

Дни науки в Челябинской области 2022  
Круглый стол «Интеллектуальное производство»  
23 ноября 2022, Челябинск, ЮУрГУ

# Интеллектуальный анализ временных рядов для приложений цифровой индустрии



ЮУрГУ

М.Л. Цымблер, Я.А. Краева, А.И. Гоглачев, А.А. Юртин

[mzym](mailto:mzym@susu.ru), [kraevaya](mailto:kraevaya@susu.ru), [goglavchevai](mailto:goglavchevai@susu.ru), [iurtinaa](mailto:iurtinaa@susu.ru)@susu.ru

Лаборатория больших данных и машинного обучения ЮУрГУ

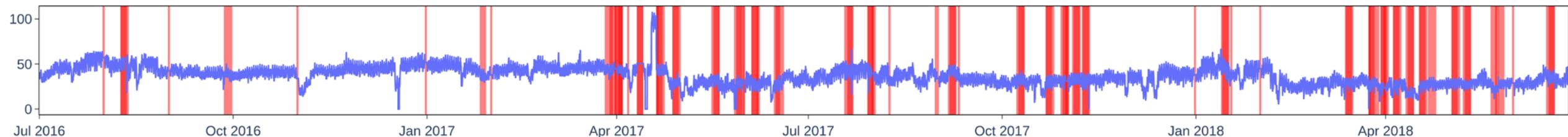
Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 20-07-00140) и Министерства науки и высшего образования РФ (государственное задание FENU-2020-0022)

# Цифровая индустрия и анализ временных рядов



# Аномалии временных рядов

## 2-годичное энергопотребление в Китае<sup>1</sup>

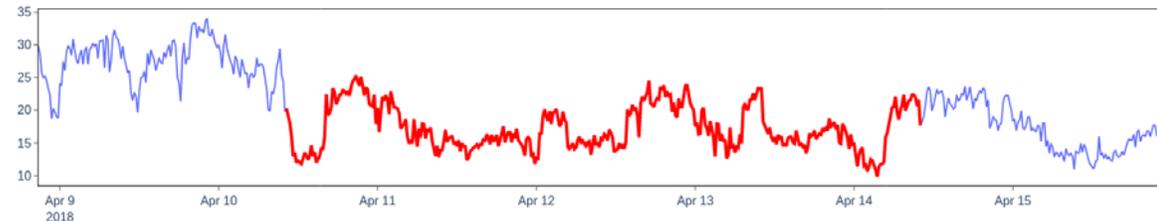
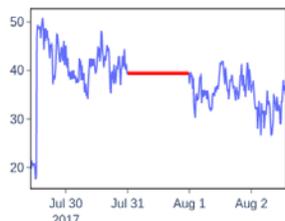


1 день

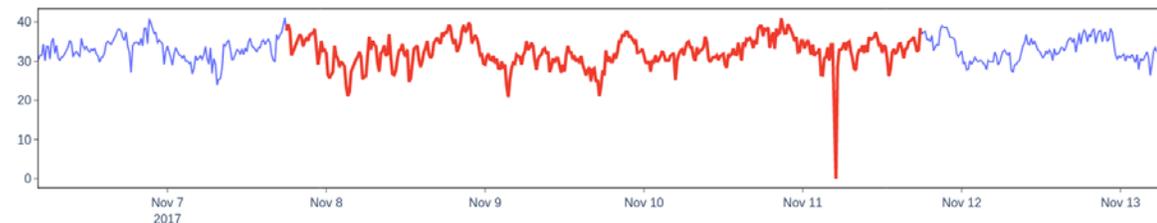
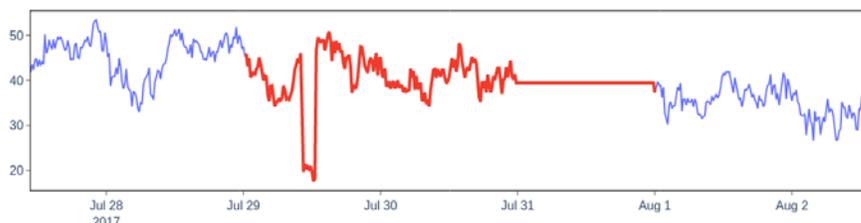
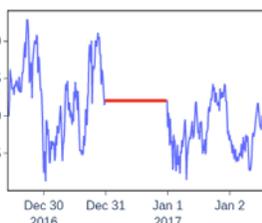
3 дня

4 дня

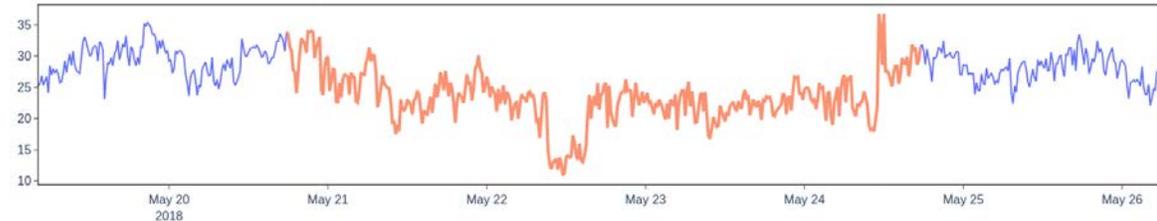
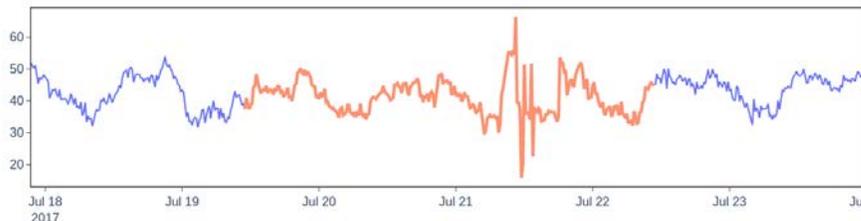
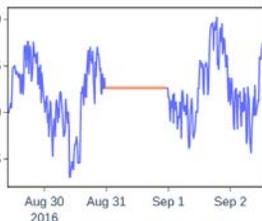
Топ-1  
аномалия



