

Санкт-Петербургский политехнический университет Петра Великого
Высшая школа прикладной математики и вычислительной физики
Кафедра прикладной математики

Курсовая работа
по дисциплине «Стохастические модели и анализ
данных»
на тему
Восстановление зависимостей

Выполнили студенты гр. 5040102/00201

Жуков А.К.

Грицаенко Н.Д.

Преподаватель

Баженов А.Н.

Санкт-Петербург

2021 год

Оглавление

Постановка задачи.....	3
Решение.....	4
Построение данных для регрессии с помощью интервальной моды.....	4
Параметры модели.....	5
Коридор совместных зависимостей	7
Граничные точки множества совместности	8
Прогноз за пределы интервала	8
Заключение	8
Приложение	9
Использованная литература	9

Постановка задачи

Дан регистратор, который может оцифровывать данные в определенном диапазоне $[-0.5, 0.5]$ V. Также есть данные размерности $1024 \times 8 \times 10 \times 10$, где 8 – число каналов, 10 – число измерений, 10 – число уровней.

Необходимо зафиксировать любой из каналов. Далее зафиксировать каждый уровень. Для каждого уровня получаются данные 10×1024 , то есть 10 измерений. Из каждого из 10 измерений нужно вырезать одинаковый интервал, затем склеить эти интервалы в один массив данных.

Возьмём интервал $[400, 600]$. Для каждого уровня получится $(600-400) \times 10 = 2000$ интервалов.

Далее для каждого массива из 2000 интервалов нужно найти *интервальную моду* [1].

Мода интервальной выборки – совокупность интервалов пересечения наибольших совместных подвыборок рассматриваемой выборки.

Таким образом, для 10 уровней получится 10 интервалов.

Для полученных интервалов нужно восстановить линейную зависимость с учётом интервальной неопределённости данных.

Модель данных будем искать в классе линейных функций:

