Санкт-Петербургский политехнический университет Петра Великого

Физико-механический институт

Высшая школа прикладной математики и вычислительной физики

Лабораторная работа № 1  
на тему:

«Обработка постоянной. Применение меры совместности к анализу данных»

|  |  |  |
| --- | --- | --- |
| Выполнил |  |  |
| студент гр. 5040102/10201 | Логинов И.А. | /\_\_\_\_\_\_\_\_\_\_\_/ |
| Руководитель |  |  |
| доцент, к.ф.-м.н. | Баженов А.Н. | /\_\_\_\_\_\_\_\_\_\_\_/ |

Санкт-Петербург

2022

# Постановка задачи

Проводится исследование из области солнечной энергетики.

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

|  |  |
| --- | --- |
|  | (1) |
|  |  |

, – эталонная эффективность эталонного и исследуемого датчика,   
, , или , – измеренные мощности. Данные датчиков находятся в файлах “Канал 1\_900nm\_0.23.csv” и “Канал 2\_900nm\_0.23.csv”.

Требуется определить параметры постоянной величины на основе двух выборок , , в частности коэффициент калибровки

|  |  |
| --- | --- |
|  | (2) |
|  |  |

при помощи линейной регрессии, интервальных данных и коэффициента Жаккара.

# Теоретическая часть

Самый известный способ получения интервальных результатов в первичных измерениях – это «обинтерваливание» точечных значений, когда к точечному базовому значению , которое считывается по показаниям измерительного прибора прибавляется интервал погрешности :

|  |  |
| --- | --- |
|  | (3) |
|  |  |

В конкретных измерениях мВ. Согласно терминологии интервального анализа, рассматриваемая выборка – это вектор интервалов, или интервальный вектор

Построение интервалов будет происходить следующим образом:

Для начала построим линейную регрессию по известному методу наименьших квадратов в виде , где – номер измерения; – прямая, аппроксимирующая экспериментальные измерения . Отклонение можно вычислить как:

|  |  |
| --- | --- |
|  | (4) |
|  |  |

Если отдельные интервалы не заключают в себе линейную регрессию, к отклонение стоит растянуть, домножить на величину , минимально возможную, для того, чтобы интервал коснулся линии регрессии.

Интервальные данные представляются в виде:

|  |  |
| --- | --- |
|  | (5) |

или кратко – множество всех интервальных данных, построенных по измерениям датчика ФП1, , .

Чтобы сделать интервальную величину более константной и в дальнейшем оценить совместность двух выборок экспериментальных измерений, следует вычесть из интервальных данных линейную зависимость (фактически из концов интервала), получим:

|  |  |
| --- | --- |
|  | (6) |
|  |  |

Для базовых значений выполняются аналогичные вычисления. Находится линейная зависимость , интервалы по формуле (5) и обработанные интервалы по формуле (6) с соответствующими индексами.

В различных областях анализа данных используют различные меры сходства множеств, иными словами, коэффициенты сходства. В данной работе используется мультимера Жаккакра, то есть ее модификация для интервальных данных:

|  |  |
| --- | --- |
|  | (7) |
|  |  |

Мера Жаккара численно характеризует меру совместности интервальных данных. В качестве рассматриваются интервальные данные объединенной выборки . – число, получаемое в результате деления пересечения интервалов на их объединение. Заметим, что если при подборе калибровочного множителя получается , то выборка совместна (имеет положительную меру совместности). Поиск оптимального можно представить так:

|  |  |
| --- | --- |
|  | (8) |

– это аргумент, у которого реализуется данный функционал, максимальная оценка коэффициента калибровки из формулы (2). Внешнюю оценку для можно найти разными способами, проще всего путем деления интервалов двух выборок , в результате чего получим интервал внешней оценки – такой интервал, в котором можно найти , перебирая с некоторым шагом и вычисляя функционал (8). Интервал, в пределах которого наблюдается является внутренней оценкой коэффициента .

# Результаты исследования

Программный код написан на языке программирования Python с использованием библиотек MatPlotLib, NumPy и Sklearn.