Топ-2  
аномалия



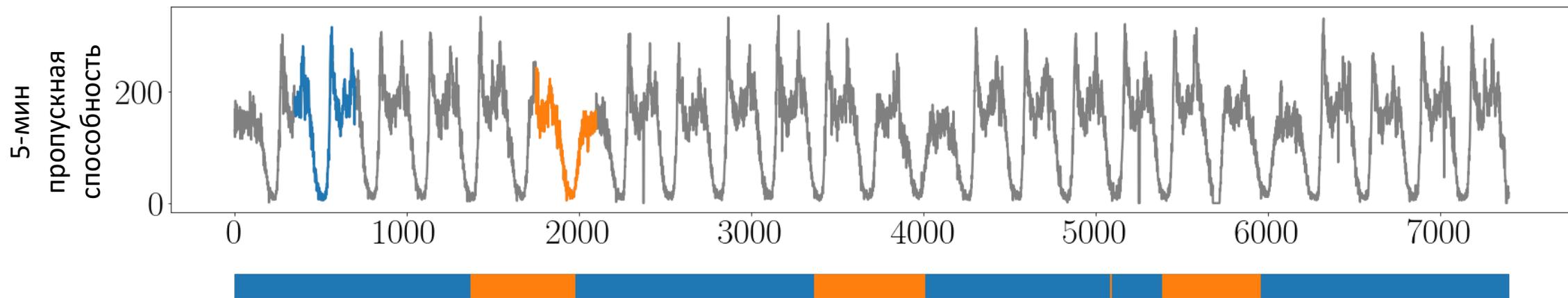
Топ-3  
аномалия



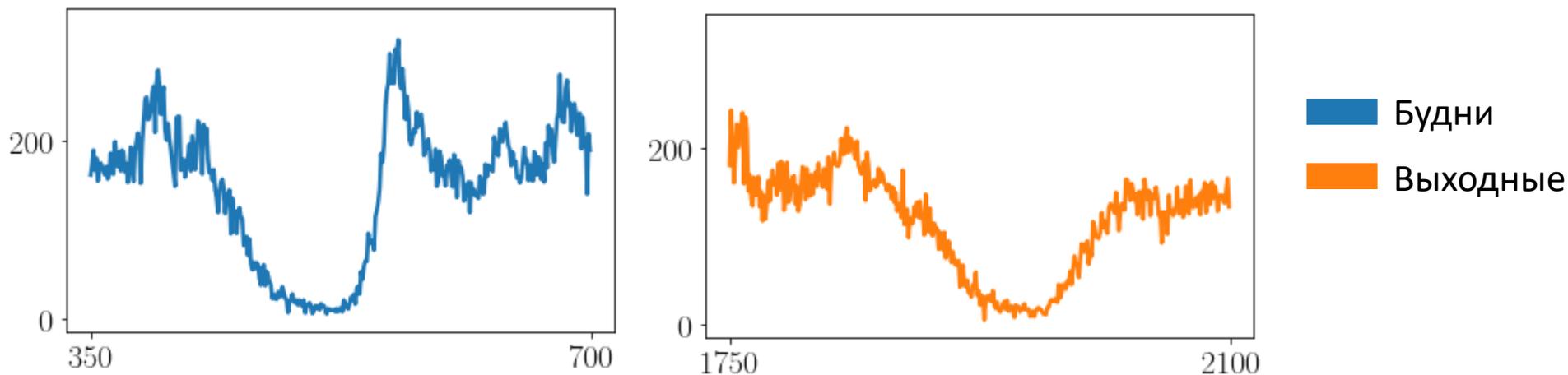
<sup>1</sup>Zhou H. *et al.* Informer: beyond efficient transformer for long sequence time-series forecasting. AAAI 2021: 11106-11115. DOI: [10.1609/aaai.v35i12.17325](https://doi.org/10.1609/aaai.v35i12.17325).

# Шаблоны временных рядов

## Месячный трафик крупного города<sup>1</sup>



## Найденные шаблоны



<sup>1</sup> Public (anonymized) road traffic prediction datasets from Huawei Munich Research Center. URL: <https://zenodo.org/record/3653880#.Y0zZi3ZBxPa>

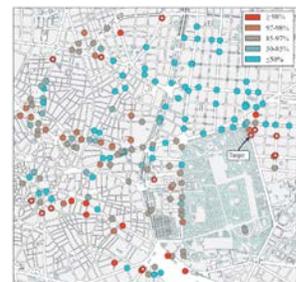
# Пропущенные значения временных рядов



## Система MAREL Carnot

измерение более 15 химико-биологических параметров каждые 20 мин. в проливе Ла-Манш

Lefebvre A. MAREL Carnot data and metadata from Coriolis Data Centre. SEANOE. 2015. DOI: [10.17882/39754](https://doi.org/10.17882/39754).



## Система управления городским трафиком Мадрида

данные более чем 3500 автоматических регистраторов дорожного движения каждые 15 мин.

Lana I. *et al.* On the imputation of missing data for road traffic forecasting: New insights and novel techniques. Transportation Research Part C: Emerging Technologies. 2018. Vol. 90. P. 18-33. DOI: [10.1016/j.trc.2018.02.021](https://doi.org/10.1016/j.trc.2018.02.021).



## Система анализа спортивных данных

датчики на бутсах форвардов и перчатках вратаря (частота 200 Гц: 15000 событий в сек.)

Mutschler C. *et al.* The DEBS 2013 grand challenge. DEBS'13. P. 289-294. DOI: [10.1145/2488222.2488283](https://doi.org/10.1145/2488222.2488283).



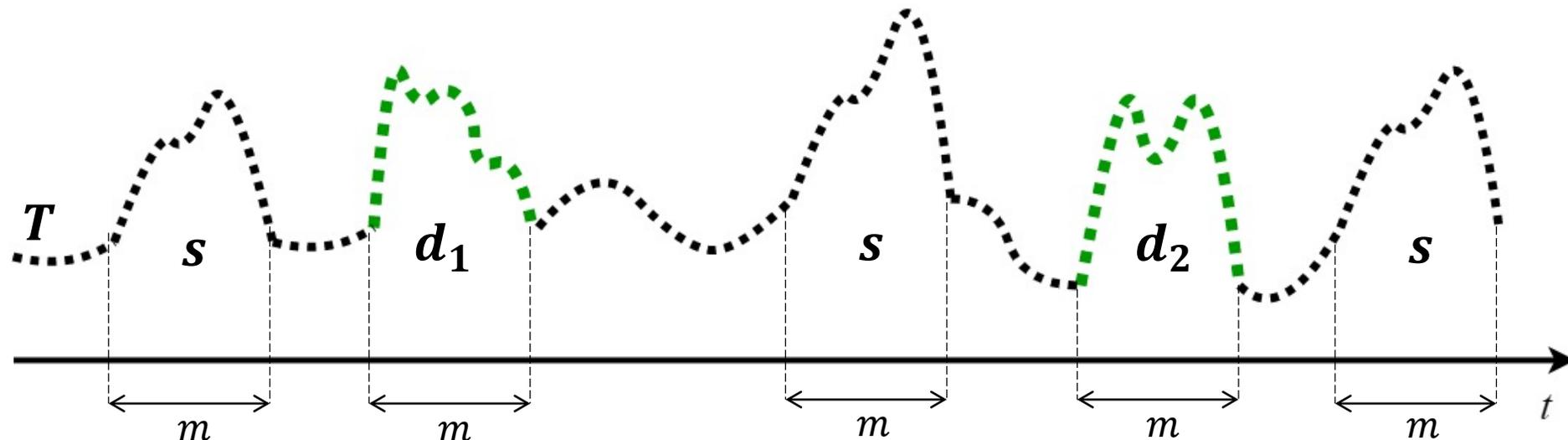
## Спутник BeiDou

теряет до 15% пакетов передаваемых данных

Liu S. *et al.* A novel BeiDou satellite transmission framework with missing package imputation applied to smart ships. IEEE Sensors Journal. 2022. Vol. 22, no. 13. P. 13162-13176. DOI: [10.1109/JSEN.2022.3177167](https://doi.org/10.1109/JSEN.2022.3177167).

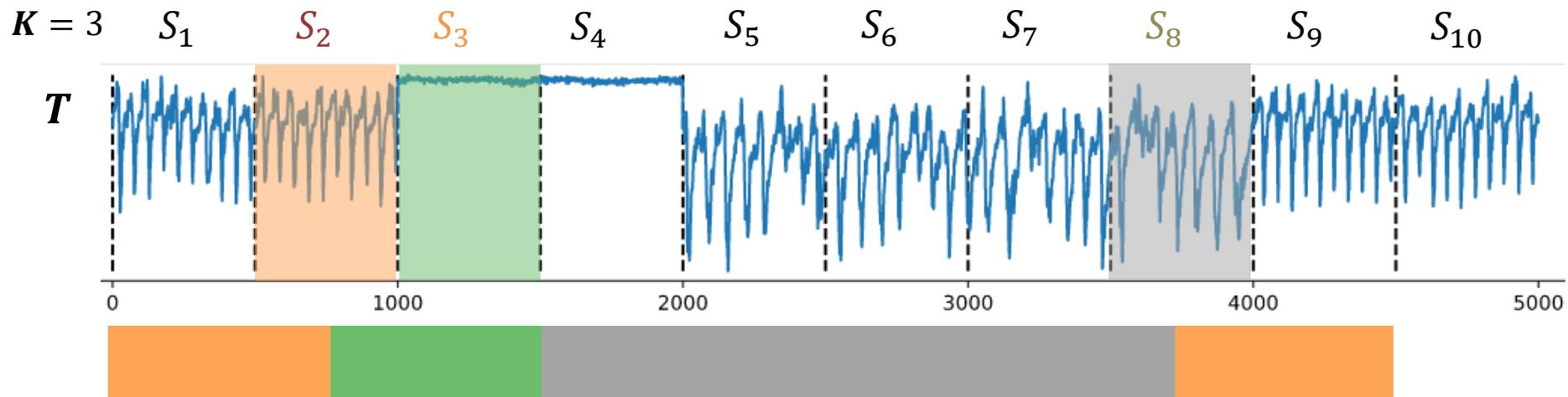
# Формализация понятия аномалии: диссонансы<sup>1</sup>

