Flexible Windowing for Correlation-Aware Ranking in Anomalous Environments

Anwesha Das[1,2] Henry Hoffmann[1] Alex Aiken[2]
[1]*The University of Chicago* [2]*SLAC National Accelerator Laboratory/Stanford University*

***Abstract*—Analyzing time-series correlations is key to anomaly detection and incident diagnosis, but existing methods that rely on fixed time windows often fail or are inaccurate in environments with (a) missing adequate ground truth, (b) asynchronous signals, or (c) mixed sampling rates containing irregularities in signals. To provide anomaly-relevant signal correlations given these challenges, we propose *AdaptWin*, a novel method to adaptively select windows for a signal-pair of the same size with distinct positions, i.e., the specific start and end timestamps can differ, for correlation-aware ranking. Our window selection is based on deviations in (1) inter-arrival times and (2) observed values. This flexible window selection approach improves ranking by better capturing signal variations aligned with anomalies. Across three real-world datasets, AdaptWin improves anomaly-relevant ranking by over 3× compared to adaptive baselines.**

***Index Terms*—Windowing, Correlations, Anomalies, Ranking**

I. Introduction

Failure diagnosis in complex systems often requires analyzing pairwise correlations in large volumes of multivariate time-series data [19, 47]. In practice, however, several challenges arise, including: (a) lack of precise ground truth [44], (b) asynchronicity in data [53], and (c) mixed sampling rates [10]. Many anomaly detection methods using multiple sources of time-series data rely on fixed-size sliding windows that assume uniform sampling and temporal alignment across signals [29]. Unfortunately, these assumptions often fail to hold in real-world settings, leading to poor performance in identifying the components responsible for anomalous behavior [24].

A core problem is selecting suitable windows for pairwise correlations, especially when the failure window is imprecise [32]. Domain experts may only know that an anomaly occurred within a coarse-grained window, but not the precise start and end times of the failure. Naively using windows across multiple signals with mixed sampling rates, asynchrony, or missing values can result in correlations that fail to capture anomaly-indicative signals [52]. Brute-force search for window selection is computationally prohibitive [46], and existing adaptive windowing methods either target univariate signals [9], or are not designed to improve anomaly-relevant correlations [31]. To this end, we study the problem of selecting windows from signal-pairs, i.e., both size and position of a window with distinct start and end times, such that their pairwise correlation leads to better anomaly-relevant ranking.

Key Insight: Anomalous conditions in non-stationary environments [2] are characterized by outliers or missing values [21] in signals. Anomalies are often identified through differences in subsequent values (δ^{V}), and/or inter-arrival times (δ^{T}). By prioritizing window regions with higher δ^{V} or δ^{T}, we can better capture time-localized deviations across signals. This process of adapting the window position and size to differences in value and time leads to more accurate rankings.

We implement this insight as *AdaptWin*, a flexible windowing method for correlation-aware ranking for incident diagnosis. Unlike traditional fixed or sliding window methods, AdaptWin adaptively selects windows for signal-pairs based on δ^{V} and δ^{T} deviations, allowing each signal's window to start and end at different times, while maintaining a consistent window size. We develop a method to filter candidate windows using these deviations and optimize bivariate correlation functions such as Pearson or Spearman correlation coefficients, and dynamic time warping. AdaptWin is robust to mixed sampling rates and works without precise failure labels, making it applicable to a wide range of supervised and unsupervised settings.

Contributions: Our contributions are as follows:

- We formulate the *flexible window* selection problem for correlation-aware signal ranking in unstable anomalous environments involving asynchrony and irregularities.
- We propose *AdaptWin*, a novel window selection method that selects windows for signal-pairs using value and time differences, allowing signal-specific start and end times.
- We evaluate AdaptWin across *three real-world* datasets: a particle accelerator, space weather monitor, and metro transportation system, and show that AdaptWin improves anomaly-relevant ranking by over 3× compared to existing fixed window and other adaptive baselines.
- We release artifacts including a curated production dataset of a particle accelerator system at https://github.com/adaptsyslearn/AdaptWin, to enable further studies.

Overall, AdaptWin provides a principled and practical solution to window selection for correlation analysis with regular, irregular, or asynchronous signals. By focusing on anomaly-relevant signal deviations, AdaptWin improves ranking accuracy across diverse domains where traditional windowing falls short.

Why need flexible windows? In many scientific or industrial facilities with no or sparse ground truth, the specific failure window is unavailable or imprecise. For e.g., an anomaly lasts for ≈10 mins within a 3-hour timeframe, but which specific 10-min relates to the anomaly is not exactly known. In such cases, various window sizes are examined by trial-and-error method [47]. Besides, data sources may not be *synchronized* leading to delays in trends across signals for which lagged or cross correlations have been used [34, 53]. Sliding windows with a certain step size are more flexible, but which windows to select for pairwise correlations that can aid overall anomaly-relevant ranking is non-trivial to determine. The problem is exacerbated in the presence of missing values (i.e., irregular or unevenly spaced observations), and mixed sampling rates [24, 51], where even defining the corresponding windows in two time-series needs a systematic method. Fixed windows mostly assume regularity, i.e., equidistant values where time differences between pairs of observations, referred to as δ^{T}, do not vary much. Imputation methods can convert an irregular data to a regular one [54], yet anomaly-relevant window selection is not really accounted for. Under this situation, an *automated* flexible windowing method to better capture intra- and inter-signal correlation patterns can benefit correlation-aware ranking.

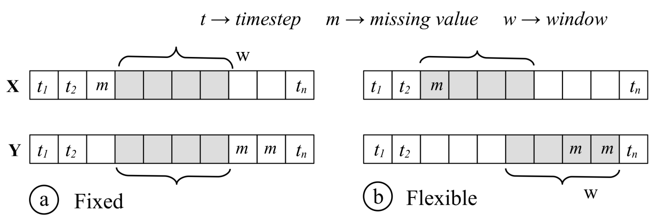


Fig. 1: Fixed vs. Flexible Windows

As an example, Fig. 1 shows two time-series X and Y with missing value(s) (*m*) at various timesteps (*t*). Using the same start and end times, i.e., *fixed* ⓐ, in both signals for choosing windows could result in unsatisfactory anomaly-relevant ranking due to the aforementioned reasons. Including certain timesteps with missing values for window selection could have better correlation with an anomaly, as shown in ⓑ. *Flexible* positioning of windows (i.e., different start and end times ⓑ) can account for any asynchronicity and irregularity, improving ranking quality. Thus, a method to select windows from X and Y that jointly optimize (maximize or minimize) a bivariate function (F) in such a way that the overall anomaly-relevant ranking quality is improved can help incident diagnosis.

A few prior works propose adaptive windows for univariate analyses [9, 15, 41], e.g., duplicate or drift detection, or domain-specific multivariate studies [28, 31] such as human activity recognition. Univariate methods are insufficient for bivariate correlations, and domain-specific methods have limited applicability. Our study addresses the need for appropriate window selection in pairwise correlations under more domain-agnostic dynamic anomalous environments [32].

Why ranking? For fault localization or incident diagnosis, one common strategy is to rank signals based on correlation scores [4, 43]–[45]. Ranking helps when the relative ordering is of greater interest than the actual statistical scores. Ranking is also robust to outliers in the data, and helps to uncover ties amongst signals. The ranking quality is then examined using the obtainable failure data (i.e., ground truth) [8]. Anomaly-relevant signal(s) if ranked higher, helps to focus on the cause-related component(s) aiding efficient diagnosis efforts. An improved window selection method is expected to rank the anomaly-relevant signals relatively higher than the anomaly-irrelevant ones. This helps domain experts/operators to analyze a small subset of signals one by one for troubleshooting. Thus, we show the accuracy of our proposed windowing method via correlation-aware ranking (no new ranking method is proposed).

Why bivariate? Pairwise correlations or functions involving two variables are often used for anomaly detection tasks [11, 14, 19, 29, 39, 47, 52]. Besides, many statistical measures such as distance functions and rank correlations are bivariate in nature [50]. Also, several multivariate associations can be derived from pairwise correlations [49]. Hence, we study bivariate functions for correlation-aware ranking.

II. Background and Motivation

We discuss the inadequacy of prevalent window selection methods for failure diagnosis (§II-A), followed by the effect of window selection on correlation-aware ranking through a real-life example (§II-B) motivating the need for our study.

