Journal of the Brazilian Society of Mechanical Sciences and Engineering (2026) 48:504
https://doi.org/10.1007/s40430-026-06488-2

TECHNICAL PAPER

Detection of oil leak and seawater influx through partial ruptures in subsea petroleum production pipelines

Daniel Centurion Barrionuevo[1,2] · Galileu Henke de Oliveira[1] · Anderson Carlos Faller[1] · Saon Crispim Vieira[1] · Benno Waldemar Assmann[1] · Jean Carlos Dias de Araújo[1] · Alexey Pavlov[2]

Received: 18 June 2025 / Accepted: 30 March 2026

Abstract

Leak detection in subsea petroleum production pipelines represents a critical challenge for offshore operations, especially under increasingly stringent environmental and regulatory requirements. This work presents a complete framework for detecting oil leaks and seawater influxes through partial ruptures in deepwater production systems. The methodology integrates the generation of synthetic time-series data using multiphase transient flow simulation and the training of supervised machine learning models, validated through real and simulated datasets of operational events. Using pressure and temperature measurements from four standard field sensors, the proposed system detects anomalies associated with both leakage and influx phenomena with high sensitivity and low false-alarm rates. A LightGBM-based classifier demonstrated superior performance, achieving detection rates above 95% for moderate or severe ruptures after statistical smoothing and operational parameter tuning. The system was validated through a large-scale testing campaign and has already been deployed for real-time monitoring of multiple producing wells. Results confirm that this software-based detection approach enhances operational safety, reduces environmental risk, and can complement existing hardware-based monitoring systems in offshore fields.

Keywords Oil leak · Seawater influx · Pipeline · Rupture · Deepwater · Machine learning

Abbreviations

RTTM	Real-time transient modeling
API	American Petroleum Institute
DHSV	Downhole safety valve, a safety equipment installed in the well and kept open by a pressurized hydraulic line
GOR	Production gas-oil ratio
P-TPT	Pressure at the WCT (TPT)
P-UP-PCK	Pressure upstream of the production choke
PCK	Production choke valve
PDG	Permanent downhole gauge, usually with piezoresistive or quartz crystal technology for pressure (accuracy in the range of 0.03%) and platinum resistance temperature detector (RTD) for temperature (accuracy of $0.5°C$)
ROV	Remote operated vehicle, subsea robot employed in offshore operations
RTTM	Real-time transient modeling
SSLDS	Subsea leak detection system
T-TPT	Temperature at the WCT (TPT)
T-UP-PCK	Temperature upstream of the production choke.
TPT	Temperature and pressure transmitter, located at the WCT, usually with piezoresistive or quartz crystal technology for pressure (accuracy in the range of 0.05%) and thermistor or platinum RTD for temperature (accuracy of $0.5°C$)
WCT	Wet Christmas tree, subsea valve assembly connecting well to pipeline

Communicated by Angela Ourivio Nieckele.

✉ Galileu Henke de Oliveira
galileu@petrobras.com.br

[1] Petrobras, Rio de Janeiro, Brazil

[2] NTNU, Trondheim, Norway

Published online: 23 May 2026

1 Introduction

Fast and reliable leak detection in subsea production pipelines and equipment has long posed a significant challenge to the oil industry. In recent years, this issue has become even more critical due to increasing regulatory and environmental pressures worldwide [1–3]. To address these concerns, industrial standards have been continuously developed and updated to establish and disseminate best practices for subsea pipeline leak detection in the petroleum sector [4].

The primary objective of a leak detection system is the prompt identification of pipeline or equipment ruptures or perforations, which—depending on the circumstances—can result in oil or gas leakage, or seawater influx. In the Brazilian offshore context, this challenge has become particularly relevant, as the majority of oil production now originates from subsea completions. Many of these tiebacks employ flexible pipelines, which are especially susceptible to degradation in environments with high CO_2 content [5], a characteristic feature of pre-salt oil fields. Consequently, there has been a concerted effort to develop and implement subsea leak detection systems (SSLDS).

The scenario addressed in this work involves a subsea oil production well directly connected to a production platform, where pipeline failure can potentially occur at any point along the tieback. The fluid transported within these pipelines typically comprises a mixture of gas, oil, and water, presenting a complex multiphase flow problem. This scenario will be better explained in Sect. 1.2.

This work is organized as follows:

1. *Introduction*—presents the problem and its motivation. A literature review on leak detection is provided, followed by a description of the main Brazilian offshore production well scenario. A subsection summarizes key definitions, and the phenomena that may occur after a rupture—namely, oil leakage and seawater influx—are explained.
2. *Methodology*—details the procedures used to develop, validate, and implement a system for detecting partial ruptures in subsea petroleum production pipelines. Synthetic time-series data were generated using multiphase flow simulations, and several machine learning algorithms were trained and compared to select the best-performing model.
3. *Test and assessment*—presents the tests performed under different evaluation metrics to assess the final model's performance.
4. *Conclusions*—summarizes the main findings and highlights the contributions of this work.

1.1 Literature review

The literature on leak detection in single-phase pipeline flow is extensive, with numerous methods developed for water, crude oil, and gasoline transport. However, as noted by Rahman et al. [6], research addressing multiphase flow—especially mixtures of gas and liquid—remains limited. This area is still largely unexplored and continues to present significant challenges for the oil industry.

Recent comprehensive reviews by Korlapati et al. [2], Abubakar et al. [7], Yuan et al. [1], and Behari et al. [8] provide systematic and bibliometric analyses of hundreds of publications, identifying prevailing research trends and technological approaches. Another valuable resource is the technical report by Krohne [9], which surveys various leak detection methods, their requirements, application scenarios, advantages, and limitations. Collectively, these works underscore the central importance of reliable leak detection for mitigating economic losses, environmental impacts, and safety hazards, as pipelines remain susceptible to deterioration, vandalism, and operational faults.

A consistent taxonomy emerges from these reviews, classifying leak detection techniques into three main categories [2]:

- *Externally based methods:* These employ dedicated hardware, such as acoustic emission sensors and optical fiber sensors, for real-time leak detection. They are recognized for their high sensitivity and precision, particularly in locating defects. Recent advances include integrating sensor arrays and using robotics for in-line inspection.
- *Vision/inspection methods:* These rely on periodic inspections, which may be performed by humans, trained animals, drones, or robots. In the oil and gas sector, inspections are often carried out by divers or remotely operated vehicles. However, these methods do not provide real-time monitoring. The exception is satellite-based oil spill detection techniques [10], which are extremely reliable for identifying any oil spill on the ocean surface; however, they do not provide information about the source of the spill, nor can they detect deepwater leaks before the oil reaches the sea surface—which could be far away from the source.
- *Internally/computationally based methods:* These approaches use process data (e.g., pressure, temperature, flow rate) as inputs to model-based state estimation or machine learning algorithms. Software-based methods are increasingly prominent due to advancements in machine learning and statistical modeling, offering potential for real-time anomaly detection. Their effectiveness,

however, depends on the availability of high-quality, representative datasets.

Many studies highlight that combining hardware and software methods—hybrid approaches—tends to yield superior results, enhancing both the accuracy and reliability of leak detection, especially for small or chronic leaks that are otherwise difficult to identify. For subsea pipelines, particular challenges include installation complexity and the need for robust, low-maintenance solutions in harsh environments. As a result, hardware-based methods are less feasible for extensive oil and gas gathering networks due to high costs and lack of widespread deployment of subsea sensors. In contrast, machine learning, especially deep learning and hybrid models, has shown rapid progress and strong potential.

Internally/computationally based methods can be further subdivided [2]:

- *Pressure/flow monitoring:* Detects leaks based on unexpected disturbances or abnormal values in pressure and flowrate measurements. Commonly used in offshore platforms, this includes monitoring pressure differentials at production chokes and abrupt or implausible changes in sensor readings.
- *Balancing systems:* Analyze mass balance at various measurement points. This approach is impractical without multiphase flowmeters installed both subsea and topside.
- *Real-time transient modeling (RTTM):* Relies on real-time numerical simulations based on physical principles, comparing simulated and measured values to detect anomalies.
- *Negative pressure wave:* Utilizes high-frequency sensors to detect rarefaction waves generated at leak initiation. While fast and sensitive for single-phase incompressible fluids, this method is less effective for slowly developing ruptures or long pipelines [8, 9, 11, 12].
- *Statistical and machine learning methods:* Employ multiple variables as input to statistical or machine learning models for leak detection. This is the approach adopted in the present work.

According to Yuan et al. [1], most machine learning methods for leak detection remain at the theoretical validation stage, though some practical implementations exist, such as neural networks trained on accelerometer signals mounted on pipeline walls [13]. For multiphase leaks, dedicated research is scarce. Ji et al. [14] demonstrated promising results with acoustic sensors using empirical mode decomposition, though externally based methods are not the focus of this study. Kam [15] and Figueiredo et al. [16] proposed mechanistic models to predict pressure and flowrate changes due to leaks. While valuable conceptually, these models may not suit the present scenario, as they do not account for temperature variations, possible water influx, or require frequent updates of well-specific physical models. Havre et al. [17] presented an RTTM system using digital twin estimations compared to real measurements to identify leak signatures.

Given the limited literature on multiphase flow leak detection and the unique context of offshore production, no established methodology currently exists for detecting small or partial ruptures in such systems. There is consensus in the literature that no single technique is sufficient for all scenarios; instead, methods are complementary. Despite the lack of comprehensive standards, machine learning approaches have emerged as promising tools for leak detection in complex, multiphase environments.

1.2 System configuration

As discussed in [18], the petroleum industry employs a wide variety of subsea production system architectures. One of the most common configurations is based on wet satellite wells, which also predominate in the pre-salt basin.

Figure 1 illustrates a typical wet-completion satellite well. In this arrangement, oil and gas flow from the reservoir through the production tubing to the wet christmas tree (WCT). From there, the fluids enter a subsea production pipeline—whose integrity is of major concern due to the risk of leakage—typically consisting of two branches: the flowline, which is nearly horizontal and lies on the seabed, and the riser, a catenary-shaped segment that extends up to the platform. At the end of the riser, just before the processing plant, a production choke valve (PCK) is installed to regulate the well's flowrate by introducing a pressure drop, a feature particularly relevant for leak detection.

A few pressure and temperature sensors are installed along the flow path: a permanent downhole gauge (PDG) is located at the bottom of the well, while a pressure and temperature transmitter (TPT) is positioned at the WCT. Additional pressure and temperature sensors are placed upstream and downstream of the choke, providing valuable information about the current flow rate. Notably, in this configuration, direct multiphase flowrate measurement is not available.

In this work, the following sensor measurements were used as inputs:

- *P-TPT:* Pressure from the TPT sensor (bar);
- *T-TPT:* Temperature from the TPT sensor (°C);
- *P-UP-PCK:* Pressure upstream of the production choke (bar);

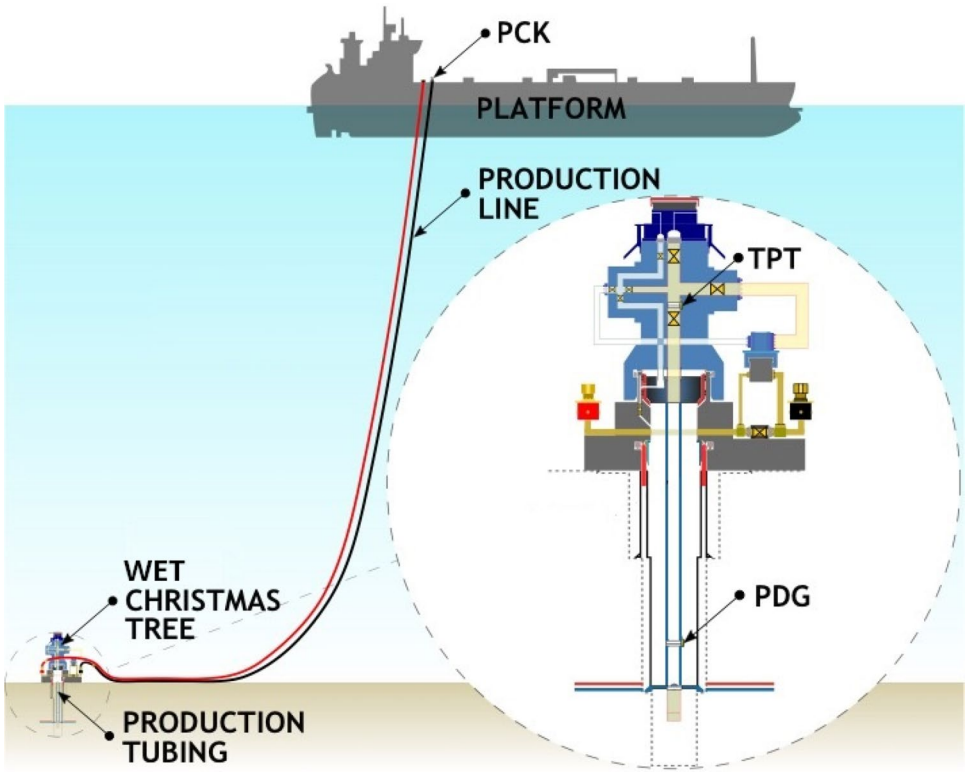

Fig. 1 Typical tieback of a satellite offshore oil well (adapted from [19])

- *T-UP-PCK:* Temperature upstream of the production choke (°C).

In many cases, the PDG sensor is unavailable due to failures. The TPT (also referred to as the *subsea sensor*) is often operational, whereas the sensors at the platform (commonly known as *topside sensors*) are consistently available. The decision to exclude PDG data was made to ensure the classifier's applicability to the majority of wells, as PDG information is often unreliable or unavailable and is less relevant to the subsea pipeline, which is the primary focus of this study.

Oil wells can operate either as naturally flowing—producing without any external energy supply—or with artificial lift, which requires external energy to transport the fluids to the platform. The most common artificial lift method for offshore wells in the present scenario is gas lift, wherein high-pressure gas is injected into the production tubing to reduce fluid density and, consequently, the downhole pressure. This study is restricted to naturally flowing wells, as they are most prevalent in pre-salt fields with high CO_2 content.

