

Санкт-Петербургский Политехнический университет  
Петра Великого  
Институт Прикладной Математики и Механики  
Кафедра «Прикладная Математика и Информатика»

Отчет  
По Курсовому Проекту  
По Дисциплине «Математическая статистика»

Выполнил:  
Студент Селянкин Федор  
Группа 3630102/70301  
Проверил:  
к.ф. – м.н., доцент  
Баженов Александр Николаевич

Санкт-Петербург  
2020 г.

## СОДЕРЖАНИЕ

Постановка задачи.....	3
Этапы работы.....	3
Реализация.....	3
Ход работы.....	3
Отбор подходящих сигналов и датчиков.....	3
Программное и ручное уточнение интересующего участка с колебаниями. ....	4
Выяснение подобию.....	4
Подобие в ряду .....	4
Подобие сигналов разных разрядов.....	6
Обсуждения.....	8
Ссылки .....	8
<b>Список Таблиц:</b>	
Таблица 1 Используемые данные.....	3
Таблица 2 Матрица корреляций частот различных сигналов для разряда 38515.....	4
Таблица 3 Корреляционная матрица для горизонтального сравнения .....	6
<b>Список Иллюстраций:</b>	
Рисунок 1 Совместный график частот для sht38515 .....	5
Рисунок 2 Совместная гистограмма частот для sht38515 .....	5
Рисунок 3 Совместный график частот для 20го сигнала .....	7
Рисунок 4 Совместная гистограмма частот для 20го сигнала.....	7

## Постановка задачи

Обработать группу разрядов и выяснить их подобие: построить гистограммы и найти матрицу корреляций частот. Провести исследования как среди различных сигналов одного разряда, так и для одного разряда различных сигналов. Выяснить какое средство является наиболее удобным для изучения «похожести» данных. Описать плюсы и минусы матрицы корреляций и гистограммы для этого исследования.

## Этапы работы

1. Отбор подходящих сигналов и датчиков.  
Не на всех графиках даже визуально можно обнаружить пилообразные выбросы, либо определение участков интереса или их исследование слишком трудоемкая задача.
2. Программное и ручное уточнение интересующего участка с колебаниями.  
В качестве начальных значений интервалов брались числа, полученные с помощью запуска программы. Далее шло ручное уточнение промежутков по следующим критериям:
  - 1) Интервал должен содержать по минимуму горизонтальный участок.
  - 2) Выходные векторы частот должны иметь одинаковую длину, для построения матрицы ковариации.
3. Выяснение подобия  
Подобие можно определить как вручную с помощью визуального сравнения исходных графиков напряжения, так и конечных графиков частот.  
Для получения количественной характеристики будем пользоваться матрицей корреляций. Дополнительно будет строить и сравнивать между собой гистограммы.

## Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm, с использованием дополнительных библиотек для отображения и расчетов. Также использовался исходный код с ресурса и скомпилированная с помощью этого кода библиотека `pyglobus` [1]. Исходный код лабораторной выложен на веб-сервисе GitHub [2].

## Ход работы

### Отбор подходящих сигналов и датчиков

Большинство данных было отбраковано кроме следующих наборов.

Разряд\сигналы	18	19	20	26
Sht38515	Да	Да	Да	Да
Sht38516	Нет	Да	Да	Да
Sht38867	Нет	Нет	Да	Нет
Sht38873	Нет	Нет	Да	Нет
Sht38875	Нет	Нет	Да	Нет
Sht38876	Нет	Нет	Да	Нет
Sht38892	Нет	Нет	Да	Нет

Таблица 1 Используемые данные

Данные комбинации позволяют сравнить как сигналы в разных разрядах, так и сигналы в одном разряде. Далее будет говорить о сравнении по вертикали для сравнения характеристик для 20го сигнала различных разрядов и сравнении по горизонтали для сравнения по разрядам sht38515.

Также были отбракованы случаи, когда пилообразные колебания наблюдаются визуально, но область интереса крайне мала, и предоставленный алгоритм выдачи частот выдает слишком короткий, относительно остальных результатов, вектор

#### Программное и ручное уточнение интересующего участка с колебаниями.

Некоторые случаи, помимо сложно определяемых границ участков интереса, имеют так же выбросы (сильные колебания значений). Например, график разряд 38515 сигнал 19. Так как область интереса для всех графиков в горизонтали совпадает, то было решено игнорировать участки выброса для всех случаев, таким образом учтя выброс при вычислении схожести графиков частот

#### Выяснение подобия

##### Подобие в ряду

Для демонстрации подобия по горизонтали используем разряд 38515 сигналы 18, 19, 20, 26