- **Диссонанс** – подпоследовательность, ближайший сосед которой дальше заданного порога  
ближайший сосед – подпоследовательность, наиболее близкая к данной и не пересекающаяся с ней
- Дано: ряд  $T$ , длина диссонанса  $m$ , порог  $r$
- Найти: диссонансы  $\mathcal{D} = \{d_1, d_2, \dots\}$ ,  $d_i \in \mathcal{D} \Leftrightarrow \min_{s \in T, s \cap d_i = \emptyset} \text{EuclidDist}(d_i, s) \geq r$

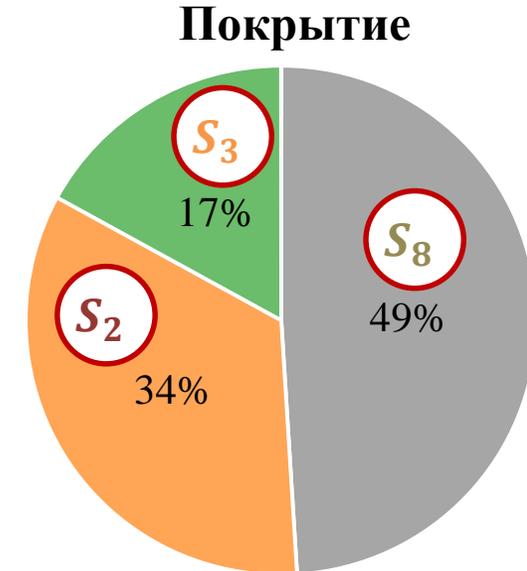


<sup>1</sup>Yankov D. et al. Disk aware discord discovery: finding unusual time series in terabyte sized datasets. Knowl. Inf. Syst. 17(2): 241–262. 2008. DOI: [10.1007/s10115-008-0131-9](https://doi.org/10.1007/s10115-008-0131-9)

# Формализация понятия шаблона: снippets<sup>1</sup>



1. Представим ряд как набор непересекающихся сегментов данной длины
2. Для каждого сегмента найдем его ближайших соседей
3. Для каждого сегмента вычислим **покрытие** (доля его ближайших соседей от подпоследовательностей ряда) и возьмем топ- $K$  сегментов, это будут **снippets ряда**

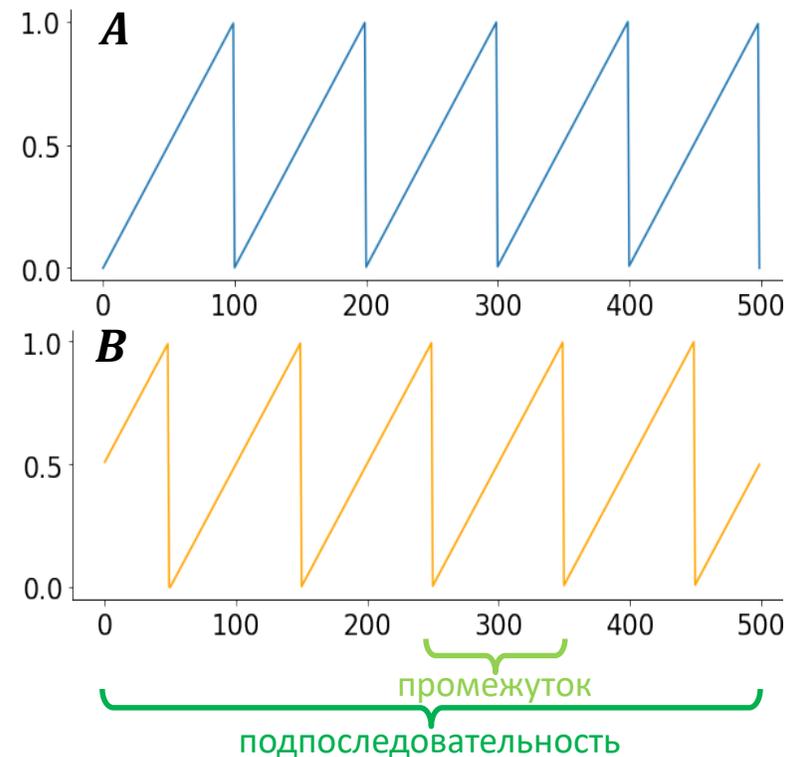


<sup>1</sup> Imani S. *et al.* Introducing time series snippets: a new primitive for summarizing long time series. Data Min. Knowl. Discov. 2020. Vol. 34, no. 6. P. 1713-1743. DOI: [10.1007/s10618-020-00702-y](https://doi.org/10.1007/s10618-020-00702-y)

# Формализация схожести для поиска сниппетов: MPdist<sup>1</sup>

- Две подпоследовательности близки в смысле **MPdist** тем больше, чем больше в них промежутков, близких в смысле **евклидова расстояния**
- **MPdist: инвариантность к фазе**

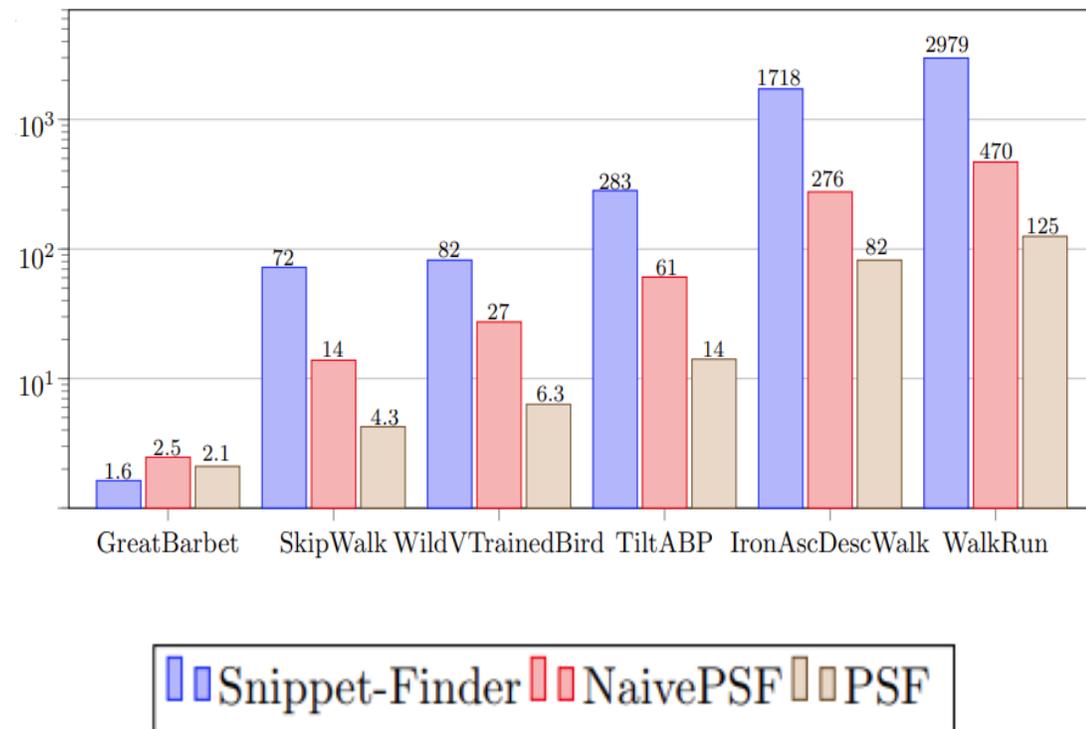
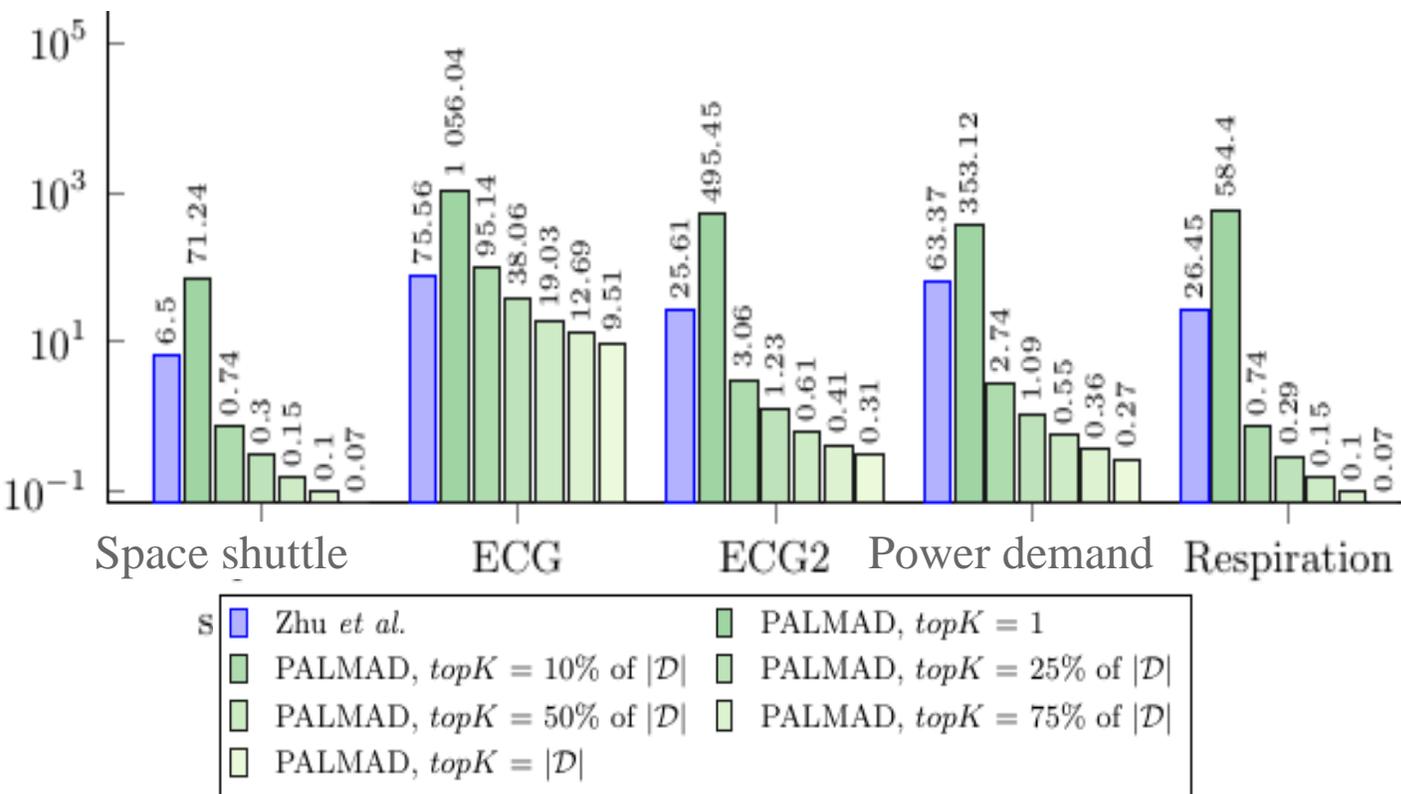
$\text{EuclidDist}(A, B)$	11.2
<b>MPdist</b> (A, B)	<b>0</b>



