Lake Baikal Summit 2024
6-7 June 2024, Moscow, Russia

Parallel unsupervised algorithms and deep learning models for anomaly detection and load prediction in large computing systems

明智是了解事件的人
Wise is the person who understands events.
Chinese proverb

Andrey Goglachev, Yana Kraeva, Alexey Yurtin, and Mikhail Zymbler

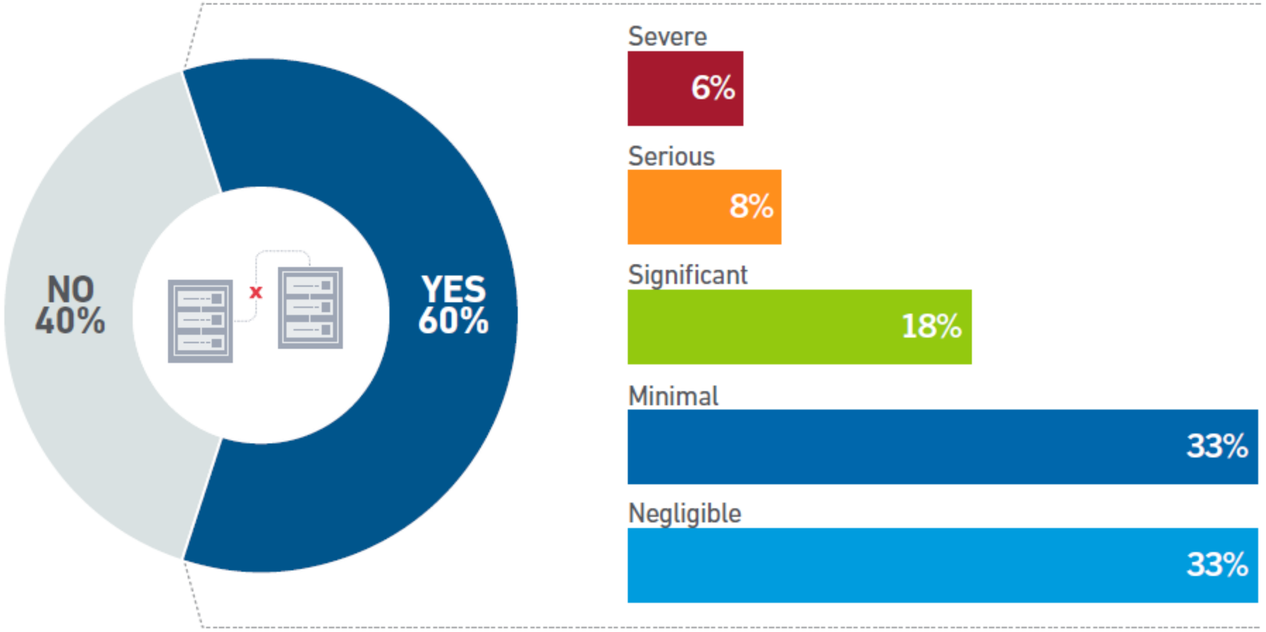
{goglachevai, kraevaya, iurtinaa, mzym}@susu.ru

Big Data and Machine Learning Lab, South Ural State University, Chelyabinsk, Russia

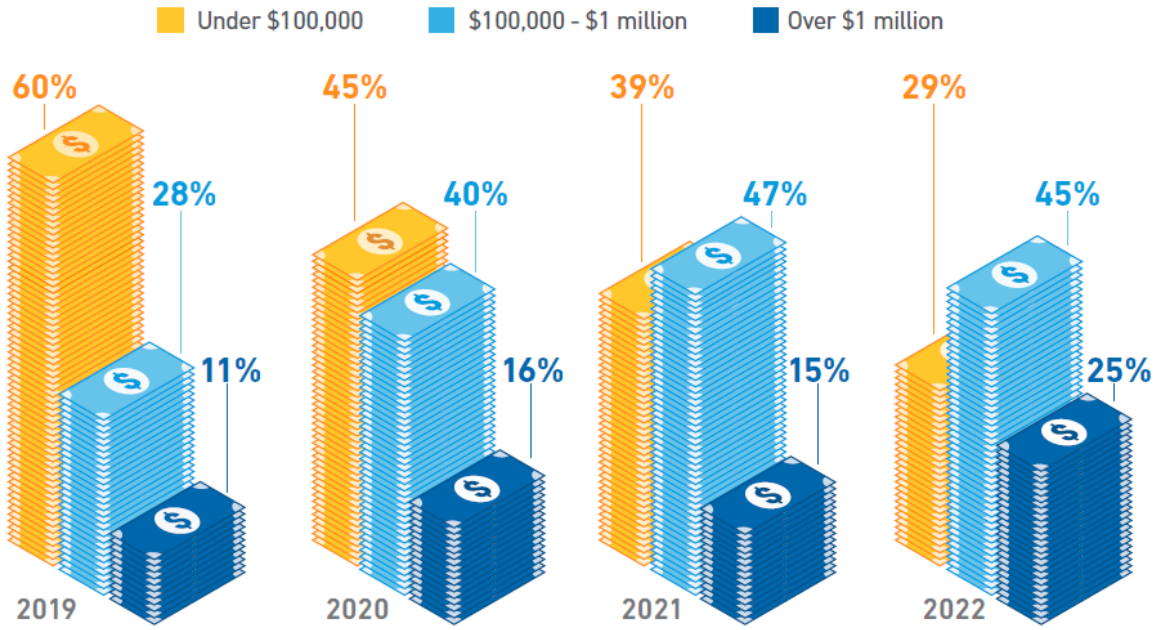
This work is financially supported by the Russian Science Foundation (grant No. 23-21-00465)

Failures and outages in data centers is a serious problem[1)]

Most organizations experienced failure/outage in recent years



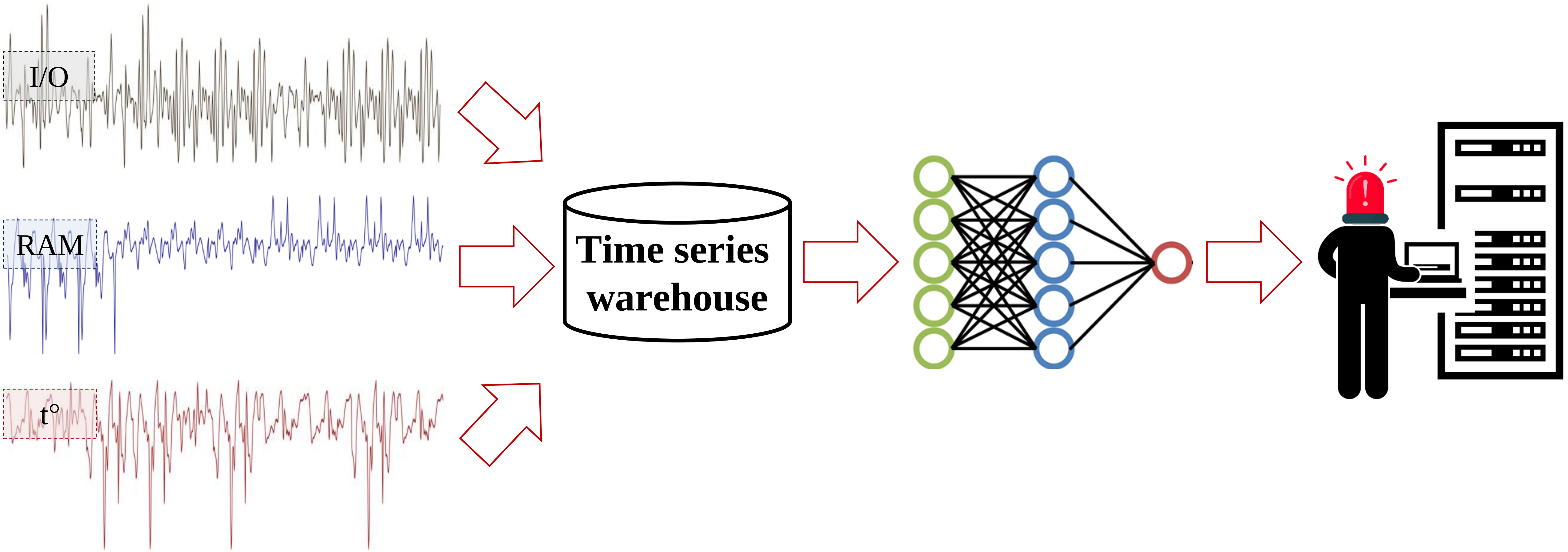
Costs of failure/outage growth (proportion of over $100K loss cases increases)



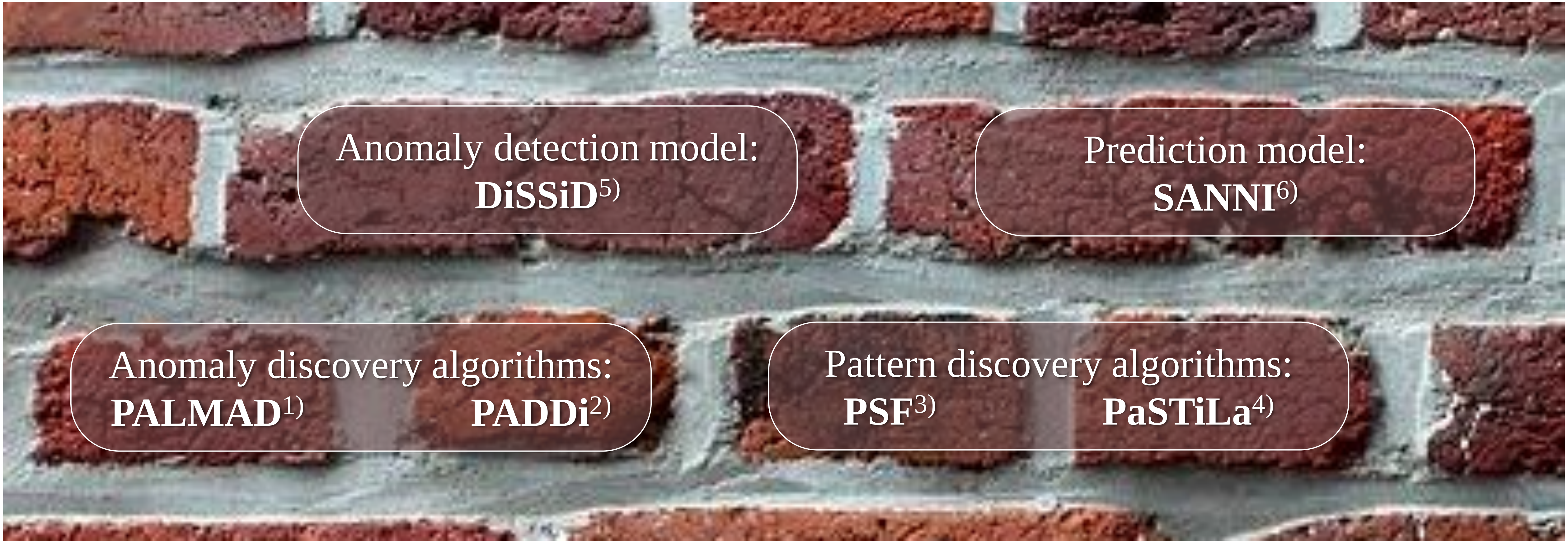
[1)] Annual outages analysis 2023: The causes and impacts of IT and data center outages in USA. Uptime Institute. [URL]

Tackle the problem: online prediction of anomaly/workload

Collect time series on workload | **Build and learn the model(s)** | **Apply the model and treat findings**



Unsupervised parallel algorithms & Deep learning models



[1)] Zymbler M., Kraeva Y. High-performance time series anomaly discovery on graphics processors. Mathematics. 2023. 11(14), 3193. DOI: 10.3390/math11143193.
[2)] Kraeva Y., Zymbler M. Anomaly detection in long time series on high-performance cluster with GPUs. Num. Meth. & Progr. 2023. 24(3), 291-304. DOI: 10.26089/NumMet.v24r320.
[3)] Zymbler M., Goglachev A. Fast summarization of long time series with graphics processor. Mathematics. 2022. 10(10). 1781. DOI: 10.3390/math10101781.
[4)] Zymbler M., Goglachev A. PaSTiLa: Scalable parallel algorithm for unsupervised labeling of long time series. LJM. 2024. 45(3), 1333-1347. DOI: 10.1134/S1995080224600766.
[5)] Kraeva Ya. Detection of time series anomalies based on data mining and neural network technologies. Bulletin of SUSU, CMSE. 2023. 12(3). 50-71. DOI: 10.14529/cmse230304.
[6)] Zymbler M., Yurtin A. Imputation of missing values of a time series based on joint application of analytical algorithms and neural networks. Num. Meth. & Progr. 2023. 24 (3), 243-259. DOI: 10.26089/NumMet.v24r318.

Unsupervised parallel algorithms for **anomaly discovery**

We formalize a time series anomaly as a ***discord*** and discover discords in parallel:

- **PALMAD**[1] discovers discords on a GPU
- **PADDi**[2] discovers discords on a multi-GPU cluster

[1] **Zymbler M., Kraeva Y. High-performance time series anomaly discovery on graphics processors. Mathematics. 2023. 11(14), 3193. DOI: [10.3390/math11143193](https://doi.org/10.3390/math11143193).**

[2] **Kraeva Y., Zymbler M. Anomaly detection in long time series on high-performance cluster with GPUs. Num. Meth. & Progr. 2023. 24(3), 291-304. DOI: [10.26089/NumMet.v24r320](https://doi.org/10.26089/NumMet.v24r320).**

[3] Zymbler M., Goglachev A. Fast summarization of long time series with graphics processor. Mathematics. 2022. 10(10). 1781. .

