OPTIMIZING SEPARATOR PERFORMANCE WITH LIQUID CARRYOVER DETECTION & MACHINE LEARNING

Class # 93003

Paul Stockwell - CEO

Process Vision Inc

1650 Marietta Boulevard NW, Unit C14

Atlanta, GA, USA

Introduction

Contamination in natural gas pipelines, particularly in the form of liquid carryover and mist formation, poses persistent challenges for operators across the gas value chain. These issues can lead to equipment damage, reduced process efficiency, and significant safety risks, yet they often occur while systems appear to be operating within specified parameters [1]. Detecting and mitigating contamination is therefore critical for maintaining operational excellence, ensuring regulatory compliance, and safeguarding downstream assets.

Traditionally, contamination control has relied on process indicators such as pressure, temperature, gas chromatography analysis and flow rate, combined into calculated metrics like Calorific Value (Btu) and the hydrocarbon dewpoint (HCDP) [2]. However, industry experience has shown that gas streams can contain liquids even when dewpoint measurements indicate “dry” gas conditions [1, 3]. This discrepancy underscores the limitations of equilibrium-based dewpoint models and sampling standards, which are designed to handle only single-phase gas and explicitly exclude multiphase flows [4, 5].

The introduction of process cameras capable of directly visualising the gas phase inside high-pressure pipelines [2] has brought a better understanding of real pipeline conditions. These systems enable operators to observe contamination events and capture data that can be synchronised with process parameters for further analysis. When combined with machine learning (ML), these visual insights provide an unprecedented opportunity to model, predict, and understand the underlying causes of mist formation and liquid carryover.

Background on Pipeline Contamination

Pipeline contamination arises primarily from two categories: liquid contaminants, including condensates, compressor oils, glycols, and process chemicals, and solid particulates, such as corrosion products and black powder [1]. Liquid carryover may result from inadequate phase separation, upstream pigging operations, surging flow conditions, or process upsets in dehydration and sweetening systems [6]. The consequences are far-reaching: liquids entering compressors can cause unscheduled trips, seal failures, corrosion, and compressor failure resulting in explosions, while downstream contamination leads to metering errors and tariff disputes [7, 1].

Field investigations have demonstrated that traditional monitoring systems, hydrocarbon dewpoint analyzers, moisture sensors, and chromatography, often fail to detect two-phase flow conditions [1]. Sampling systems following API 14.1 and ISO 10715 are designed to exclude liquids to protect analyzers, meaning that operators are usually “running blind” to liquid ingress [4,5]. Consequently, liquids and mist routinely pass through custody transfer points undetected, undermining fiscal accuracy and increasing risk.

The inability of existing instrumentation to verify gas-phase integrity has driven the development of direct optical observation systems. Visual access into pressurized pipelines has revealed frequent mist and stratified liquid flows even when dewpoint and flow parameters indicate normal operation [2]. Such findings challenge long-held

assumptions about “dry gas” and have prompted a paradigm shift toward data-driven, image-assisted contamination monitoring.

Process Camera Contamination Monitoring Systems

These monitoring systems include a permanently installed process camera and illumination system. It is mounted on a standard tapping point above the pipeline, isolated via a valve, allowing continuous video capture without intruding into the flow or disrupting pigging operations. An example of a process camera installation can been seen in Figure 1. The system provides a video feed of the internal gas stream, enabling operators to directly visualize the onset of mist, liquid slugs, or particulate contamination [1].

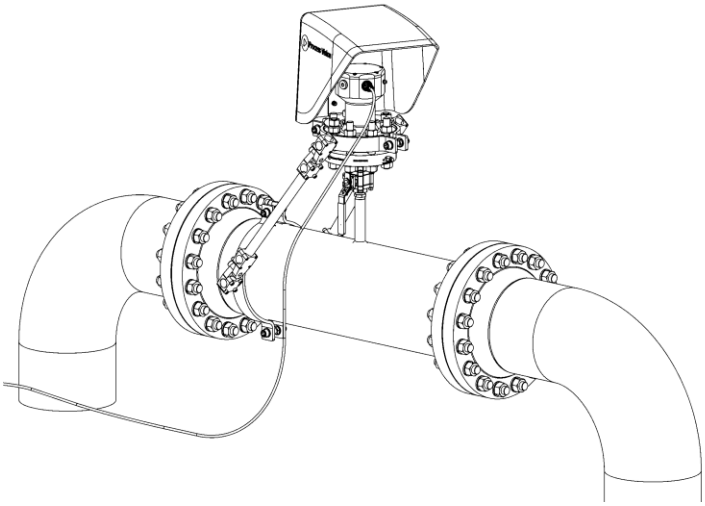

Figure 1 - Example of LineVu Installation

Each video frame is processed to extract quantitative metadata that serves as an optical proxy for contamination, including:

Brightness – total pixel light intensity, indicating overall reflectivity of the flow

Brightness variance – contrast variation across pixels, sensitive to turbulent or mixed-phase regions

Laplacian variance – a measure of image sharpness, reflecting edge density and droplet definition

Under normal dry-gas conditions, the camera system reports a stable, low-brightness baseline. When contamination occurs, these metrics exhibit distinct temporal and spatial changes, brightness spikes for reflective liquid slugs, variance increases for mist and turbulence, and sharpness peaks for droplet edges [3].

Beyond event detection, timestamped video is provided which then can be synchronised with process data (pressure, flow, temperature) for post-event analysis, enabling root-cause identification and improved separator performance assessment. The system has been deployed in major transmission networks, gas processing facilities, and compressor stations across Europe, the Middle East, and North America [8].

Machine Learning Background and Relevance

Machine learning has emerged as a powerful analytical approach in the process industries, offering tools to uncover non-linear and multivariate relationships that traditional regression methods fail to capture [9]. In natural gas operations, ML has been successfully applied to leak detection, equipment failure prediction, and flow regime classification [10, 11]. Among available techniques, Random Forests (RF) are particularly effective for process data because of their ability to handle noisy inputs, identify key variables, and quantify feature importance [12].

In the context of contamination monitoring, Random Forest regression models have been trained to predict camera-derived features such as brightness variance and Laplacian variance from a variety of conventional process data, achieving coefficients of determination exceeding 0.9 in field trials [3]. These results demonstrate that video-derived brightness metrics can serve as quantitative proxies for contamination levels, providing a bridge between visual evidence and numerical process control.

Furthermore, feature-importance and sensitivity analyses from such models can highlight the dominant process variables influencing contamination, thereby guiding operators toward targeted interventions. The integration of ML-driven predictive analytics with direct visual feedback therefore represents a step-change in process

understanding, moving from reactive detection toward proactive contamination management and digital twin development for gas processing assets.

This paper describes the development of a machine learning model that integrates process camera data and a wide range of process inputs to enable continuous monitoring of pipeline conditions. The model provides predictive capabilities, smart alarms, and insight into the factors influencing mist formation and liquid carryover events, including identification of the process inputs most strongly associated with historical event occurrences. The approach involves ingesting raw process data and camera feeds, performing end-to-end model training and validation, and automating the workflows through a custom internal tool. The scalable system supports real-time benchmarking, alerting, and performance tracking across multiple sites. The approach is designed for integration into Integrity Management Programs and digital-twin environments, linking model outputs directly with operational decision-making.

By combining visual and process data, the methodology in this paper offers a previously unavailable view into pipeline behavior, representing a significant step toward predictive, data-driven gas processing improvement and supporting the broader goal of operational excellence through digitalization.

Nomenclature

- MMWC Millimetre Water Column
- MMSCFD Million Standard Cubic Feet per Day
- PSIG Pounds per Square Inch Guage
- PSI Pounds per Square Inch
- DEG C (°C) Degrees Celsius
- PERCENT (%) Percentage

Brightness Data and How it Relates to Pipeline Contamination

The brightness level, brightness variance, and Laplacian variance are image meta data metrics that provide insights into the presence and nature of contamination in the pipeline.