<sup>1</sup> Gharghabi S. *et al.* An ultra-fast time series distance measure to allow data mining in more complex real-world deployments. *Data Min. Knowl. Discov.* 2020. Vol. 34. P. 1104–1135. DOI: [10.1007/s10618-020-00695-8](https://doi.org/10.1007/s10618-020-00695-8)

# Параллельные алгоритмы поиска аномалий и шаблонов

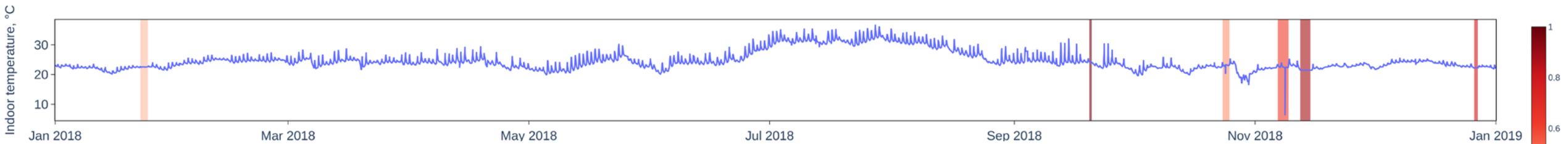
Время работы в сравнении с лучшими аналогами (сек., лог шкала)



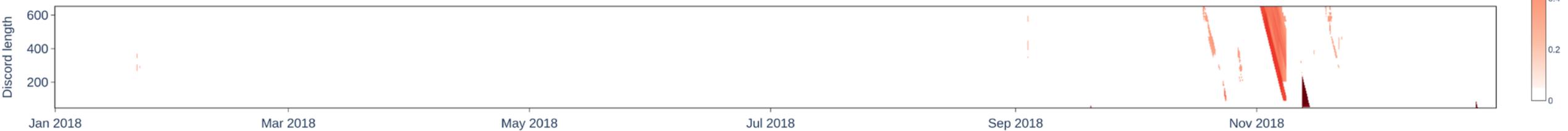
- Опережает аналоги по количеству найденных аномалий за единицу времени

- Опережает аналоги по быстродействию поиска шаблонов

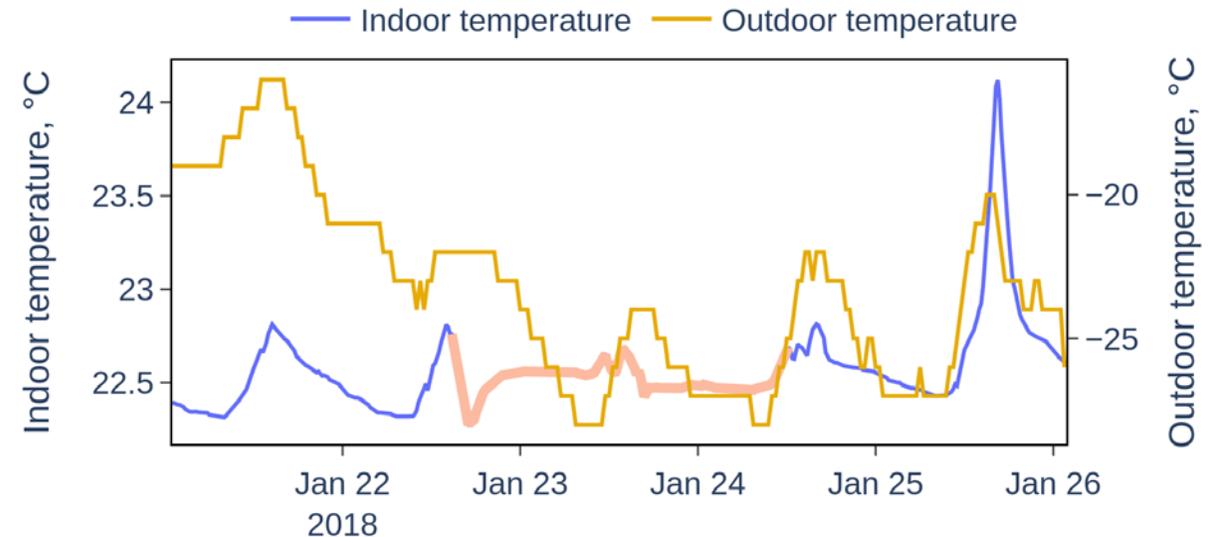
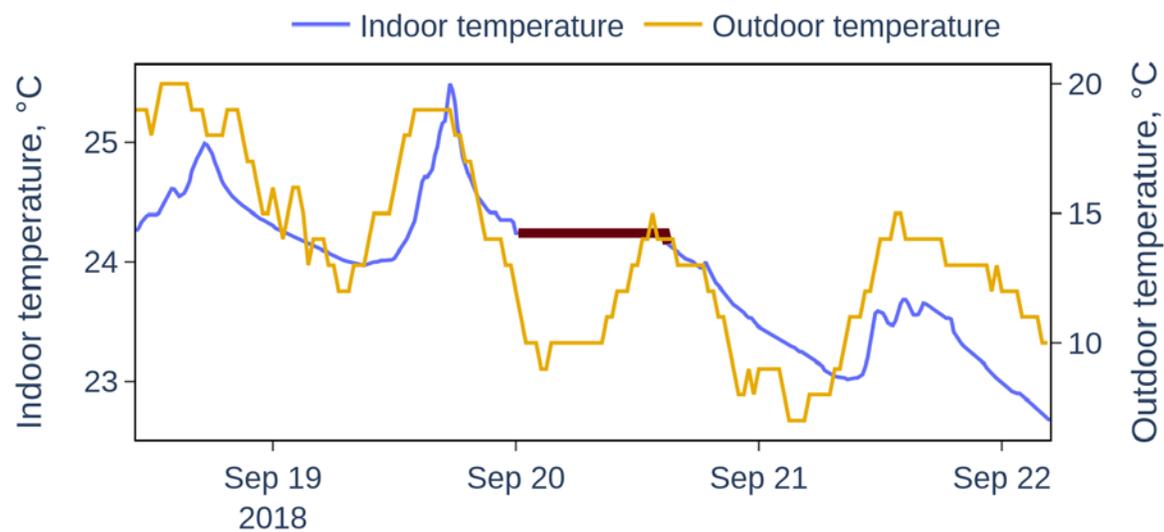
# Выявление аномалий в системе отопления ЮУрГУ



