

Санкт-Петербургский политехнический университет Петра Великого
физико-механический институт
кафедра "Прикладная математика"

Отчёт

по лабораторной работе №3
по дисциплине
"математическая статистика"

Работу

выполнил:

И.С. Золотухин

Группа:

5030102/90101

Преподаватель:

К.ф.-м.н., доцент

Баженов А. Н.

Санкт-Петербург
2022

Содержание

1. Введение	4
2. Теория	5
2.1. Анализ данных с интервальной неопределённостью.	5
2.2. Линейная регрессия	5
2.2.1. Описание модели	5
2.2.2. Метод наименьших модулей	5
2.3. коэффициент Жаккара	6
2.4. Процедура оптимизации	6
3. Реализация	6
4. Результаты	6
5. Литература	12

Список иллюстраций

1.1. Схема установки для исследования фотоэлектрических характеристик . .	4
4.1. Исходные данные	6
4.2. Интервальное представление исходных данных ch1	7
4.3. Интервальное представление исходных данных ch2	7
4.4. Линейная модель дрейфа данных ch1	8
4.5. Линейная модель дрейфа данных ch2	8
4.6. Гистограммы значений множителей коррекции w ch1	9
4.7. Гистограммы значений множителей коррекции w ch2	9
4.8. Скорректированная модель данных ch1	10
4.9. Скорректированная модель данных ch2	10
4.10. Гистограмма скорректированных данных ch1	11
4.11. Гистограмма скорректированных данных ch2	11
4.12. Значение коэффициента Жаккара от калибровочного множителя от R_{21} .	11
4.13. Гистограмма объединённых данных при оптимальном значении R_{21} . . .	12

1. Введение

Постановка задачи Исследование из области солнечной энергетики. на рис.1 показана схема установки для исследования фотоэлектрических характеристик.

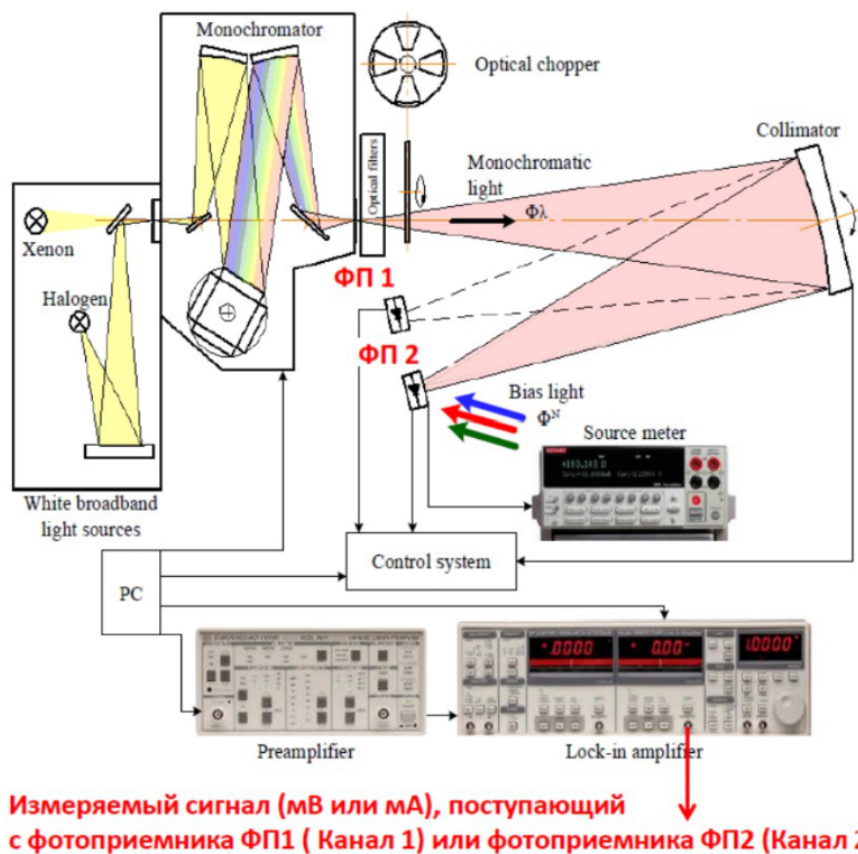


Рисунок 1.1. Схема установки для исследования фотоэлектрических характеристик

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

$$QE_{\text{ФП2}} = \frac{I_{\text{ФП2}}}{I_{\text{ФП1}}} * QE_{\text{ФП1}}$$

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

Исходные данные. Имеется 2 выборки данных с интервальной неопределённостью. Одна из них относится к эталонному датчику ФП2, другая - к исследуемому датчику ФП1.