Brightness level is a measure of the overall light intensity in the scene, with higher values indicating more light or brighter conditions. In the LineVu system, the camera is paired with an integrated active illumination unit, so the observed scene is lit by a controlled internal light source rather than relying on ambient lighting which would be inconsistent. The illumination unit consists of multiple high-powered LEDs distributed around the central lens to provide an even light coverage while allowing for a centrally aligned camera view. As a result, changes in measured brightness primarily reflect changes in optical behavior within the gas stream, such as scattering and specular reflection from aerosols or liquid films, rather than external lighting variation. Liquid contamination tends to cause spiked peaks in brightness due to the reflective nature of liquids. In contrast, mist or fog-like contamination results in a flatter brightness profile with lower peaks, as the mist scatters and diffuses the light. Brightness variance quantifies the contrast or variation in brightness across the image pixels. Liquid contamination, with its distinct liquid droplets or streams, creates high contrast with bright highlights and dark shadows, leading to higher brightness variance. Mist or fog, being more homogeneous in nature, exhibits lower brightness variance due to its lack of sharp contrasts. Laplacian variance is a measure of sharpness or focus, indicating the presence of well-defined edges in the image. Liquid contamination often results in higher

Laplacian variance due to the sharp edges between the liquid and gas phases. However, turbulent mist can also exhibit higher Laplacian variance as the mist propagates and forms distinct edges.

Each metric is computed over a 60 second sliding window, with the minimum and maximum capturing the range of short-term fluctuations and the temporal moving average capturing the underlying trend in contamination behaviour:

- The maximum values represent brief periods of intense optical change, which may correspond to sudden contamination events such as splashes, droplets passing through the field of view, or turbulent bursts of mist.
- The minimum values, in contrast, indicate the baseline or clean-state conditions of the pipeline, where little to no contamination is present and the image remains stable and uniform.
- The average values provide a smoothed measure of general contamination levels over time, helping to distinguish persistent contamination (such as sustained fog or residue buildup) from transient spikes.

Together, these min-max-average statistics give a more complete picture of both the intensity and stability of contamination dynamics within the monitored segment.

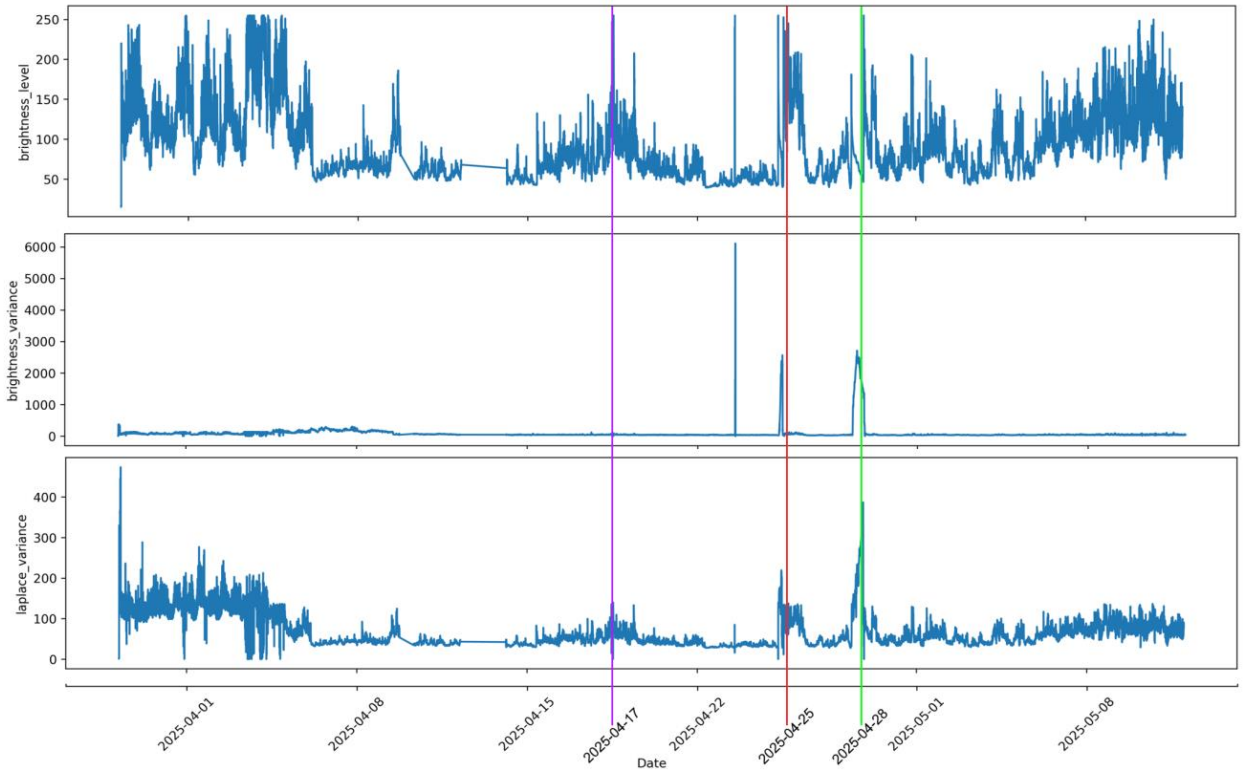

Figure 2 - Brightness data extracted from frames of pipeline video, showing how metrics for Brightness Level, Brightness Variance, and Laplacian Variance can vary against each other. Three events are highlighted in purple, red, and green respectively.

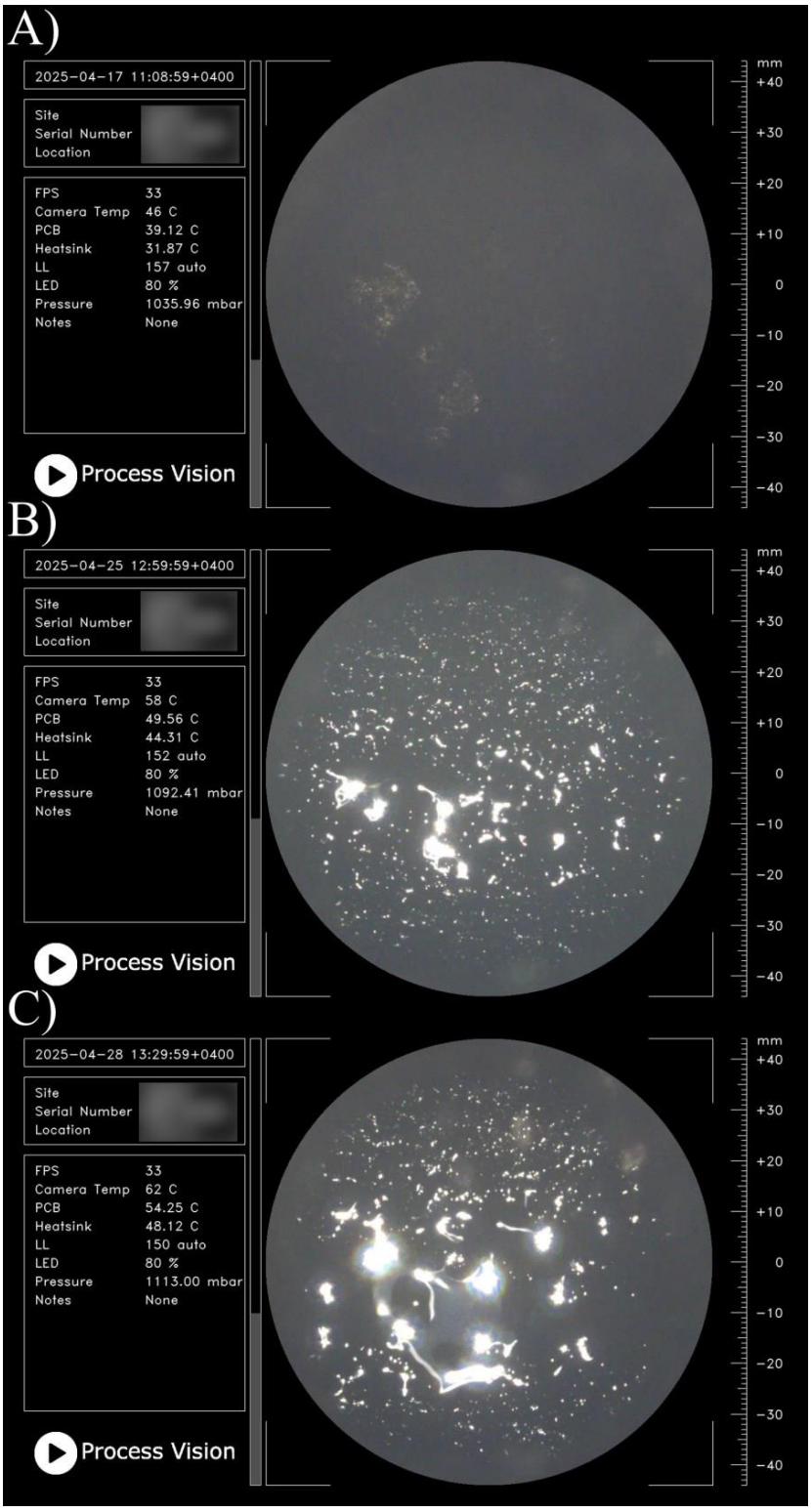

Figure 3 – Image snapshots of pipeline events:

A) View of the pipeline during a mist event marked by the purple line of Figure 2,

B) View during an event marked by the red line of Figure 2 some liquid contamination present,

C) View during an event marked by the green line in Figure 2, a large amount of liquid contamination present

Figure 2 illustrates how the three image-derived metrics, brightness level, brightness variance, and Laplacian variance, fluctuate during live operation, with three representative contamination events highlighted in red and green. While each metric describes a distinct optical property, they are inherently correlated under stable

conditions, typically rising and falling together as illumination and flow dynamics vary. However, during contamination events, the relationships between these metrics diverge, revealing characteristic patterns that enable event detection and classification through comparative analysis. Figure 3 presents snapshots from three representative events corresponding to the colored markers in Figure 2.

