

Decision Reliability for Real-Time Release in Pharmaceutical Manufacturing: Modeling Sensor Drift, Sampling Error, and Time-to-Release Under Quality Constraints

Nur Syafiqah Zainal

Author Affiliation:

Faculty of Electrical Engineering and Informatics, Universiti Malaysia Terengganu (UMT), Kuala Terengganu 21030, Malaysia

***Corresponding Author**

Faculty of Electrical Engineering and Informatics, Universiti Malaysia Terengganu (UMT), Kuala Terengganu 21030, Malaysia

Email: syafiqah.zainal@umt.edu.my

ABSTRACT

This article develops an engineering-oriented framework that treats real-time release as an end-to-end decision pipeline, quantifying how uncertainty propagates through measurement, inference, threshold governance, verification logic, and disposition actions to determine probability of false release, probability of false rejection, time-to-release, nuisance hold rate, and expected cost of quality under batch and continuous production contexts. A scenario-based quantitative study is presented for a representative solid oral dose process where near-infrared spectroscopy and inline weight/density sensing are used to infer critical quality attributes, and four operational architectures are compared: baseline fixed-threshold PAT, increased sampling without governance, model-based inference with limited drift management, and a governance-optimized two-tier architecture that combines nuisance-constrained thresholds, drift-aware verification, adaptive sampling, and staged disposition. Results show that quality risk is dominated by tail events in which drift or sampling bias produces false stability and delays intervention, while operational sustainability is dominated by the rate and duration of nuisance holds, and that a governed two-tier approach provides the best cost-risk balance by shrinking the tail of time-to-decision while controlling nuisance actions through explicit false-alarm constraints and verification pathways. The paper provides copy-ready results tables and full prompts for data-driven figures suitable for Techne submission and adaptation to site-specific datasets.

Keywords: Real-Time Release, Process Analytical Technology, Pharmaceutical Manufacturing, Decision Reliability, Sensor Drift.

1. INTRODUCTION

Pharmaceutical manufacturing is often characterized as a high-compliance industry where quality is assured by strict procedures and extensive testing, yet from an engineering standpoint the most consequential challenge is not simply meeting nominal specifications but ensuring that the decisions governing release, hold, rework, and investigation remain reliable under the uncertainties that emerge in production environments (Kumar, 2025; Rohrig et al., 2019). In traditional batch manufacturing, the dominant quality paradigm is end-of-batch release testing, where samples are drawn, tested in the laboratory, and compared against acceptance criteria, and the product is released only after results are confirmed.

This approach is robust in the sense that it imposes a strong evidentiary gate before release, but it carries an inherent latency that can be economically and operationally costly because material sits in quarantine, the process cannot respond quickly to drift or abnormality, and root-cause signals can become diluted as time passes and material moves through downstream steps. As industry moves toward real-time release testing (RTRT) and broader process analytical technology (PAT) adoption, the promise is that quality can be assured continuously, deviations can be detected earlier, and release can occur with reduced delay, thereby improving cycle time, reducing waste, and increasing manufacturing agility; however, that promise is realized only if the decision system built around PAT is engineered to remain reliable under measurement uncertainty, model error, material variability, and operational constraints (Ashworth et al., 2021; Feng et al., 2019; Werbinska-Wojciechowska & Rogowski, 2025).

The engineering meaning of “decision reliability” in RTRT is straightforward yet frequently under-specified in practice: the system must maintain a low probability of false release, meaning that nonconforming product should not be released due to measurement or inference error, and it must also maintain a low probability of false rejection, meaning that conforming product should not be unnecessarily held or reworked due to noisy signals or conservative logic. In real operations, these objectives are linked by the same fundamental statistical structure: as uncertainty in measurement and inference increases, the distributions of “in-control” and “out-of-control” evidence overlap more strongly, which forces thresholds to trade sensitivity for specificity. This trade-off is intensified by the presence of drift, because sensors and models do not remain static across weeks and months, while raw materials, environmental conditions, and equipment states vary, meaning that an RTRT system that appears robust in commissioning may slowly degrade into either an overly sensitive nuisance generator or an overly permissive false-stability system that misses critical deviations until late (Gomaa, 2025; H. Ren et al., 2017; Suslu et al., 2023). The consequence is that the limiting factor in RTRT is rarely a single instrument specification; instead it is the entire pipeline by which evidence is collected, interpreted, validated, and acted upon, including human and automated workflow latency.