A. Window Selection

Fixed-size windows with time lags of one or more values (i.e., sliding windows) selected via random shuffling or moving windows are generally used in time-series analysis [1, 21, 29]. A suitable single window size is derived for correlations empirically guided by available failure ground truth data, i.e., manually created labels [24] or via trial-and-error method [29]. Pairwise correlations result in a score [27, 52] that can be used for ranking-related tasks. Usually, the same window position is used for all correlating signals assuming equally spaced values in time-series of uniform sampling rate [19, 47].

In supervised domains, reliable ground truth helps estimate precise windows, i.e., the known start and end times of an anomaly is fairly accurate [43]. However, in unsupervised domains with sparse ground truth where manual labeling is cumbersome or *error-prone*, only a coarse-grained approximate timeframe is available [17, 44]. In such cases, random trial-and-error method for window selection is inadequate [19, 39]. Besides, asynchronicity in signals require flexible windows to account for delays in trends to improve correlation scores as evident from prior work [2, 47]. Also, using a few arbitrary window sizes suffice for signals with equidistant values of similar sampling rates to analyze failures [11, 29]. Appropriate window selection becomes non-trivial in face of mixed sampling rates with sizable irregularities in signals [20]. Missing data gives rise to sparsity that can affect bivariate correlations.

Why not brute-force? A strawman solution of finding all windows for a signal-pair with c entries each and m window sizes is *quadratic* $\mathcal{O}(c^2\mathtt{m})$. The time complexity is $\mathcal{O}(pc^2\mathtt{m})$ for p signal-pairs. Clearly, a brute-force method is inefficient for large c, m, or p values even without anomaly relevance or real-time needs. Prior studies on matrix profiles build on this aspect to propose methods that improve runtime [30, 46]. Besides, a brute-force search or sliding window method does not innately account for anomaly relevance. Further analysis is needed for incident diagnosis, e.g., pruning time-series graphs [4].

Our window selection is an *offline* process as many sensor-based physical systems need to diagnose failures using accurate hints about faulty subsystem(s) without any real-time needs. Thus, our goal is to improve *accuracy*, i.e., ranking quality,

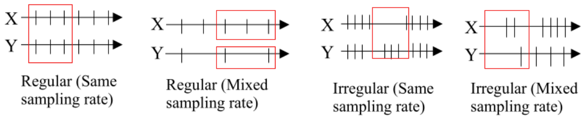


Fig. 2: Sampling Rates and Missing Data

TABLE I: Pairwise Correlation (R: Rank Variations)

Bivariate Relation	Pearson's Coefficient (PC)						Dynamic Time Warping (DTW)					
	1-hour			20-mins			1-hour			20-mins		
	T^S	T^D	R	T^S	T^D	R	T^S	T^D	R	T^S	T^D	R
S1→S4	**0.23**	0.01	219→543	0.13	**0.13**	340→340	**5.15**	11.59	196→44	5.44	2.30	183→541
S2→S4	0.28	0.09	117→414	0.25	0.01	157→548	7.04	14.67	73→12	3.80	8.52	359→48
S3→S4	0.06	0.19	472→261	0.30	0.11	101→371	7.77	7.43	63→71	3.97	**7.29**	346→67

T^S/T^D: Scores of windows related to Same and Different start/end timestamps

though we show that our method's efficiency (runtime) is no worse than the usual state-of-the-art methods.

Fig. 2 shows four possible cases of correlations considering regular and irregular time-series with same and mixed sampling rates for a signal-pair `X` and `Y`. The correlation strength varies based on the chosen windows' contents. Correlation score-based ranking may not be sufficiently accurate unless fitting windows are chosen for such cases [10, 32]. For causal discovery or incident diagnosis in unstable domains, a minor shift in the window can make a major difference in the final accuracy [14, 44]. Thus, we study flexible window selection for correlations that can be adapted for both regular and irregular signals.

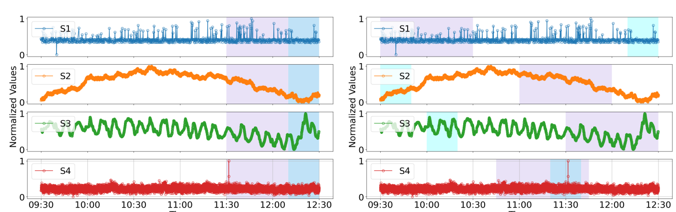
Prior Techniques: Analyzing a range of window sizes, or creating a labeling tool is often used to circumvent the lack of precise ground truth or labeled data [44, 47]. Delays in signal trends have been addressed by considering many possible delay values or aligning delayed trends to a reference frame [2, 47]. Aggregation functions are used for mixed sampling rates (e.g., mean/median) that are not always straightforward to determine in a given context [10]. In face of unevenly spaced values, irregular data is converted to a regular one via missing data imputation [39, 54] (e.g., backfill) or the elapsed time between values (δ^T) is used as an added feature. The former involves data augmentation while the later may lose contextual information. These approaches to fill missing values are not correlation-aware by themselves, and can lead to suboptimal performance under various anomalous conditions [24].

The common scenario is that, for an anomalous incident, anomaly- or cause-related signals deviate from their usual behaviour for certain timesteps leading to a change in the correlation strength. Our intuition is to prioritize timesteps of relatively higher deviation to form windows and then impute only within the selected window. Estimating deviations over time to select windows for a signal-pair can improve overall correlation-aware ranking across a set of signals.

B. A Motivating Example

For fault localization, some methods rank signals based on correlation scores using the selected windows [13, 44]. The ranking quality is then assessed using the available manually logged failure ground truth, e.g., *Water cooler `A` failed for ≈30-mins between `10:00` and `18:00` hrs*. Even after consultation with domain experts it can be infeasible to deduce finer information, e.g., *from `11:00` to `11:30` hrs, signals A^1 and A^2 showed anomalous behaviour*, where A^1 and A^2 are two of the many signals of subsystem `A`, i.e., dense or precise labels can be unavailable in some systems, conforming to prior studies [19, 44]. With better window selection, we expect `A`-related signal(s) to appear amongst the top ranked signals.

We show rank variations based on window selection in a production system illustrating the importance of flexible windowing. Fig. 3 shows four signals of a complex particle accelerator [18] logging over 5000 signals, namely, `S1` from an amplifier system, `S2` and `S3` from two water cooling systems, and `S4` from the beam monitoring system, respectively, over a period of three hours (`9:30` to `12:30`). The two highlighted areas correspond to 1-hour and 20-min windows, respectively. Generally, correlations are obtained using the same start and end timestamps across signals, as shown in Fig 3a. However, correlations with a delay help to capture the relation between disparate sensors of a system by accommodating asynchronous windows (overlapping or non-overlapping) [53]. For e.g., there is a 45-mins lag between `S3` and `S4` for the 1-hour window in Fig. 3b. Such slightly delayed correlations help to excavate anomaly-indicative signals, or interpret collective anomalies [1] (i.e., multiple incidents together determine an anomaly) that may be missed if the window position is always fixed. Changing the window need not necessarily change some statistical functions, such as, information gain [10]. However, the strength of correlation depends on *how long* and *which observations* are used for any signal-pair. Flexibility in window positioning can enable such overlaps across various time-series.



(a) Same Windows (b) Flexible Windows

Fig. 3: Correlation windows. a) Same/Fixed, b) Dissimilar

Tab. I shows the bivariate scores of two correlation functions, namely, Pearson's correlation (PC) and dynamic time warping (DTW), corresponding to the timeframes marked in Fig. 3. PC measures the linear relationship, and DTW [50] computes the similarity score between two signals. T^S and T^D indicate the correlation scores of signals `S1` to `S3` with `S4` over 1-hour and 20-min windows for same and different window positions, respectively. `S1` to `S3` are then ranked based on these scores across 5511 signals (i.e., range [1, 5511]). `R` lists the updated ranks based on score changes when transitioning from fixed to flexible window selection. In case of `S1`→`S4` relation with *fixed* 1-hour windows, PC is 23× higher while DTW is 2.25×

lower than correlation with *dissimilar* windows. For a 20-min window, S1→S4 correlation remains unchanged for PC, while S3→S4 correlation is 1.83× higher using dissimilar windows for DTW. *We do not analyze rank accuracy in relation to faults here, but show that ranks can alter considerably*. Such changes are magnified when the set of signals is large with marginal score differences, where ranks can rise or fall by 7×/8×.

As evident from Tab. I, regardless of the correlation function, the impact of windowing depends on the nature of time-series (e.g., sampling rate, range of values), and window size. Even a small score change (e.g., 0.03) is non-trivial for a dataset with signals in $\mathcal{O}(10^3)$ orders of magnitude for which an automated method can help enable accurate correlation-aware ranking.

