

Adaptive unsupervised anomaly detection with dynamic recalibration for scalable industrial asset reliability monitoring

Dhirendra Singh

dhireens@gmail.com

UPES: University of Petroleum and Energy Studies <https://orcid.org/0009-0005-6680-230X>

Durga Prasad Panday

UPES: University of Petroleum and Energy Studies

Manish Kumar

UPES: University of Petroleum and Energy Studies

Research Article

Keywords: Predictive maintenance, Anomaly detection, Unsupervised learning, Dynamic recalibration, Asset reliability, Fault diagnosis, Petrochemical systems

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10515853/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Adaptive unsupervised anomaly detection with dynamic recalibration for scalable industrial asset reliability monitoring

Dhirendra Singh^{1*}, Durga Prasad Panday¹, Manish Kumar^{1,2}

¹Sustainability Cluster, School of Advanced Engineering, UPES, Dehradun-248007, Uttarakhand, India

²Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, Campus Monterrey, Monterrey 64849, Nuevo León, Mexico

*Corresponding author: Dhirendra Singh, Email: Dhireens@gmail.com; ORCID: 0009-0005-6680-230X

Article type: Original Research Article

Highlights

- A lightweight unsupervised approach which can identify anomalies in the pump without failure data labels.
- Dynamic recalibration automatically adjusts thresholds with changes in operating baseline.
- Long horizon applicability has been validated with three years of petrochemical pump historian data.
- All 44 of the functional failures found were in the retrospective validation set.
- Average warning lead time of 72 h allows for decisions to be made based on reliability of maintenance.

Abstract

Anomaly-detection methods typically fail to detect changes in operating conditions or are unresponsive in detecting changes in the aging of industrial assets. This paper outlines an unsupervised lightweight machine learning-based approach to adaptively detect anomalies in the reliability of industrial assets. It incorporates in its framework operating-range normalisation, principal component analysis, Euclidean-distance anomaly scoring, Gaussian-based dynamic thresholds, sensor-contribution ranking and autonomous baseline recalibration. The validation was performed by taking 10 sensor variables initially screened in a critical P-101 A/B pump in an HDPE slurry polymerisation process, sampled approximately every 15 min and nine variables were selected for sensor-quality screening and then used for the validation. The framework identified all 44 reported functional failures resulting in a 100% recall, 84.62% precision, an average warning lead time of 72 h and an F1-score of 91.67%. The

findings show that a self-recalibrating unsupervised model can successfully detect faults in a scalable manner, with an interpretable model, low computational complexity and can adapt to ageing assets and varying operating baselines without labelled fault information.

Keywords: Predictive maintenance; Anomaly detection; Unsupervised learning; Dynamic recalibration; Asset reliability; Fault diagnosis; Petrochemical systems

1. Introduction

Industrial Asset Reliability is vital to the continuity of production and is essential to the safety of the processes and protection of the environment in petrochemical, oil and gas, utilities and advanced manufacturing systems. Failure of a pump, compressor, heat exchanger, pipeline or a safety-instrumented component in a tightly coupled technological system not just be a local equipment failure, it can create abnormal states of the process, unplanned shutdowns, hazardous releases and cascading disruptions of operations. The transformation from reactive maintenance to condition-based or predictive maintenance, with the help of historical data, industrial Internet of Things (IIoT) infrastructure and machine-learning analytics, has gained momentum due to these concerns. Due to these concerns, the shift from the reactive maintenance towards condition-based or predictive maintenance with the support of historical data, industrial Internet of Things (IIoT) infrastructure and machine-learning analytics has accelerated.

Traditional maintenance concepts such as reliability-centred maintenance, failure mode and effects analysis and time based preventive maintenance are still relevant, but is static. These factors such as equipment ageing, maintenance interventions, product change, restart, feedstock variation and environmental/process disturbances cause normal operating behaviour changes in industrial plants. Fixed alarms and manually set thresholds can thus trigger nuisance alarms when the operating state changes or not trigger alarms when the asset slowly moves away from a previous normal operating state. It is therefore necessary to have adaptive, interpretable and robust models to describe reliability properties of data streams under non-stationary conditions in industry [3,4].

Predictive maintenance and fault diagnosis have been widely used, with the aid of machine learning. There have been models based on supervised and deep learning that have had good performance on prediction of pipeline failures, temperature prediction, prognosis of rotating equipment and maintenance tasks in the power grid [5-7]. Supervised models, however, do need failure histories, are typically asset-specific, and sometimes fail to exist. However, for critical assets, failures are exactly the

events the plants try to avoid, making it uncommon to have a labelled failure record that is thorough and consistent. Hybrid and deep architectures are used to achieve higher classification accuracy, but require greater computational resources, deployment times and model development [8-10]. These constraints restrict the scalability in a variety of equipment sets.

Predictive-maintenance research in recent years increasingly focuses on the aspects of adaptive modelling, real-time industrial data streams, integration with digital-twins, handling of concept drift and explainable anomaly detection. In the past few years, reviews of anomaly detection in the IIoT have noted that static models can become inaccurate in the presence of concept drift and that a large number of explainability methods still offer statistical feature attribution instead of root cause diagnosis [18]. Likewise, adaptive predictive-maintenance in IIoT has demonstrated that models that are able to update as equipment and environmental conditions change, can outperform non-adaptive models [19]. It is also important to consider industrial application to balance detection sensitivity, interpretation and computational efficiency in the framework of digital-twin and hybrid deep-learning frameworks [20,21]. These advances reinforce the need for generic but self-recalibrating methods that will help sustain performance without the need for copious amounts of labelled failure data, or complex architectures.

However, for industrial use it is not enough that an algorithm can detect an anomaly, it is also necessary that the detection logic is still useful when the definition of normal operation evolves. For the use in industry, the question of reliability then is not only whether an algorithm can classify an anomaly, but also whether the detection logic is still useful when the definition of normal operation changes. A scalable framework needs to use the data generated during operation to learn the normal behaviour, identify significant deviations, update the normal baseline when a new stable healthy state occurs, and generate information which is interpretable and is used by the operator and the maintenance. It should also do this without having to tweak thresholds for each asset, or relying on labelled examples of each failure mode.

Unsupervised anomaly detection is by no means the novelty of this study. The current unsupervised methods are able to detect deviations from historical normal behaviour, however they are constrained by a number of fixed thresholds, the need for asset-specific tuning, manually selected alarm thresholds and are often computationally intensive architectures. The proposed framework, however, is intended to be a general and scalable framework for monitoring reliability that integrates a low-dimensional statistical model with the ability for dynamically adjusting the model's parameters through autonomy.

The model is trained with a baseline of the normal operating condition, and identifies deviations from the baseline; the new stable healthy operating state will update the centroid and thresholds. This enables the method to be usable even if equipment becomes 'old' or the process changes, maintenance is carried out or the environment and operations fluctuate.

There are three contributing factors. First, the study proves a lightweight unsupervised model can attain fault-detection accuracy that is similar to the results of more advanced predictive-maintenance models, despite the fact that no labelled failure information is needed for training, and computational requirements are lower. Second, the dynamic recalibration mechanism decreases reliance on fixed engineering thresholds, and enables long-term deployment through various asset states. Third, sensor-contribution analysis adds interpretability by assigning anomaly scores to the real-world deviations in the process, which helps investigate the root cause and make maintenance decisions. The features enable the framework to outperform standard unsupervised models for anomaly detection, and make it more appropriate for large scale industrial asset reliability applications.

