mTADS: Multivariate Time Series Anomaly Detection Benchmark Suites

NorwAI Norwegian Research Center for AI Innovation

David Baumgartner, Vilhjalmur Arnar Vilhjalmsson, Kenth Engø-Monsen, Helge Langseth, Heri Ramampiaro

NTNU

Detecting anomalous events in time series data, ranging from manufacturing processes to health care monitoring, is important. Thus, the problem of uncertainty in real-world datasets and the anomalies contained in it, make it difficult to validate and compare results between algorithms. Finding anomalies requires, therefore, an understanding of the data. This can be defined as *'The problem of finding patterns in data that do not conform to expected or normal behaviour . . .* [1]'.

W present two benchmark suits to fill gaps in today's landscape of datasets for anomaly detection in multivariate time series data. One suite focuses only on fully synthetic sequences for testing algorithms with complete knowledge of the sequences. The second suite bridges between fully synthetic and real-world sequences. It provides a few extensive sequences with close-to-reality complexity.

Benchmark Suites and visualizer: `https://github.com/2er0/mTADS`

Fully synthetic benchmark suite (FSB)

The FSB suite contains 70 sequences covering many scenarios and anomalies. Some sequences are generated via our tooling but match the output structure defined by GutenTAG, which we used to generate 60% of the FSB dataset.

Table 1: All anomaly types mapped to their occurrence in all sequence types and to which detection type the anomaly belongs.

Anomaly type	Sine (16)	ECG (6)	CBF (6)	RW (8)	Increasing (13)	Saw (2)	Wave (14)	Corr (5)	Detection type
Platform	•			•					range
Mean	•		•	•					range
Frequency	•	•							range
Pattern	•	•							range
Pattern-shift	•								range
Amplitude	•	•							range
Extremum	•		•	•					point
Variance	•			•					range
Trend	•		•	•					range
Signal-cancelation					•				point range
Signal-reset						•	•		point range
Signal-cut					•				point
Signal-cut-match							•		point
Disconnect								•	range