$$y = \beta_1 + \beta_2 x$$

С неотрицательной первой производной: $\beta_2 > 0$

Ниже приведём графики исходных данных

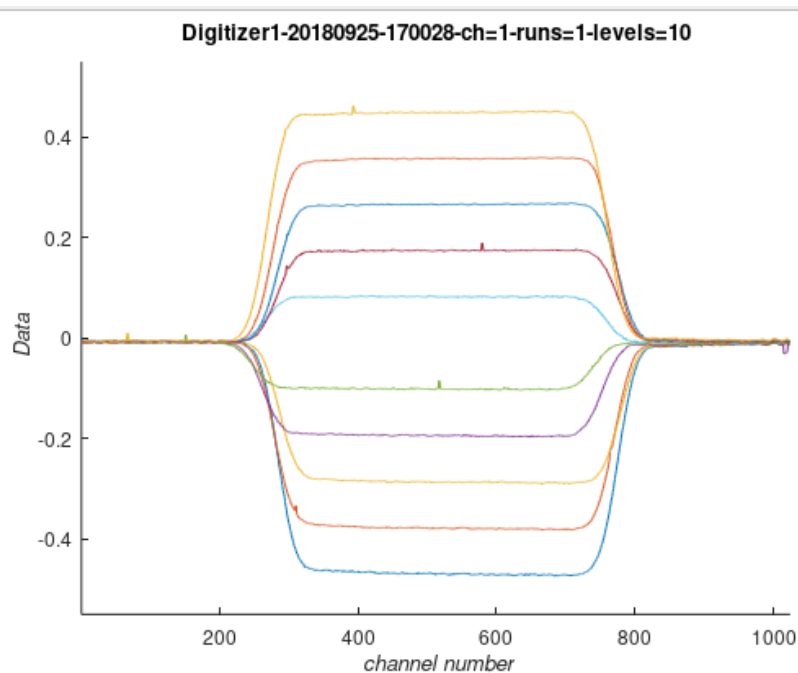


Рисунок 1 Исходные данные: 10 уровней

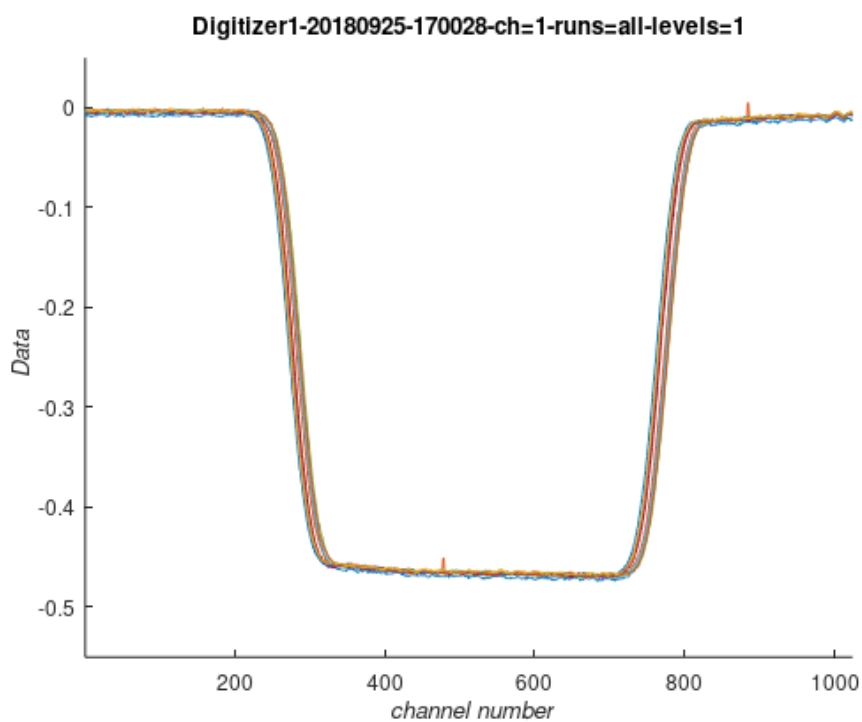


Рисунок 2. 10 измерений для одного уровня

Решение

Построение данных для регрессии с помощью интервальной моды

Соберем 2000 интервалов для первого уровня и построим интервальную моду:

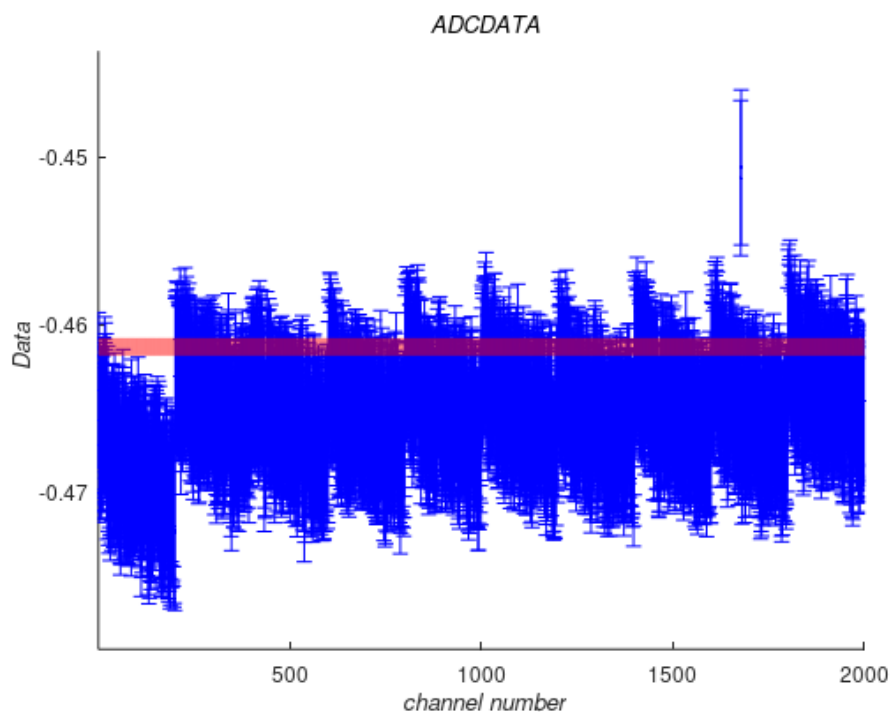


Рисунок 3. Интервальная мода для всех измерений 1 уровня 1 канала

Аналогично строим интервальную моду для остальных 9 уровней первого канала, получаем 10 интервальных мод:

Таблица 1. Интервальные моды для первого канала

Уровень	1	2	3	4	5	6	7	8	9	10
inf(mode)	-0.4659	-0.3756	-0.2837	-0.1907	-0.0991	0.0845	0.1764	0.2682	0.3593	0.4516
sup(mode)	-0.4567	-0.3670	-0.2744	-0.1807	-0.0890	0.0943	0.1859	0.2777	0.3687	0.4610