This work extends and tests the framework on data from the historians of a critical P-101 A/B pump of a high-density polyethylene (HDPE) slurry polymerisation process for three years. It is made up of data cleaning, operating-range normalisation, principal component analysis (PCA), Euclidean-distance anomaly scoring, Gaussian warning and alert thresholds, sensor-contribution ranking and autonomous baseline recalibration. Evaluation of the model is based on the documented functional failures, without the use of failure labels during model learning.

2. Industrial system and data

2.1 Process description and equipment under study

The industrial case is related to an HDPE slurry polymerisation process with a critical P-101 A/B pump. The pump is a part of the heat-removal loop of Reactor R-101 and plays a key role in safe and efficient polymerisation temperature control. In this study, the most important failure mode taken into account is the polymer lumping in slurry. These aggregates can trap the pump impeller, cause reduced flow and if large enough, can cause shutdown of the plant. Production continuity and process safety is therefore crucial in the early detection of abnormal behaviour that is related to impeller obstruction, flow reduction and pressure instability.

The chosen process setting is important for reliability engineering and research on system-safety since the pump is part of a complicated technological process. Pump failure can impact heat removal, operation of the reactor and downstream risk. The goal was not to create a model based on a specific failure signature from the past, but rather to check if a generic unsupervised approach could learn healthy behaviour and still be able to detect any important multivariate failures while still remaining sensitive to changes in operating regime.

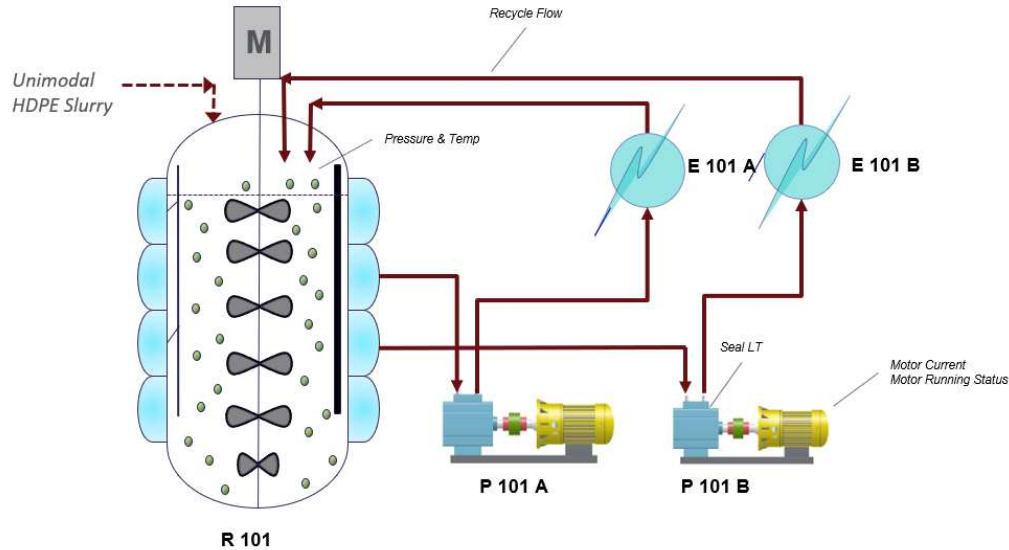


Fig. 1. Schematic representation of the HDPE slurry polymerisation process and the position of the P-101 A/B pump in the heat-exchange loop.

2.2 Historian data and process variables

Industrial historians that are integrated into distributed control and supervisory control systems continuously record variables of the process and enable event reconstruction, model validation and long horizon reliability analysis [11]. Sampling frequency of the selected sensor variables was about 15 min resulting in a total of approximately 35,040 observations for each year of the period of record and over 105,120 observations for each sensor variable during the entire period of record. This multi-year record enabled the framework to be assessed during stable operation, when starting up from a cold start, during maintenance transitions and when approaching failure.

To select the input variables, two alternative methods were followed. Firstly, process and instrumentation documentation was researched to locate variables within the boundaries of the pump and/or variables which are directly associated with the pump operation. Second, exploratory statistical

screening was done to check for variability, operating range and inter-variable relationships. There were initially 10 variables that were screened. Motor current was also monitored, but was not used in the PCA modelling due to the distribution of the motor current not having reliable sensor behaviour in the range of operation. The other 9 variables were left out for the development of the model. A note of correction: Variables for vibration were retained while variables for motor current were not because of concerns with the quality of the sensors.

Variable	Process measurement	Unit	Model use
X1	Motor current	A	Screened; excluded from PCA due to sensor reliability issue
X2	Pump DE vibration	m/s ² pk	Retained
X3	Pump NDE vibration	m/s ² pk	Retained
X4	Pump hexane flush inlet	m ³ /h	Retained
X5	Seal pot level	%	Retained
X6	Pump flow	m ³ /h	Retained
X7	Suction temperature	°C	Retained
X8	Suction pressure	Psi	Retained; also used for functional-failure definition
X9	Discharge pressure	Psi	Retained
X10	Discharge temperature	°C	Retained

Table 1. Process variables used for screening and model development.

2.3 Data cleaning and failure definition

The historian data was pre-processed before developing the model. For a few data points missing that did not correspond to any failure event, local window means were used as a pragmatic means of filling in missing data, which is suitable for long time series in industrial environments [12]. The model was trained without using the extended flat-line segments, physically impossible values and clearly corrupted sensor segments in order to avoid the model learning the behaviour of defective sensors as "normal" operation. These parameters for pre-processing and normalisation were then applied to a uniform degree throughout the inference process.

The definition of functional failure was based on the reliability concept—when an item fails to perform a required function [13]. Operational safety and process engineering considerations determined that a functional failure for the P-101 A/B pump was a suction pressure lower than 1200 psi. A running-status indicator (RS) was therefore programmed to be 1 if the suction pressure was greater than 1200 psi, and 0 if it was less. These labels were not used for model training, but were only used for post hoc validation of the unsupervised framework.

3. Proposed adaptive anomaly-detection framework

3.1 Framework overview

The proposed framework is for the scalable reliability monitoring for industrial applications with limited labelled failure data, and with evolving operating baselines. The workflow involves 5 steps: sensor screening, and data cleaning; statistically defined operating-range normalisation; PCA transformation to a reduced correlated-feature space; Euclidean-distance anomaly scoring from a current normal baseline centroid; Gaussian-based thresholding/warning/alerting with dynamic recalibration. A sensor-contribution layer ranks the variables associated with each alert and gives assistance in formulating a root cause diagnosis.

The framework is generic as it does not rely on a failure signature that is specific to a pump. The inputs required are the historian data, engineering relevant process variables, initial operating window which is stable and a process definition of functional failure to be used for validation. The methodology is generalized for implementing this approach on other industrial assets, with the same modelling logic and after the asset-specific sensor selection and validation.

3.2 Operational range estimation

The first step was to establish an operating range for each of the sensors kept. If the design limits were not available or not appropriate for all variables, design limits were estimated from the filtered historical data based on the mean and standard deviation of each sensor. The lower statistical limit (LSL) and upper statistical limit (USL) for sensor i were determined to be:

$$\text{Low}_i = \mu_i - 2\sigma_i \quad (1)$$

$$\text{High}_i = \mu_i + 2\sigma_i \quad (2)$$

μ_i is the mean and σ_i is the standard deviation of sensor i for the duration of its operation that is screened. This interval is a common scale for the comparison of variables having different physical units, values and ranges.