Event 1 (purple line in Figure 2; snapshot A in Figure 3) represents the onset of mist contamination. The image shows a light haze developing across the field of view, with minimal liquid present. This corresponds to a gradual increase in brightness level as scattered light from the mist raises overall luminance. The Laplacian variance increases moderately due to the turbulence within the mist, which introduces weak edges and subtle textural variation. In contrast, brightness variance remains low, reflecting the uniform nature of mist and its tendency to reduce contrast across the image.

Event 2 (red line in Figure 2; snapshot B in Figure 3) captures a liquid contamination event characterized by visible deposition along the pipe floor. All three metrics exhibit pronounced peaks in response. Brightness level increases sharply due to strong specular reflections from the liquid surface, while brightness variance rises as these bright highlights contrast with darker background regions. Laplacian variance also increases, reflecting the formation of sharp boundaries and distinct edges associated with reflective liquid interfaces.

Event 3 (green line in Figure 2; snapshot C in Figure 3) shows a more substantial liquid contamination event, with greater accumulation along the pipe floor than in Event 2, evident by the transition from many small, isolated bright speckles produced by droplets to larger, continuous bright regions produced by pooled liquid, which creates broader specular reflections in the black and white image. Although all metrics increase, the magnitude and temporal patterns differ. The brightness level fluctuates as larger reflective surfaces shift with the liquid flow, while brightness variance remains consistently elevated, driven by strong contrast between reflective and non-reflective regions. The Laplacian variance shows the most significant increase, consistent with the presence of larger, well-defined liquid features producing sharper transitions in the image.

These examples demonstrate that variations in brightness-based image metrics can effectively characterize both the presence and nature of pipeline contamination. The distinct temporal and relational patterns across the metrics provide insight into whether contamination is mist-like, liquid-based, or transitional. As such, these image-derived indicators can serve as proxies for traditional process measurements, linking visual contamination data to process conditions. This forms the foundation for integrating image analytics with process parameters, such as flow rate, pressure, and temperature, to enhance predictive monitoring and contamination management in industrial pipeline systems.

PROCESS DATA AND HOW IT RELATES TO BRIGHTNESS DATA

Overview of Process Variables

The process dataset provides a quantitative foundation for understanding how pipeline operating conditions influence the brightness data derived from the process-camera system. The dataset comprises a range of pressures, differential pressures, flow rates, temperatures and level measurements, all collected over the same operating period as the optical data. Each parameter represents a distinct physical aspect of the system, capturing both steady-state and transient operating regimes.

The intention was that relevance would be learned during model training: inputs that did not contribute predictive signal would naturally receive negligible weight, while informative variables would be emphasised. This allowed

the model to surface unexpected but meaningful relationships that may have been missed under a manual variable selection approach.

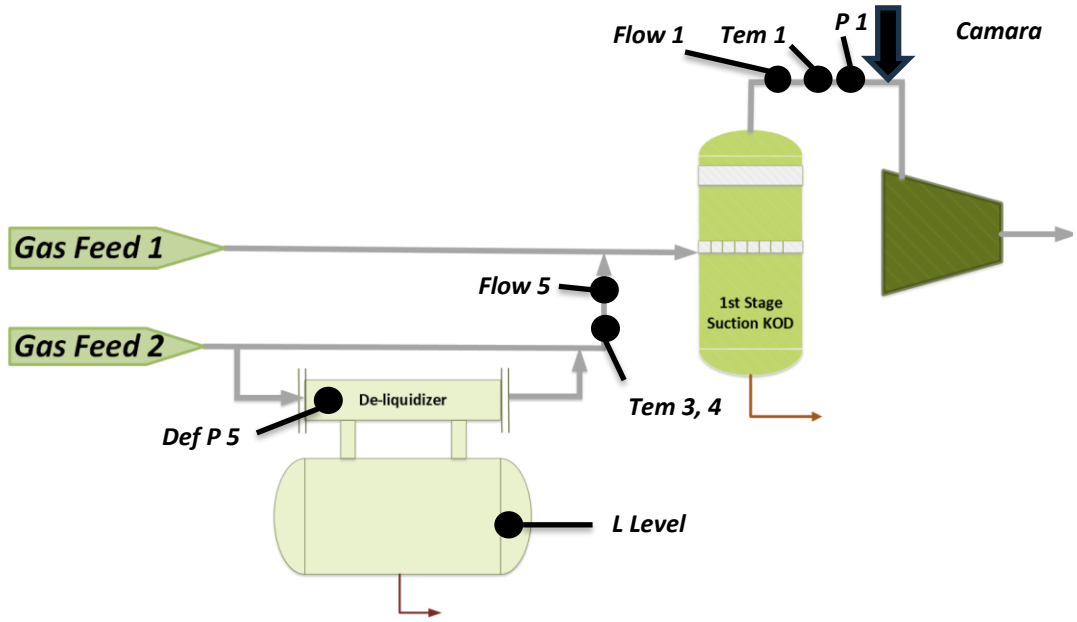

Figure 4 - Separator and Compressor

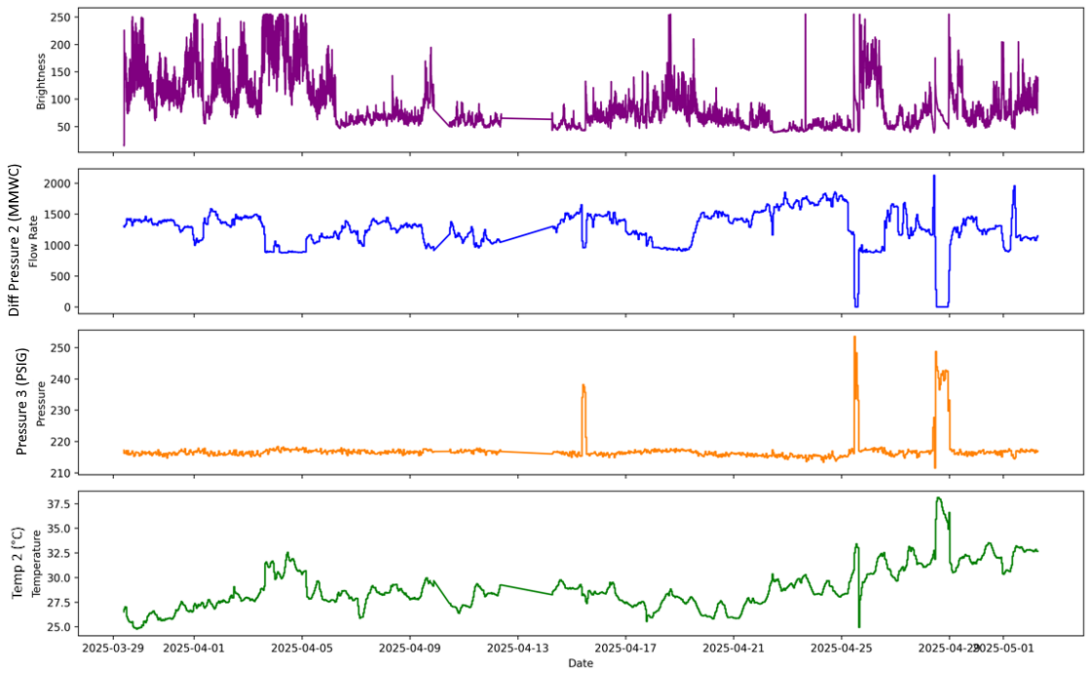

Figure 5 – Comparison of brightness under a subset of core changing process conditions, flowrate, pressure, and temperature. Missing data (i.e. 2025-04-13) shown as a straight line between closest available points.

