ANOMALY DETECTION IN SPECTRAL DATA STREAMS: THE DEGASSING CYCLE SCENARIOS

Anonymous authors
Paper under double-blind review

ABSTRACT

Monitoring the elemental composition of materials is of great importance in many scientific and industrial applications. Optical Emission Spectroscopy (OES) is a technique that enables this through the analysis of electromagnetic radiation emitted by a material at different wavelengths, producing spectral data streams. Unlike previous studies on unsupervised anomaly detection in spectral data streams, which mainly focused on air leaks under constant pressure conditions in a vacuum chamber, we collected simulated real-world data within the BV5 Benchmark. This benchmark involves degassing cycles with argon leaks, air leaks, or both. The main objective in BV5 is to distinguish abnormal from normal degassing cycles in real time. However, following previous benchmark settings, we also collected simulated real-world datasets for argon leaks under constant pressure conditions and for temporal anomaly scenarios involving either argon or air leaks. We then provide a progressive evaluation of online anomaly detectors, using only spectral data, across benchmarks. Our results show that OBKNN (TNone), IFASD, OIF, and KitNet consistently outperform competing approaches in spectral anomaly detection. Furthermore, KitNet and OBKNN (TZNorm) exhibit a progressive decline in performance compared to the original benchmark. Lastly, a comparative analysis of methods in degassing cycle scenarios highlights that spectral data can be combined with additional features to enhance anomaly detection performance. To this end, we utilized the collected spectral data and additional features in BV5 to evaluate state-of-the-art online anomaly detection methods. We introduce the Subset Framework (SF), which complements spectral data with an additional feature. We show that incorporating data related to vacuum conditions using SF, such as the electrical current from the OES plasma, improves the detection performance of many online methods. Experiments demonstrate that SF improves the effectiveness of the majority of online detectors, notably SFOBKNN (TNone), SFOBKNN (TZNorm), and SFIFASD, with increases of 0.384, 0.323, and 0.217 in VUS-PR, respectively.

1 INTRODUCTION

Real-time analysis of elemental material composition is critical in many domains, including vacuum gas monitoring, food safety, and environmental applications. Notably, in vacuum gas monitoring, data can be obtained from Optical Emission Spectroscopy (OES), which enables the analysis of the electromagnetic radiation (mainly ultraviolet, visible, and near-infrared) emitted by a material after the de-excitation of its particles. This process produces a set of ordered wavelengths with their corresponding intensities, known as spectral data. Consequently, an OES-based sensor can help characterize baseline operating conditions through these spectral measurements. Anomaly detection (AD) can be performed on this data to identify deviations from normal spectral signatures or their expected temporal behavior. Such deviations are often associated with abnormal operating conditions, such as system faults and contamination events.

Traditionally, an *outlier* is defined as an observation that deviates sufficiently from other observations to arouse suspicion that it was generated by a different mechanism (Hawkins, 1980). Furthermore, terms such as *anomaly*, *novelty*, and *outlier* are often used interchangeably. However, their precise meanings depend on the context. Depending on the application, such as in physical system measurements, outliers resulting from noisy measurement data are not necessarily anomalies. This noise from

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

measurements is considered a normal result of present conditions in the system, as is the case with OES for vacuum gas analysis.

Spectral data from OES-based sensors are generated continuously; therefore, they can naturally be treated as data streams, where a potentially infinite flow of data instances arrives at a high rate. Anomaly detection in this context is typically unsupervised (without *a priori* labels), as obtaining labels can be costly or infeasible in a real-time environment.

Moreover, in the related context of unsupervised anomaly detection for multivariate data, where multiple features are evaluated, anomalies can be categorized as spatial or temporal (Ahmad et al., 2017). A spatial anomaly occurs when an individual data instance is considered abnormal with respect to a local window, independent of its position in time. In contrast, an anomaly is considered temporal when normal behavior is defined by the sequence of data over time. In this case, an instance is anomalous only within a specific temporal context and may appear normal when observed in isolation.

While numerous anomaly detection methods have been proposed for multivariate data, spectral data streams remain relatively underexplored. Although benchmarks exist for anomaly detection in spectral data streams, these benchmarks are limited to constant operating vacuum conditions, such as the scenarios in Varela et al. (2026) and the BV4 benchmark (see Appendix A), where the pressure in a single or double vacuum chamber remains unchanged under normal conditions. Furthermore, the leaks used are specific to air, and no other types of compounds are tested.

To address these limitations, we introduce the BV5 benchmark, which allows the evaluation of state-of-the-art methods, considering different types of anomalies for unsupervised anomaly detection in spectral data streams. The main focus of the BV5 datasets, among other aspects, is on degassing cycles where pressure conditions vary throughout the scenario, as well as on the use of different leak compounds rather than just air. Furthermore, the benchmark provides not only spectral data but also complementary data regarding vacuum conditions.

Accordingly, the contributions of this work include the collection and subsequent progressive evaluation of datasets for unsupervised anomaly detection in spectral data streams within the BV5 benchmark and previous benchmarks. Furthermore, we focus on degassing cycle scenarios in the BV5 benchmark, since under these conditions, spectral data requires further consideration given the instrumental noise caused by changes in measurement conditions. Therefore, we highlight the advantages of incorporating additional information beyond spectral data (notably with the Subset Framework) to enhance the performance of online anomaly detectors by avoiding the use of noisy instrumental spectral data within the models.

2 Related Work

Unsupervised anomaly detection in spectral data streams is a relatively underexplored task. The use of spectral data combined with additional features in a streaming setting is an even less explored area. Few methods have been proposed for the spectral data stream case, including Online Bootstrapping K-Nearest Neighbor (OBKNN) (Varela et al., 2026), a technique based on K-Nearest Neighbors (KNN), bootstrapping, and the z-test to assess the abnormality of incoming spectral instances.

For multivariate anomaly detection, several adaptations of the K-Nearest Neighbors (KNN) (Ramaswamy et al., 2000) algorithm for data streams have been proposed. Micro-Cluster Outlier Detection (MCOD) (Kontaki et al., 2011) optimizes the identification process by defining micro-clusters within sliding windows, where instances failing to be associated with any cluster are flagged as abnormal. This structure allows the model to adapt to distribution drift as the stream evolves. Furthermore, methods such as Lifespan-Aware Probing (LEAP) (Cao et al., 2014) and Stream Outlier Miner (EStorm) (Angiulli & Fassetti, 2007) exploit the temporal order of instances in a window to efficiently distinguish inliers from outliers.

Density-based techniques, such as STAtionary REgion Skipping (STARE) (Yoon et al., 2020), use kernel density estimation to identify the top-n local anomalies. This approach reduces computational costs by skipping updates in regions with stable data distributions. Additionally, the Sparse Gaussian Process Q-Function (SGP-Q) (Gu et al., 2020) models normal behavior through kernel functions and evaluates new observations using a Q-function based on absolute error and prediction likelihood.

Randomized Subspace Hashing (RSHash) (Sathe & Aggarwal, 2016) is an approach for subspace outlier detection that evaluates instance similarity using hash functions.

Subspace partitioning methods, particularly those derived from the decision tree-based Isolation Forest (IF) method (Liu et al., 2008), are extensively used in online settings. Notably, Half-Space Trees (HSTrees) (Tan et al., 2011) use an ensemble of trees to construct mass profiles of the feature space, where instances located in low-mass regions indicate anomalies. IForestASD (IFASD) (Ding & Fei, 2013) incorporates a retraining mechanism triggered by detected distribution shifts. The Robust Random Cut Forest (RRCF) (Guha et al., 2016) uses the Collusive Displacement (CoDisp) score to quantify anomalous behavior. Furthermore, the Online Isolation Forest (OIF) (Leveni et al., 2024) enhances efficiency by dynamically merging or splitting tree regions as data instances enter or exit the sliding window. Other partitioning-based methods include XStream (Manzoor et al., 2018), which uses Half-Space Chains (HSC) to perform density estimation in a grid space of projected values, facilitating detection in evolving streams with variable feature sets.

Deep learning methods have also been adapted for stream processing. KitNet (Mirsky et al., 2018) uses correlative dimension clustering to create an ensemble of autoencoders, where anomalies are detected based on reconstruction error. ARCUS (Yoon et al., 2022) learns from non-overlapping windows and maintains a pool of autoencoders, merging or adding models as new behaviors emerge in the stream.

