

# Flexible Windowing for Correlation-Aware Ranking in Anomalous Environments

Anwasha Das<sup>1, 2</sup>      Henry Hoffmann<sup>1</sup>      Alex Aiken<sup>2</sup>

<sup>1</sup>The University of Chicago, IL      <sup>2</sup>SLAC National Accelerator Laboratory/Stanford University

IEEE International Conference on Data Mining (ICDM)

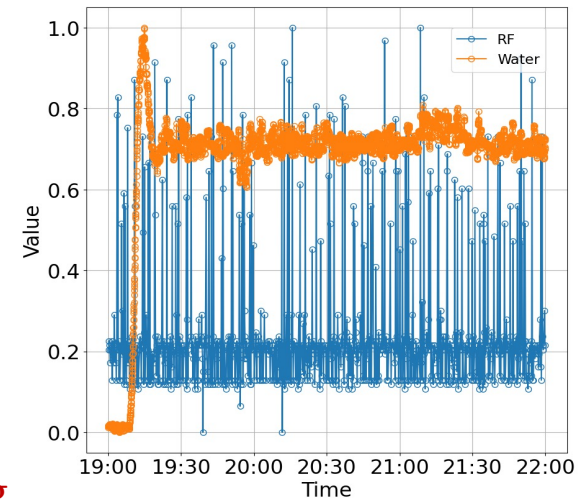
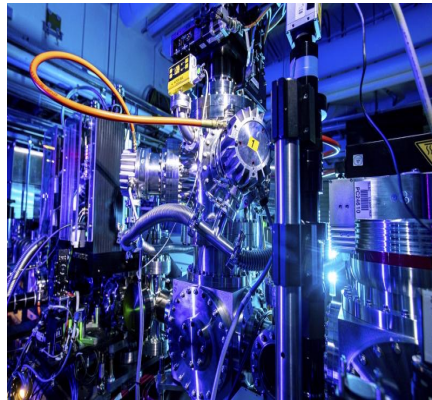


12 – 15<sup>th</sup> November 2025



# Time-Series Correlations for Anomaly Detection

- Incident diagnosis and anomaly detection in systems and facilities
  - **Multivariate** time-series data monitored from diverse sources
  - **Pairwise signal correlations** often used to rank signals (e.g., fault localization)
  - **Time-windows** of signals analyzed as opposed to full signals



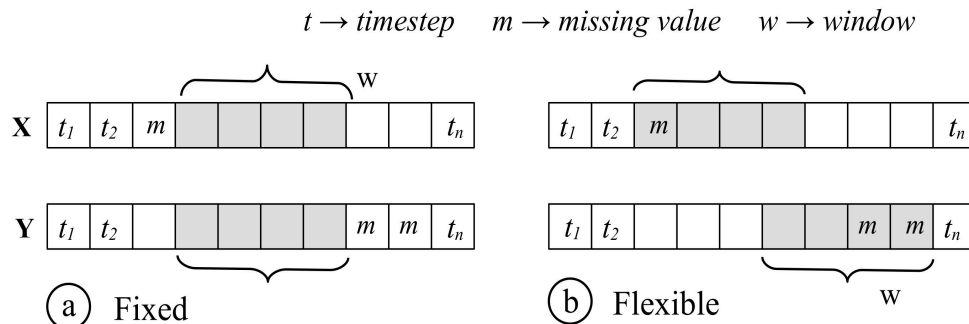
- What kind of task? Accurate anomaly-relevant signal **ranking**
  - **Bivariate function** leading to correlation scores
  - Multivariate associations can be derived from bivariate relations
  - Score-based ranking (assess **only top-k** signals)