As shown in Figure 5, variations in brightness closely follow fluctuations in several process variables over time. When the flow rate decreases, a corresponding rise in brightness can be observed, implying a reduction in gas velocity and increased likelihood of mist or droplet entrainment. Conversely, elevated pressure and temperature tend to stabilize brightness at lower levels, indicating drier flow conditions. These dynamic interactions confirm

that optical metrics are responsive to underlying process states rather than random noise, validating their use as a real-time contamination indicator.

Statistical Distribution and Parameter Interpretation

To better understand the inherent variability of the operating conditions, Figure 6 illustrates the frequency distributions of the sixteen primary process variables and the minimum brightness value.

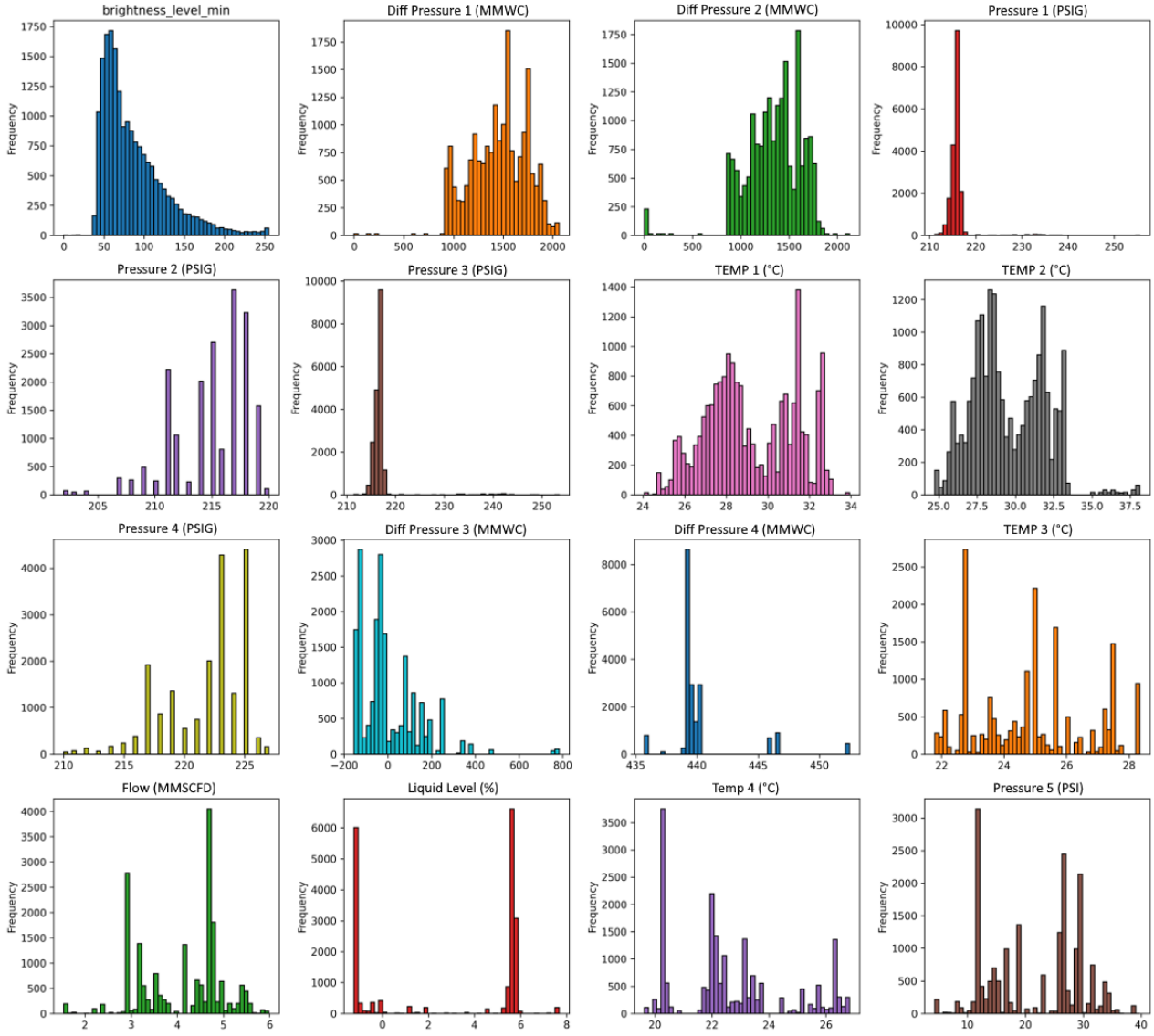

Figure 6 – Frequency Distributions of Sixteen Process Variables.

Figure 6 displays the frequency distributions of sixteen distinct process variables, revealing diverse distribution patterns from unimodal and skewed to multimodal and sharply peaked, highlighting the inherent variability and characteristic operating points of each parameter.

The distributions range from unimodal to multimodal and sharply peaked, revealing how different instruments and control loops shape plant behaviour. For example, Flow (MMSCFD), representing gas flow rate, exhibits a distinctly bimodal pattern, corresponding to high- and low-throughput modes.

Temperature sensors such as Temp 1(°C) and Temp 2(°C) display multimodal behaviour associated with compressor recycle conditions, while pressure sensors cluster around nominal operating ranges, indicating tight control. Process data was collected from instrumentation upstream and downstream of the process camera installation point. The brightness level min variable shows a right-skewed distribution, with most data points concentrated at lower intensities, characteristic of clean, dry-gas flow. The long tail toward higher brightness

values represents intermittent contamination events where optical scattering increases due to droplets or mist. This behavior reinforces the interpretation of brightness as a proxy variable for contamination intensity.

Hysteresis and Dynamic Response

Inspection of the time-series trends, seen in Figure 5, reveals hysteresis-type behaviour between process conditions and optical response. During load-change cycles, brightness rises rapidly following pressure or flow disturbances but decays more slowly when operating parameters return to baseline. This asymmetry arises because liquid condensation and evaporation occur at different thermodynamic thresholds; once formed, fine droplets persist even when temperature and pressure normalize [2].

This hysteresis effect highlights the time-lagged relationship between process perturbations and optical recovery, a phenomenon consistent with condensation, droplet re-entrainment, and separator drain-back dynamics [1]. It demonstrates that simple linear or instantaneous relationships are insufficient and that subsequent modelling must capture non-linear, temporal dependencies.

Correlation Analysis and Model Justification

To quantify the strength and direction of relationships between optical and process data, a Pearson correlation matrix was generated, shown in Figure 7.