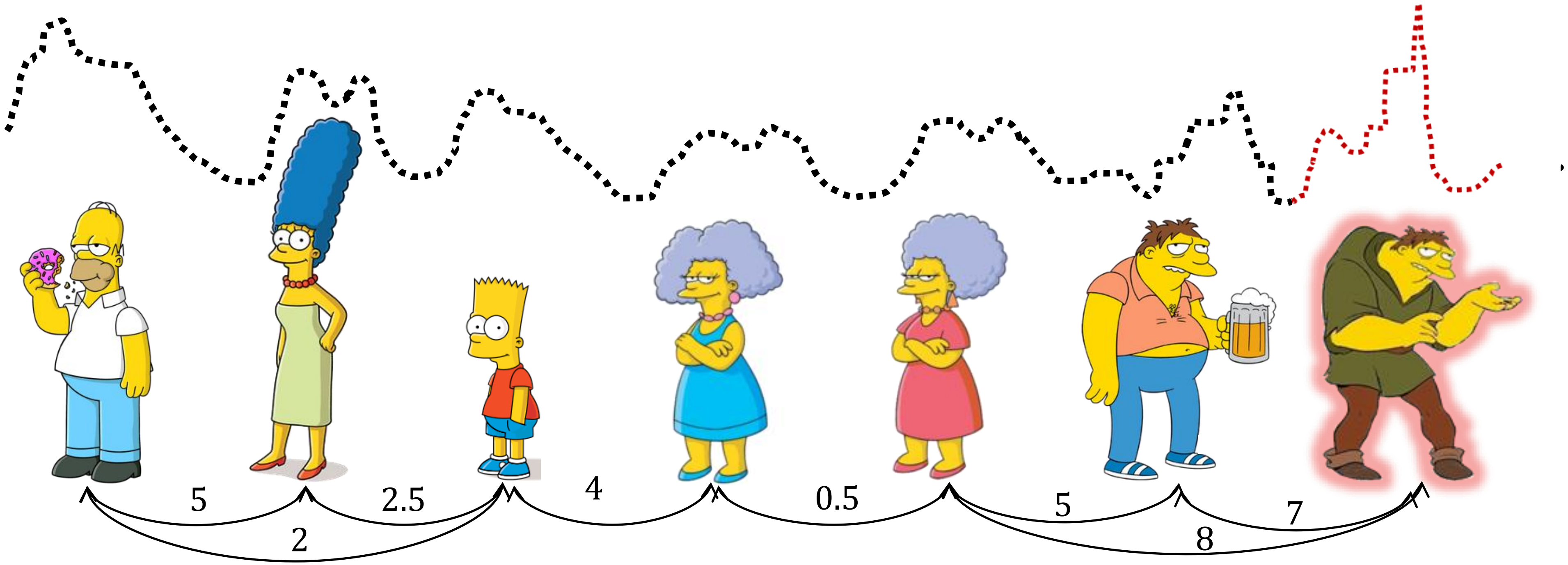
4) Zymbler M., Goglachev A. PaSTiLa: Scalable parallel algorithm for unsupervised labeling of long time series. LJM. 2024. 45(3), 1333-1347. .

5) Kraeva Ya. Detection of time series anomalies based on data mining and neural network technologies. Bulletin of SUSU, CMSE. 2023. 12(3). 50-71.

6) Zymbler M., Yurtin A. Imputation of missing values of a time series based on joint application of analytical algorithms and neural networks. Num. Meth. & Progr. 2023. 24 (3), 243-259.

Discord formalizes anomaly of any domain

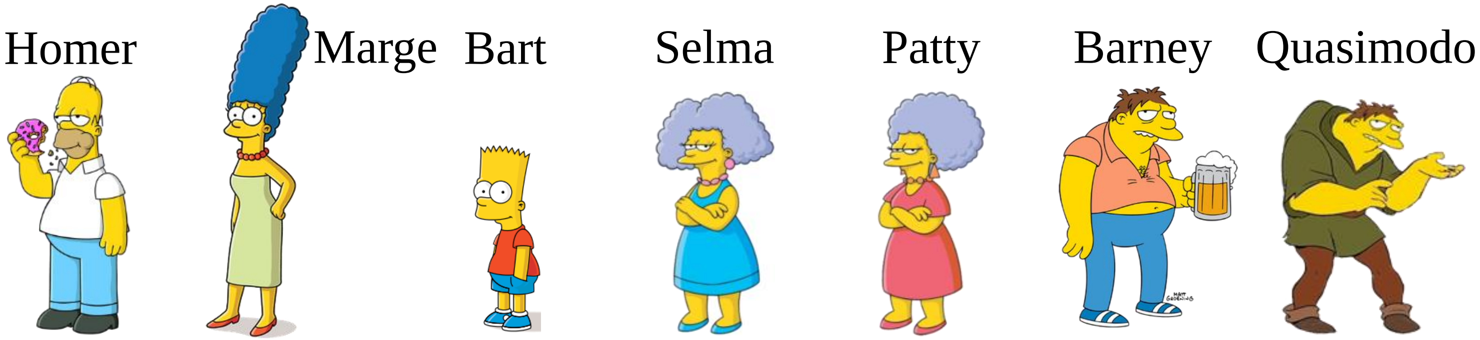
- *Discord*[1)] is the given-length subsequence whose distance to its nearest neighbor is greatest
- *Nearest neighbor* is the same-length subsequence whose distance to the given subsequence is smallest



[1)] Keogh E. *et al.* HOT SAX: Efficiently finding the most unusual time series subsequence. ICDM 2005. pp. 226-233. DOI: 10.1109/ICDM.2005.79

Discord concept

Distance matrix:
the close neighbors,
the similar they are



	Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
	0						
		0					
			0				
				0			
					0		
						0	
							0

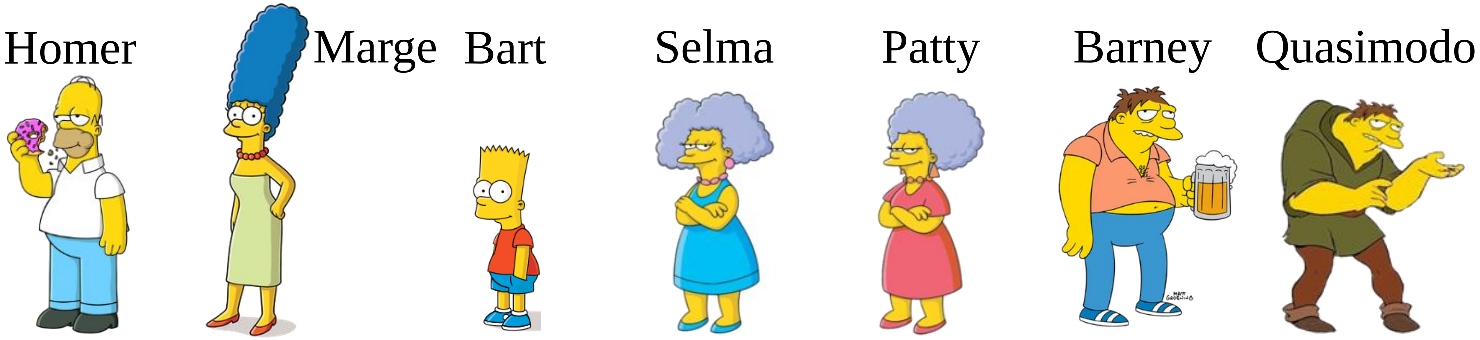
Discord concept

Distance matrix
with calculated distances
to neighbors

Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	7
8	10	9	8	8	7	0

Discord concept

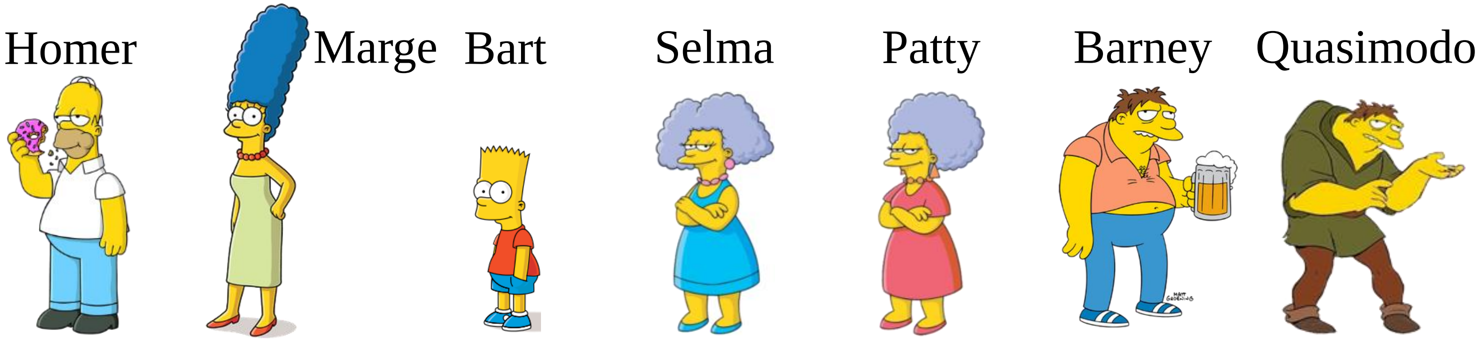
Distance matrix
with
distances to their nearest neighbors
(i.e. column-wise minima)



Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
0	5	**2**	4	4	6	8
5	0	2.5	3	3	6	10
2	**2.5**	0	4	4	6	9
4	3	4	0	**0.5**	**5**	8
4	3	4	**0.5**	0	**5**	8
6	6	6	5	5	0	**7**
8	10	9	8	8	7	0

Discord concept

Distance matrix
with the
farthest distance
to the nearest neighbor
(i.e. maximum
among
column-wise minima)



Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	7
8	10	9	8	8	7	0

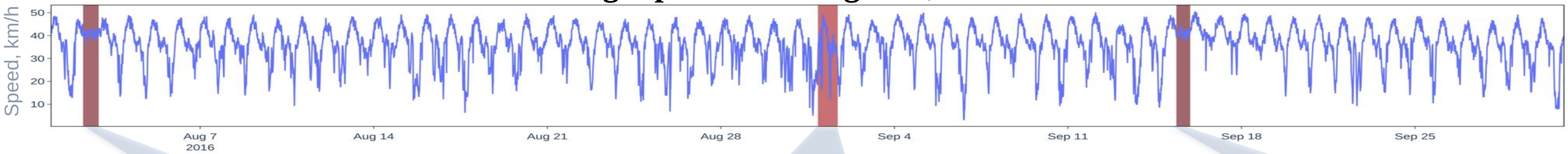
Discord concept

Discord is an object with the **farthest nearest neighbor** (i.e. argument of the maximum among column-wise minima)

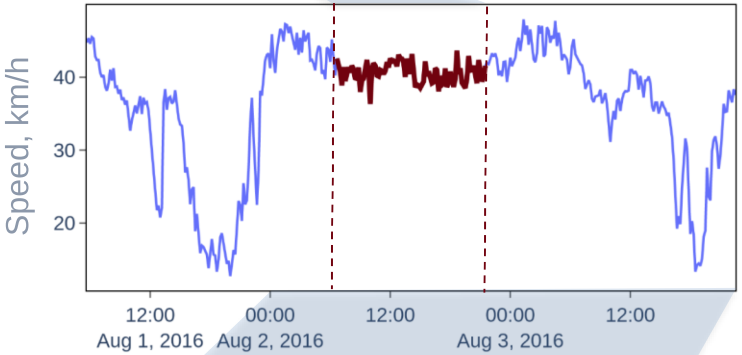
Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	7
8	10	9	8	8	7	0

PALMAD and PADDi grab anomalies in real time series

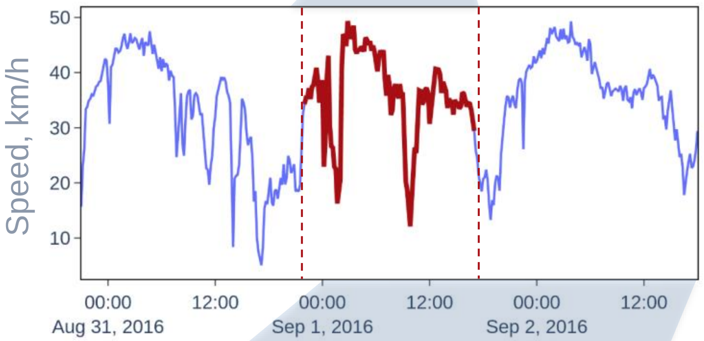
Average speed in Guangzhou, China[1)]