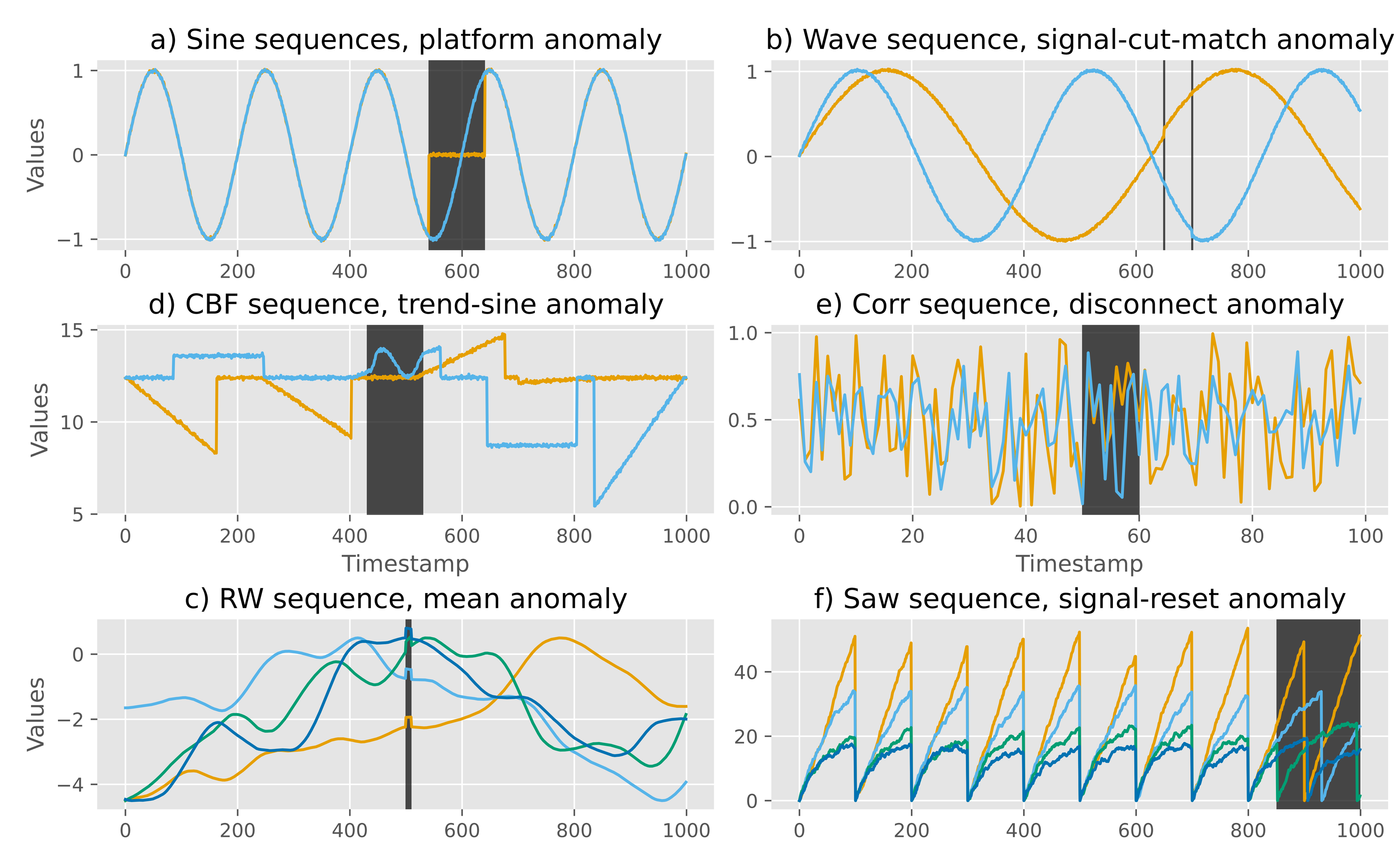


Figure 1: Six sequences from the FSB suite showing a sine, wave, random walk, CBF, ECG, and a saw sequence with various injected anomalies. A shaded area indicates a ranged and a vertical line point anomaly.

Semi-realistic benchmark suite (SRB)

The second benchmark suite aims at bridging the gap between fully synthetic and real-world datasets. A fully synthetic sequence has multiple disadvantages, such as the inability to capture real-world complexity, variability and uncertainty.

This SRB dataset uses an extended Lotka-Volterra equation to simulate N predators and M prays and injects anomalies dynamically during the simulation.

