

## **Презентация бакалаврской работы по теме**

**«Разработка сервиса интерполяции и  
экстраполяции данных»**

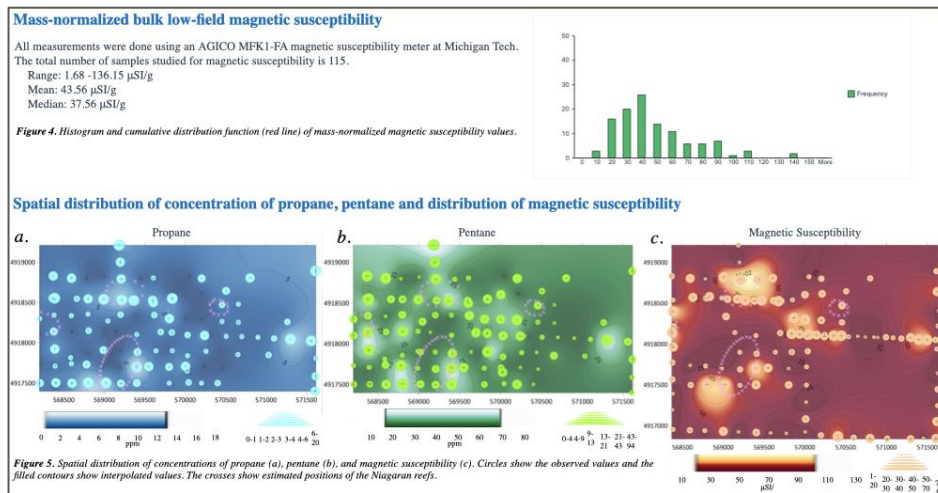
---

**Автор: Попов А.В., студент группы М34351**

**Руководитель: Лукин М.А., к.т.н.**

В геостатистике для анализа данных очень часто используется пространственная интерполяция и экстраполяция — предсказание значений в неизвестных объектах по набору данных.

Эти задачи решаются при помощи методов интерполяции. В зависимости от области, на которой применен метод, вычисляется интерполяция или экстраполяция набора данных.



# Сравнение инструментов

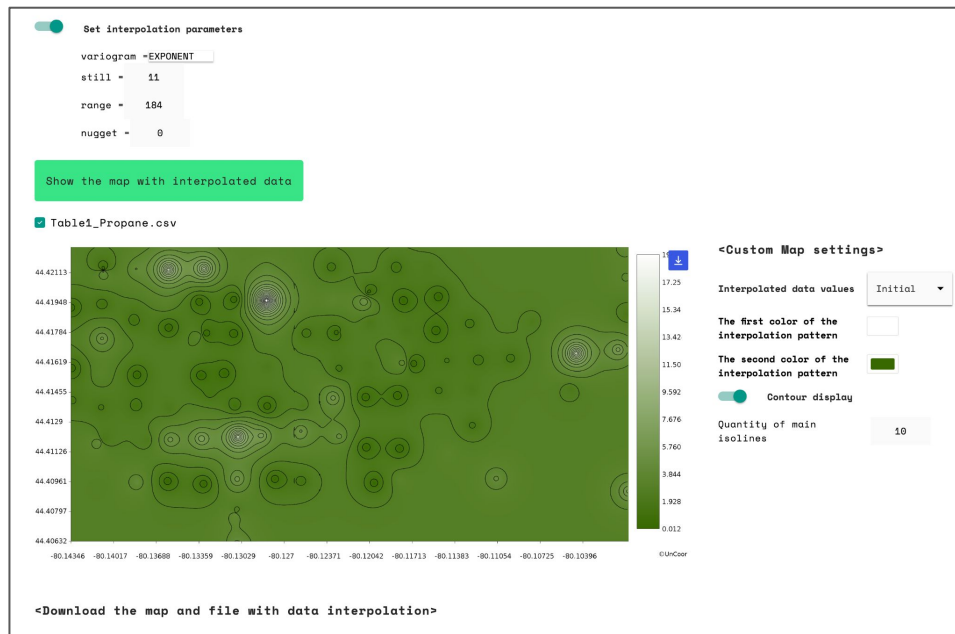
Все инструменты для решения этой задачи имеют ограниченный набор функций, требуют знаний программирования, сложной настройки или являются платными.

