Санкт-Петербургский государственный университет

Кафедра системного программирования

# Генератор временных рядов с пространственными взаимосвязями

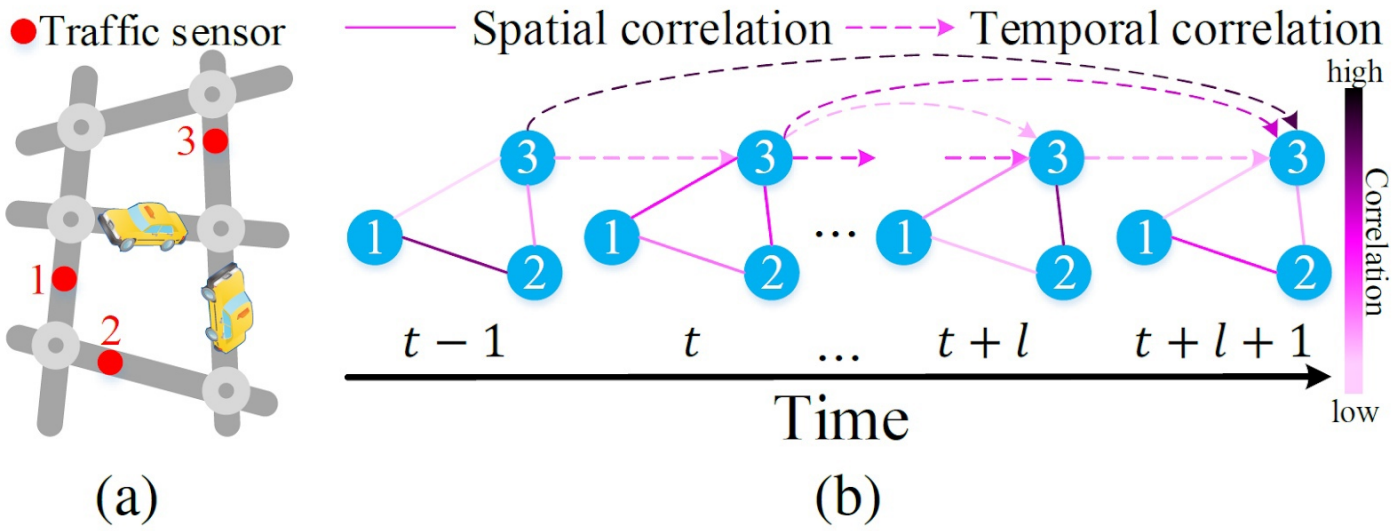
Кропачева Алена Михайловна, группа 21.Б08-мм

**Научный руководитель:** профессор каф. системного программирования, д. т. н., доцент Д.В. Кознов,
**Консультант:** старший инженер ООО «Техкомпания Хуавэй» Д.В. Гирдюк,

Санкт-Петербург
2024

# Введение

- Временные ряды широко используются в различных областях, и для их анализа требуется большое количество данных
- Для задач из области пространственно-временного моделирования необходимы специальные наборы данных с коррелирующими временными рядами



# Мотивация

- Временные ряды, представленные во многих открытых наборах данных, достаточно однородны и независимы друг от друга
- Подходящие наборы данных либо полностью отсутствуют, либо имеют небольшой размер
- Синтетические временные ряды предоставляют информацию о процессах и параметрах

# Требования

- Данные генерируются на основе различных авторегрессионных моделей временных рядов
- Все случайные процессы имеют набор параметров, которые задаются извне и определяют близость временных рядов
- Параметры процессов могут быть изменены в некоторый момент времени
- Пользователь может создать собственный случайный процесс

# Существующие решения

| Метод | Преимущества | Недостатки |
|---|---|---|
| GRATIS: Generating Time Series with diverse and controllable characteristics | Генерация настраиваемых временных рядов | Временные ряды независимы |
| Correlated synthetic time series generation for energy system simulations | Генерация коррелирующих временных рядов | Корреляция применяется одинаково ко всей системе |
| timeseries-generator | Генерация рядов с желаемыми особенностями | Используется только один процесс |
| mockseries | Генерация рядов с желаемыми особенностями, можно менять процесс | Временные ряды независимы |