3.3 Normalisation

The sensor variables measured were retained and each one had a different unit and scale, so normalisation was needed prior to PCA to avoid having high magnitude variables dominating the reduced feature space [14]. Operating limit normalisation:

$$x'_{\{i,t\}} = (x_{\{i,t\}} - Low_i) / (High_i - Low_i) \quad (3)$$

where $x_{\{i,t\}}$ is the data from sensor i at time t . For online scoring, baseline minimum and maximum values were also recorded for the subsequent observations to be transformed consistently with baseline:

$$x'_{\{i,t\}} = (x_{\{i,t\}} - Min_i) / (Max_i - Min_i) \quad (4)$$

3.4 PCA transformation and anomaly scoring

Suppose that x_t is the normalised multivariate sensor vector at time t . The current stable normal operating window (SLOW) was used to fit a PCA, which was then applied to reduce x_t to a lower dimensional k dimensional score vector y_t . The PCA was fitted on the current stable normal operating window (SLOW) and applied to reduce x_t to a lower dimensional k dimensional score vector y_t . The first two components accounted for a large proportion of the baseline variance for each window of recalibration and therefore were used for anomaly scoring and visual interpretation for this study, $k = 2$.

Given the loading matrix W_k of the first k principal component loadings the PCA score vector is given by:

$$y_t = W_k^T x_t \quad (5)$$

The current normal operating window centroid in PCA space is found to be:

$$c = (1/n) \sum_{t=1}^n y_t \quad (6)$$

The anomaly score, at the time t , is then equal to the Euclidian distance between the PCA score at t and the baseline centroid:

$$D_t = \|y_t - c\|_2 \quad (7)$$

A small distance is said to be similarity to the current learned normal state, while a large distance is said to be multivariate deviation from normal operation.

3.5 Justification for PCA-based scoring and Gaussian thresholding

Our goal was to create a generic, transparent and computationally efficient anomaly-detection framework, and not an asset-specific black-box classifier, so that's why we chose PCA. Process variables such as flow, pressure, temperature, vibration and seal variables are correlated and control the behaviour of an industrial pump. This is an appropriate application of PCA as it is used to map correlated multivariate sensor data into a reduced dimensional space, while retaining the dominant variance structure of normal operation [15]. This provides the framework to identify unusual interactions of sensor behaviour rather than the univariate violations of thresholds.

It uses significantly less computing power than deep autoencoders, recurrent neural networks or transformer-based models, and is more easily recalibrated, and has greater interpretability for maintenance engineers. This helps to meet the envisioned objective of the framework as a scalable reliability-monitoring model which is replicated to other asset types while minimizing site-specific model engineering. Architectures that can deliver high accuracy is more complex, and require large numbers of labelled or simulated anomaly samples to train, or have high computation cost, limiting the architectures to deployment in industrial plants where asset populations are not necessarily homogeneous [18-21].

A simple, interpretable and dynamically updateable alarm logic was provided with the use of the Gaussian thresholding. The distribution of the Euclidean distances from the baseline centroid was modelled as a normal distribution in the anomaly-scoring step. This assumption is substantiated with practical and theoretical elements. The distance scores are affected by a number of small variations across correlated sensor variables, so that the overall spread of distance scores for the baseline distance values is fairly evenly distributed around a mean distance operating radius. The histogram of distances in this study had little skewness in the baseline data and a normal approximation was used. Second, Gaussian thresholding is in harmony with accepted industrial process-monitoring practice, which involves reporting statistical unusual behaviour when the process falls outside of three standard deviations on either side of the mean.

The warning threshold was thus taken as $\mu_D + 3\sigma_D$ and the alert threshold was taken as $\mu_D + 5\sigma_D$, where μ_D and σ_D are the mean and standard deviation of the baseline distance scores, respectively:

$$T_{\text{warning}} = \mu_D + 3\sigma_D \quad (8)$$

$$T_{\text{alert}} = \mu_D + 5\sigma_D \quad (9)$$

The $\mu_D + 5\sigma_D$ warning threshold is used to alert to higher deviations and is more sensitive to early abnormal behavior but less sensitive to normal variation, while the $\mu_D + 3\sigma_D$ threshold alerts to a higher deviation and has more sensitivity to early abnormal behavior and less sensitivity to normal variation. While euclidean distances are non-negative in real life and can, in theory, have a chi or Rayleigh distribution under a perfect multivariate-normal assumption, the Gaussian approximation is still good for the high percentile point that is used for warning generation purposes. The robustness was also assessed by comparing the normal based warning threshold with the empirical 99.7th percentile distance value of the baseline data and it was shown that these two values are quite similar. So, the Gaussian fit could be used in industrial anomaly detection to get a reliable, lightweight and easily recalibrated thresholding mechanism.

This thresholding selection is also consistent with the overall goal of the study to show that a simple, yet computationally very efficient model can attain good reliability performance without the massive computing power of deep architectures or the labeled data needs of supervised models.

3.6 Dynamic recalibration

Dynamic recalibration is the key component in the framework. Upon a restart, maintenance action, operating transition or post-anomaly recovery, the algorithm looks for a new stable operating window of $\sim 100h$. When the candidate window meets the data quality and stability, the PCA fitting, centroid estimation and gaussian threshold fitting are performed again. A new baseline is then computed and the next baseline used for the future scoring.

This logic will minimise false alarms arising from legitimate operating changes whilst maintaining the sensitivity for degradation when compared to the latest healthy operating condition. It is not possible to go back to design during unstable/close to failure times as this normalise incipient degradation. The stable-window requirement and/or data-quality screening thus serve as a protection against undesirable updating of the baseline.

3.7 Sensor contribution and root-cause interpretation

After each alert, the PCA loadings and standardised deviations from each of the sensors were analysed to improve the interpretability of the diagnosis. The PCA loadings show which of the variables are most important to the main modes of normal operating variation, and the z-scores at the sensor level show which process variables exhibited the largest deviations from normal for every alert. This interpretation is two layered which enables the model to differentiate between statistical anomaly detection and engineering diagnosis.

For instance, alerts with high scores for negative flow deviation and discharge-pressure deviation would indicate a slurry restriction/impediment or obstruction of the impeller. Transient process disturbances are more likely to be the cause of alerts that are primarily based on discharge-temperature excursions. This interpretation layer further enhances the applicability of this framework as it enables maintenance personnel to progress from “anomaly detected” to “probable process area requiring inspection”.

3.8 Implementation and validation metrics

This framework has been carried out in python with the help of the libraries pandas, NumPy, scikit-learn, and SciPy for the time-series analysis, numerical computation, PCA and Gaussian fitting, respectively [16]. Validation was undertaken to demonstrate its ability to detect functional failures (where documented) as well as minimize false alarms during normal operating transitions and provide adequate lead time for maintenance planning.

True positive was defined as model alerts made before a known functional failure during the evaluation period. A false positive was when a model alert did not result in a functional failure during the evaluation period. There were documented functional failures that were not preceded by a model alert, termed as False negatives. Precision, recall and F1-score were defined as:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (10)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (11)$$

$$\text{F1} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall}) \quad (12)$$

The first warning/alert was defined as the lead time, which was defined as the period of time between the first warning/alert and the subsequent failure. This is directly relevant to the reliability-centred

maintenance because it is the time window in which the pump is inspected, loaded or the pump is changed or planned intervention is required.

4. Results

4.1 Sensor screening and baseline visualisation

All of the process variables that remained in the process were found to be within plausible ranges by exploratory screening. Motor current was omitted due to unreliable sensor behaviour in comparison to its operating range. Variables of vibration, hexane flush, seal-pot level, flow, pressure and temperature variables were kept. This allowed the baseline learned by the model to not be affected by the known defect in the sensors' quality.

The first stable normal baseline was identified from 2017-01-09T12:45:00Z to 2017-01-13T16:30:00Z. During this 100 h period, the operation of the pumps remained steady and the conservatively operated variables remained within the operation limits. Two dimensional baseline for 78.36% of variance was obtained by PCA of this window. The initial warning threshold, alert threshold and maximum standard deviation of distances were 0.3077, 0.4637 and 0.0862 respectively.