Как удобная бесплатная альтернатива им был создан сервис Uncorr.

|                                    | <i>Surfer</i> | Методы программирования, <i>Microsoft Excel</i> / <i>Google Spreadsheets</i> | <u><i>Uncorr 1.0</i></u> | Другие онлайн сервисы |
|------------------------------------|---------------|--|--------------------------|-----------------------|
| Количество методов                 | >10           | >10  | 3                        | 1                     |
| Возможности настройки              | Да            | Да   | Да                       | Нет                   |
| Качественные графики               | Да            | Да   | Нет                      | Нет                   |
| Поддерживает большие наборы данных | Да            | Да   | Нет                      | Нет                   |
| Платное                            | Да            | Нет  | Нет                      | Нет                   |
| Сложность использования            | Средняя       | Высокая  | Низкая                   | Средняя               |

## Недостатки Uncorr 1.0:

- Мало методов интерполяции.
- Артефакты в графиках.
- Только маленькие наборы данных.
- Внешний вид графиков не подходит для научных работ.
- Нестандартный для компании набор технологий.



Разработать веб-приложение, в котором будет:

- Весь функционал предыдущей версии.
- **Основные методы** интерполяции.
- Возможность обрабатывать **большие наборы данных** (до миллиона точек) за **ограниченное время** (до шестидесяти секунд на запрос).
- **Качество графиков и возможности их обработки** сравнимы с программой Surfer от Golden Software.
- Создать инструмент для финальной обработки диаграмм, полученных в сервисе forctool.com. Проработать возможность интеграции.

Кроме того, приложение должно быть реализовано с использованием **стандартных для компании Sudo технологий**.

1. Разработка архитектуры приложения.
2. Разработка и реализация библиотеки с методами интерполяции.
3. Разработка и реализация серверной части.
4. Разработка и реализация клиентской части.
5. Добавление возможности интеграции с другими сервисами.

# 1. Разработка архитектуры

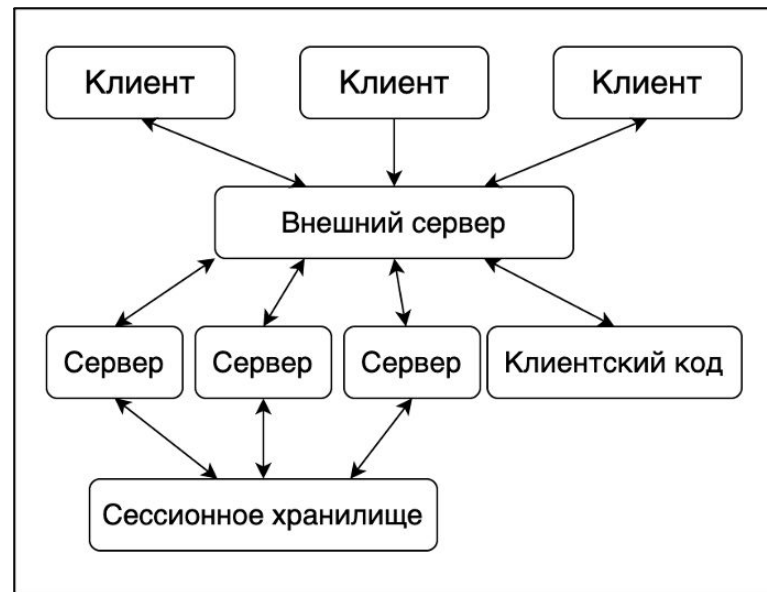
Приложение состоит из модулей, помещенных в Docker-контейнеры:

- **Серверная часть:** python + fastAPI + gstools + libs
- **База данных для хранения серверных сессий:** Redis
- **Клиентская часть:** typescript + reduxToolkit + react + plotly.js + webpack
- **Внешний сервер:** Nginx

Решена инженерная проблема — **горизонтальное масштабирование**, благодаря тому, что:

- модули стали более изолированы друг от друга.
- Используемые библиотеки **уменьшили требования на каждый сервер**.

Это важно, потому что обработка запроса иногда занимает до десятков секунд.



## 2. Разработка библиотеки

Стандартизированы оптимальные алгоритмы из популярных библиотек.

Метод IDW реализован самостоятельно, потому что не была найдена подходящая под задачу реализация.

Доступны параметризуемые методы интерполяции:

- Inverse Distance Weighting
- Linear
- Nearest neighbour
- Radial basis function
- Kriging: simple, ordinary и universal

Больше информации можно найти здесь:

<https://pypi.org/project/sdinterp/>

| Метод             | Время работы, с |
|-------------------|-----------------|
| IDW               | $10,0 \pm 0,5$  |
| Linear            | $8,3 \pm 0,1$   |
| Nearest Neighbour | $3,0 \pm 0,1$   |
| RBF               | $26,0 \pm 1,0$  |
| Simple Kriging    | $33,2 \pm 2,0$  |
| Ordinary Kriging  | $37,0 \pm 2,0$  |
| Universal Kriging | $39,5 \pm 2,0$  |



### 3. Разработка серверной части

Серверная часть предоставляет **публичный интерфейс**, позволяющий:

- Обработать большие наборы данных
- Собрать по ним статистику
- Выполнять параметризируемую интерполяцию различными методами

Также присутствует автоматически созданный **веб-интерфейс**.