Finally, comparative studies (Ntroumpogiannis et al., 2023; Cao et al., 2024) have evaluated some of these algorithms across diverse datasets. These analyses demonstrate that online methods such as XStream and MCOD offer significant performance advantages. Additionally, offline methods adapted for online processing with sliding windows, such as SWKNN based on KNN (Ramaswamy et al., 2000), SWLOF based on LOF (Breunig et al., 2000), and IFASD based on IF (Liu et al., 2008), remain competitive. Moreover, previous work has shown that OBKNN, KitNet, and IFASD are accurate methods specifically for spectral data streams (Varela et al., 2026). However, no particular comparative study exists for the combination of spectral data and multivariate data to detect anomalies in real-time.

3 Problem Statement

3.1 Spectral Data

Spectral data is a set of ordered wavelengths with their respective intensities. This data can be obtained indirectly by analyzing aggregated signals with the Fourier Transform (Camelo et al., 2021). Alternatively, several techniques allow the direct collection of spectral data. Notably, this work focuses on Optical Emission Spectroscopy (OES) in low-density plasma, specifically with Inductively Coupled Plasma (ICP).

The ICP-OES technique typically begins with the generation of a low-density plasma, which then interacts with the sample to be analyzed. This interaction ionizes the sample, causing its particles to become electrically charged. As these particles return to a stable state (de-excitation), they emit electromagnetic radiation, which is subsequently decomposed and analyzed by an optical sensor, mainly in the ultraviolet (UV), visible, and near-infrared (NIR) regions. This enables the real-time processing of electromagnetic radiation from a sample, revealing its material composition. This method requires low sample concentrations while offering high accuracy and precision (Helaluddin et al., 2016) as well as high sensitivity to changes in the signal (Cho et al., 2025). Spectral data can be collected in isolated chamber conditions that include:

- Constant (C): Pressure, temperature, and gas composition remain stable during normal operating conditions.
- Degassing (D): Pressure conditions change over time from atmospheric pressure to vacuum conditions.

3.2 Stream of Spectral Data with Complementary Features

Furthermore, in order to handle spectral data with complementary information in a streaming setting, intensities across different electromagnetic wavelengths and additional features are processed

continuously over time. Formally, let $\mathbb{W} = \{1, \ldots, \mathrm{w}\}$ denote the set of ordered wavelengths, let $\mathbb{D} = \{1, \ldots, \mathrm{d}\}$ denote the set of additional features in a multivariate setting, and let $\mathbb{R}$ denote the domain of real numbers. A stream of spectral data with complementary information can be represented as:

$$\begin{array}{rcl} S : \mathbb{N} \times \mathbb{W} & \to & \mathbb{R} \\ (t, w) & \mapsto & S(t, w) \end{array}$$

and

$$\begin{array}{rcl} X : \mathbb{N} \times \mathbb{D} & \to & \mathbb{R} \\ (t, d) & \mapsto & X(t, d) \end{array}$$

where $S(t, w)$ represents the intensity value at time $t \in \mathbb{N}$ for wavelength $w \in \mathbb{W}$, and $X(t, d)$ represents the value of the additional feature $d \in \mathbb{D}$ at time t. Notice that $x(t) = X(t, 1)$ corresponds to the univariate case where only a single additional feature is monitored alongside the spectral data.

Since using multiple features combined with spectral data for real-time processing can be computationally prohibitive, we propose to mainly focus on the univariate case, where two perspectives can be considered to handle this data stream:

- A spectral data stream with an additional feature: $\{(S(1, \cdot), x(1)), \ldots, (S(t, \cdot), x(t))\}_{t \in \mathbb{N}}$.
- A collection of $\mathrm{w} + 1$ univariate data streams: $\{S(\cdot, 1), \ldots, S(\cdot, \mathrm{w}), x(\cdot)\}$.

Following previous work, this study focuses on performing anomaly detection in a *spectral data stream*, therefore using only $\{S(1, \cdot), \ldots, S(t, \cdot)\}_{t \in \mathbb{N}}$, but also in a *spectral data stream with an additional feature*, $\{(S(1, \cdot), x(1)), \ldots, (S(t, \cdot), x(t))\}_{t \in \mathbb{N}}$, for the case of degassing cycle scenarios. It has been shown that handling data as a *spectral data stream* offers benefits in both effectiveness and interpretability (Varela et al., 2026). The approach of using a *spectral data stream with an additional feature*, as will be shown, allows us to further enhance performance for degassing cycle scenarios without affecting the direct interpretability derived from the analysis of the raw spectral instances.

3.3 Integration of an Additional Feature in Spectral Data Streams: *Subset Framework (SF)*

Since additional information complementary to spectral data can provide insights regarding its measurement conditions in a system, a framework called the Subset Framework (SF) is proposed. This framework evaluates the value of an additional feature to determine if a spectral instance should be processed during the data stream.

Specifically, this preprocessing step makes use of an additional feature, denoted by $x(t)$. This feature is utilized to evaluate whether to consider a spectral instance according to a specific reference range, typically defined by domain experts, represented by the interval $[l_c, u_c]$.

Therefore, the spectra $S(t, \cdot)$ are grouped into a subset $\mathcal{L}_0$, which is used for the different anomaly detection methods:

$$\mathcal{L}_0 = \{S(t, \cdot) : x(t) \in [l_c, u_c]\}\,. \tag{1}$$

Spectra associated with feature values outside this range are not considered for modeling or updating models, and a default score of 0 is assigned to these instances in the stream.

4 Experimental Setup

This section presents the evaluation methodology for assessing online anomaly detection methods applied to *spectral data streams* across the various benchmarks, as well as to *spectral data streams with an additional feature* using SF for degassing cycle scenarios in the BV5 benchmark. All source code used in the experiments is available in our anonymous repository.

4.1 ALGORITHMS

The comparative analysis includes the Online Bootstrapping K-Nearest Neighbors (OBKNN) framework (Varela et al., 2026), incorporating minor modifications to update the statistics in the z-score test (see Appendix C). We evaluate two configurations of OBKNN: one with z-normalization (TZNorm) and one without spectral pretreatment (TNone), under different hyperparameter settings.

We include subspace partitioning approaches such as Robust Random Cut Forest (RRCF) (Guha et al., 2016), IForestASD (IFASD) (Ding & Fei, 2013), Online Isolation Forest (OIF) (Leveni et al., 2024), HSTrees (Tan et al., 2011), and XStream (Manzoor et al., 2018). We also evaluate distance-based methods such as EStorm (Angiulli & Fassetti, 2007), density-based methods such as SWLOF (Breunig et al., 2000) and RSHash (Sathe & Aggarwal, 2018), and the deep learning approach KitNet (Mirsky et al., 2018).

We retrieved the implementations from various Python sources. We obtained HSTrees and OIF from the CapyMoa library (Gomes et al., 2025), and used the repository provided by Hartl for SWLOF, while OBKNN was obtained from its original source code. Finally, all other methods were sourced from the PySAD library (Yilmaz & Kozat, 2020).

4.2 DATASETS

We used the original spectral datasets from Varela et al. (2026), which correspond to air leaks as anomalies in a single vacuum chamber operating under constant conditions (9 spectral datasets). We also included additional spectral datasets from the BV4 benchmark (see Appendix A), which mainly involves a double-connected chamber and different types of anomalies, such as temporal and spatial anomalies, under constant pressure conditions across scenarios, as well as scenarios with sudden drift (9 spectral datasets in total). Finally, we extended both the original and BV4 benchmarks with the BV5 benchmark[1], which also involves a double-connected chamber as in BV4 but incorporates degassing cycle scenarios and argon leak scenarios. In addition, BV5 includes additional features describing the state of the vacuum system for each scenario (12 spectral datasets and their associated additional features).

4.2.1 DATASETS FOR HYPERPARAMETER TUNING

To perform hyperparameter tuning for all experiments with spectral or multivariate data streams, one spectral dataset (DA3) from the BV4 benchmark and one multivariate dataset (116_TAO) (Bennett et al., 2000) from the TSB-AD-M benchmark for curated time series (Liu & Paparrizos, 2024) were used. All details on the grid-based tuning process can be found in Appendix D.

4.2.2 BV5 BENCHMARK WITH EXTRA FEATURES

The BV5 benchmark aims to extend the previous collection of datasets with new simulated real-world scenarios. Therefore, with the help of domain experts, a new experimental setup was designed. The setup in BV5 consists of two vacuum chambers. Chamber 1 acts as the "gas monitoring chamber", while a second chamber, Chamber 2, acts as the "fabrication chamber". The setup supports different leak sources, including argon and air leaks, and allows the configuration of different initial humidity conditions during degassing cycles. More details on the vacuum setup can be found in Appendix B.