Для частот этих сигналов была построена корреляционная матрица

	18	19	20	26
18	1.000000	0.795036	0.982749	0.991681
19	0.795036	1.000000	0.800807	0.819166
20	0.980522	0.800807	1.000000	0.987105
26	0.991681	0.819166	0.987105	1.000000

Таблица 2 Матрица корреляций частот различных сигналов для разряда 38515

совместный график частот

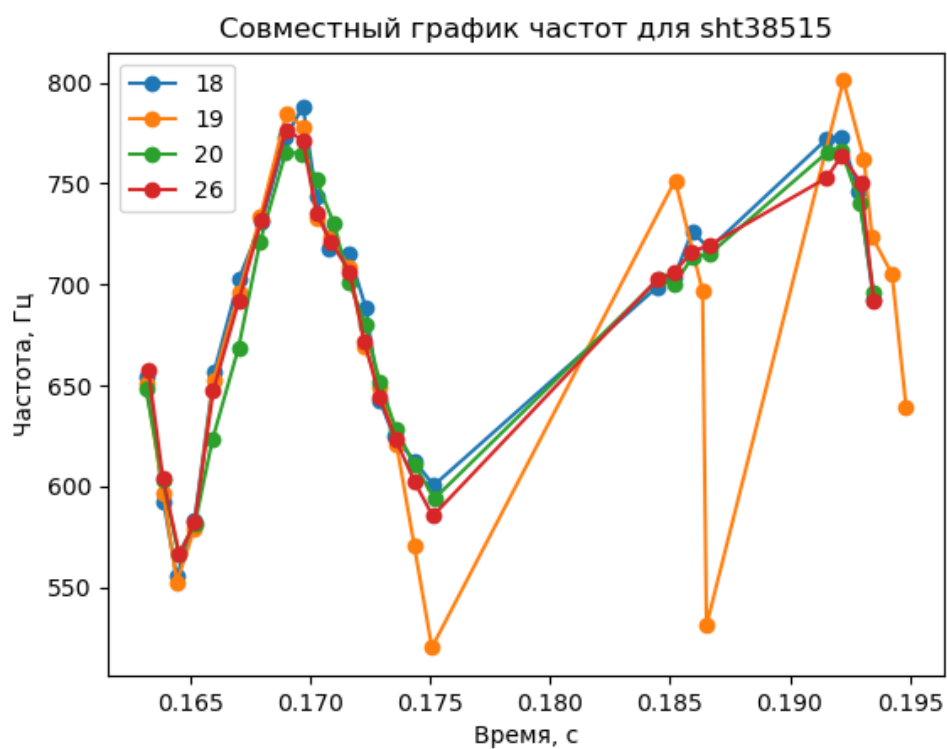


Рисунок 1 Совместный график частот для sht38515

и гистограмма распределения частот

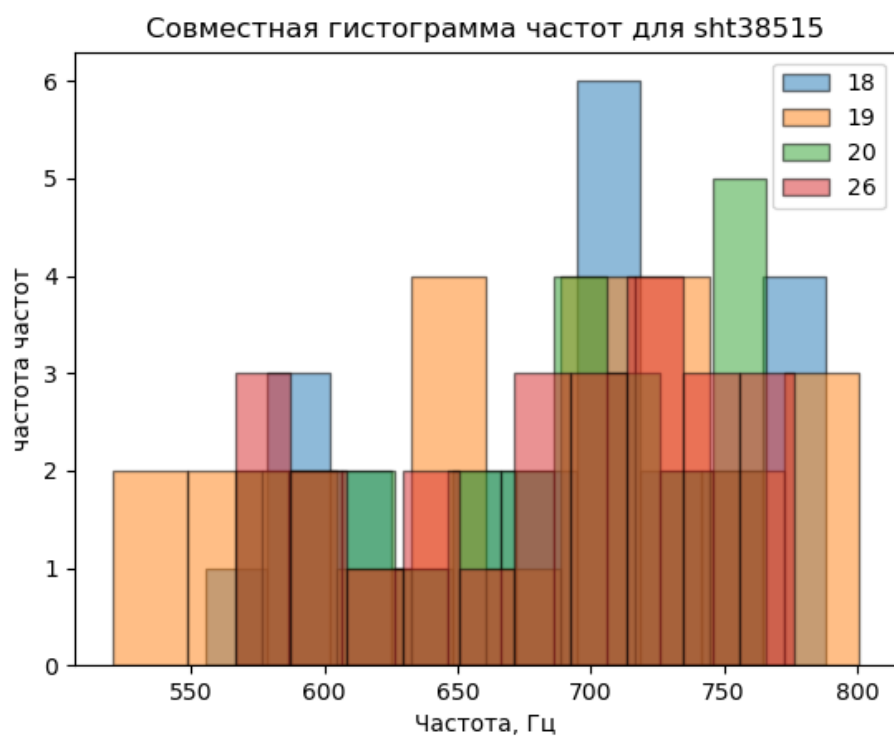


Рисунок 2 Совместная гистограмма частот для sht38515

Из этих представлений можно сделать следующие вывод о довольно высокой степени похожести распределения частот. На гистограмме видна область в районе 700-800 Гц, в которой сконцентрировано наибольшее число значений.