III. Related Work

Adaptive Window Selection: Prior work on adaptive window selection consists of a) univariate and b) multivariate models. Concept drift estimation and duplicate or change point detection methods that identify changes in data distribution [9, 15, 25, 41] fall under a). Some imputation methods do not select windows but handle missing data or construct latent space [51, 54] that can be used to post-process selected windows, and are subsumed in a). Time-series segmentation methods finding meaningful windows using temporal statistics [7], or methods to pick multiple windows of different incidents in the same time-series [23] also fall under a). These univariate methods are inadequate for bivariate associations.

Some of the developed multivariate methods [10, 16, 33] are domain-specific, involving feature aggregation, segmentation, or human activity recognition [28, 31]. Flexwin [31] picks windows from seismograms based on shape and similarity, while Optwin [28] chooses windows by maximizing class separability in signals to improve classifier performance. Joint time-series models using chaining, segmentation, and discord or motif discovery methods [7, 22, 30, 48] are domain-agnostic, subsumed under b). Some of these methods focus on improving efficiency (reducing runtime), or do not consider irregularities. Few multivariate models may be adapted to bivariate analyses but they do not address ranking accuracy. In contrast to these, our generic method to select windows for pairwise correlations focus on improving accuracy in anomaly-relevant ranking.

Correlation-Aware Ranking: Prior studies on causal inference, fault localization or incident diagnosis with time-series data involve score-based ranking [4, 13, 19, 39, 44, 52]. The scores are obtained using similarity measures, reconstruction errors, null hypothesis testing, and point adjustment-like techniques. Apart from using fixed windows that do not slide, sliding windows with a step size are commonly used. The importance of selecting suitable windows for failure diagnosis is exemplified in recent works affirming our motivation [8, 20]. Ranking suggests incidents of interest, or cause-related signals behind failures. F1-score, Z-score, or a domain-specific scoring function (e.g., weighted precision and recall) is often used to rank. Some studies detect the anomalous window or analyze correlations in comparison to normal data [12, 39], but do not assess anomaly relevance for incident diagnosis. Few studies use synthetic anomalies, or use only regular sensor data, while others can be expensive with respect to the achieved accuracy (smaller window size implies more windows to be examined). These studies show that based on the intricacies of the facility, window selection is not straightforward [47]. We do not propose any novel ranking scheme, but our window selection method can add to the wealth of these prior efforts.

IV. Methodology

In this section, we discuss the assumptions made, and the characteristics of anomalous conditions (§IV-A), followed by a description of our proposed window selection method (§IV-B).

A. Assumptions and Anomaly Characteristics

Our design conforms to the following realistic assumptions:

1) The choice of window is sensitive to the characteristics of anomalies, aside from optimizing a correlation function. Having too few values of signals may maximize correlation (local extrema), but may not be relevant for detecting anomalies. Similarly, abrupt changes, or longer than usual time differences between observations implying anomalous conditions are important to capture even if the correlation score does not improve significantly.
2) The sampling rates of continuous-time signals may not be known apriori. Reflective of real-world settings, signals with collective or contextual anomalies are considered unlike point anomalies [1, 21] (i.e., subsequence of values together indicate an anomaly).
3) The *exact* start or end timestamps of failure incidents may not be available. Approximate windows about failures based on the available coarse ground truth are used.
4) Multiple incidents can occur in close temporal proximity, i.e., using too narrow windows can miss important correlations, while very large windows may dampen helpful correlations to the point that they become undetectable.

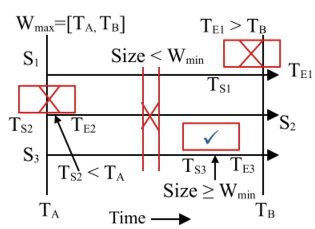


Fig. 4: Feasible Window

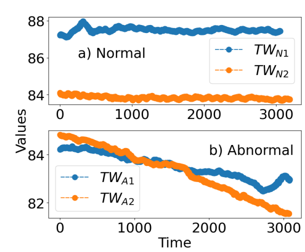


Fig. 5: (Ab)Normality

Conditions leading to a fault usually persist for some time. Using too few values for correlation may not reflect an actual fault even with an optimal correlation score. Hence, a minimum-size window, W_{min}, across the considered set of signals helps to have a lower bound on the possible sizes of the selected window. In Fig. 4, T denotes time, S signal, T_S and T_E start and end times for three signals, S_1, S_2, and S_3, respectively. Fig. 4 shows that windows that cross the boundaries of original window [T_A, T_B] of size W_{max}, or whose size is below W_{min} are not considered, while those within [T_A, T_B] with a size of W_{min}

or higher are allowed (e.g., $\mathtt{S}_3$). The flexibility of choosing windows is limited to the range of the original window.

For window selection, we leverage two key characteristics of time-series often observed in anomalous conditions:

- *Inter-Observation Time*: The inter-arrival times or time differences between two adjacent values are referred to as the inter-observation times (δ^{T}). The most frequent inter-observation time in a time-series suggests the sampling interval (SI), i.e., the *mode* value. Longer than usual δ^{T} can indicate abnormalities in many production environments. Unusual δ^{T} can also be intentional, such as, when a logger is configured to skip storing values unless there is a major change in the sensed measurement. Imputation helps in such cases (e.g., forwardfilling). Unevenness in data can imply a faulty device or intermittent defects (i.e., an anomaly fixes itself before reappearing). Capturing such irregularities helps to uncover interesting correlations. We use unique δ^{T}s with their respective frequencies to deduce a minimum size window ($\mathrm{W}_{\mathrm{min}}$) across signals.
- *Observational Difference*: Gradual or abrupt changes imply instabilities. Unstable conditions relate to unusual correlations. Fig. 5 shows a specific time-series during two non-overlapping normal (TW_{N}) and abnormal windows (TW_{A}), respectively. The overall difference in the range of observations between normal and abnormal times is not high. However, slow changes with relatively low variance can indicate anomalies, as evident from Fig. 5b. Based on this insight, the difference in subsequent values (δ^{V}), as opposed to variance of a randomly chosen full window is explored for window selection. We use unique δ^{V}s with their respective frequencies to consider candidate windows for capturing interesting correlations.

TABLE II: Summary of Major Notations

Notation	Description
$\mathrm{W}_{\mathrm{max}}$	Original coarse window size (fixed during window selection)
$\mathrm{W}_{\mathrm{min}}$	Minimum size window
$\mathrm{W}_{\mathrm{act}}$	Actual chosen finer window size
$\mathtt{F}$	Correlation function
δ^{T}, δ^{V}	Inter-arrival time, Observational difference
T_{st}, T_{E}	Start, End timestamps of a window (W)
SI, $\mathrm{SI}^{\mathrm{com}}$	Sampling Interval of a signal (S), Common SI of a signal-pair
λ, γ, α, θ	Frequencies of δ^{T}, δ^{V} and their filtering thresholds
L_{k}	Location indices for candidate windows

Tab. II lists the notations used to describe our approach.

B. Approach Overview

Fig. 6 gives an overview of our approach comprising window selection and correlation analysis. A coarse window [T_{A}, T_{B}] of size $\mathrm{W}_{\mathrm{max}}$ and a set of signal-pairs are considered based on the available failure ground truth. This original window *does not change* during the process of window selection. AdaptWin first identifies the sampling interval (SI) of each signal based on which it selects a common minimum size window ($\mathrm{W}_{\mathrm{min}}$). AdaptWin then obtains anomaly-relevant windows (W_{X}, W_{Y}) that optimize a correlation function $\mathtt{F}$ for a signal-pair $\mathtt{X}$ and $\mathtt{Y}$ from [T_{A}, T_{B}]. Pairwise correlation scores are obtained for all signal-pairs. Based on score-based ranking of signals, and the available ground truth about anomalies, we examine the accuracy of ranking for the studied failure instances. If anomaly-relevant signals are ranked higher (top $\mathtt{K}$), the related window selection method is deemed more accurate as it localizes the failure across fewer subsystems helping operators and domain experts to narrow down their search space for failure diagnosis. Not all possible signal-pairs are considered. The specific signal-pairs used for correlation depends on the context of the dataset and nature of failures. For e.g., a system performance-indicative signal is correlated with all other signals to assess subsystem correlations, or a signal representing full system reliability is correlated with signals of a specific sensor type to determine inter-signal relationships of that sensor, and so on.

