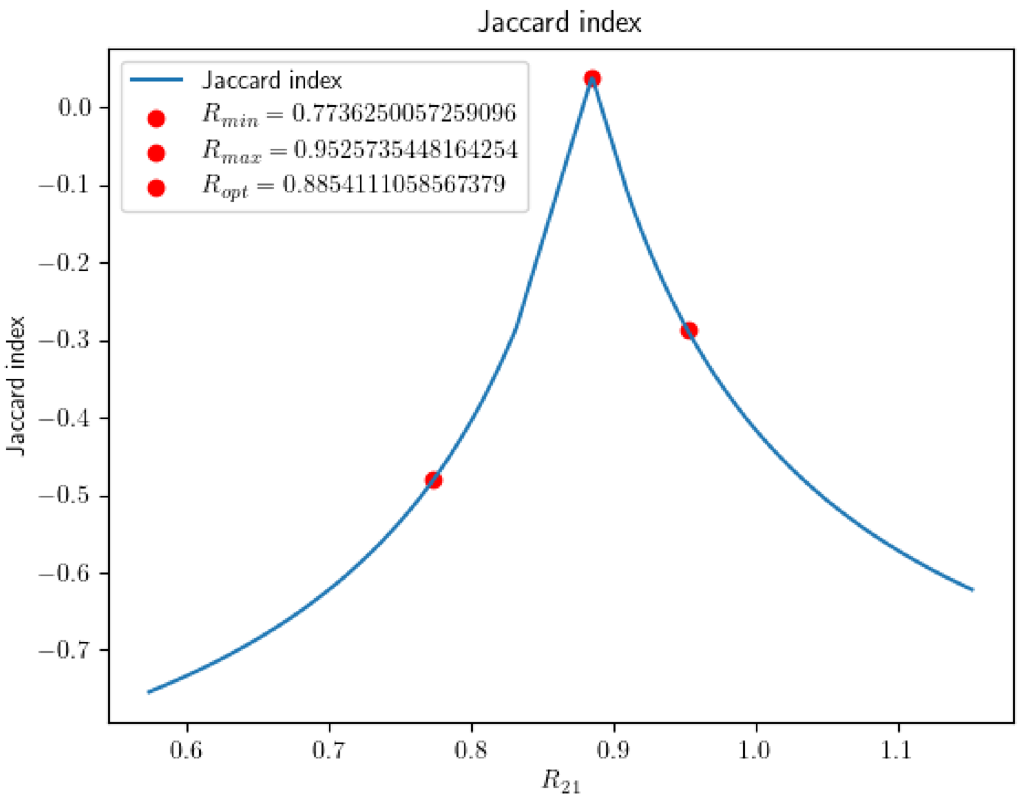

Рис. 5. Зависимость коэффициента Жаккара от R_{21}

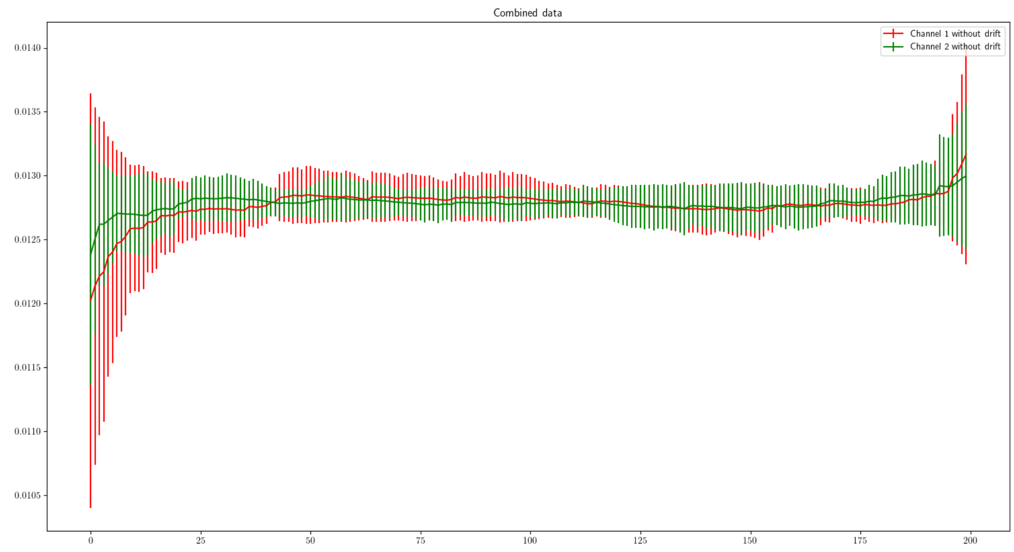

Рис. 6. Результат наложения данных при максимальном коэффициенте Жаккарда

5. Обсуждение

Исходя из представленных графиков, можно судить о том, что все описанные в теории этапы выполнены правильно.

- Простая линейная регрессия и обынтерваливание проведены так, что каждый интервал содержит соответствующую точку аппроксимирующей прямой, при этом аппроксимирующая прямая лежит визуально близко к исходным данным.
- В результате отсечения наклонной части действительно получились визуально горизонтальные графики.
- График зависимости коэффициента Жаккара от искомого множителя ожидаемо имеет один локальный максимум. При этом видно, что оценка интервала R_{21} с точки зрения меры Жаккара действительно очень грубая: значение коэффициента Жаккара в нижней оценке приблизительно равно -0.5. Данное число привело бы к абсолютно неприемлемому результату интервальной регрессии: хоть точечные наборы и получились бы визуально похожими, этого нельзя было бы сказать про интервалы. Учитывая характер полученных данных, важно удостовериться именно в максимальном совпадении интервалов.
- На последнем рисунке видно, что значительная часть интервалов совпадает практически идеально, что также является показателем качественно выполненной работы.

6. Приложения

1. Репозиторий с кодом программы и кодом отчёта:
   https://github.com/kystyn/interval2