Задача. Требуется определить коэффициент калибровки

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

2. Теория

2.1. Анализ данных с интервальной неопределённостью.

Представление данных. В первую очередь представим данные таким образом, чтобы применить понятия статистики данных с интервальной неопределённостью. Один из распространённых способов получения интервальных результатов в первичных измерениях - это "обинтерваливание" точечных значений, когда к точечному базовому значению \dot{x} , которое считается по показаниям измерительного прибора, прибавляется интервал погрешности ϵ :

$$x = \dot{x} + \epsilon$$

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 вектора $x = (x_1, x_2, \dots, x_n)$.

2.2. Линейная регрессия

2.2.1. Описание модели

Для оценки постоянной необходима предварительная обработка данных. Зададимся линейной моделью дрейфа.

$$I_{\Phi\Pi} = A + B * n, n = 1, 2, \dots, N$$

Поставим и решим задачу линейного программирования. Найдем коэффициенты А,В и вектор w множителей коррекции данных.

$$\sum_{i=1}^n |w_i| \rightarrow \min$$

$$-w_i * \epsilon \leq b_0 + b_1 * x_i - y_i, i = 1, \dots, n$$

$$-w_i * \epsilon \leq -b_0 - b_1 * x_i + y_i, i = 1, \dots, n$$

$$w_i \geq 1$$

Построим "спрямленные" данные выборки, вычтя из исходных данных "дрейфовую" компоненту.

$$I_{\Phi\Pi}^c = I_{\Phi\Pi} - B * n, n = 1, 2, \dots, N$$

2.2.2. Метод наименьших модулей

Для наиболее точного приближения данных с фотоприёмника y_i линейной регрессией $I_{\Phi\Pi}$ используется метод наименьших модулей:

$$\|f(x_i) - y_i\|_{l^1} \rightarrow \min, n = 1, 2, \dots, N$$

2.3. коэффициент Жаккара

Коэффициент Жаккара - мера сходства множеств. В интервальных данных рассматривается неоторая 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, коэффициента Жаккара может достигать значения только в интервале $[-1;1]$.