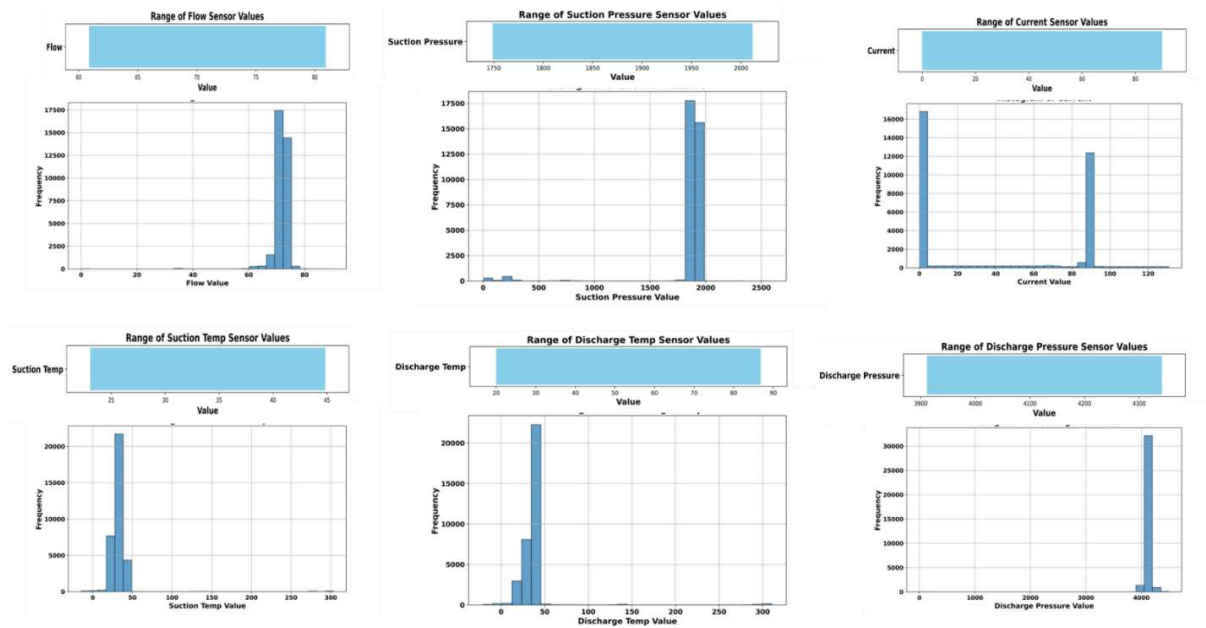


Fig. 2. Histograms of sensor data collected from the historian system during operational periods. The current signal was excluded from model building because of sensor reliability issues.

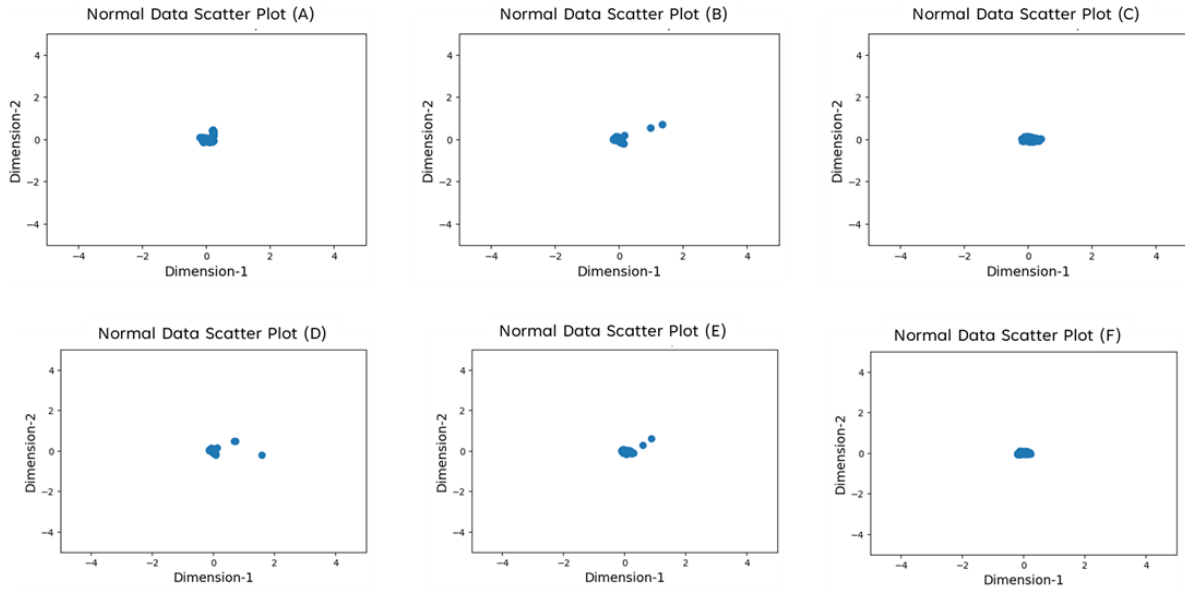


Fig. 3. PCA-transformed scatter plots of normal operating baselines identified during dynamic recalibration. Each cluster represents a stable operating window in two-dimensional PCA space.

4.2 Dynamic recalibration events

A variety of normal operating baselines were found over the three-year period of validation. The events for recalibration, explained variance, warning thresholds, alert thresholds and maximum SDs of the baseline distance scores are summarized in Table 2. The cumulative variance explained by the first two PCA components was between 76.13% and 92.03%, thus showing that a significant amount of the structure in different operating states was captured in this representation. The differences in the thresholds between baselines also demonstrate the need for not using a fixed threshold for long-term deployment.

Recalibration event	Normal operating range (start-end)	Explained variance	Warning threshold	Alert threshold	Year
Initial baseline	2017-01-09T12:45:00Z - 2017-01-13T16:30:00Z	78.36%	0.3077	0.4637	2017
Recalibration 1	2017-07-22T02:00:00Z	76.13%	0.3185	0.4634	2017

	- 2017-07-26T06:00:00Z				
Recalibration 2	2017-11-23T21:45:00Z - 2017-11-28T01:30:00Z	91.20%	0.1676	0.2548	2017
Recalibration 3	2018-01-02T01:00:00Z - 2018-01-06T04:45:00Z	77.58%	0.2554	0.3654	2018
Recalibration 4	2018-05-09T18:00:00Z - 2018-05-13T21:45:00Z	89.23%	0.3990	0.6170	2018
Recalibration 5	2018-11-06T14:15:00Z - 2018-11-10T18:00:00Z	83.75%	0.2515	0.3786	2018
Recalibration 6	2019-01-25T09:30:00Z - 2019-01-29T13:30:00Z	89.36%	0.3352	0.5015	2019
Recalibration 7	2019-09-07T09:45:00Z - 2019-09-11T13:30:00Z	92.03%	0.3422	0.5350	2019

Table 2. Key parameters for identified normal operating baselines and recalibration events.

The value in the proposed approach is illustrated by the changes in the threshold after recalibration. The model did not impose changes to the 2017 baseline to fit future data. Instead it recalculated the centroid and the thresholds after stable operating windows were found, thus retaining sensitivity to degradation while not triggering the alarm system when the system was in normal operating mode.