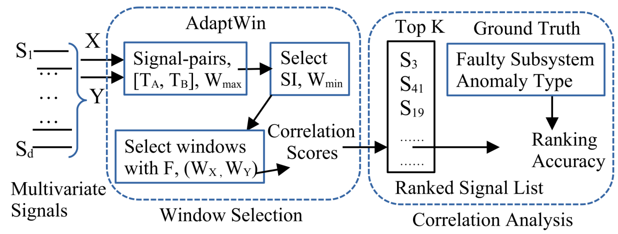


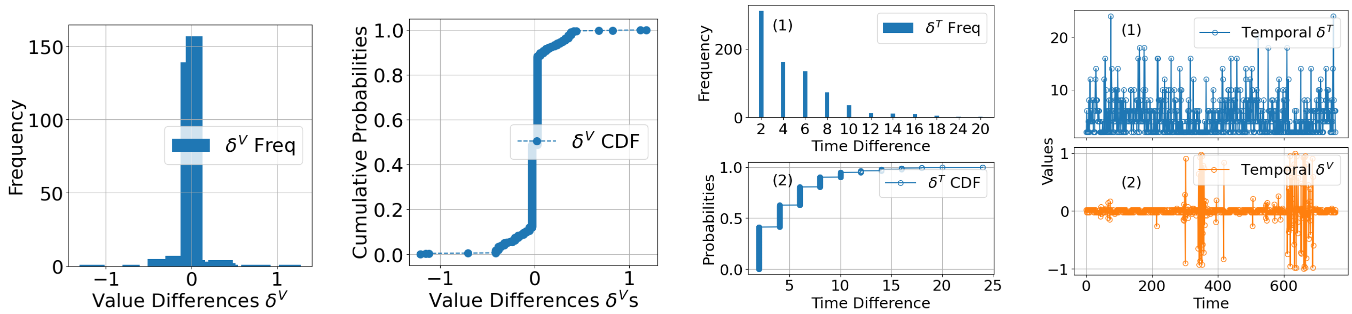
Fig. 6: Overview

Algo. 1 shows AdaptWin's minimum window size selection approach. The idea is to estimate a sensible coarse-grained timeframe based on non-uniformity in signals. In a dataset with $\mathtt{d}$ signals, for each time-series of a signal-pair, the distinct inter-arrival times (δ^{T}) with their frequencies (λ) are arranged in descending order of δ^{T}. δ^{T} with the highest frequency (i.e., mode) forms the SI of a signal. A few missing timesteps similar to the size of SI resulting in short delays are often noticed during normal conditions. Thus, *fewer* but *longer* delays are prioritized over *many shorter* delays relative to SI. For e.g., with a 2-secs SI, a single 22-secs delay can be more helpful for correlation over 170 delays of 4-secs each. Based on this insight, a threshold γ greater than SI, is derived from the distribution of δ^{T}s, e.g., a function of inter-quartile range (IQR).

Algorithm 1 Minimum Window Size Selection

Require: $\mathrm{T_W}$=[$\mathrm{T_A}$, $\mathrm{T_B}$]$^{\mathrm{W_{max}}}$, $\mathrm{S_L}$=[S_1,… S_d], g ▷ *d* Signals
Ensure: W_{min}, [t^{i}_{span}, SI^{i}] for i^{th} signal ▷ Min. window size
procedure MIN WIN SIZE($\mathrm{T_W}$, $\mathrm{S_L}$, g)
  for each [S_i, S_j] $\in$ $\mathrm{S_L}$ **do** ▷ Signal-pair
    [(δ^T, λ)] $\leftarrow$ Sort δ^{T}s with their frequencies λ
    [SI^i, SI^j] $\leftarrow$ Mode [(δ^T, λ)] ▷ Sampling Interval
    γ $\leftarrow$ Derive threshold from δ^T distribution
    if regular signal **then** $\mathrm{t_{span}} \leftarrow (\frac{1}{g} * W_{max})$ **else**
    $\mathrm{t_{span}} \leftarrow \sum (\delta^T * \lambda), \forall\, (\delta^T > \gamma)$ ▷ Irregular
    Pair_{min} $\leftarrow$ Max $\mathrm{t_{span}}$ of [S_i, S_j]
  W_{min} $\leftarrow$ Max of all [Pair_{min}]
return W_{min}, [t^{i}_{span}, SI^{i}]

δ^{T}s exceeding γ are used to estimate the cumulative duration of irregularities, $\mathrm{t_{span}}$, using the *sum of product* rule, i.e., the products of δ^{T}s with their respective λs are added. For regular signals, δ^{T}s are either same or close to SI, for which $\mathrm{t_{span}}$ is defined as $\frac{1}{g}^{th}$ of $\mathrm{W_{max}}$. This ensures that at least a fraction ($\frac{1}{g}$) of the total number of entries in the original window



(a) δ^{V} Freq. (b) δ^{V} CDF (c) δ^{T} Freq. and CDF (d) δ^{T} and δ^{V} over time

Fig. 7: Frequency and Cumulative Distribution Function (CDF) of δ^{V} and δ^{T} indicating the level of symmetry and skewness.

is used for correlation in case of fewer or no irregularities. The minimum window size of a signal-pair ($\mathrm{Pair_{min}}$) is the maximum $\mathrm{t_{span}}$ of the two signals. This is repeated for all signal-pairs. To accommodate instabilities as much as possible, the maximum of all $\mathrm{Pair_{min}}$s from the examined signal-pairs forms the common $\mathrm{W_{min}}$.

The intuition behind deriving $\mathrm{t_{span}}$ is to capture a timeframe that is long enough to include the major patches of unevenness in the window. The number of missing values in the window can be estimated using $\mathrm{t_{span}}$ and SI. For a simple case of two regular signals, Algo. 1 guarantees $\mathrm{W_{min}}$ to be $\frac{1}{g}^{th}$ of $\mathrm{W_{max}}$.

For illustration, Fig. 7 shows the statistics for a specific 1-hour window of a signal comprising frequency (*freq.*) and cumulative distribution function (CDF) of distinct δ^{V}s and δ^{T}s, respectively. In Fig. 7c (1) (top), δ^{T}=2 secs is the SI (x-axis) with the highest frequency. As per Algo. 1, $\mathrm{W_{min}}$=11.6 mins for threshold γ=8 (where δ^{T}s exceeding γ range between 10 and 24). As seen from Figs. 7b and 7c (2), δ^{V} CDF has a normal distribution, while δ^{T} CDF has a gamma distribution hinting at the skewness in δ^{T}s over δ^{V}s. The histogram plots in Figs. 7a and 7c (1) show more asymmetry in the distribution of δ^{T} over δ^{V} hinting at signal irregularities. Higher symmetry in both δ^{V} and δ^{T} hint at less irregularity providing more room for window selection without major change in the correlation accuracy. Fig. 7d shows adjacent δ^{T}s and δ^{V}s over time indicating that the timesteps of higher δ^{V}s need not correspond to higher δ^{T}s, and vice-versa. There exists timesteps where change in both δ^{T} and δ^{V} coincide (i.e., deviate in both time and value).

Based on this empirical insight, *timesteps with changes in both δ^V and δ^T gain precedence for window selection over those with changes in δ^V only, followed by timesteps with changes in δ^T only*. The last condition helps when there are no trends (i.e., values barely change) yet gaps exist in the data. The key idea is to capture timesteps with relatively higher deviations in fewer iterations to form windows. To do so, we filter relatively larger δ^{V}s and δ^{T}s to form an ordered list of location indices that are inspected for window formation starting with size $\mathrm{W_{min}}$. We extend the candidate windows by a fixed step size, until the correlation function `F` reaches an extrema for a signal-pair.

Algo. 2 shows AdaptWin's overall window selection method that identifies the window position in each signal maximizing function `F`. Apart from $\mathrm{W_{min}}$, SI, and the threshold-based filtered δ^{T}s are obtained from Algo. 1. For each signal, the unique observational differences (δ^{V}) with their frequencies (α) are arranged as per the decreasing order of δ^{V}, and a threshold θ is defined from λ^{V} distribution (e.g., p^{th}%ile). As frequent minor deviations in values are normal, fewer yet larger δ^{V}s are preferred. Thus, [$\delta^{\mathrm{V}} > \theta$] related to larger magnitudes are considered. From the threshold-based filtered δ^{T}s and δ^{V}s, the location indices ($\mathrm{L_k}$) are arranged as per the aforementioned order of precedence, i.e., indices where both δ^{V} and δ^{T} change are prioritized over indices with change in either δ^{V} or δ^{T}. For each signal-pair, we consider these location indices (L_k) as various start times (T_{st}) for inspecting candidate windows. For each start time in [L_k], temporary windows of size $\mathrm{W_{min}}$ are formed. To assess F-values, signals are resampled with an identified common SI for a signal-pair, converting an irregular window to a regular one. The windows are extended by β steps while examining indices in [L_k] of both signals until `F` reaches a maximum, or an expected value. This gives window positions of a signal-pair, i.e., the specific start and end times.

