Benchmarking Anomaly Detection in Streaming Time Series: Current Landscape and Where to Go Next

Magali Parrino[†,1,2], **Antoine Ajenjo**[§,1], **Emmanuel Remy**[§,1], **Pierre Stephan**[§,1], **Pierre Senellart**[§,2], **Paul Boniol**[§,2]

[†]PhD student (presenting author), [§]PhD supervisor

PhD expected duration: July 2025 – July 2028

[1]EDF R&D, Chatou, France

`{first.last}@edf.fr`

[2]DI ENS, ENS, PSL University, CNRS, Inria, Paris, France

`{first.last}@inria.fr`

Abstract

Streaming Time Series Anomaly Detection (Streaming TSAD) is essential for monitoring safety-critical and industrial systems where measurements arrive sequentially and data distributions may evolve over time. The Streaming TSAD community has focused on developing fast, light and machine learning adaptive architectures [1, 2] while not considering the works proposed in the general TSAD literature [3], designed for offline anomaly detection and deemed too computationally heavy [2].

In this work [4], we challenge this premise through **StrAD**, the first large-scale empirical benchmark comparing 19 static TSAD models and 10 state-of-the-art streaming methods across 17 real-world multivariate datasets. To systematically assess adaptation under data non-stationarity, we introduce **TSB-*drift***, a curated subset of TSB-AD [3], identifying real-world time series exhibiting diverse drift regime. Our empirical findings reveal a striking result: static models evaluated sequentially without parameter updates (*Online TSAD*) significantly outperform native streaming methods, even under substantial concept drift, although the gap narrows. This performance gap stems from an architectural bias: existing streaming algorithms historically focus on point-wise outliers and fail to model complex, sequence-level collective anomalies that high-capacity offline models capture effectively.

While fixed online models currently provide better sequence modeling, relying on a single frozen architecture remains brittle under long-term non-stationarity. To conclude, we outline our ongoing research perspective: shifting away from the pursuit of a single universal streaming model toward dynamic unsupervised model selection across a pool of online models.

References

[1] Y. Cao, Y. Ma, Y. Zhu, and K. M. Ting. Revisiting streaming anomaly detection: benchmark and evaluation. *Artificial Intelligence Review*, 58(1):8, 2025.

[2] F. Iglesias Vázquez, A. Hartl, T. Zseby, and A. Zimek. Anomaly detection in streaming data: A comparison and evaluation study. *Expert Systems with Applications*, 233:120994, 2023.

[3] Q. Liu and J. Paparrizos. The Elephant in the Room: Towards A Reliable Time-Series Anomaly Detection Benchmark. In *Advances in Neural Information Processing Systems 38 (NeurIPS 2024)*, 2024.

[4] M. Parrino, A. Ajenjo, E. Remy, P. Stephan, P. Senellart, and P. Boniol. In a Streaming World, Should You Stand Still? A Comprehensive Benchmark of Anomaly Detection in Streams. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '26)*, 2026.