1.3 Definitions

The following definitions, adapted from [18], are used throughout this work:

- *Sensor:* equipment that returns the value of a physical variable at a specific position. As explained in Sect. 1.2, four different sensors were considered in this work.
- *Timestamp:* the time at which a recording is produced by any sensor.
- *Measurement value:* the numerical value recorded by a sensor at a specific timestamp.
- *Measurement unit:* the engineering unit associated with the value recorded by a sensor at a specific timestamp.
- *Time series:* a sequence of consecutive sensor measurement values within a given time range.
- *Event:* a set of time series from different sensors in the same well, collected over a closed time interval and representing a specific scenario.
- *Window interval:* the interval of consecutive timestamps of an event that are between an initial and a final timestamp.
- *Window:* the measurement values of a specific sensor corresponding to the timestamps contained in a window interval.
- *Feature point:* a vector containing all features computed from the sensor measurements at a single timestamp.
- *Level-shift anomaly:* an event representing an anomaly in which sensor measurements change from a steady-state normality baseline to a different level.
- *Dynamic Event:* also referred to as a *dynamic anomaly event*, an event representing an anomaly in which the

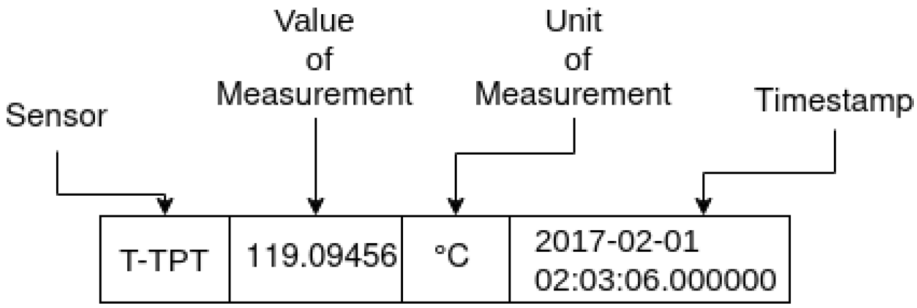

Fig. 2 Sensor measurement example (adapted from [18])

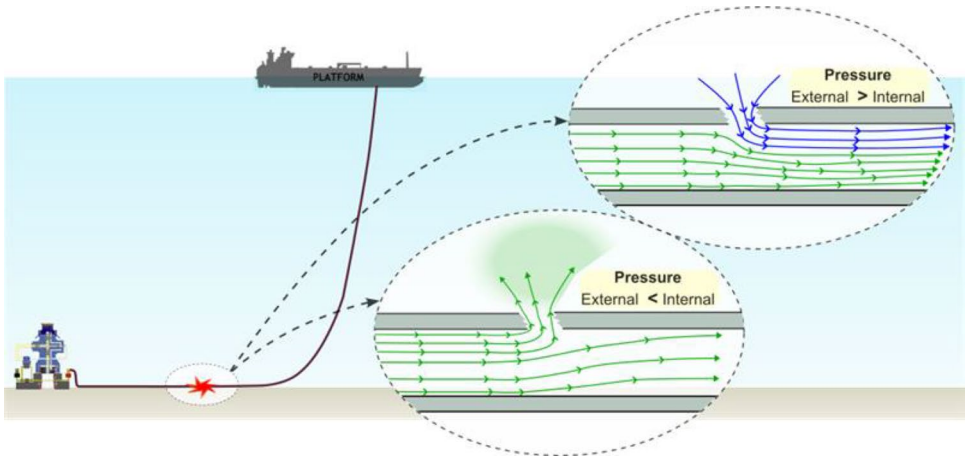

Fig. 3 Conceptual representation of a leak or an influx through a partial rupture in a subsea pipeline (adapted from 19)

sensor measurements vary continuously over time, exhibiting a pattern-based, time-dependent behavior.

Figure 2 shows an example of a sensor-provided recording, including its sensor name, measurement value, measurement unit, and timestamp.

1.4 Definition of leak and influx

Leaks result from a total or partial rupture of the pipeline. A total rupture completely separates the pipeline into two ends, causing the entire well flowrate to be diverted to the ocean (or, depending on the local pressure conditions, allowing seawater to enter the well). Although extremely harmful, this type of occurrence can be detected using simple techniques, such as monitoring the low pressure differential across the production choke at the topside. A partial rupture, on the other hand, does not sever the pipeline entirely, and the well remains connected to the platform. Detecting a partial rupture is considerably more challenging, as oil and gas continue to flow to the platform—albeit at a reduced flowrate. In general, the smaller the leakage or seawater influx, the more difficult it becomes to identify the event.

A leak in a subsea pipeline (illustrated in Fig. 3) occurs when, following a rupture or the formation of a hole, the internal pipeline pressure exceeds the local external seawater pressure. This condition decreases the pressure profile along the entire production system, as observed in [15, 16], and is reflected in reduced readings from both subsea and topside pressure sensors. The magnitude of this reduction depends on several factors, including rupture size, pressure differential, well productivity, and choke position. Temperature variations also occur and may be either positive or negative, depending on fluid properties and the heat-transfer behavior along the pipeline. While the total flowrate from the reservoir tends to increase, the flowrate at the platform always decreases.

An influx (also depicted in Fig. 3) occurs when the internal pipeline pressure is lower than the external seawater pressure. Its primary signature is a marked temperature decrease at the topside, caused by the arrival of cold seawater. An increase in pressure at the Christmas tree and a reduction in the flowrate from the reservoir are also expected. Under specific circumstances—identified through transient-flow numerical simulations—flow instabilities may arise when a heavier fluid (seawater) begins to fill the production riser. This can temporarily increase internal pipeline pressure and potentially cause the influx to revert to a leak.

In this work, we present a machine-learning-based technique for detecting partial ruptures in subsea pipelines. The system was designed to operate within a typical subsea architecture, where only pressure and temperature measurements are continuously available and no flowrate sensors are installed. It is intended specifically for rupture detection during steady-state production conditions and is not

configured to operate during transient phases or periods of operational change.

In the next sections, we describe the methodology used to develop the machine learning model, beginning with data generation and exploratory analysis, and proceeding through model training and parameter tuning. Subsequently, we present the operational metrics associated with leak detection, followed by the model assessment and a discussion on its deployment.

Finally, it is important to emphasize that this work addresses rupture detection in multiphase flow systems, which remains a relatively unexplored research area, as discussed in Sect. 1.1. Moreover, it focuses on a scenario of critical economic relevance to the Brazilian offshore industry, in which no real occurrences of leakage under the conditions assumed in this study have been recorded, and only a limited number of sensors are available. To the best of the authors' knowledge, this is the first study to combine multiphase flow simulations with machine learning techniques to develop a solution capable of detecting leaks or seawater influxes in such systems while maintaining a low false-alarm rate.

2 Methodology

This section presents the methodology used to develop, validate, and implement the partial rupture detection system. A synthetic rupture time-series dataset was generated using multiphase flow simulations. In addition, a time-series dataset containing production well events was also used.

A relative deviation feature, which is well suited for detecting level-shift events such as leaks and influxes, was selected and applied to all time series.

Subsequently, three different machine learning algorithms were used to train a leak and influx classifier, and their performances were compared. Bayesian search was employed for hyperparameter optimization to obtain the best model for each algorithm.

Then, still using the training data, an operational tuning step was performed by determining the optimal moving average of the model's probability output in order to reduce the number of false positives while maintaining high rupture detection accuracy.

Finally, the selected model was evaluated using different testing approaches with data reserved exclusively for testing.

2.1 Leak and influx data generation

The approach adopted for the present SSLDS relies on comparing a suspected anomaly, i.e., a potential rupture, with a large collection of previously cataloged leakage and influx data. Because ruptures in multiphase subsea pipelines are rare, real process data for such events in multiphase subsea pipelines are generally unavailable, making SSLDS training particularly challenging. To the authors' knowledge, no public dataset currently exists for pipeline rupture detection in multiphase flow systems, and even if one were available, it would need to precisely match the scenario described in Sect. 1.2.

To overcome this limitation, artificial data were generated through transient multiphase-flow simulations using the OLGA software (version 7.3.3). A set of simulation models representing the targeted wells was selected. For each model, a leak was introduced as a *pressure source* positioned at different points along the pipeline and set to the local seawater hydrostatic pressure. Starting from a steady-state condition without leakage, the pressure source was activated. When the hydrostatic pressure exceeded the internal pipeline pressure, seawater was injected into the flowline (influx); otherwise, the production fluid (a homogeneous mixture of oil, gas, and water) flowed out of the rupture (leak).

One important assumption in the simulation model is that the leaked mass consists of oil, gas, and water in the same proportions as found locally in the pipeline. While reasonable for bubble or slug flow regimes, this assumption may not hold for stratified flow. If a rupture occurs at the top of the pipe, leakage would predominantly consist of gas; if at the bottom, it would mainly consist of liquid. Analyzing such phase-segregated effects would require substantially more complex modeling and was deemed beyond the scope of this work.

For each well model, multiple simulations were executed using randomly sampled parameters within precribed ranges:

- *Rupture diameter:* between 0.1 and 4.0 inches (in pipelines with 6-inch internal diameter);
- *Rupture position:* two discrete locations, near the WTC or at the riser-flowline connection;
- *Well productivity index:* between 10 and 500 $Sm^3/d/bar$;
- *Reservoir pressure:* between 425 and 625 bar;
- *Watercut:* between 0 and 50%;
- *Gas-oil ratio (GOR):* between 150 and 500 Sm^3/Sm^3;
- *Production choke relative opening:* between 1 and 100%;
- *Production choke downstream pressure:* between 15 and 50 bar;
- *Well OLGA model:* nine representative well models with their respectives subsea pipelines from the pre-salt

basin. These wells have liquid flowrates typically ranging from 2000 to 7000 Sm^3/d.

After simulation, all cases exhibiting numerical issues or physically implausible behavior were discarded. This is a manual screening which involved reviewing all results and removing those with unrealistic outcomes; for example, some simulations of large ruptures showed substantial pressure differences between the pipeline interior and the external environment yet produce no leakage. Unfortunately, such anomalies are a common challenge in large-scale numerical simulation campaigns.

The resulting dataset contains time series of several variables at different measurement points. The most relevant variables for the subsequent analysis were the pressure and temperature readings from both subsea and topside instruments. The final database comprises leak events, influx events, and normal-operation periods. While the time series from the selected sensors served as input to the machine learning task, the oil, gas, and water flowrates through the rupture were recorded to label each event as leak or influx.

The final dataset consists of 3089 leak events and 1677 influx events, totaling 4766 events. Each event contains a 270-minute time series of the selected variables. All simulations begin in an initial steady-state regime without any rupture. At the 60-min mark, the rupture is formed within 10 s, initiating a leak or influx characterized by a transient phase followed by a new steady-state regime.

All data were sampled at a frequency of one per minute. This sampling rate reflects operational practice, system dynamics and strikes a balance between temporal resolution and computational efficiency. A faster sampling rate would significantly increase the computational cost of subsequent tasks while offering limited additional benefit, given that operational responses to rupture alarms (e.g., shutting in the well) typically occur on timescales much longer than one minute.

Figures 4 and 5 illustrate typical examples of leak and influx events. In both cases, the upper plot shows the evolution of the sensors P-TPT (solid blue), P-UP-PCK (dashed blue), T-TPT (solid red), and T-UP-PCK (dashed red). The primary (left) axis corresponds to the pressure variables, while the secondary (right) axis corresponds to the temperature variables. The lower plot shows the liquid flowrate at standard conditions at the rupture (blue), at the topside (orange), and in the reservoir (green). For leak events, the rupture flowrate is negative (fluid exiting the system); for influx events, it is positive (fluid entering). Standard conditions (15.5°C, 1 atm) are used for flowrate because they provide a consistent mapping to mass flowrate and are widely used in engineering practice. Liquid flowrate is expressed in Sm^3/d.

Figure 4 shows a simulated leakage event. Prior to rupture, the well produced 2086 Sm^3/d of liquid, observed uniformly at all points along the flowpath in steady state. After rupture at the 60-min mark, leakage reached 1075 Sm^3/d, increasing the reservoir liquid flowrate to 2717 Sm^3/d and reducing the topside flowrate to 1642 Sm^3/d. The difference corresponds to the leak flowrate. P-TPT, P-UP-PCK, and T-UP-PCK decreased slightly, while T-TPT exhibited a small increase.

Figure 5 shows a simulated influx event. Before rupture, the well produced approximately 2750 Sm^3/d of liquid. After rupture, 2200 Sm^3/d of seawater entered the pipeline,

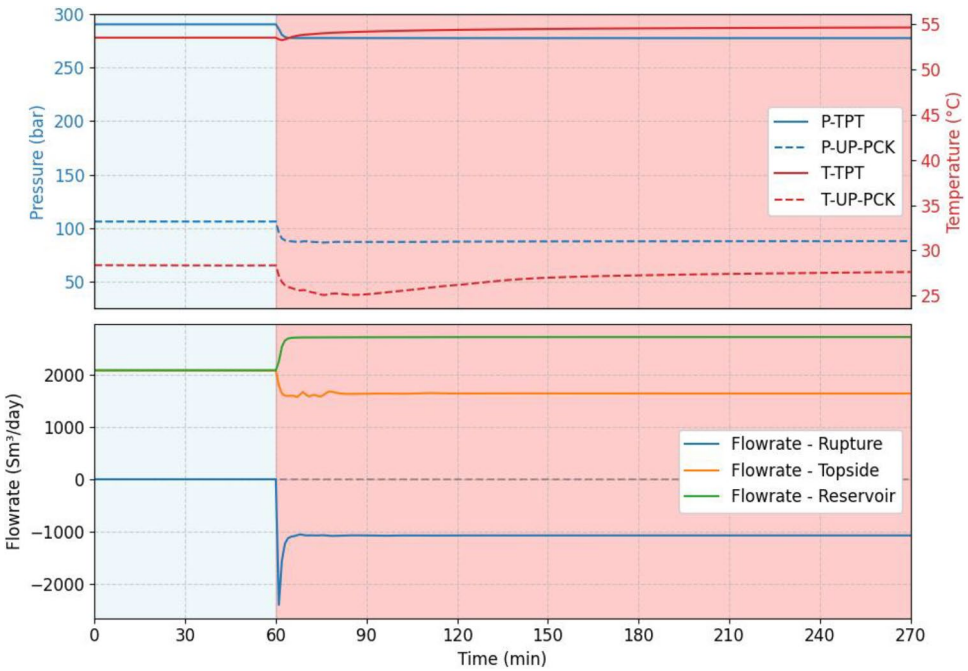

Fig. 4 Simulated leakage event after rupture formation

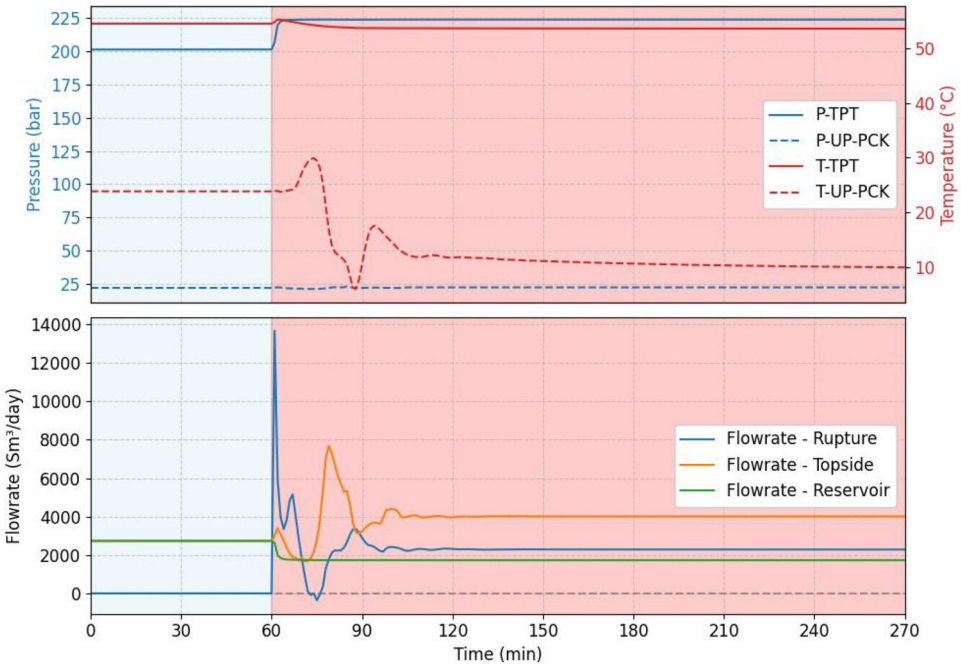

Fig. 5 Simulated seawater influx event after rupture formation

leading to a reduction in reservoir flowrate. A substantial decrease in T-UP-PCK occurred only after about 30 min—the time required for the cold seawater to reach the topside. P-TPT increased slightly, while the remaining sensors were only minimally affected.