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

2.4. Процедура оптимизации

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

$$JK(I_1^c(n) * R \cup I_2^c(n)) \rightarrow \max$$

где $I_1^c I_2^c$ - 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

Лабораторная работа была сделана на jupyter notebook

4. Результаты

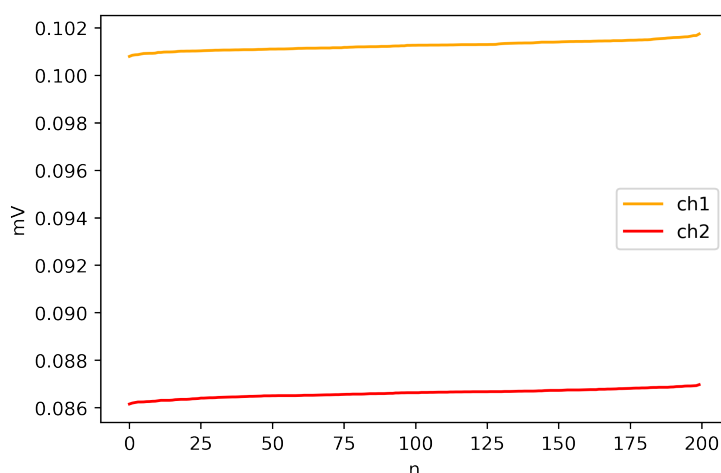


Рисунок 4.1. Исходные данные

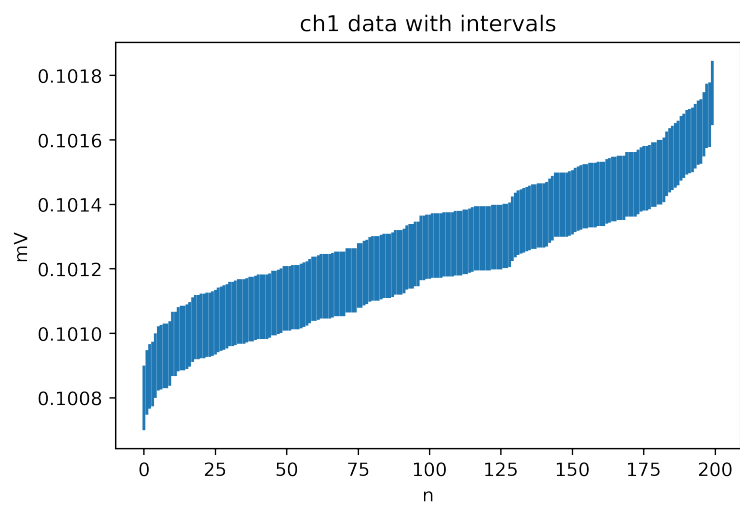


Рисунок 4.2. Интервальное представление исходных данных ch1