Algorithm 2 Flexible Window Selection

Require: $\mathrm{T_W}$=[$\mathrm{T_A}$, $\mathrm{T_B}$]$^{\mathrm{W_{max}}}$, $\mathrm{S_L}$=[S_1,…S_d], g, `F`, β
Ensure: $\mathrm{W_{act}}$, (W_i, W_j) for pair ($\mathrm{S_i}$, $\mathrm{S_j}$) ▷ Selected Windows
procedure ADAPTWIN($\mathrm{T_W}$, $\mathrm{S_L}$, g, F, β)
  W_{min}, ($\mathrm{t_{span}}$, SI)^i ← MinWinSize ($\mathrm{T_W}$, $\mathrm{S_L}$, g) ▷ Algo. 1
  for each $\mathrm{S}_i \in \mathrm{S_L}$ **do**
    [(δ^V, α)] ← Sort δ^{V}s with their frequencies α
    θ ← Derive threshold from δ^{V}s' spread
    [$\mathrm{val_{diff}}$] ← Select (δ^{V}, α), ∀ ($\delta^{\mathrm{V}} > \theta$)
    [L_k] ← Arrange location indices from filtered [$\mathrm{t_{span}}$] and [$\mathrm{val_{diff}}$] as per precedence order ([δ^{V} δ^{T}]>δ^{V}>δ^{T})
  for each ([S_i, S_j] ∈ F) **do** ▷ Signal-pair
    $\mathrm{SI^{com}}$ ← Common SI from [$\mathrm{SI^i}$, $\mathrm{SI^j}$]
    Select candidate windows ∀ ($\mathrm{T_{st}} \in [\mathrm{L_k}]$) with $\mathrm{W_{min}}$
    Resample candidate windows with $\mathrm{SI^{com}}$
    Extend windows by β while checking `F`
    Stop if `F` reaches max; $\mathrm{W_{act}}$, W_i=[T^i_{st}, T^i_E], W_j=[T^j_{st}, T^j_E]
return $\mathrm{W_{act}}$, (W_i, W_j)

For a signal-pair, if SIs are similar (e.g., 2-min vs. 1-min), the maximum interval is chosen, but if one of the SIs is too large compared to the other (e.g, 5-mins≫30-secs), the minimum of the SIs is chosen as the common sampling interval, $\mathrm{SI^{com}}$. This balances the trade-off between losing information via aggregation, and changing the signal shape via new data augmentation. As location indices can change post-resampling, instead of resampling the entire original window at the outset (T_W), only intermediate candidate windows are resampled. We find that the number of shortlisted location indices are $\approx$15$\times$ lower than the number of entries in W_{max} (i.e., all possible start times). By using the indices of relatively higher deviations and not using every timestep as a start time exhaustively, we reduce the number of iterations for window formation. Thus, AdaptWin selects W_i for signal i of size W_{act} with function $\mathtt{F}$.

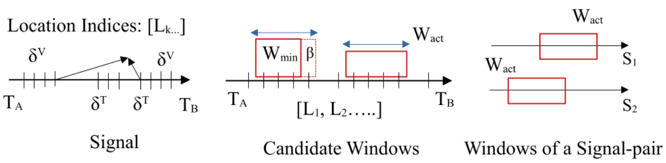


Fig. 8: Window Selection

Fig. 8 shows that for each time-series i, AdaptWin forms an ordered list of location indices corresponding to the selected δ^Vs and δ^Ts (L_k), that are analyzed for window selection for a signal-pair. As $\mathtt{F}$ can be a non-monotonic function, if $\mathtt{F}$-value is lower than its value in the previous iterations (for a maximization problem), or does not increase for a few β steps, further window extensions are skipped. Theoretically, $\beta \geq W_{min}$ is possible. However, in this case W_{min} is chosen from across all signal-pairs that may not be fine-grained enough for incremental updates during pairwise correlations. When $\beta >1$ or a small fraction of total values, and $\beta < W_{min}$, extending a window by β can speedup the window selection process.

Algo. 3 shows our correlation analysis post window selection (Fig. 6). For a set of signals of dataset D, the correlation scores (C_s) are obtained using the selected windows of the considered signal-pairs, and function $\mathtt{F}$. We assess the quality of top $\mathtt{K}$ ranked signals based on these scores using failure ground truth.

Algorithm 3 Correlation Analysis

Require: $W_i = [T^i_{st}, T^i_E]$, $\forall\ i \in$ D, $\mathtt{F}$, $\mathtt{K}$, ground truth ▷ Algo. 2
Ensure: Ranking Quality
procedure CORRELATION(W_i, $\mathtt{F}$, $\mathtt{K}$, ground truth)
    [C_s] ← Pairwise correlation [(W_i, W_j) ∈ D, $\mathtt{F}$]
    [K] ← Top $\mathtt{K}$ ranked signals using [C_s]
    Ranking Quality ← Assess [K] with ground truth
return Ranking Quality

Time Complexity: For a signal-pair with $\mathtt{c}$ entries each in W_{max}, filtering δ^T and δ^V take $\mathcal{O}(\mathtt{c})$ time. With $\mathtt{m}$ and $\mathtt{n}$ location indices of a signal-pair, window selection takes $\mathcal{O}(\mathtt{c+mn})$ time. For $\mathtt{p}$ pairs, the overall complexity is $\mathcal{O}(\mathtt{p(c+mn)})$. For $\mathtt{d}$ signals, the number of pairs $\mathtt{p}$ for anomaly-relevant correlations in our study is $\mathcal{O}(\mathtt{d})$. Generally, $\mathtt{p}$ is not $\mathcal{O}(\mathtt{d}^2)$.

Methods applying a single chosen fixed window take $\mathcal{O}(1)$ for a pair, and $\mathcal{O}(\mathtt{p})$ for $\mathtt{p}$ pairs. The sliding window method for a signal-pair of $\mathtt{c}$ entries each takes $\mathcal{O}(\mathtt{c}^2)$ time, and for $\mathtt{p}$ pairs $\mathcal{O}(\mathtt{pc}^2)$. As $\mathtt{m} \ll \mathtt{c}$ and $\mathtt{n} \ll \mathtt{c}$ in AdaptWin, the number of iterations are much lower than sliding window or brute-force method (§Sec. I) improving runtime. Thus, AdaptWin improves accuracy with time complexity no worse than the commonly used windowing methods. The efficiency can be further improved with the state-of-the-art methods [30, 46] that reduce runtime, which is beyond the scope of this work.

V. Evaluation

We describe datasets, baselines, correlation functions, and evaluation metrics used (§V-A), followed by our experimental results using AdaptWin with practical case studies (§V-B).

A. *Datasets, Baselines, Functions and Metrics*

Datasets: We use the following diverse real-world datasets:

- *Particle Accelerator*: This dataset relates to a complex particle accelerator system (PAS) [26] with heterogeneous subsystems comprising regular and irregular signals of mixed sampling rates. The signals encompass over 25 distinct sampling rates ranging from 0.1 to 120 Hz, some of which have missing entries. The domain experts document the high-level anomaly information, i.e., coarse-grained failure windows, e.g., magnet or pump trips and water system faults that serve as sparse ground truth.
- *Space Weather*: The SWAN logs [40] contain sensor data measured with vector magnetograms while monitoring the Sun [5]. A sudden increase in X-ray flux or magnetic energy indicates anomalies such as solar flares. The signals relate to time and location data including magnetic field parameters containing missing values leading to unevenness. The verified solar flare reports specify the duration of such anomalies that serve as our ground truth.
- *Air Compressor*: This dataset has sensor readings from an urban metro (e.g., pressure/temperature), i.e., the air production unit (APU) [6] of a train containing zeroes but no missing data. Air and oil leakage problems break down the compressor leading to anomalies [42]. The available expert curated report has the specific anomalies with their duration that is leveraged as ground truth for evaluation.