# Постановка задачи

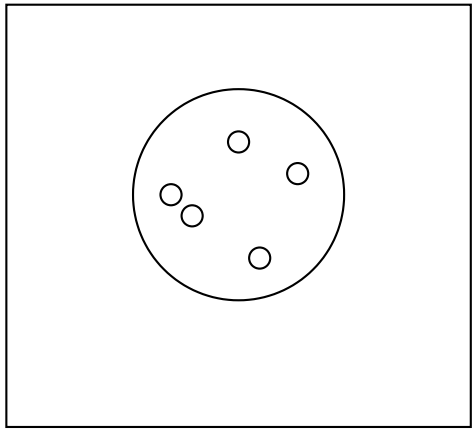
**Целью** работы является разработка приложения для генерации наборов данных временных рядов с пространственными взаимосвязями.

**Задачи**:

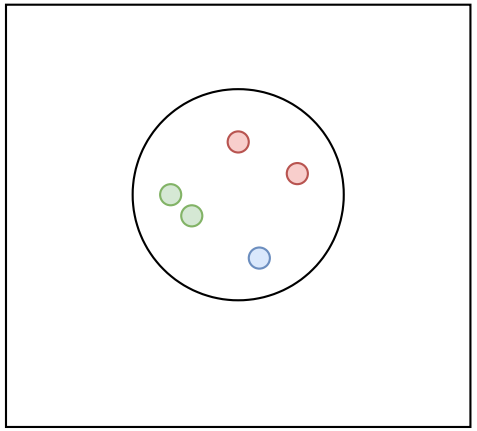
1. Реализовать набор процессов для генерации комплексных моделей временных рядов
2. Реализовать различные способы сэмплирования исходных данных
3. Реализовать различные способы генерации параметров процессов на основе исходных данных
4. Провести тестирование генератора временных рядов