На рис.1 представлены экспериментальные данные, измеренные двумя датчиками, на рис.2 и рис.3 – те же данные, но в другом масштабе. На рис. 4 и 5 показаны построенные согласно описанной выше теории интервальные данные и линейная регрессия с коэффициентами

.

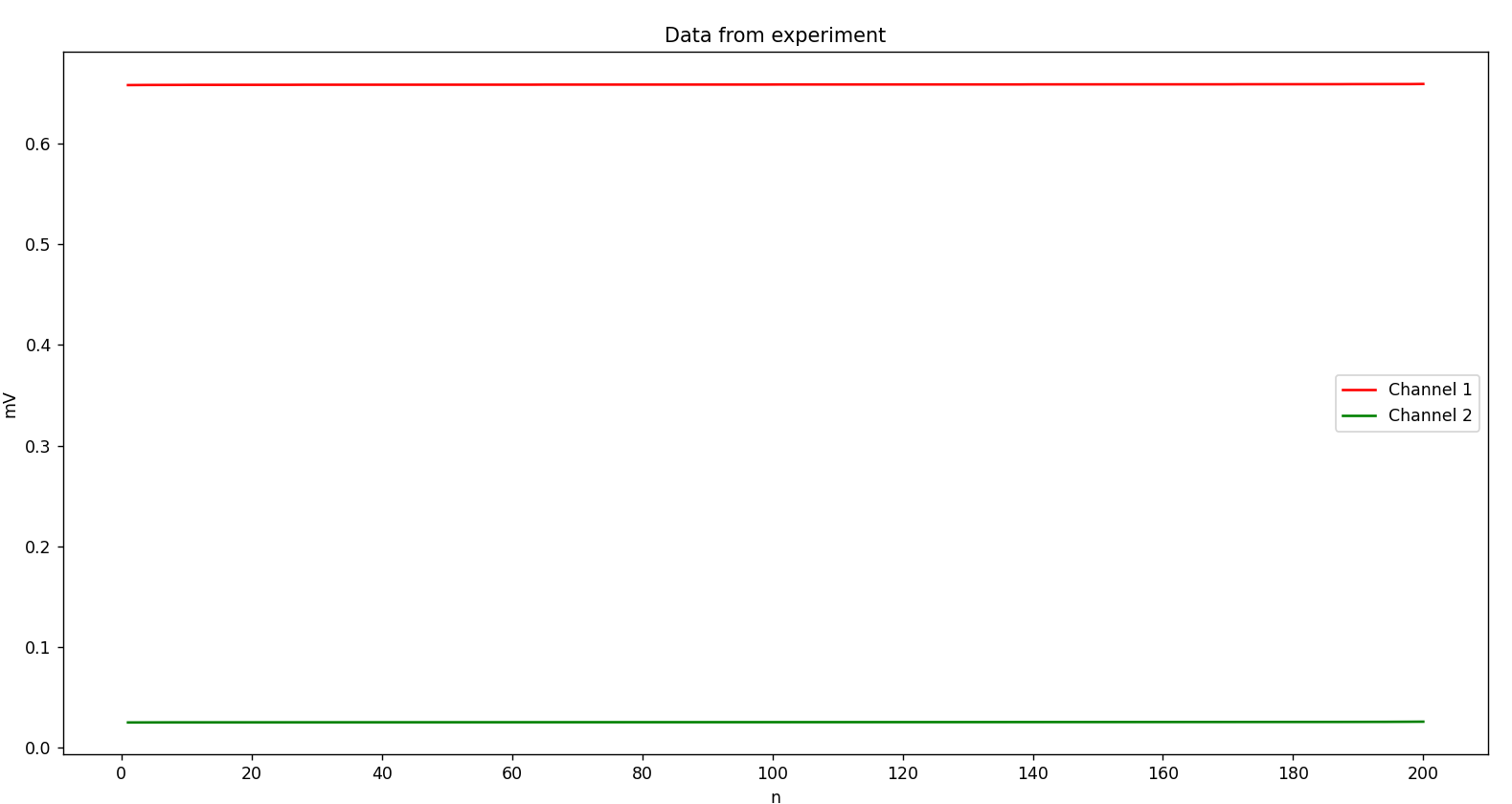


Рис. 1. Две выборки экспериментальных данных, измеренным датчиками

1. Измеренные данные:

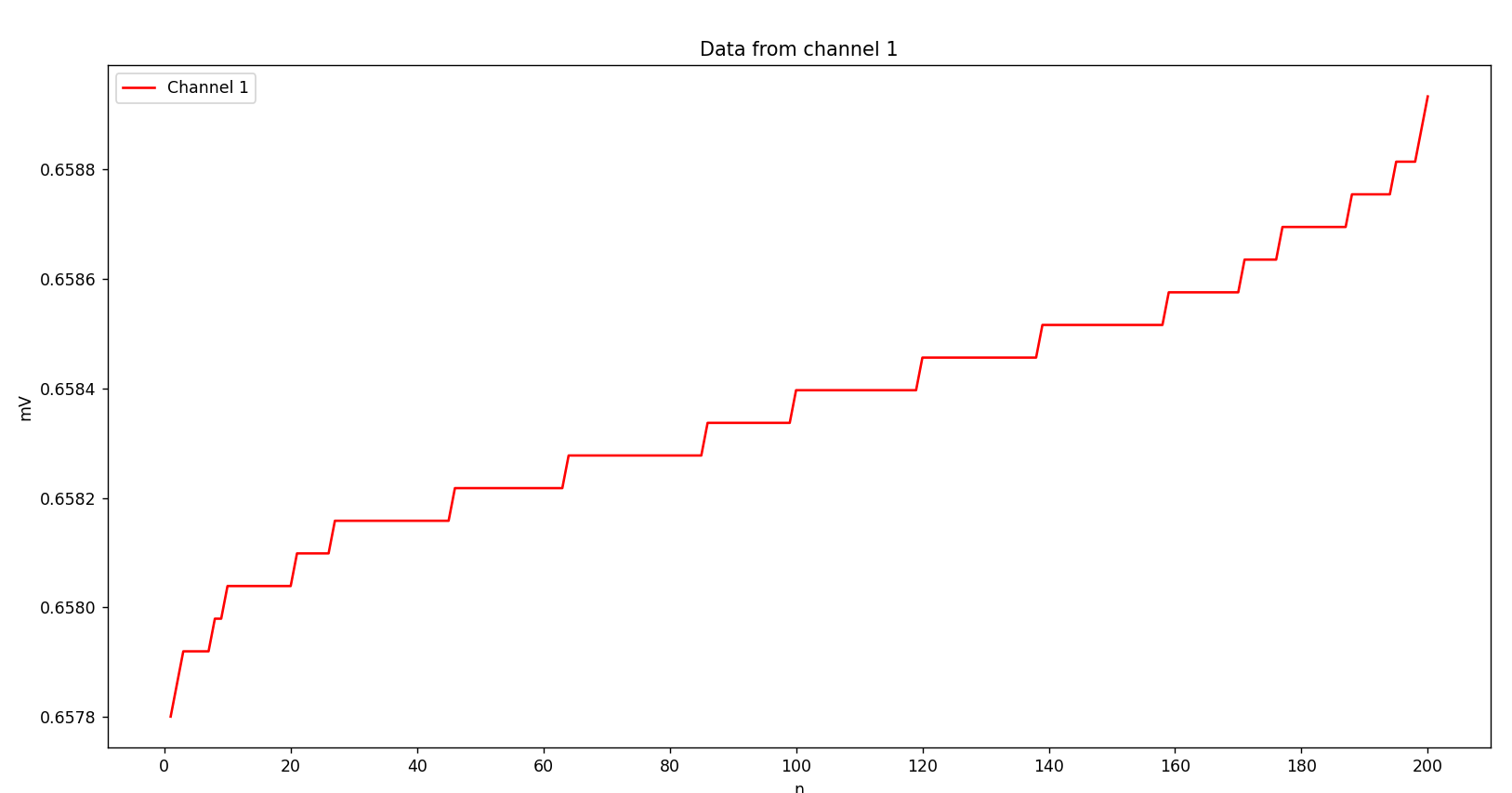


Рис. 2. Данные, измеренные датчиком ФП1