The BV5 collection was divided into four groups with a total of twelve datasets (see Table 1), recording the spectra (with 2048 wavelengths) and 54 additional features (excluding the Ground Truth and Timestamp columns). All scenarios contain 4200 instances (collected at a rate of one spectrum per second).

The first group of scenarios (SAR1, SAR2, and SAR3) focuses on spatial anomalies caused by argon leaks introduced at the midpoint of each experiment. The leak durations are 60, 300, and 600 seconds, respectively.

In the second group, temporal anomalies are represented by the TNH (air) and TAR (argon) scenarios. In these scenarios, leaks occur every 300 seconds with a duration of 60 seconds. At the midpoint of

[1]Four datasets from BV5 are provided; the others will be released after paper review.

the experiment, this pattern changes to an anomalous behavior in which leaks occur every 60 seconds with a duration of 30 seconds.

The third group corresponds to fixed abnormal degassing cycle scenarios and includes HNH, LNH, HAR, and LAR. These simulations involve six normal cycles from atmospheric pressure to vacuum, followed by one abnormal cycle with a leak introduced at the midpoint of the degassing cycle, after which the remaining cycles are normal. HNH and LNH use air leaks under high and low initial humidity conditions, respectively. HAR and LAR are equivalent but with argon leaks. For these four scenarios, the connection between the monitoring chamber and the fabrication chamber lasts 30 seconds once the fabrication chamber pressure drops below 0.1 mbar during each cycle.

The last group corresponds to random abnormal degassing cycle scenarios and includes RNH, RAR, and RMX. RNH includes six abnormal cycles, while RAR and RMX include five abnormal cycles each at different points in time. These scenarios use air leaks in RNH, argon leaks in RAR, and either argon or air leaks in RMX for the different abnormal cycles. In these scenarios, the chambers are connected for 60 seconds once the fabrication chamber pressure drops below 0.1 mbar during each cycle.

Table 1: Collected datasets in BV5. All scenarios contain 4200 instances in total.

Scenario	Anomaly Characteristics			Condition (Constant or Degassing)	Leak Characteristics					
	Total Dur. (sec.)	# Anom.	Type		Period (sec.)		Duration (sec.)		Leak Type	% Anom.
					Norm.	Abnorm.	Norm.	Abnorm.		
SAR1	60	1	Spatial	C	–	–	–	60	Argon	1.43%
SAR2	300	1	Spatial	C	–	–	–	300	Argon	7.14%
SAR3	600	1	Spatial	C	–	–	–	600	Argon	14.29%
TNH	449	1	Temporal	C	300	60	60	30	Air	10.69%
TAR	449	1	Temporal	C	300	60	60	30	Argon	10.69%
HNH	30	1	Spatial	D	–	–	–	30	Air	0.71%
HAR	30	1	Spatial	D	–	–	–	30	Argon	0.71%
LNH	30	1	Spatial	D	–	–	–	30	Air	0.71%
LAR	30	1	Spatial	D	–	–	–	30	Argon	0.71%
RNH	360	6	Spatial	D	–	–	–	360	Air	8.57%
RAR	300	5	Spatial	D	–	–	–	300	Argon	7.14%
RMX	300	5	Spatial	D	–	–	–	300	Air or Argon	7.14%

4.3 METRICS

We summarized the anomaly scores using the Volume Under the Surface Precision-Recall (VUS-PR) metric. VUS-PR is threshold-free, robust, and well-suited for highly unbalanced datasets (Paparrizos et al., 2022). We set the sliding window size of VUS-PR to the median anomaly length minus one. To assess statistical significance, we plotted a Critical Difference Diagram (CDD) ($\alpha = 0.05$) using the implementation in Herbold (2020). The CDD displays the average ranks across datasets and connects statistically equivalent methods with black lines.

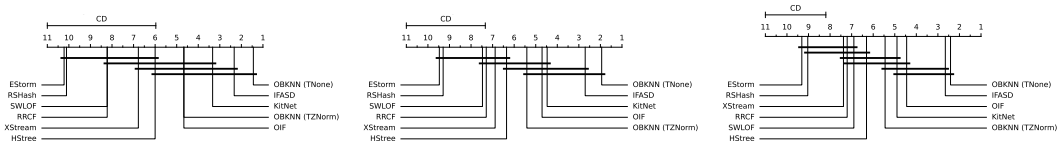
4.4 EXPERIMENTAL EVALUATION

We used the hyperparameters obtained from the grid search tuning for the spectral and multivariate datasets, as mentioned in Section 4.2.1, to perform a progressive overall evaluation that included datasets from the original benchmark (Varela et al., 2026) and the BV4 and BV5 benchmarks. For the Subset Framework (SF) with the BV5 benchmark, we used a separate set of default hyperparameters (Table 7 in Appendix E), since SF introduces new experimental conditions with a different processing setting. Additionally, we used the same tuned hyperparameters mentioned in Section 4.2.1 to contrast the results in the multivariate setting from BV5. Each method was executed three times for each spectral dataset, including the experiments with SF and multivariate data. Finally, we summarized the anomaly scores obtained from the sequential evaluation of the methods on each dataset using the VUS-PR metric. To ensure a fair comparison among the methods, we excluded the scores generated during an initial training window.

5 Overall Results and Discussion

As shown in the critical difference diagrams based on the mean VUS-PR (Figure 1), we observe the evolution of the different approaches when evaluated cumulatively across datasets: from the original spectral benchmark (Figure 1a), to the original benchmark and BV4 benchmark[2] (Figure 1b), and finally when consolidating the evaluation of all datasets across the three benchmarks (Figure 1c). The OBKNN technique without data transformation, namely OBKNN (TNone), and without updating the z-test statistical mean and standard deviation for abnormal cases (as its best hyperparameters show in Table 6 from Appendix D), performs best across the CDDs. Additionally, IFASD, OIF, and KitNet are close in mean rank to OBKNN (TNone). Therefore, these methods do not present statistically significant differences in performance, as indicated by their grouping within the same clique in these diagrams.

Furthermore, OBKNN (TZNorm), which uses z-normalization and updates the mean and standard deviation with anomalous instances (as its best hyperparameters show in Table 6 from Appendix D), performed similarly to the OBKNN (TNone) configuration in the original benchmark (Figure 1a). However, its performance progressively decreased when considering not just the simple leak scenario but also more complex scenarios, such as those in BV4 with drift and temporal anomalies (Figure 1b), and the temporal and degassing cycle scenarios in BV5 (Figure 1c). Similar behavior is observed for KitNet, although in the latter case, the decrease in performance is less pronounced, with KitNet remaining in the best-performing group across all comparisons.



(a) Original Benchmark (9 datasets)

(b) Original + BV4 (17 datasets)

(c) Original + BV4 + BV5 (29 datasets)

Figure 1: Critical difference diagrams for the different benchmarks over three runs using VUS-PR.

5.1 Subset Framework in BV5

For some scenarios, using extra information beyond spectral data can enhance the performance of anomaly detectors. To this end, we applied the Subset Framework (SF), introduced in Section 3.3, in BV5. Therefore, with the help of domain experts, we defined the regulation current (mA) measured from the plasma in OES as this extra feature, setting its range between 7.75 and 8.25, which allows us to discard spectral data outside this range.

As shown in Table 2, under this setting, the number of anomalous instances effectively considered in each BV5 dataset decreases, dropping to zero for non-degassing scenarios, such as SAR1, SAR2, SAR3, TNH, and TAR. Thus, given that measurement conditions are stable for these scenarios and that they contain zero anomalous instances, SF only becomes relevant for the remaining degassing cycle scenarios.

Therefore, for the degassing cycle scenarios in SF, the effective anomaly percentage, measured as the number of anomalous spectral instances relative to all spectral instances in the analyzed range, varies from 7.64% to 11.16% (see Table 2). This remains within the same percentage range as the original datasets, where the maximum percentage was 14.29%, preserving the assumption that anomalies are rare and that there are few in a dataset.

5.2 SF versus Spectral Data Only for Degassing Cycle Scenarios in BV5

We performed a comparative analysis of the different online anomaly detection methods in terms of the mean VUS-PR for degassing cycle scenarios in BV5. The comparison was made between using SF (spectral data with the current as an extra feature) and without SF (using spectral data only).

[2]DA3 from BV4 was excluded from the evaluation since this dataset was used for tuning hyperparameters.

Table 2: Effective data used with SF, with extra feature as current, set between 7.75 and 8.25 mA.