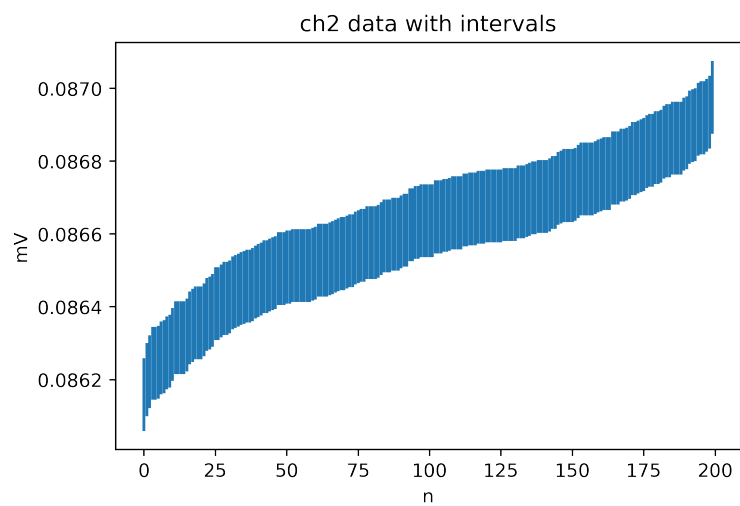


Рисунок 4.3. Интервальное представление исходных данных ch2

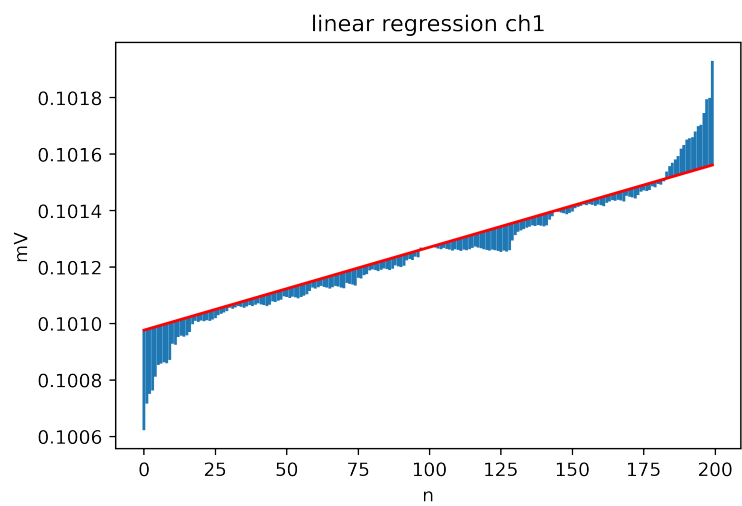


Рисунок 4.4. Линейная модель дрейфа данных ch1

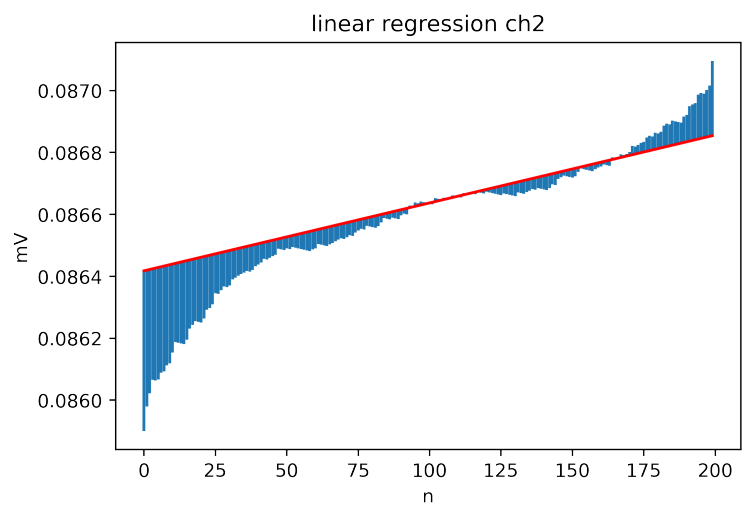


Рисунок 4.5. Линейная модель дрейфа данных ch2

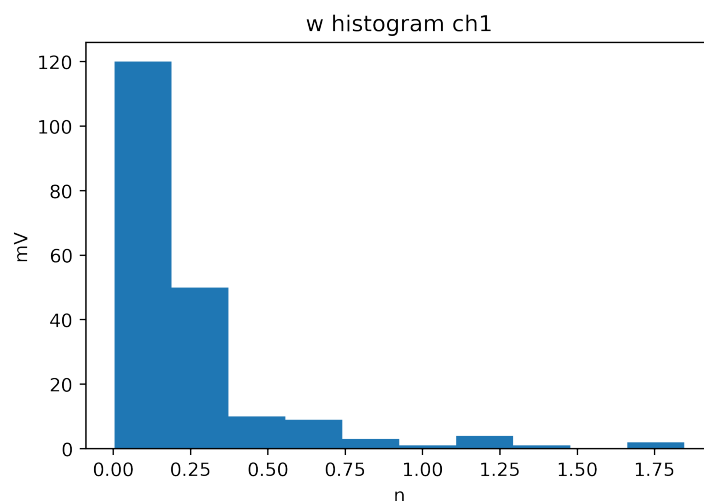


Рисунок 4.6. Гистограммы значений множителей коррекции w ch1

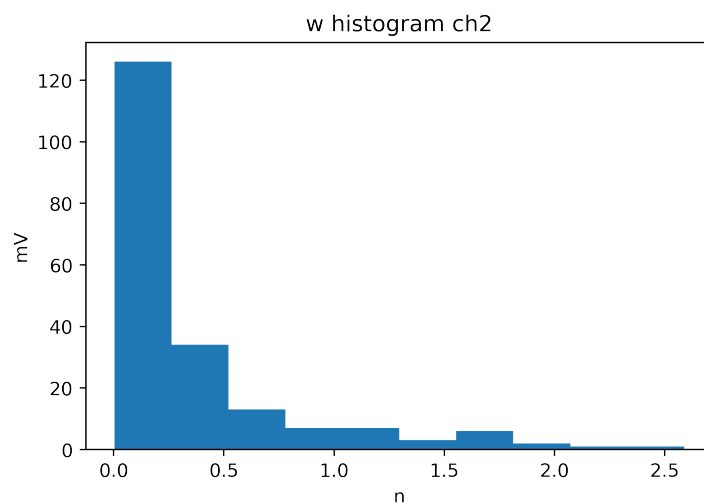


Рисунок 4.7. Гистограммы значений множителей коррекции w ch2

Результаты линейного приближения токов.