Выброс довольно существенно влияет на «похожесть» данных. На гистограмме столбцы соответствующие 19му сигналу, имеют два пика ( все остальные имеют по одному) и захватывают намного больше значений. На графике частот видно насколько сильно отклоняется 19ый сигнал от остальных в разряде. Также, матрица корреляций в столбцах и строках, относящихся к 19му сигналу, имеет значения намного ниже, чем остальные столбцы.

Таким образом, если бы имели только один или два массива данных ставился бы вопрос о достоверности обоих из них. Но имея уже 4 набора данных уже можно говорить о достоверности и правдоподобности некоторых из них. Благодаря гистограмме и графику частот можно визуальное оценить «выбивающийся» набор данных, а с помощью матрицы корреляций можно задать некоторое пороговое значение расхождения данных и, если какая-либо строка или столбец имеют значения меньше чем остальные, можно сделать предположения о его неправдоподобности.

Если бы были использованы данные до фильтрации их выбросов, то расхождения были бы еще сильнее.

#### Подобие сигналов разных разрядов

Для данного исследования был использован 20 сигнал разрядов 38515, 38516, 18873, 38875, 38876, 38892.

Для частот этих сигналов была построена корреляционная матрица

	sht38515	sht38516	sht38873	sht38875	sht38876	sht38892
sht38515	1.000000	0.286882	-0.474476	-0.255016	-0.318144	-0.586844
sht38516	0.286882	1.000000	-0.384219	-0.871594	-0.901507	-0.589096
sht38873	-0.474476	-0.384219	1.000000	0.630942	0.597008	0.311612
sht38875	-0.255016	-0.871594	0.630942	1.000000	0.800183	0.552075
sht38876	-0.318144	-0.901507	0.597008	0.800183	1.000000	0.440868
sht38892	-0.586844	-0.589096	0.311612	0.552075	0.440868	1.000000