# Метод генерации временных рядов

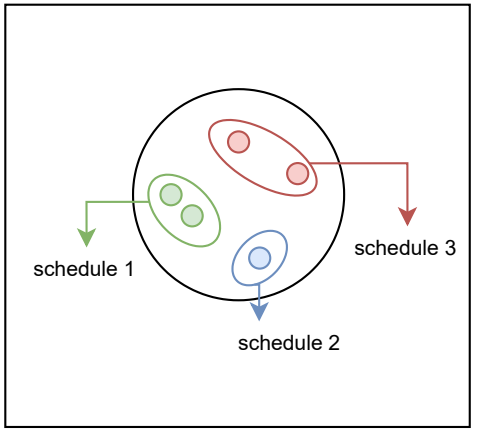
1. Generate source points



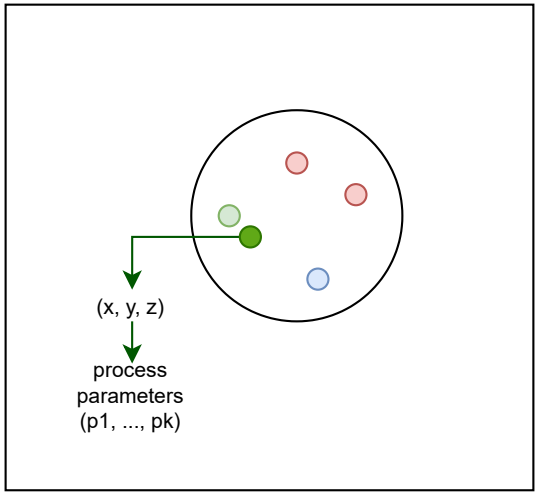
2. Cluster points



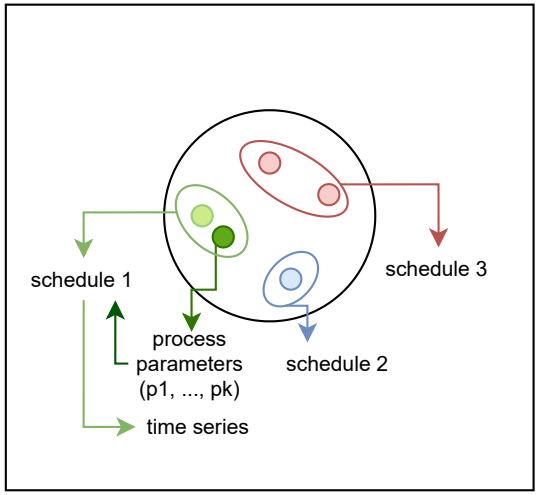
3. Generate schedule for each cluster




4. Generate process parameters




5. Generate time series according to schedules



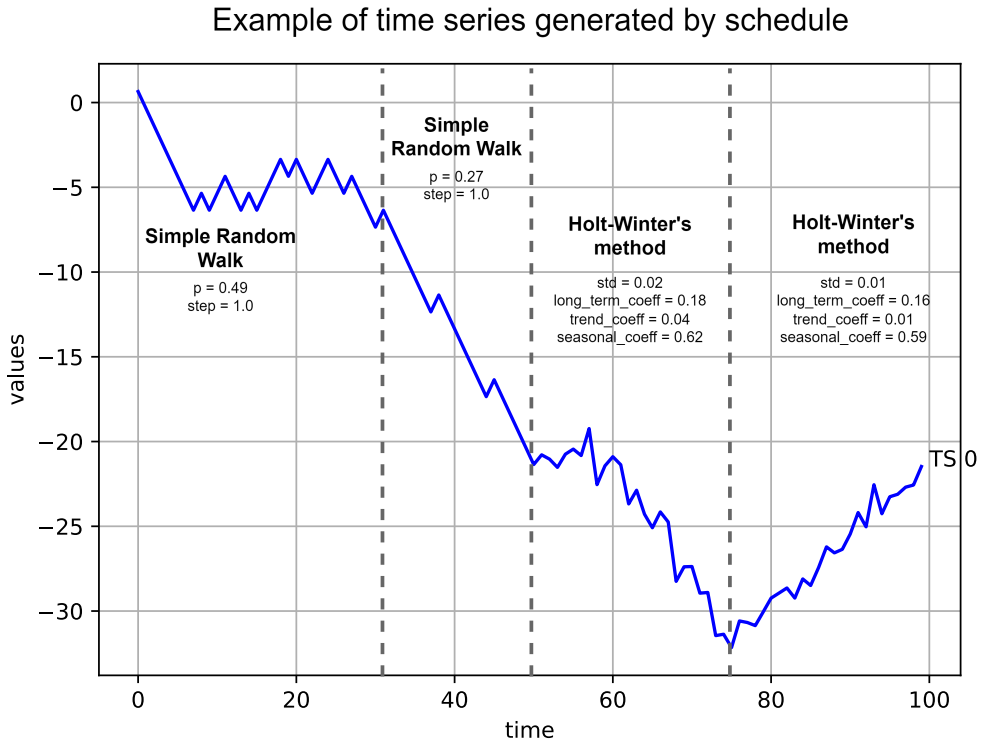
# Процессы, порождающие временные ряды

Для генерации временных рядов в программе используются следующие процессы:

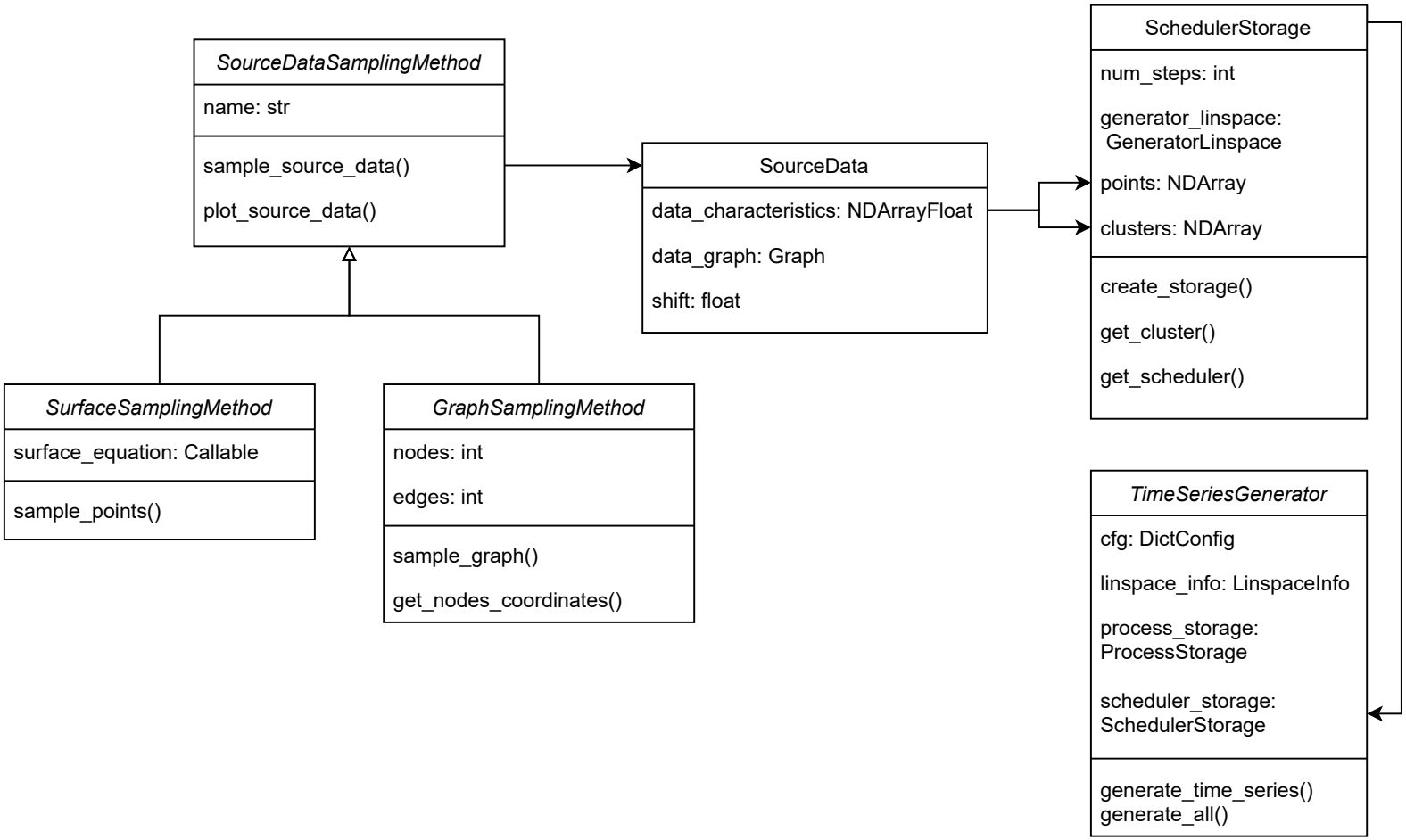
- Белый шум
- Случайное блуждание
- ETS-модели:
  - Простое экспоненциальное сглаживание
  - Метод Хольта
  - Метод Хольта-Уинтерса