Тепловая карта найденных аномалий длительностью 0.5-2 суток

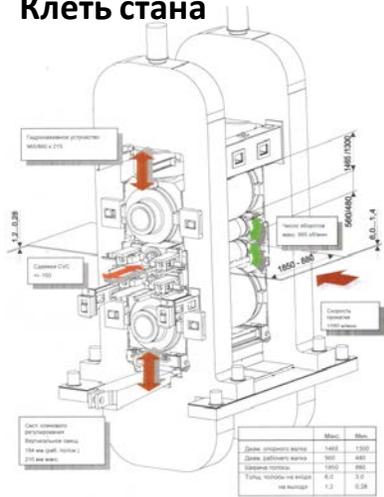


Примеры найденных аномалий длительностью 0.5-2 суток

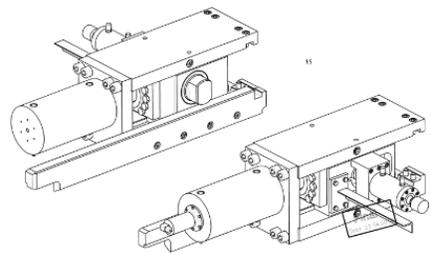


# Стан холодной прокатки ММК: аномалии

Клеть стана



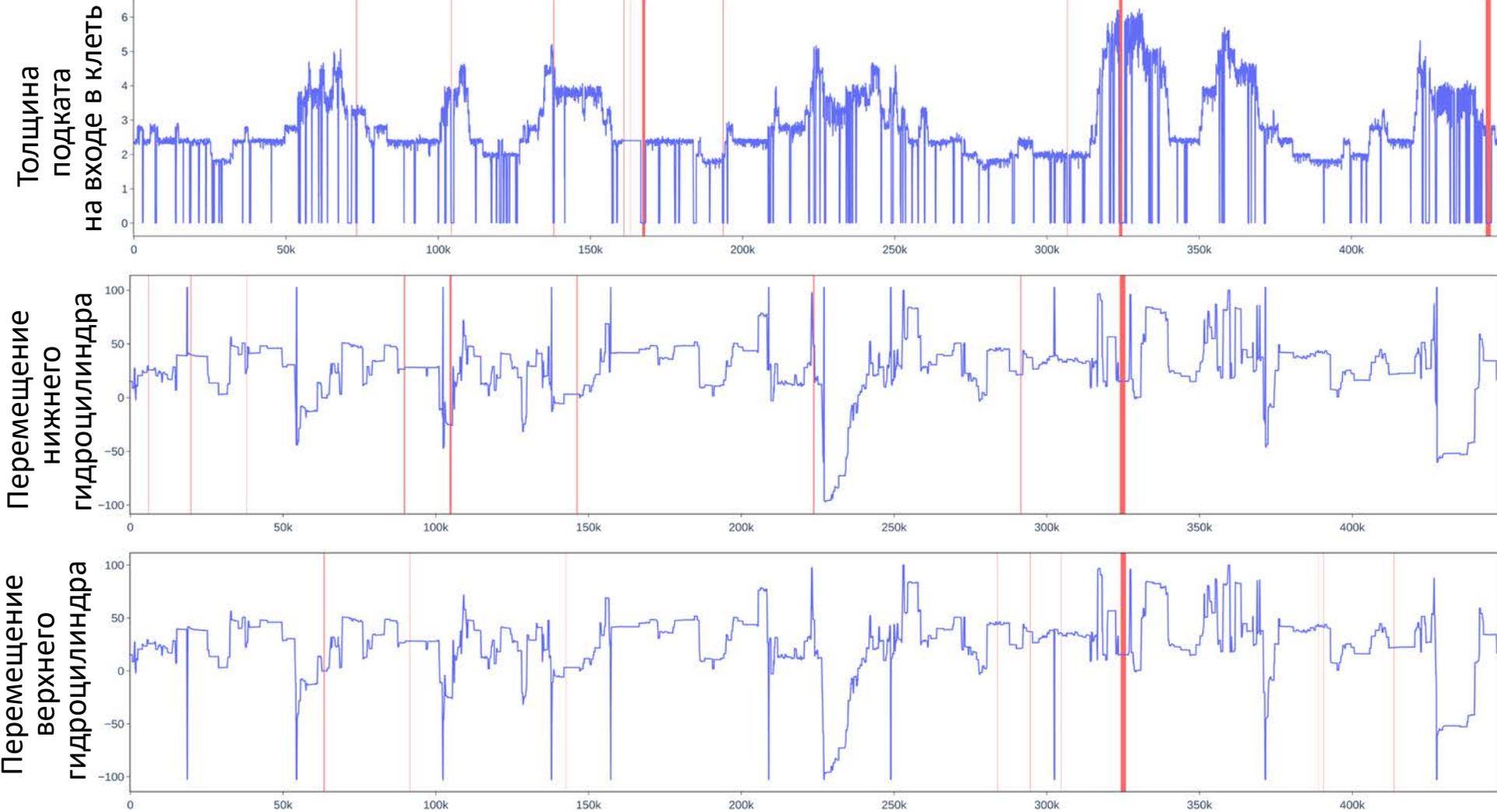
Система валков с непрерывно изменяемой кривизной (CVC, continuously variable curvature)



Разрушение плит CVC



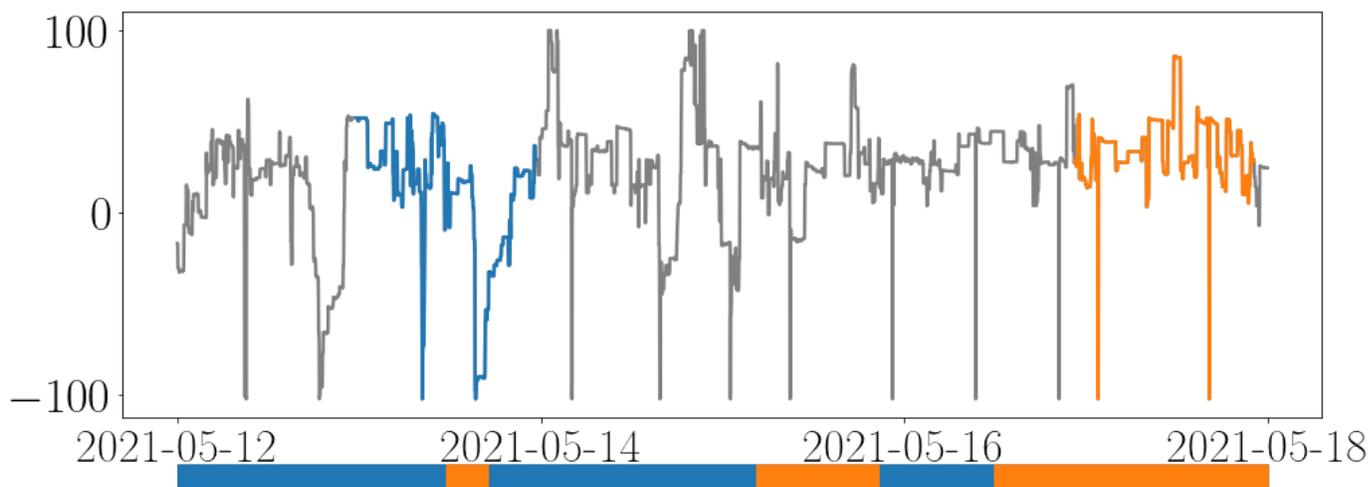
Аномалии за период 12.05.2021 – 18.05.2021 длительностью 1-30 мин