*How to select windows from signals for accurate anomaly-relevant ranking?*

# Window Selection in Anomalous Environments

## → Prevalent window selection methods

- Fixed size, sliding window with a step size, multiple arbitrary sizes
- Suitable when (a) in Figure)
  - Adequate/**precise** failure ground truth (label) is available
  - **Same/similar** sampling rates, mostly **regular** (no missing data)
- Inadequate (inaccurate results) when (b) in Figure)
  - Failure ground truth → Unavailable or **coarse-grained**
  - **Mixed sampling rates**, lots of **irregularities**
  - **Asynchronous** data from **heterogenous** sources

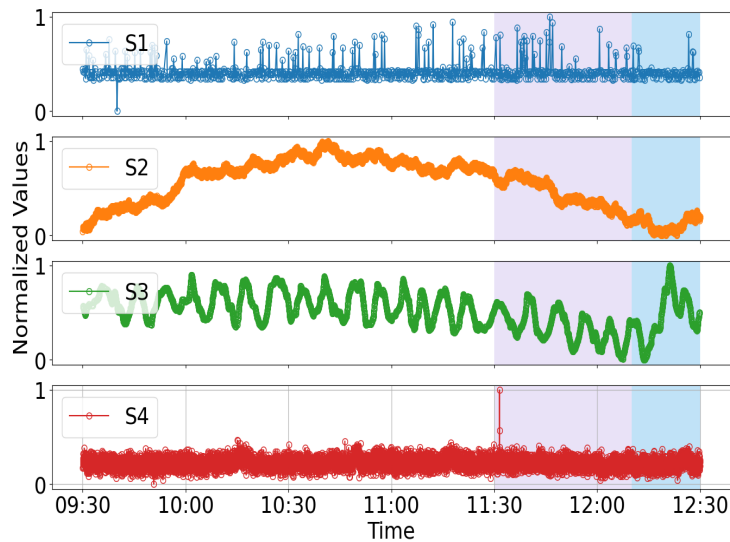


*Even for regular signals there can be lags in temporal variation !!*

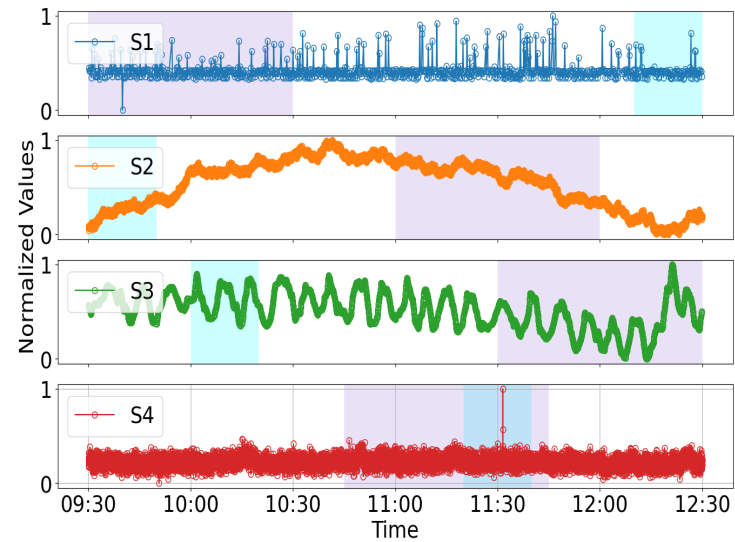
*How to select windows from pairwise signals in presence of asynchrony, irregularity, and mixed sampling rates?*

# Window Selection in Anomalous Environments

- Select windows with available coarse-labels
  - Such that bivariate scores are better aligned with anomalous times



Fixed Windows  
(same positions, i.e., start/end times)

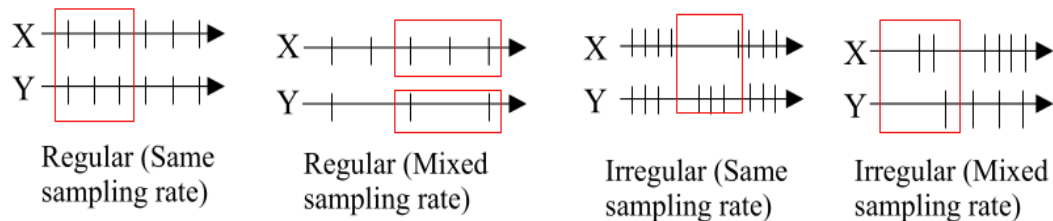


Flexible Windows  
(different positions)