Приведём график получившихся интервальных мод:

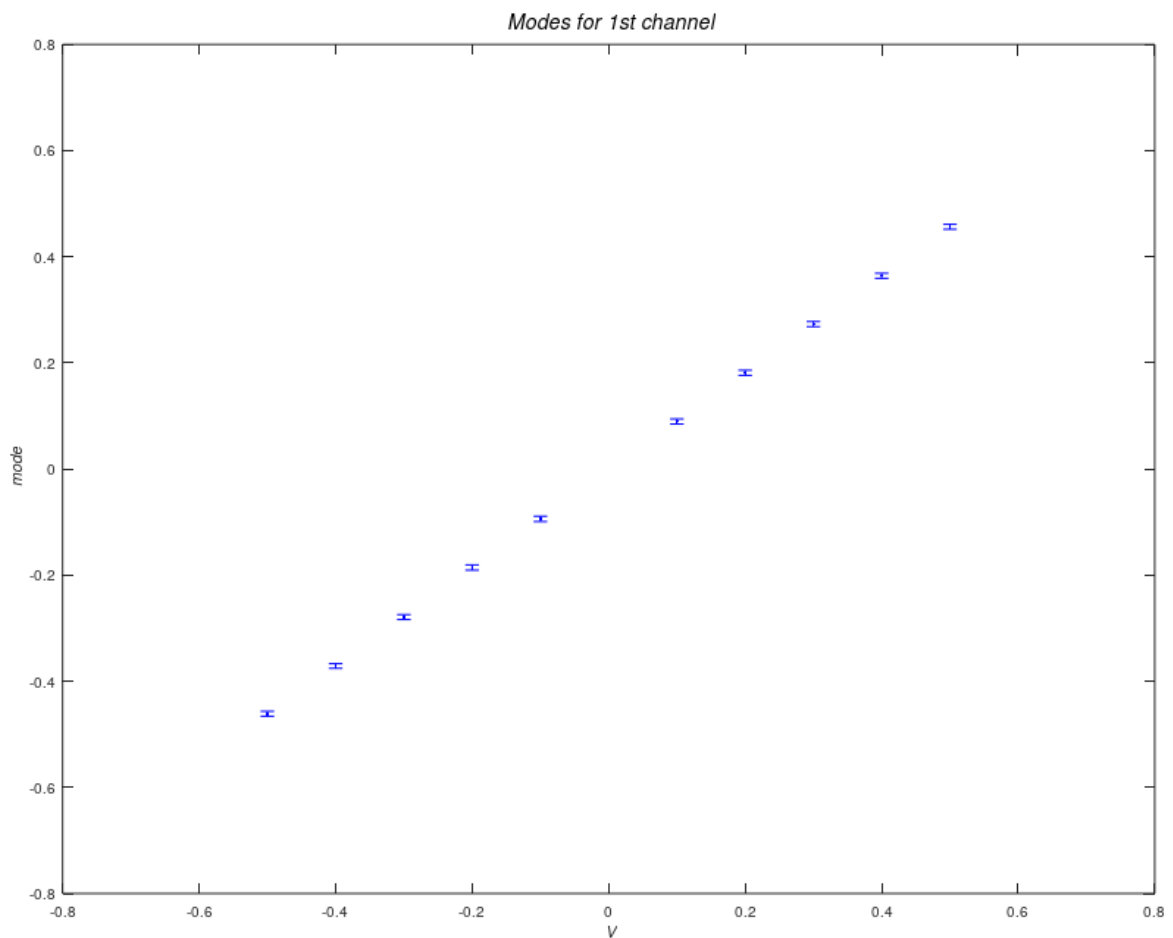


Рисунок 4. Интервальные моды для первого канала

Параметры модели

Поставим задачу оптимизации и решим её методом линейного программирования [\[1\]](#):

$$mid y_i - w_i \cdot rad y_i \leq X\beta \leq mid y_i + w_i \cdot rad y_i, \quad i = 1, m,$$

$$\sum_{i=1}^m w_i \rightarrow \min,$$

$$w_i \geq 0, \quad i = 1, m,$$

$$w, \beta = ?$$

Где $m = 10$, y_i – интервальные моды, X – матрица $m \times 2$, в первом столбце которой элементы равные 1, во втором – значения x_i (номера уровней).