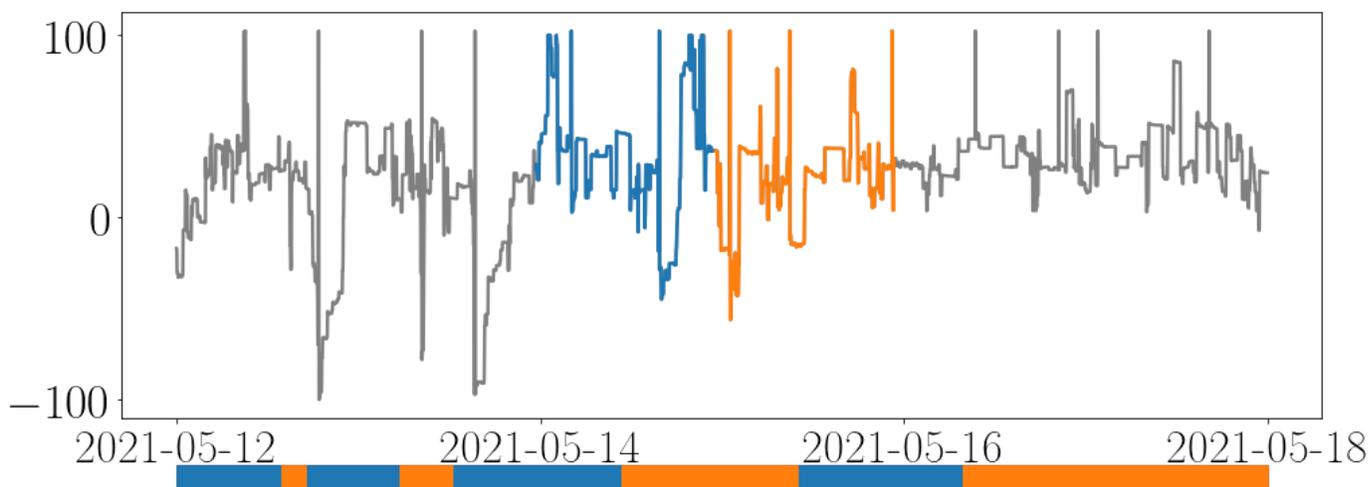
# Стан холодной прокатки ММК: шаблоны

Перемещение верхнего гидроцилиндра системы осевой сдвижки валков SVC

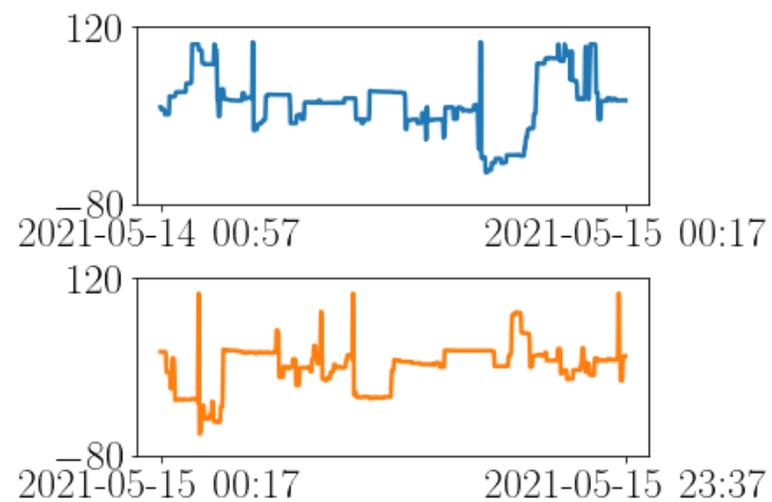
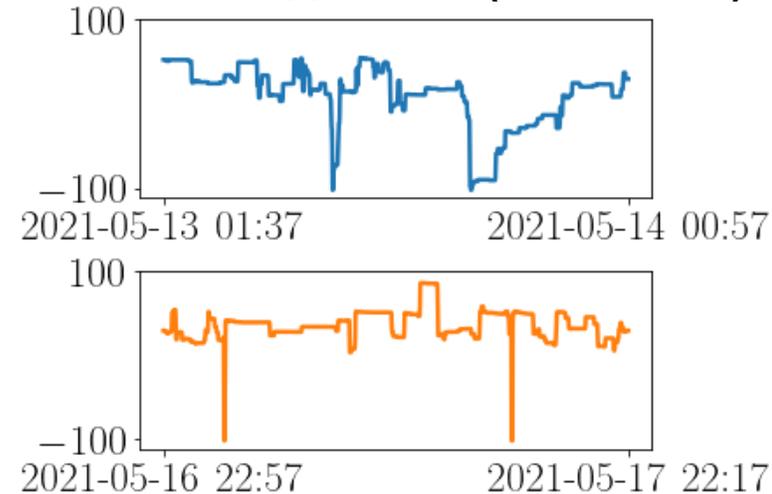
## Разметка данных



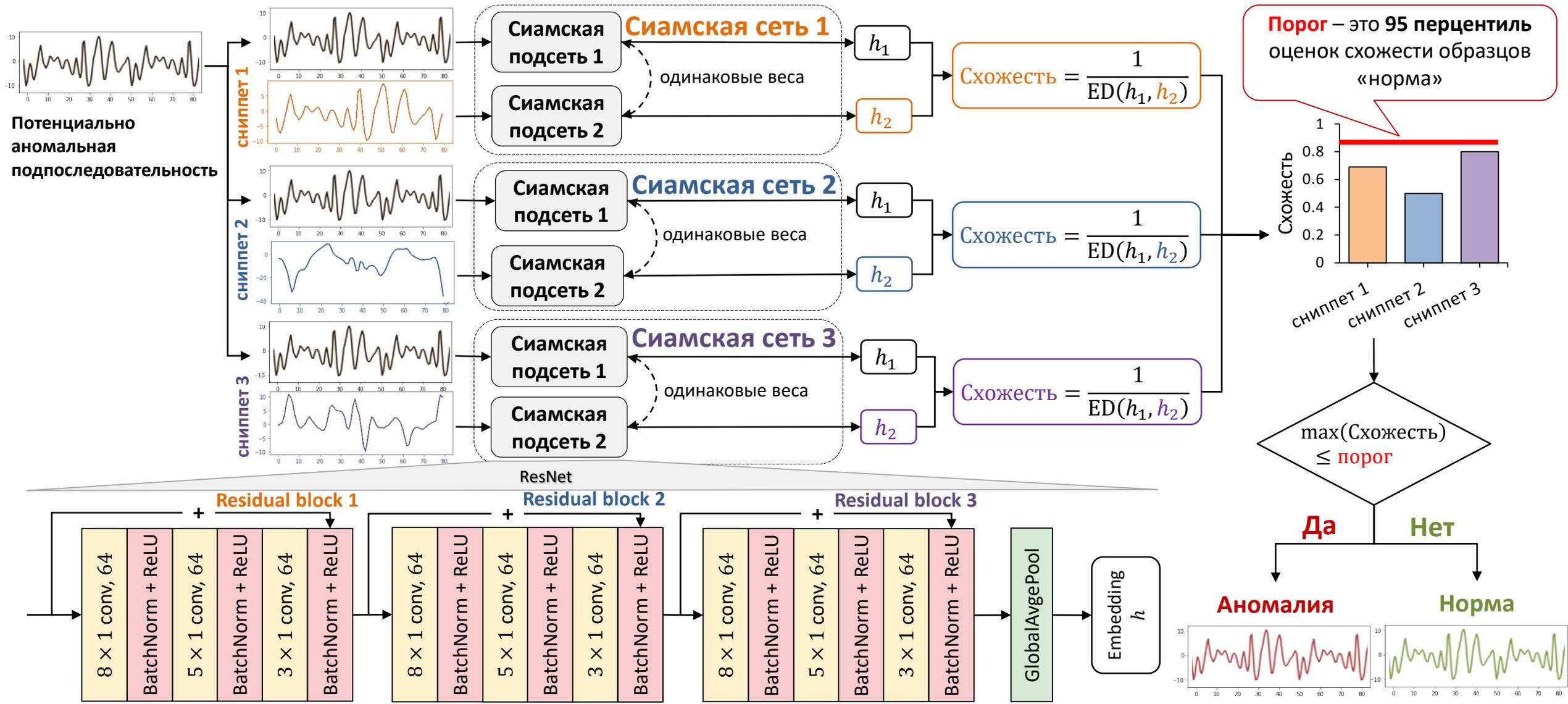
Перемещение нижнего гидроцилиндра системы осевой сдвижки валков SVC