*Fixed-windows: Time-localized deviations may not be well aligned with anomalies if same start and end times of windows are considered across diverse signals !!*

# Our Method: AdaptWin

- Leverage anomaly-indicative symptoms (dips/spikes, graduate/abrupt changes)
  - Inter-arrival times ( $\delta^V$ ), Value-differences ( $\delta^T$ )
  - *Key idea: Adapt the window position and size based on filtered  $\delta^V$  and/or  $\delta^T$*
  - Filter? Configurable threshold based on distribution of  $\delta^V$ s and  $\delta^T$ s
  - Minimum window size selection ( $W_{\min}$ ): Lower bound
  - Goal: Better **accuracy** in ranking quality (**offline** process)



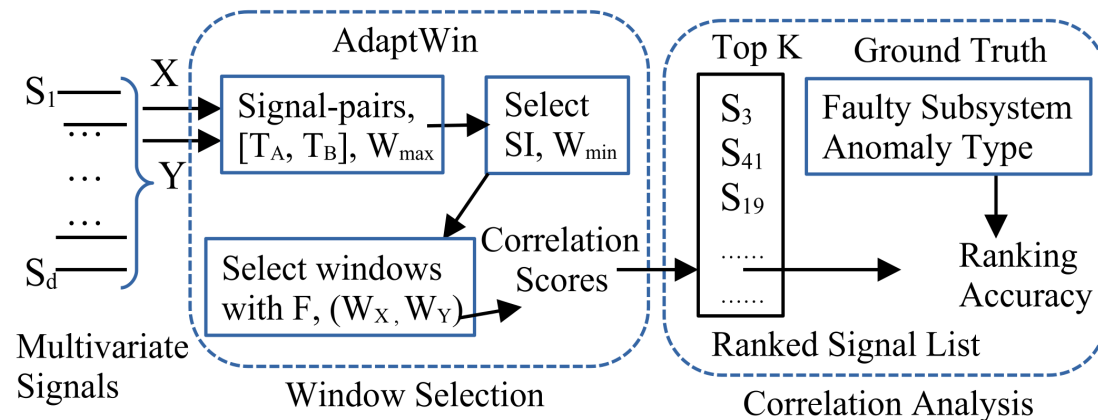
## → Non-goals

- No new ranking method
- No better efficiency: Fast window selection with scale
- No real-time requirements

# AdaptWin

## → Window Selection

- Specific **signal-pairs given** for known failure ground truth (context-relevant/manual)
- Select *minimum* and then *actual* windows
- Obtain pairwise scores with correlation functions (e.g., correlation coefficients)



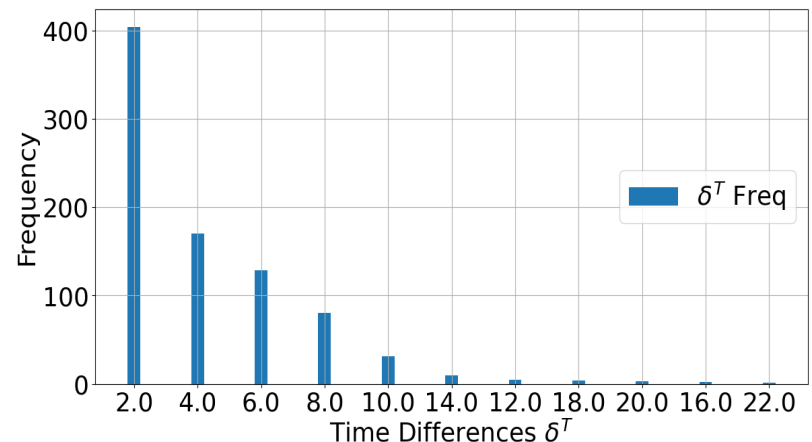
## → Correlation Analysis

- Analyze ranks with selected windows
- Ground truth validation (top K)
- Comparative analysis (e.g., fixed etc.)

