

# Проектирование расчётного модуля для построения различных статистических показателей климатических данных в рамках проекта “BioSense”

Сопов Н.Е.

Алтайский государственный университет, г. Барнаул  
nsopov79@gmail.com

## Аннотация

Статья посвящена проектированию расчётного модуля для проекта “BioSense”. Описываются расчёты на основе исходных данных различных статистических показателей: средние значения (суточные, годовые), максимумы, минимумы и стандартные отклонения. Особое внимание уделяется методам корреляционного анализа данных из нескольких датчиков, которые измеряют одинаковые параметры в одной локации. Проектирование данного модуля позволит автоматизировать реализацию запросов с анализом климатических данных.

*Ключевые слова:* Проектирование, backend, Python, web-приложение, статистические показатели, NumPy, обработка данных, корреляция данных.

В настоящее время ведётся разработка web-приложения “BioSense” для хранения, обработки и анализа данных, поступающих с климатических датчиков. Система предназначена для мониторинга различных параметров окружающей среды и их влияния на биологические объекты. Важнейшим этапом в создании системы является проектирование расчётного модуля, отвечающего за построение различных статистических показателей и срочных данных на основе собранных измерений. Значимость проектируемого модуля backend части проекта обусловлена необходимостью автоматизации вычислительных процессов, обеспечения необходимой точности расчётов и предоставления возможности проводить глубокий анализ взаимосвязей между различными климатическими параметрами.

Этапы работы проектируемого модуля должны быть выстроены в строгую логическую последовательность и осуществлять выполнение собственной подзадачи:

### 1. Определение подходящих данных.

На этом этапе происходит фильтрация данные, подходящих для выбранных параметров. Критерии отбора включают временные метки, целостность данных и релевантность (например, данные с пометкой “не релевантные” исключаются). Фильтр для выбора данных настраивается пользователем. На стороне сервера выборка осуществляется с помощью запросов к базе данных, реализованных с использованием библиотеки SQLAlchemy [1].

### 2. Предварительная обработка данных.

Данные очищаются от “подозрительных” выбросов, близких к нерелевантным значениям. Производится преобразование единиц измерения и синхронизация временных меток. Этап важен для получения качественных и достоверных результатов.

### 3. Расчет статистических показателей.

После предварительной обработки вычисляются выбранные параметры: срочные данные, среднесуточные и среднегодовые значения, медиана, moda, размах значений, максимумы и минимумы за заданные периоды, а также скользящие средние для сглаживания

временных рядов. При выборе пользователем нескольких источников измерений, выполняется корреляционный анализ для оценки согласованности данных. Этот этап включает вычисление коэффициентов корреляции (например, Пирсона) и расчет средних значений по нескольким характеристически схожим датчикам для повышения точности оценки. Величина коэффициента корреляции используется для оценки качества данных и позволяет отфильтровать недостоверные или несогласованные измерения перед отправкой результатов расчётов пользователю. Высокий коэффициент корреляции указывает на согласованность измерений, низкий – на возможные систематические ошибки или наличие недостоверных данных. На языке Python [2] (используемом на backend составляющей проекта) эти вычисления эффективно выполняются с помощью функционала библиотек NumPy [3] и SciPy [4].

Аналогичные, но менее эффективные для больших объемов данных расчеты, могут быть проведены в Excel [5]: среднесуточные значения вычисляются с помощью функции “СРЗНАЧЕСЛИ”. Построение графиков также возможно с помощью встроенных инструментов. Таким образом, вычисление среднесуточной температуры в Excel служит простой аналогией алгоритму, с помощью NumPy и SciPy для обработки больших объемов данных. Расчет статистических параметров позволяет визуализировать данные, например, показать зависимость среднесуточной температуры от времени. Пример такого графика, построенного в Excel, представлен на рисунке 1.



Рисунок 1. График среднесуточной температуры

Спроектированный модуль обеспечит частичную автоматизацию анализа климатических данных, предоставляя пользователям необходимые статистические показатели и результаты корреляционного анализа.

## Список литературы

1. SQLAlchemy [Электронный ресурс]. – URL: <https://docs.sqlalchemy.org/en/20/>. Дата обращения 22.11.2024.
2. Python [Электронный ресурс]. – URL: <https://docs.python.org/3.12/index.html>. Дата обращения 22.11.2024.
3. NumPy [Электронный ресурс]. – URL: <https://numpy.org/doc/>. Дата обращения 22.11.2024.
4. SciPy [Электронный ресурс]. – URL: <https://docs.scipy.org/doc/scipy/>. Дата обращения 22.11.2024.
5. Excel [Электронный ресурс]. – URL: <https://support.microsoft.com/ru-ru/office>. Дата обращения 22.11.2024.