## Шаблоны данных (сниппеты)

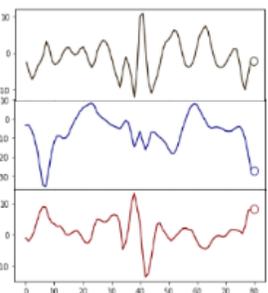


# Нейросетевая модель выявления аномалий временного ряда

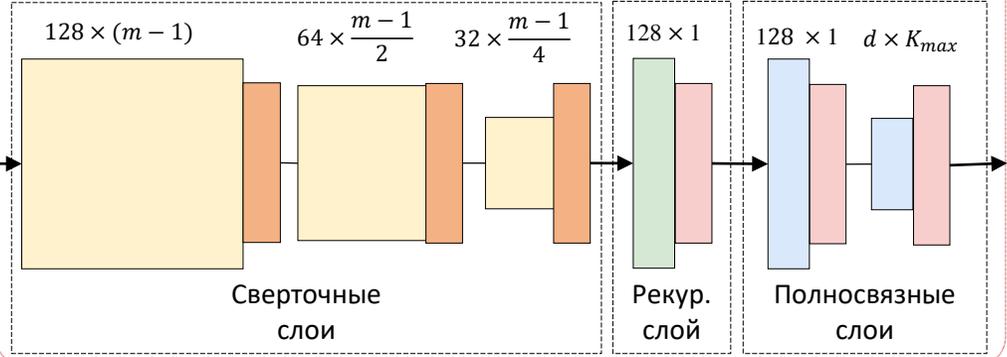


# Нейросетевая модель восстановления пропусков временного ряда

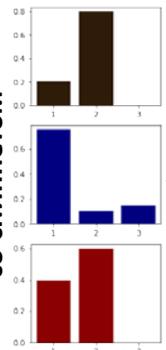
Подпоследовательности с пропусками



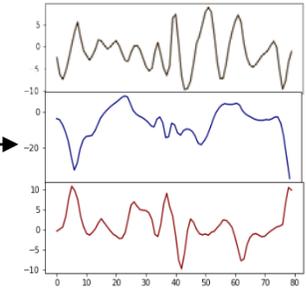
## Распознаватель



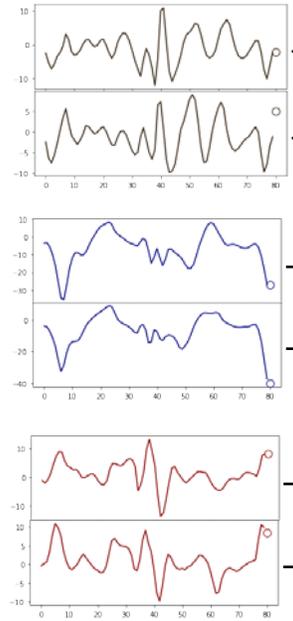
Вероятность сходства со снippetом



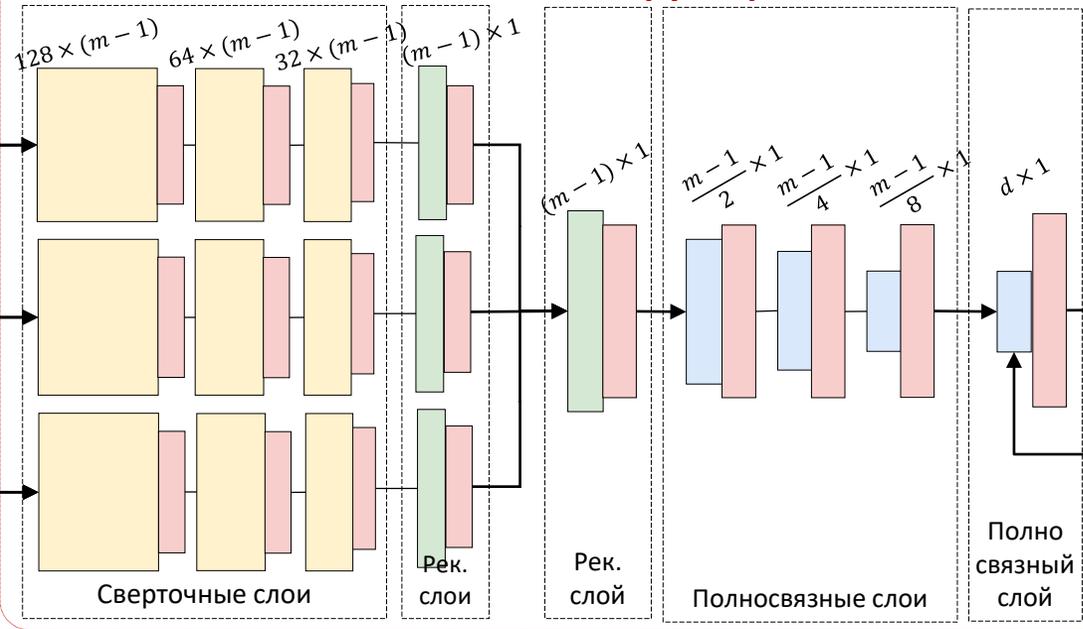
## Сниппеты



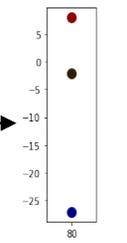
Подпоследовательности с пропущенными значениями