Через него можно взаимодействовать с API и узнать как им пользоваться.

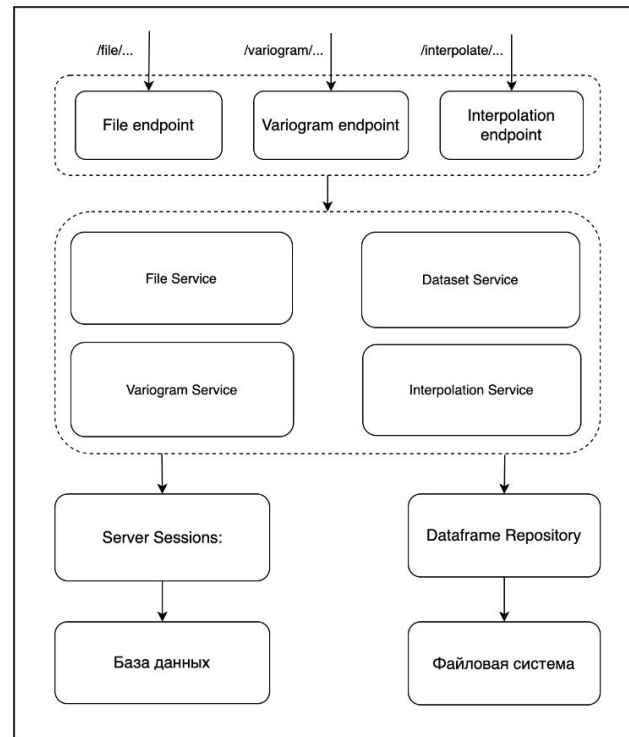
The screenshot displays the 'interpolate' web interface. At the top, it shows the method 'GET' and the endpoint '/interpolate/idw idw'. Below this is a 'Parameters' section with a 'Try it out' button. The parameters are listed in a table with columns 'Name' and 'Description'. The parameters include 'filename' (string, required), 'columns' (string, required), 'normalized' (boolean, default value: false), 'xMin' (number), 'xMax' (number), 'yMin' (number), 'yMax' (number), 'p' (integer, default value: 2), and 'k' (integer, default value: 10). Each parameter has a corresponding input field.

| Name       | Description          |
|------------|----------------------|
| filename   | filename             |
| columns    | columns              |
| normalized | Default value: false |
| xMin       | xMin                 |
| xMax       | xMax                 |
| yMin       | yMin                 |
| yMax       | yMax                 |
| p          | Default value: 2     |
| k          | Default value: 10    |

### 3. Разработка серверной части

**Состояние сервиса** хранится как кэш в отдельной базе данных. Оно нужно только для ускорения обработки запросов.

Хранится только последние рассчитанные данные каждого типа, но для каждой пары из пользователя и набора данных кэш независим.