# Расписание временного ряда

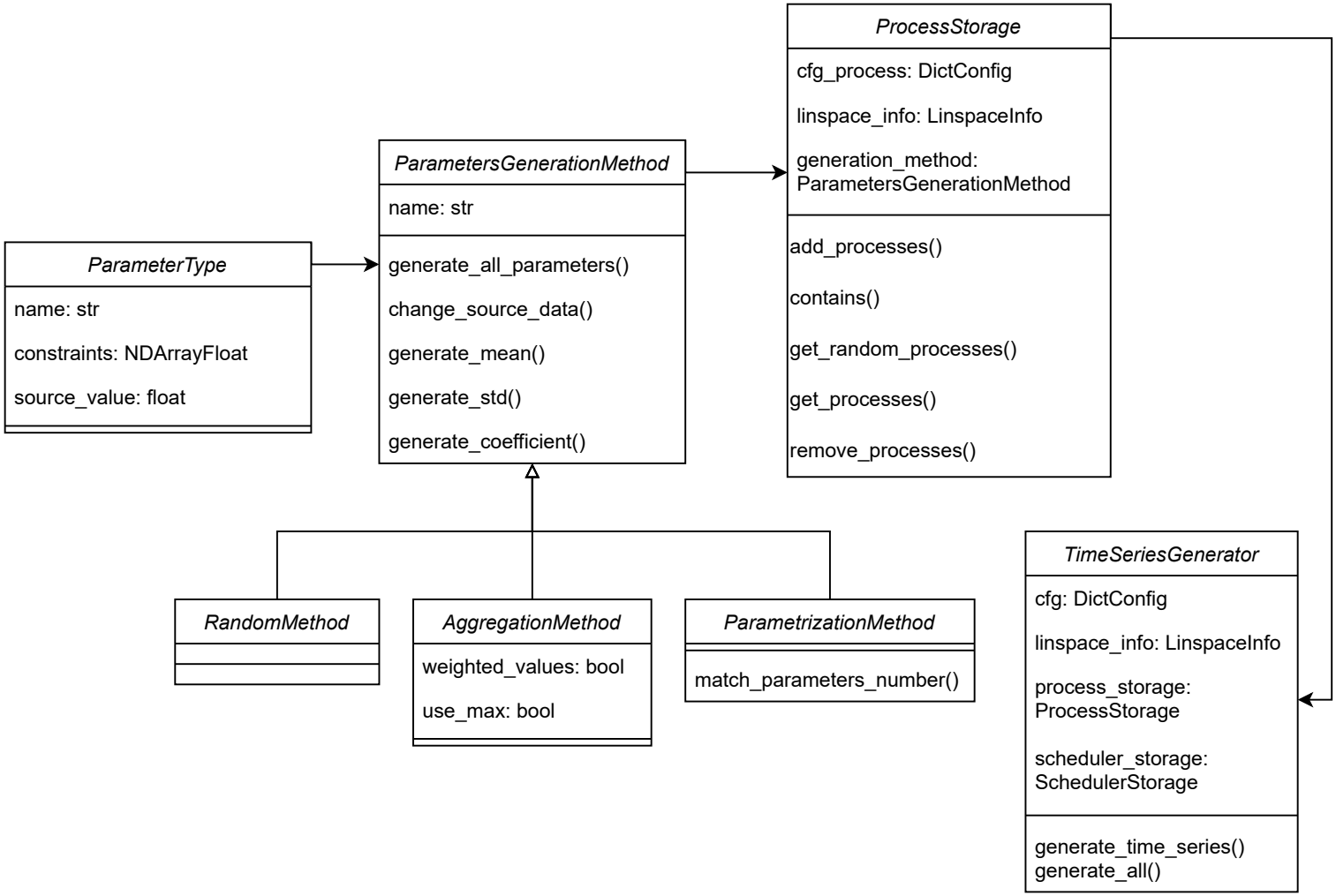
Расписание временного ряда — это комплексная модель, которая задает порядок процессов, порождающих временной ряд. Метод создания расписания: выбор $k$ случайных процессов, генерация $i_k$ наборов параметров для каждого процесса.