Both events in Figs. 4 and 5 exhibit Level-Shift anomaly characteristics (see Sect. 1.3), consisting of an initial normal steady state, a transient phase triggered by the anomaly, and a final post-anomaly steady state. All simulated cases in the dataset follow this general pattern.

Occasionally, however, ruptures displayed features of a dynamic anomaly (see Sect. 1.4). In such situations, seawater influx begins, it ascends in the riser and, because the seawater is denser than the production fluids, a temporary pressure increase is observed at the rupture. This reduces the pressure differential and halts the influx, allowing leakage to occur instead. When leakage through the rupture reduces internal pressure again, influx resumes. This cycle repeats, producing alternating periods of leak and influx. Only 8 of the 4766 cases exhibited this behavior; an example is shown in Fig. 6. Although intermittent, these events were labeled as *influx* because the majority of each event exhibited seawater entry, and the sensor behavior (P-TPT, T-TPT, P-UP-PCK, T-UP-PCK) more closely resembled pure influx patterns than pure leaks. Whether such cases may occur in practice is still an open question.

2.2 Reference dataset for non-rupture events (3W dataset)

As counterexample data—i.e., events not associated with oil leaks or seawater influxes but capable of producing similar effects on the relevant variables—the authors used the 3W dataset [19], a public dataset containing instances of normal production and several undesirable operational events in oil production wells.

The events included in the 3W dataset are:

0. *Normality:* normal production, characterized by measurements fluctuating around a constant level, without large oscillations or monotonic trends.
1. *Abrupt increase of watercut:* event resulting from the arrival of the reservoir water-front at the well location.
2. *Spurious closure of DHSV:* the downhole safety valve (DHSV) may close unintentionally, shutting the well in.
3. *Severe Slugging:* a hazardous, cyclic phenomenon in the flowline-riser system characterized by alternating arrivals of large liquid slugs and gas pockets at the platform, typically occurring over minutes or hours.
4. *Instability:* alternating large volumes of liquid and gas, similar to severe slugging, but with lower intensity and poorly defined periodicity.
5. *Rapid production loss:* sharp reduction in the well productivity index (e.g., due to gravel pack collapse) or fast reservoir depletion, causing a significant flowrate decline.

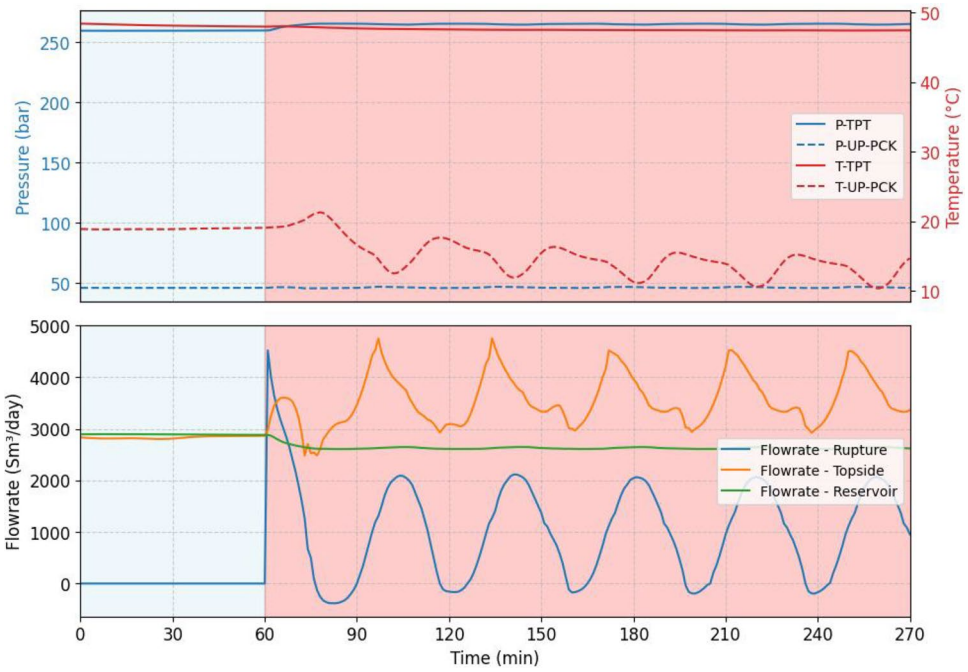

Fig. 6 Simulated seawater influx alternating with oil leakage with no steady-state condition after rupture formation

6. *Quick restriction in production choke:* rapid and unintended choke closure or restriction, potentially leading to hazardous consequences.
7. *Scaling in production choke:* inorganic deposition near the choke valve, resulting in impaired or faulty valve operation.
8. *Hydrate in production flowline:* gas hydrates formation in the pipeline, indicating risk of blockage.

Most of these events are level-shift anomalies, similar to leak and influx cases, and consist of a pre-fault steady-state phase, a transient phase, and a post-fault steady-state phase. Three events, however (0—normality; 3—severe slugging; and 4—instability), are continuous processes with constant or intermittent behavior. Severe Slugging and Instability correspond to dynamic anomalies, while Normality includes only noise and common operational variability. Although the original dataset distinguishes between transient and post-fault phases, in this work both phases were assigned the same label. Examples of each 3W event are shown in Fig. 7.

The 3W dataset is composed of real, simulated, and sketched data. Real data were collected directly from the Plant Information system of several Petrobras wells. Simulated events were generated using a multiphase flow simulator (similar to the rupture dataset described in Sect. 2.1). Sketched events were constructed using expert knowledge combined with mathematical formulations and heuristic rules. All Normality events are real, whereas most other events are simulated. Table 1 summarizes the dataset composition and includes the leak and influx events from the rupture dataset, labeled here as events 9 and 10, respectively.

The 3W dataset includes additional variables beyond those chosen in Sect. 1.2, such as PDG pressure and gas-lift measurements (injection flowrate, injection pressure, and injection temperature). As explained previously, PDG data were excluded and gas-lift wells were not considered. Thus, the events were retained, but the variables associated with those sensors were removed. Another difference is that the 3W dataset does not include upstream choke temperature measurements (T-UP-PCK), only downstream temperature. Because these values are expected to be very similar under most conditions, the downstream temperature was used as a proxy for upstream temperature.

Unlike the fully simulated rupture dataset, the 3W dataset contains many missing values (particularly in real time series), frozen sensors (i.e., readings that stop updating), and invalid values, such as pressure measurements equal to zero. All events containing missing, frozen, or invalid values were removed.

It should be noted that the rupture detection system developed in this work does not operate with missing values. All four required sensors must be functional, and no missing values exist in the rupture dataset.

2.3 Feature engineering

Because leaks and influxes manifest as level-shift anomalies, the detection methodology relies on comparing the current measurement value to a reference value that represents the expected steady-state baseline of each sensor. For every

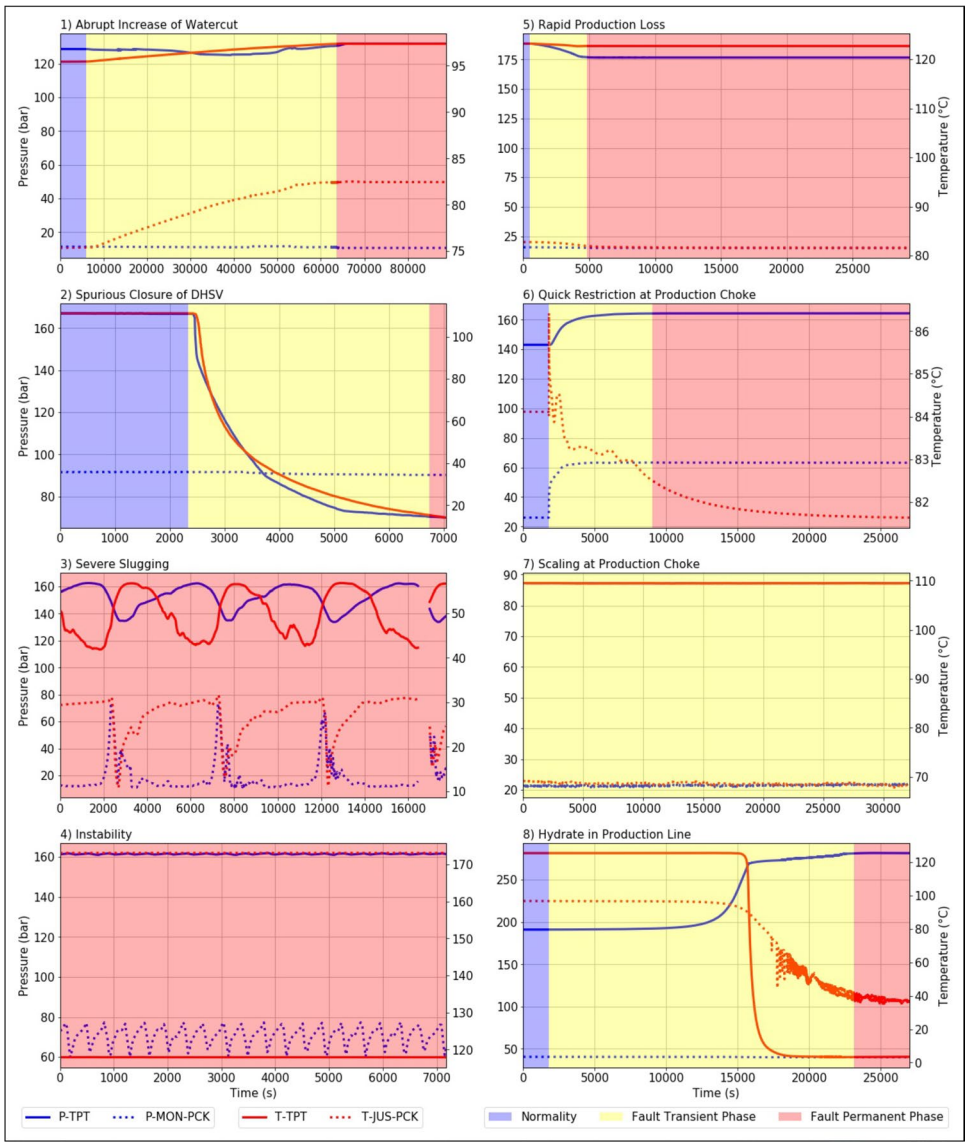

Fig. 7 Examples of 3W dataset events, adapted from [18]

time series, the following feature—referred to as the *relative deviation*—was computed:

$$x_{\mathrm{pct}}(t) = \frac{x(t) - x_{\mathrm{ref}}}{x_{\mathrm{ref}}} \times 100\%, \tag{1}$$

where $x_{\mathrm{pct}}(t)$ is the relative deviation at timestamp t, x_{ref} is the sensor's expected steady-state value under normal operation, and $x(t)$ is the current sensor measurement. For clarity, x_{ref} is termed the *reference value*, while $x(t)$ represents the *current value*. Temperature measurements (T-TPT and T-UP-PCK) were converted from degrees Celsius to Kelvin prior to feature computation.

This *steady-state deviation* approach was selected for several reasons:

- *Level-shift anomaly behavior:* Following a rupture, sensor values gradually transition to a new steady-state level. The relative deviation feature is specifically designed to capture this shift.
- *Uncertainty in rupture formation dynamics:* In the simulations, rupture formation occurs within a few seconds. In reality, ruptures may develop over minutes, hours, or even days, producing highly variable transient responses. Comparing steady-state conditions mitigates sensitivity to these uncertainties.
- *Rupture formation across shut-offs and ramp-ups:* If a rupture develops while the well is shut in and the well is later restarted under identical operating conditions (e.g., same choke opening, separator pressure, and reservoir pressure), the event remains detectable through deviations between pre-shut-off and post-ramp-up steady states.

Table 1 Composition of rupture and 3W dataset time series: event label, name, number of real instances, number of simulated instances, number of sketched instances and total count

Label	Event	Real	Simulated	Sketched	Total
0	Normality	597	0	0	597
1	Abrupt increase of watercut	5	114	10	129
2	Spurious closure of DHSV	22	16	0	38
3	Severe slugging	32	74	0	106
4	Instability	344	0	0	344
5	Rapid production loss	12	439	0	451
6	Quick restriction in production choke	6	215	0	221
7	Scaling in production choke	4	0	10	14
8	Hydrate in production line	3	81	0	84
9	Oil leak—subsea pipeline rupture	0	3089	0	3089
10	Seawater influx—subsea pipeline rupture	0	1677	0	1677
–	Total	1025	5705	20	6750

Table adapted from [18]

- *Extended detection window:* A rupture is expected to shift all relevant variables to new steady-state levels. Thus, detection remains possible even long after the transient phase has ended—something not guaranteed if detection relied solely on transient behavior.

