

Санкт-Петербургский политехнический университет  
Петра Великого

Институт прикладной математики и механики  
Высшая школа прикладной математики и вычислительной физики

**Отчёт**  
**по курсовой работе**  
**по дисциплине**  
**«Математическая статистика»**

Выполнил студент:

Кондратьев Д. А.  
группа: 3630102/70301

Проверил:

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

Санкт-Петербург  
2020 г.

## Содержание

<b>1. Постановка задачи</b>	<b>2</b>
<b>2. Теория</b>	<b>3</b>
2.1. Используемые параметры . . . . .	3
2.2. Подготовка данных . . . . .	3
<b>3. Реализация</b>	<b>4</b>
<b>4. Результаты</b>	<b>5</b>
<b>5. Обсуждение</b>	<b>6</b>
<b>6. Литература</b>	<b>6</b>
<b>7. Приложение</b>	<b>6</b>

## Список таблиц

1	Таблица интенсивностей . . . . .	3
---	----------------------------------	---

## Список иллюстраций

1	Обработанный файл с выделенными областями . . . . .	4
2	Двумерные поля Африки и севера России . . . . .	5
3	Совместное двумерное поле Африки и севера России . . . . .	5

## 1. Постановка задачи

Есть набор 2D данных в текстовом формате – следы жизни в геологических объектах. Образцы взяты с двух разных регионов:

- русского севера;
- центральной Африки.

На объект подавалось излучение от ближнего ультрафиолетового до видимого. Длина волны — первая переменная  $x_1$ .

Когда свет с заданной  $x_1$  попадал в объект, его поглощали молекулы и в свою очередь, излучали свет с длинами волны  $x_2$  примерно в том же диапазоне.

То, что они излучали записывается в виде графика  $I(x_1 = \text{const}, x_2)$ .

Далее,  $x_1$  варьируются, и формируется  $I(x_1, x_2)$ . Функция 2-х переменных.

Пики на графике  $I$  можно идентифицировать с излучением протеиногенных аминокислот, т.е. это остатки органической жизни.

Известна область для каждой аминокислоты в координатах  $(x_1, x_2)$ .

Для классификации двух типов данных ранее был предложен параметр  $K$  (на основании С, А, В, Т), который позволяет достаточно уверенно проводить разделение этих типов. При этом не использовались данные по переменной  $M$ .

Необходимо:

- построить двумерное поле  $(M, K)$  для Африки и Арктики;
- проанализировать полученный результат.

## 2. Теория

### 2.1. Используемые параметры

Буквенное обозначение	Тип компонента	$E_{x_{max}}(nm)$	$E_{m_{max}}(nm)$
C	Humic-like	320 – 350	420 – 480
A	Humic-like	250 – 260	380 – 480
M	Mariane Humic-like	310 – 320	380 – 420
B	Tysone-like, Protein-like	270 – 280	300 – 320
T	Tryptophane-like, Protein-like or phenol-like	270 – 280	320 – 350

Таблица 1. Таблица интенсивностей

Параметр  $K$ , ранее использовавшийся для сравнения данных, выражает отношение сложной и простой органики и вычисляется по формуле:

$$K = \frac{C + A}{B + T} \quad (1)$$

### 2.2. Подготовка данных

По полученным данным были получены изображения, далее были обрезаны релеевские облучения и выделены области в соответствии с таблицей интенсивностей [1]. После были посчитаны суммарные интенсивности каждой области, по которым в дальнейшем будет вестись исследование.

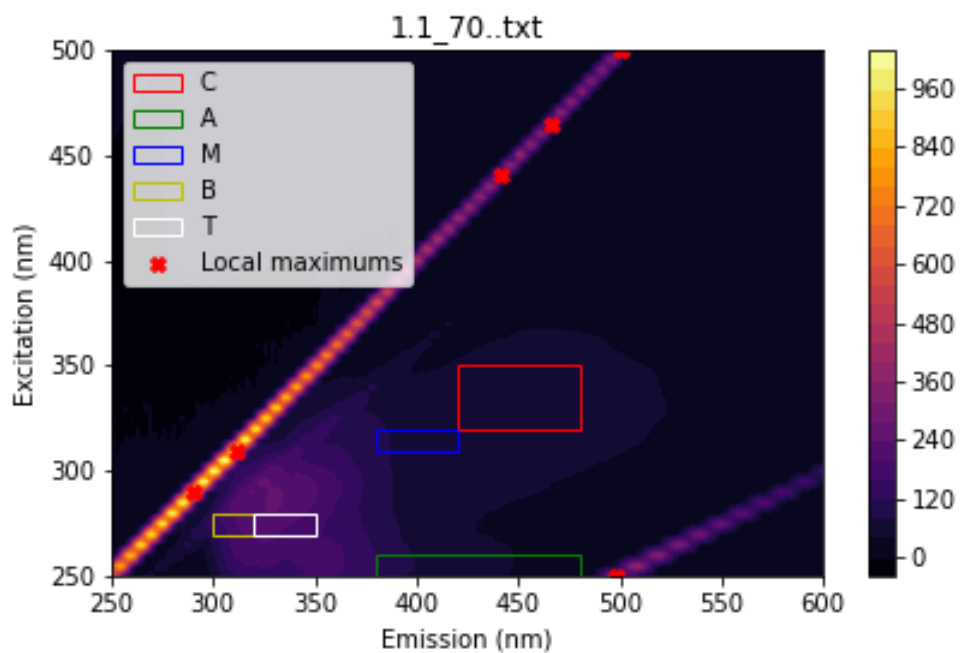


Рис. 1. Обработанный файл с выделенными областями

### 3. Реализация

Курсовая работа выполнена на программном языке *Python 3.8* в среде разработки *Jupyter Notebook 6.0.3*. В работе использовались следующие пакеты языка *Python*:

- *numpy* — для обработки исходных данных и работы с массивами;
- *matplotlib* — для визуализации результатов.