# AdaptWin

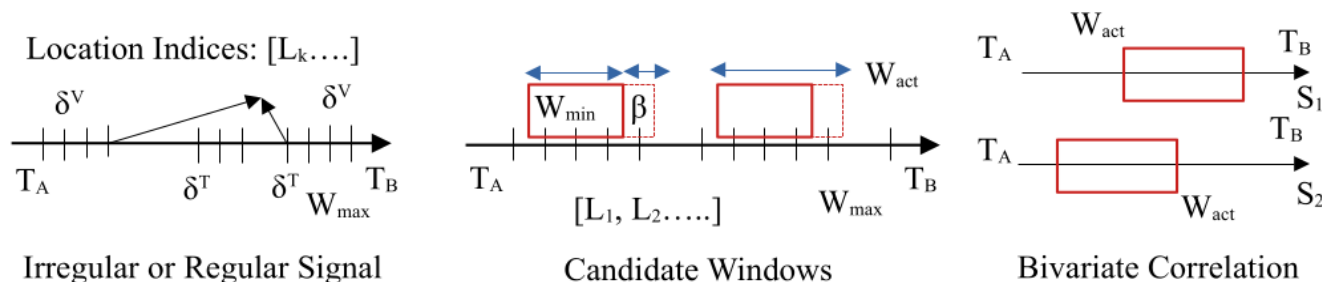
- Set of signal-pairs  $[S_i, S_j]$ , Correlation Function  $F$ , Original window size:  $W_{\max}$ 
  - $[\lambda, \delta^T]$  ( $\lambda$ : frequency of distinct  $\delta^T$  s)
  - Identify sampling interval (SI) (mode: most frequent  $\delta^T$ )
- Minimum Window Selection
  - Regular signal ( $\delta^T$  do not vary):  $t_{\text{span}} = A$  fraction of  $W_{\max}$  ( $1/g$ )
  - Irregular signal:
    - Sort and filter  $\delta^T$  s
    - Temporal gaps:  $t_{\text{span}} = \Sigma (\delta^T * \lambda)$  for filtered  $\delta^T$  s
    - $\text{Pair}_{\min} = \max$  of ( $t_{\text{span}}$  s),  $W_{\min} = \max$  of ( $\text{Pair}_{\min}$  s)

*Intuition: Long enough to capture major patches of unevenness in the window !!*



# AdaptWin

- Actual Window Selection
  - ➔ Unique observational differences: [Frequency  $\alpha$ ,  $\delta^V$ ]
  - ➔ Sort and filter  $\delta^V$  s
  - ➔ *Prioritize* time-steps with changes in **both  $\delta^V$  and  $\delta^T$**  > those with  **$\delta^V$  only** > those with  **$\delta^T$  only**:  $[(\delta^V, \delta^T) > \delta^V > \delta^T]$
  - ➔ Prioritizing *reduces number of candidate windows* (not all possible start time-steps)
    - ➔ Shortlisted Location Indices  $[L_k]$ : *Lower iterations*
- ➔ For each *signal-pair*: Common SI (max or min of SI-pair)
  - ➔ Candidate windows  $W$  with start times in  $[L_k]$  with size  $W_{\min}$
  - ➔ Resample with common SI
  - ➔ Extend  $W$  by fixed time-steps ( $\beta$ ) while optimizing  $F$
  - ➔ Obtain actual size:  $W_{\text{act}}$  with specific window positions