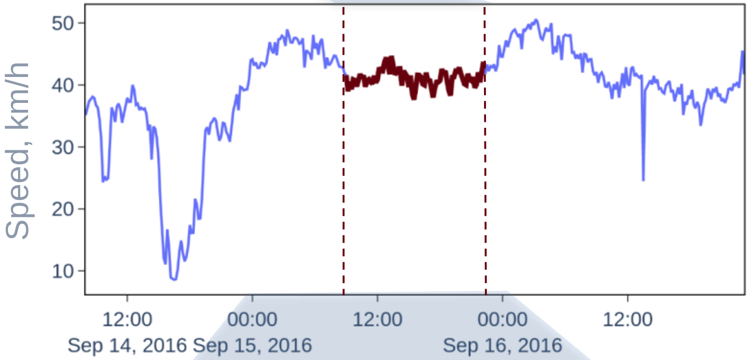
Top-2 discord



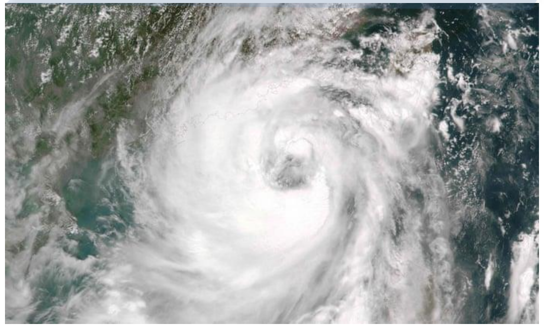
Top-3 discord



Top-1 discord



Typhoon Nida



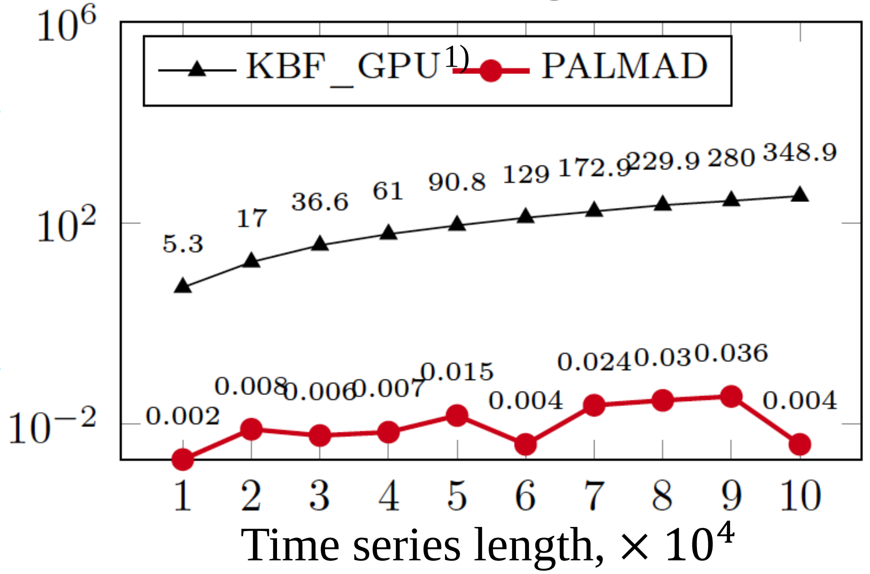
Day of Victory over Japan

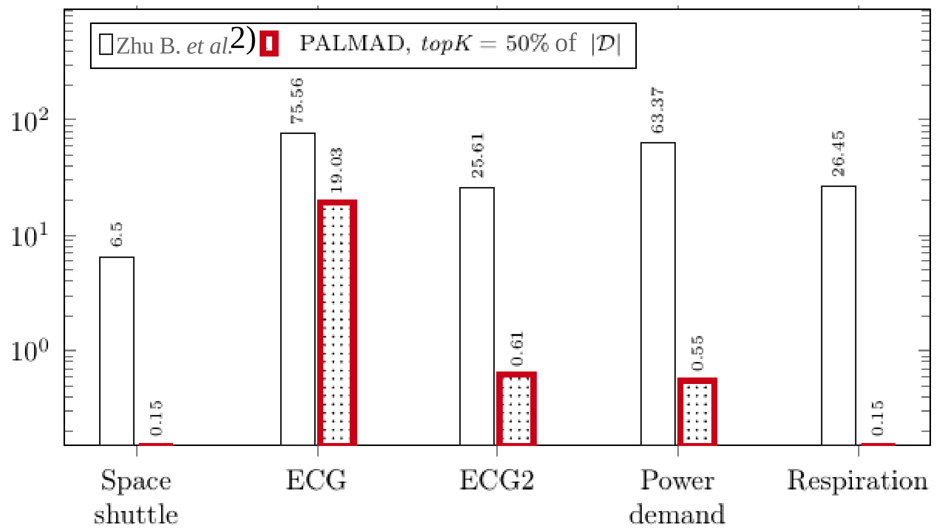
Mid-Autumn Festival

[1)] Chen X, Chen Y, He Z. Urban traffic speed dataset of Guangzhou, China. 2018. DOI: 10.5281/zenodo.1205229.

PALMAD and PADDi outperform S.O.T.A. analogs

Average runtime per discord on a single GPU

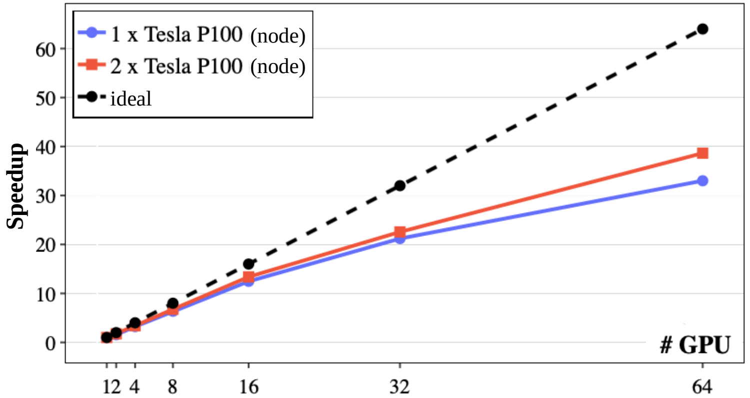




[1)] Thuy T.T.H. *et al.* A new discord definition and an efficient time series discord detection method using GPUs. ICSED 2021. pp. 63-70. DOI: 10.1145/3507473.3507483

[2)] Zhu B. *et al.* A GPU acceleration framework for motif and discord based pattern mining. IEEE TPDS. 2021. 32(8). 1987-2004. DOI: 10.1109/TPDS.2021.3055765

PADDi is the only algorithm for discord discovery on HPC clusters with multi-GPU nodes



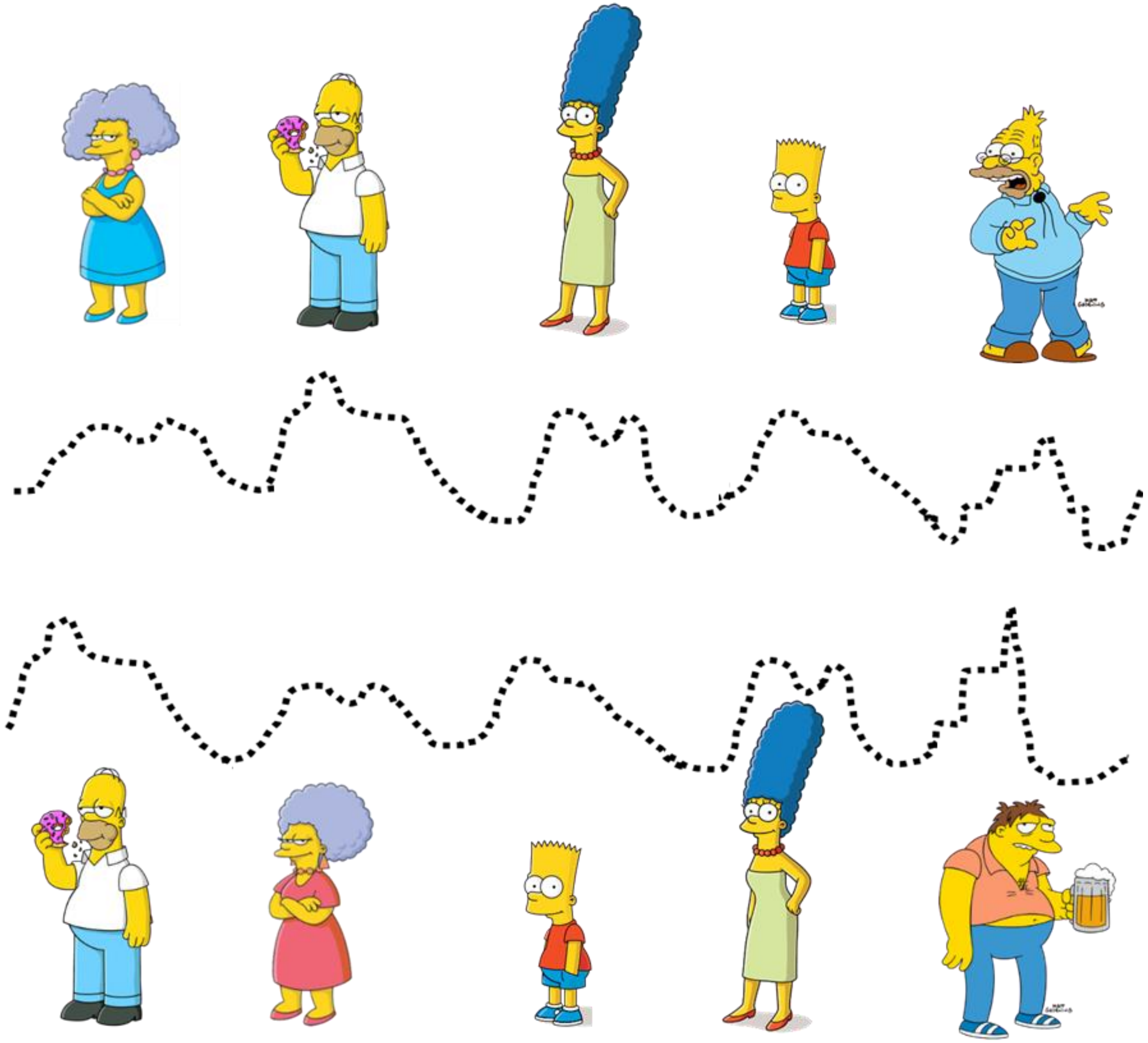
Unsupervised parallel algorithms for **pattern discovery**

We formalize a time series pattern as a ***snippet*** and discover snippets in parallel:

- **PSF**[3] discovers snippets on a GPU
- **PaSTiLa**[4] discovers snippets on a multi-GPU cluster

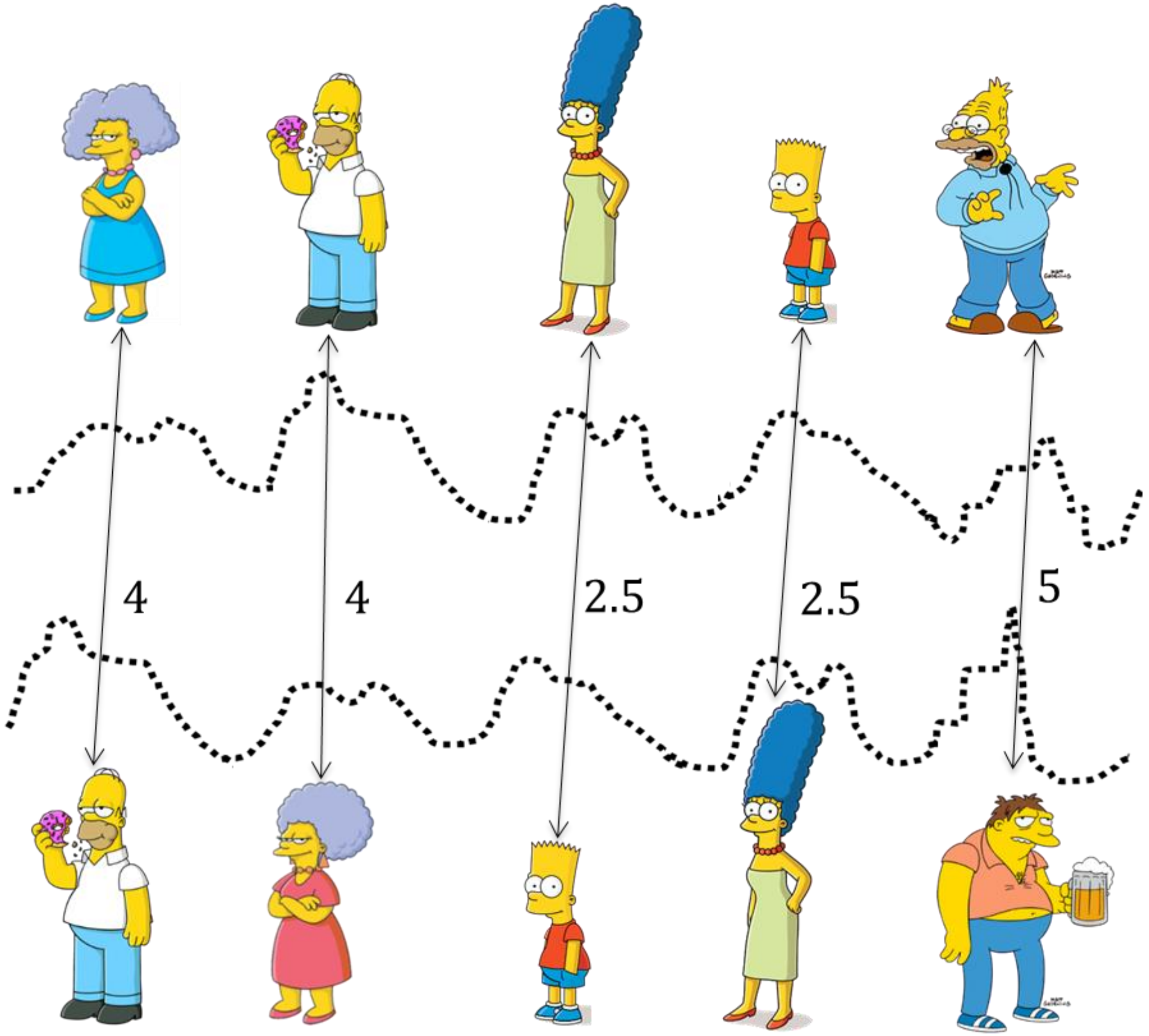
1) Zymbler M., Kraeva Y. High-performance time series anomaly discovery on graphics processors. Mathematics. 2023. 11(14), 3193.
2) Kraeva Y., Zymbler M. Anomaly detection in long time series on high-performance cluster with GPUs. Num. Meth. & Progr. 2023. 24(3), 291-304.
3) Zymbler M., Goglachev A. Fast summarization of long time series with graphics processor. Mathematics. 2022. 10(10). 1781. DOI: 10.3390/math10101781.
4) Zymbler M., Goglachev A. PaSTiLa: Scalable parallel algorithm for unsupervised labeling of long time series. LJM. 2024. 45(3), 1333-1347. DOI: 10.1134/S1995080224600766.
5) Kraeva Ya. Detection of time series anomalies based on data mining and neural network technologies. Bulletin of SUSU, CMSE. 2023. 12(3). 50-71.
6) Zymbler M., Yurtin A. Imputation of missing values of a time series based on joint application of analytical algorithms and neural networks. Num. Meth. & Progr. 2023. 24 (3), 243-259.