An alternative approach would be to analyze the rate of change of each sensor, identifying ruptures through transient dynamics rather than steady-state comparisons. However, this strategy suffers from important limitations: (i) simulations would need to incorporate multiple rupture-formation profiles to cover the full range of possible rates of change; (ii) rupture formation during shut-ins or ramp-ups could not be detected; and (iii) the detection window would be restricted to the transient period.

The main drawback of the steady-state comparison approach is the need to determine the reference value (x_{ref} in Eq. 1), which is not always straightforward in real-time operation. The challenges associated with reference-value estimation are discussed in Sect. 3.3. In this study, the reference value for each event was extracted directly from its initial measurements. For the rupture dataset, the reference value was defined as the average sensor measurement over the 30 min preceding the rupture. For the 3W dataset, the reference value was taken as the steady-state average preceding the fault for level-shift anomalies. For normality and dynamic anomaly events (instability and severe slugging), the average over the entire time series was used.

2.4 Exploratory data analysis

An exploratory data analysis was carried out by computing the relative deviation (using Eq. 1) between the final steady state (after rupture formation) and the initial steady state (before rupture formation). The following measurement values, described in Sect. 1.2, were used: pressure at the WCT (P-TPT), temperature at the WCT (T-TPT), pressure upstream of the production choke (P-UP-PCK), and temperature upstream of the production choke (T-UP-PCK). For each leak or influx event, a 4D feature point was formed from the relative deviations of these four sensors.

It should be noted that relative deviations could also be computed at all timestamps after rupture formation (as performed in the training phase, see Sect. 2.6). However, for the purpose of exploratory analysis, doing so would produce an unnecessarily large number of points. Our aim here is specifically to characterize the *final* relative deviation—after the system reaches its new steady-state level—and thereby understand the general behavior of rupture events predicted by the multiphase-flow simulations.

Figure 8 presents all possible two-dimensional scatterplots of the computed relative deviations. Each point corresponds to a single leak or influx event. A visual inspection of the figure shows that oil leaks (red points) and seawater influxes (blue points) occupy distinct regions with very limited overlap, suggesting that the four proposed relative deviations are sufficient to discriminate between the two types of events.

These results are consistent with the physical interpretations discussed in Sect. 1.4. Oil leaks decrease pressure in both sensors (P-TPT and P-UP-PCK) and typically lead to lower topside temperature (T-UP-PCK). However, as previously mentioned, the WCT temperature (T-TPT) exhibits mixed behavior: depending on the case, it may increase, decrease, or remain approximately constant.

Conversely, seawater influx produces a pronounced decrease in T-UP-PCK, with some cases reaching relative deviations of approximately -8% (a significant change in Kelvin units; for instance, -8% of a 50 °C reference value corresponds to a temperature drop of about 25 °C). Influx events also tend to increase WCT pressure (P-TPT) and, in most cases, decrease WCT temperature (T-TPT), while the behavior of P-UP-PCK is more varied.

2.5 Noise data augmentation

To make the simulated data more representative of real operational conditions, random noise was added to all simulated events, including both the rupture dataset and the simulated cases from the 3W dataset. This noise was derived from a normal-production event after applying the feature

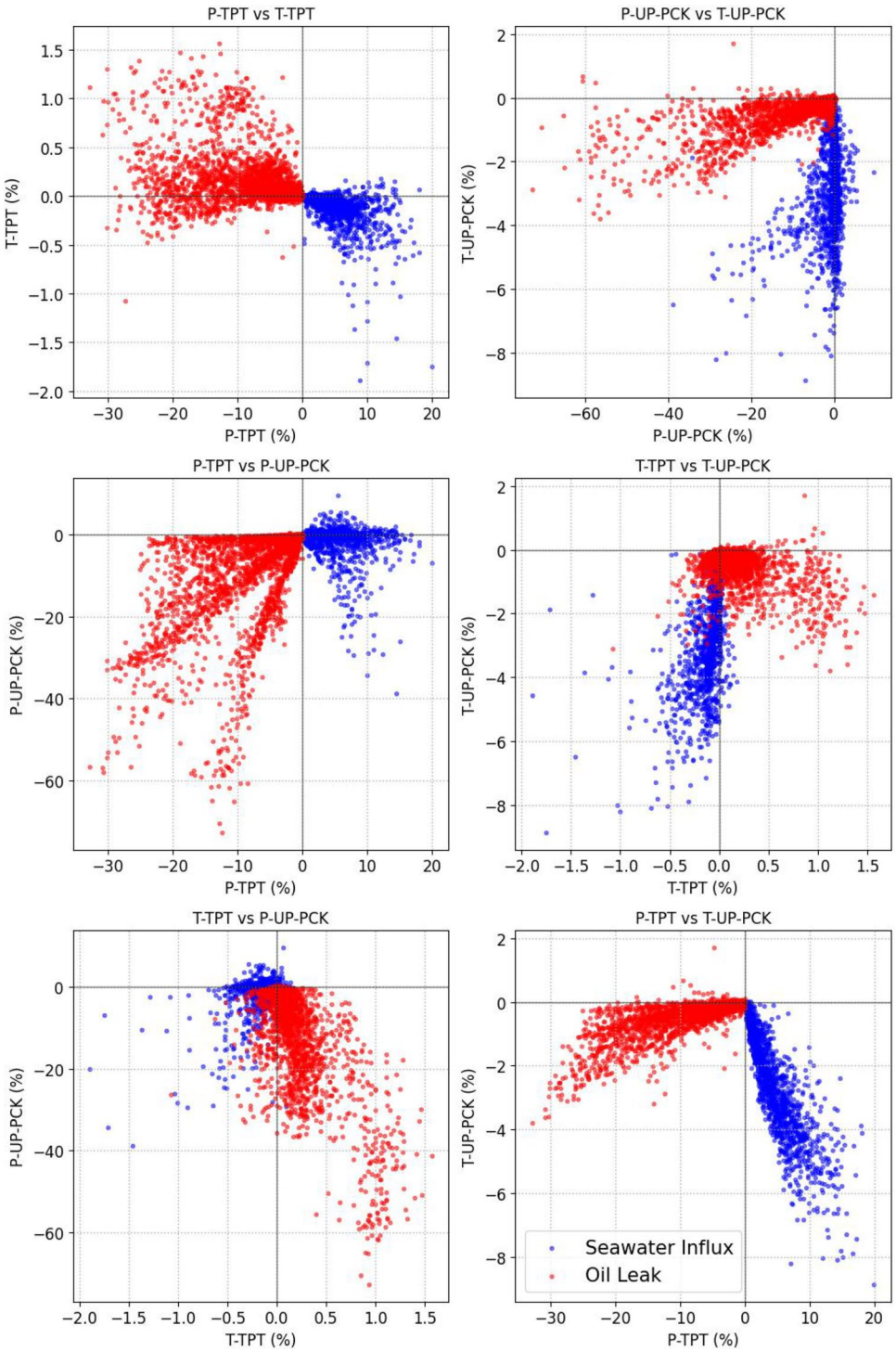

Fig. 8 Relative deviations between the steady states before and after rupture formation for all leak and seawater influx events

transformation in Eq. 1 to all timestamps. To apply the transformation, the average of each sensor during normality was used as the reference value, and the covariance matrix of the resulting relative-deviation time series was computed. The obtained covariance matrix is shown below:

$$\begin{bmatrix} 0.1771 & 0.0021 & 0.0915 & 0.0073 \\ 0.0021 & 0.0038 & 0.0003 & -0.0003 \\ 0.0915 & 0.0003 & 0.2476 & 0.0089 \\ 0.0073 & -0.0003 & 0.0089 & 0.0020 \end{bmatrix}, \tag{2}$$

where the first, second, third, and fourth rows and columns correspond to P-UP-PCK, T-UP-PCK, P-TPT, and T-TPT, respectively. This noise was added directly to the relative

deviation values, as described in Sect. 2.6, and it was also used to generate additional normality data points.

The addition of noise improved detection robustness in normality scenarios. Although the 3W dataset already contains noise, its statistical characteristics may differ from those of the wells considered in this study. Furthermore, the magnitude and structure of measurement noise are not constant; they depend on the intrinsic precision of each sensor and on well-specific flow properties. Incorporating noise into the simulated dataset therefore makes the training conditions more realistic and enhances the generalization capability of the detection model.

2.6 Model training

The events from both datasets (simulated ruptures and 3W) were first divided into training, validation, and test sets using a 60%, 20%, and 20% split, respectively, while maintaining class balance. The training set was used for model fitting and hyperparameter optimization. The validation set was used for model comparison and selection. The test set remained completely isolated during development, serving exclusively for the final performance assessment of the chosen model.

The detection system was designed to be as simple as possible. The goal of this work is to construct a reliable system capable of detecting leak and influx events with high accuracy, while avoiding unnecessary complexity or excessive input features. Accordingly, the probability of leak or influx at each timestamp was estimated using only the four sensors described in Sect. 1.2 (P-TPT, P-UP-PCK, T-TPT, and T-UP-PCK). At the same time, it is desirable that both classifier inputs and outputs can be adjusted by an operator to reduce false alarms. This operational aspect is discussed further in Sect. 3.3.

The following machine learning algorithms were compared: support vector machine (SVM) [20], random forest (RF) [21], and LightGBM (LGBM) [22]. All of these techniques are suited for classification tasks and can be configured to output class probabilities from 0 to 1, which is essential because it allows operators to adjust alarm thresholds based on sensor quality and well-specific operational conditions.

Table 2 Comparison of two data preprocessing methodologies

Aspect	Methodology A	Methodology B
Algorithm	SVM, Random forest	LightGBM
Treatment of level-shift events	One feature point per event	One feature point per timestamp
Treatment of normality and dynamic events	One feature point every 20 min	One feature point per timestamp
Feature matrix size	Smaller	Larger
Noise injection	Generates additional noise points	Adds noise to simulated or sketched events

Two different preprocessing methodologies were employed. Methodology A was used for SVM and Random Forest, while Methodology B was used for LightGBM.

Methodology A constructs one feature point per level-shift event by extracting its steady-state relative deviation (Sect. 2.3). The purpose is to capture the most relevant information of each event, namely the variation between steady states. For events without a well-defined post-fault steady state (Normality, Instability, Severe Slugging), feature points were sampled every 20 min. Additional synthetic "normal" feature points were generated by sampling from a Gaussian distribution using the covariance matrix (Eq. 2), with the number of noise-generated points matching the number of normal points already present. After concatenation, the resulting feature matrix is:

$$X_{\text{pct}} = \begin{bmatrix} x_{\text{pct}}^{(\text{P-TPT},1)} & x_{\text{pct}}^{(\text{P-UP-PCK},1)} & x_{\text{pct}}^{(\text{T-TPT},1)} & x_{\text{pct}}^{(\text{T-UP-PCK},1)} \\ x_{\text{pct}}^{(\text{P-TPT},2)} & x_{\text{pct}}^{(\text{P-UP-PCK},2)} & x_{\text{pct}}^{(\text{T-TPT},2)} & x_{\text{pct}}^{(\text{T-UP-PCK},2)} \\ \vdots & \vdots & \vdots & \vdots \\ x_{\text{pct}}^{(\text{P-TPT},M)} & x_{\text{pct}}^{(\text{P-UP-PCK},M)} & x_{\text{pct}}^{(\text{T-TPT},M)} & x_{\text{pct}}^{(\text{T-UP-PCK},M)} \end{bmatrix} \in \mathbb{R}^{M\times 4},$$

where each row corresponds to a feature point from a level-shift event, a sampled timestamp from a dynamic event, or a noise-generated sample. The total number of rows M is the sum of all such points.

Methodology B computes feature points for every timestamp in all events, labeling points from leak and influx events as normal when sampled before rupture initiation. In intermittent events (e.g., Fig. 6), all timestamps were labeled as influx, as justified in Sect. 2.1. After computing all feature points, Gaussian noise following the covariance matrix in Eq. 2 was added to all points originating from simulated or sketched events. The resulting feature matrix X_{pct} has one row per timestamp across all events.

Both methodologies use the reference-value approach described in Sect. 2.3. Table 2 summarizes the main differences.

A separate preprocessing approach was used for LightGBM because of its computational efficiency: it is able to handle large matrices extremely quickly, whereas SVM and Random Forest become significantly slower as matrix size grows. For SVM, which is distance-based, Z-scaling [23] was applied so that pressure and temperature deviations—represented in different magnitudes due to Kelvin conversion—contribute comparably to the distance metric.

Hyperparameter tuning for each algorithm was conducted using Bayesian optimization [24], where the score for each iteration was the sample-weighted binary cross-entropy loss [25] computed via 5-fold cross-validation on the training set. Bayesian search uses probabilistic inference to guide

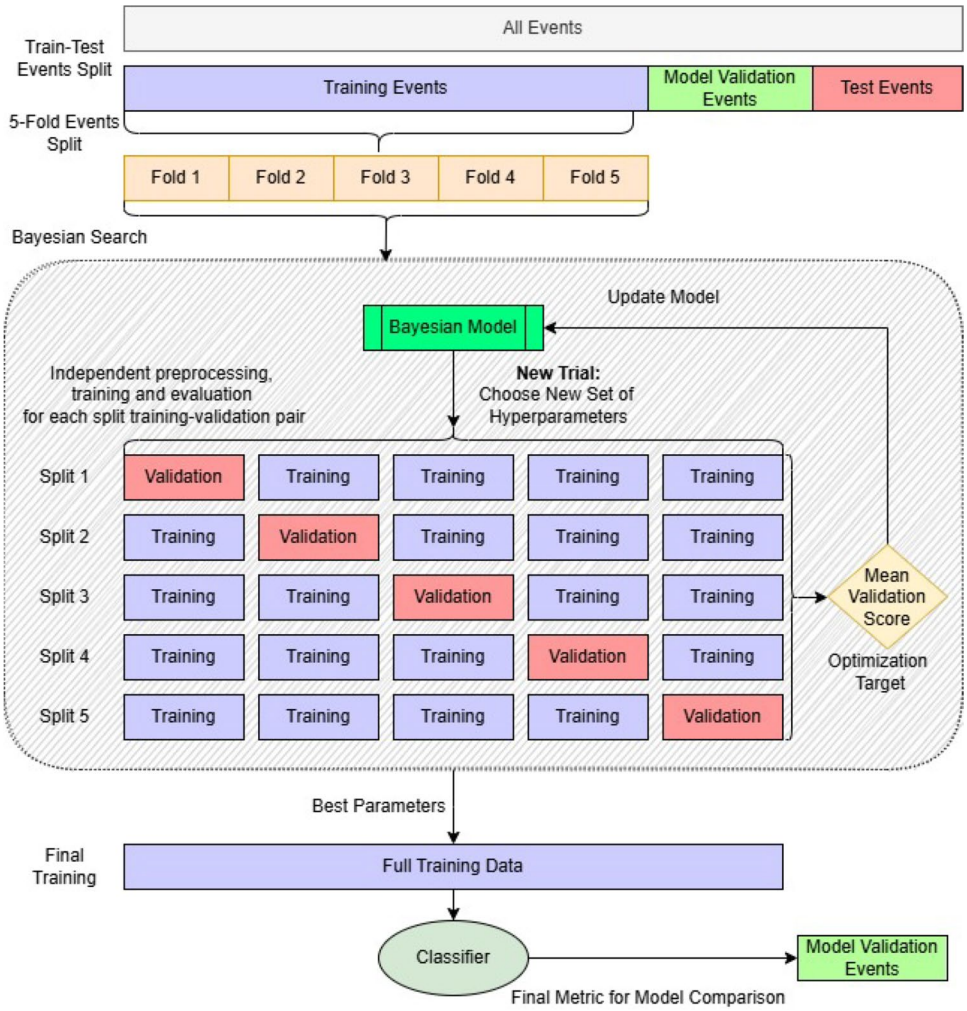

Fig. 9 Model development workflow (adapted from scikit-learn documentation, Sect. 3.1)