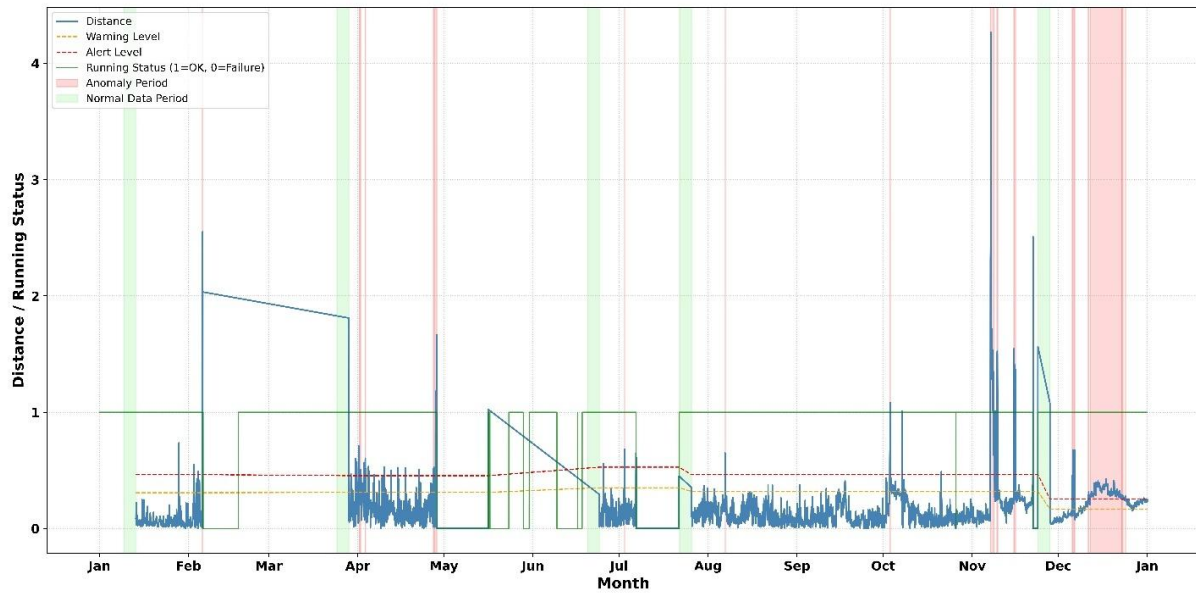


Fig. 4. Deviation plot for 2017 showing anomaly scores with dynamically adjusted warning and alert thresholds.

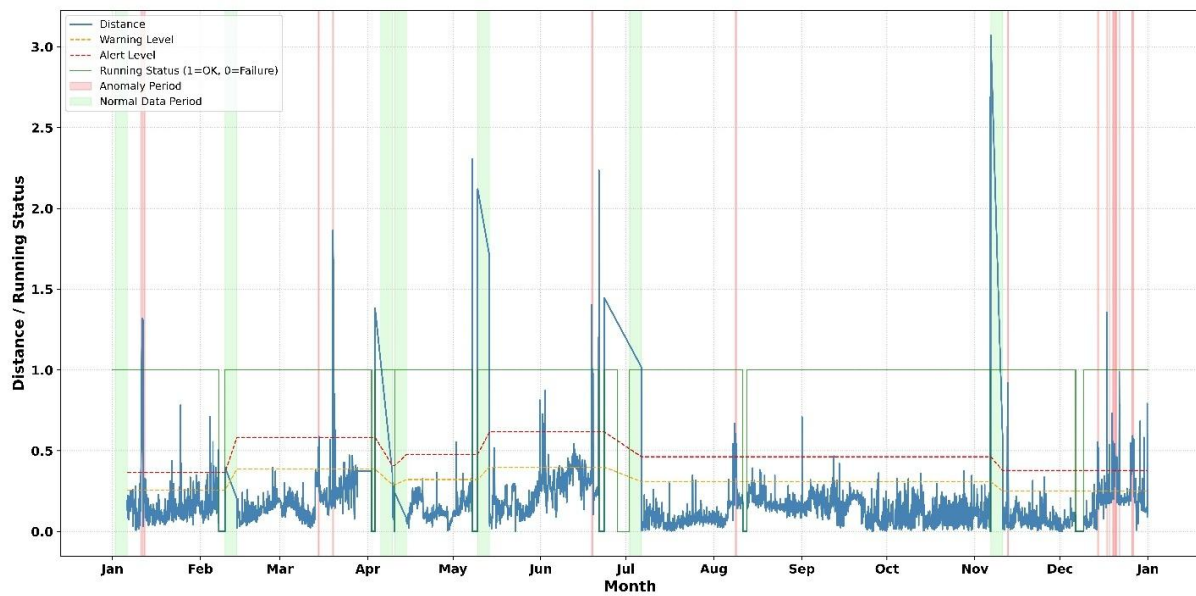


Fig. 5. Deviation plot for 2018 showing adaptive anomaly detection and recalibration across operational changes.

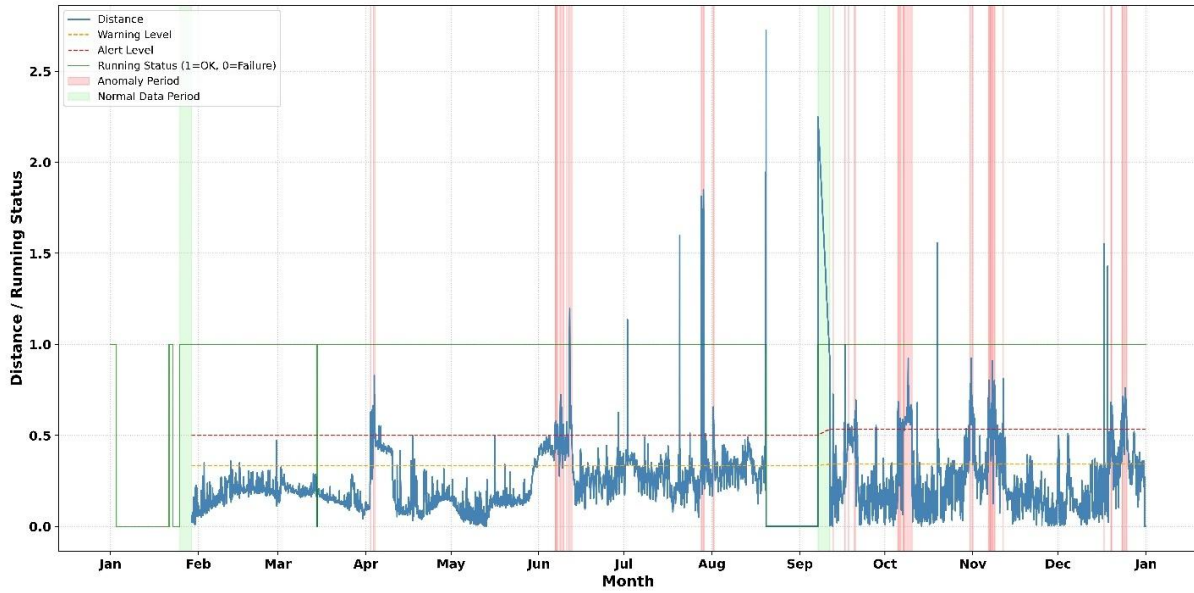


Fig. 6. Deviation plot for 2019 showing the persistence of adaptive thresholding across the final validation year.

4.3 Alert classification and diagnostic interpretation

The framework identified all of the 44 functional failures in the retrospective validation set. The model achieved 44 true positives, 8 false positives and zero false negatives. This resulted in 100% recall, 84.62% precision and an F1-score of 91.67%. Compared to the average warning lead time of about 72 h, there was a more realistic time frame for action (inspection, load adjustment, etc. pump changeover and planned maintenance).