Scenario	Total Dur. Anom. (sec.)	# Anom.	# Instances (sec.)	Condition (Constant or Degassing)	% Anom. SF	% Anom. Total
SAR1	0	0	0	C	0.00%	1.43%
SAR2	0	0	0	C	0.00%	7.14%
SAR3	0	0	0	C	0.00%	14.29%
TNH	0	0	93	C	0.00%	10.69%
TAR	0	0	53	C	0.00%	10.69%
HNH	23	1	271	D	8.49%	0.71%
HAR	23	1	289	D	7.96%	0.71%
LNH	25	1	283	D	8.83%	0.71%
LAR	22	1	288	D	7.64%	0.71%
RNH	319	6	2858	D	11.16%	8.57%
RAR	268	5	2864	D	9.36%	7.14%
RMX	268	5	2859	D	9.37%	7.14%

As shown in Figure 2, almost all methods improve their performance when using SF.[3] In particular, SFOBKNN (TNone) shows an increase of 0.384 in mean VUS-PR, reaching 0.576. It is followed by SFOBKNN (TZNorm) and SFIFASD, which show increases of 0.323 and 0.217, respectively, making them the next best-performing methods. We believe the general increase in performance of many methods with SF occurs because the framework reduces the inclusion of noisy spectral instances that occur during each degassing cycle, when the OES sensor calibrates to generate the spectra after every connection and disconnection of the two vacuum chambers. Consequently, SF prevents the methods from training with these noisy instances at the beginning of the connection of the two chambers, and performance is not negatively impacted. Although OBKNN and IFASD have mechanisms to handle noisy data, such as bootstrapping and subsampling, respectively, the number of noisy spectral instances in the original datasets (without using SF) is considerably high. Therefore, despite these mechanisms, these methods fail on the original datasets, which is not the case for SFOBKNN and SFIFASD, as they better handle the noisy data. On the other hand, methods like KitNet, which do not have such internal mechanisms to avoid training with noisy data, fail to show good performance on the original data as well as when applying SF.

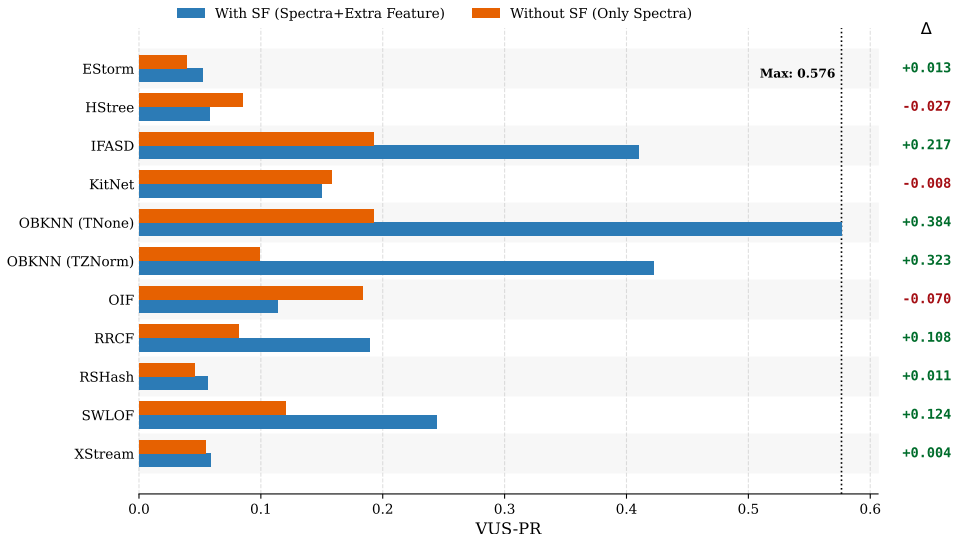


Figure 2: Mean VUS-PR over three runs, with and without SF for degassing cycle scenarios.

[3]Hereafter, approaches using SF are prefixed with "SF" to distinguish them from the original techniques.

5.3 Using Multivariate Data Streams for Degassing Cycle Scenarios in BV5

We also performed a comparative analysis of the online anomaly detectors using extra features associated with plasma generation for OES, which are available in the BV5 datasets, to confirm that the combination of spectral data with extra features (SF) is a better option than using multivariate data directly under degassing cycle scenarios. Therefore, we applied multivariate online detection for these plasma-related features, notably three specific features (voltage, current, and pressure of the monitoring chamber), and we calculated the VUS-PR. We also compared the performance using only the regulation current in OES as a univariate data stream to perform online anomaly detection. Since some methods, such as KitNet and the OBKNN configuration with z-transformation, OBKNN (TZNorm), do not support univariate stream processing, their results are omitted in this case.

As shown in Figure 3, the best VUS-PR result was achieved by OBKNN (TZNorm) with 0.131 using plasma features (voltage, current, and pressure), followed by KitNet using the same plasma features. We observe that, in general, under this setting, no method performed better than SF, where SFOBKNN (TNone) achieved 0.576 as the best performance in Figure 2. Therefore, SF remains a better alternative than using multivariate or univariate data directly. Detailed performance results can be seen in Appendix F, and further experiments for non-degassing cycles can be found in Appendix G.

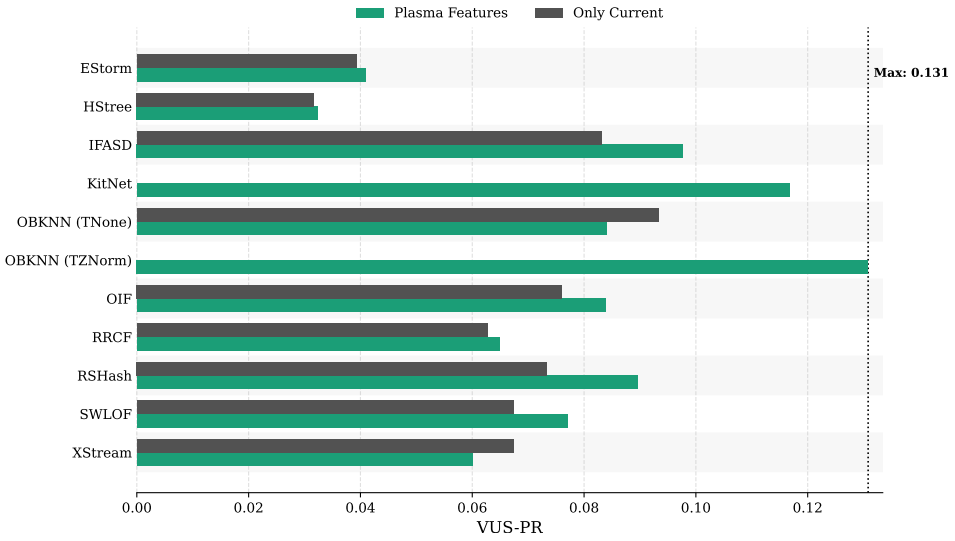


Figure 3: Mean VUS-PR over three runs using multiple features for degassing cycle scenarios.

6 Conclusions

In conclusion, we performed a comparative study of unsupervised online anomaly detection methods, progressively including spectral datasets from the original, BV4, and BV5 benchmarks. We showed that the OBKNN (TNone) configuration, IFASD, OIF, and KitNet methods consistently outperform competing approaches across benchmarks. Furthermore, methods such as KitNet and the OBKNN (TZNorm) configuration exhibit a progressive decline in performance.

We also presented the Subset Framework (SF) as a mechanism to better handle noisy spectral data related to different spectral measurement conditions. We showed that SF improves the performance of many online anomaly detectors, likely by reducing the inclusion of noisy spectral instances during model training. Thus, SF especially improves techniques such as SFOBKNN (TZNorm), SFOBKNN (TNone), and SFIFASD.

Lastly, we demonstrated that using SF is generally more appropriate than using multivariate data directly for degassing cycle scenarios. However, a limitation of the present work is that SF is currently restricted to a single extra feature and is specific to the regulation current of OES. Furthermore, evaluating broader multivariate data in BV5 beyond plasma features could further strengthen the comparison with SF.

AI USE STATEMENT

In this work, we used generative AI tools to assist with writing and data analysis in Python. We have not used generative AI tools to generate new ideas for this paper. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness by one of the authors, and all writing recommendations were verified. We take responsibility for the final content of this work, including the text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

To ensure the reproducibility of our results, we have made the source code and four datasets for the BV5 benchmark publicly available via an anonymous repository (the remaining datasets in BV5 will be released after the paper revision). The datasets for the BV4 benchmark are available on Hugging Face. Datasets from the original spectral benchmark are available in the original public repository. The TAO dataset used for tuning is available on the official TSB-AD website. Additionally, all hyperparameters used for the online anomaly detectors are explicitly detailed in the anonymous repository, as well as in Appendix D and Appendix E.