Table 3 Hyperparameter search space used in Bayesian optimization. Ranges indicate sampling intervals; distributions indicate how values were sampled

Algorithm	Parameter	Min	Max	Distribution
SVM (RBF)	C	10^{-2}	10^{5}	Log-uniform
	γ	10^{-4}	1	Log-uniform
Random forest	Num. estimators	50	500	Uniform (int)
	Max depth	2	50	Uniform (int)
	Min samples split	2	20	Uniform (int)
	Min samples leaf	1	20	Uniform (int)
	Max features	{2,3,4}	–	Categorical
LightGBM	Num. Leaves	4	128	Log-uniform (int)
	Learning rate	10^{-4}	10^{-1}	Log-uniform
	Min child weight	10^{-3}	20	Log-uniform
	λ_1	10^{-8}	10	Log-uniform
	λ_2	10^{-8}	100	Log-uniform
	Feature fraction	0.5	1.0	Uniform
	Bagging fraction	0.5	1.0	Uniform
	Bagging frequency	1	7	Uniform (int)

exploration of the hyperparameter space. The 5-fold cross-validation ensured that each subset of training events was used exactly once as a validation fold. After 200 optimization iterations, the best hyperparameters were selected, and a final model was trained using the full training set. Figure 9 illustrates the model development workflow. The complete hyperparameter search space is provided in Table 3.

All metric scores used during training, validation, and model selection were sample-weighted. The weighting scheme assigned 45% of the score to normal points, 45% to rupture points (leak or influx, depending on the classifier), and the remaining 10% to points from all other events, distributed proportionally. This approach ensures equal importance for normal and rupture classifications, with reduced influence from rarer events whose behavior may overlap with rupture patterns but are of secondary importance for leak detection.

For final model selection, Methodology B was applied to the validation events to generate a feature matrix $X_{\text{pct}}^{\text{val}}$.

Table 4 Comparison of machine learning models for leak detection (weighted metrics)

Model	Accuracy	Recall	Precision	F1-score
SVM	0.871	0.838	0.888	0.862
Random forest	0.855	0.916	0.751	0.825
LightGBM	0.901	0.919	0.858	0.888

Table 5 Comparison of machine learning models for seawater influx detection (weighted metrics)

Model	Accuracy	Recall	Precision	F1-score
SVM	0.931	0.936	0.910	0.923
Random forest	0.924	0.962	0.866	0.912
LightGBM	0.956	0.953	0.950	0.951

The performance of each classifier for leak and influx detection is reported in Tables 4 and 5. In both tasks, LightGBM consistently outperformed SVM and Random Forest across nearly all metrics and was therefore selected as the model for both classifiers.

One may ask why Methodology B was used to validate all classifiers. The rationale is that, after noise augmentation, Methodology B preserves more event information—not only steady-state deviations but also temporal variability—making it more representative of the scenario encountered during model deployment. Although LightGBM may appear to benefit from using the same preprocessing strategy in both training and validation, we attribute its superior performance primarily to its ability to efficiently process large feature matrices, a capability not shared by SVM and Random Forest.

2.7 Operational parameter tuning

New classifiers were trained using the LightGBM algorithm with the combined training and validation events. However, an issue emerged when evaluating classifier performance on an event-wise basis: both probability outputs frequently exhibited isolated peaks above 0.5 during noisy normal operation. This behavior is illustrated in Fig. 10. Although the leak probability correctly jumps to 1.0 immediately after the rupture at the 60-minute mark, several spurious peaks appear beforehand. Performance on the 3W dataset reinforced this observation: at least one false alarm occurred in 48.8% of purely normal events for the leak classifier and 40% for the influx classifier.

In long-term operation, excessive daily false alarms pose a critical challenge, as they diminish user trust and compromise system usefulness. At the same time, it is essential to preserve the classifier's ability to detect ruptures reliably. A fundamental limitation of the machine-learning model is that it operates on individual timestamp samples, optimizing its ability to correctly classify each individual leak or non-leak timestamp. In practical deployment, however, the main requirement is that an alarm must trigger at least once during a rupture event. Although detection time is also important—particularly for subsea petroleum leak scenarios—sacrificing a few minutes of response time in exchange for fewer false alarms is often an acceptable trade-off.

Given these considerations, an additional parameter-tuning step was performed using the training and validation sets, now evaluating metrics at the *event* level rather than at the timestamp level. An initial attempt involved applying a moving average directly to the raw pressure and temperature data before feeding them into the classifier. However, this approach increased the number of false alarms in the 3W dataset. We attribute this effect to outlier measurements, which can distort averaged sensor values and unintentionally resemble small leak signatures.

Instead, a moving average was applied to the classifier's probability output. The results for leak and influx detection are presented in Tables 6 and 7, respectively. For each moving-average window size N (in minutes), the following metrics were computed:

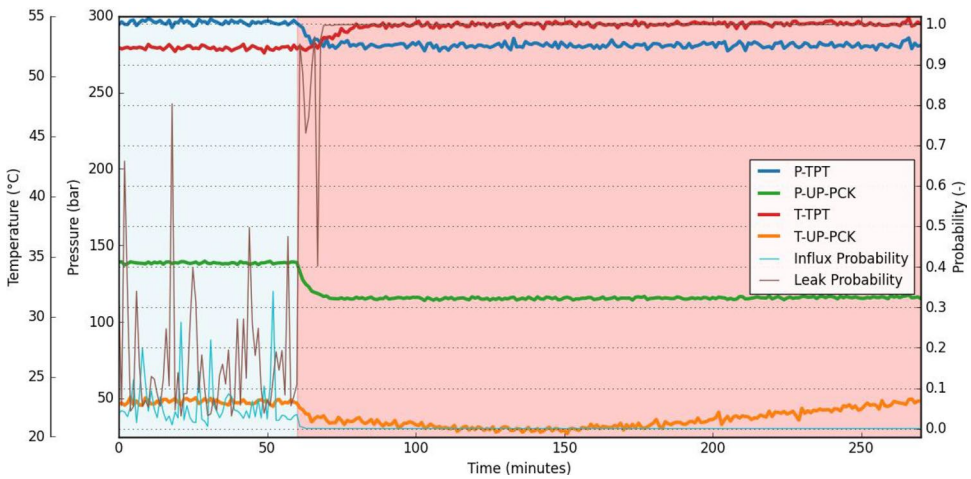

Fig. 10 Classifiers performance in a leak training event

Table 6 Leak detection event-based summary for different moving-average windows N

N	Alarms after rupture per event (%)	Alarms before rupture per event (%)	Detection time (min)	Normality false alarms per window (%)	Normality false alarms per event (%)
1	100.0	89.0	3.3 ± 3.7	2.86	48.8
5	97.7	1.5	8.9 ± 16.6	0.94	24.3
10	95.2	0.2	11.4 ± 15.9	0.73	14.7
15	93.9	0.0	13.4 ± 13.7	0.63	10.9
30	92.6	0.0	20.7 ± 14.3	0.30	4.4

Table 7 Seawater influx detection event-based summary for different moving-average windows N

N	Alarms after rupture per event (%)	Alarms before rupture per event (%)	Detection time (min)	Normality false alarms per window (%)	Normality false alarms per event (%)
1	100.0	40.1	12.6 ± 11.3	0.80	40.0
5	98.3	0.9	17.7 ± 13.8	0.05	1.9
10	98.2	0.4	23.0 ± 17.2	0.04	0.8
15	98.0	0.1	25.9 ± 18.2	0.02	0.2
30	97.5	0.0	33.6 ± 18.3	0.01	0.2

- the percentage of leak or influx events correctly detected after the rupture,
- the percentage of false alarms before rupture initiation,
- the mean and standard deviation of the detection time,
- the false-alarm rate for normality events, both per 10-min window and per event.

Based on these results, $N = 10$ min was selected for both alarms. This window size preserved high detection performance while substantially reducing false positives, achieving a balanced compromise between sensitivity and robustness.

It is important to note that this parameter can be adjusted for individual wells. For wells exhibiting high instability or noise, operators may increase N or modify additional parameters—such as the probability threshold or the on-delay time (the minimum duration the probability must remain above threshold before triggering an alarm). Conversely, for wells with very stable behavior and low measurement noise, smaller window sizes may be appropriate.

3 Test and assessment

After completing all the procedures described in the previous subsection, a final pair of LightGBM classifiers—one for leak detection and one for influx detection—was trained using the combined training and model-validation events. These classifiers were then evaluated using the test events, which had not been used at any stage of training, hyperparameter tuning, or model selection.

The results of these evaluations are presented in the following subsections. The first set of results is based on individual timestamp samples: each measurement set from the four sensors of every test event was passed to both classifiers, and a 10-min moving average was applied to the resulting probability outputs. For this reason, this part of the evaluation is referred to as the *window-based assessment*, as each prediction corresponds to a 10-min window.

The second set of results is based on entire events, i.e., the full time series representing a well-production episode. Accordingly, this part is termed the *event-based assessment*. In this framework, a detection—whether true or false positive—is considered to have occurred if the classifier's probability output exceeds a chosen threshold within any 10-min window. Two thresholds were analyzed: 0.5 and 0.8. For rupture events, detection time and leaked liquid volume prior to detection were also evaluated.

For all tests, the raw pressure and temperature measurements were transformed into relative deviations using Eq. 1. Reference values for this transformation were estimated using the methodology described in Sect. 2.3, and applied consistently to both rupture and 3W events. Each resulting feature point (i.e., each timestamp's vector of relative deviations) was then passed to the leak and influx classifiers, which return probability values. A 10-minute moving average was subsequently applied to these probabilities to yield the smoothed outputs used in the analysis.

In all assessments, normality results refer exclusively to the normality events of the 3W dataset. Normality periods preceding faults in other events were excluded, since the primary interest is ensuring that no false alarms occur before rupture formation in leak and influx cases.

Table 8 Percentage of 10-min window samples for which the leak and influx classifiers exceeded the probability thresholds

Event	Leak classifier		Influx classifier	
	$p > 0.5$	$p > 0.8$	$p > 0.5$	$p > 0.8$
0—normality	1.24	0.06	0.04	0.00
1—abrupt increase of watercut	1.77	0.36	1.67	0.70
2—spurious closure of DHSV	0.00	0.00	0.00	0.00
3—severe slugging	10.87	1.22	8.88	1.01
4—flow instability	4.25	1.09	0.45	0.00
5—rapid production loss	27.26	3.37	1.00	0.43
6—quick restriction in PCK	0.00	0.00	0.00	0.00
7—scaling in production choke	0.00	0.00	0.00	0.00
8—hydrate in production line	0.00	0.00	6.62	0.00
9—oil leak	85.46	71.28	0.18	0.02
10—seawater influx	0.00	0.00	85.49	78.44

3.1 Performance assessment: window-based metrics

This subsection presents results in which each 10-min window sample—corresponding to the moving-averaged probability outputs of the leak and influx classifiers—is evaluated independently.

Table 8 reports the percentage of window samples for which the leak or influx probability exceeded a given threshold. For a threshold of 0.5, the leak classifier correctly detected 85.4% of all window samples labeled as leak, while the influx classifier correctly detected 85.5% of all window samples labeled as influx. When the threshold was increased to 0.8, these values decreased to 71.3% and 78.4%, respectively, as expected for a more conservative alarm criterion.

Both classifiers exhibited very low false-positive rates for normality samples. The influx classifier produced 0.04% and 0.00% false positives for thresholds of 0.5 and 0.8, respectively. The leak classifier produced 1.24% and 0.06% false positives for the same thresholds.

The most challenging event for the leak classifier was *rapid production loss*, for which 27.26% of window samples were incorrectly flagged as leaks at a threshold of 0.5; this dropped to 3.37% at a threshold of 0.8. Both classifiers also produced relatively high false-positive rates in the *severe slugging* event: at a threshold of 0.5, 10.87% and 8.88% of window samples were incorrectly classified as leak and influx, respectively. The influx classifier also showed a higher false-positive rate for the *hydrate in production line* event (6.62% at threshold 0.5), likely due to its strong dynamic variability.

3.2 Performance assessment: event-based metrics

The event-based assessment differs from the window-based analysis presented earlier. Here, performance is evaluated on a per-event basis: an event is considered *detected* if, at any timestamp, the moving-averaged probability output of the classifier exceeds the chosen alarm threshold. In contrast, the window-based assessment examines each 10-min window independently.

Table 9 Event detection rate (%) of leak and influx classifiers

Event	Leak classifier		Influx classifier	
	$p > 0.5$	$p > 0.8$	$p > 0.5$	$p > 0.8$
0—normality	24.7	3.1	1.0	0.0
1—abrupt increase of watercut	66.7	20.8	29.2	20.8
2—spurious closure of DHSV	0.0	0.0	0.0	0.0
3—severe slugging	57.1	28.6	76.2	33.3
4—flow instability	22.7	9.1	6.8	0.0
5—rapid production loss	94.4	51.1	2.2	4.4
6—quick restriction in PCK	0.0	0.0	0.0	0.0
7—scaling in production choke	0.0	0.0	0.0	0.0
8—hydrate in production line	0.0	0.0	68.7	0.0
9—oil leak	94.8	85.8	0.5	0.4
10—seawater influx	0.0	0.0	97.9	95.8

Table 9 presents the percentage of events for which at least one leak or influx probability exceeded the threshold. At a threshold of 0.5, the leak classifier detected almost 95% of all oil leak events. However, it also triggered false alarms in 24.7% of pure normality events. Increasing the threshold to 0.8 reduced false alarms in normality to 3.1%, but detection of leak events also fell to 85.8%. The leak classifier additionally showed higher false-positive rates for events 1, 3, 4, and 5 (abrupt increase of watercut, severe slugging, flow instability, and rapid production loss).

