EXPERIMENTAL SETTINGS

The machine used in the experiments has two AMD7742 64-core CPUs with 1024GB memory.

Subsequences in each time series are preprocessed with z-score normalization. All final scores output by detectors are normalized in [0, 1]. For detectors that rely on randomization, we report the average result of 10 trials on each time series.

[**Datasets**] Synthetic datasets noisy_sine and ARMA are from a previous work [9]. Real-world datasets include MIT-BIH Supraventricular Arrhythmia Database (MBA) [7, 13] and other datasets which are from various domains have been studied in earlier works [9, 10, 23].

Some datasets have two versions, e.g., ann_gun and stdb_308; and each version uses one of the two variables. When either version produces similar AUC for most detectors, we have chosen to use one only. Some datasets are trivial, e.g., chfdb_chf0175 and qtdbsel102; and all detectors have the perfect result (AUC=1), so we do not show them in Table 7.

We label anomalous periods for each time series following the previous work [9, 10, 23]. Details are given in Table 9. Positions of anomalies in MBA datasets can be seen in folder "MBA_Annotation".

The period of some dataset varies slightly at different time steps in the series; but it has no effect on the detection accuracy of all algorithms. Our algorithm works well when choosing subsequence length to be roughly the length of the period.

Brief descriptions of some datasets are given as follows.

dutch_pwrdemand: This time series has power consumption for a Dutch research facility for the year 1997 (one power measurement every 15 minutes for 365 days). It shows a characteristic weekly pattern that consists of 5 power usage peaks corresponding to the 5 weekdays followed by 2 days of low power usage on the weekends. Anomalous weeks occur when one or more of the normal usage peaks during a week do not occur due to holidays [9]. There are a total of 672 points each week (672=7x24x60/15) so the period length is 672. The series starts on Wednesday, January 1st, so each week period starts on Wednesday. There are a total of 6 anomalous weeks. Some papers [1, 9, 10] use this dataset with fewer anomalous weeks because they treat continuous anomalous weeks as one anomaly. Additional information about this dataset is given in [31].

ann_gun: The only anomalous period is first used in Keogh's work [10], as shown in Figure 4. Other anomalous periods in this series were later identified [1], and they are shown in Figure 5.

Patient_respiration: Foloowing previous work [9], we use the subset of nprs44 dataset [10], beginning at 15500 and ending at 22000. There are one apparent anomaly and one subtle anomaly in this dataset as shown in Figure 6.

TEK: Following previous work[9], we also concatenate dataset TEK14, TEK16 and TEK17 as TEK of length 15000. In Keogh's work [10], a total of 4 anomalies are marked. But in TEK14, 2 anomalous snippets belongs to the same period as shown in Figure 7. Since we regard each anomaly as an anomalous subsequence of one complete period, it is treated as one anomalous periodic subsequence of length 1000. So there are a total of 3 anomalous subsequences in our annotations of this dataset.

MBA803,MBA805,MBA806,MBA820,MBA14046: Following previous work [1], all are the subset of full MBA datasets, begining at 1,ending at 100K.

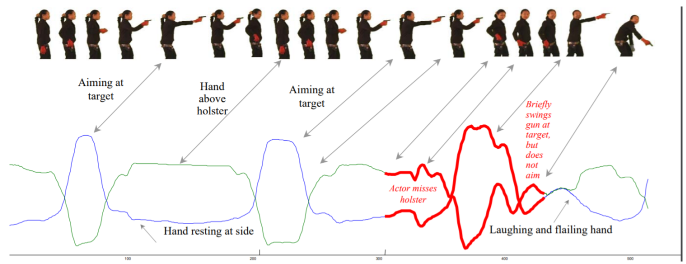

Figure 4: One anomaly period in the ann_gun dataset. The diagram is extracted from [10]

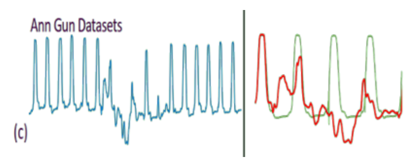

Figure 5: Additional anomalous periods in the ann_gun dataset, as identified by [1]. The diagram is extracted from [1].