Similarity measure for time series



Euclidean distance is for the structural similarity

Euclidean distance ≈ 8



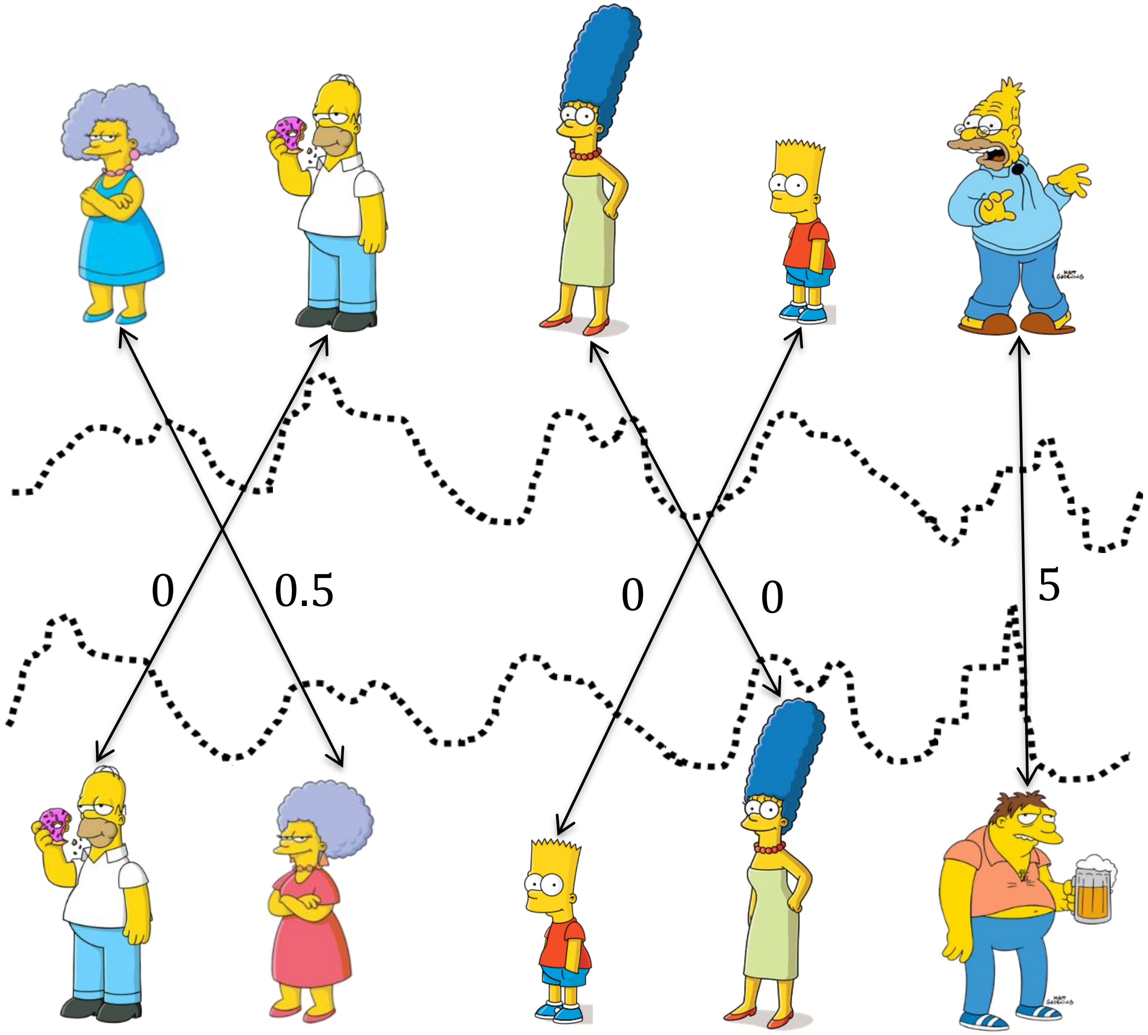
Structural similarity compares time series point-by-point

Complexity: $\boldsymbol{O(n)}$

$$\mathrm{ED}(A,B)=\sqrt{\sum_{i=1}^{n}(a_i-b_i)^2}$$

MPdist distance is for the behavioral similarity

MPdist[1] distance $\approx$ **0**



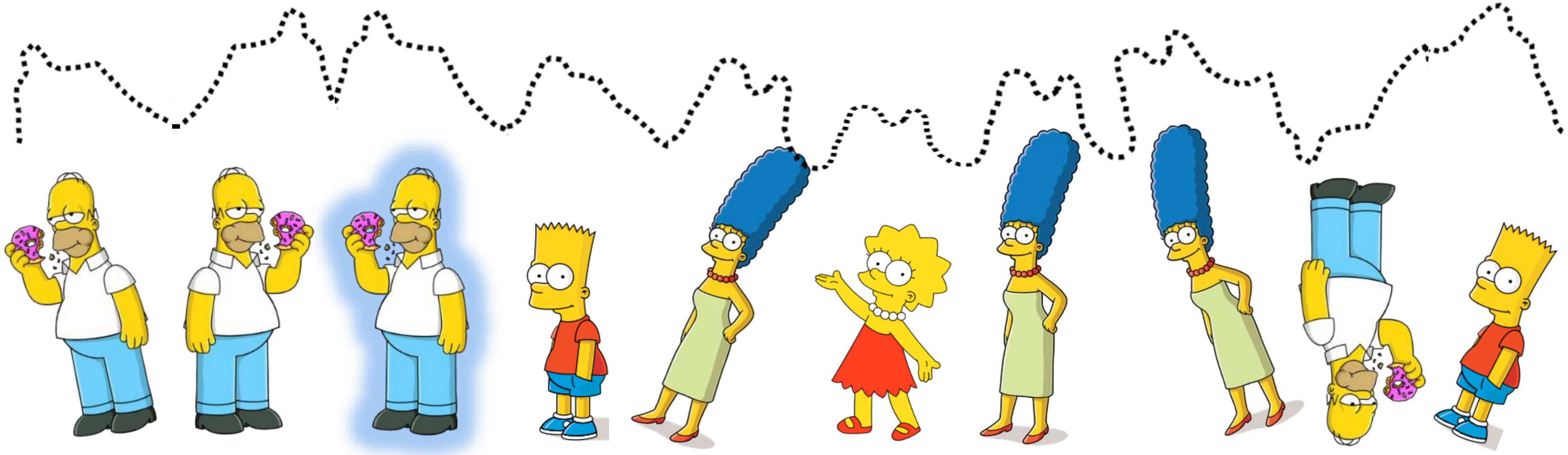
Behavioral similarity is proportional to the number of subsequences that are close w.r.t. the Euclidean distance (no matter their locations)

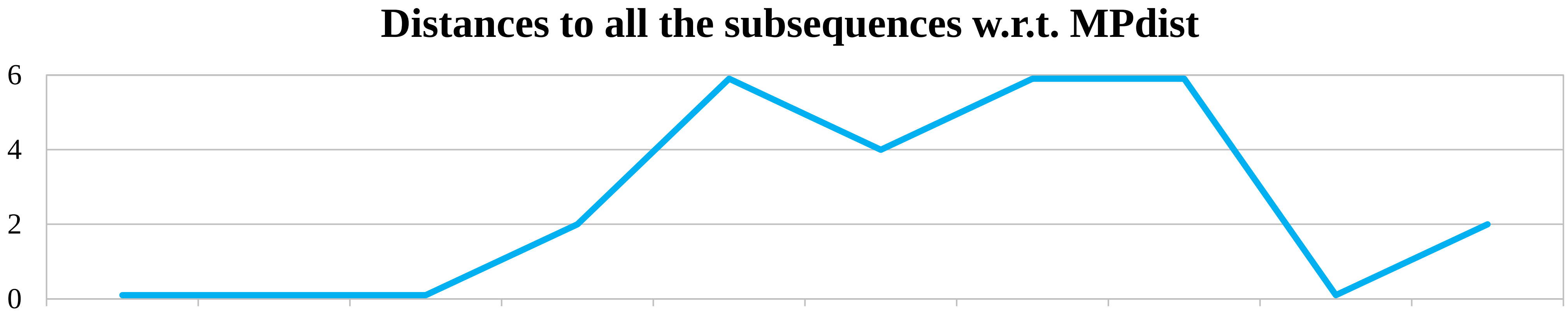
Complexity: $O(n^2)$

$$\text{MPdist}_\ell^k(A,B) = \text{AscSort}(P_{ABBA})(k),$$
$$P_{ABBA} = P_{AB} \cdot P_{BA},$$
$$P_{AB} = \{\text{ED}^2_{\text{norm}}(A_{i,\ell}, B_{j,\ell})\}_{i=1}^{n-\ell+1},$$
$$B_{j,\ell} = \arg \min_{1 \le q \le n-\ell+1} \text{ED}^2_{\text{norm}}(A_{i,\ell}, B_{q,\ell}),$$
$$\text{where } 0 < k < n,\ k_{\text{default}} = \lceil 0.1n \rceil$$

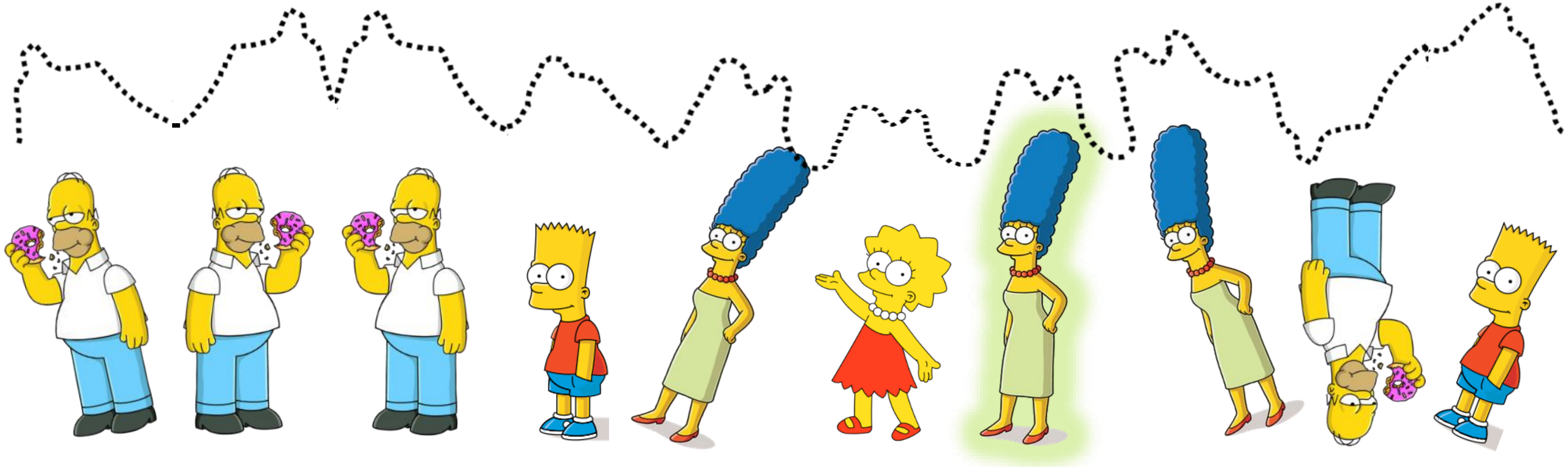
[1] Gharghabi S. *et al*. An ultra-fast time series distance measure to allow data mining in more complex real-world deployments. DMKD. 2020. (34). pp. 1104-1135. DOI: 10.1007/s10618-020-00695-8