REFERENCES

Subutai Ahmad, Alexander Lavin, Scott Purdy, and Zuha Agha. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, 262:134–147, 2017.

Fabrizio Angiulli and Fabio Fassetti. Detecting distance-based outliers in streams of data. In *CIKM*, pp. 811–820, 2007.

Andrew F Bennett, Boon S Chua, D Ed Harrison, and Michael J McPhaden. Generalized inversion of tropical atmosphere–ocean (TAO) data and a coupled model of the tropical pacific. part ii: The 1995–96 la nina and 1997–98 el nino. *Journal of Climate*, 13(15):2770–2785, 2000. URL `https://www.pmel.noaa.gov/`.

Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. LOF: identifying density-based local outliers. In *SIGMOD*, pp. 93–104. ACM, 2000.

Miguel Camelo, Paola Soto, and Steven Latre. A general approach for traffic classification in wireless networks using deep learning. *TNSM*, 19(4):5044–5063, 2021.

Lei Cao, Di Yang, Qingyang Wang, Yanwei Yu, Jiayuan Wang, and Elke A Rundensteiner. Scalable distance-based outlier detection over high-volume data streams. In *ICDE*, pp. 76–87. IEEE, 2014.

Yang Cao, Yixiao Ma, Ye Zhu, and Kai Ming Ting. Revisiting streaming anomaly detection: benchmark and evaluation: Y. cao et al. *Artificial Intelligence Review*, 58(1):8, 2024.

Chulhee Cho, Sijun Kim, Isak Lee, Youngseok Lee, Soonyong Cha, Dongho Cha, and Shinjae You. Comparison of leak sensors for detection sensitivity in a vacuum chamber: Self-plasma optical emission spectroscopy and residual gas analyzer. *IEEE Sensors Journal*, 25(17):33765–33778, 2025.

Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *ICML*, pp. 233–240, 2006.

Zhiguo Ding and Minrui Fei. An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window. *IFAC*, 46(20):12–17, 2013.

Andrey A. Efanov, Sergey A. Ivliev, and Alexey G. Shagraev. Welford's algorithm for weighted statistics. In *REEPE*, pp. 1–5, 2021. doi: 10.1109/REEPE51337.2021.9387973.

Heitor Murilo Gomes, Anton Lee, Nuwan Gunasekara, Yibin Sun, Guilherme Weigert Cassales, Justin Liu, Marco Heyden, Vitor Cerqueira, Maroua Bahri, Yun Sing Koh, et al. Capymoa: Efficient machine learning for data streams in python. *arXiv preprint arXiv:2502.07432*, 2025.

Minghao Gu, Jingjing Fei, and Shiliang Sun. Online anomaly detection with sparse gaussian processes. *Neurocomputing*, 403:383–399, 2020.

Sudipto Guha, Nina Mishra, Gourav Roy, and Okke Schrijvers. Robust random cut forest based anomaly detection on streams. In *ICML*, pp. 2712–2721. PMLR, 2016.

Alexander Hartl. dSalmon. URL https://github.com/CN-TU/dSalmon.

Douglas M Hawkins. *Identification of outliers*, volume 11. Springer, 1980.

ABM Helaluddin, Reem Saadi Khalid, Mohamed Alaama, and Syed Atif Abbas. Main analytical techniques used for elemental analysis in various matrices. *TJPR*, 15(2):427–434, 2016.

Steffen Herbold. Autorank: A python package for automated ranking of classifiers. *JOSS*, 5(48):2173, 2020. doi: 10.21105/joss.02173. URL https://doi.org/10.21105/joss.02173.

Maria Kontaki, Anastasios Gounaris, Apostolos N Papadopoulos, Kostas Tsichlas, and Yannis Manolopoulos. Continuous monitoring of distance-based outliers over data streams. In *ICDE*, pp. 135–146. IEEE, 2011.

Filippo Leveni, Guilherme Weigert Cassales, Bernhard Pfahringer, Albert Bifet, and Giacomo Boracchi. Online isolation forest. In *ICML*, volume 235, pp. 27288–27298. JMLR, 21–27 Jul 2024.

Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *ICDM*, pp. 413–422. IEEE, 2008.

Qinghua Liu and John Paparrizos. The elephant in the room: Towards a reliable time-series anomaly detection benchmark. In *NeurIPS*, volume 37, pp. 108231–108261, 2024.

Jie Lu, Anjin Liu, Fan Dong, Feng Gu, Joao Gama, and Guangquan Zhang. Learning under concept drift: A review. *TKDE*, 31(12):2346–2363, 2018.

Emaad Manzoor, Hemank Lamba, and Leman Akoglu. xstream: Outlier detection in feature-evolving data streams. In *SIGKDD*, pp. 1963–1972. ACM, 2018.

Yisroel Mirsky, Tomer Doitshman, Yuval Elovici, and Asaf Shabtai. Kitsune: an ensemble of autoencoders for online network intrusion detection. In *NDSS*. The Internet Society, 2018. doi: 10.14722/ndss.2018.23204.

Antonios Ntroumpogiannis, Michail Giannoulis, Nikolaos Myrtakis, Vassilis Christophides, Eric Simon, and Ioannis Tsamardinos. A meta-level analysis of online anomaly detectors. *VLDB*, 32 (4):845–886, 2023.

John Paparrizos, Paul Boniol, Themis Palpanas, Ruey S Tsay, Aaron J Elmore, and Michael J Franklin. Volume under the surface: A new accuracy evaluation measure for time-series anomaly detection. *VLDB*, 15(11):2774–2787, 2022.

Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim. Efficient algorithms for mining outliers from large data sets. In *SIGMOD*, pp. 427–438. ACM, 2000.

Saket Sathe and Charu C Aggarwal. Subspace outlier detection in linear time with randomized hashing. In *ICDM*, pp. 459–468. IEEE, 2016.

Saket Sathe and Charu C Aggarwal. Subspace histograms for outlier detection in linear time. *KAIS*, 56:691–715, 2018.

Swee Chuan Tan, Kai Ming Ting, and Tony Fei Liu. Fast anomaly detection for streaming data. In *IJCAI*, volume 22, pp. 1511. AAAI Press, 2011.

Nicolas Rojas Varela, Julien Ah-Pine, and Engelbert Mephu Nguifo. OnlineBootKNN: An unsupervised framework for detecting anomalies in spectral data streams. In *AAAI*, volume 40, pp. 15797–15805, 2026.

Selim F Yilmaz and Suleyman S Kozat. Pysad: A streaming anomaly detection framework in python. *arXiv preprint arXiv:2009.02572*, 2020.

Susik Yoon, Jae-Gil Lee, and Byung Suk Lee. Ultrafast local outlier detection from a data stream with stationary region skipping. In *SIGKDD*, pp. 1181–1191. ACM, 2020.

Susik Yoon, Youngjun Lee, Jae-Gil Lee, and Byung Suk Lee. Adaptive model pooling for online deep anomaly detection from a complex evolving data stream. In *SIGKDD*, pp. 2347–2357. ACM, 2022.

A BV4 BENCHMARK: EXPERIMENTAL DATA COLLECTION

The BV4 Benchmark[4] is a collection of nine spectral datasets designed to simulate realistic conditions for spectral anomalies. A schematic overview of the experimental setup is shown in Fig. 4, where Chamber 1 acts as the "gas monitoring chamber" and Chamber 2 as the "fabrication chamber". This configuration enabled the acquisition of spectral data characterizing the elemental composition of gases inside the vacuum system.

As the initial condition for each scenario, various pumps were used to establish the idle condition of the chambers. In the gas monitoring chamber (Chamber 1), a primary pump reduced the pressure from atmospheric level to 10^{-2} mbar, followed by a secondary pump that achieved a pressure of 10^{-6} mbar. In the fabrication chamber (Chamber 2), the primary pump reduced the pressure to 10^{-2} mbar, while the secondary pump reached 10^{-4} mbar. The system includes two valves: one for connecting the two chambers and another for introducing controlled air leaks. Both an optical sensor and a pressure sensor were installed in the gas monitoring chamber.

Nine experiments were conducted, as detailed in Table 3. Most experiments lasted approximately one hour and comprised 4200 instances, with the exception of scenario TA3, which contained 9000 spectral instances. All data were recorded at a sampling rate of 1 spectrum per second. In all cases, leak intensities were fixed at 4.15×10^{-3} sccm (standard cubic centimeters per minute). Each instance consists of a spectrum with 2048 wavelengths, ranging from 189.81 nm to 888.68 nm, and is indexed by its corresponding timestamp.