A 100% recall score is within the range of this dataset, asset and failure definition. It demonstrates failure sensitivity over three years of varying operation in the data available for validation. It is not to be read as a guarantee that the model will identify all the failure modes for all asset classes and or all operating situations. Such a widespread statement needs to be confirmed.

The true positive alerts represented showed that many of these were dominated by a negative flow deviation and discharge-pressure deviation, which is physically consistent with slurry restriction/impeller obstruction. A few false positives were related to either a discharge-temperature excursion or short-lived operational disturbances that did not develop into a functional failure based on the suction-pressure. The following cases show the value of using the combination of anomaly scoring and interpretation of the sensors.

Alert	Date/time range	Top sensor	Z-score	Second sensor	Z-score	Classification
1	2018-01-11T00:15Z to 2018-01-11T20:00Z	Discharge Temp	89.33	Suction Pressure	-1.60	False positive
2	2018-01-12T04:30Z to 2018-01-12T12:15Z	Discharge Temp	-20.01	Flow	-2.10	True positive
5	2018-03-19T14:15Z to 2018-03-19T21:45Z	Flow	-18.77	Discharge Pressure	6.41	True positive
6	2018-06-18T21:45Z to 2018-06-19T09:30Z	Flow	-22.93	Discharge Pressure	4.42	True positive
10	2018-11-12T16:00Z to 2018-11-12T19:00Z	Flow	-5.04	Discharge Pressure	3.06	False positive
21	2018-12-26T11:45Z to 2018-12-27T00:30Z	Flow	-6.17	Discharge Pressure	1.87	False positive

Table 3. Representative alerts, sensor contributions and classifications.

The PCA loading structure was found to be different for the pressure flow variables compared to the temperature-related variables, and this was used to make a root cause interpretation of the PCA anomaly score rather than using the anomaly score as a black box output. For instance, Alert 1 was characterized by a false positive caused by temperature deviation in the discharge and Alert 6 was characterized by a true positive, which was mainly due to flow reduction and discharge-pressure

change. This behavior is consistent with that expected in engineering of slurry-loop restriction, and it demonstrates the ability of the framework to aid in operational diagnosis.

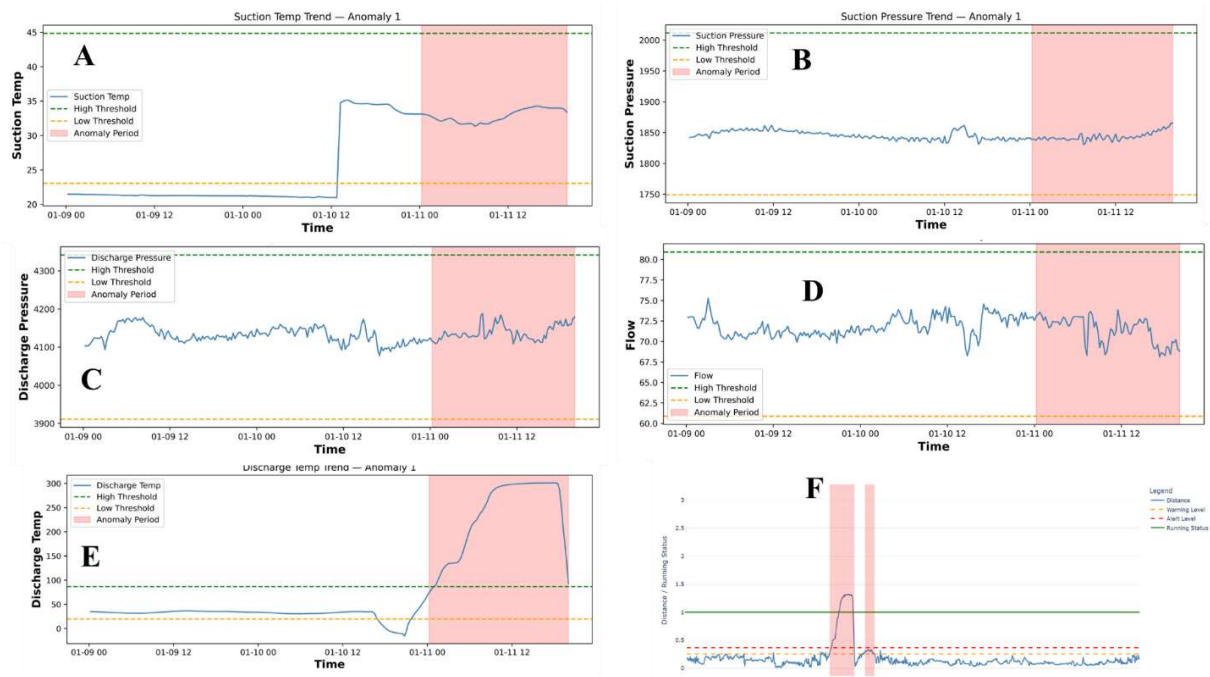


Fig. 7. Detailed sensor behaviour and deviation plot for Alert 1, a false-positive event dominated by discharge-temperature deviation.

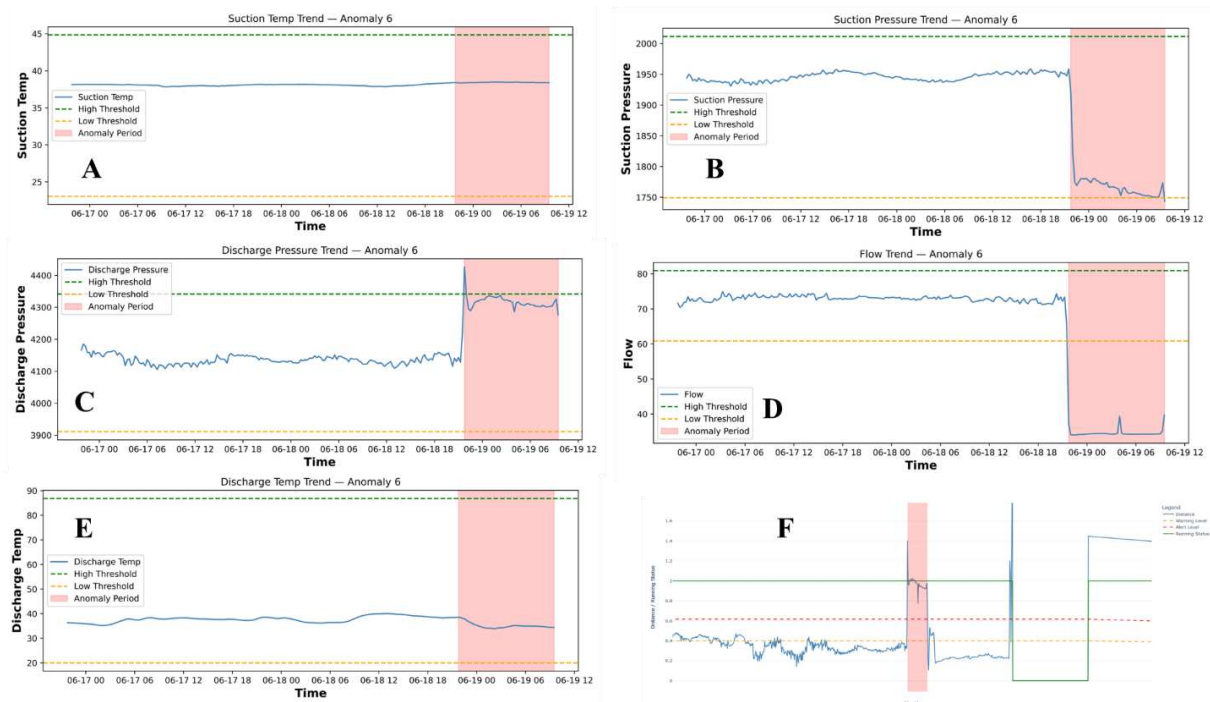


Fig. 8. Detailed sensor behaviour and deviation plot for Alert 6, a true-positive event dominated by flow and discharge-pressure deviations.