# Генерация исходных данных

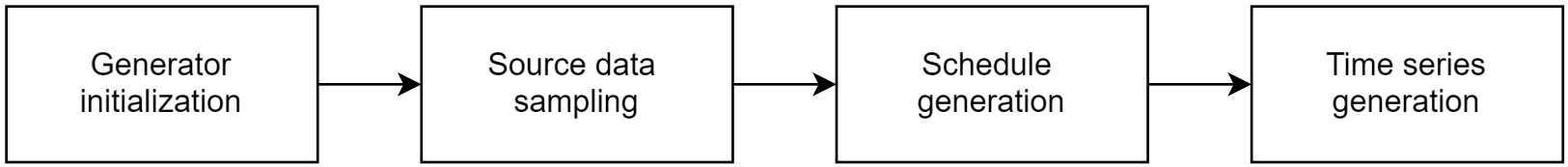


# Параметризация процессов



# Реализация

Сценарий генерации временных рядов:

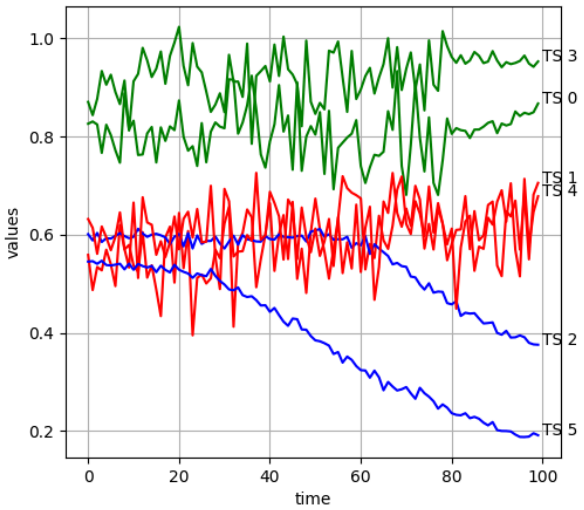


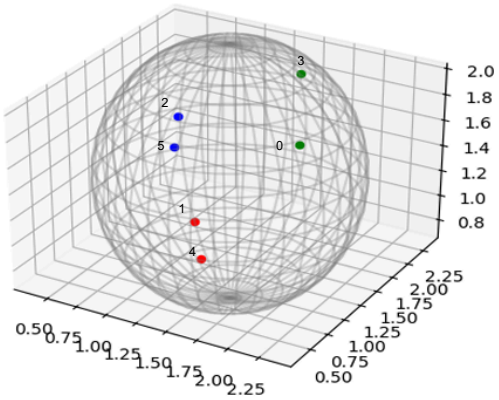

Для реализации были использованы следующие технологии:

- Python 3.10
- NumPy, hydra, scikit-learn, NetworkX — сэмплирование исходных данных и вычисление значений временных рядов
- pytest — Unit-тестирование
- mypy, black, isort — форматирование кода
- GitHub CI

Код реализации: https://github.com/stil-mmo/time-series-generator

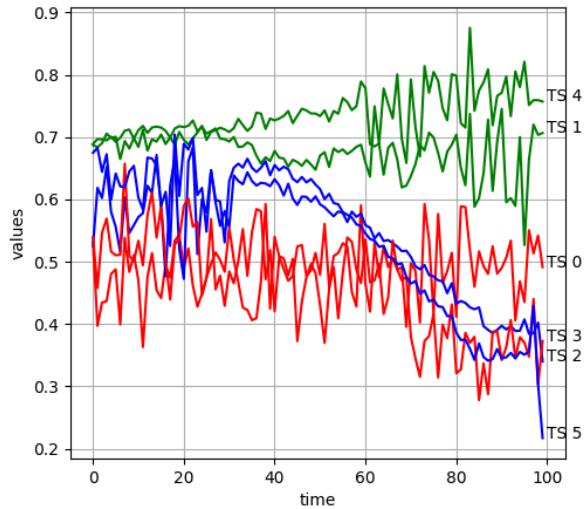
# Тестирование

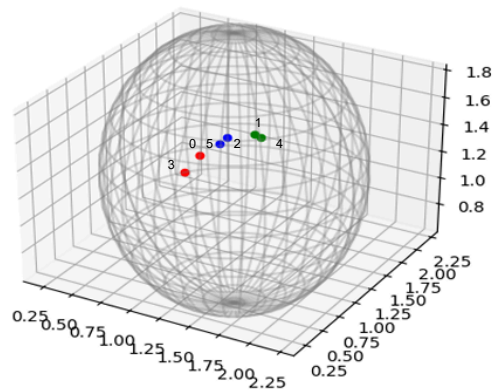


| | DTW | MSE | MAE | SMAPE |
|---|---|---|---|---|
| C1 | 0.94 | 0.03 | 0.16 | 38.81 |
| C2 | 0.47 | 0.01 | 0.07 | 11.82 |
| C3 | 0.78 | 0.02 | 0.12 | 14.50 |