Решение задачи оптимизации:

$$w = [1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0]$$

$$\beta = [0, 0.9175]$$

Построим график $y = \beta_1 + \beta_2 x$:

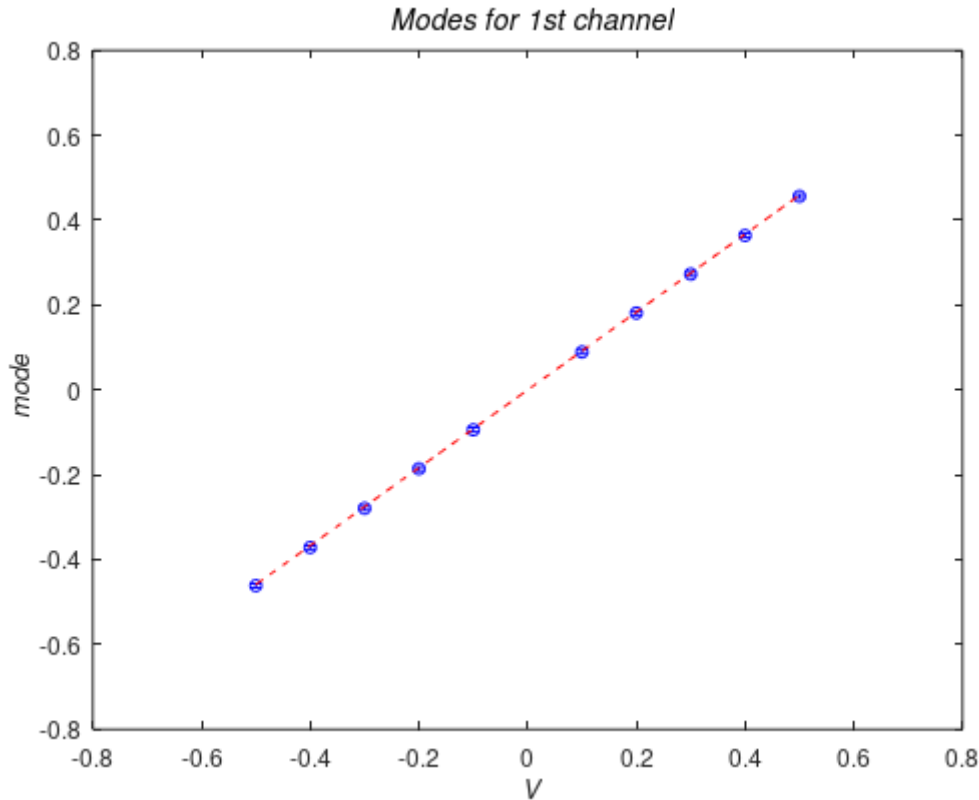


Рисунок 5. Решение задачи оптимизации ($y = 0 + 0.9175x$)

Построим информационное множество параметров модели. Поскольку информационное множество задачи построения линейной зависимости по интервальным данным задаётся системой линейных неравенств, то оно представляет собой выпуклый многогранник [2].

Обозначим на графике несколько точечных оценок:

- Центр наибольшей диагонали информационного множества:

$$\hat{\beta}_{\max \text{dig}} = \frac{1}{2}(b_1 - b_2),$$

где b_1 и b_2 – наиболее удалённые друг от друга вершины многогранника

- Центр тяжести информационного множества:

$$\hat{\beta}_{\text{gravity}} = \frac{1}{n} \sum_{i=1}^n b_i,$$

где b_i – вершина многогранника, n – их количество.