Таблица 3 Корреляционная матрица для горизонтального сравнения

совместный график частот

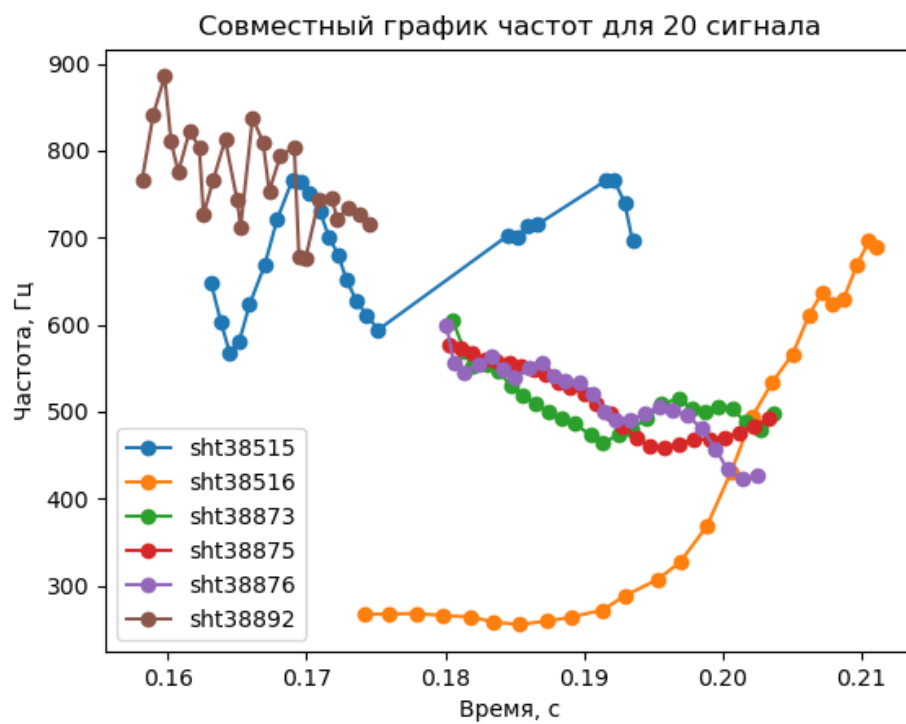


Рисунок 3 Совместный график частот для 20го сигнала

и гистограмма распределения частот

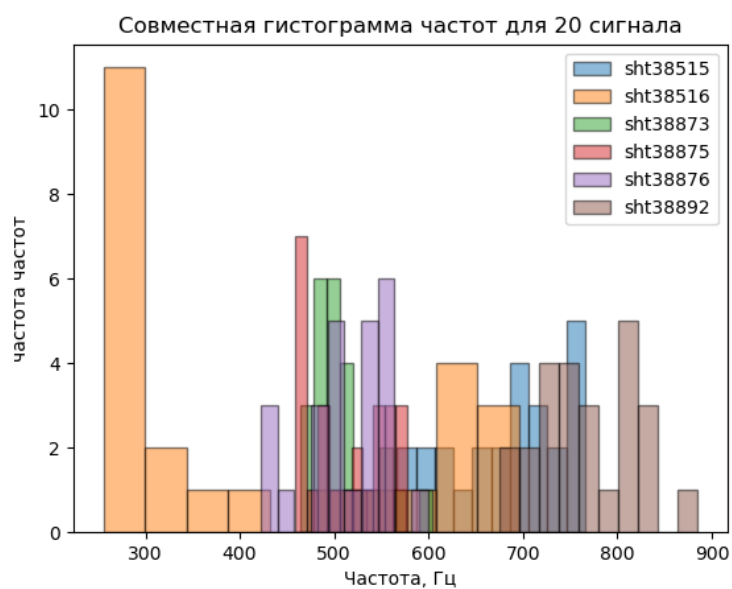


Рисунок 4 Совместная гистограмма частот для 20го сигнала

Из данных представлений можно сделать следующий выводы.

Из совместной гистограммы можно получить информацию о распределении частот среди различных сигналов. Видно, что наибольшую плотность от примерно 450 до 600. На этом участке перекрывается наибольшее число частот сигналов. Из графика частот эту информацию так же можно получить, но там это менее наглядно из-за расхождения временных отрезков сигналов.

В то же время из совместного графика частот быстрее можно обнаружить близкие сигналы, например сигналы sht38873, sht38875 и sht38876 близки как по одной, так и по другой оси.

По матрице корреляций можно сделать вывод как о близости некоторых сигналов, но в основном значения близки к нулю или вообще отрицательные, что говорит о том, что поступающие данные довольно различные. Тем не менее по матрице корреляций можно увидеть, что данные sht38515 и sht38516 немного похожи и действительно, если посмотреть на окончания линий соответствующим этим графикам, то можно заметить некоторое подобие, к тому же оба графика возрастают.

Положительные значения ячеек соответствующим корреляциям данных sht38873, sht38875, sht38876 не удивительны, так как они почти совпадают на графиках, но с этими данными также имеет положительную корреляцию sht38892. И действительно. Поведение этих графиков (в среднем все графики убывают) довольно похоже. Несмотря на то, что график sht38892 лежит намного выше. Сравнив с оставшимися двумя наборами (sht38515, sht38516) и еще раз изучив таблицу можно сделать следующий вывод: по матрице корреляции можно сделать предположение, что графики либо вместе растут/убывают, если коэффициент корреляции положительный или он движется в разные стороны.

## Обсуждения

Матрица корреляций является наиболее удобным инструментом для анализа подобия данных. С её помощью можно судить как о степени похожести, так и о взаимном поведении данных. Минусом является то, что для ее использования вектора данных должны обладать одинаковой длиной, что может быть проблемой, так как может потребоваться отсечение полезных данных либо добавления искусственных данных.

Совместная гистограмма позволяет быстро визуально понять какие значения принимаются чаще всего. Так же позволяет понять границы распределения для всех предоставленных данных. Минусом может служить то, что при большом объеме данных теряется простота визуального анализа.

И гистограмма, и матрица удобно программируются, что позволяет, например задавать критерий подобия довольно просто. Например, для матрицы это просто значение в ячейке. В случае гистограммы можно говорить только о сходстве принимаемых значений, так как характер линий она передать не может.

Совместный график слабо подходит для оценки подобия, кроме разве, что очевидных случаев. Тем не менее, его полезно использовать для подтверждения гипотез, полученных при анализе гистограммы и матрицы.

## Ссылки

GitHub репозиторий с pyglobus. (б.д.). Получено из github.com: <https://github.com/dev0x13/globus-plasma>

GitHub с исходным кодом курсового проекта. (б.д.). Получено из github.com: <https://github.com/SelyankinFyodor/math-statistics>