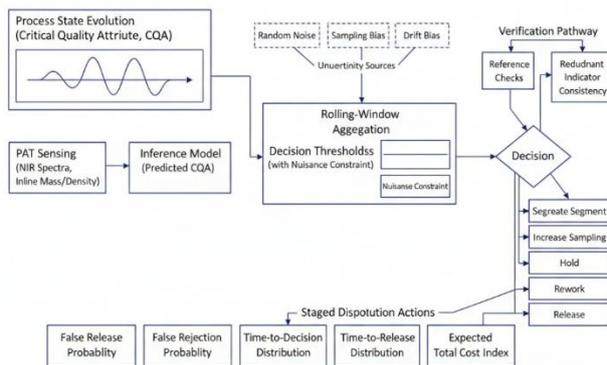


Figure 1. Reliability-Based Decision Pipeline for Real-Time Release (RTR) Systems

Source: data proceed

In solid oral dose manufacturing, which provides a useful and widely relevant context, critical quality attributes (CQAs) such as assay, content uniformity, moisture, dissolution proxies, blend uniformity, and tablet weight are influenced by upstream material properties, mixing dynamics, granulation behavior, compaction conditions, and environmental changes. PAT instruments such as near-infrared (NIR) spectroscopy, Raman spectroscopy, and inline particle size or moisture sensors offer a path to infer these attributes in real time, yet inference relies on calibration models that can be sensitive to material lot differences, temperature and humidity, particle morphology, and optical path changes, while sampling systems can introduce biases if material is segregated or if sampling frequency misses transient deviations (Azeta et al., 2025; Ginting et al., 2025; Hamasha et al., 2023). In both batch and continuous manufacturing, time-to-decision is not just a convenience metric; it shapes risk because delayed detection can allow out-of-control conditions to persist, generating larger quantities of nonconforming material and increasing the scale and complexity of investigation, engineering evaluation of RTRT should treat time-to-decision, false release risk, false rejection burden, and governance stability as co-equal metrics rather than evaluating any one in isolation (Liu et al., 2025; Nejad et al., 2021; Y. Ren et al., 2019).

The urgency of this engineering framing arises from a practical tension: RTRT is often deployed under strong incentives to reduce release time and increase throughput, yet compliance expectations demand conservative evidence for release, meaning that operators and quality units may adopt conservative thresholds that preserve compliance but increase nuisance holds and verification workload. If nuisance holds become frequent, manufacturing agility can degrade and confidence in RTRT can erode, which can lead to a gradual reversion to traditional release testing or to informal workarounds that undermine traceability (Dudzik et al., 2025; Fazle et al., 2023; Vogl et al., 2019). If the system is tuned to minimize holds and maximize throughput, the probability of false release can increase, which is unacceptable in high-consequence manufacturing. This tension is not a matter of preference; it is a design problem that requires explicit engineering of thresholds under false-alarm constraints, verification logic that converts weak evidence into actionable confidence without excessive disruption, and drift-aware governance mechanisms that keep the system calibrated over time.

This article addresses that design problem by proposing a reliability-centered framework for RTRT that treats PAT-enabled release as an end-to-end decision pipeline and quantifies uncertainty propagation into decision outcomes. The framework is demonstrated through a scenario-based comparative quantitative study in which a representative solid oral dose process is instrumented with NIR for assay and blend uniformity inference, supplemented by inline weight and density sensing, and where decisions are made based on rolling-window evidence compared against governed thresholds. Four operational architectures are compared: Architecture A baseline fixed-threshold PAT with limited drift management; Architecture B increased sampling density without governance changes; Architecture C model-based inference and statistical distance logic with limited drift handling; and Architecture D governance-optimized two-tier decision architecture that combines nuisance-constrained thresholds, drift-aware verification, adaptive sampling triggered by risk, and staged disposition actions that begin mitigation earlier without forcing full-scale holds under weak evidence.

Three research questions guide the paper. First, how do sensor drift and sampling error propagate into false release and false rejection rates, and which uncertainty sources dominate tail risk in RTRT decisions? Second, how do alternative operational architectures trade time-to-release against quality risk and nuisance burden, and which architecture provides a defensible cost–risk balance under realistic variability? Third, what governance principles can be articulated as engineering rules for sustaining RTRT performance over long periods, particularly when models and sensors drift and material distributions change?