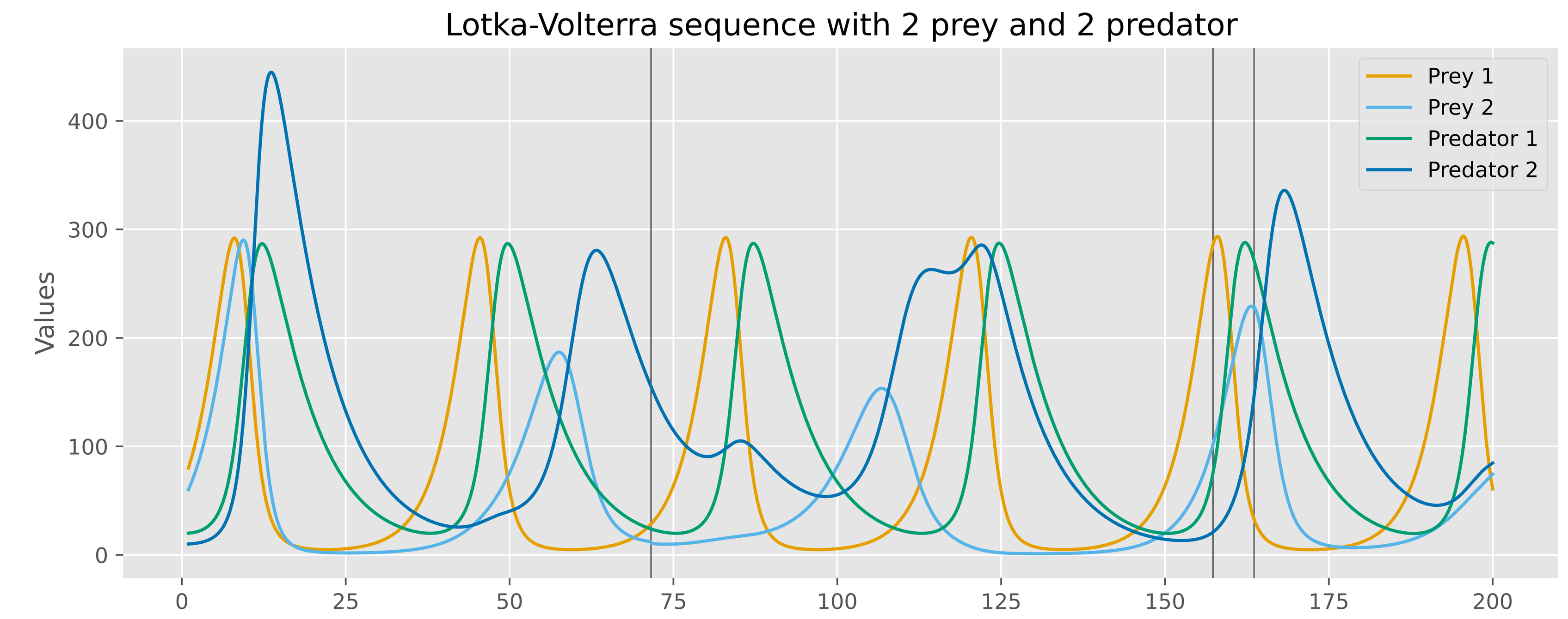
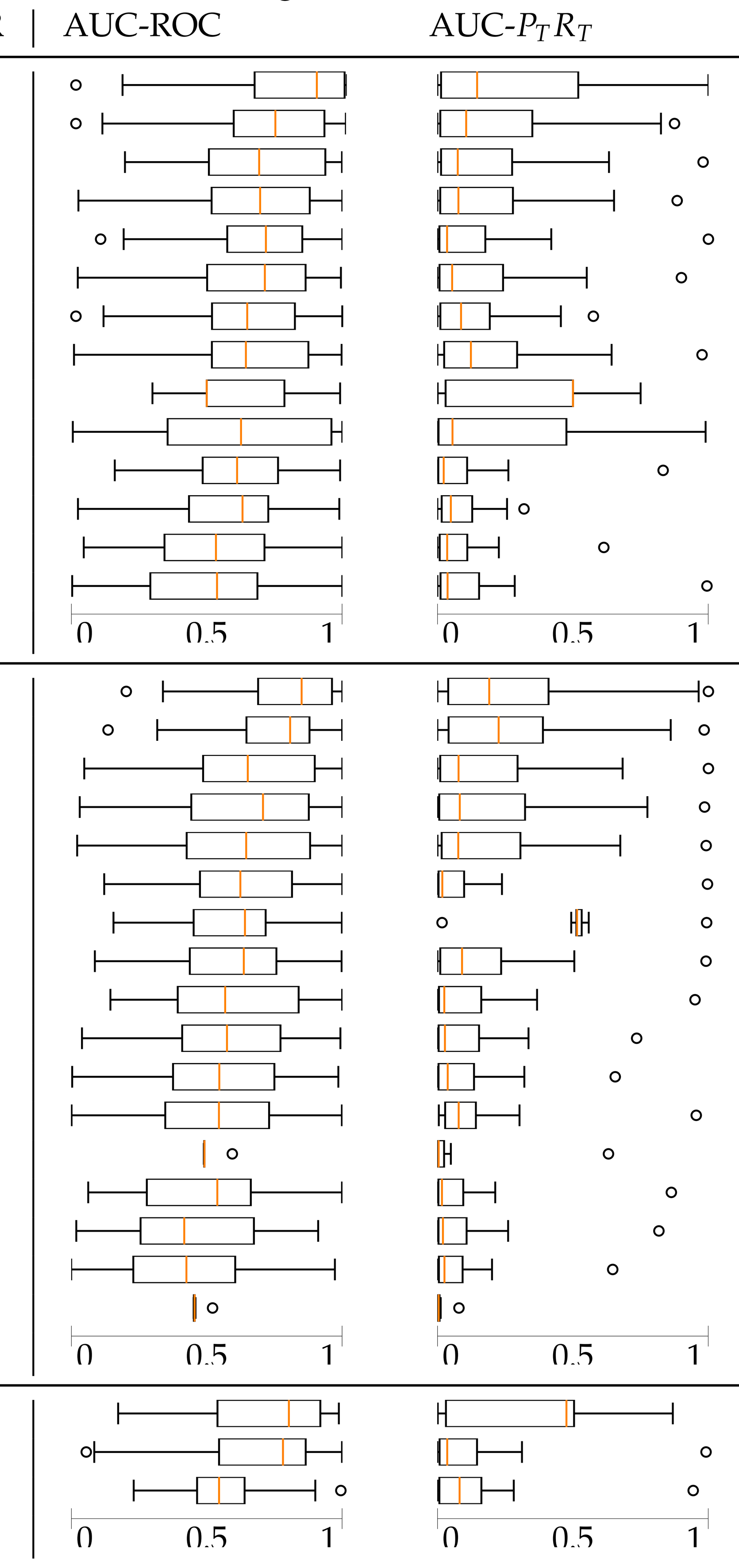