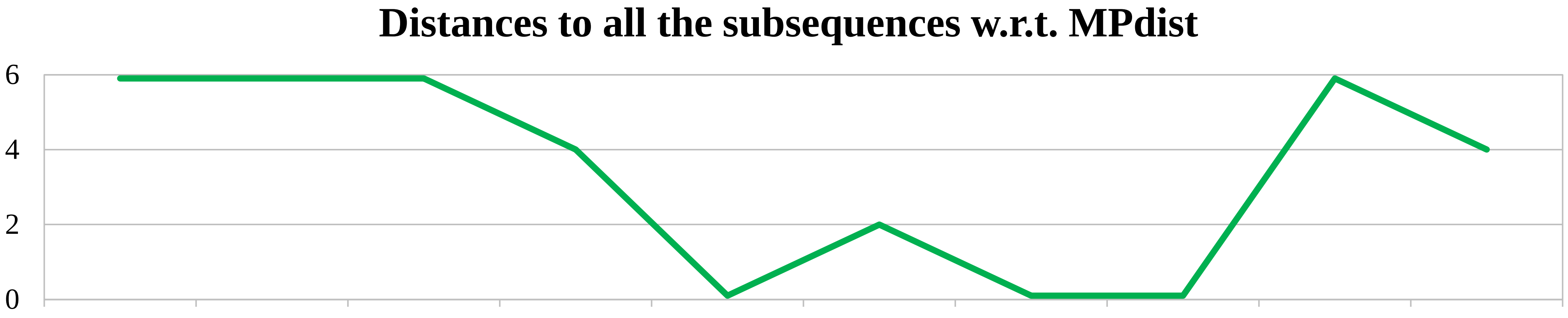
Distance profile of the potential behavioral pattern



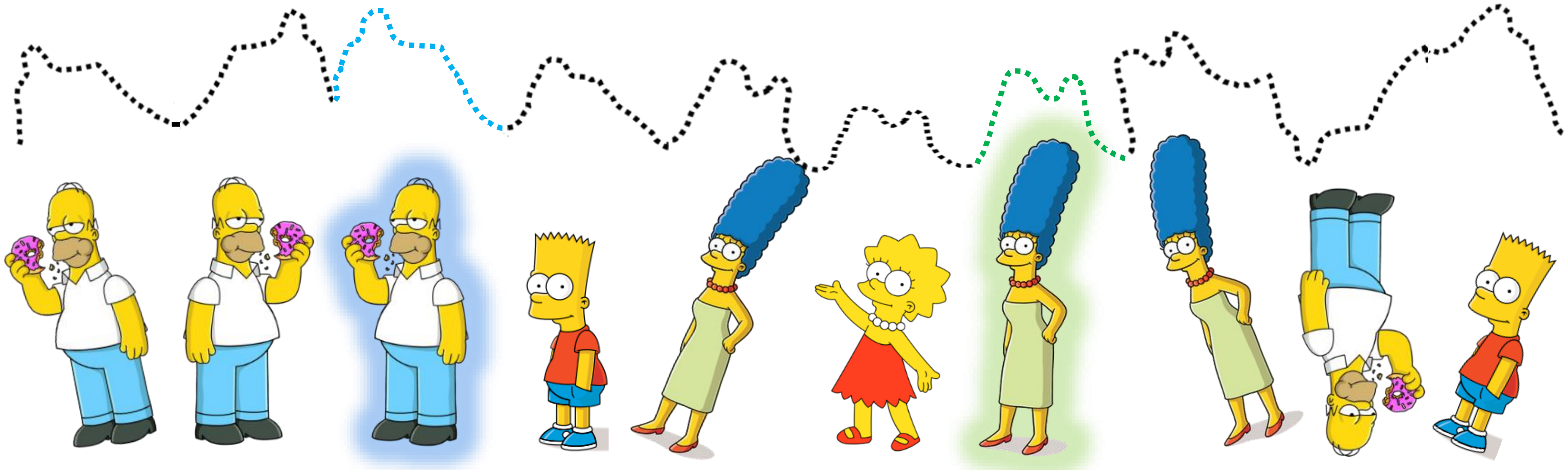


Distance profile of the potential behavioral pattern

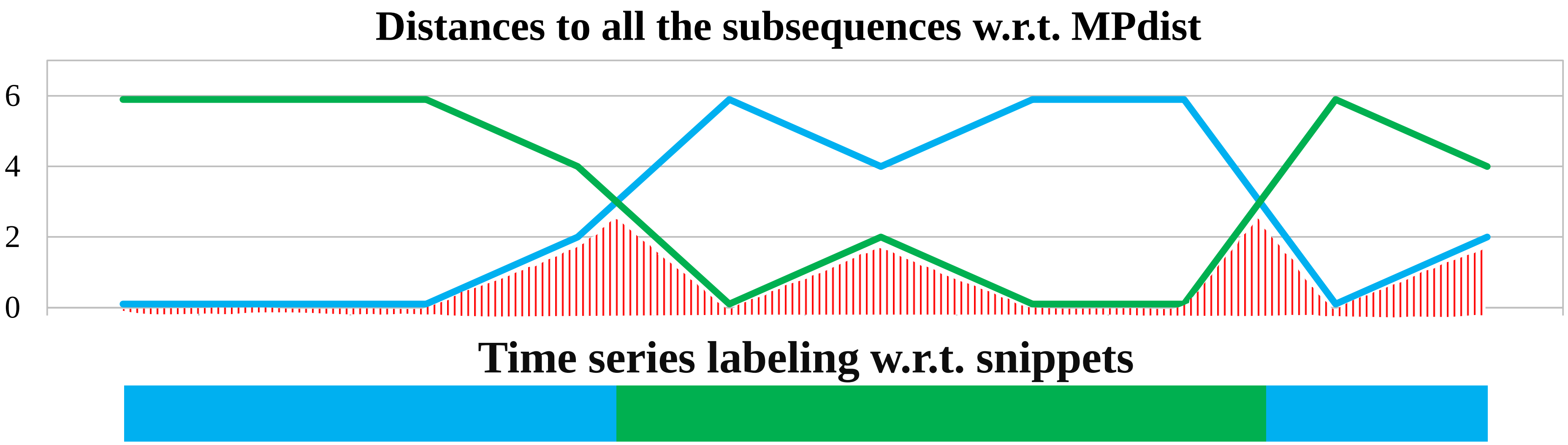


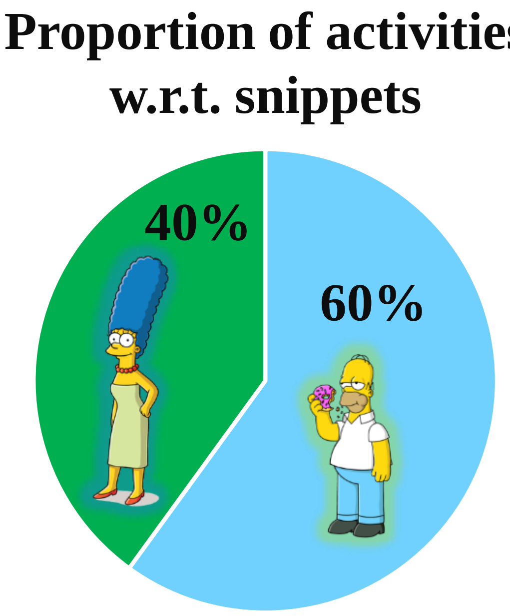


Snippets formalize behavioral patterns of any domain



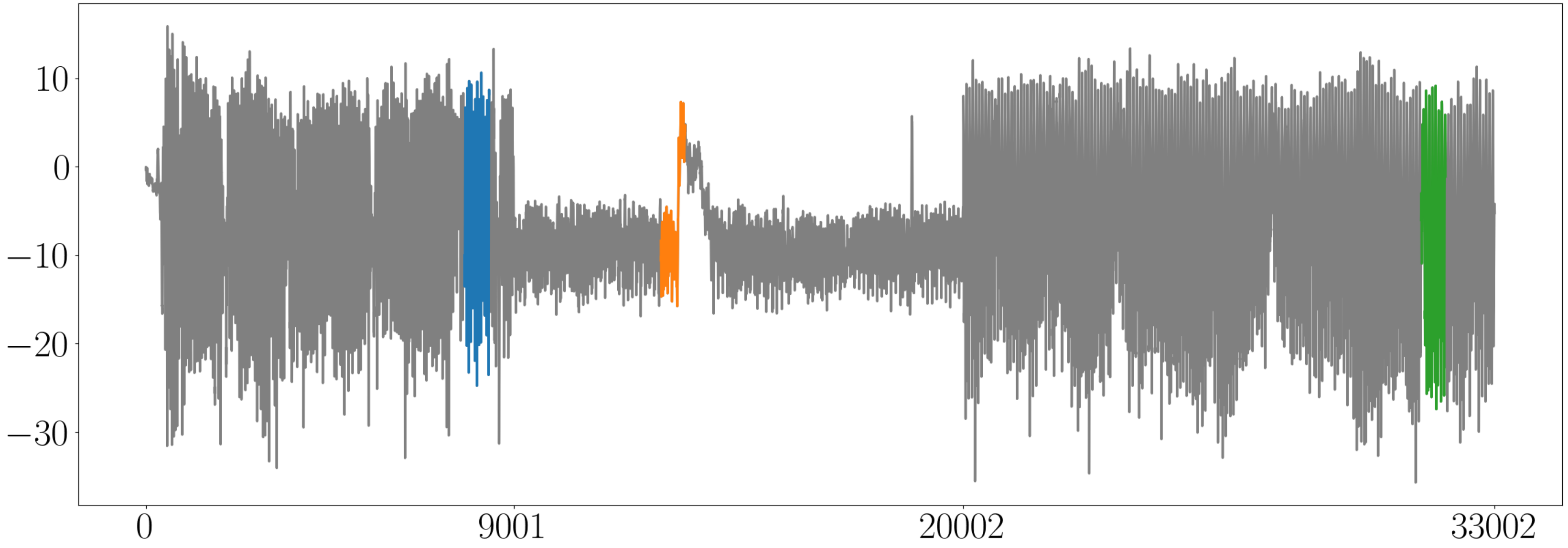
Snippets are segments for which the area under the MPdist curve is minimal