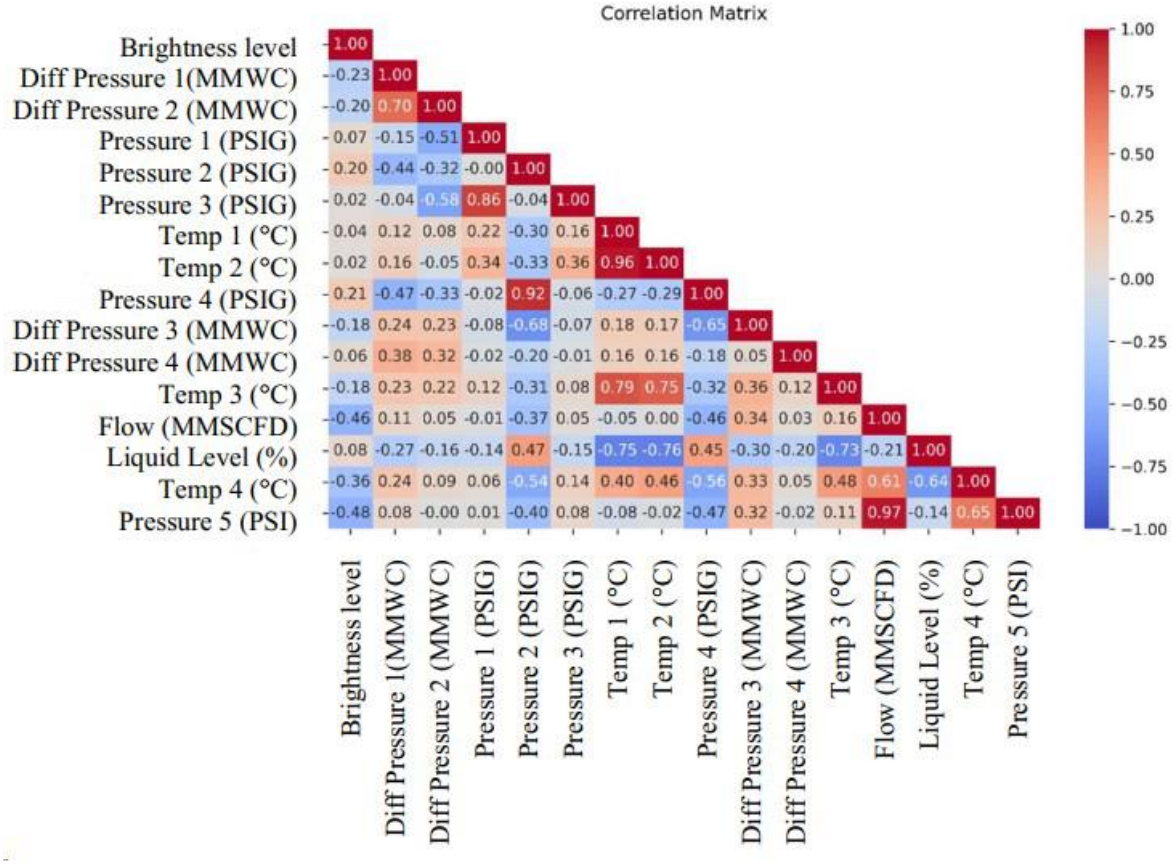

Figure 7 – Correlation Matrix for Brightness Vs Process Parameters

The correlation analysis indicates that no single process variable strongly dictates the brightness response. Moderate negative correlations (≈ –0.46 to –0.33) between temperature and brightness suggest that cooler gas, closer to its dewpoint, aligns with increased mist formation. Conversely, positive correlations (≈ 0.23–0.38) for several differential-pressure tags imply that separator loading and pressure imbalances coincide with higher brightness levels. Flow rate shows a weak inverse relationship (≈ –0.46), consistent with reduced flow enhancing droplet residence time.

These moderate but physically consistent correlations indicate that brightness behavior arises from the combined influence of multiple interacting variables rather than from a single dominant factor. Such

interdependence supports the application of ensemble machine-learning models, such as Random Forest regression, capable of capturing multivariate, non-linear relationships and feature interactions [9].

Furthermore, the relatively independent behavior of brightness compared with most process tags confirms that optical data provides new and complementary information beyond conventional instrumentation. This justifies its integration into digital-monitoring frameworks as a contamination indicator.

Summary

The combined analysis of the figures in this section demonstrate that optical brightness data responds coherently to process changes and can be quantitatively linked to thermodynamic and conditions within the gas stream. The observed hysteresis and multi-parameter correlations validate the need for non-linear, data-driven modelling to predict contamination events. Together, these findings establish the technical rationale for fusing process and visual data in the subsequent machine-learning framework.

MODEL CONSTRUCTION METHODS

Overview and Objectives

To quantify how operating conditions influence optical contamination signatures, a machine-learning model was developed to predict the behavior of the process-camera brightness signal directly from pipeline process data.

Rather than manually preselecting a limited set of “expected” drivers, the full set of available site process data tags for the relevant unit and time period was requested and aligned to the camera data. Variable relevance was then learned during model training, with non-informative inputs contributing negligible influence, while informative tags were automatically emphasised. This approach reduces analyst bias and allows unexpected but physically meaningful predictors to emerge.

The objective was twofold:

1. To determine whether process-measurement variables alone could explain variations in optical brightness.
2. To create a validated, reproducible framework capable of scaling across multiple installations as part of a digital-twin architecture.

The approach employed a Random Forest (RF) regression algorithm, chosen for its ability to model complex, non-linear relationships while retaining interpretability, a key requirement for adoption in operational environments [12].

Random Forest Modelling Approach

Random Forest is an ensemble-learning technique that averages the outputs of multiple decision trees, each trained on a random subset of the dataset.

This approach reduces over-fitting, handles noisy multivariate data, and captures high-order interactions between process variables [13].

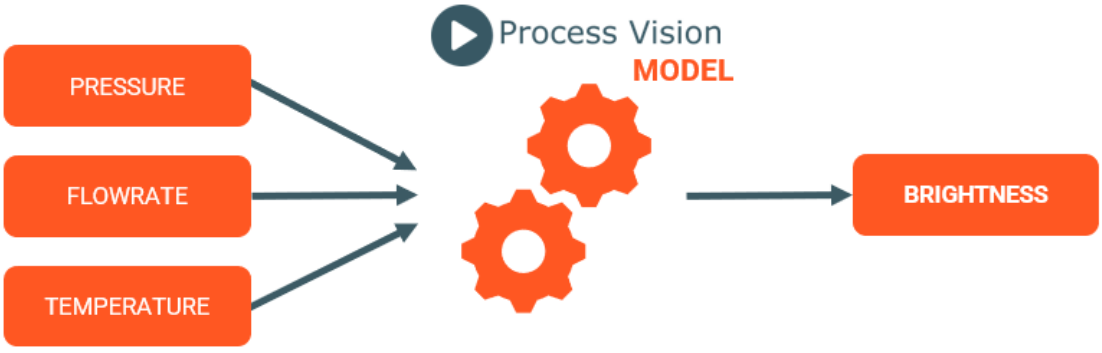

Figure 8 – Simplified model of data inputs and outputs to create a digital twin for brightness

As illustrated in Figure 8, process variables, including pressure, flow rate, temperature, and differential pressure, form the input feature set (X), while camera-derived metrics such as brightness level min, brightness variance, and Laplacian variance serve as target outputs (y). The trained RF model therefore functions as a digital twin

of the optical system, predicting how the camera's brightness response evolves under changing operating conditions.

The algorithm's ensemble nature provides resilience against measurement noise and missing data, making it particularly suitable for field applications in which process signals may fluctuate or intermittently drop out.

Training, Validation, and Hyper-parameter Strategy

Operational datasets are partitioned into 80% training and 20% hold-out test subsets. This ratio provides a practical balance between model learning capacity and unbiased evaluation, ensuring that the RF algorithm is exposed to sufficient operational variability while retaining unseen data for validation [14]. The larger training fraction captures the diversity of pressure, flow, and temperature conditions encountered during normal and upset operation, while the independent test subset evaluates the model's ability to generalize to unseen process states.

A convergence based tuning strategy was used: the number of trees and other hyperparameters were increased stepwise. This approach follows the best-practice guidance of Probst, Boulesteix and Hornik (2017) [13] and avoids unnecessary model complexity while retaining sufficient depth to capture non-linear behavior in the data.

Cross-validation within the training refined the parameters governing tree depth, minimum samples per leaf, and feature subset size.

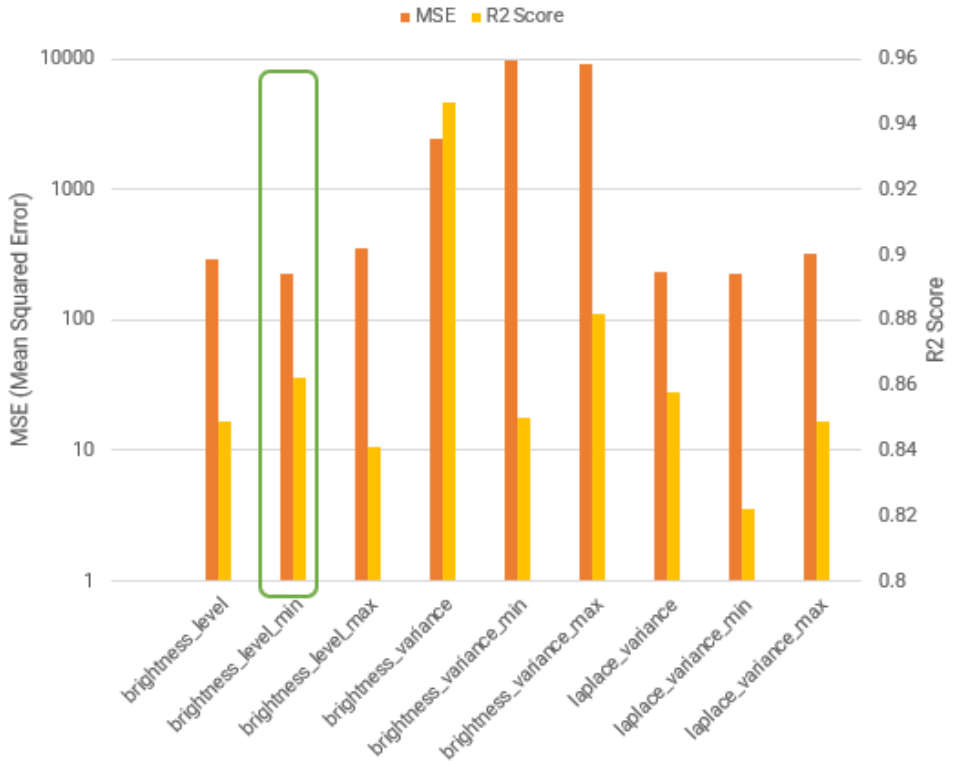

Figure 9 – Model Performance Comparison Across Image Features

Figure 9 shows the Mean Squared Error (MSE) and R^2 scores for nine different image features, with brightness variance min and brightness variance max achieving the highest R^2 scores (~0.95-0.96) but also the highest MSE values (~9000-10000), while Laplace variance min shows the lowest MSE (~5) but reduced predictive performance (R^2 ~0.82). Brightness level min has the best middle ground between low error and high predictive performance, highlighted by the green outline.