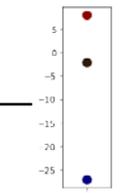
## Реконструктор



Восстановленные значения



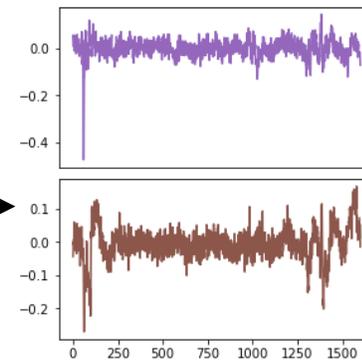
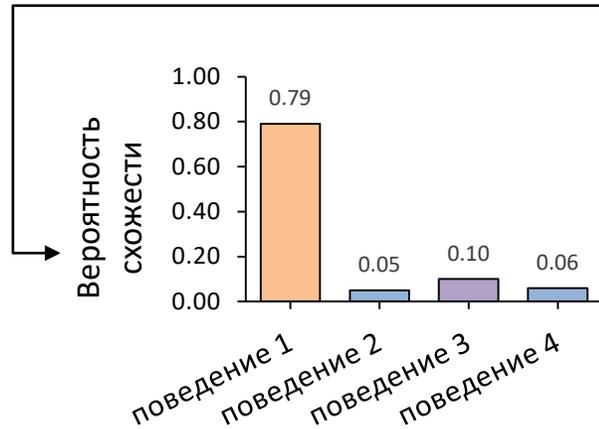
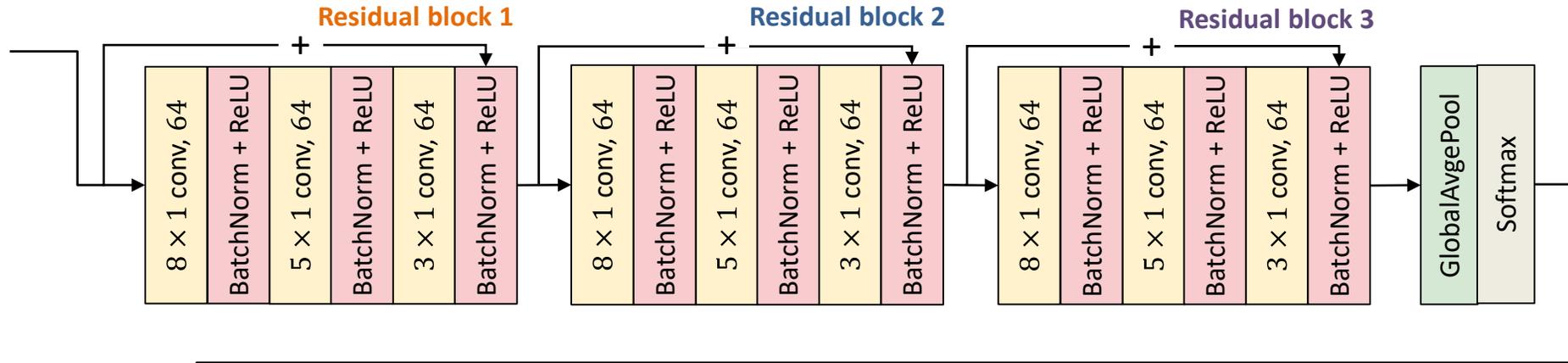
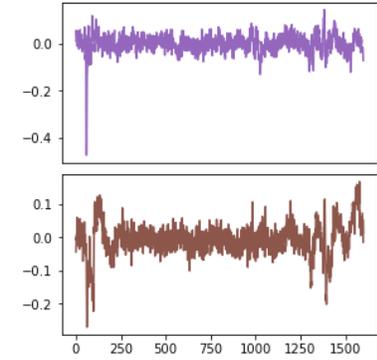
Последние точки сниппетов



- Conv1D
- MaxPool+ReLU
- Leaky ReLU
- GRU
- Fully connected

# Нейросетевая модель распознавания режимов работы

Показания датчиков



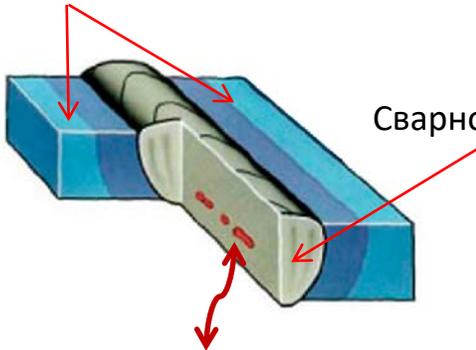
поведение 1

# Выявление дефектов труб в процессе сварки

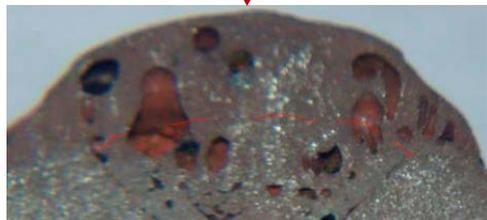
Электродуговая сварка промышленных труб



Кромки металлического листа

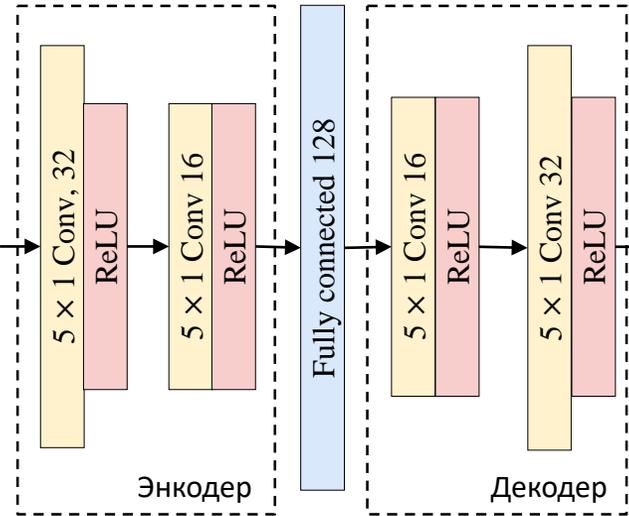
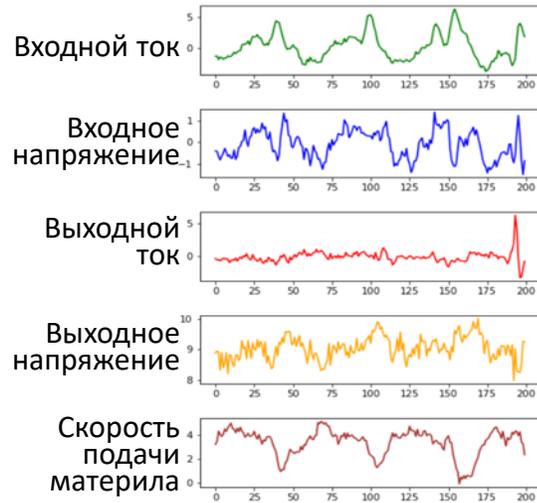


Сварной шов



Шлаковые включения

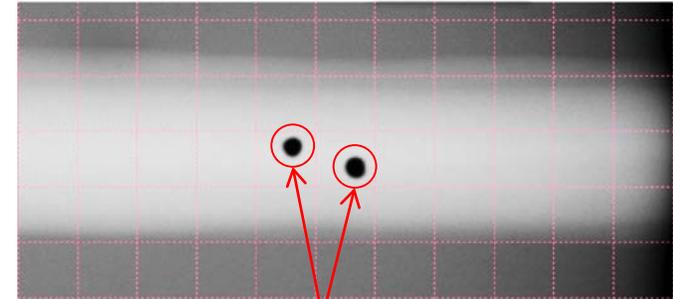
Данные датчиков сварочной установки за 25 с



Сварной шов без видимых дефектов

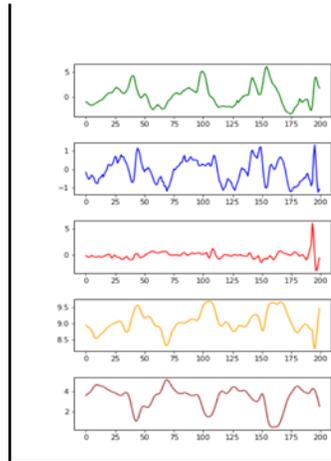


Рентгеновский снимок сварного шва



Шлаковые включения

Прогноз модели

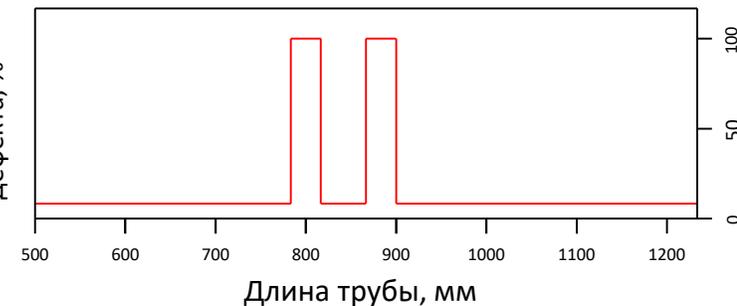


Сравнение реальных и прогнозных данных

MAE

Вероятность возникновения дефекта, %

Прогноз дефекта



# Мы готовы к сотрудничеству

- Разработка параллельных алгоритмов и нейросетевых моделей для анализа временных рядов цифровой индустрии

- Поиск аномалий и шаблонов
- Восстановление и прогноз
- ...

- Контакты:  
Михаил Леонидович Цымблер,  
[mzym@susu.ru](mailto:mzym@susu.ru)



Лаборатория больших данных и машинного обучения (ЛБДМО)



**М.Л. Цымблер**  
д.ф.-м.н.,  
зам. зав. ЛБДМО



**Я.А. Краева**  
программист, аспирант  
обладатель стипендий  
Президента РФ,  
Заксобрания области



**А.И. Гоглачев**  
программист, аспирант  
лауреат конкурса  
«Молодая наука  
ЮУрГУ»



**А.А. Юртин**  
программист, аспирант  
победитель конкурса  
«УМНИК»