# Evaluation

## ➤ Datasets from three domains:

- High-Energy Physics<sup>1</sup> (Particle Accelerator System; PAS)
- Space Weather (SWAN)
- Train Air Compressor (APU)

## ➤ Overall Failure Duration: 6-mins to 2-hours

Dataset	Signals	Size	Sampling Rate (Hz)	Missing values?	# Failure Instances	Fault Type
PAS	5511	300K	[0.1 – 120]	Yes	14	Water Cooler, Magnet, Pump Trips
SWAN	51	20K	0.08	Yes	11	Solar Flares
APU	15	500K	0.04	No	10	Air/Oil Leaks

- Ground truth example: “*Water cooler Z failed for  $\approx 30$ mins between 10:00 and 18:00 hrs*”,  
Unavailable details: Signals X and Y of Z was anomalous between 11:00 and 11:30 hrs

<sup>1</sup>Linac Coherence Light Source (LCLS/SLAC)

# Evaluation

## ➤ **Baselines:**

- *Domain-agnostic (univariate)*: a) Fixed, b) Adwin

- *Domain-specific (multivariate)*: a) Optwin, b) Flexwin

## ➤ **Evaluation Metrics for top-K Ranking:**

- Precision@K (P)

- Recall@K (R)

- Normalized Discounted Cumulative Gain (NDCG@K)

## ➤ **Bivariate Correlation Functions:**

- Pearson Correlation (PC)

- Spearman Correlation Coefficient (SC)

- Dynamic Time Warping (DTW)

# Results

→  $W_{\max} = 3\text{-hour}$ , Top-15 signals, AdaptWin obtains better ranking **quality** (RQ1)

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	<b>0.3</b>	<b>0.56</b>	<b>0.5</b>	<b>0.22</b>	<b>0.5</b>	<b>0.6</b>	<b>0.25</b>	<b>0.66</b>	<b>0.6</b>	<b>0.8</b>	<b>0.76</b>	<b>0.67</b>	<b>0.4</b>	<b>0.6</b>	<b>0.4</b>	<b>0.2</b>	<b>0.4</b>	<b>0.5</b>	<b>0.35</b>	<b>0.53</b>	<b>0.5</b>	<b>0.4</b>	<b>0.5</b>	<b>0.7</b>	<b>0.4</b>	<b>0.56</b>	<b>0.6</b>

→ **Variation** of  $W_{\max}$  and bivariate function (RQ2):

→ NDCG@15: Performance is similar for certain other windowing methods

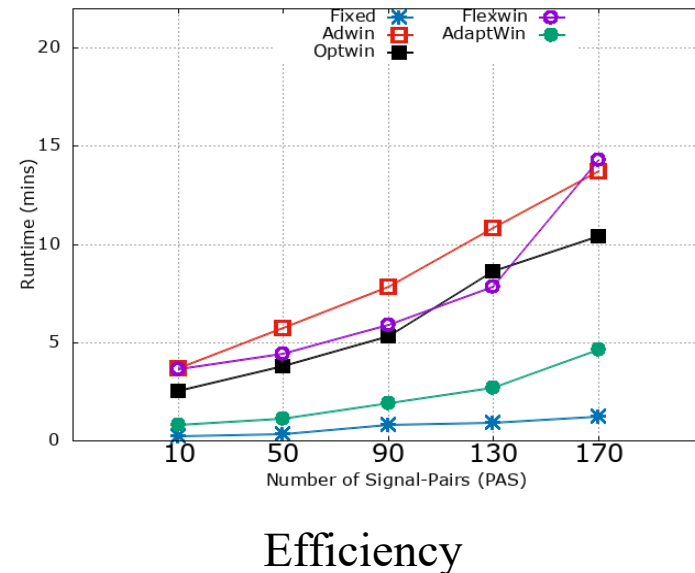
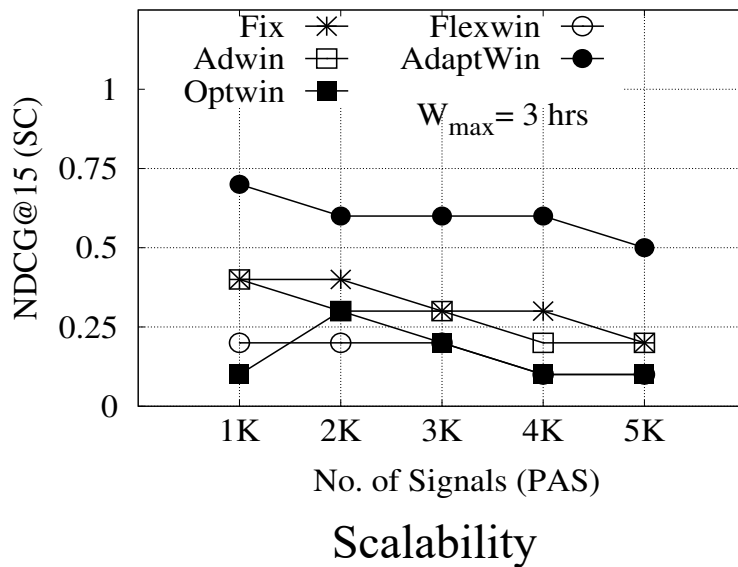
→ Position of relevant signals better with AdaptWin