Figure 2: One short example from the SRB suite containing three anomalies and almost some dying-out races with 300.000 steps.

Results

Table 2: Performance for all methods (cf. [3]) over all sequences of the FSB suite. Only the most extreme outliers are shown. Sorting based on mean AUC-ROC value.

Learn.	Method	ERR
UNSUPERVISED	k-Means	0%
	COF	0%
	IF-LOF	0%
	CBLOF	0%
	MSTAMP	0%
	GDN	0%
	iForest	0%
	KNN	0%
	Torsk	61%
	PCA	0%
	DAMP	0%
	HBOS	0%
	COPOD	0%
	PCC	0%
SEMI-SUPERVISED	RBForest	19%
	Health-ESN	0%
	DeepAnT	6%
	AE	0%
	DAE	0%
	Telemanom	6%
	LaserDBN	11%
	Fast-MCD	0%
	LSTM-AD	7%
	DeepNAP	6%
	RobustPCA	0%
	USAD	7%
	TAnoGan	6%
	EncDec-AD	6%
	HybridKNN	6%
	OmniAnomaly	6%
	MSCRED	81%
SUPER.	NF	0%
	HIF	6%
	MultiHMM	44%



Conclusion

The two presented benchmark sets highlight that none of the algorithms performs superior, even in this highly controlled environment over both benchmark suits. Our discoveries align with the findings in [2]. Next, the results show that neither of the metrics is favoured with this publication and that we provide a bias-free environment for algorithms and metrics.

References

[1] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):15:1–15:58, 2009.

[2] V. Jacob, F. Song, A. Stiegler, B. Rad, Y. Diao, and N. Tatbul. Exathlon: a benchmark for explainable anomaly detection over time series. *Proc. of the VLDB Endowment*, 14(11):2613–2626, 2021.

[3] S. Schmidl, P. Wenig, and T. Papenbrock. Anomaly detection in time series: a comprehensive evaluation. *Proc. of the VLDB Endowment*, 15(9):1779–1797, 2022.