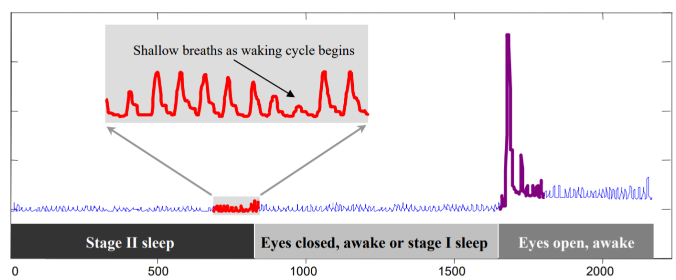

Figure 6: Anomalies in Patient_respiration dataset. The diagram is extracted from [10]

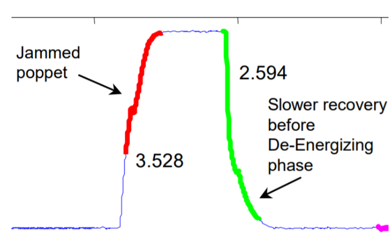

Figure 7: Anomalies in a period of TEK14 dataset. The diagram is extracted from [10]

Table 9: Locations of anomalous periodic subsequences in each dataset in terms of index i in $Y_{i,m}$, where the period length is m.

Dataset	period length	anomalous period index i
noisy_sine	300	6,11,21,31
ARMA	500	101,102,161
GPS_trajectory	2200	3,6
Patient_respiration	150	7,34
TEK	1000	2,10,13
dutch_pwrdemand	672	1,13,18,19,20,52
ann_gun	150	3,15,16,17,19
mitdb_100_180	250	8
mitdbx_108	370	12,28,29,30,31
stdb_308	400	7
lstdb_20221_43	170	5
lstdb_20321_40	200	5
MBA803	105	
MBA805	100	see details
MBA806	75	in folder:
MBA820	100	MBA_Annotation
MBA14046	90	

[**Algorithms**] The STOMP [33] implementation of MP is used; NormA is from http://helios.mi.parisdescartes.fr/ themisp/norma/; IDK-based detectors are our implementations based on [28]; and WFD is from github.com/GAMES-UChile/Wasserstein-Fourier. Others are from scikit-learn.org. All are in Python.

As for the 1Line method, we use one of the following five types of basic vectorized primitive functions in Matlab as an anomaly score for each sliding window of size ω:

(i) $\pm$diff(Y): the difference between the current point and the previous point. Here $\omega = 1$.
(ii) $\pm$movmax(Y, $\boldsymbol{\omega}$)
(iii) $\pm$movmin(Y, $\boldsymbol{\omega}$)
(iv) $\pm$movmean(Y, $\boldsymbol{\omega}$)
(v) $\pm$movstd(Y, $\boldsymbol{\omega}$)

where Y is the time series; and the maximum, minimum, mean or standard deviation is computed for each window of $\boldsymbol{\omega}$ points.

We run these 5 one-liner on each dataset and report the median AUC (out of the five values) in Table 7. Low median values indicate that the datasets are hard to detect using the 1Line method; otherwise, the datasets have anomalies that can be easily detected.

[**Measures**] The detection accuracy of an anomaly detector is measured in terms of AUC (Area under ROC curve). As all the anomaly detectors are unsupervised learners, all models are trained with the given datasets with no labels. Only after the trained models have made predictions, the ground truth labels are used to compute the AUC for each dataset.

Given a periodic time series Y of length n and period length m, a subsequence $Y_{i,m}$ of Y is a subset of contiguous values of length m, for $i = 1, \ldots, s$, where $s = \lfloor n/m \rfloor$. A distribution-based (non-sliding-window) algorithm outputs a score of each periodic subsequence $Y_{i,m}$. Then AUC can be calculated based on scores α_i for $Y_{i,m}$ $\forall i = 1, \ldots, s$.

An anomaly detector using the sliding window having window size ω produces a total of $n - \omega + 1$ subsequences from Y. When calculating AUC, scores of the sliding subsequences are transformed into periodic subsequence scores as follows: Let S_j be the anomaly score of subsequence $Y_{j,\omega}$, where $1 \leq j \leq (n - \omega + 1)$. The final score corresponds to a periodic subsequence $Y_{i,m}$ is the maximum score of S_j $\forall j$ such that at least half of $Y_{j,\omega}$ is included in $Y_{i,m}$.