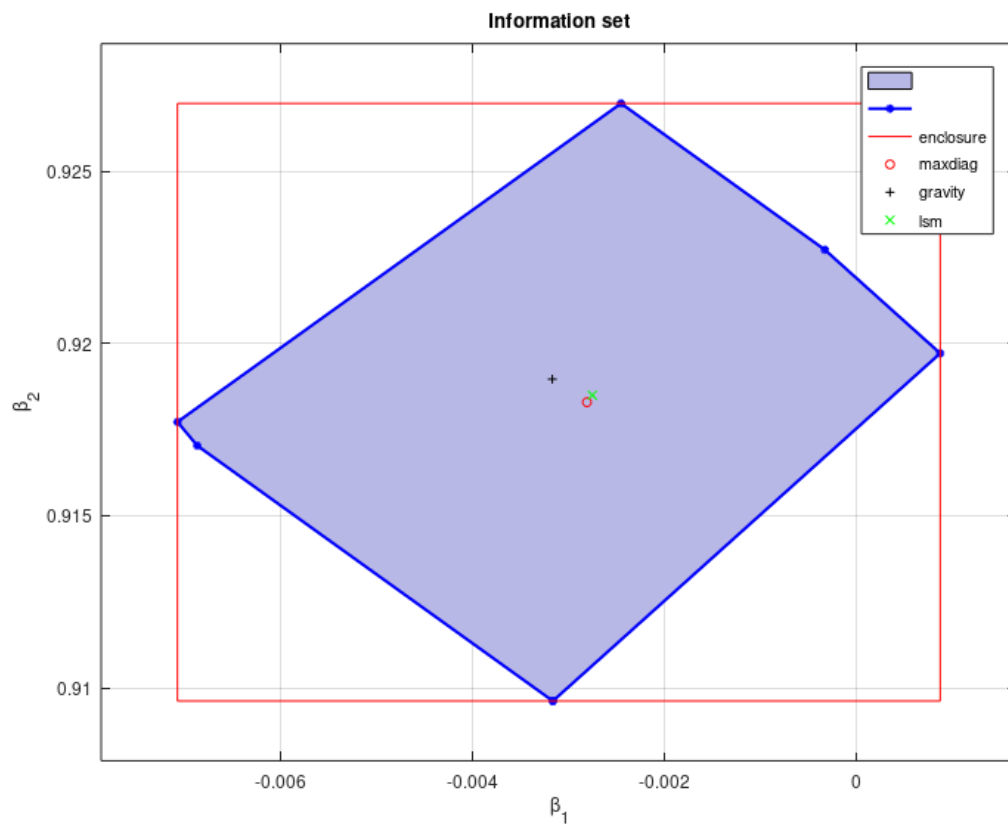


Рисунок 6. Информационное множество линейной модели

$$\beta_1 = [-0.0071, 0.00087]$$

$$\beta_2 = [0.90962, 0.92698]$$

Коридор совместных зависимостей

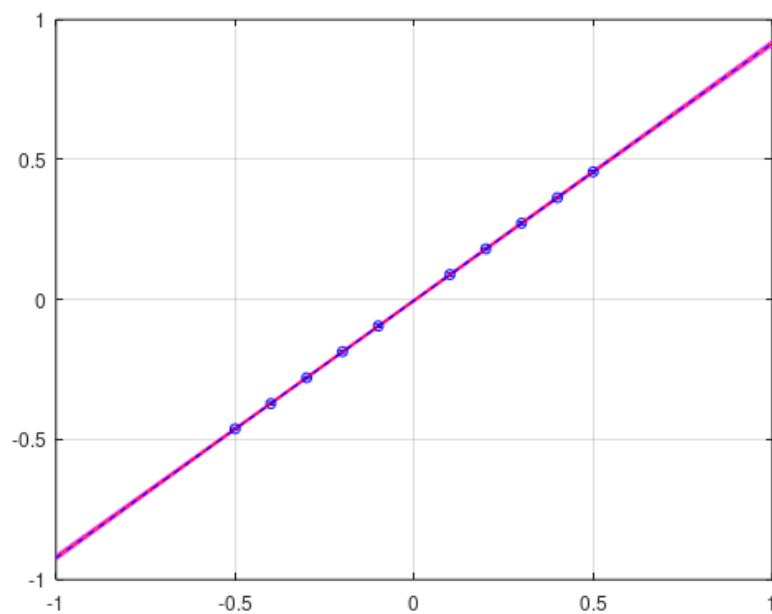


Рисунок 7 Коридор совместных зависимостей, весь диапазон

Рассмотрим подробнее, что происходит вокруг каждой точки:

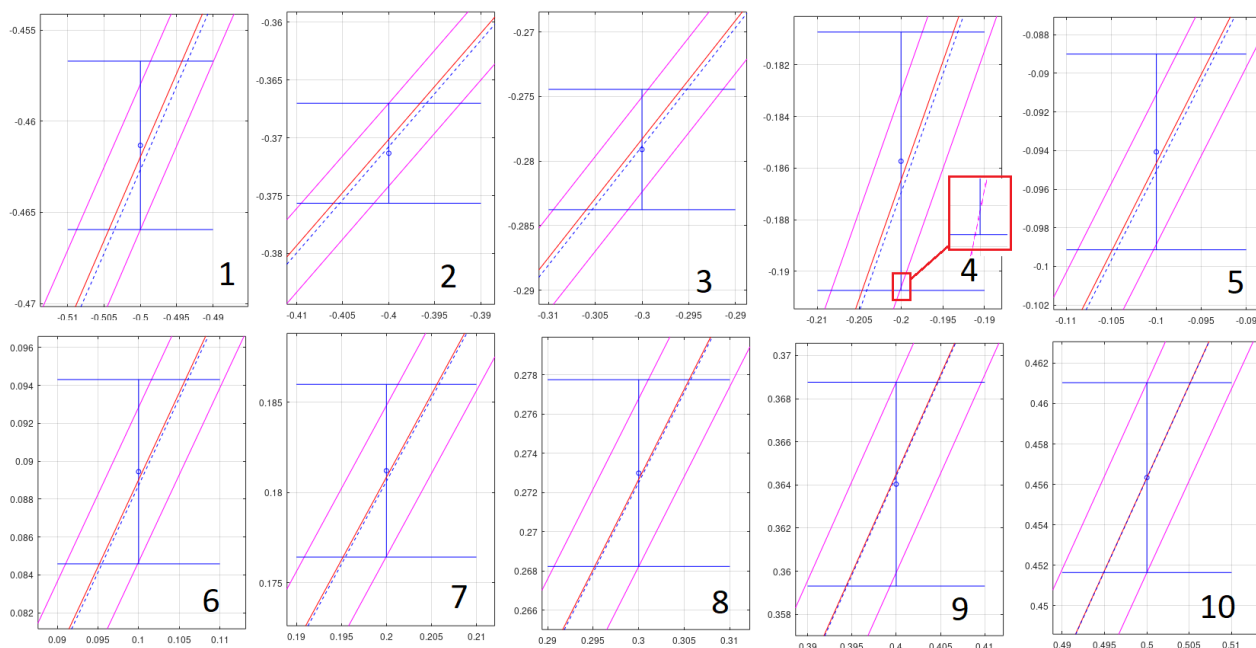


Рисунок 8 Коридор совместных событий в окрестности каждого наблюдения

Граничные точки множества совместности

По рисунку 8 видим, что граничными оказались точки с номерами 1, 2, 6, 7, 8, 9, 10.

Прогноз за пределы интервала:

С помощью построенной выше модели

$$\hat{y}(x) = [-0.0071, 0.00087] + [0.90962, 0.92698]x$$

Можно получить прогнозные значения выходной переменной:

Возьмём 5 точек:

$$x_p = [-0.75, -0.25, 0.25, 0.75, 5]$$

Тогда $y_p = \hat{y}(x_p)$

x_p	y_p	$rad\ y_p$
-0.75	[-0.70234, -0.68135]	0.0105
-0.25	[-0.23885, -0.22654]	0.0062
0.25	[0.22031, 0.23262]	0.0062
0.75	[0.67512, 0.69611]	0.0105
5	[4.5410, 4.6358]	0.0474

Неопределённость прогноза растёт по мере удаления от области, в которой производились исходные измерения. Это обусловлено видом коридора зависимости, расширяющимся за пределами области измерений.

Заключение

В ходе первой части работы с помощью вычисления интервальной моды нами были найдены данные для построения регрессии.

В ходе второй части работы была построена линейная модель данных, была сформирована и решена задача линейного программирования. Также было получено информационное множество для параметров линейной модели, построен коридор совместности и обнаружены граничные точки коридора совместности.

По полученной модели были вычислены прогнозы за пределами области измерений.

Приложение:

Ссылка на проект с кодом реализации:

<https://github.com/Nikitagritsaenko/Stochastic-models-and-data-analysis>

Использованная литература

1. А.Н. Баженов, С.И. Жилин, С.И. Кумков, С.П. Шарый. Обработка и анализ данных с интервальной неопределённостью. РХД. Серия «Интервальный анализ и его приложение». Ижевск. 2021. с.200.
2. А.Н. Баженов – Лекции по обработке данных с интервальной неопределённостью (2021) <https://cloud.mail.ru/public/rUwf/V8qPtjC1H>