5. Discussion

The results demonstrate not only that dynamic recalibration is not a luxury to be added on to unsupervised anomaly detection, but rather a reliability requirement to be met for long-term industrial deployment. If a fixed model is placed on the 1st 2017 baseline, the latter would have to re-tune the thresholds manually for different healthy operating states or the model would have to flag them as abnormal. The proposed framework, on the other hand, refreshed its centroids and thresholds upon identification of stable operating windows, thus maintaining the anomaly detection in line with the present health of the asset. This is in line with recent research that adaptive models perform better in systems that vary their operating conditions and have environmental variations and concept drift [18,19].

A framework is also proposed that takes into account the labelled failure scarcity. If the failure history of a critical asset is large and representative, supervised models can achieve good performance, but a large failure history from critical assets is not readily available as repeated failures are not acceptable in practice. The proposed method learns from normal operation and fails instances are only used for later validation. This is more easily implemented in plants with missing or incomplete or delayed labelled fault classes. Furthermore, its lightweight construction reduces the complexity of calculations and connection with other systems compared to deep-learning or hybrid digital-twin systems, which require more complex training, high data volume and higher computing requirements [20,21].

In the retrospective validation set of 44 functionally failed items, the model was 100% accurate. From a system-safety point of view, it is especially critical to avoid failure where the asset is used to remove heat or pressure or to contain. This is a result to be taken with a pinch of salt though. The recall value is based on a one pump, one HDPE process, one failure definition and one historical validation record. The proposed recalibration logic should be considered an indication of a good attempt to preserve sensitivity in this case, and not as a blanket guarantee of failure detection in all cases.

That doesn't have 8 false positives that are excessive, it has 8 false positives that are meaningful. Some were related to transient temperature or pressure changes which did not lead to functional failure by suction pressure. In reliability, it's not as big of a deal if only a few false alerts get investigated; it's more important that an operator doesn't miss a failure, but if fewer false alerts are being investigated,

the operator will trust the system more. Future implementations should thus include process mode tags, maintenance work orders, operator annotations and contextual information to be able to differentiate transient disturbances from sustained degradation.

The sensor-contribution layer adds to the interpretability, and enhances the utility of the framework. An anomaly score is only a scalar that is pointing to a change, but not a place where operators should look. Ranking the deviating sensors is a way to connect the world of statistical A/D and the engineering diagnostics. Physically consistent true-positive alerts that are dominated by flow reduction and by deviation from discharge pressure are related to slurry restriction and impeller obstruction, respectively. The deviations from the discharge temperature provide false positives, suggesting that there are some process disturbances that can cause statistical anomalies without a functional failure. This type of explanation is more operational than other, generally post hoc, feature attributions.

The study also helps in developing scalable industrial reliability monitoring. There is no need for manually set thresholds for each failure mode as in labelled training failures or asset-specific deep architectures, as proposed. Rather, it needs a stable normal window, pertinent sensor variables, screening and engineering failure definition of data quality for validation. What this implies is that (for example) the same architecture is deployed in a range of asset categories as well – but each one will have to be validated against that asset.

There are limitations. The first step with respect to validation was to validate one pump system in one HDPE slurry polymerisation process. The framework is intended to be generic, but it would be beneficial to see it validated in other equipment types and/or other process environments. Secondly, the criterion for failure was chosen as being when the suction pressure fell below 1200 psi. Future research should link alert performance to maintenance work orders, inspections, parts replaced, and actual failure modes. Third, the motor current was omitted as there were sensor reliability issues; better sensor current measurement could provide useful electromechanical information in future deployments. Fourth, it was a retrospective study. Prospective deployment needs to be done to ensure that the model will increase maintenance actions, decrease downtime, and enhance safety in real time.

Future studies should specifically test the predictions of future multi-asset validations instead of general climate change assertions that are not directly tested in this data set. The next steps that are most relevant are to assess the framework across a variety of pumps, compressors, heat exchangers and other rotating equipment, incorporate maintenance logs and operator feedback, add process-mode and ambient operating-condition tags, evaluate the framework in the presence of gaussian thresholds and

empirical/chi-square thresholds in the presence of different baseline distributions, and evaluate the framework in edge/cloud and/or digital-twin architectures for near-real-time industrial deployment.

6. Conclusion

The aim of this study was to design and test a lightweight, generic and self-recalibrating unsupervised machine-learning framework for adaptive detection of anomalies in industrial asset reliability. The framework includes operating-range normalisation, dimensionality reduction using PCA, anomaly scoring using Euclidean distance, Gaussian warning and alert thresholds, ranking of sensor contributions and dynamic baseline recalibration.

The framework was applied to three years of historic historical data for a critical P-101 A/B pump of a HDPE slurry polymerisation process, and resulted in a 100% recall, 84.62% precision, 91.67% F1 and an average warning lead time of 72 h for all 44 recorded functional failures in the retrospective validation dataset. The results illustrated here show that adaptive unsupervised learning is sensitive to varying operating baselines without the need for labelled failure data to be used for training.

The key contribution is demonstrating a self-recalibrating anomaly-detection framework that, with reduced computational complexity, can achieve scalable and interpretable reliability monitoring, compared to many sophisticated deep-learning solutions. The approach can provide early warning, root cause investigations, maintenance prioritisation and operator decision making. The framework needs to be validated in the future through prospective studies in various asset classes, with the inclusion of maintenance and operator feedback data, and tested in actual industrial deployment in real-time.

Funding

This work has not been supported by any specific grant funding from public, commercial or not-for-profit funding body.

Declaration of competing interest

The authors have no known competing financial interests or personal relationships that could have a potential or apparent influence on the work reported in this paper.

Data availability

The information in this study is from confidential industrial historian records. The aggregated data and processed data that support the conclusions of the work presented in this paper, obtained by the corresponding author upon reasonable request and with institutional and industrial approval.

Ethics approval

Not applicable. No human, animal or personal data was involved in this study. It was derived from information from the sensors on industrial equipment.

Author contributions

D.S.: Conceptualization, Methodology, Data curation, Software, Formal analysis, Investigation, Visualization, Writing - original draft. Writing - review and editing, D.P.P.: Supervision, Methodology, Validation, Project administration. M.K.: Conceptualization, Methodology, Supervision, Writing - review and editing. All the authors have read and approved the final manuscript.

Acknowledgements

The authors appreciate the assistance of the plant personnel in collecting and interpreting the data used for developing the plant historians.

References

- [1] Ahmed Murtaza A, Saher A, Zafar MH, Moosavi SKR, Aftab MF, Sanfilippo F. Paradigm shift for predictive maintenance and condition monitoring from Industry 4.0 to Industry 5.0: A systematic review, challenges and case study. *Results Eng* 2024;24:102935. <https://doi.org/10.1016/j.rineng.2024.102935>.
- [2] Ahmed S, Terefe T, Hailemariam D. Machine learning for base transceiver stations power failure prediction: A multivariate approach. *e-Prime Adv Electr Eng Electron Energy* 2024;100814. <https://doi.org/10.1016/j.prime.2024.100814>.
- [3] Alsulieman A, Ge X, Zeng Z, Butenko S, Khan F, El-Halwagi M. Dynamic risk analysis of evolving scenarios in oil and gas separator. *Reliab Eng Syst Saf* 2024;243:109834. <https://doi.org/10.1016/j.ress.2023.109834>.

