Detection of Slowly Developing Anomalies in Engineering Systems Based on Matrix Profile Family Algorithms and Multivariate Time Series Representation Methods

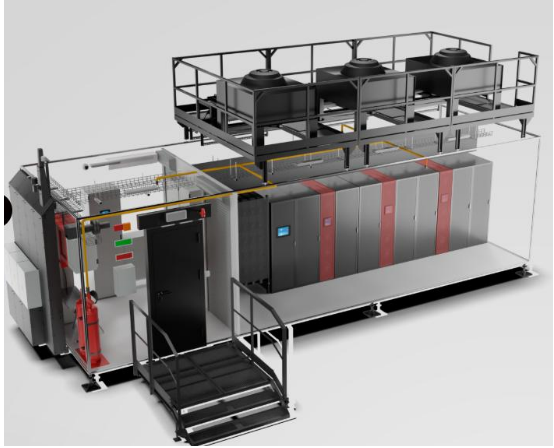
Anastasiia Kalita[1], Sergey Ivanov[1], Vasilii Borisov[1], Alexander Manzhurov[1], Stanislav Polyakov[1], Mikhail Ronkin[1], Mikhail Hushchyn[2], Pavel Solovev[3]

[1] Ural Federal University, Yekaterinburg [2] HSE University, Moscow [3] DATARK LLC, Yekaterinburg

The 9th International Conference in Deep Learning in Computational Physics

Slowly Developing Anomalies

- Engineering systems (industrial plants, data-center HVAC) continuously stream large multivariate telemetry.
- Threshold monitoring reacts only after a violation: readings stay within limits while signal dynamics has already left the normal regime.
- Typical slow faults: refrigerant leakage, heat-exchanger fouling, sensor drift - hours to weeks of gradual deviation.
- Labeled anomalies are rare in practice.



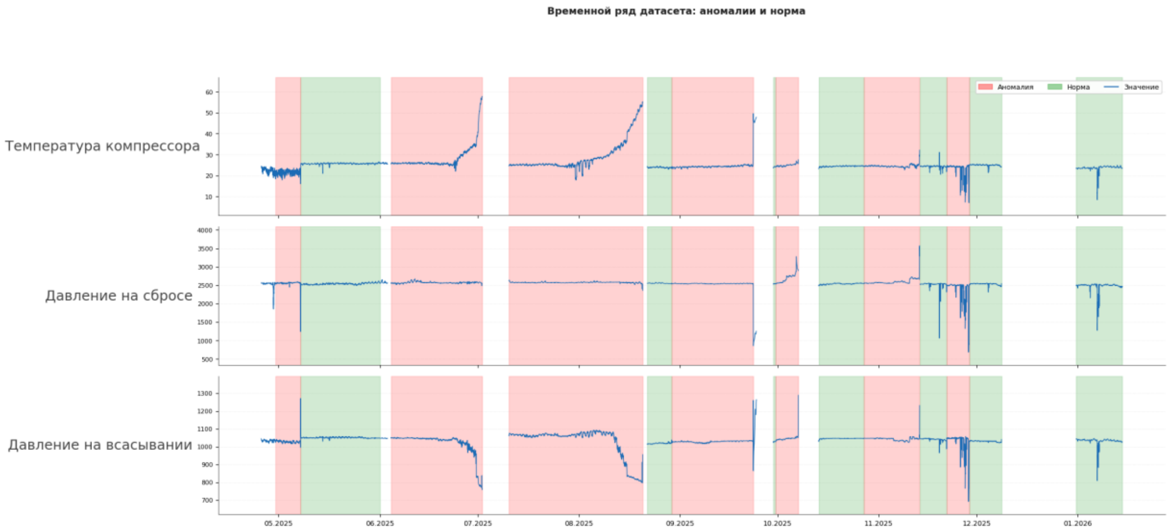
Data-center HVAC cooling units - the monitored system

Why Matrix Profile?

Deterministic, no supervised training stage, streaming-capable, the distance to the nearest non-trivial neighbor flags a pattern never seen in normal history.

Goal: compare Matrix Profile algorithms combined with representation schemes vs. ML/DL methods for early detection of slow anomalies.

Dataset: Data-Center HVAC Telemetry



Telemetry channels, green - normal operation, red - verified anomalous intervals

Dataset

- 60 channels: temperatures, refrigerant pressures, compressor, fans, energy
- 30-min sampling interval
- Point-wise Normal / Anomaly labels from verified incidents

Preprocessing

Resampling → removal of constant and highly correlated channels → per-channel z-normalization → chronological train / val / test split

Problem Statement & Evaluation Protocol

Primary metric — Detection Delay

$$\Delta t = t(\text{first alarm}) - t(\text{anomaly onset}),$$
$$\text{FPR} \leq 0.05$$

- First trigger after anomaly onset, missed segments penalized with the full segment length.
- Reflects the industrial priority, early warning at a controlled false-alarm rate.

Hyperparameters & threshold

- Window m: grid {48, 96, 144, 240, 336, 480, 672}, max MCC on validation.
- Threshold: among all ROC cutoffs with FPR ≤ 0.05, pick the one minimizing delay on validation.

Robustness estimation

Sliding-window cross-validation on the training set (step = 5 %). All metrics reported as mean ± std over iterations.

Secondary metrics

- MCC — threshold metric robust to class imbalance (uses TP, TN, FP, FN).
- FPR.

Matrix Profile Family Algorithms

MP_i = min over j ($|i-j| > m/2$) of dist($Z_{i:i+m}$, $Z_{j:j+m}$) - z-normalized Euclidean distance to the nearest non-trivial neighbor.

STOMP	Batch algorithm.
DAMP	Streaming left matrix profile, neighbors searched only in the past.
mSTOMP	Native multivariate extension with post-sorting.
Multidim MP	Native multivariate extension with pre-sorting.