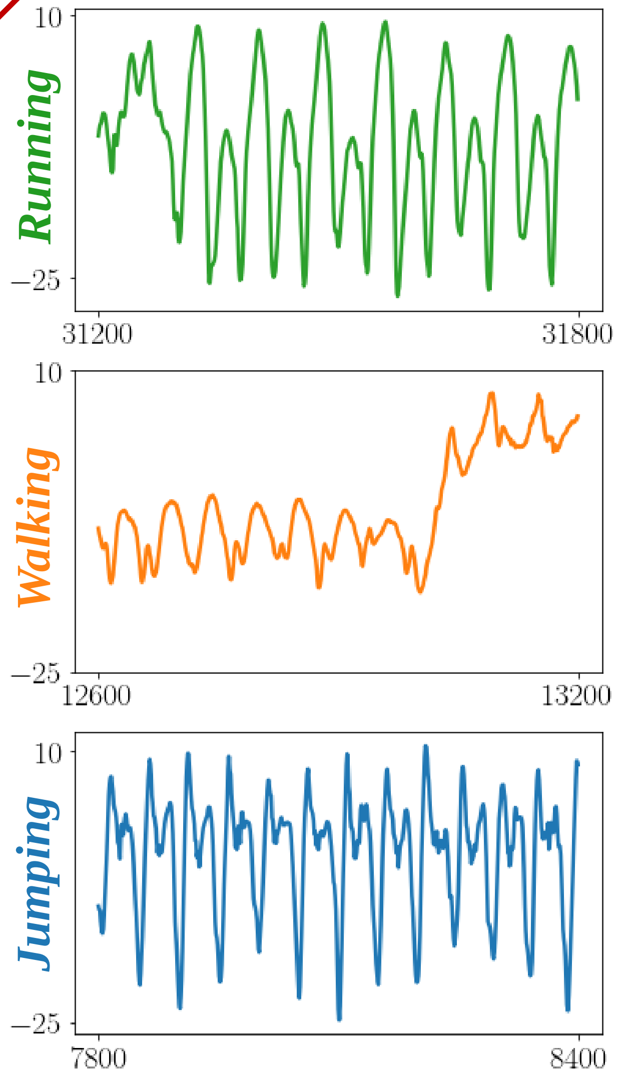


PSF and PaSTiLa grab behavioral patterns in real time series

Measurements of a wearable accelerometer during an athlete's training



Snippets discovered

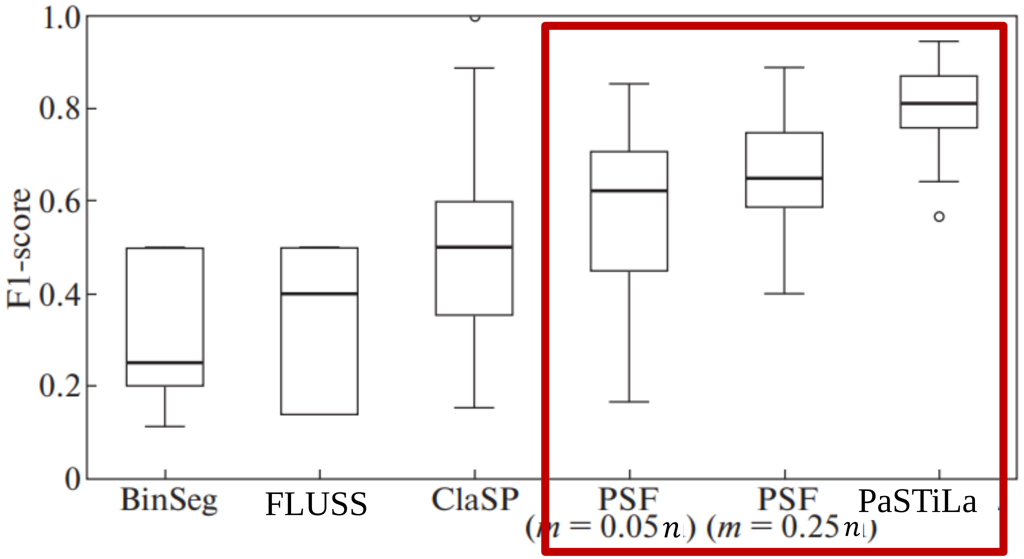


Labeling of the training w.r.t. snippets

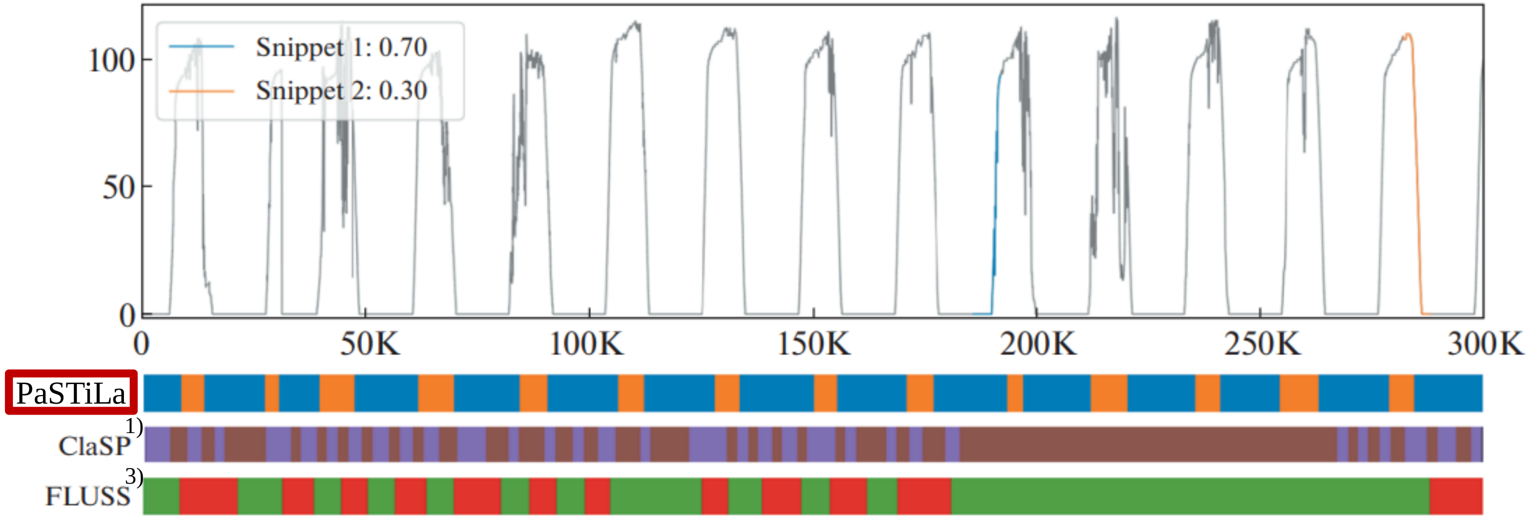
Training schedule (ground truth)

PSF and PaSTiLa outperform S.O.T.A. analogs

Accuracy over pre-labeled data,
Time Series Segmentation Benchmark[1)]



Accuracy over unlabeled data,
Solar Power time series[4)]



Our approach is more accurate in determining the day-night cycles

1) Ermshaus A. *et al*. ClaSP: Parameter-free time series segmentation. Data Min. Knowl. Discov. 37, 1262–1300 (2023). DOI: 10.1007/S10618-023-00923-X
2) Truong C. *et al*. Selective review of offline change point detection methods. Signal Process 167, 107299 (2020). DOI: 10.1016/J.SIGPRO.2019.107299
3) Gharghabi S. *et al*. Domain agnostic online semantic segmentation for multi-dimensional time series. Data Min. Knowl. Discov. 33, 96–130 (2019). DOI: 10.1007/S10618-018-0589-3
4) Rakshitha G. *et al*. Solar Power Dataset (4 Seconds Observations). DOI: 10.5281/zenodo.4656027.

Behavioral patterns are the key to online processing

Preprocessing

Take a representative fragment of time series to be processed → **Discover behavioral patterns** → Build a training set(s) for deep learning model(s)

↓

Online processing

Determine a behavioral pattern to which the subsequence that came from a sensor, is most similar →

Deep learning model for anomaly detection:

How much does the subsequence differ from all the patterns?

Deep learning model for load prediction:

According to all the patterns, what should be the next subsequence?

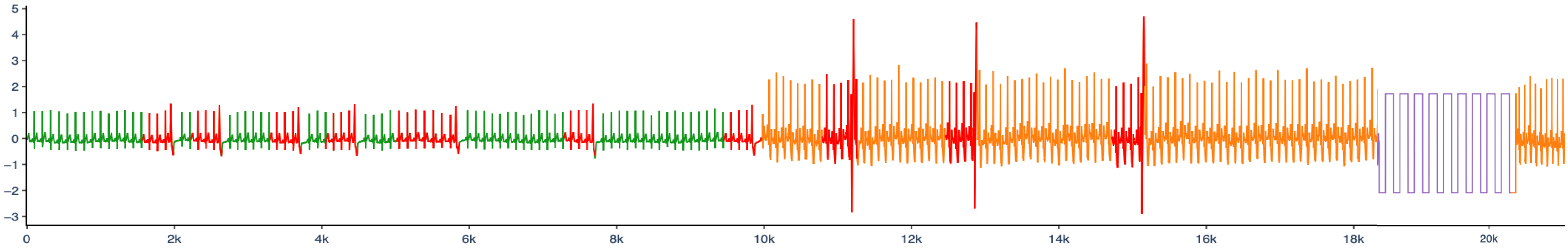
Deep learning model for **online anomaly detection**

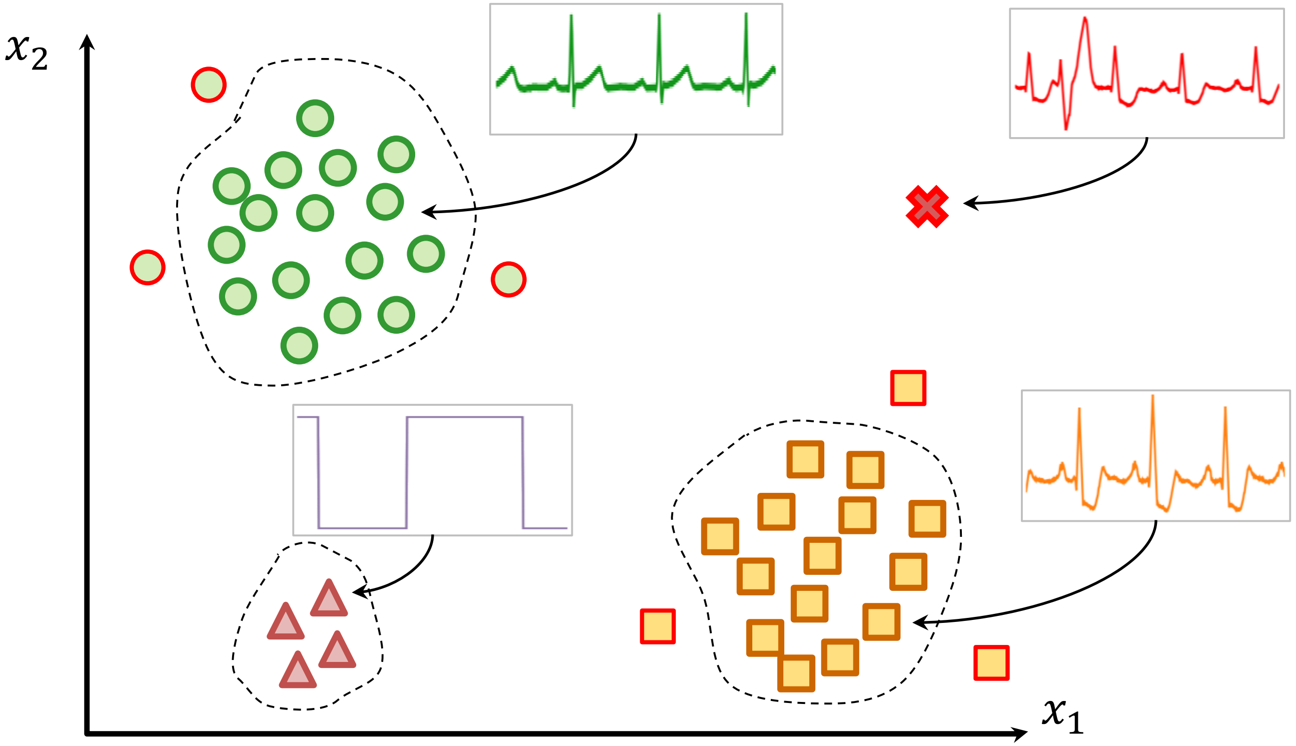
DiSSiD learns to differ subsequences of typical behavior from abnormal ones:

- snippets represent typical behavior
- discords represent abnormal behavior

[1)] Zymbler M., Kraeva Y. High-performance time series anomaly discovery on graphics processors. Mathematics. 2023. 11(14), 3193.
[2)] Kraeva Y., Zymbler M. Anomaly detection in long time series on high-performance cluster with GPUs. Num. Meth. & Progr. 2023. 24(3), 291-304.
[3)] Zymbler M., Goglachev A. Fast summarization of long time series with graphics processor. Mathematics. 2022. 10(10). 1781.
[4)] Zymbler M., Goglachev A. PaSTiLa: Scalable parallel algorithm for unsupervised labeling of long time series. LJM. 2024. 45(3), 1333-1347.
[5)] Kraeva Ya. Detection of time series anomalies based on data mining and neural network technologies. Bulletin of SUSU, CMSE. 2023. 12(3). 50-71. DOI: 10.14529/cmse230304.
[6)] Zymbler M., Yurtin A. Imputation of missing values of a time series based on joint application of analytical algorithms and neural networks. Num. Meth. & Progr. 2023. 24 (3), 243-259.

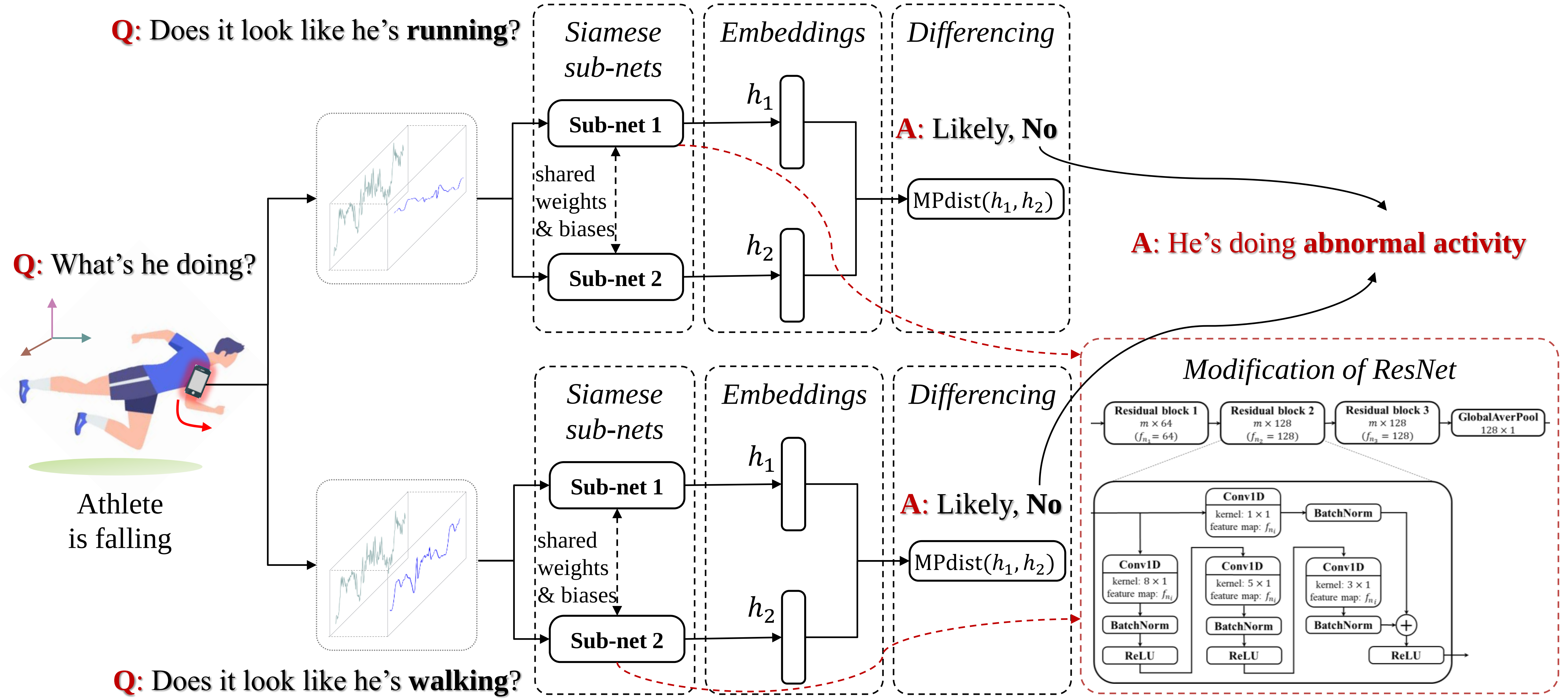
DiSSiD differs normal data from the opposite ones