- Для первого фотоприемника:

$$B = 2.937 * 10^{-6}, A = 0.1010$$

- Для второго фотоприёмника:

$$B = 2.193 * 10^{-6}, A = 0.08642$$

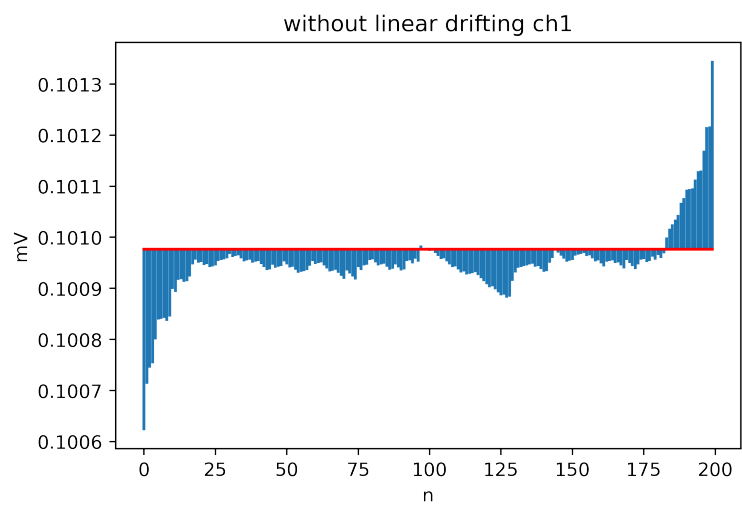


Рисунок 4.8. Скорректированная модель данных ch1

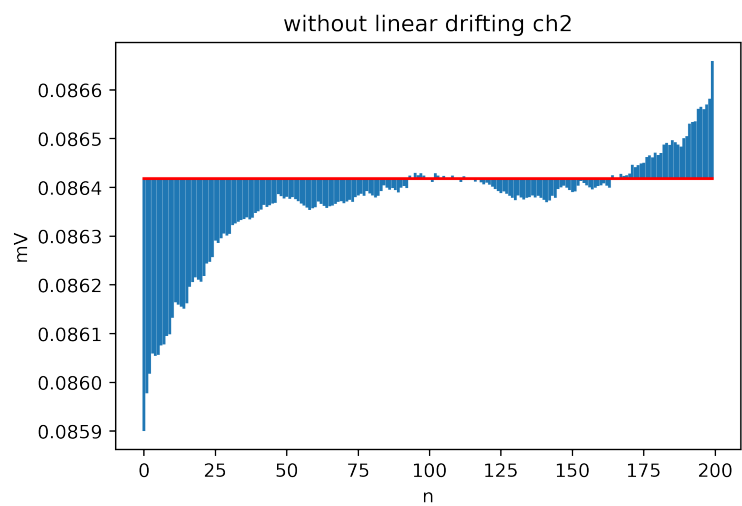


Рисунок 4.9. Скорректированная модель данных ch2

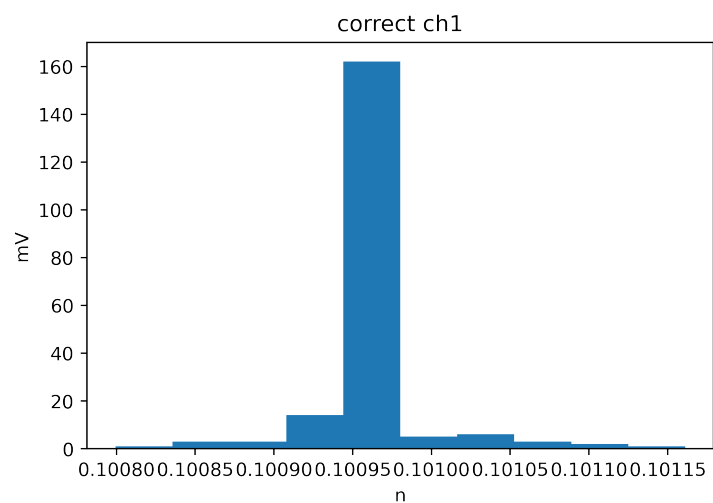


Рисунок 4.10. Гистограмма скорректированных данных ch1

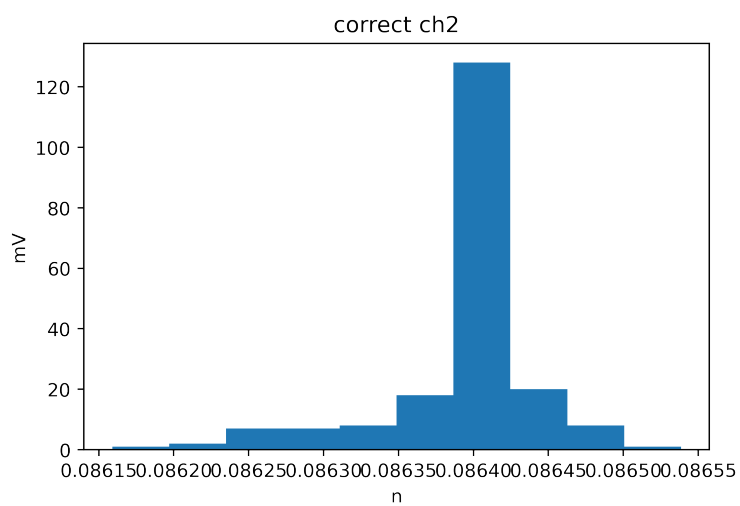