| | | DTW | MSE | MAE | SMAPE |
|---|---|---|---|---|---|
| C1 | C2 | 1.16 | 0.03 | 0.12 | 25.60 |
| C1 | C3 | 3.03 | 0.13 | 0.31 | 46.64 |
| C2 | C3 | 1.86 | 0.06 | 0.22 | 29.25 |

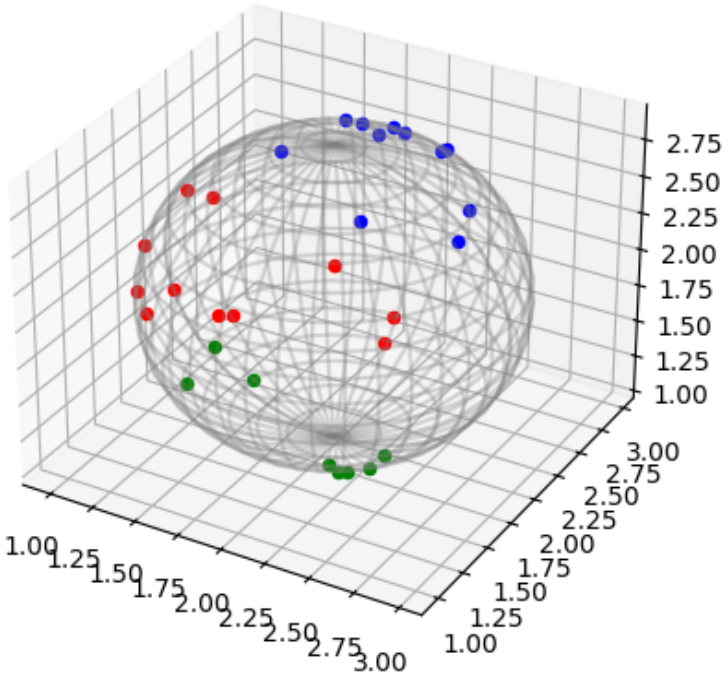


| | DTW | MSE | MAE | SMAPE |
|---|---|---|---|---|
| C1 | 0.33 | 0.003 | 0.04 | 8.36 |
| C2 | 0.55 | 0.01 | 0.09 | 18.76 |
| C3 | 0.48 | 0.01 | 0.06 | 8.88 |

| | | DTW | MSE | MAE | SMAPE |
|---|---|---|---|---|---|
| C1 | C2 | 0.62 | 0.01 | 0.09 | 19.51 |
| C1 | C3 | 1.23 | 0.03 | 0.13 | 21.66 |
| C2 | C3 | 1.51 | 0.04 | 0.17 | 29.57 |