- typical activities (big-sized sets of snippets and their nearest neighbors)
- rare activity (small-sized sets of snippets and their nearest neighbors)
- noises within activities
- abnormal activity (discords)

DiSSiD: Discord, Snippet, and Siamese Net-based Detector of anomalies

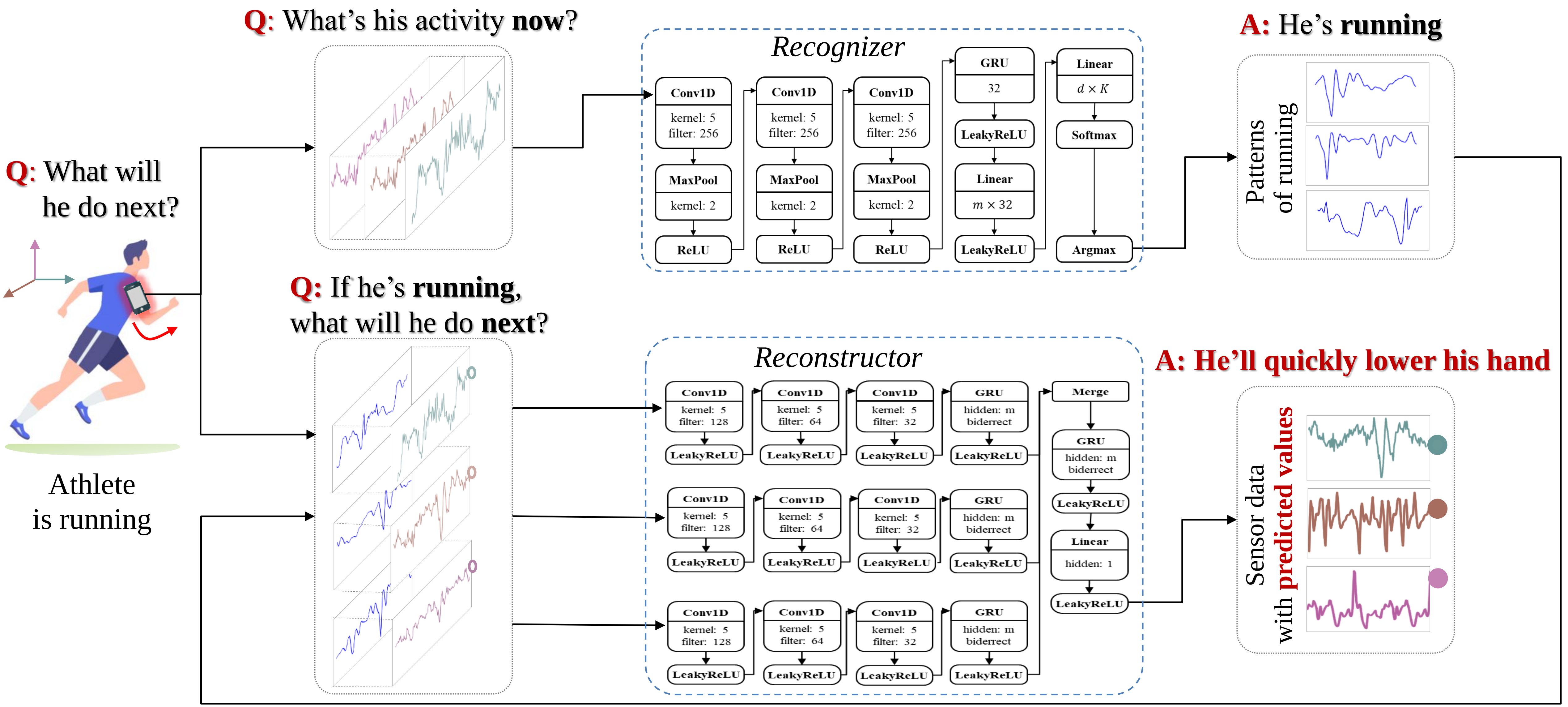


Deep learning model for **online prediction**

SANNI learns to predict future subsequences based on past ones, classified by typical behavior using previously discovered snippets

1) Zymbler M., Kraeva Y. High-performance time series anomaly discovery on graphics processors. Mathematics. 2023. 11(14), 3193.
2) Kraeva Y., Zymbler M. Anomaly detection in long time series on high-performance cluster with GPUs. Num. Meth. & Progr. 2023. 24(3), 291-304.
3) Zymbler M., Goglachev A. Fast summarization of long time series with graphics processor. Mathematics. 2022. 10(10). 1781. .
4) Zymbler M., Goglachev A. PaSTiLa: Scalable parallel algorithm for unsupervised labeling of long time series. LJM. 2024. 45(3), 1333-1347.
5) Kraeva Ya. Detection of time series anomalies based on data mining and neural network technologies. Bulletin of SUSU, CMSE. 2023. 12(3). 50-71.
6) Zymbler M., Yurtin A. Imputation of missing values of a time series based on joint application of analytical algorithms and neural networks. Num. Meth. & Progr. 2023. 24 (3), 243-259. DOI: 10.26089/NumMet.v24r318.

SANNI: Snippet & ANN-based online prediction



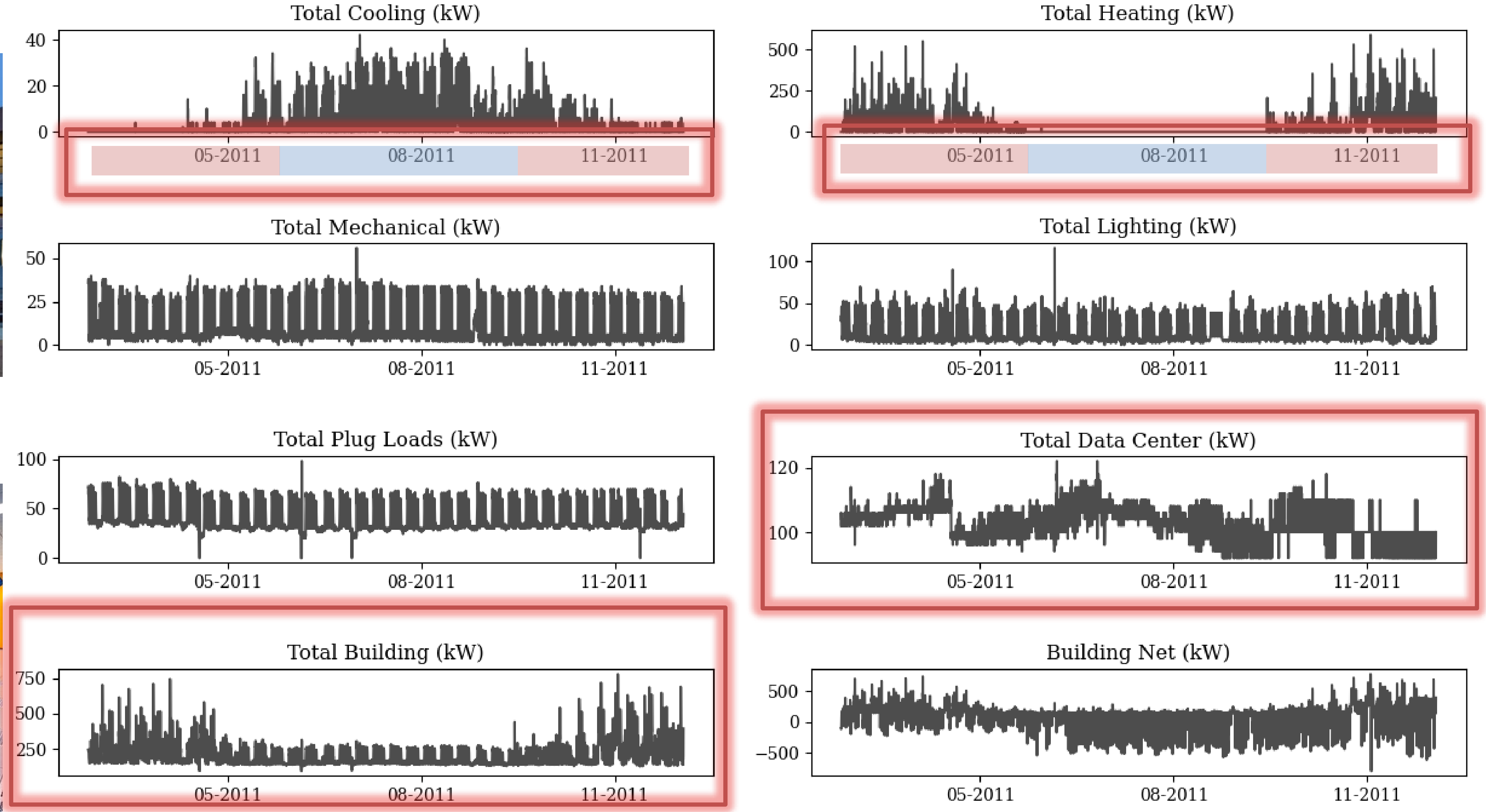
Case study: NREL[1]