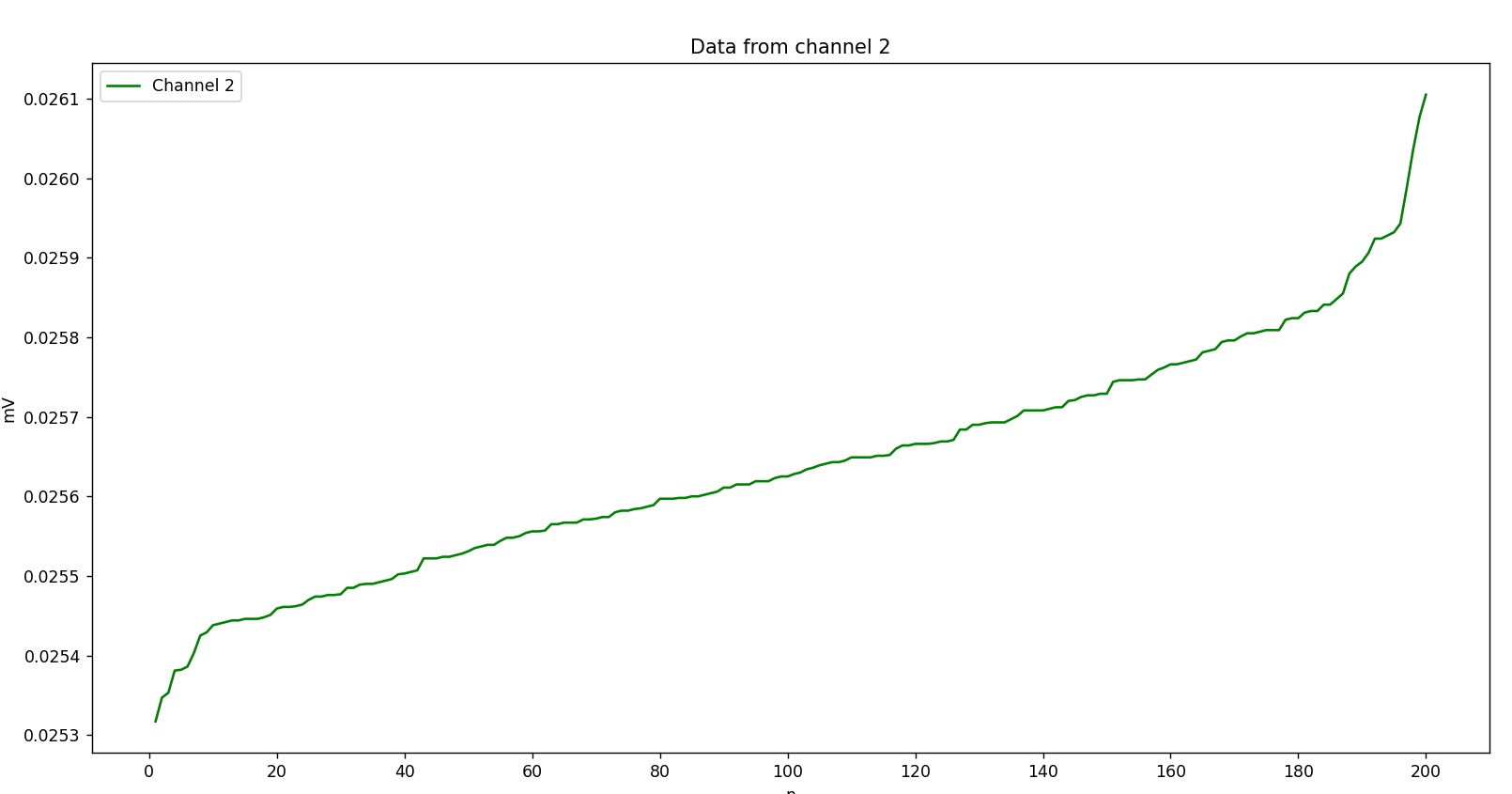


Рис. 3. Данные, измеренные датчиком ФП2

2. Интервальные данные:

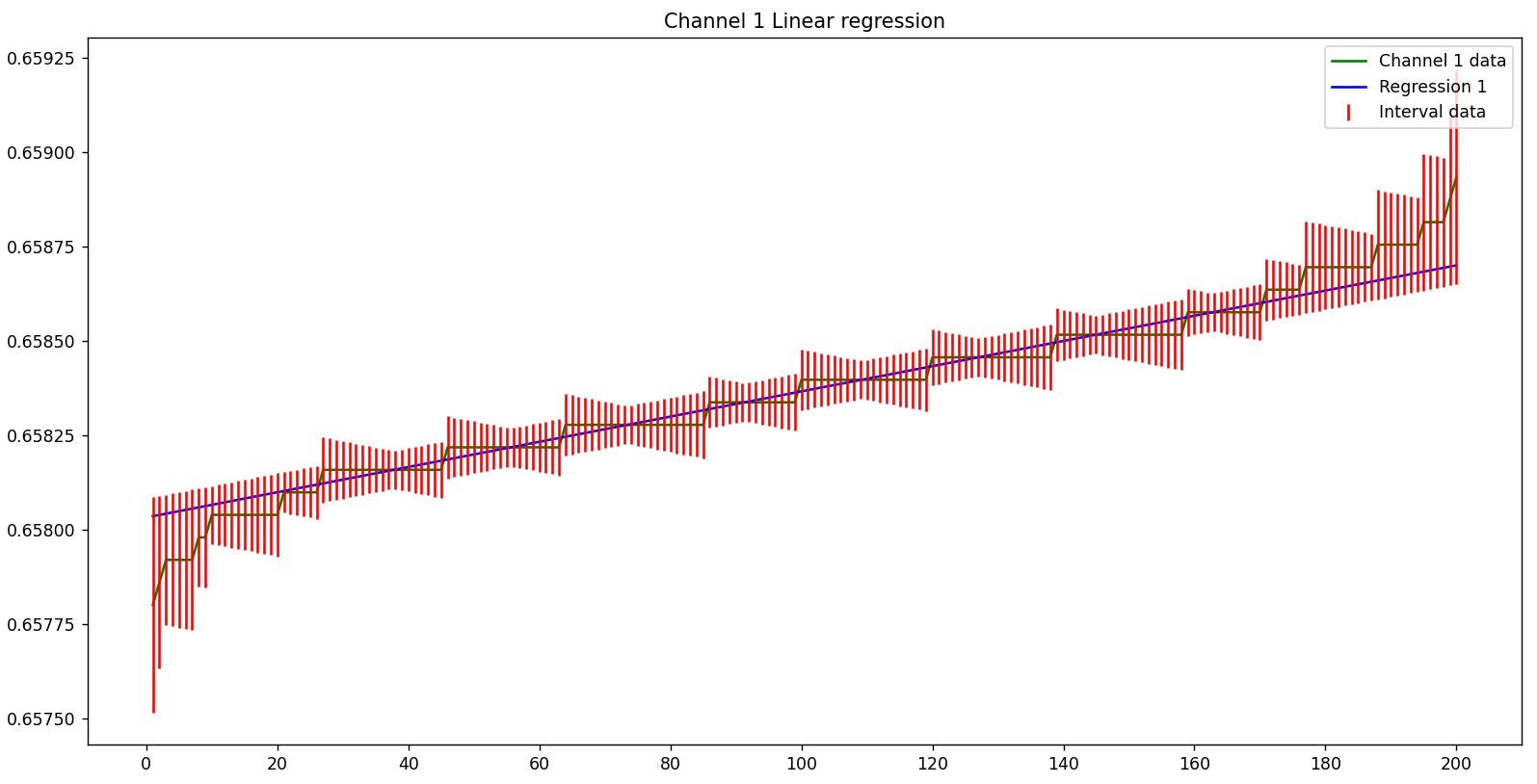


Рис. 4. Интервальные данные первой выборки и линейная регрессия

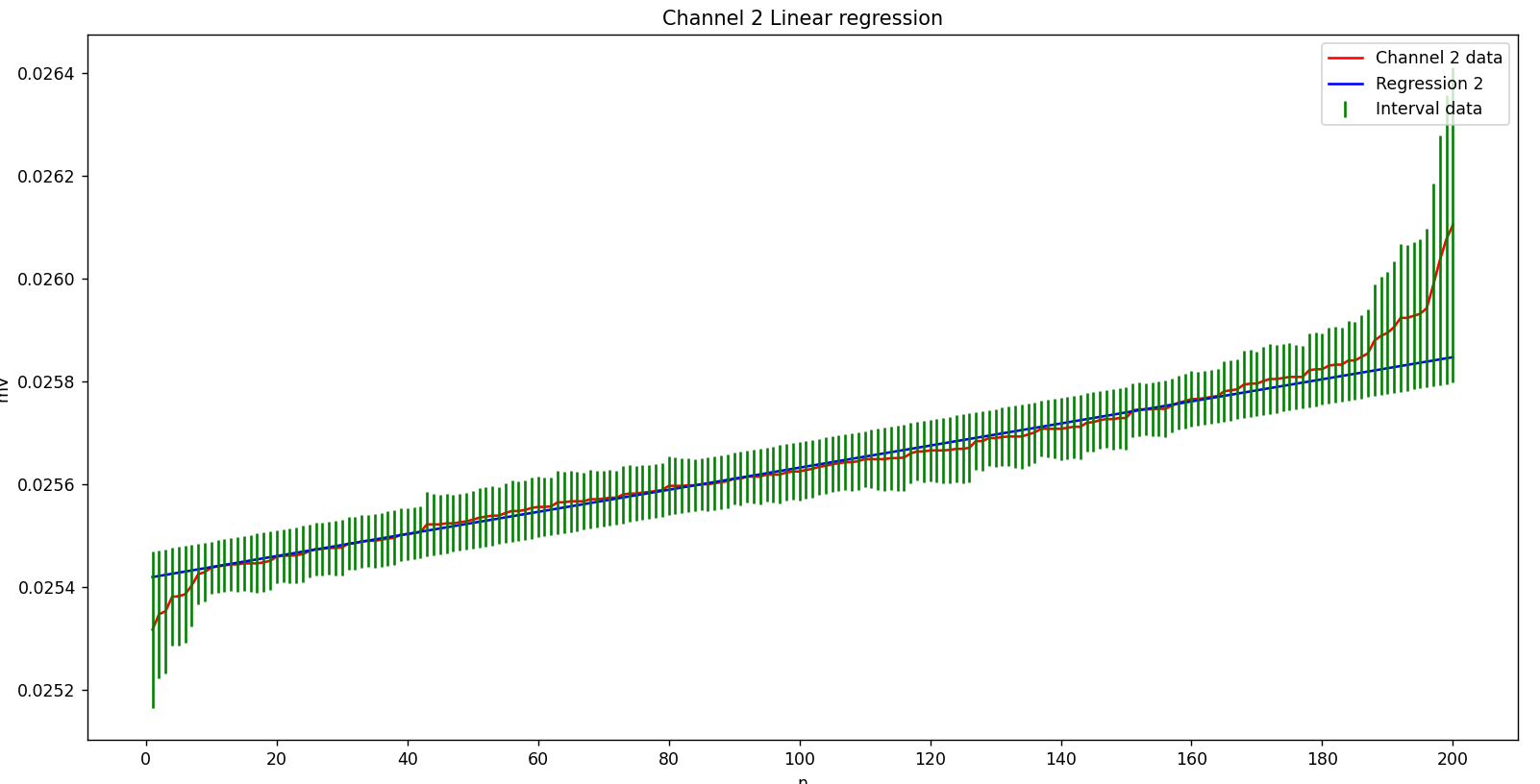


Рис. 5. Интервальные данные второй выборки и линейная регрессия

На рис. 6 визуализирован пример совместных выборок , что выполняется при , обеспечивающим .