Model performance was assessed using the coefficient of determination (R^2) and Mean Squared Error (MSE) metrics. Figure 9 compares the results for nine optical-feature targets. Models predicting brightness-variance metrics achieved the strongest correlation with process data ($R^2 \approx 0.95–0.96$) but also exhibited higher absolute MSE because of the large numerical range of those features. In contrast, models predicting Laplacian-variance min produced the lowest MSE (≈ 5) but a reduced R^2 (~ 0.82). These differences reflect how each optical descriptor captures a distinct aspect of contamination: brightness variance responds to fluctuations in overall light scatter from mist or foam, while Laplacian variance quantifies image sharpness associated with droplet

edges or particulates. As such, both metrics are complementary and were modelled separately to preserve interpretability.

Although R^2 and MSE are both used to evaluate regression quality, they describe different facets of model performance. A high R^2 indicates that the model captures the overall pattern and variability of brightness behavior, yet it can coincide with a relatively high MSE if the model occasionally misestimates the magnitude of peaks or troughs during transient contamination events. Conversely, a low MSE with moderate R^2 suggests accurate absolute predictions but limited ability to reflect subtle dynamic variations in brightness. In this context, the brightness-variance models tended to maximise R^2, effectively reproducing the temporal pattern of mist activity, whereas the Laplacian-variance models achieved the smallest MSE, providing precise amplitude prediction for sharply defined optical events. Balancing these metrics is therefore crucial: a model optimized solely for R^2 may deliver strong interpretive insight but weaker quantitative accuracy, while one tuned for MSE may provide precise brightness magnitudes yet miss early-stage pattern development. The adopted RF configuration represents a deliberate compromise, maintaining high explanatory value while ensuring numerical fidelity suitable for deployment in real-time contamination monitoring.

Software Workflow and Deployment Architecture

The modelling workflow was designed for repeatability and scalability, enabling consistent training and deployment across multiple sites [15].

It follows a streamlined four-step process:

1. Data ingestion and alignment – process and camera datasets are synchronized to a common one-minute interval, filtered for noise, and interpolated where necessary.
2. Feature engineering – derived variables such as dewpoint margin, flow-rate gradient, and brightness-window statistics, enhance sensitivity to transient contamination behavior.
3. Model training and validation – RF regressors are trained on 80 % of the data and tested on the remaining 20 %, with key metrics and configurations logged for reproducibility.
4. Model deployment – validated models are archived and can run centrally or locally, producing real-time predictions of brightness and contamination likelihood for integration with control-room dashboards.

This modular workflow links model development, validation, and field deployment, enabling continuous improvement as additional operational data are collected.

Model Outputs, Performance Assessment, and Interpretation

Each Random-Forest model generates a consistent set of outputs that provide both statistical validation and engineering insight. Model accuracy was assessed using the coefficient of determination (R^2) and Mean Squared Error (MSE). Together, these metrics measure how effectively the model reproduces overall brightness

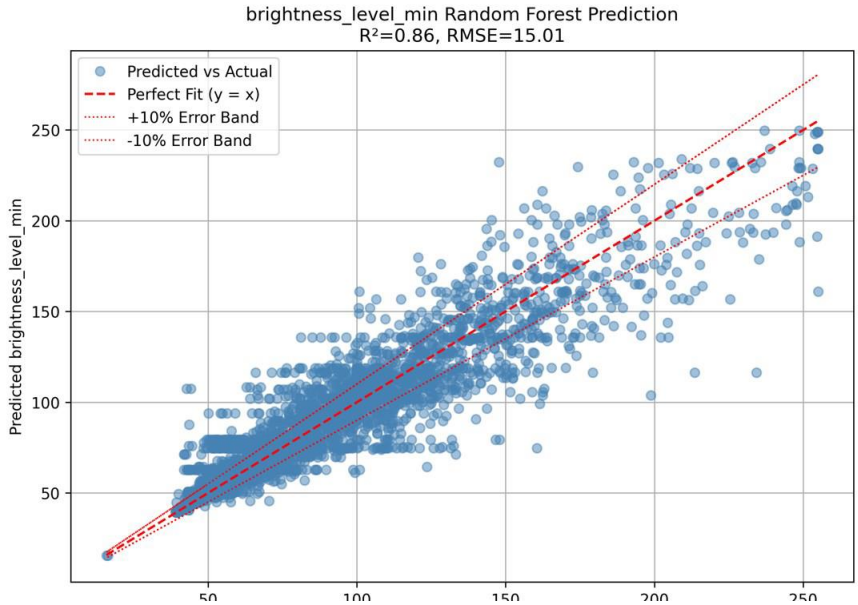

Figure 10 – Random Forest Model Performance for brightness level min Prediction.

trends while maintaining quantitative precision. High R^2 values across all optical features confirm that the model captures the non-linear dependencies between process conditions and contamination behavior.

The scatter plot in Figure 10 compares predicted versus actual values for brightness level min, demonstrating the RF model's performance with an R^2 of 0.86 and RMSE of 15.01, with most data points falling within ±10% error bands indicating good predictive accuracy.

As shown in Figure 10, predicted and measured brightness level min values align closely, with $R^2 = 0.86$ and RMSE ≈ 15. Most predictions fall within ±10 % of observed values, indicating that the model generalizes effectively across operating regimes and reliably reproduces the brightness excursions associated with contamination events.

Analysis of feature importance reveals that gas-temperature margin, separator differential pressure, and flow rate contribute most strongly to the prediction, reflecting their thermodynamic and hydraulic influence on mist formation. Secondary parameters such as downstream pressure and flow stability play supporting roles in shaping the onset and recovery of contamination events. The results align with established field observations that small temperature deficits near the hydrocarbon dewpoint, coupled with pressure fluctuations, are leading indicators of carryover formation.

Further examination through sensitivity testing confirmed that the model behaves consistently with process physics: a decrease in temperature margin or flow velocity increases predicted brightness variance, representing higher droplet density and optical scattering. These findings demonstrate that the RF framework does not merely fit statistical correlations but successfully captures the physical dependencies linking process parameters and optical signatures.

Overall, the modelling approach balances statistical performance with interpretability. The models reproduce brightness behavior directly from process data while revealing which operational variables most strongly drive contamination. This foundation supports the broader objective of integrating predictive modelling into digital-twin workflows for real-time gas-quality monitoring and informs the feature-importance and sensitivity analyses discussed in the following section.

ANALYSIS OF MODEL RESULTS

The evaluation of model performance provides insight not only into statistical accuracy but also into the physical predictability of contamination events using process data. A model’s performance, whether strong or weak, offers diagnostic value: a high-performing model indicates that process variables adequately describe the conditions driving contamination, while poorer model performance suggests that the event type or flow behavior may be weakly coupled to the available process measurements. In this sense, reduced model fidelity does not necessarily reflect poor training data but instead highlights the intrinsic limits of process-data predictability for certain contamination mechanisms [15].