	*	6.9	*	*	*	*	*	*	*
*		*	*	*	*	*	*	1.4	*
*	*		*	*	*	*	*	*	6.2
*	7.9	*		*	*	*	*	*	*
*	*	*	*		*	*	*	*	11.4
*	*	13.6	*	*		*	*	*	*
*	*	*	*	*	*		14.1	*	*
*	*	14.0	*	*	*	*		*	*
*	1.4	*	*	*	*	*	*		*
*	*	6.2	*	*	*	*	*	*	

6.9
1.4
6.2
7.9
11.4
13.6
14.1
14.0
1.4
6.2

STOMP & DAMP operate on a univariate signal → the multivariate series must first be reduced to one dimension.

Six Representation Schemes

Scheme	Type	Resulting univariate anomaly signal
Per-channel	aggregation	Element-wise max of per-channel normalized profiles
PCA	linear	First principal component of the normalized series
UMAP	nonlinear manifold	Norm of low-dimensional embedding
AE	neural, reconstruction	Reconstruction MSE
VAE	neural, probabilistic	Reconstruction MSE + KL term
Contrastive	self-supervised	Euclidean distance to the centroid of normal embeddings

All encoders are trained only on normal observations (200 epochs, Adam)

Comparison Methods & Experimental Conditions

#	Methods	Features	Training data
1	Matrix Profile (14 configs)	full set	normal only
2	Matrix Profile (14 configs)	selected	normal only
3	IF, OC-SVM, DAGMM, OmniAnomaly, TranAD, MTAD-GAT, TS2Vec, PatchTST	full set	normal only
4	SVC, Random Forest + DL models	full set	normal + anomalous

Feature selection (MP only)

Channels ranked by per-channel STOMP profile MCC on validation.

Same threshold rule for all

Every method is tuned on validation and evaluated at the fixed operating point FPR ≤ 0.05.

Results - Condition 1 (Full Feature Set)

Configurations satisfying the median FPR ≤ 0.05 constraint (4 of 14):

Method	Delay, min ↓	FPR ↓	MCC ↑
STOMP | UMAP	5397 ± 1345	0.02 ± 0.02	0.22 ± 0.13
STOMP | VAE	5158 ± 229	0.03 ± 0.01	0.19 ± 0.09
STOMP | Contrastive	5667 ± 653	0.03 ± 0.02	0.14 ± 0.09
DAMP | PCA	4345 ± 1361	0.04 ± 0.01	0.07 ± 0.06

Filtered out

mSTOMP, Multidim MP and all DAMP variants except DAMP | PCA systematically exceed FPR ≤ 0.05.

Best MCC — STOMP | UMAP (0.22), most reproducible delay — STOMP | VAE (± 229 min).
For the three leading variants the grid search selects the maximum window m = 672.

Results - Condition 2 (Feature Selection)

Method	Delay, min ↓	FPR ↓	MCC ↑
STOMP | Contrastive	**4960 ± 134**	**0.01 ± 0.01**	**0.27 ± 0.10**
STOMP | UMAP	5952 ± 1111	0.03 ± 0.02	0.16 ± 0.17
DAMP | PCA	4843 ± 1393	0.04 ± 0.01	0.04 ± 0.03
DAMP | Contrastive	5770 ± 1546	0.03 ± 0.01	0.09 ± 0.05
DAMP | VAE	6500 ± 2036	0.05 ± 0.01	0.02 ± 0.05

Comparison Methods: Semi and Fully Supervised

Method	Delay, min ↓	FPR ↓	MCC ↑	Condition
SVC	5335 ± 85	0.01 ± 0.01	0.29 ± 0.03	4 (labeled)
Random Forest	6717 ± 3271	0.04 ± 0.02	0.26 ± 0.10	4 (labeled)
DAGMM	12162 ± 3239	0.03 ± 0.02	0.11 ± 0.05	3 (normal only)
Isolation Forest	15165 ± 2652	0.00 ± 0.00	0.03 ± 0.02	3 (normal only)
One-Class SVM	16362 ± 805	0.00 ± 0.00	0.08 ± 0.03	3 (normal only)
OmniAnomaly / TranAD / MTAD-GAT / TS2Vec / PatchTST	—	—	0 segments detected	3

5 of 8 deep-learning methods detected no anomalous segment in the semi-supervised setting: reconstruction/contrastive architectures adapt to normal behavior and the residual on slow drift stays below the FPR-constrained threshold.

Statistical Significance & Noise Robustness

Wilcoxon signed-rank tests

- Top-3 semi-supervised variants ($\alpha = 0.05$).
- H_0 not rejected in any comparison, differences within random variability.
- Contrastive detects both segments most consistently: 1.33 vs 0.83 (UMAP) and 1.17 (VAE).

Additive Gaussian noise (SNR 20 / 15 / 10 dB)

- STOMP Contrastive - delay unchanged at every SNR, both segments always detected.
- UMAP embedding is most noise-sensitive.

Rank	Method	Mean |Δ%| across metrics & SNR levels
1	**STOMP | Contrastive**	**17.25 %**
2	STOMP | VAE	18.64 %
3	STOMP | UMAP	45.43 %

Conclusions

1. Matrix Profile + representation learning matches supervised quality without labels.

2. Deep-learning baselines fail on slow drift in the semi-supervised setting: 5 of 8 detect nothing. Native multivariate MP extensions give no advantage over 1-D projection.

3. Nonlinear and contrastive representations beat the linear projection.

4. Contrastive STOMP is also the most noise-robust variant

Thank you for your attention!

Detection of Slowly Developing Anomalies in Engineering Systems
Based on Matrix Profile Family Algorithms

Code & experimental data: github.com/KalitaAS/Matrix-Profile