Spatial anomalies in this setup correspond to spectral instances whose signatures differ from previously observed instances in the data stream. In contrast, temporal anomalies correspond to spectral instances that are similar in structure to previous observations but occur at unusual points in time.

The first three scenarios (SA1, SA2, SA3) focus on spatial anomalies, where air leaks are introduced in the fabrication chamber at the midpoint of the experiments, with durations of 60, 300, and 600 seconds, respectively, while the chambers remain connected throughout the process. The subsequent three scenarios (TA1, TA2, TA3) simulate temporal anomalies by varying leak intervals. Consequently, normal leak behavior occurs every 300 seconds (TA1, TA2) or 600 seconds (TA3), whereas abnormal cases shift to intervals of 60 seconds (TA1, TA2) or 300 seconds (TA3) at the midpoint of the experiment, with varying leak durations. In all cases, the chambers remain connected.

Finally, scenarios DA1, DA2, and DA3 incorporate spatial anomalies alongside sudden concept drift. This sudden drift is defined as an abrupt and unexpected change in the data distribution (Lu et al., 2018) and occurs when the two chambers are connected or disconnected before and after the insertion of a leak at the midpoint of the experiment. This change alters the spectral baseline of the normal condition at both the beginning and the end of the experiment.

After recording the data for each scenario, the ground truth was manually defined based on the start and end points of the introduced abnormal patterns. These include air leaks of varying durations in SA1, SA2, SA3, DA1, DA2, and DA3, as well as shifts in leak timing patterns in TA1, TA2, and TA3. It is important to note that the definition of an abnormal pattern is case-dependent and may vary across different applications.

[4] https://huggingface.co/datasets/a25578073/spectral-benchmark-bv4

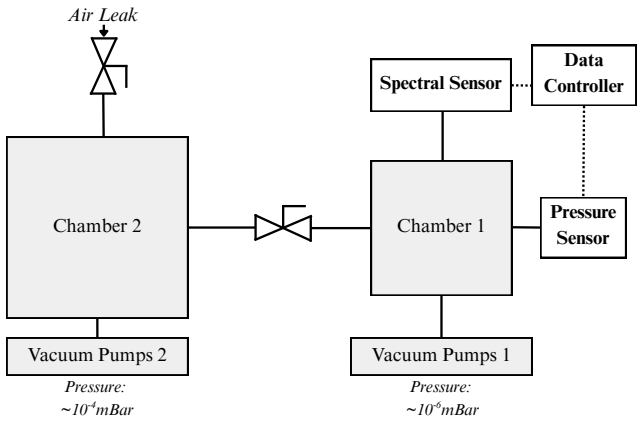


Figure 4: Schema BV4 Benchmark for Spectral Data Collection.

Table 3: Set of Collected Datasets in BV4.

Scenario	Anomaly Characteristics			# Instances (sec.)	Leak Characteristics (sec.)				Concept Drift	% Anom.
	Total Dur. (sec.)	# Anom.	Type		Period		Duration			
					Norm.	Anom.	Norm.	Anom.		
SA1	60	1	Spatial	4200	–	–	–	60	No	1.43%
SA2	300	1	Spatial	4200	–	–	–	300	No	7.14%
SA3	600	1	Spatial	4200	–	–	–	600	No	14.29%
TA1	300	1	Temporal	4200	300	60	60	3	No	7.14%
TA2	300	1	Temporal	4200	300	60	60	12	No	7.14%
TA3	120	2	Temporal	9000	600	300	60	60	No	1.33%
DA1	60	1	Spatial	4200	–	–	–	60	Yes	1.43%
DA2	300	1	Spatial	4200	–	–	–	300	Yes	7.14%
DA3	600	1	Spatial	4200	–	–	–	600	Yes	14.29%

B BV5 Benchmark: Complementary Description

The schema for data collection in BV5 is illustrated in Figure 5. In this schema, the different valves, pumps, and sensors used are shown. Notably, the BV5 setup contained five valves. One dedicated valve connected to Chamber 1 allows modifying pressure conditions if necessary. Another valve allows the two chambers to be connected or isolated, depending on the experimental scenario. The remaining valves allow inserting leaks with different compositions (air or argon) or defining different humidity conditions using degassing materials, notably plastic.

This setup includes optical and pressure sensors, as well as valves and pumps that can be operated automatically from specialized software. This automation was used notably for degassing cycle scenarios, as well as for scenarios with temporal anomalies, ensuring consistent behavior across all scenarios.

Finally, the updated configuration includes an additional pressure sensor in Chamber 2. The minimum vacuum pressure levels are fixed at around 10^{-6} mbar for Chamber 1 and 10^{-4} mbar for Chamber 2 under stable conditions. All data were recorded at a sampling rate of 1 sample per second. Each instance consists of a spectrum (ranging from 189.73 to 895.79 nm) indexed with precise timestamps. In all data collections, the leak intensities are fixed at 4.15×10^{-3} sccm.

After recording the data for each scenario, the ground truth for degassing cycle scenarios was defined using specialized software as a reference, which tracked the opening and closing of the leak valves (variable "ETATREALTIMELIFE" in the datasets). The ground-truth label ("ANOMALY?") was then refined to match the duration of each automation. For non-degassing scenarios, the ground truth was manually defined using the start and end points of the introduced abnormal patterns.

Additional features beyond spectral data are provided to describe the state of the vacuum system for the different collected scenarios. These features include information such as the pressure in the various vacuum chambers over time, the *Region Study* variable corresponding to the maximum value within a predefined wavelength range, and the regulation voltage (V) and regulation current (mA) for the plasma of the OES sensor, among others.

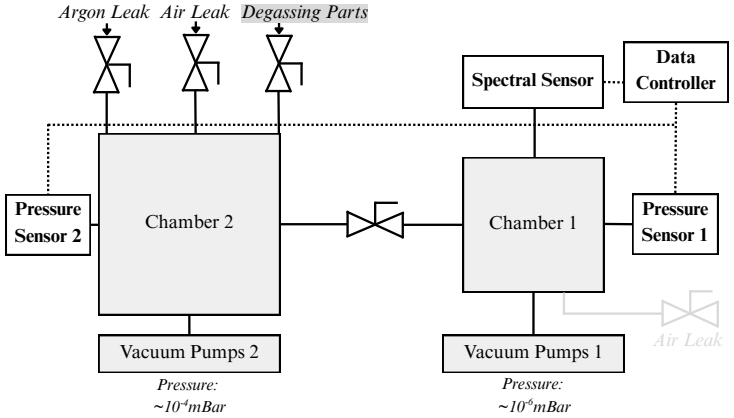


Figure 5: Schema BV5 for data collection.

C Adaptation of OBKNN for Multivariate Data Streams and Other Modifications

OBKNN was originally designed to exploit the joint behavior of wavelengths in spectral data streams. To extend this formulation to general multivariate data with different channels, we allowed omitting the z-normalization step. The `transf` hyperparameter controls this adaptation (see Table 5). This permits defining a variant of OBKNN called OBKNN (TNone), which processes instances without transformation.

Furthermore, the original implementation of OBKNN used Welford's algorithm (Efanov et al., 2021) to incrementally update the historical statistics μ^s_{t+1} and σ^s_{t+1}. We incorporated an alternative mechanism to update these statistics through the `update_mode_stats` hyperparameter (Table 5). We implemented the Exponential Moving Average (EMA) and its associated variance (EMV). This approach applies a decay factor $\alpha_{\text{EMA}} \in [0, 1]$ (the `alpha_ema` hyperparameter in Table 5) to give more weight to recent observations:

$$\mu^s_{t+1} = (1 - \alpha_{\text{EMA}})\mu^s_t + \alpha_{\text{EMA}} a_{t+1} \tag{2}$$

The update for the moving standard deviation follows this recurrence:

$$\sigma^s_{t+1} = \sqrt{(1 - \alpha_{\text{EMA}})\left((\sigma^s_t)^2 + \alpha_{\text{EMA}}(a_{t+1} - \mu^s_t)^2\right)} \tag{3}$$

Finally, the original OBKNN implementation used all minimum distances to update the statistics. However, in stable processes with no changes in the data distribution, it could be beneficial to avoid updating these statistics for anomalous cases. Therefore, we introduced the `update_distance_with_abnormal` hyperparameter (see Table 5). This allows the model to skip updating μ^s_{t+1} and σ^s_{t+1} if the current minimum distance a_{t+1} is associated with an abnormal instance.