NATIONAL RENEWABLE ENERGY LABORATORY

Golden, Colorado, USA

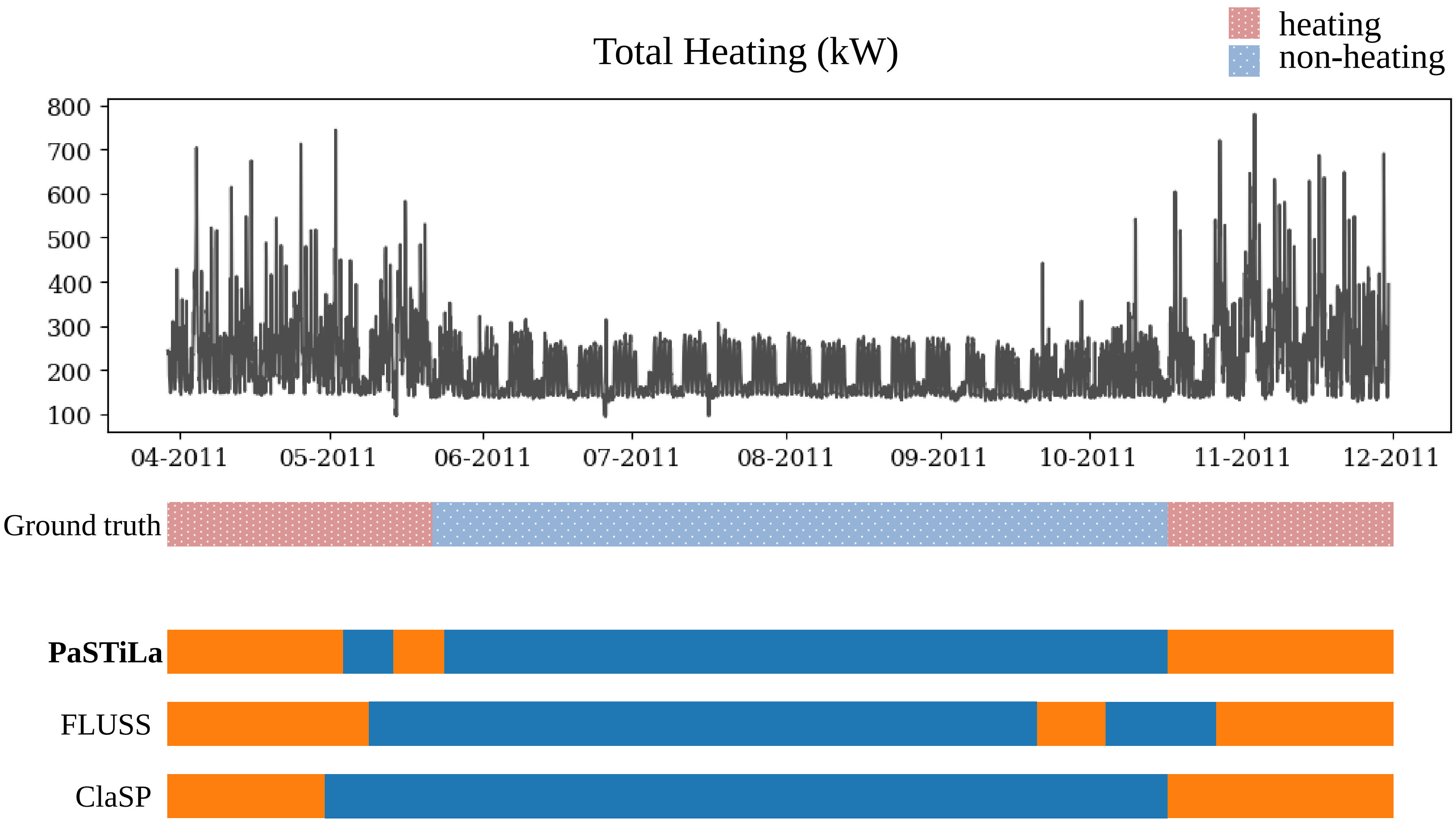
Whole Building

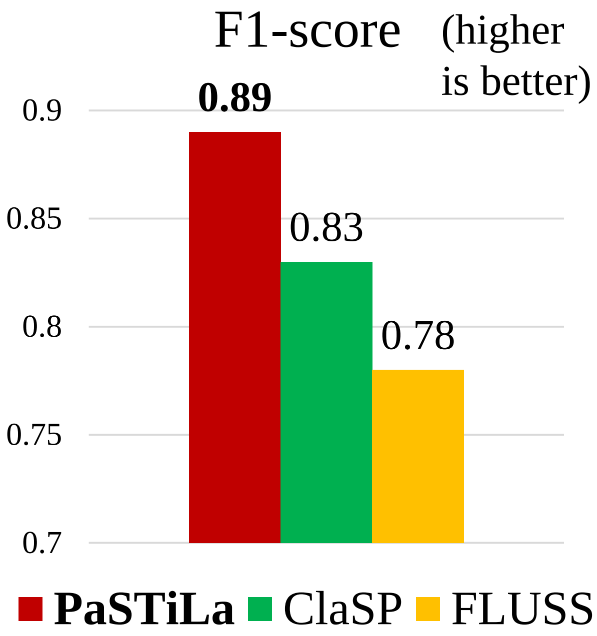
Data Center



[1] Sheppy M. *et al.* National Renewable Energy Laboratory (NREL) Research and Support Facility (RSF) Measured Data 2011. DOI: [10.25984/1845288](https://doi.org/10.25984/1845288)

case: Behavioral pattern discovery

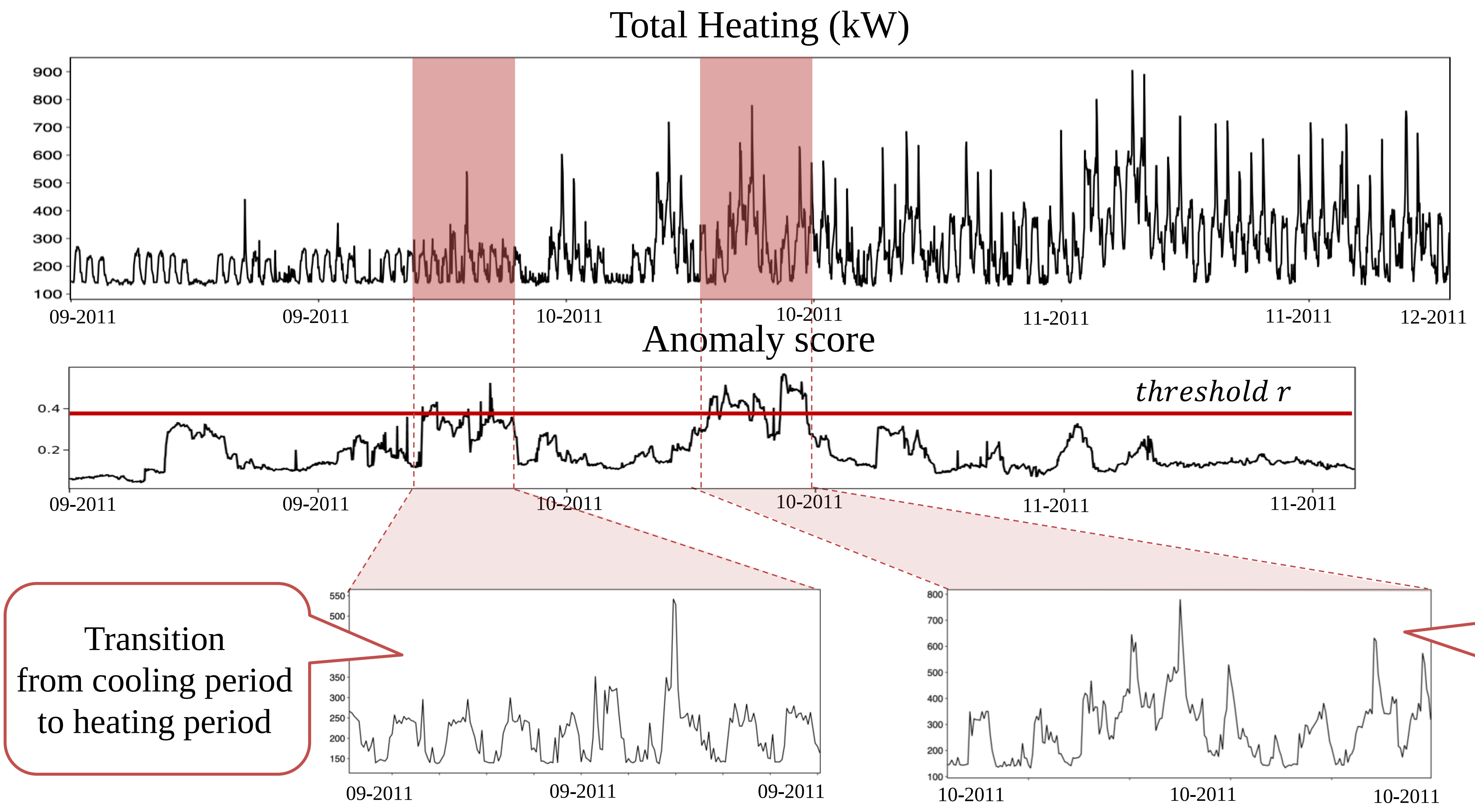


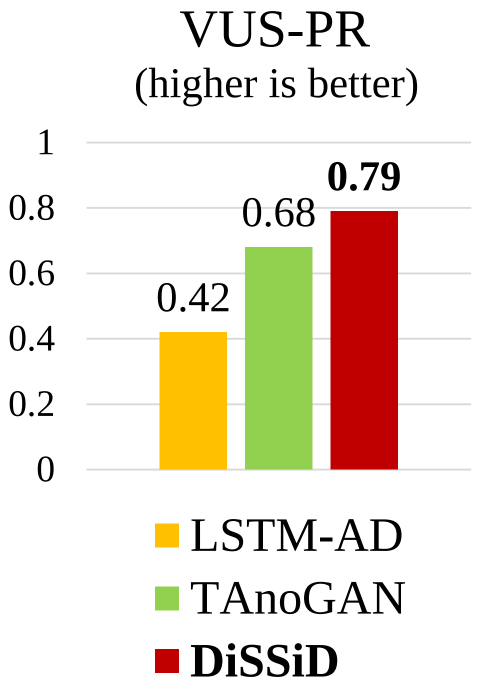


Our approach better detects change points

NREL case: Anomaly detection

NATIONAL RENEWABLE ENERGY LABORATORY
GOLDEN, COLORADO, USA

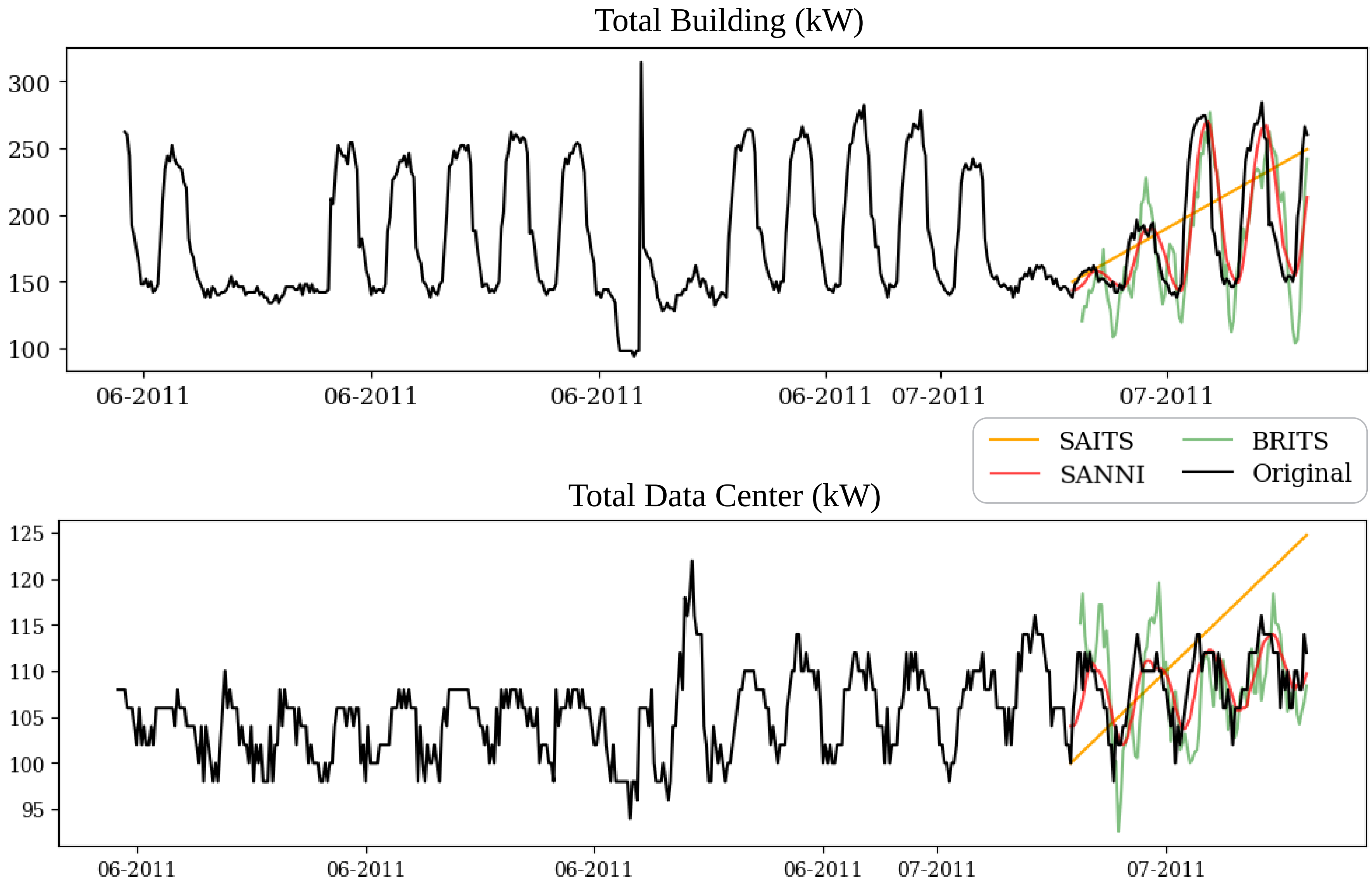


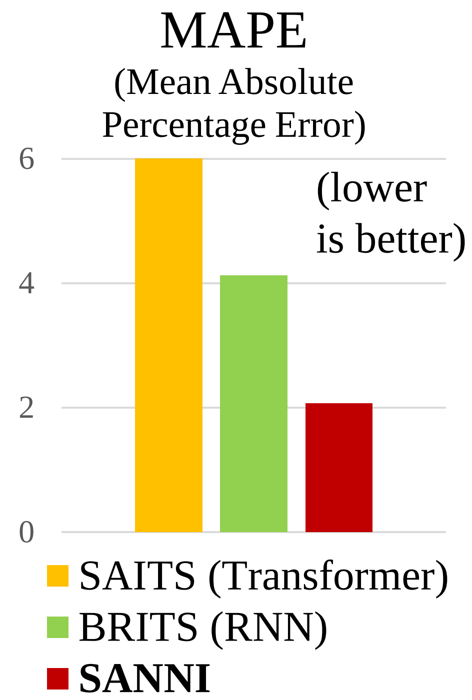


NREL case: Load prediction

NATIONAL RENEWABLE ENERGY LABORATORY

Golden, Colorado, USA





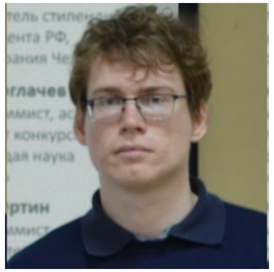
Our approach better predicts behavior

We’re ready to apply our ideas to your tasks if given a chance

- **Parallel unsupervised algorithms, which outperform S.O.T.A. rivals**
  - *Discord discovery:* PALMAD, PADDi (on GPU and multi-GPU clusters, respectively)
  - *Snippet discovery:* PSF, PaSTiLa (on GPU and multi-GPU clusters, respectively)
- **Deep learning models, which outperform S.O.T.A. rivals**
  - *Anomaly detection:* DiSSiD
  - *Prediction:* SANNI

Mikhail Zymbler
Dr.Sci.
Assioc. Prof.

Yana Kraeva
Cand.Sci.



Andrey Goglachev
MSc,
PhD student

Alexey Yurtin
MSc,
PhD student