# Тестирование



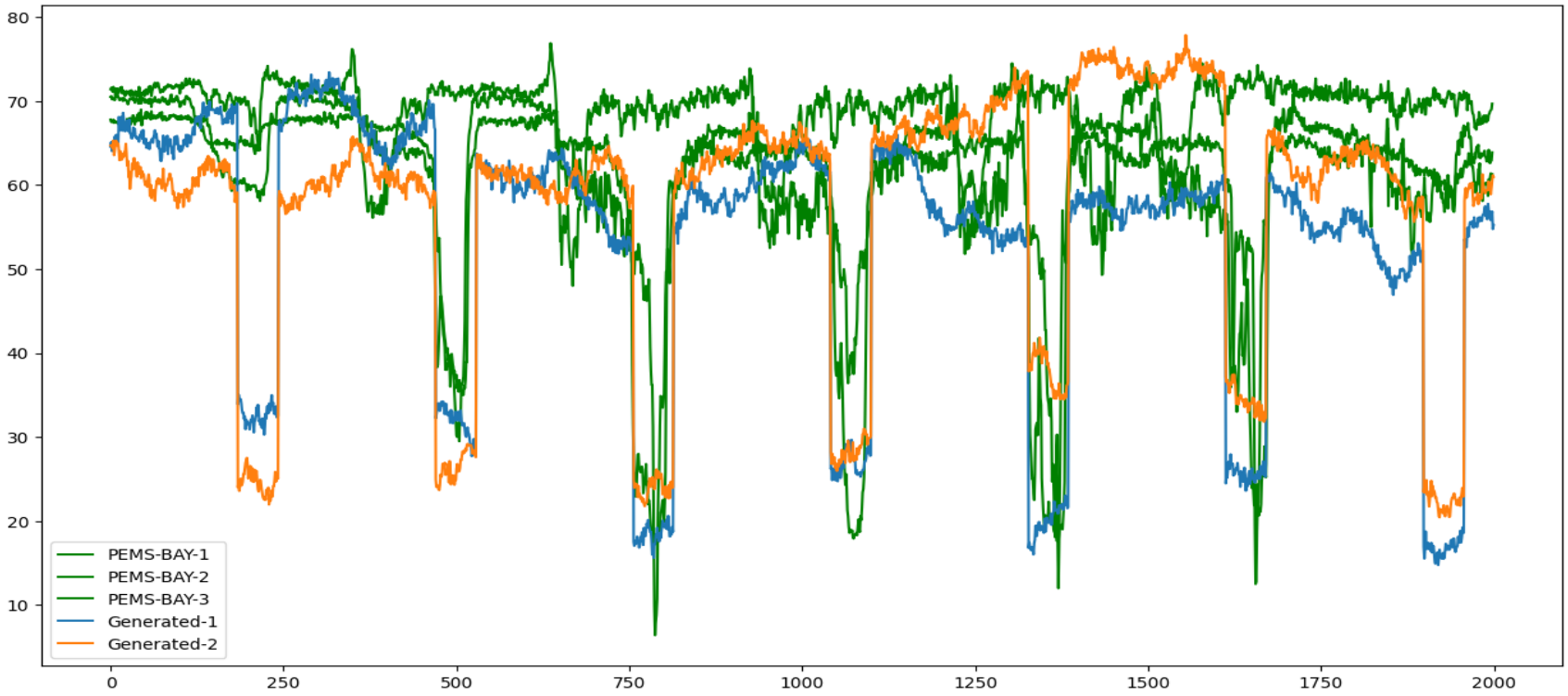

| | DTW | MSE | MAE | SMAPE |
|---|---|---|---|---|
| C1 | 1.73 | 0.05 | 0.19 | 8.26 |
| C2 | 2.92 | 0.13 | 0.29 | 17.51 |
| C3 | 2.46 | 0.09 | 0.23 | 26.07 |

| | | DTW | MSE | MAE | SMAPE |
|---|---|---|---|---|---|
| C1 | C2 | 3.54 | 0.20 | 0.35 | 20.28 |
| C1 | C3 | 5.02 | 0.52 | 0.49 | 39.76 |
| C2 | C3 | 3.57 | 0.21 | 0.34 | 33.19 |

# Тестирование

План эксперимента:

1. Генерация синтетических временных данных на основе реального датасета
2. Обучение моделей на синтетических и реальных данных
3. Сравнение метрик



# Результаты

1. Проведен обзор существующих решений для генерации временных рядов
2. Спроектирована модель генератора временных рядов, которая включает в себя:
    1. Набор стохастических процессов, порождающих временные ряды
    2. Метод построения комплексных моделей для генерации временных рядов
    3. Методы генерации исходных данных
    4. Методы генерации параметров процессов
3. Выполнена программная реализация модели
4. Составлен план тестирования результатов работы генератора
5. Подготовлена статья и сделан доклад на конференции SYRCoSE 2024