Tab. III shows the statistics of datasets used. We list the number of signals, the overall log *size*, the range of sampling rates (SR), the presence of missing data if any (Miss), the number of failure instances (FI) and the fault types, respectively. The anomaly duration across these datasets ranges between 6-mins to 2-hrs based on which suitable W_{max} values are chosen.

TABLE III: Datasets and Failure Statistics

Dataset	#Signals	Size	SR[1](Hz)	Miss?[2]	#FI[3]	Fault Type
PAS	5511	300K	[0.1-120]	✓	14	Cooler/Magnet/Pump Trips
SWAN	51	20K	0.08	✓	11	Solar Flares
APU	15	500K	0.04	×	10	Air and Oil Leaks

[1] Sampling Rate [2] Missing Values [3] Failure Instances

Baselines: We use the following domain-agnostic methods:

1) *Fixed* (Fix): A fixed size window (common practice).
2) *Adwin* [3]: This univariate method [9] selects windows based on evolving changes in data. Though not bivariate, we consider Adwin as it is a well explored adaptive technique used for anomaly detection tasks [38].

TABLE IV: Ranking quality with various window selection methods (Top-15, W_{max}=3 hrs)

Method	Pearson's Correlation (PC)									Spearman's Correlation (SC)									Dynamic Time Warping (DTW)								
	PAS			APU			SWAN			PAS			APU			SWAN			PAS			APU			SWAN		
	P	RC	NDCG	P	RC	NDCG	P	RC	NDCG	P	RC	NDCG	P	RC	NDCG	P	RC	NDCG	P	RC	NDCG	P	RC	NDCG	P	RC	NDCG
Fix	0.06	0.25	0.2	0.16	0.25	0.3	0.06	0.33	0.3	0.06	0.25	0.3	0.1	0.2	0.2	0.06	0.25	0.3	0.19	0.18	0.3	0.12	0.3	0.3	0.2	0.16	0.47
Adwin	0.13	0.05	0.2	0.06	0.05	0.3	0.06	0.2	0.4	0.06	0.25	0.4	0.01	0.01	0.3	0.06	0.25	0.4	0.23	0.33	0.4	0.16	0.25	0.5	0.26	0.25	0.4
Optwin	0.06	0.07	0.1	0.02	0.02	0.12	0.16	0.33	0.1	0.13	0.33	0.4	0.06	0.16	0.12	0.13	0.16	0.35	0.12	0.32	0.14	0.15	0.26	0.21	0.2	0.3	0.3
Flexwin	0.04	0.06	0.1	0.02	0.3	0.45	0.18	0.25	0.5	0.2	0.4	0.3	0.3	0.4	0.3	0.18	0.34	0.4	0.05	0.14	0.21	0.03	0.21	0.19	0.24	0.3	0.31
AdaptWin	**0.3**	**0.56**	**0.5**	**0.22**	**0.5**	**0.6**	**0.25**	**0.66**	**0.6**	**0.8**	**0.76**	**0.67**	**0.4**	**0.6**	**0.4**	**0.2**	**0.4**	**0.5**	**0.35**	**0.53**	**0.5**	**0.4**	**0.5**	**0.7**	**0.4**	**0.56**	**0.6**

In addition, we adapt two domain-specific non-univariate methods as baselines to assess the robustness of AdaptWin:

1) *Optwin* (Opt) [35]: This multivariate method selects windows to improve classification via Kullback-Leibler (KL) divergence [28]. We customize Optwin for anomalous incidents where windows have several fault types leading to multiple classes. Optwin takes the original and minimum window size as input, similar to W_{max} and W_{min} in Algo. 2. We replace their objective function with our correlation function, i.e., their temporal segmentation strategy is used for examining correlations.
2) *Flexwin* (Flex) [36]: This bivariate method [31] iteratively rejects segments of windows based on several filters. Candidate windows are positioned around local minima, analogous to AdaptWin's location indices related to δ^T and δ^V. For fair comparison, we use Flexwin for cross-correlation threshold- and time lag-based filtering of windows only, that is relevant in our context.

Optwin and Flexwin are designed for purposes different from AdaptWin, yet, we consider these to see the kind of ranking quality achievable via other multivariate windowing strategies relying on techniques of KL divergence, and cross-correlation.

Correlation Functions: We use two popular functions whose variants are heavily used for data analyses [4, 47]:

- *Correlation Coefficient*: Pairwise rank correlation coefficient is a statistical metric used for analyzing linear, non-linear, negative, or positive associations. Specifically, the Pearson's (PC) and Spearman's (SC) correlation coefficients [49] are used for experiments.
- *Dynamic Time Warping* (DTW): DTW determines the similarity of two signals. DTW has been used for data mining in various domains including anomaly detection [22].

Evaluation Metrics: We use the following metrics, generally found effective for ranking and recommendation tasks [14]:

1) *Precision@K* (P): Precision is the number of signals in top-K that are actually related to failure instances.
2) *Recall@K* (RC): RC is the ratio of anomaly-related signals in top-K to the relevant signals in the dataset.
3) *Normalized Discounted Cumulative Gain* (NDCG@K): This measures the relative deviation of the ranked signals compared to the expected ranking as per ground truth.

B. Experiments

Experimental Setup: We use an Intel(R) Xeon(R) Silver 4110 CPU running @ 2.10 GHz with 376 GB of main memory. We set g of Algo. 1 to 1/8 or 1/10, and β of Algo. 2 to 1/12th or 1/18th of W_{max} based on the size and regularity of windows. Fixed-size windows (Fix) use W_{max} in our experiments. We seek to answer the following research questions:

- **RQ1:** *What is the ranking quality obtained via AdaptWin*?
- **RQ2:** *What is the impact of window size and correlation function on the quality of AdaptWin*?
- **RQ3:** *How scalable is AdaptWin with increasing signals*?
- **RQ4:** *What is the runtime of AdaptWin*?

Ranking Quality (RQ1): Tab. IV shows the results with 3-hour windows, where AdaptWin's NDCG is over 2× while recall and precision over 3× compared to `Fix` for several instances. `Adwin` considers too many changes that can lead to false positives, whereas `Optwin` and `Flexwin` can consider timesteps that are not anomaly relevant. Overall, AdaptWin's ranking quality is relatively better across all the datasets.

TABLE V: Window Size and Function Variation (NDCG@15)

Method	PAS (SC)				SWAN (SC)				PAS (2.5-hr)		
	2-hr	4-hr	5-hr	6-hr	2-hr	4-hr	5-hr	6-hr	PC	SC	DTW
Fix	0.4	0.2	0.3	0.1	0.3	0.3	0.2	0.2	0.38	0.4	0.4
Adwin	0.2	0.12	0.12	0.1	0.2	0.1	0.02	0.01	0.25	0.2	0.1
Optwin	0.2	0.1	0.1	0.07	0.22	0.19	0.15	0.15	0.1	0.14	0.2
Flexwin	0.3	0.1	0.1	0.07	0.35	0.2	0.21	0.1	0.35	0.3	0.26
AdaptWin	**0.6**	**0.5**	**0.5**	**0.45**	**0.7**	**0.55**	**0.51**	**0.48**	**0.52**	**0.6**	**0.6**

W_{max} and `F` Variation (RQ2): Tab. V shows the ranking accuracy with variation in window size and correlation function for PAS and SWAN. We skip APU as it has inadequate data for this experiment. First, we fix function SC and vary the window size from 2 hrs to 6 hrs. As evident from columns 2 to 9, increasing W_{max} does not necessarily degrade ranking accuracy as several anomaly-irrelevant signals can also have similar correlations as the relevant ones. In some cases, for a change in window size, NDCG remains the same. Next, we fix the window size to 2.5 hrs and vary the three correlation functions for PAS dataset. As seen in columns 10 to 12, NDCG values for AdaptWin across functions is relatively better than the baselines (we omit SWAN as findings are similar). NDCG considers the position of the relevant signal too. Having a relevant signal show up in top-15 is useful in itself, a higher rank, e.g., 3rd vs. 5th, can further help diagnosis efforts.