Рисунок 4.11. Гистограмма скорректированных данных ch2

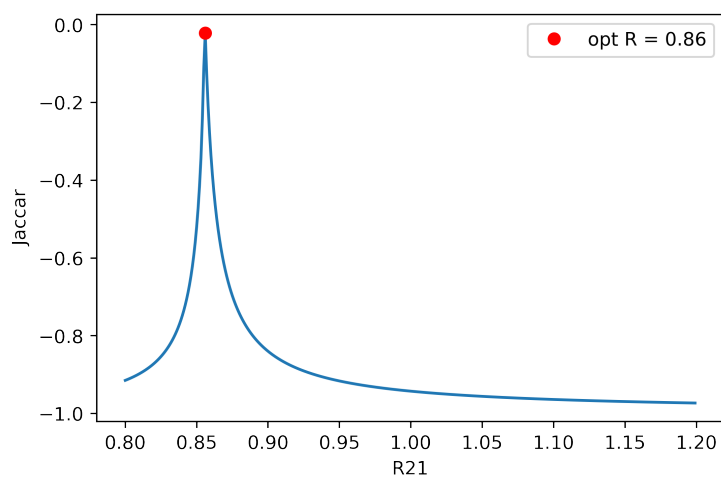


Рисунок 4.12. Значение коэффициента Жаккара от калибровочного множителя от R_{21}

Результаты исследования:

$$R_{opt} = 0.86, JK(0.87) = -0.021$$

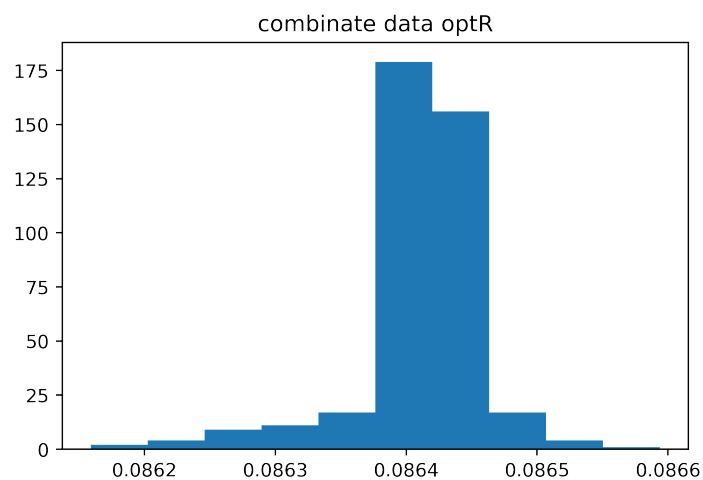


Рисунок 4.13. Гистограмма объединённых данных при оптимальном значении R_{21}

5. Литература

[1] А.Н. Баженов, С.И. Жилин, С.И.Куков, С.П.Шарый обработка и анализ интервальной неопределённостью.2022