Ссылка на исходный код курсовой работы приведена в приложении.

## 4. Результаты

Исходя из полученных данных  $(M, K)$  были построены следующие двумерные поля:

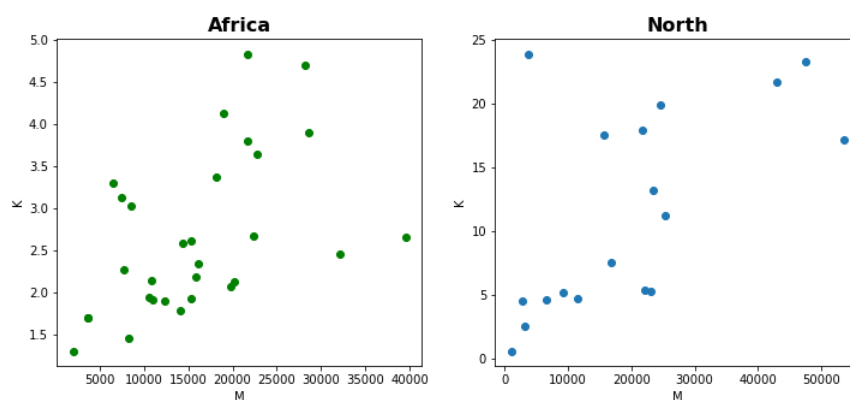


Рис. 2. Двумерные поля Африки и севера России

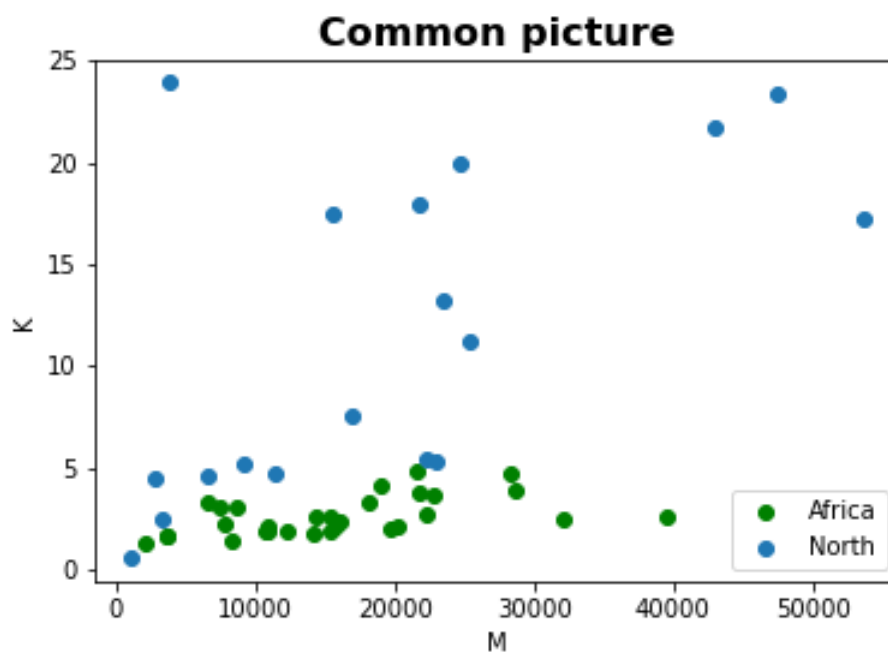


Рис. 3. Совместное двумерное поле Африки и севера России

## 5. Обсуждение

Исходя из полученных результатов можно сделать следующие выводы:

- Для севера России характерен большой разброс точек по сравнению с Африкой, где они расположены более кучно.
- Для почти всех координат севера России справедливо утверждение, что при увеличении  $M$   $K$  возрастает.
- Для определения области лучше опираться на параметр  $K$ , так как он дает более четкую картину поведения аминокислот, характерную для конкретную область.
- Также, используя данные поля, можно поставить задачу классификации и построить алгоритм, определяющий область по координатам  $(M, K)$ .

## 6. Литература

- 1) Документация *numpy*. URL: <https://numpy.org/doc/stable/reference/>
- 2) Документация *matplotlib*. URL: <https://matplotlib.org/3.2.1/contents.html>

## 7. Приложение

- 1) Код лабораторной. URL: [https://github.com/DmitriiKondratev/MatStat/blob/master/Course\\_work/Course\\_work.ipynb](https://github.com/DmitriiKondratev/MatStat/blob/master/Course_work/Course_work.ipynb)
- 2) Код отчёта. URL: [https://github.com/DmitriiKondratev/MatStat/blob/master/Course\\_work/Course\\_work\\_report.tex](https://github.com/DmitriiKondratev/MatStat/blob/master/Course_work/Course_work_report.tex)