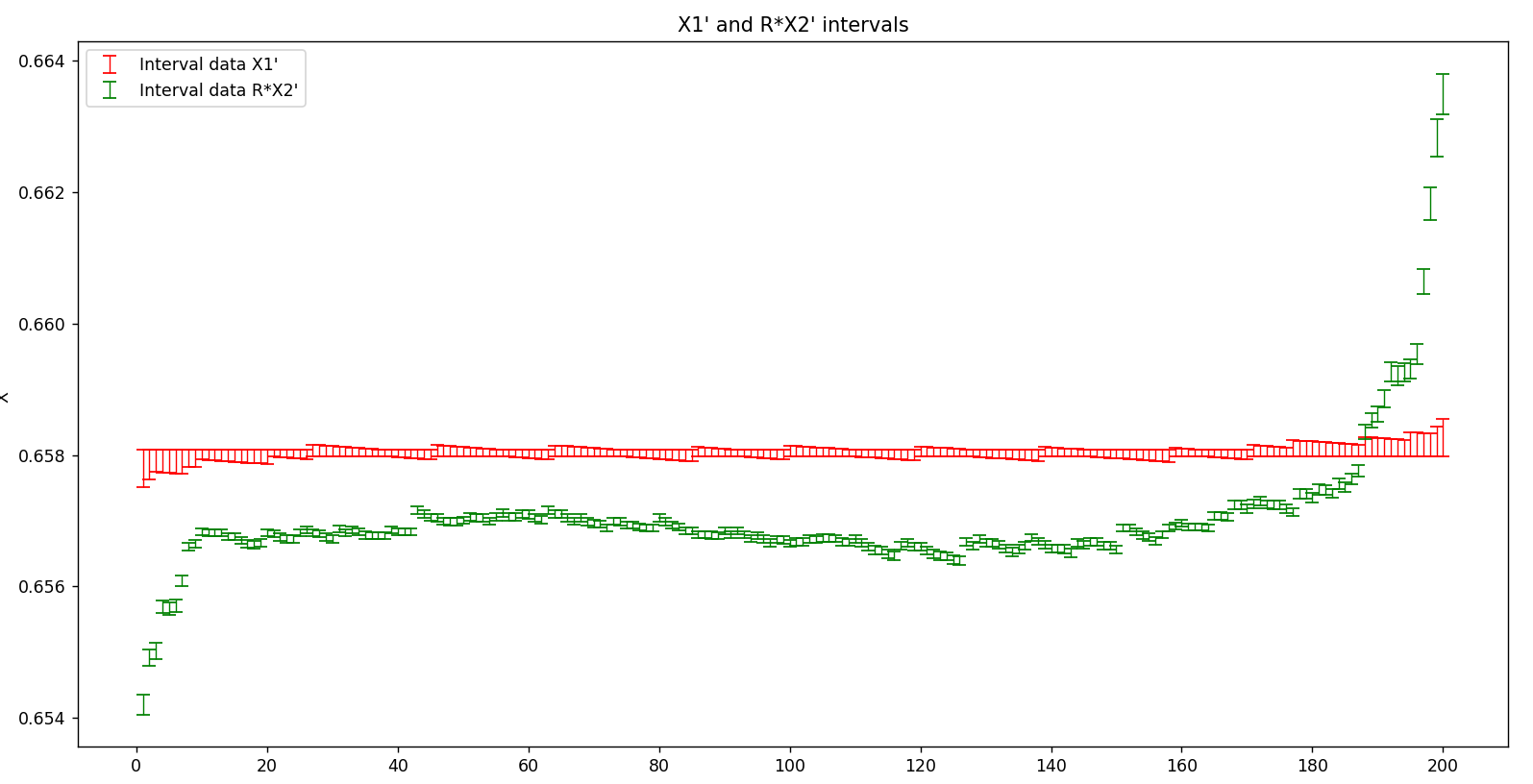


Рис. 6. Обработанные интервальные данные совместной выборки при R,   
обеспечивающем совместность выборок

**3. Мультимера Жаккара:**

На рис. 7 показана зависимость коэффицциента Жаккара от коэффициента калибровки . Согласно внешней оценке оптимальное значение осуществлялся в диапазоне . Как интервал можно представить . В нашем эксперименте, максимум коэффициента Жаккара имеет значение 0.00472.

Это связано с наличием различных погрешностей, которые на практике невозможно устранить, но несмотря на их присутствие, поведение коэффициента Жаккара позволило найти оптимальный калибробочный коэффициент .

Таким образом, можно сказать, что область, где является оценкой искомой величины .