Prediction Behaviour and Event Representation

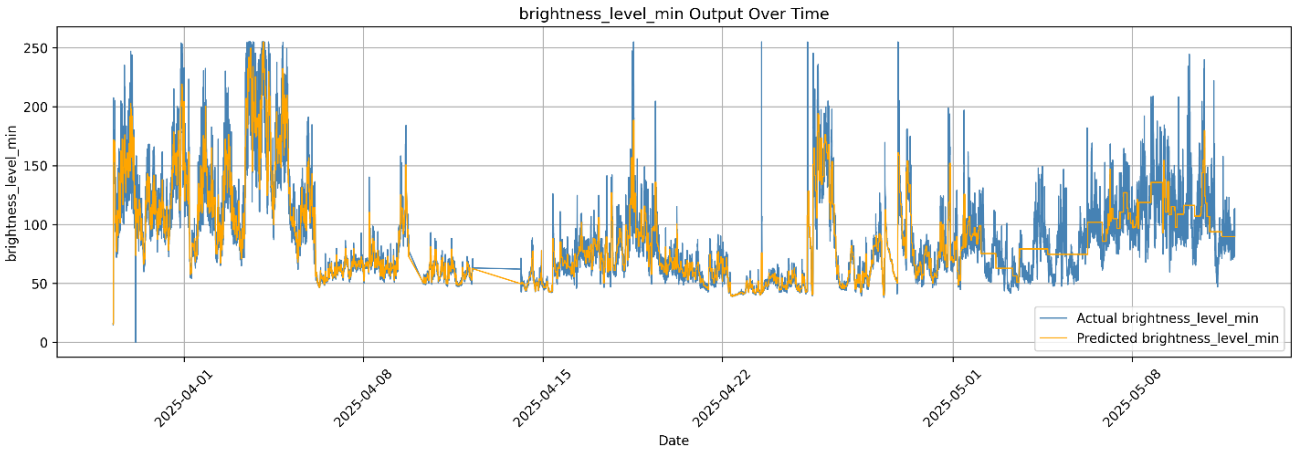

Figure 11 – Time series comparison of actual (blue) and predicted (orange) brightness level min values from March 28 to May 8, 2025.

The model captures general trends well but shows reduced accuracy in the later period, where predictions become less responsive to rapid fluctuations in the actual data due to lack of process data for the period.

As shown in Figure 11, the RF model captures the general evolution of brightness level min over the analyzed period, reproducing both high-brightness excursions and quiescent baseline intervals. Within the period analyzed, the predicted brightness (orange) follows the measured signal (blue) closely, correctly identifying several large-amplitude contamination events. In later periods, deviations appear as the model becomes less

responsive to rapid local fluctuations, primarily due to reduced variability in the corresponding process-data inputs.

These deviations are operationally meaningful: sharp differences between predicted and measured brightness often coincide with unmodelled transient events, for instance, liquid slugs, separator drain failures, or condensation surges, that are poorly represented by slowly varying process parameters.

Hence, the prediction error itself can serve as an indicator of abnormal process dynamics, providing an additional diagnostic layer for event detection.

Feature Importance and Relationship to Process Data

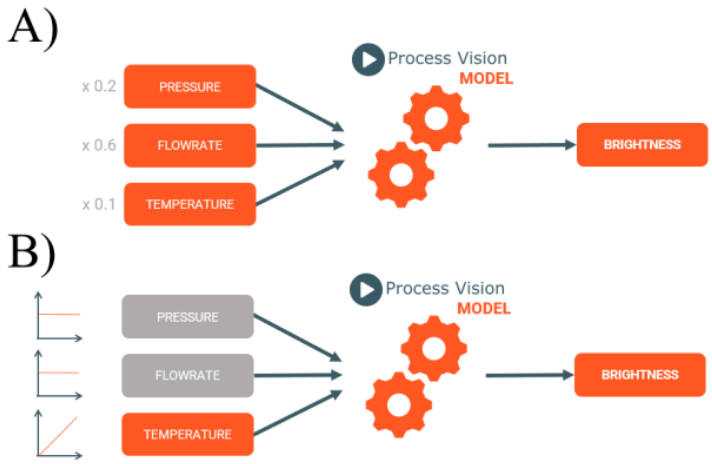

Figure 13 – A) Simplified model of how features can contribute differently to predictions, B) Simplified model of how sensitivity can be analysed through selected input variation

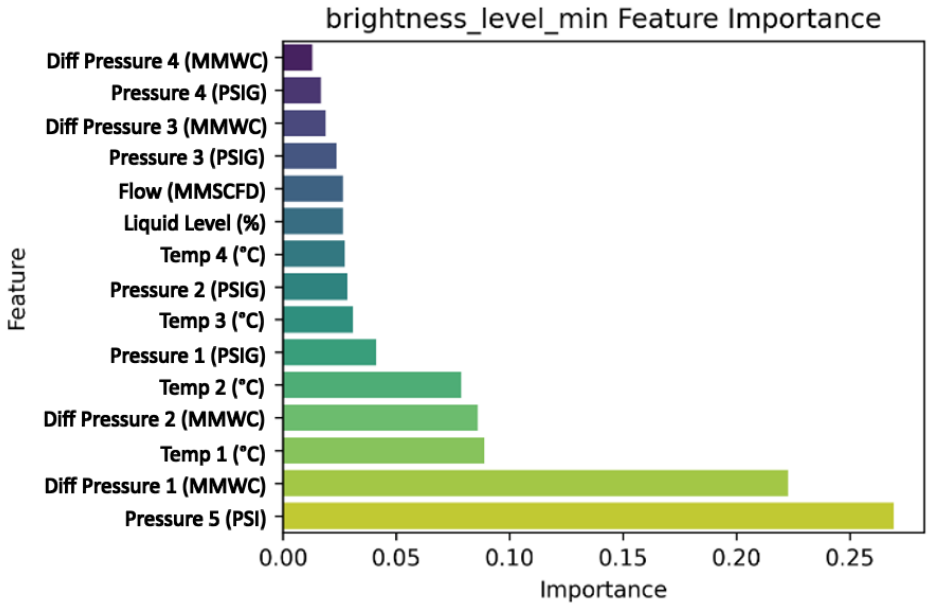

Figure 12 – Feature Importance for Laplacian Variance Vs Process Parameters

Feature-importance analysis identifies which process variables exert the greatest influence on predicted brightness. As depicted conceptually in Figure 13 A, each input contributes with different relative weighting to the model output. Quantitatively, the ranking in Figure 14 reveals that downstream differential pressure (Pressure 5) and upstream flow indicators (Diff Pressure 1 and Diff Pressure 2) dominate model predictions,

accounting for over 45 % of total feature importance. Temperature-related variables, including Temp 1 and Temp 2, also contribute significantly, indicating the thermal sensitivity of mist or droplet formation in the gas stream.

These results reinforce established thermodynamic understanding that mist generation intensifies when gas temperature approaches dewpoint conditions [2]. Pressure and flow effects further suggest that hydraulic loading in separators and pipe segments directly influences entrainment intensity. The combined interpretation is that the RF model captures both the steady-state driving conditions (via temperature and pressure) and the dynamic transport mechanisms (via flow-rate changes) that determine the visibility of contamination.

Interpretation and Operational Implications

The model's overall performance demonstrates that process data can be used as a reliable proxy for optical contamination behavior, providing predictive capability even when direct imagery is unavailable. At the same time, regions of reduced accuracy, particularly during unsteady operations, highlight the limits of static process variables in capturing transient fluid-dynamic effects. From an operational perspective, this balance between predictive reliability and residual deviation offers two valuable insights:

1. Predictive stability: under steady-state operation, process-driven models can estimate contamination risk with high confidence.
2. Event detection: large residual deviations between predicted and actual brightness can act as an early indicator of unmodelled events or system upsets.

Together, the feature-importance and sensitivity results provide both quantitative ranking and physical justification for the variables that most strongly influence contamination. These findings support the integration of data-driven models within digital-twin environments to enhance gas-quality assurance and enable proactive intervention.

Discussion

The results presented in this paper demonstrate the feasibility and practical value of using process-camera data combined with machine-learning methods to interpret and predict contamination behavior in gas pipelines. The approach moves beyond conventional process analysis by directly linking the optical manifestation of mist and liquid carryover to measurable process variables. This integration represents a major advancement in how operators can understand and manage gas quality.

Interpreting Predictive Performance in an Operational Context

The predictive models developed in this paper show strong agreement between measured and predicted optical brightness, confirming that existing process instrumentation captures much of the thermodynamic and hydraulic behavior that governs contamination events. However, deviations between model output and observed brightness, especially during transient operating states, indicate that some contamination mechanisms remain uncaptured by process variables alone. Rather than representing model weakness, these discrepancies highlight the inherent limits of process data in describing two-phase flow phenomena.

From an operational standpoint, this limitation underscores the value of process cameras: without direct visual confirmation, critical transient events would remain invisible to operators. Process imagery thus acts as both a validation and an extension of process instrumentation, providing the ground truth against which all inferred data can be assessed.

Integration Into Integrity Management Programs

The ASME B31.8S (2022) standard [16] mandates that operators identify, assess, and mitigate all potential threats to pipeline integrity, including those arising from fluid-phase behaviour. Liquid carryover, as extensively documented in field studies [2], constitutes one of the most persistent and under-reported integrity threats. When uncontrolled, it accelerates corrosion, erodes compressor blades, and contributes to unplanned downtime.