The influx classifier achieved superior performance: even at a threshold of 0.8, it detected almost 96% of all seawater influx events. False positives were mainly associated with events 3 and 8 (severe slugging and hydrate formation), but performance on normality events was excellent, with zero false alarms at the higher threshold.

An important additional observation (not shown in Table 9) is that, for a threshold of 0.5, *no leak or influx event generated an alarm before rupture formation*. This indicates that, for all rupture scenarios evaluated, the classifier triggered only after the anomaly began, never prematurely.

The relatively high false-positive rates observed for certain other fault events do not compromise the purpose of the rupture detection system. As explained in Sect. 2.6, lower weighting was assigned to these events during training because they are relatively infrequent. Flow instability—and, to a lesser extent, severe slugging—constitutes the most challenging class for the leak classifier due to its continuous, highly variable nature. This challenge can be mitigated operationally by adjusting detection parameters (moving-average window, on-delay time, alarm threshold)

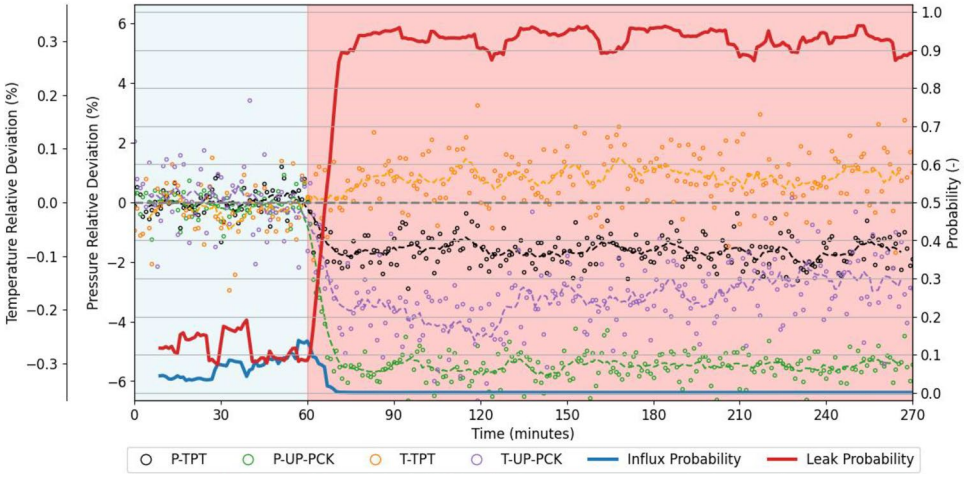

Fig. 11 Leak test case (leak flowrate of 471 Sm^3/d in a well producing 3855 Sm^3/d before the fault)

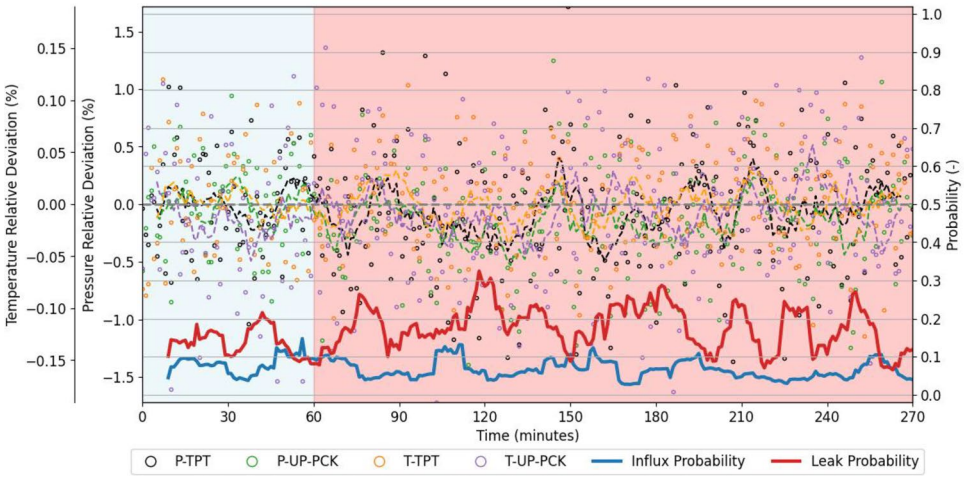

Fig. 12 Small leak test case (leak flowrate of 17 Sm^3/d in a well producing 3968 Sm^3/d before the fault)

or by retraining classifiers with increased weighting for these cases.

Other fault events—such as abrupt increases in watercut, rapid production loss, or hydrate formation—are level-shift anomalies that occur only occasionally. They may trigger isolated false alarms, but these occurrences could be handled by implementing dedicated alarms for such events, preventing unnecessary activation of the leak or influx alarms.

The next figures illustrate classifier performance on full events. Circles represent the relative deviation features (Eq. 1) for each sensor: black, green, orange, and purple correspond to P-TPT, P-UP-PCK, T-TPT, and T-UP-PCK, respectively. Dashed lines show the 10-min moving averages of these deviations—these averages were plotted only to help visualize the trend of each sensor. Solid red and blue lines represent the moving-averaged leak and influx probabilities.

Figure 11 shows a leak test case with a leak flowrate of 471 Sm^3/d in a well producing 3855 Sm^3/d before rupture. Before the 60-min rupture, relative deviations fluctuate around zero, and probabilities stay near 0.1, due to noise injection. After rupture, P-TPT, P-UP-PCK, and T-UP-PCK decrease, while T-TPT increases. The leak probability spikes rapidly above 0.9, while the influx probability approaches zero.

Figure 12 shows a small leak (17 Sm^3/d) in a well producing 3968 Sm^3/d. The relative deviations barely change, and the leak probability never exceeds 0.3. Such small leaks are extremely difficult to detect: the natural noise magnitude exceeds the physical effect of the rupture. Probabilities at this level could only be considered meaningful in exceptionally stable wells, under conditions with minimal sensor noise and negligible production oscillations—conditions that are rarely encountered in practice.

Figure 13 shows a large leak (3370 Sm^3/d) in a well producing 1600 Sm^3/d. This scenario reflects a reservoir with high production capacity combined with a very restricted production choke. Although the alarm may be triggered immediately, the sensor measurements subsequently move so far from their pre-rupture values that the resulting feature point lies outside the region covered by most simulated cases. This behavior is qualitatively similar to that shown

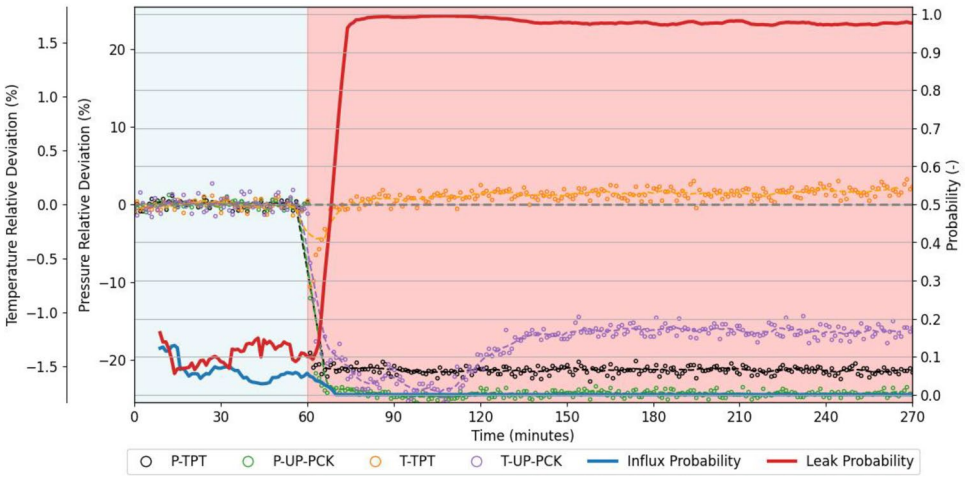

Fig. 13 Large leak test case (leak flowrate of 3370 Sm^3/d in a well producing 1600 Sm^3/d before the fault)

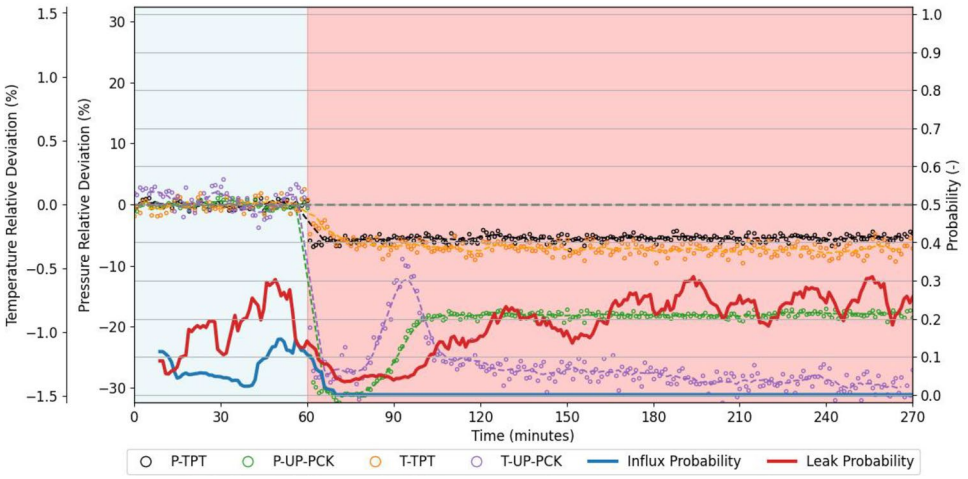

Fig. 14 Failed leak detection (leak flowrate of 2103 Sm^3/d in a well 1857 Sm^3/d before the fault)

in Fig. 11, but with much larger pressure deviations. In this particular case, the leak probability still spikes to nearly 1.0, ensuring correct detection despite the extreme transient response.

Despite the improvements introduced in the current version of the leak classifier, two large-leak test events remained undetected. This issue had already been observed in earlier versions of the classifier, where extremely large leaks were sometimes missed because their feature points lay far outside the distribution of most simulated cases.

Figure 14 shows the first undetected case: a liquid leak flowrate of 2103 Sm^3/d in a well producing 1857 Sm^3/d before rupture. In this event, both pressure measurements dropped sharply, yet the leak probability never exceeded 0.5, reaching only about 0.3.

Figure 15 shows the second case, even more extreme, with a leak flowrate of 4005 Sm^3/d in a well producing 1733 Sm^3/d before rupture. All sensors exhibited very large relative deviations, but the leak probability unexpectedly decreased rather than rising.

In both events, the final steady-state feature points were clearly distant from the main leak cluster. This can also be observed in Figure 8, where a few points lie detached from the principal distribution. We believe the root cause is an atypical thermal response: T-TPT decreased significantly (− 0.5% in the first case and − 1% in the second), despite the very large leak flowrates—larger than the original well production. In such scenarios, the choke is highly restricted and heat transfer from the reservoir would normally dominate WCT temperature. However, as explained in Sect. 1.4, depending on fluid properties, depressurization effects may dominate instead, causing cooling and producing an unusual temperature signature not represented in the simulated training data.

Although these failures occurred in only two cases, they highlight an important limitation of machine-learning-based detection: if the real event differs substantially from the training data, the classifier may fail to detect it. Thus, for very large leaks, simpler safeguards (e.g., absolute pressure thresholds) should complement the ML-based system.

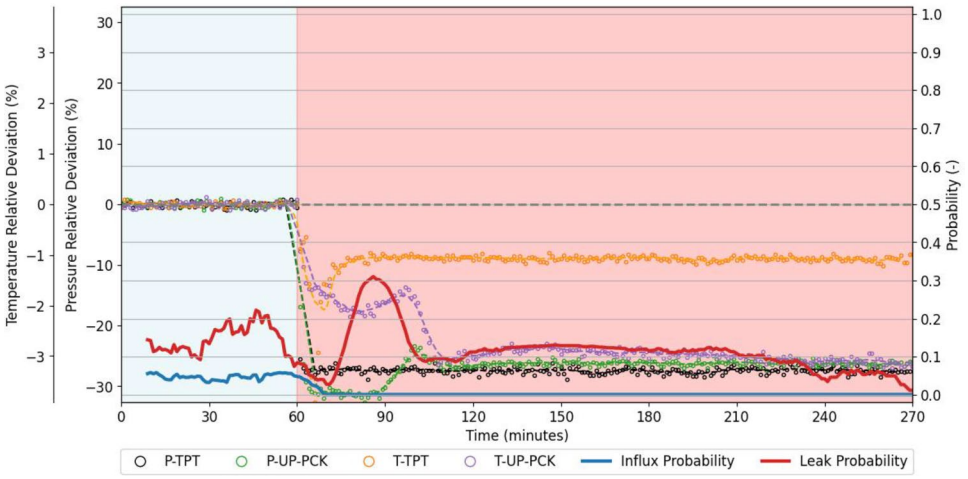

Fig. 15 Failed leak detection (leak flowrate of 4005 Sm^3/d in a well 1733 Sm^3/d before the fault)

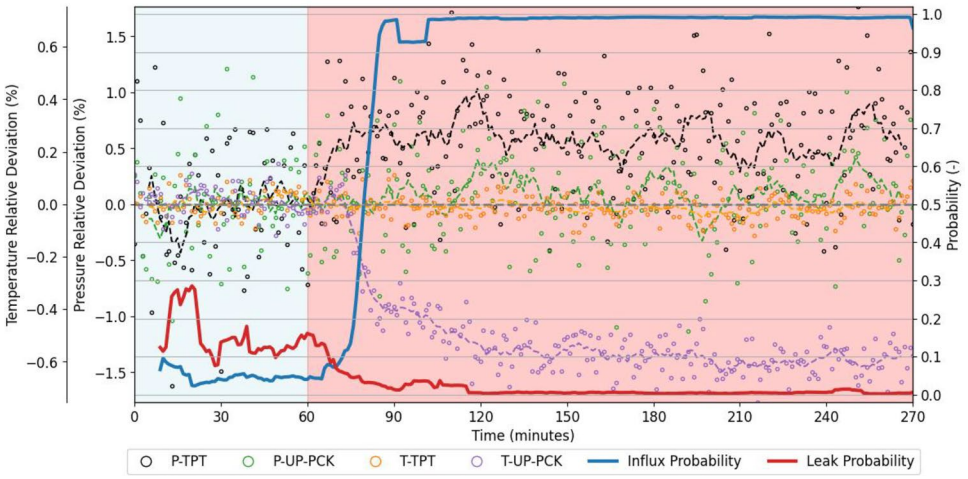

Fig. 16 Seawater influx test case (influx flowrate of 171 Sm^3/d in a well 4288 Sm^3/d before the fault)