- [4] Wahono T, Purniawan A, Mukhlash I, Putri E. Risk-based asset integrity management in the oil and gas industry from traditional to machine learning approaches: A systematic review. *Results Eng* 2025. <https://doi.org/10.1016/j.rineng.2025.107287>.
- [5] Al-Sabaei AM, Alhussian H, Abdulkadir SJ, Jagadeesh A. Prediction of oil and gas pipeline failures through machine learning approaches: A systematic review. *Energy Rep* 2023;10:1313-1338. <https://doi.org/10.1016/j.egy.2023.08.009>.
- [6] Cabarcos A, Paz C, Suarez E, Vence J. Application of supervised learning algorithms for temperature prediction in nucleate flow boiling. *Appl Therm Eng* 2024;240:122155. <https://doi.org/10.1016/j.applthermaleng.2023.122155>.
- [7] Kazim M, Pirim H, Le C, Le T, Yadav O. Multilayer GNN for predictive maintenance and clustering in power grids. *iScience* 2025. <https://doi.org/10.1016/j.isci.2025.113490>.
- [8] Duan D, Han S, Wang Z, Pang C, Yao L, Liu W, et al. Multivariate state estimation-based condition monitoring of slurry circulating pumps for wet flue gas desulfurization of power plants. *Eng Fail Anal* 2024;159:108099. <https://doi.org/10.1016/j.engfailanal.2024.108099>.
- [9] Wang K, Yuan X, Cao Z, Chen G, Yuan X, Yang C, et al. Capturing principal features in slow industrial processes for anomaly detection application. *J Process Control* 2025;153:103487. <https://doi.org/10.1016/j.jprocont.2025.103487>.
- [10] Kim J, Cho S-B. Unsupervised video anomaly detection by memory network with autoencoders in Euclidean and non-Euclidean spaces. *Pattern Recognit* 2025;167:111759. <https://doi.org/10.1016/j.patcog.2025.111759>.
- [11] Morsi I, Mohy El-Din L. SCADA system for oil refinery control. *Measurement* 2014;47:5-13. <https://doi.org/10.1016/j.measurement.2013.08.032>.
- [12] Xiao H, Cao R, Chen Z, Hong C, Wang J, Yao M, et al. Handling missing data in large-scale TBM datasets: Methods, strategies, and applications. *Intell Geoeng* 2025;2:109-125. <https://doi.org/10.1016/j.ige.2025.07.001>.
- [13] IEC 60050. International Electrotechnical Vocabulary. Geneva: International Electrotechnical Commission; 1990.
- [14] Nicosia GP. Machine learning, optimization, and data science. Cham: Springer Nature; 2020.

- [15] Wang X, Chen Y. Unsupervised detection of multivariate geochemical anomalies using a high-performance deep autoencoder Gaussian mixture model. *J Geochem Explor* 2025;271:107671. <https://doi.org/10.1016/j.gexplo.2025.107671>.
- [16] López-Flores FJ, Ochoa-Barragán R, Raya-Tapia AY, Ramírez-Márquez C, Ponce-Ortega JM. Machine learning with Python. In: *Machine Learning Tools for Chemical Engineering: Methodologies and Applications*. Elsevier; 2025. p.31-55. <https://doi.org/10.1016/B978-0-443-29058-9.00004-5>.
- [17] Crespo Márquez A, Serra Parajes J, de la Fuente Carmona A, Sola Rosique A. Integrating complex asset health modelling techniques with continuous time simulation modelling: A practical tool for maintenance and capital investments analysis. *Comput Ind* 2021;133:103507. <https://doi.org/10.1016/j.compind.2021.103507>.
- [18] Ben Haha MA, Bohli A, Haddour N, Bouallegue R. A systematic review of industrial IoT anomaly detection and the forensic interpretability gap. *Electronics* 2026;15(11):2240. <https://doi.org/10.3390/electronics15112240>.
- [19] Subashree S, Rajakumaran M, Pushpa G, Manivannan R. Adaptive machine learning models for predictive maintenance in industrial internet of things (IIoT) systems. *Sci Rep* 2026;16:12451.
- [20] Mavrelis I, et al. Hybrid deep learning for predictive maintenance in industrial machinery using LSTM and MLP models. *Machines* 2026;14(2):191.
- [21] Khan T, et al. Data-driven digital twin framework for predictive maintenance of smart manufacturing systems. *Machines* 2025;13(6):481. <https://doi.org/10.3390/machines13060481>.
- [22] El Hammoumi Z, et al. Predictive maintenance approaches: A systematic literature review. *Eng Proc* 2025;112(1):70.
- [23] Elkateb S, Métwalli A, Shendy A, Abu-Elanien AEB. Machine learning and IoT-based predictive maintenance approach for industrial applications. *Alex Eng J* 2024;88:298-309. <https://doi.org/10.1016/j.aej.2023.12.065>.
- [24] Gao J, Liu Y, Zhao Q, Ai X, Shan L. Research on sensor anomaly detection and fault location in nuclear power plants based on GAN-IAAKR model. *Ann Nucl Energy* 2025;219:111462. <https://doi.org/10.1016/j.anucene.2025.111462>.
- [25] Guo L, Gao J, Yang J, Kang J. Criticality evaluation of petrochemical equipment based on fuzzy comprehensive evaluation and a BP neural network. *J Loss Prev Process Ind* 2009;22:469-476. <https://doi.org/10.1016/j.jlp.2009.03.003>.

- [26] Mohammed A. Data driven-based model for predicting pump failures in the oil and gas industry. *Eng Fail Anal* 2023;145:10719. <https://doi.org/10.1016/j.engfailanal.2022.107019>.
- [27] Ratnayake RM. Application of a fuzzy inference system for functional failure risk rank estimation: RBM of rotating equipment and instrumentation. *J Loss Prev Process Ind* 2014;29:216-224. <https://doi.org/10.1016/j.jlp.2014.03.002>.
- [28] Wu S, Yang P, Li D, Chen G, Wang Z. Towards predictive maintenance: A performance evaluation framework for cooling towers in HVAC systems. *Build Environ* 2025;284:113443. <https://doi.org/10.1016/j.buildenv.2025.113443>.
- [29] Xie L, Håbrekke S, Liu Y, Lundteigen MA. Operational data-driven prediction for failure rates of equipment in safety instrumented systems: A case study from the oil and gas industry. *J Loss Prev Process Ind* 2019;60:96-105. <https://doi.org/10.1016/j.jlp.2019.04.005>.
- [30] Yucesan YA, Dourado A, Viana FA. A survey of modeling for prognosis and health management of industrial equipment. *Adv Eng Inform* 2021;50:101404. <https://doi.org/10.1016/j.aei.2021.101404>.
- [31] Dereci U, Tuzkaya G. An explainable artificial intelligence model for predictive maintenance and spare parts optimization. *Supply Chain Anal* 2024;8:100078. <https://doi.org/10.1016/j.sca.2024.100078>.

Appendices

Code	Meaning	Unit
X1	Motor current	A
X2	Pump DE vibration	m/s ² pk
X3	Pump NDE vibration	m/s ² pk
X4	Pump hexane flush inlet	m ³ /h
X5	Seal pot level	%
X6	Pump flow	m ³ /h
X7	Suction temperature	°C
X8	Suction pressure	psi
X9	Discharge pressure	psi
X10	Discharge temperature	°C

So, for example, **X6** represents pump flow, **X8** represents suction pressure, and **X9** represents discharge pressure. The manuscript also states that **X1 motor current was monitored but excluded from the PCA model due to sensor reliability issues.**