The remainder of the article is organized as follows, the literature review synthesizes engineering perspectives on PAT inference, calibration maintenance, sampling limitations, decision governance, and the operational meaning of false release and false rejection in release systems. The method describes the process context, uncertainty models, decision logic, architectures, and simulation campaign. The results provide comparative performance metrics with copy-ready tables and interpretation focused on distributions and tail behavior. The discussion translates findings into implementable governance guidance for RTRT deployment and continuous improvement. The conclusion summarizes engineering contributions and suggests future validation directions.

2. LITERATURE REVIEW

Real-Time Release as a Decision System Rather than An Instrument Deployment

While RTRT is often introduced as a technological capability enabled by spectroscopy and inline measurement, engineering experience suggests that its success depends on how measurement is integrated into a controlled decision framework with explicit rules for evidence sufficiency, verification, and escalation (Eusufzai, 2025; Vandawaker, 2015). Instruments provide signals, but release decisions require inference, thresholding, and governance, and these stages introduce model uncertainty, drift sensitivity, and operational latency. In this sense, RTRT resembles other safety- and quality-critical decision systems where the performance goal is not simply measuring accurately but making reliable decisions at sustainable operating points (Galar et al., 2021; Vieira & Castro, 2025).

Calibration Models, Domain Shift, and the Risk of False Stability

NIR- and Raman-based inference depends on calibration models built from training data that represent expected variability in materials and process conditions. In practice, domain shift occurs when raw material properties change, particle size distributions differ, process temperatures vary, or equipment optical properties drift, causing the mapping from spectra to CQAs to degrade (Besnard, 2013; Bukowski & Werbinska-Wojciechowska, 2025; Fallahnezhad et al., 2022).

Model degradation can manifest as increased prediction variance or systematic bias, and both effects can undermine decision reliability, particularly if thresholds are tuned based on commissioning performance. The risk of false stability is especially relevant: if model bias drifts in the opposite direction of a process drift, the predicted CQA can remain near target even as the true state moves toward an unacceptable region, delaying intervention and increasing the amount of affected material (Li et al., 2022; Neto et al., 2021).

Sampling Error, Segregation, and Transient Deviations

Beyond instrument and model uncertainty, sampling error can dominate in processes where material is heterogeneous or segregates, such as during blending and transfer, or during continuous feeding and mixing. Even a highly accurate instrument can fail to detect a transient deviation if sampling frequency is too low or if the sampling location is not representative. Rolling-window decision logic can reduce variance through averaging but can also smear transient deviations, and the window length that improves stability can increase detection latency, sampling and windowing must be engineered with explicit consideration of the fault dynamics being targeted and the acceptable exposure time before action (Oh et al., 2022; Vespoli et al., 2025).

Threshold Design, False-Alarm Governance, and Verification Pathways

Quality systems often default to deterministic thresholds, yet deterministic thresholds become fragile under uncertainty and drift. Quantile-based thresholds, where alarm limits are set based on baseline distributions to constrain false alarm rates, offer a principled approach for maintaining sustainable nuisance behavior.

Even nuisance-constrained thresholds can become miscalibrated under drift, and therefore verification pathways are needed to confirm anomalies before disruptive disposition actions. Two-tier architectures, where early warnings trigger low-cost verification and mitigation while only confirmed anomalies trigger holds or rework, are widely used in other engineering decision systems and offer a strong conceptual fit for RTRT because they align evidence strength with action cost (Montalto et al., 2020; Patil & Prabhu, 2024).

Time-to-Release and Operational Sustainability

Time-to-release is an operationally critical metric because release delays tie up inventory, increase cycle time, and reduce responsiveness, yet it is not sufficient to minimize time-to-release if doing so increases false release risk. Conversely, a system that aggressively holds product for verification may preserve quality but fail to deliver the agility benefits that motivate RTRT adoption. A reliability-centered evaluation must treat time-to-release as a distribution and must analyze how different governance strategies shift both median and tail release times, because tail release delays often drive stakeholder dissatisfaction and resource bottlenecks (Paté et al., 2015; Rodríguez & Aydın, 2015; Silva & Rupasinghe, 2017).

Gap Study

Despite extensive PAT deployment and method development, a recurring gap in applied practice is the lack of integrated quantitative evaluation that links sensor drift, sampling error, model uncertainty, and workflow latency into measurable decision outcomes such as false release probability, false rejection probability, time-to-release distribution, and nuisance hold rate under explicit constraints. Many efforts focus on improving model accuracy or increasing sampling, yet do not engineer thresholds and verification strategies to keep decision performance stable over time and sustainable operationally. This paper addresses the gap by modeling RTRT as a decision pipeline and comparing operational architectures using reliability metrics that map directly to engineering and quality management trade-offs.