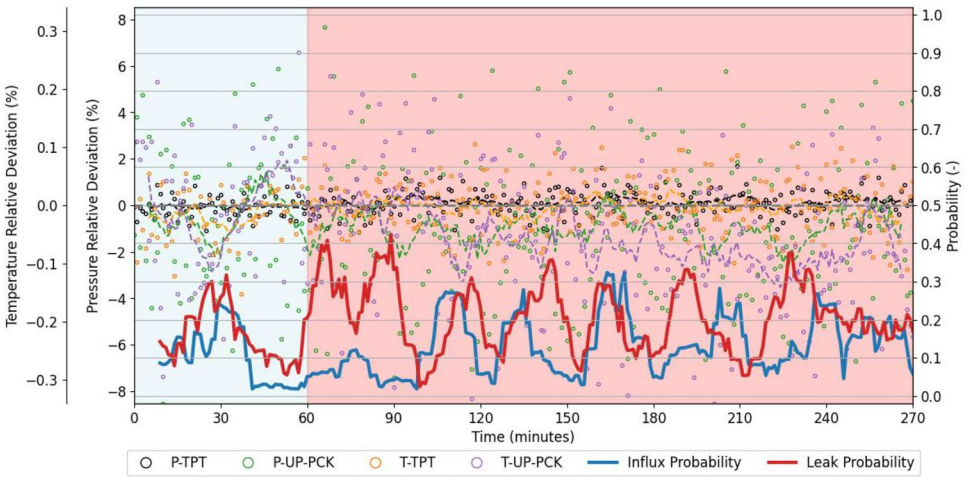

Fig. 17 Seawater influx test case (influx flowrate of 24 Sm^3/d in a well 4026 Sm^3/d before the fault)

These simpler alarms are unsuitable for small leaks—where ML excels—but remain essential as backup mechanisms.

The influx classifier showed consistently strong performance. Figure 16 shows a case with seawater influx of 171 Sm^3/d; detection takes longer than leak detection because T-UP-PCK only changes once cold seawater reaches the topside.

Small influxes are similarly difficult to detect: Fig. 17 shows a 24 Sm^3/d influx in a well producing 4026 Sm^3/d, where probability trends remain nearly unchanged.

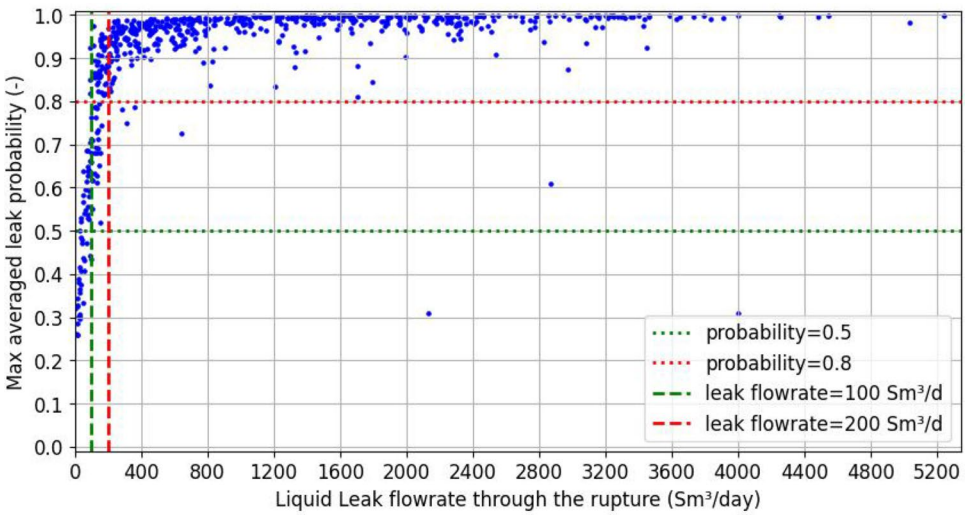

Fig. 18 Maximum leak probability versus steady-state liquid leak flowrate

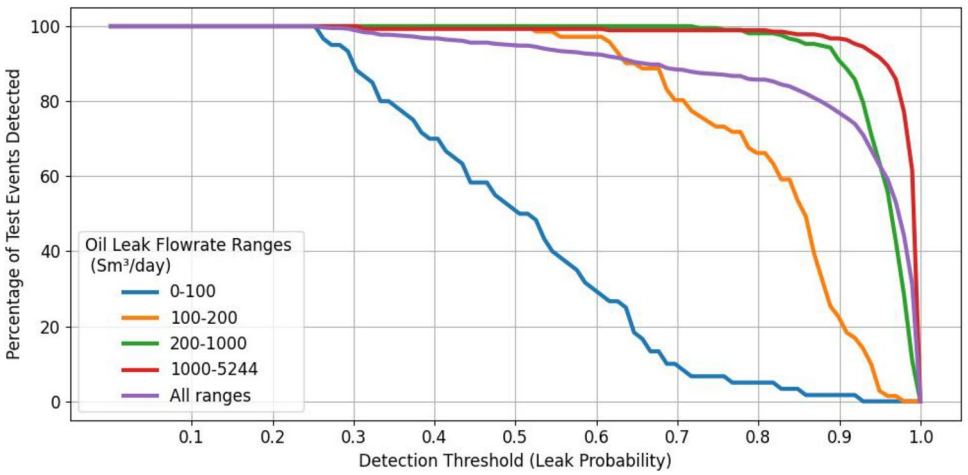

Fig. 19 Leak detection recall for different thresholds and flowrate ranges

Figures 18 and 19 summarize leak classifier performance across flowrates and probability thresholds. A large number of leaks events with flowrate below 100 Sm^3/d were not detected with a 0.5 threshold, whereas most leaks above 100 Sm^3/d were with this same threshold value. For a 0.8 threshold, the transition point shifts to approximately 200 Sm^3/d. Almost all undetected cases were small leaks, except the two anomalous large-leak cases.

Figures 20 and 21 present analogous results for influx detection. The classifier reliably detected nearly all influxes above 50 Sm^3/d (with a 0.8 threshold), and no event above this flowrate had a maximum probability below 0.65.

Across all evaluations, the influx classifier outperformed the leak classifier: it produced fewer false alarms, achieved higher detection rates, and demonstrated lower flowrate sensitivity.

Figure 22 summarizes detection times for all rupture events detected by the classifiers. All distributions are right-skewed with long tails. Leak detection with a 0.5 threshold occurred mostly between 5 and 20 min (mean 12.1 min; median 8 min). With a 0.8 threshold, the mean increased to 18.1 min. Influx detection times were longer, typically between 10 and 40 min at a 0.5 threshold (mean 24.0 min) and between 15 and 70 min at 0.8 (mean 36.2 min). This behavior aligns with the slower thermal response of T-UP-PCK.

Ultimately, leakage volume is often more relevant than detection time. Figure 23 shows histograms of total leaked volume up to detection, calculated by integrating the leak flowrate from rupture onset to alarm. With a 0.5 threshold, 95% of cases leaked less than 24 m^3 (mean 8.0; median 4.3). With a 0.8 threshold, 95% were below 37 m^3 (mean 14.0; median 10.5).

The results in Tables 8 and 9 show that false alarms may occur frequently depending on well stability. Therefore, automatic well-shutdown ("trip") actions should not be triggered immediately upon detection. Sudden shut-ins are operationally risky. Instead, alarms should be complemented with a time delay that allows operators to analyze the cause, acknowledge the alarm, or cancel it. If no intervention occurs within the configured time window, automated safety actions can then proceed.

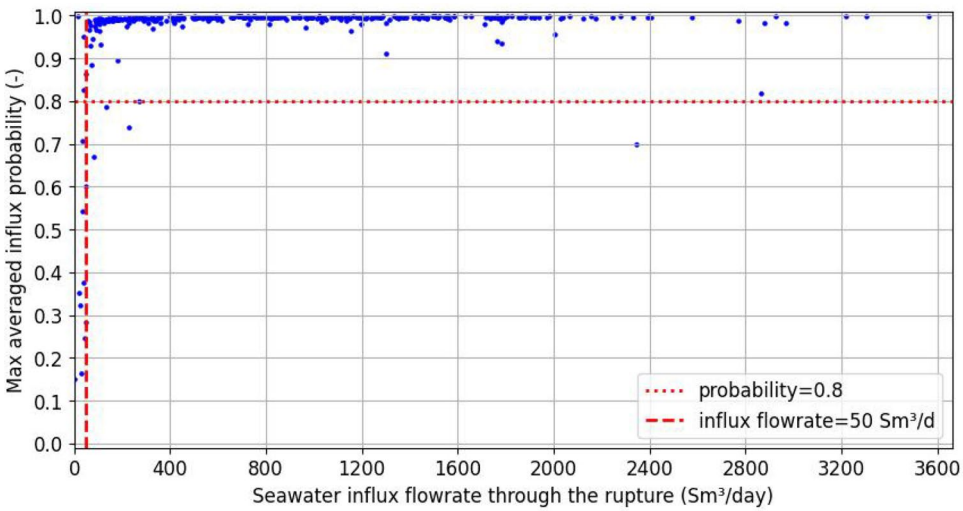

Fig. 20 Maximum influx probability versus steady-state water influx flowrate

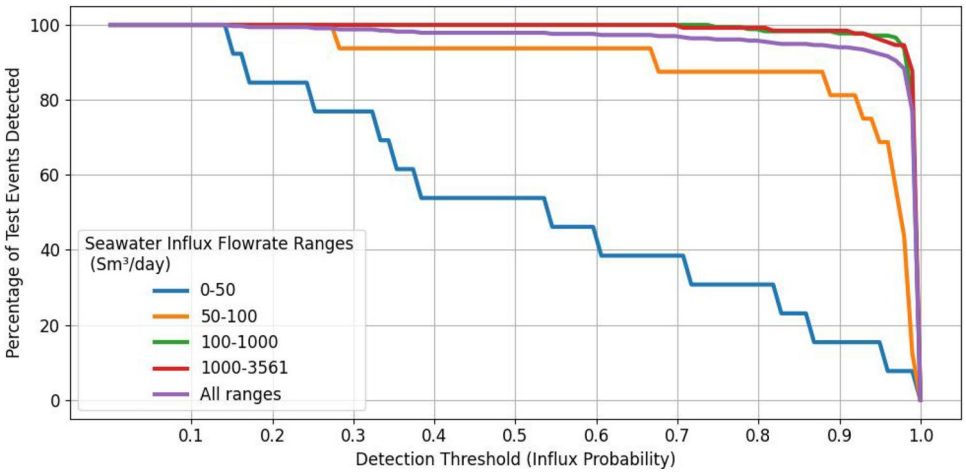

Fig. 21 Influx detection recall across thresholds and flowrate ranges

3.3 Operational deployment and results

As discussed in previous subsections, the classifier uses the relative deviations of four sensors (P-TPT, T-TPT, P-CHK, and T-CHK) as inputs and returns the probability of an oil leak or seawater influx as output. This subsection describes how the system has been deployed in practice.

Initially, each sensor's reference value x_{ref} was manually defined by the well monitoring team. This approach worked for wells with stable conditions, i.e., with nearly constant choke openings and slowly varying reservoir properties such as reservoir pressure, GOR, and watercut, but proved inadequate for wells subject to frequent operational changes. Thus, several automatic estimation methods were evaluated.

For some wells, an automated procedure was implemented to update the reference value based on historical data. Nevertheless, the method that has proven most robust and is currently used for the majority of wells is a simple moving-window calculation: a six-hour window in which the first four hours are used to compute the effective average, while the last two hours act as a buffer (offset) to avoid using data too close to the current timestamp. Other window lengths and offsets may be adopted when appropriate. Although simple and easy to implement, this approach has limitations: (i) ruptures that develop very slowly or occur during shut-ins longer than two hours may not be detected, and (ii) the averaging period must correspond to a stable interval without operational changes or transient behavior.

Overall, estimating reference values remains a rich topic for future research. A promising direction would be to combine data-driven approaches—based on the historical behavior of each well and of its surrounding field—with physics-based digital twins, such as multiphase-flow simulations tuned to measured data. Such hybrid models could enable more accurate and condition-dependent reference values for each sensor.

Alarm suppression mechanisms were also implemented to prevent false alarms during intentional operational changes. The main suppression rule concerns choke-opening variations: a moving comparison—similar to the one used for reference estimation—can identify significant

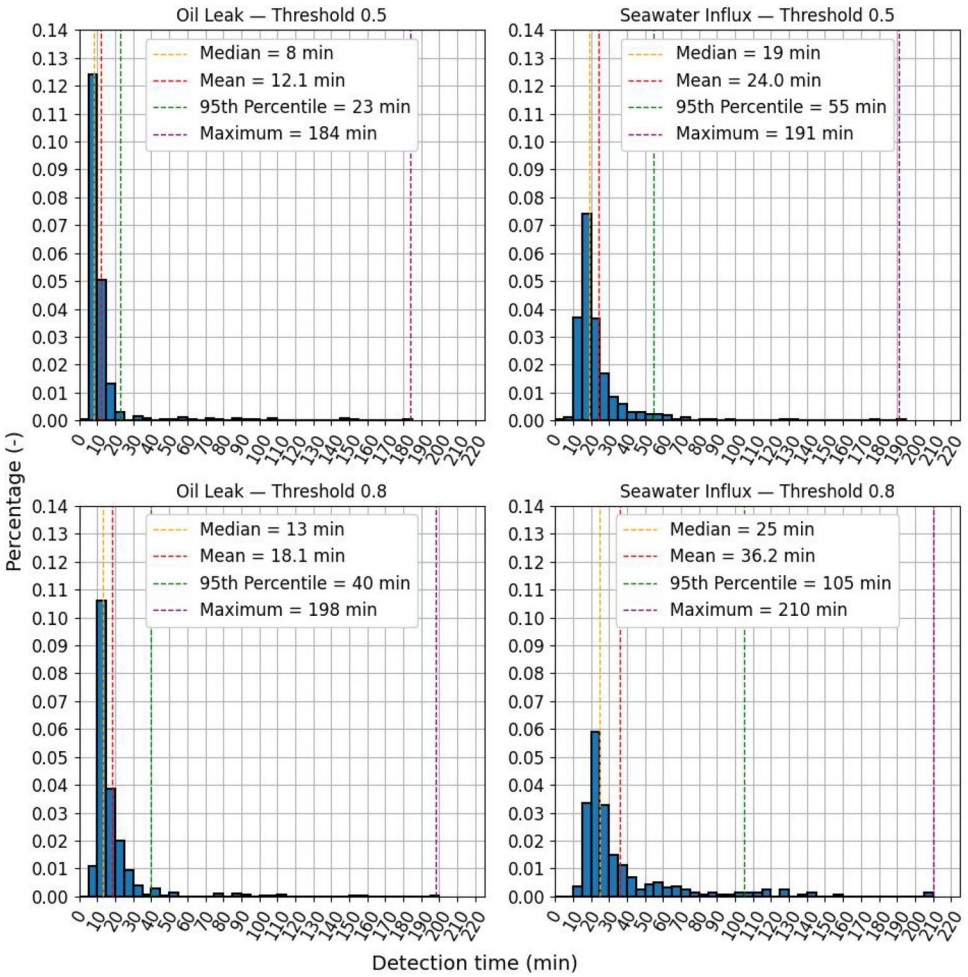

Fig. 22 Detection-time histograms for leak and influx classifiers (thresholds 0.5 and 0.8)

operational changes, thereby temporarily suspending leak and influx alarms. Additional suppression rules may apply during ramp-ups, shut-ins, and periods in which no stable reference can be computed.