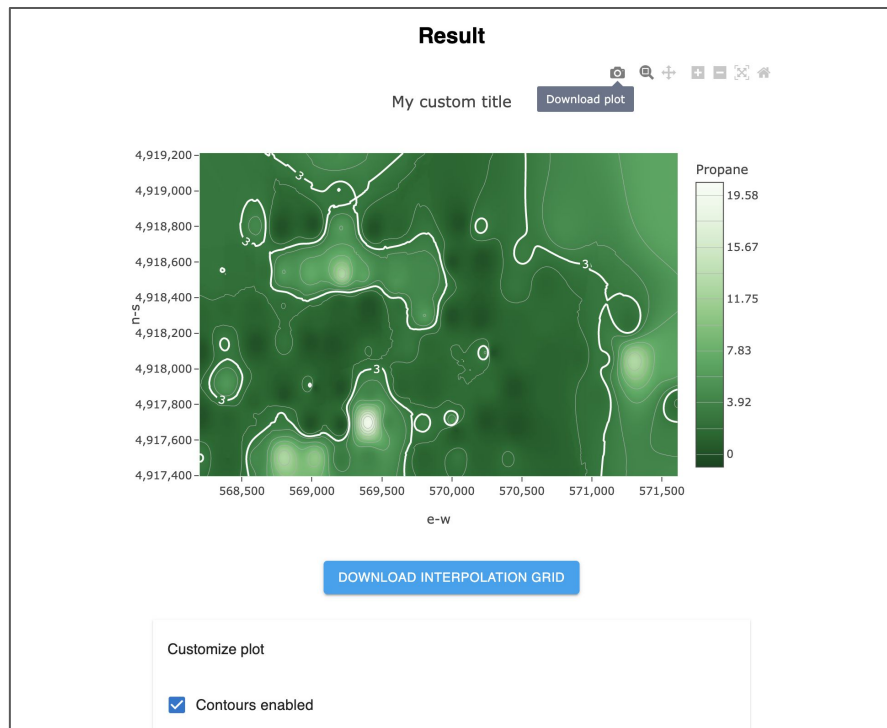
## 4. Разработка клиентской части

Интерфейс в формате Jupyter / Zeppelin.

Раздел с документацией.

Решенные проблемы:

- **Отрисовка графиков в браузере** - уменьшает нагрузку на сервера и дает возможность гибкой настройки.
- Подобраны **значения по умолчанию** для методов интерполяции и настроек графиков. Это сильно упрощает работу с сервисом



## 4. Разработка клиентской части

DOWNLOAD INTERPOLATION GRID

Customize plot

☒ Contours enabled

Amount of contours

10

☒ Show contour label

Contours color

RGB(255, 255, 255)

Additional contours

+

☒ Use default colorscale

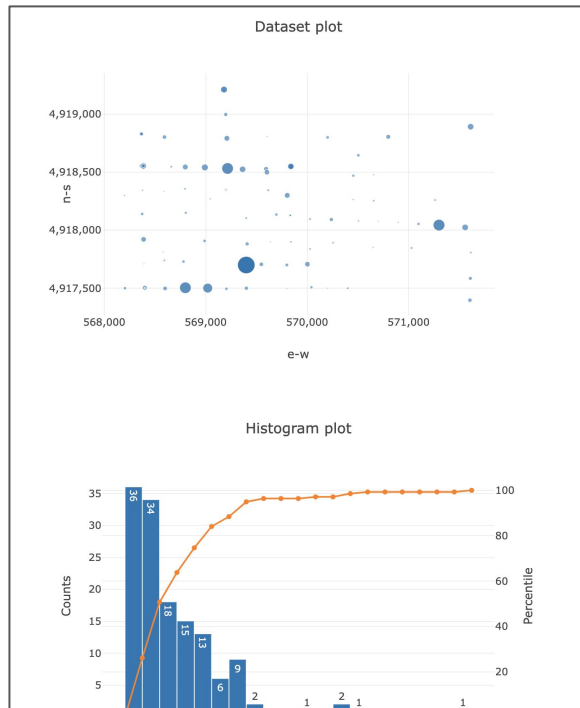
Colorscale

RdBu

Custom range

+

SUBMIT



Number of points

138

|                | Min     | Max     | Average | Median  |
|----------------|---------|---------|---------|---------|
| <b>e-w</b>     | 568198  | 571614  | -       | -       |
| <b>n-s</b>     | 4917397 | 4919212 | -       | -       |
| <b>Propane</b> | 0       | 19.593  | 1.945   | 2.66834 |

Coordinate system

UTM

Utm zone

17T

SUBMIT

Map plot

## 5. Интеграция

Возможности интеграции с другими сервисами:

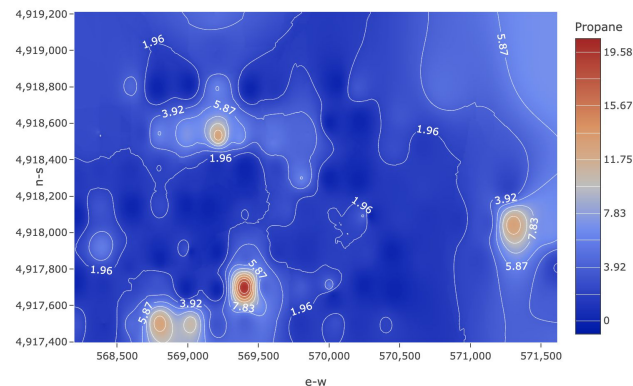
- Встроить график интерполяции в другую страницу.
- Скопировать ссылку на страницу.

Автоматизированы популярные пользовательские сценарии.