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	<b>0.6</b>	<b>0.5</b>	<b>0.5</b>	<b>0.45</b>	<b>0.7</b>	<b>0.55</b>	<b>0.51</b>	<b>0.48</b>	<b>0.52</b>	<b>0.6</b>	<b>0.6</b>

*AdaptWin is relatively more accurate overall !*

# Results

→ **Scalability**: With increasing signals, the ranking quality is much better than Fix (**RQ3**)



→ **Efficiency**: With increasing signals, runtime increases from 0.8 to 4.6 mins for AdaptWin, which is lower than other multivariate methods (**RQ4**)

*The runtime considering better ranking quality is worthwhile for AdaptWin !*

# Results

- Time Complexity:  $p$  signal-pairs with  $c$  entries each
- Fixed Size:  $O(p)$
  - Sliding Windows:  $O(pc^2)$
  - Brute-force:  $O(pc^2m)$ ,  $m$  unique window sizes
  - AdaptWin:  $O(p(c+mn))$ ,  $m$  and  $n$  are number of location indices of a pair,  $[m \ll c, n \ll c]$
- Case Study: Boosts ranks of anomaly-indicative signal(s) for a faulty **water** cooler

Signal	Fix	Adwin	Optwin	Flexwin	Adaptwin
Energy	3	36	41	47	21
Magnet	17	42	37	23	52
Water1	33	37	49	56	<b>12</b>
Water2	35	29	44	40	<b>18</b>
Water3	49	47	31	61	<b>10</b>

# Conclusion

- Adaptive window selection from a signal-pair
  - 3x better anomaly-relevant ranking
  - Suitable? Asymmetry, Aperiodicity, High dimensional space, Sparse labels
  - Accuracy vs. runtime trade-off worthwhile with scale
- Future Work
  - Optimization with pairwise evaluations
  - Other tasks beyond ranking, e.g., regression problems
  - Better bounds based on the nuances of anomalies

*AdaptWin selects windows for correlations that improve anomaly-relevant ranking !*

*Thank You*

# **Backup Slides**

# Related Work

- Univariate methods
  - Concept drift estimation, Change point detection, Imputation
  - Time-series segmentation, duplicate detection
  - Not adequate for bivariate relations
- Multivariate methods
  - Human activity recognition, Localization in seismograms
  - Feature aggregation, Lagged Correlations, Trend Analysis
  - Domain-specific or do not consider anomalous environments
- Some of the univariate and multivariate methods do not consider *irregularities* or *mixed* sampling rates