To assess AdaptWin's performance beyond PC, SC, and DTW, we use Jaccard similarity (JS) as `F`, a distance measure used in anomaly detection study [19]. The ranking accuracy using JS is shown in Fig. 9 for PAS. AdaptWin has lower false

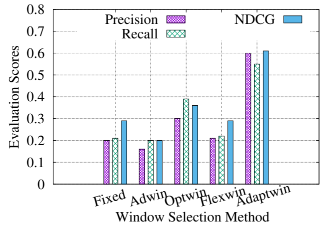


Fig. 9: Ranking (JS)

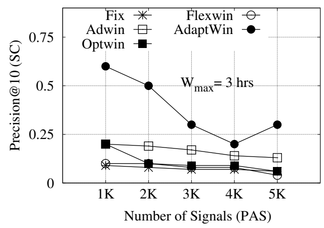


Fig. 10: Precision

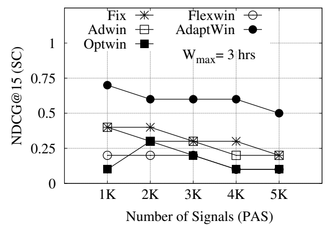


Fig. 11: Accuracy

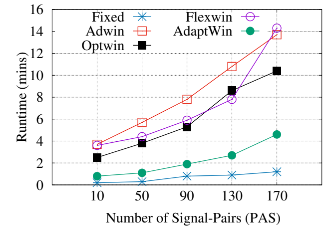


Fig. 12: Runtime (DTW)

positives that improves the ranking quality compared to the baselines (findings are similar for other datasets).

Scalability (RQ3): To analyze ranking quality with scale, we vary the number of signals in PAS as PAS has more signals with higher irregularity than others. Figs. 10 and 11 show the precision@10 and NDCG@15 for function SC. Increase in signals by 5× can lead to as much as 70% decrease in precision, and 2× drop in NDCG. AdaptWin's accuracy is better than the baselines. More signals can include some correlations that are similar to anomaly-irrelevant ones, negatively affecting accuracy. Thus, better ranking accuracy can be harder to achieve with higher number of signals. With a large search space, having a relevant signal appear in top-`K` can be non-trivial even if `K` is large (e.g., 50). AdaptWin can aid fault localization compared to naive methods in such high dimensional signal spaces. We do not try to further optimize pairwise evaluations for a set of signal-pairs as efficient ranking is not our goal [37].

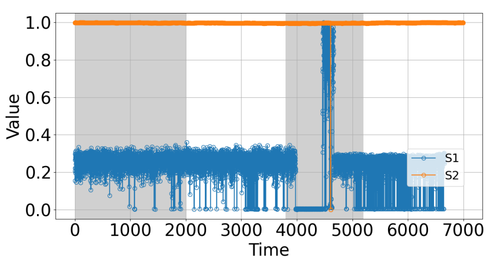


Fig. 13: Signal Deviations

We show how AdaptWin accounts for anomaly relevance through an example. Signals `S1` and `S2` are more correlated during timesteps [0-2000] in Fig. 13. However, AdaptWin picks [3800-5400] as dips and spikes are more causal of anomaly even if the window shows relatively weaker correlation. For abrupt as well as gradual changes (as in Fig. 5), signals with stronger correlation can rank lower and vice versa. AdaptWin's sensitivity to diverse deviations alongside optimizing `F` helps to improve overall anomaly-relevant ranking.

For normal windows (no faults), change in ranks is ≈25% lower than failure windows (not shown for brevity). We assume the availability of coarse failure windows for window selection, which is often the case in practice (§ IV-A). The ranking quality with any windowing method is quite variable as windows can be quite diverse in terms of temporal data distribution.

Efficiency (RQ4): While we strive for better accuracy, Fig. 12 shows AdaptWin's runtime with various number of signal-pairs for DTW on PAS, where W_{max} is 3-hrs. The other datasets with more uniform sampling rates have lower runtimes. `Fixed` has the lowest runtime as expected. With increasing signal-pairs, `Adwin` takes longer time as several candidate windows are assessed that are limited by AdaptWin. AdaptWin amortizes runtime by filtering candidate windows that can be fewer than the baselines in face of asynchrony, as seen in Fig. 12. The decrease in AdaptWin's runtime compared to baselines is larger with a larger number of signal-pairs.

Case Studies: Table VI shows the obtained ranks for a set of signals related to two failure cases across the studied windowing methods, namely, a) a faulty water cooler in PAS, and b) oil leakage in APU. In these cases, water- and oil-based signals are anomaly-relevant, i.e., energy, magnet, and pressure-related signals can rank lower. As evident, AdaptWin (*Adapt*) uplifts the ranks of the anomaly-relevant signals in relation to other adaptive baselines. AdaptWin tends to be beneficial when a failure affects multiple signals of diverse subsystem(s).

TABLE VI: Case Studies (Signal Ranking)

Water System Failure						Oil Leakage Fault					
Signal	**Fix**	**Adwin**	**Opt**	**Flex**	**Adapt**	**Signal**	**Fix**	**Adwin**	**Opt**	**Flex**	**Adapt**
Energy	3	36	41	47	21	Pressure1	14	10	15	7	11
Magnet	17	42	37	23	52	Pressure2	11	15	8	6	13
Water1	33	37	49	56	**12**	Pressure3	9	12	9	5	14
Water2	35	29	44	40	**18**	Oil1	12	9	13	12	**7**
Water3	49	47	31	61	**10**	Oil2	10	10	12	11	**8**

Practical Impact: In noisy time-series, different windows of the same signal can lead to different correlations due to which distinguishing between normal and failure times is hard. AdaptWin helps when the data is aperiodic or lacks symmetry in distribution, and faults lead to temporal deviations for at least a fraction of the coarser failure window (usual scenario). Anomalies with gradual changes can lead to more location indices for candidate window selection, while those with longer failure durations can help better capture time-localized deviations. Spurious correlations though feasible are rare when dense and sparse time-series are correlated for a dynamic system with multiple incidents in close temporal proximity. Sliding window-based analysis do not intrinsically prioritize δ^Vs for which anomaly relevance can be low. If the window size is too small, a method may fail to learn important signal associations. For other scenarios where identifying atypical signal correlations is needed, windows can afford to accommodate some delays. AdaptWin can enable such flexibility in window selection for various application domains.

VI. Conclusion

We present AdaptWin, an alternative approach of generic window selection from time-series for pairwise correlations in anomalous environments. Our design selects windows that improve correlation-aware ranking involving regular and irregular signals with mixed sampling rates. AdaptWin enhances anomaly-relevant ranking by over 3× compared to baselines. Our method can be used in domains with sparse ground truth involving multivariate signals for failure diagnosis.

Acknowledgment

We are grateful to the anonymous reviewers for their feedback and acknowledge the use of resources from the SLAC Shared Science Data Facility (S3DF). This work is supported by LCLS development program, the U.S. DOE under contract DE-AC02-76SF00515 at SLAC NAL, and NSF award #2119184.

References

[1] S. Abbady, C. Ke, J. Lavergne, J. Chen, V. V. Raghavan, and R. Benton, "Online mining for association rules and collective anomalies in data streams," in *IEEE BigData*, 2017.

[2] A. Abdulaal, Z. Liu, and T. Lancewicki, "Practical approach to asynchronous multivariate time series anomaly detection and localization," in *ACM SIGKDD*, 2021.

[3] *Adwin*, https://github.com/scikit-multiflow/scikit-multiflow.

[4] Y. Ang, Q. Huang, A. K. H. Tung, and Z. Huang, "A stitch in time saves nine: Enabling early anomaly detection with correlation analysis," in *IEEE ICDE*, 2023.

[5] R. A. Angryk, P. C. Martens, B. Aydin *et al.*, "Multivariate time series dataset for space weather data analytics," *Scientific data*, p. 227, 2020.

[6] *Metro Train*, https://doi.org/10.5281/zenodo.7766691.

[7] Y. Bai, L. Wang, Y. Liu, Y. Yin, and Y. Fu, "Dual-side auto-encoder for high-dimensional time series segmentation," in *ICDM*. IEEE, 2020.

[8] T. Bi, Y. Zhang, Y. Pan, Y. Zhang, M. Ma, X. Jiang, L. Han, F. Wang, X. Liu, and P. Wang, "FaultInsight: Interpreting hyperscale data center host faults," in *ACM SIGKDD*, pp. 141–152.

[9] A. Bifet and R. Gavaldà, "Learning from time-changing data with adaptive windowing," in *SIAM SDM*, 2007.

[10] R. Cachucho, M. Meeng, U. Vespier, S. Nijssen, and A. J. Knobbe, "Mining multivariate time series with mixed sampling rates," in *ACM UbiComp*, 2014.

[11] M. A. M. Carrasco and C. Wu, "An unsupervised framework for anomaly detection in a water treatment system," in *IEEE ICMLA*, 2019.