D Hyperparameter Tuning for Anomaly Detection in Spectral and Multivariate Data Streams

To optimize hyperparameters for spectral and multivariate data streams, we performed a grid search (the search space is detailed in Table 5). We ran the online anomaly detection algorithms sequentially

in a streaming setting on two selected datasets: DA3 from the BV4 benchmark for spectral data, and 116_TAO (Bennett et al., 2000) from the TSB-AD-M benchmark for curated multivariate time series (Liu & Paparrizos, 2024). The main characteristics of the datasets used for tuning are summarized in Table 4. Specifically, the TAO dataset contains 10000 instances and 3 features, whereas the DA3 dataset contains 4200 instances and 2048 features (wavelengths).

Each hyperparameter combination was evaluated three times for each dataset. We varied the processing window sizes to 2%, 5%, and 20% of the dataset length for the methods. To ensure a fair assessment, we excluded anomaly scores from the initial training phase. The optimized configurations were selected based on the best mean AUC-PR (Davis & Goadrich, 2006) across all runs and datasets, while minimizing the standard deviation between scenarios. This procedure yielded a set of optimized hyperparameters for online anomaly detectors on both spectral and multivariate data streams (Table 6). For the OBKNN method, we kept the best configuration with z-normalized instances, OBKNN (TZNorm), as well as the best configuration without data transformation, OBKNN (TNone).

Table 4: Characteristics of datasets used for tuning hyperparameters.

Dataset	Features	# Abn. Instances	# Total Instances	% Anomaly
116_TAO	3	492	10000	4.92
DA3_20250610	2048	600	4200	14.29

Table 5: Hyperparameter search space, for window sizes (`p_window_size`) at 2%, 5%, and 20% of the dataset size.

Method	Hyperparameter Space
EStorm	`max_radius`: [0.1, 300, 900, 1800, 4000]
HSTrees	`number_of_trees`: [15, 25], `max_depth`: [10, 15], `anomaly_threshold`: [0.3, 0.5], `size_limit`: [0.1]
IForestASD	`initial_window_X`: [None]
KitNet	`max_size_ae`: [10, 20], `learning_rate`: [0.1, 0.2], `hidden_ratio`: [0.75, 0.9]
OBKNN	`ensemble_size`: [30, 240], `dmetric`: ['cityblock'], `transf`: ['NONE', 'ZNORM'], `update_distance_with_abnormal`: [True, False], `update_mode_stats`: ['ema', 'welford'], `alpha_ema`: [0.01], `algorithm`: ['brute'], `chunk_size`: [30], `type_dist`: ['largest'], `alpha_z_test`: [0.05, 0.95], `n_jobs`: [-1], `no_bootstrapp`: [False], `no_z_score`: [False]
OIF	`num_trees`: [32, 64], `max_leaf_samples`: [32, 64], `growth_criterion`: ['adaptive', 'fixed'], `n_jobs`: [-1]
RRCF	`num_trees`: [4, 32], `tree_size`: [256, 512]
RSHash	`feature_mins`: [0], `feature_maxes`: [∞, 2000], `decay`: [0.015, 0.05], `num_components`: [50, 100], `num_hash_fns`: [1, 5]
SWKNN	`k`: [1, 10, 50], `k_is_max`: [False], `metric`: ['cityblock'], `min_node_size`: [5], `max_node_size`: [20], `split_sampling`: [5]
SWLOF	`k`: [1, 10], `k_is_max`: [False], `simplified`: [False], `metric`: ['cityblock', 'euclidean']
XStream	`num_components`: [50, 100], `n_chains`: [50, 100], `depth`: [15, 25]

Table 6: Best hyperparameters according to AUC-PR metric for spectral and multivariate datasets.

Method	Best Hyperparameters Configuration
EStorm	p_window_size: 0.2, max_radius: 900
HSTrees	p_window_size: 0.02, anomaly_threshold: 0.5, max_depth: 10, number_of_trees: 25, size_limit: 0.1
IForestASD	p_window_size: 0.2, initial_window_X: None
KitNet	p_window_size: 0.05, hidden_ratio: 0.9, learning_rate: 0.1, max_size_ae: 10
OBKNN (TNone)	p_window_size: 0.2, algorithm: brute, alpha_ema: 0.01, alpha_z_test: 0.05, chunk_size: 30, dmetric: cityblock, ensemble_size: 30, n_jobs: -1, no_bootstrapp: False, no_z_score: False, transf: NONE, type_dist: largest, update_distance_with_abnormal: False, update_mode_stats: ema
OBKNN (TZNorm)	p_window_size: 0.2, algorithm: brute, alpha_ema: 0.01, alpha_z_test: 0.05, chunk_size: 30, dmetric: cityblock, ensemble_size: 30, n_jobs: -1, no_bootstrapp: False, no_z_score: False, transf: ZNORM, type_dist: largest, update_distance_with_abnormal: True, update_mode_stats: welford
OIF	p_window_size: 0.2, growth_criterion: fixed, max_leaf_samples: 64, n_jobs: -1, num_trees: 64
RobustRandomCutForest	p_window_size: 0.2, num_trees: 4, tree_size: 256
RSHash	p_window_size: 0.2, decay: 0.05, feature_maxes: [∞, 2000], feature_mins: [0], num_components: 50, num_hash_fns: 1
SWLOF	p_window_size: 0.02, k: 10, k_is_max: False, metric: euclidean, simplified: False
XStream	p_window_size: 0.02, depth: 25, n_chains: 100, num_components: 50

E Hyperparameters for Online Anomaly Detectors with Subset Framework (SF)

Default hyperparameters were used in the Subset Framework since this is a different processing setting from spectral or multivariate data streams. Therefore, hyperparameters were set as default in the original multivariate data stream implementation of the method, and new hyperparameters were included, notably the range of values used in SF for the extra feature, min_value and max_value, which in our experiments were set to 7.75 and 8.25 with current as the extra feature (see Table 7).

F Detailed Results: Degassing Cycle Scenarios in BV5

Table 8 shows detailed results of the mean VUS-PR with its standard deviation over the three runs presented in Section 5.2 for degassing cycle scenarios with and without SF, but also using only extra features such as the current, or extra features related to the generation of plasma in OES (current, voltage, and pressure).

Table 7: Default Hyperparameters for Subset Framework (SF) with Current as Extra Feature.

Method	Hyperparameters Configuration
SF_EStorm	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, max_radius: 0.1
SF_HSTrees	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, number_of_trees: 25, max_depth: 15, anomaly_threshold: 0.5, size_limit: 0.1
SF_IForestASD	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, initial_window_X: None
SF_KitNet	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, max_size_ae: 10, learning_rate: 0.1, hidden_ratio: 0.75
SF_OIF	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, num_trees: 32, max_leaf_samples: 32, growth_criterion: adaptive, n_jobs: -1
SF_RobustRandomCutForest	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, num_trees: 4, tree_size: 256
SF_RSHash	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, feature_mins: [0], feature_maxes: [10000], decay: 0.015, num_components: 100, num_hash_fns: 1
SF_SWLOF	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, k: 10, k_is_max: False, simplified: False, metric: euclidean
SF_XStream	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, num_components: 100, n_chains: 100, depth: 25
SF_OnlineBootKNN_TNone	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, alpha_ema: 0.01, ensemble_size: 240, chunk_size: 240, dmetric: cityblock, transf: NONE, update_distance_with_abnormal: False, update_mode_stats: ema, algorithm: brute, type_dist: largest, alpha_z_test: 0.05, n_jobs: -1, no_bootstrapp: False, no_z_score: False
SF_OnlineBootKNN_TZNorm	min_value: 7.75, max_value: 8.25, p_window_size: 0.02, alpha_ema: 0.01, ensemble_size: 240, chunk_size: 240, dmetric: cityblock, transf: ZNORM, update_distance_with_abnormal: True, update_mode_stats: welford, algorithm: brute, type_dist: largest, alpha_z_test: 0.05, n_jobs: -1, no_bootstrapp: False, no_z_score: False

G Other Results: Non-Degassing Cycle Scenarios in BV5

Furthermore, the mean results over three runs with standard deviations for the BV5 non-degassing cycle scenarios are presented in Table 9 and plotted in Figure 6 for only spectra (since SF does not have enough data to process), and in Figure 7 for the application to multivariate data.