The classifier was integrated into Petrobras' Intelligent Production Surveillance System, an event-detection framework similar to the one described by Yanez et al. [26]. It currently monitors the possibility of partial ruptures in production pipelines of approximately 100 oil wells connected to 16 floating production units (FPUs). The alarms are continuously monitored by onshore teams, 24 h a day. Nine of these FPUs, covering 55 wells, also rely on an automated shutdown system capable of closing subsea valves in the event of a detected oil leak or seawater influx.

On average, the SSLDS coverage is around 60% of total operating time, owing mainly to its inability to classify partial ruptures during suppression periods associated with transients or shutdown conditions. Historically, fewer than 2 unintended alarms (false positives) per production unit per month have been recorded, which has contributed to operator confidence in the system. To date, no partial rupture has been observed in any monitored well in the pre-salt region; thus, no true positives or false negatives have been recorded so far.

4 Conclusions

This work presented a methodology for developing an internally based rupture-detection system for subsea production pipelines. The approach consists of four main steps: (i) generation of a synthetic rupture dataset through extensive transient multiphase-flow simulations; (ii) training of machine-learning models to detect oil leaks and seawater influxes using both simulated data and additional operational datasets; (iii) assessment of model performance with independent test events; and (iv) implementation of a configurable detection system suitable for real-time operation.

The results demonstrated that partial ruptures can be detected using only the standard pressure and temperature sensors typically available in subsea production systems. Detection was highly reliable for liquid leak rates above 200 Sm^3/d and seawater influx rates above 50 Sm^3/d,

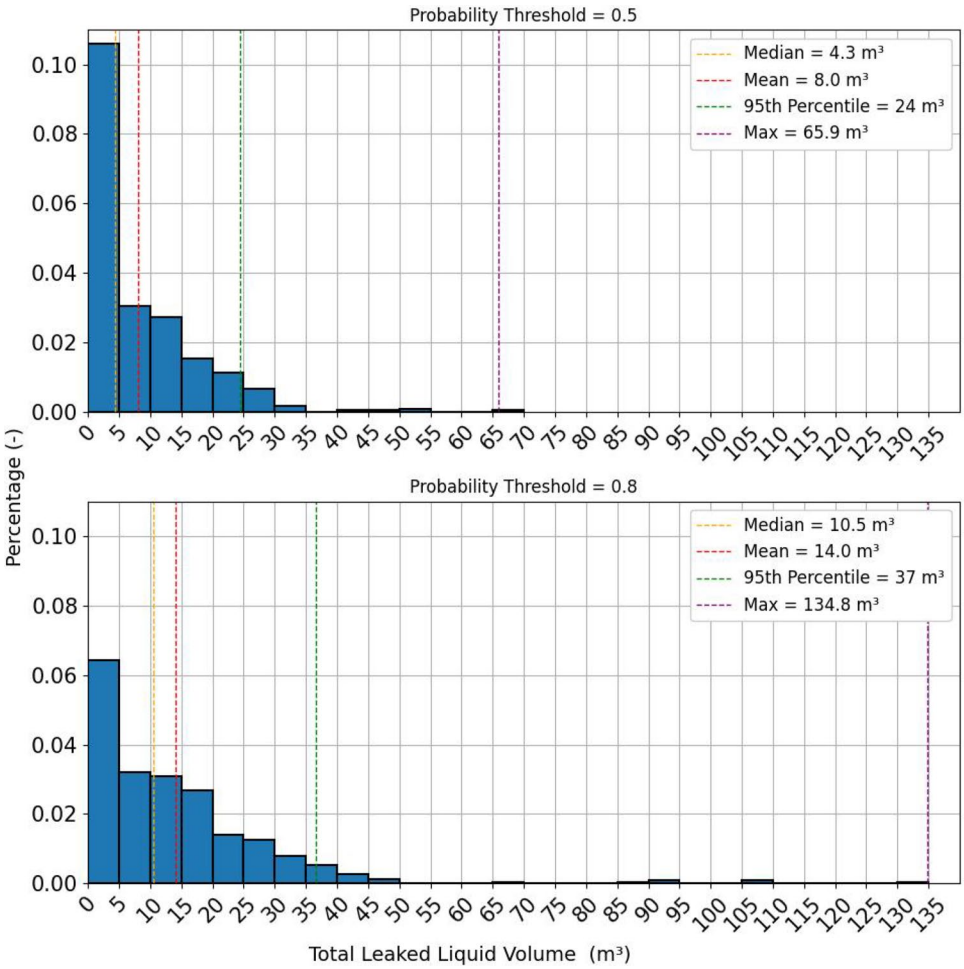

Fig. 23 Histograms of total leaked volume for leak alarm thresholds of 0.5 and 0.8

with near-perfect detection rates during testing. Detection times were also satisfactory for operational use.

Some conditions must, however, be satisfied for successful detection: accurate well simulations for training, functional subsea and topside sensors, reasonably stable well behavior, and suitable reference-value estimation for each sensor. These aspects suggest several directions for future improvement:

- *Improved rupture simulations:* although the multiphase-flow simulator used is state-of-the-art, validation against experimental or field data is needed. Effects related to rupture geometry, axial position, and flow regime may substantially influence sensor responses.
- *Reference-value estimation:* the moving-window method currently used is effective but limited, especially during operational changes or slow-forming ruptures. More advanced data-driven, physics-based, or hybrid approaches could provide more accurate reference estimates.
- *Classifiers independent of reference values:* rate-of-change-based classifiers may complement or replace steady-state comparison methods.
- *Handling missing data:* the method assumes all four sensors are available, which is not always the case. Imputation schemes or alternative classifiers using fewer sensors should be explored.
- *Use of additional variables:* incorporating downhole measurements, choke-opening data, or other topside variables could enhance robustness.
- *Higher-frequency measurements:* these could enable detection of pressure-wave signatures during rupture formation and improve early detection.
- *Detection during transient phase:* distinguishing rupture behavior from operational transients (e.g., choke changes or shut-ins) remains a significant challenge.

Large ruptures generally cause substantial parameter deviations and are therefore easier to detect, often requiring only simple single-sensor threshold checks. The machine-learning system performed well for these cases, though a few extreme scenarios were missed due to atypical combinations

of sensor deviations, as shown in Fig. 18. Conversely, very small ruptures remain difficult to detect, even with the proposed system, as their effects can be smaller than natural measurement noise. In those situations, complementary monitoring technologies—such as low differential-pressure trips, satellite imaging, or ROV inspection—remain important. The developed solution is best suited for intermediate ruptures or leaks (flow rates between 200 and 1000 Sm^3/d). Thus, the proposed solution should be viewed as an additional layer of protection within a broader integrity-management framework, rather than a replacement for existing detection methods. The rupture dataset generated for this work may serve as a foundation for benchmarking and for guiding future research and model improvements.

The methodology developed here has already supported the successful deployment of a real-time rupture-detection system for a large set of producing wells. The observed stability of the system and its low false-alarm rate have contributed to strong acceptance among monitoring and technical teams, increasing confidence in its capability to provide early detection and reduce operational and environmental risk.

Acknowledgements The authors acknowledge Petrobras for technical support and the Production Surveillance team for operational validation of the SSLDS system.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could be perceived as potential conflict of interest in connection with this work.

References

1. Yuan J, Mao W, Hu C, Zheng J, Zheng D, Yang Y (2023) Leak detection and localization techniques in oil and gas pipeline: a bibliometric and systematic review. Eng Fail Anal 146:107060. https://doi.org/10.1016/j.engfailanal.2023.107060
2. Korlapati NVS, Khan F, Noor Q, Mirza S, Vaddiraju S (2022) Review and analysis of pipeline leak detection methods. J Pipeline Sci Eng 2(4):100074. https://doi.org/10.1016/j.jpse.2022.100074
3. Bureau of Safety and Environmental Enforcement: 30 CFR Part 250—Oil and Gas and Drilling Operations. https://www.ecfr.gov/current/title-30/chapter-II/subchapter-B/part-250, eCFR.gov. Accessed 01 Dec 2025
4. API 17TR16—Subsea Hydrocarbon Production Leak Detection Systems Using Process Data. First edition; guidelines for design, operation, maintenance and personnel training for subsea hydrocarbon production leak detection systems using process data
5. Vasconcelos Ferreira Lobo C, Ribeiro ML (2023) Scc-co2 failure mode: literature review and available technologies used to inspect and access the integrity of the flexible pipes. In: Offshore technology conference, vol. Offshore Technology Conference, pp 041–056001. https://doi.org/10.4043/32252-MS
6. Rahman MA, Barooah A, Khan MS, Hassan R, Hassan I, Sleiti AK, Hamilton M, Gomari SR (2024) Single and multiphase flow leak detection in onshore/offshore pipelines and subsurface sequestration sites: an overview. J Loss Prev Process Ind 90:105327. https://doi.org/10.1016/j.jlp.2024.105327
7. Abubakar A, Abisoye OA, Alabi IO, Solomon A, Oyefolahan IO (2025) Systematic literature review and bibliometric analysis of pipeline monitoring and leakage detection techniques. Discov Mech Eng 4(1):17. https://doi.org/10.1007/s44245-025-00102-w
8. Behari N, Sheriff MZ, Rahman MA, Nounou M, Hassan I, Nounou H (2020) Chronic leak detection for single and multiphase flow: a critical review on onshore and offshore subsea and arctic conditions. J Nat Gas Sci Eng 81:103460. https://doi.org/10.1016/j.jngse.2020.103460
9. KROHNE (2022) Principles of leak detection. Technical report, KROHNE Group. Available as PDF from KROHNE website
10. Brekke C, Solberg AHS (2005) Oil spill detection by satellite remote sensing. Remote Sens Environ 95(1):1–13. https://doi.org/10.1016/j.rse.2004.11.015
11. Martins JC, Seleghim JP (2010) Assessment of the performance of acoustic and mass balance methods for leak detection in pipelines for transporting liquids. J Fluids Eng 132(1):011401. https://doi.org/10.1115/1.4000736
12. El-Shiekh TM (2010) Leak detection methods in transmission pipelines. Energy Sources Part A Recov Utiliz Environ Effects 32(8):715–726. https://doi.org/10.1080/15567030903058618
13. Spandonidis C, Theodoropoulos P, Giannopoulos F, Galiatsatos N, Petsa A (2022) Evaluation of deep learning approaches for oil & gas pipeline leak detection using wireless sensor networks. Eng Appl Artif Intell 113:104890. https://doi.org/10.1016/j.engappai.2022.104890
14. Ji J, Li Y, Liu C, Wang D, Jing H (2018) Application of emd technology in leakage acoustic characteristic extraction of gas-liquid, two-phase flow pipelines. Shock Vib 2018(1):1529849. https://doi.org/10.1155/2018/1529849
15. Kam SI (2010) Mechanistic modeling of pipeline leak detection at fixed inlet rate. J Petrol Sci Eng 70(3):145–156. https://doi.org/10.1016/j.petrol.2009.09.008
16. Figueiredo AB, Sondermann CN, Patricio RAC, Bodstein GCR, Rachid FBF (2017) A leak localization model for gas-liquid two-phase flows in nearly horizontal pipelines. In: International mechanical engineering congress and exposition, vol. Volume 7: fluids engineering, pp 007–09006. https://doi.org/10.1115/IMECE2017-71512
17. Havre K, Trudvang C, Kjørrefjord G, Smith S, King CC, Vinicombe J, Roberts T (2022) Use of rigorous multiphase flow models for leak detection and online flow assurance. In: Canadian energy technology conference, vol. SPE Canadian Energy Technology Conference, pp 021–025001. https://doi.org/10.2118/208899-MS
18. Barrionuevo DC (2023) Missing value imputation with similarity-based modeling applied to oil well fault classification. Master's thesis, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil. Concentration area: Electronics–Signal Processing. Research line: Signal Processing. Project: Applications of Signal Processing: Health and Telecommunications
19. Vargas REV, Munaro CJ, Ciarelli PM, Medeiros AG, Amaral BG, Barrionuevo DC, Araújo JCD, Ribeiro JL, Magalhães LP (2019) A realistic and public dataset with rare undesirable real events in oil wells. J Petrol Sci Eng 181:106223. https://doi.org/10.1016/j.petrol.2019.106223
20. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. IEEE Intell Syst Appl 13(4):18–28. https://doi.org/10.1109/5254.708428

21. Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324
22. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017) Lightgbm: a highly efficient gradient boosting decision tree. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in neural information processing systems, vol 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
23. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, Springer series in statistics, 2nd edn. Springer, New York. https://doi.org/10.1007/978-0-387-84858-7
24. Bergstra J, Bardenet R, Bengio Y, Kégl B (2011) Algorithms for hyper-parameter optimization. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira FCN, Weinberger KQ (eds) Advances in neural information processing systems 24: 25th annual conference on neural information processing systems 2011, pp 2546–2554. Curran Associates, Inc., Granada. http://papers.nips.cc/paper/4443-algorithms-for-hyper-parameter-optimization.pdf
25. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge. https://www.deeplearningbook.org/
26. Yanez J, Fuenmayor R, Srivastava P, Sadek N, Vivas P (2024) Applying digital rapid response ranking and event detection system in oriente basin. https://doi.org/10.2118/218125-MS