[12] F. Chen, Y. Zhang, Z. Qin, L. Fan, R. Jiang, Y. Liang, Q. Wen, and S. Deng, "Learning multi-pattern normalities in the frequency domain for efficient time series anomaly detection," in *IEEE ICDE*, 2024.

[13] X. Chen, L. Deng, F. Huang, C. Zhang, Z. Zhang, Y. Zhao, and K. Zheng, "DAEMON: Unsupervised anomaly detection and interpretation for multivariate time series," in *ICDE*. IEEE, 2021, pp. 2225–2230.

[14] X. Chen, L. Deng, Y. Zhao, and K. Zheng, "Adversarial autoencoder for unsupervised time series anomaly detection and interpretation," in *ACM WSDM*, 2023.

[15] U. Draisbach, F. Naumann, S. Szott, and O. Wonneberg, "Adaptive windows for duplicate detection," in *IEEE ICDE*, 2012, pp. 1073–1083.

[16] A. Ermshaus, P. Schäfer, and U. Leser, "Window size selection in unsupervised time series analytics: A review and benchmark," in *AALTD Workshop, ECML PKDD*. Springer, 2022.

[17] C. Fahy, S. Yang, and M. Gongora, "Scarcity of labels in non-stationary data streams: A survey," *ACM Comput. Surv.*, vol. 55, no. 2, 2023.

[18] L. Felsberger, A. Apollonio, T. Cartier-Michaud *et al.*, "Explainable deep learning for fault prognostics in complex systems: A particle accelerator use-case," in *CD-MAKE*. Springer, 2020.

[19] W. Gu, X. Sun, J. Liu, Y. Huo, Z. Chen, J. Zhang, J. Gu, Y. Yang, and M. R. Lyu, "KPIRoot: Efficient monitoring metric-based root cause localization in large-scale cloud systems," in *IEEE ISSRE*, 2024.

[20] X. Hao, Y. Chen, C. Yang, Z. Du, C. Ma, C. Wu, and X. Meng, "From chaos to clarity: Time series anomaly detection in astronomical observations," in *ICDE*. IEEE, 2024.

[21] O. Ibidunmoye, A. Rezaie, and E. Elmroth, "Adaptive anomaly detection in performance metric streams," *IEEE Trans. Netw. Serv. Manag.*, 2018.

[22] M. Imamura and T. Nakamura, "Parameter-free spikelet: Discovering different length and warped time series motifs using an adaptive time series representation," in *ACM SIGKDD*. ACM, 2023, pp. 857–866.

[23] S. Imani and E. Keogh, "Multi-window-finder: Domain agnostic window size for time series data," *Proceedings of the MileTS*, vol. 21, 2021.

[24] S. Y. Jhin, J. Lee, and N. Park, "Precursor-of-anomaly detection for irregular time series," in *ACM SIGKDD*. ACM, 2023, pp. 917–929.

[25] Y. Kawahara, T. Yairi, and K. Machida, "Change-point detection in time-series data based on subspace identification," in *IEEE ICDM*, 2007.

[26] *Linac Coherence Light Source*, https://lcls.slac.stanford.edu/.

[27] Z. Li, Y. Zhao, J. Han, Y. Su, R. Jiao, X. Wen, and D. Pei, "Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding," in *ACM SIGKDD*, 2021.

[28] J. Liono, A. K. Qin, and F. D. Salim, "Optimal time window for temporal segmentation of sensor streams in multi-activity recognition," in *ACM MobiQuitous*, 2016.

[29] J. Liu, T. Yang, Z. Chen, Y. Su, C. Feng, Z. Yang, and M. R. Lyu, "Practical anomaly detection over multivariate monitoring metrics for online services," in *IEEE ISSRE*, 2023.

[30] Y. Lu, T. V. A. Srinivas, T. Nakamura, M. Imamura, and E. J. Keogh, "Matrix profile XXX: MADRID: A hyper-anytime and parameter-free algorithm to find time series anomalies of all lengths," in *ICDM*, 2023.

[31] A. Maggi, C. Tape, M. Chen, D. Chao, and J. Tromp, "An automated time-window selection algorithm for seismic tomography," *Geophysical Journal International*, 2009.

[32] D. Marcato, G. Arena, D. Bortolato, F. Gelain *et al.*, "Machine learning-based anomaly detection for particle accelerators," in *IEEE CCTA*, 2021.

[33] S. Maya, A. Yamaguchi, K. Nishino, and K. Ueno, "Lag-aware multivariate time-series segmentation," in *SIAM SDM*, 2020.

[34] A. J. Oliner, A. V. Kulkarni, and A. Aiken, "Using correlated surprise to infer shared influence," in *IEEE/IFIP DSN*, 2010, pp. 191–200.

[35] *Optwin*, https://github.com/cruiseresearchgroup/OPTWIN/.

[36] *Pyflex*, https://github.com/adjtomo/pyflex.

[37] N. B. Shah and M. J. Wainwright, "Simple, robust and optimal ranking from pairwise comparisons," *J. Mach. Learn. Res., JMLR*, 2017.

[38] B. Shayesteh, C. Fu, A. Ebrahimzadeh, and R. H. Glitho, "Auto-adaptive fault prediction system for edge cloud environments in the presence of concept drift," in *IEEE IC2E*, 2021, pp. 217–223.

[39] Y. Sun, G. Pang, G. Ye, T. Chen, X. Hu, and H. Yin, "Unraveling the 'anomaly'in time series anomaly detection: A self-supervised tri-domain solution," in *ICDE*. IEEE, 2024, pp. 981–994.

[40] *Space Data*, https://doi.org/10.7910/DVN/EBCFKM.

[41] M. D. L. Tosi and M. Theobald, "OPTWIN: Drift identification with optimal sub-windows," in *ICDE Workshops, ICDEW*. IEEE, 2024.

[42] B. Veloso, R. P. Ribeiro, J. Gama, and P. M. Pereira, "The MetroPT dataset for predictive maintenance," *Scientific Data*, vol. 9, no. 1, 2022.

[43] Y. Wang, Z. Zhu, Q. Fu, Y. Ma, and P. He, "MRCA: Metric-level root cause analysis for microservices via multi-modal data," in *IEEE/ACM Automated Software Engineering, ASE*, 2024, pp. 1057–1068.

[44] C. Wu, N. Zhao, L. Wang, X. Yang, S. Li, M. Zhang, X. Jin, X. Wen, X. Nie, W. Zhang, K. Sui, and D. Pei, "Identifying root-cause metrics for incident diagnosis in online service systems," in *IEEE ISSRE*, 2021.

[45] L. Yan, B. Zheng, J. Qing, W. You, T. Chen, Z. Xu, S. Liu, K. Zeng, T. Ye, and X. Zhou, "Anomaly diagnosis with siamese discrepancy networks in distributed cloud databases," in *IEEE ICDE*, 2025, pp. 4053–4065.

[46] C. M. Yeh, A. Der, U. S. Saini, V. Lai *et al.*, "Matrix profile for anomaly detection on multidimensional time series," in *IEEE ICDM*, 2024.

[47] G. Zhang, C. Li, K. Zhou, L. Liu, C. Zhang, W. Chen, H. Fang, B. Cheng, J. Yang, and J. Xing, "DBCatcher: A cloud database online anomaly detection system based on indicator correlation," in *IEEE ICDE*, 2023.

[48] L. Zhang, N. Patel, X. Li, and J. Lin, "Joint time series chain: Detecting unusual evolving trend across time series," in *SIAM SDM*, 2022.

[49] L. Zhang, "Systematically exploring associations among multivariate data," in *Conference on Artificial Intelligence, AAAI*, 2020.

[50] Z. Zhang, S. Zhang, Y. Jiang, and Z. Zhou, "LIFE: Learning individual features for multivariate time series prediction with missing values," in *IEEE ICDM*, 2021.

[51] Y. Zhao and M. Udell, "Missing value imputation for mixed data via Gaussian Copula," in *ACM KDD*, 2020.

[52] Y. Zheng, H. Y. Koh, M. Jin, L. Chi *et al.*, "Correlation-aware spatial-temporal graph learning for multivariate time-series anomaly detection," in *IEEE Trans. on Neural Networks Learn. Syst.*, 2023.

[53] S. Zhong, V. M. A. de Souza, and A. Mueen, "Combining filtering and cross-correlation efficiently for streaming time series," *ACM Trans. Knowl. Discov. Data*, 2022.

[54] L. Zhou, M. Poli, W. Xu, S. Massaroli, and S. Ermon, "Deep latent state space models for time-series generation," in *ICML*. PMLR, 2023.