We can observe that under non-degassing cycles, the performance is relatively equivalent for only spectra (without SF) versus multivariate data (0.682 against 0.652 VUS-PR). However, the use of spectral or multivariate data should also be evaluated based on the information each type of data provides. On one hand, spectral data allows monitoring the elemental composition of the material; on the other hand, multivariate data allows monitoring the conditions of the vacuum system.

G.1 Used Servers

The experiments were conducted on servers running Debian GNU/Linux 12. For the majority of the experiments, a server equipped with a 16-core / 32-thread Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40

Table 8: Comparison of mean VUS-PR performance (± standard deviation over the 7 degassing cycle datasets) of methods across degassing cycle scenarios. For each column, the best and second-best scores are highlighted in bold and underlined, respectively. The final row summarizes the highest performance achieved in each configuration.

	Subset Framework (SF)	Spectral Data	Multivariate Data	
Method	Spectra + EF	Only Spectra	Only Current	Plasma Features
EStorm	0.052 ± 5.0e-02	0.039 ± 3.5e-02	0.039 ± 3.5e-02	0.041 ± 3.4e-02
HStree	0.058 ± 6.1e-02	0.085 ± 1.0e-01	0.032 ± 2.9e-02	0.032 ± 2.8e-02
IFASD	0.410 ± 3.0e-01	**0.193** ± 1.1e-01	<u>0.083</u> ± 2.2e-02	0.098 ± 2.6e-02
KitNet	0.150 ± 1.5e-01	0.158 ± 1.1e-01	-	<u>0.117</u> ± 2.7e-02
OBKNN (TNone)	**0.576** ± 1.9e-01	<u>0.192</u> ± 9.5e-02	**0.093** ± 2.7e-02	0.084 ± 2.0e-02
OBKNN (TZNorm)	<u>0.422</u> ± 2.0e-01	0.099 ± 3.4e-02	-	**0.131** ± 7.6e-02
OIF	0.114 ± 3.6e-02	0.184 ± 1.2e-01	0.076 ± 2.5e-02	0.084 ± 2.5e-02
RRCF	0.189 ± 9.3e-02	0.081 ± 3.5e-02	0.063 ± 2.4e-02	0.065 ± 2.3e-02
RSHash	0.057 ± 6.0e-02	0.046 ± 4.0e-02	0.073 ± 2.1e-02	0.090 ± 1.9e-02
SWLOF	0.244 ± 5.5e-02	0.120 ± 6.2e-02	0.067 ± 4.7e-02	0.077 ± 3.7e-02
XStream	0.058 ± 6.1e-02	0.055 ± 3.5e-02	0.067 ± 6.6e-02	0.060 ± 5.8e-02
Best Result	0.576 ± 1.9e-01	0.193 ± 1.1e-01	0.093 ± 2.7e-02	0.131 ± 7.6e-02

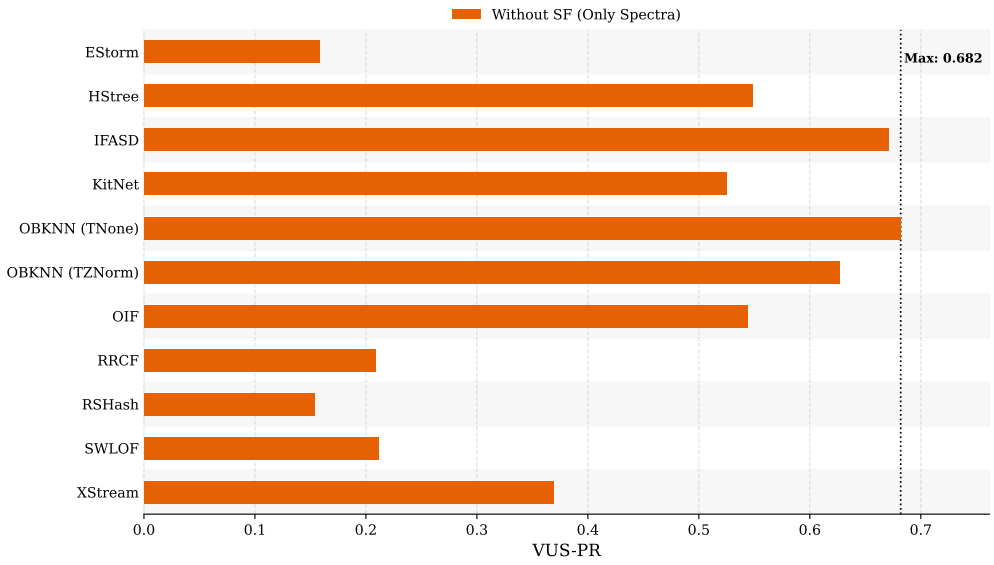


Figure 6: Mean VUS-PR over three runs, without SF for non-degassing cycle scenarios.

GHz was used. The only exception was the SFOBKNN method, for which a server equipped with a 24-core / 48-thread Intel(R) Xeon(R) Gold 5118 CPU @ 2.30 GHz was used.

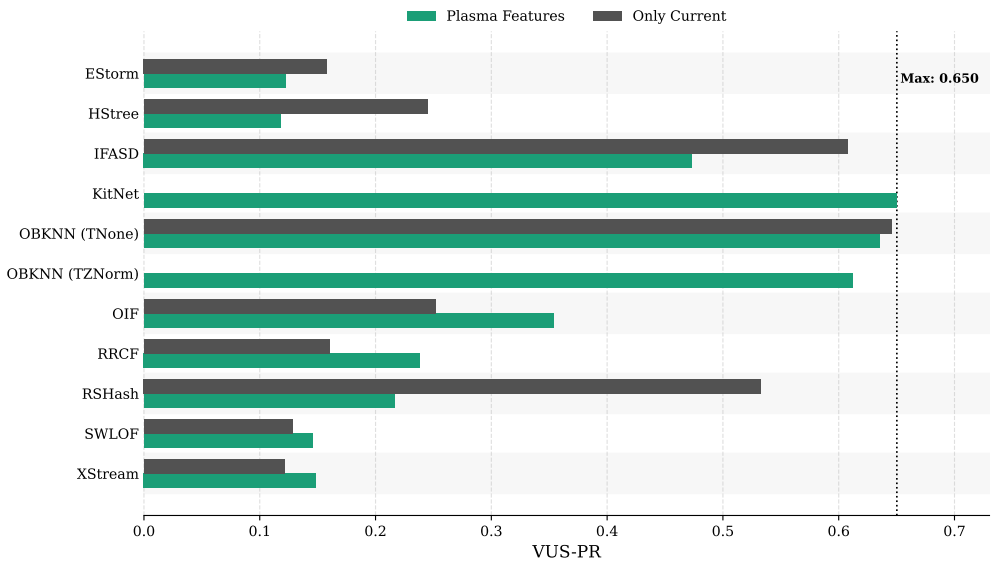


Figure 7: Mean VUS-PR over three runs using different features for non-degassing cycle scenarios.

Table 9: Comparison of mean VUS-PR performance (± standard deviation over the 5 non-degassing cycle datasets) of methods across non-degassing cycle scenarios. For each column, the best and second-best scores are highlighted in bold and underlined, respectively. The final row summarizes the highest performance achieved in each configuration.

	Spectral Data	Multivariate Data	
Method	Only Spectra	Only Current	Plasma Features
EStorm	0.158 ± 8.0e-02	0.158 ± 8.0e-02	0.123 ± 6.5e-02
HStree	0.549 ± 3.4e-01	0.245 ± 2.6e-01	0.118 ± 8.1e-02
IFASD	<u>0.671</u> ± 3.8e-01	<u>0.607</u> ± 3.5e-01	0.473 ± 2.8e-01
KitNet	0.525 ± 3.5e-01	-	**0.650** ± 4.2e-01
OBKNN (TNone)	**0.682** ± 4.0e-01	**0.646** ± 4.5e-01	<u>0.635</u> ± 4.0e-01
OBKNN (TZNorm)	0.627 ± 4.0e-01	-	0.612 ± 3.4e-01
OIF	0.544 ± 3.3e-01	0.252 ± 1.5e-01	0.354 ± 3.3e-01
RRCF	0.209 ± 7.5e-02	0.160 ± 5.1e-02	0.238 ± 5.0e-02
RSHash	0.153 ± 6.8e-02	0.533 ± 1.8e-01	0.217 ± 6.5e-02
SWLOF	0.211 ± 5.1e-02	0.128 ± 6.5e-02	0.145 ± 4.6e-02
XStream	0.369 ± 2.8e-01	0.122 ± 7.9e-02	0.149 ± 1.2e-01
Best Result	0.682 ± 4.0e-01	0.646 ± 4.5e-01	0.650 ± 4.2e-01