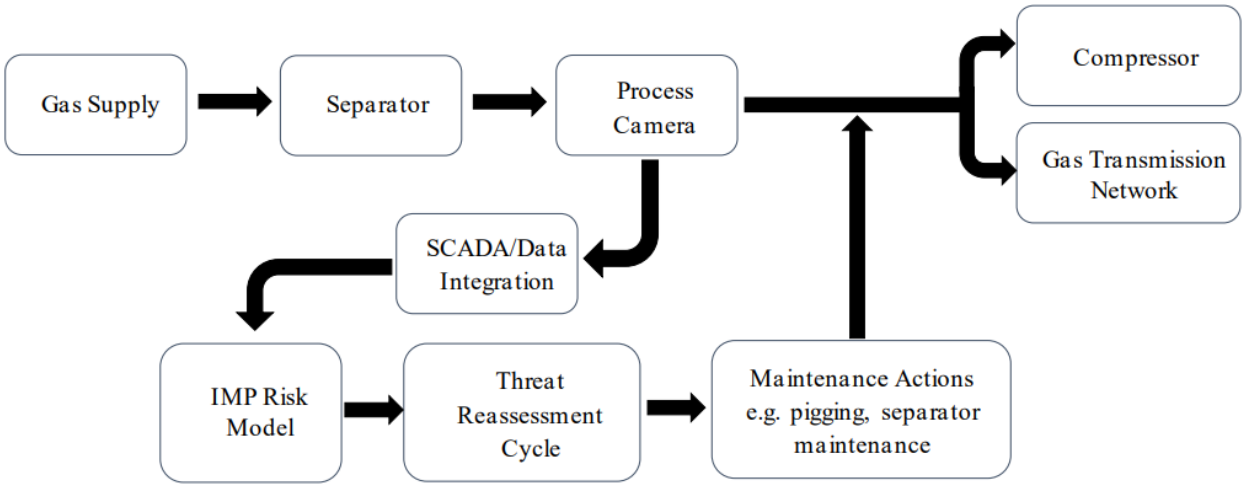

Figure 15 – Integrated Gas Processing and Maintenance System with Risk-Based Feedback. This flowchart illustrates a gas processing pipeline from supply to transmission, incorporating a process camera and SCADA/data integration for monitoring, with a feedback loop utilizing an IMP risk model and threat reassessment to inform maintenance actions that influence the ongoing process.

As shown in Figure 15, process-camera systems can be integrated into existing Integrity Management Program (IMP) frameworks through the site's SCADA or data-management architecture. Each carryover event detected by the camera can automatically feed into the operator's risk model, triggering event reassessment and targeted maintenance actions such as separator inspection or pigging campaigns. This integration not only enhances compliance with ASME B31.8S but also enables dynamic, evidence-based risk mitigation.

Where traditional IMP data sources rely primarily on pressure, flow, and corrosion-rate indicators, process-camera imagery adds a direct physical observation layer, converting what was previously inferred risk into observable fact. The result is a closed-loop monitoring system in which every visual detection contributes to a continuously updated threat model.

The Role of Process Cameras in Digitalisation and Root-Cause Analysis

The findings of this paper reinforce that digital twins and AI-based models are only as valuable as the data that feed them. Process cameras such as LineVu provide the missing data modality

required to validate model predictions and ground them in observable phenomena. Without visual verification, many apparent "steady-state" periods would conceal low-level mist activity or separator malfunction, leading to incorrect model calibration and false confidence in data quality.

In effect, process cameras transform the model from a theoretical exercise into a verifiable operational tool. They enable root-cause analysis that ties every statistical anomaly to a physical event, closing the feedback loop between data science and engineering. This synergy between visual data, machine learning, and traditional process metrics represents the practical realization of digitalization in gas-quality management. While integration with existing IMP and SCADA systems is technically straightforward, data-latency and historian sampling rates may limit real-time performance on some assets

Conclusion

This paper presents a data-driven methodology for understanding and predicting contamination in gas pipelines by linking process measurements with real-time visual data from process cameras. Using video brightness as a proxy for mist and liquid carryover, a Random-Forest framework was developed to model the optical behavior of the gas stream directly from process variables. The approach demonstrated that process data contain

sufficient information to explain much of the optical response associated with contamination, providing a new perspective on the thermodynamic and hydraulic mechanisms behind mist formation.

The results confirmed that models built on this framework achieve strong predictive performance while remaining interpretable for engineering analysis. Feature-importance and sensitivity analyses identified temperature margin, differential pressure, and flow rate as the most influential parameters, consistent with established physical understanding. Equally, instances of reduced model performance revealed the limits of process data alone and highlighted the diagnostic value of direct visual observation. This dual insight, predictive capability and deviation-based event detection, provides operators with a more comprehensive picture of system behaviour.

The integration of process-camera data with machine learning introduces an analytical dimension that was previously unavailable through instrumentation alone. Without direct visual feedback, subtle liquid carryover events and early mist formation remain hidden, preventing effective root-cause analysis and accurate validation of process assumptions. Process cameras bridge this gap, providing the visual ground truth needed for confident interpretation of complex process interactions.

As demonstrated through this work, combining process and visual data represents a practical and scalable enhancement to Integrity Management Programs and introduces smart alarms. Process-camera insights can be automatically integrated into operator SCADA systems, linking contamination events to maintenance

decisions, separator inspection, or pigging schedules. This supports compliance with frameworks such as ASME B31.8S, while enabling proactive risk management and operational optimization.

Ultimately, the fusion of optical and process intelligence forms the basis for predictive, data-driven gas-quality assurance. It enables continuous visibility of pipeline conditions, turning what was once reactive fault-finding into proactive contamination prevention. This integration of

machine learning, process analytics, and visual verification marks a decisive step toward digitalization and operational excellence in the natural gas industry. Future work will focus on multi-site retraining and adaptive model updating to further enhance predictive reliability.

Acknowledgements

Harry Thorpe and Vincent Strong – Process Vision

References

[1] “Stockwell, P. and Parker, S. (2024) ‘Errors in Hydrocarbon Dewpoint Can Lead to Large Losses for Gas Processing Plants’, Laurance Reid Gas Conditioning Conference, Norman, OK.”.

[2] “Stockwell, P. (2024) ‘Rethinking Gas Quality Measurements: The Case for Direct Observation’, World Pipelines, December issue, pp. 22–29.”.

[3] “Chanda, S., Gupta, K., Thorpe, H. and Stockwell, P. (2025) ‘Unlocking Business Value by Introducing Process Cameras’, SPE-229273-MS, ADIPEC Conference, Abu Dhabi.”.

[4] “API (2023) Manual of Petroleum Measurement Standards, Chapter 14.1 – Natural Gas Sampling. American Petroleum Institute, Washington DC.”.

[5] “ISO (2022) ISO 10715:2022 Natural gas – Sampling guidelines, International Organization for Standardization, Geneva.”.

[6] “Amine Experts (2022) Amine Plant Failure Survey Report.”.

[7] “Health and Safety Executive (2018) Compressor Dry Gas Seal Failure Analysis Report, HSE UK.”.

[8] “Process Vision (2025) SDS-0007-MA-2 LineVu Discovery Product Sheet, Process Vision Ltd., Basingstoke.”.

[9] “Usiabulu, G.I., et al. (2023) ‘Early Leak Detection in Gas Processing Plants Using Machine Learning’, Journal of Engineering Research and Reports, 25(6), pp. 18–27.”.

[10] “Qiao, Y., Zhang, X. and Sun, C. (2022) ‘Deep learning-based flow regime identification in multiphase pipelines using video data’, Journal of Natural Gas Science and Engineering, 104: 104–118.”.

[11] “AIChE Journal (2024) Special Issue: Artificial Intelligence in Process Engineering, Vol. 70, No. 5.”.

[12] “Breiman, L., 2001. Random forests. Machine Learning, 45(1), pp.5–32.”.

[13] “Probst, P., Boulesteix, A.-L. and Hornik, K. (2017) ‘To tune or not to tune the number of trees in random forest?’, Journal of Machine Learning Research, 18, pp. 1-18.”.

[14] “Hox, J., Moerbeek, M. and van de Schoot, R. (2017) Multilevel Analysis: Techniques and Applications, 3rd edn. Routledge, New York.”.

[15] “Hernandez, R., Gao, Y. and Zitney, S. (2021) ‘Data-Driven Modelling for Oil and Gas Production Systems’, Energy and AI, 4, p. 100069.”.

[16] “American Society of Mechanical Engineers (ASME) (2022) ASME B31.8S-2022: Managing System Integrity of Gas Pipelines. New York: ASME.”.