Какая-то другая страница

Result

Interpolation plot

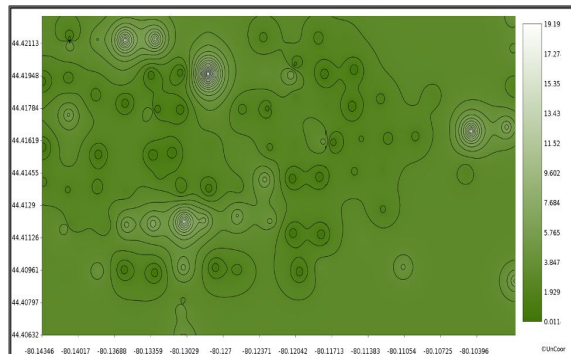


DOWNLOAD INTERPOLATION GRID

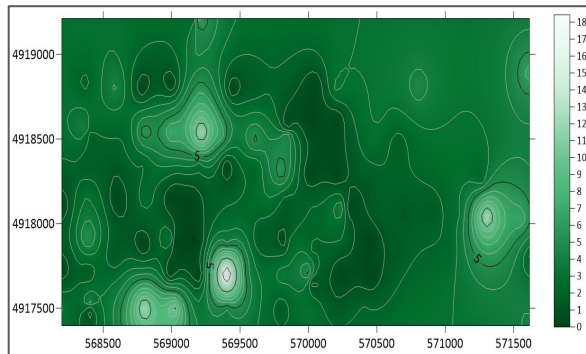
- Решены проблемы, затрудняющие развитие сервиса - использован стандартный для компании Sudo набор технологий.
- Решена проблема масштабирования сервиса.
- Создана и выложена в публичный доступ библиотека с оптимальными методами интерполяции, которая поддерживает работу с большими наборами данных.
- Соблюдены требования на наборы данных и скорость работы в сервисе благодаря правильному использованию кеша.
- Добавлены качественные настраиваемые интерактивные графики.
- Добавлена возможность интеграции с другими сервисами.
- Получена **правильная** интерполяция и экстраполяция.

# Сравнение графиков интерполяции

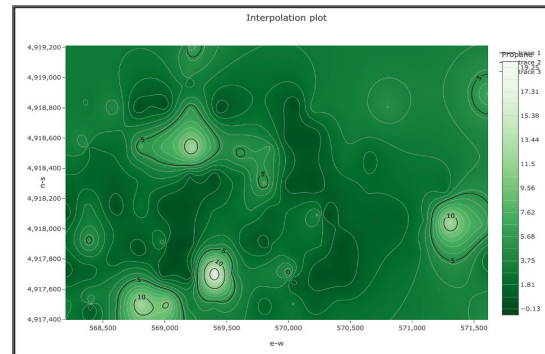
Uncorr



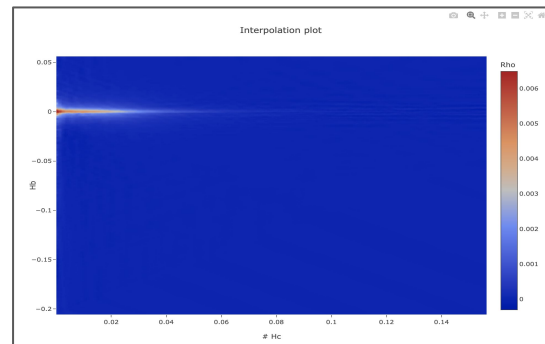
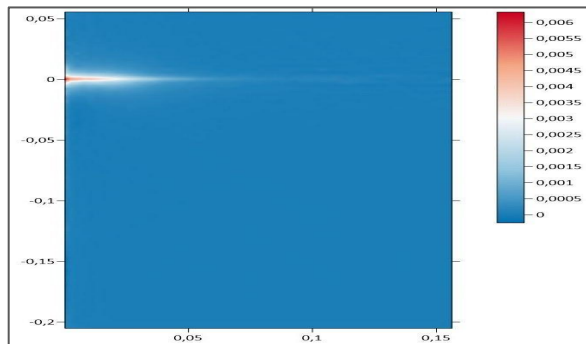
Surfer



Uncorr 2.0



Не поддерживает



- Создана и выложена в публичный доступ библиотека, в которой собраны и стандартизированы оптимальные реализации методов интерполяции. Аналогов ей по разнообразию методов нет.
- Реализован сервис по интерполяции и экстраполяции данных, удовлетворяющий актуальным требованиям.
- Сервис внедрен в компанию Sudo.
- Запланирована интеграция в сервис [forctool.com](https://forctool.com).



# **ІТМО**

---

**Спасибо за внимание!**

---

# Экстраполяция

Решена задача экстраполяции

Методы имеют две фазы:  
обучение и предсказание. На  
втором этапе можно получить  
аппроксимацию для  
произвольной части  
плоскости.