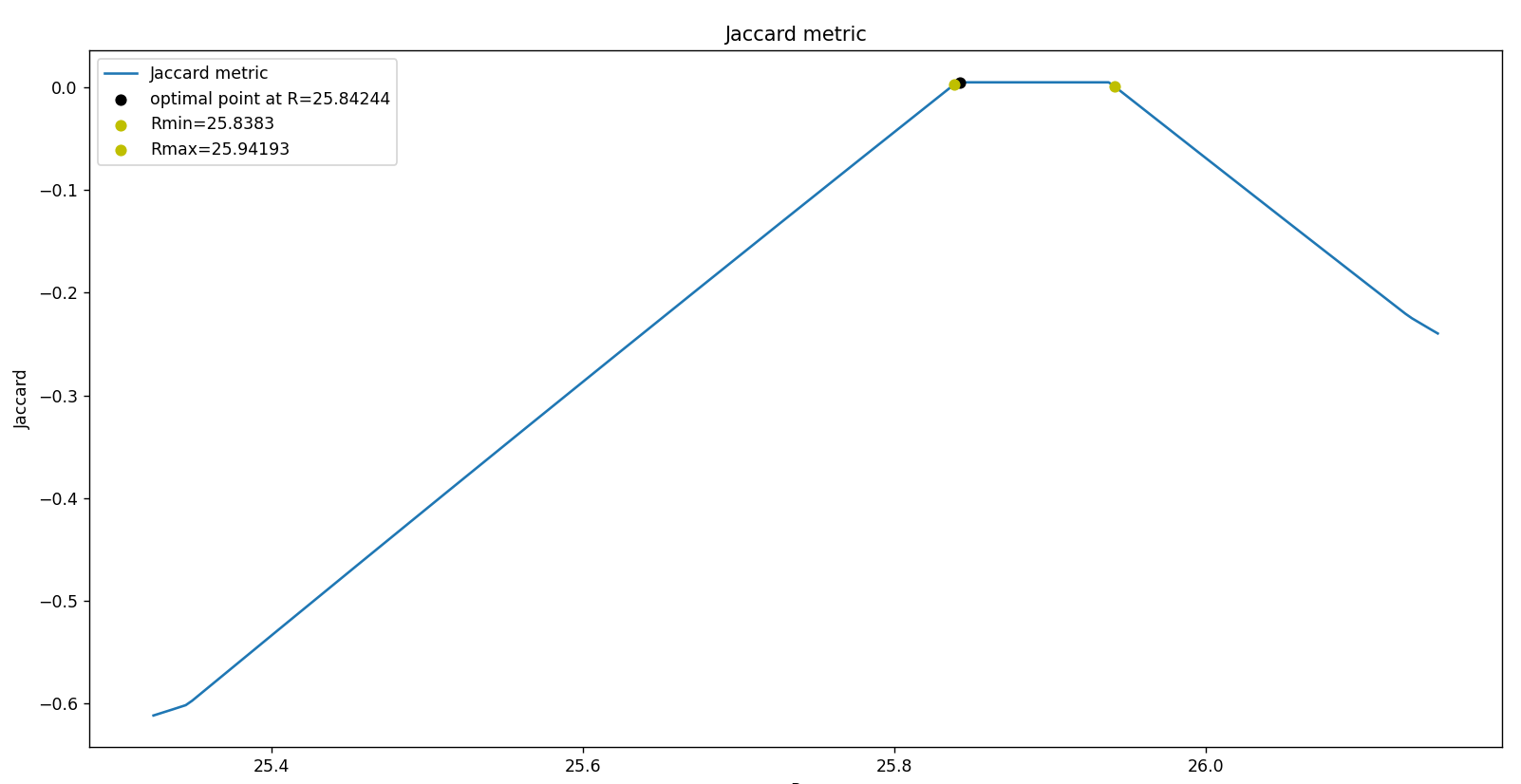


Рис. 7. Значения коэффициента Жаккара от коэффициента калибровки

**Ссылка на GitHub с реализацией**

<https://github.com/ivanandreich/interval_data_analysis/tree/master/1>

Файлы данных:

Канал 1\_900nm\_0.23.csv

Канал 2\_900nm\_0.23.csv

**Коэффициенты линейной регрессии**

|  |  |  |
| --- | --- | --- |
| № выборки |  |  |
| 1 | 3.3368e-06 | 0.6580321 |
| 2 | 2.1512e-06 | 0.0254172 |