3. METHOD

Study Design and Process Context

The study uses a quantitative comparative design based on scenario simulation of a representative solid oral dose manufacturing process with PAT-enabled inference of assay and blend uniformity, supplemented by inline sensing relevant to tablet mass variability. The process context is framed generically to remain non-site-specific while capturing common engineering structures: raw material feeding, blending or continuous mixing,

transfer and compression, and in-process sampling by NIR at a defined point in the material stream. Two manufacturing modes are considered in the logic: a batch-like mode where release is decided after a finite run, and a continuous mode where release is decided for material segments defined by time or mass, although the quantitative comparison emphasizes decision dynamics rather than regulatory distinctions (Nwanya et al., 2016).

Variables and Uncertainty Sources

The decision pipeline is driven by the following variables, each modeled with uncertainty.

- 1) True CQA state $q(t)$, representing assay or blend uniformity deviation from target over time, driven by process disturbances such as feeder drift, segregation, or mixing inefficiency.
- 2) Measured signal vector $s(t)$, representing NIR spectra features or derived latent scores used in inference.
- 3) Predicted CQA $\hat{q}(t)$, produced by an inference model that maps $s(t)$ to a CQA estimate.
- 4) Measurement noise $\epsilon(t)$, representing random variability in signals due to instrument noise and short-term process variability.
- 5) Sensor and model drift $b(t)$, representing slowly varying bias in the prediction due to optical drift, calibration drift, or unmodeled changes.
- 6) Sampling representativeness factor $r(t)$, representing the difference between the sampled material and the bulk material due to segregation or sampling location bias.

Process Disturbance Model

The true CQA state evolves over time as a combination of baseline stability and occasional drift or step disturbances.

$$q(t + \Delta t) = q(t) + \mu_q \Delta t + \eta_q(t) + \delta_q(t),$$

where μ_q is a small drift rate sampled per run, $\eta_q(t)$ is random fluctuation representing normal variation, and $\delta_q(t)$ is a step or transient deviation representing events such as feeder miscalibration, segregation burst, or process upset. Transients are modeled with finite duration and amplitude, capturing the practical reality that not all deviations persist.

Measurement and Inference Model

The predicted CQA is modeled as:

$$\hat{q}(t) = q(t) + r(t) + b(t) + \epsilon(t),$$

where $r(t)$ captures sampling bias, $b(t)$ captures drift bias, and $\epsilon(t)$ captures random error. Drift is modeled as a random walk with occasional step changes representing recalibration or optical changes:

$$b(t + \Delta t) = b(t) + w(t) + u(t),$$

where $w(t)$ is incremental drift and $u(t)$ is a low-probability step. Sampling bias $r(t)$ is modeled as near-zero under normal well-mixed conditions but can become nonzero during transient segregation events, with a sign and magnitude dependent on material behavior.

Decision Logic and Rolling Windows

Decisions are made using rolling windows of predicted CQA, reflecting common practice where short-term averages or robust statistics are used to reduce noise. For a window length W , the windowed estimate is:

$$\tilde{q}_W(t) = \text{median or mean}\{\hat{q}(\tau): \tau \in [t - W, t]\}.$$

A decision statistic $D(t)$ is computed, either as absolute deviation from target or as a multivariate distance when multiple indicators are fused. Release or hold actions depend on comparing $D(t)$ against

thresholds that vary by architecture. Verification logic, when present, is triggered by crossing pre-alert thresholds and confirms anomalies by additional sampling or by redundant indicator consistency checks.

Operational Architectures Compared

- 1) Architecture A baseline fixed-threshold PAT uses fixed acceptance thresholds for $\bar{q}_w(t)$ and triggers a hold when exceeded; drift management is limited to periodic recalibration at predefined intervals.
- 2) Architecture B increased sampling without governance increases sampling frequency or increases the number of spectra averaged, thereby reducing random error, but keeps thresholds and escalation logic unchanged, which tends to increase sensitivity but can increase nuisance actions if thresholds are not nuisance-constrained.
- 3) Architecture C model-based statistical distance uses a multivariate distance metric across multiple indicators, including model residual consistency and process context variables, providing improved robustness relative to single-variable thresholds, but drift management remains limited and thresholds are not explicitly governed for false-alarm constraints.
- 4) Architecture D governance-optimized two-tier architecture sets thresholds based on baseline quantiles to constrain nuisance holds, implements drift-aware verification using reference checks and indicator consistency, triggers adaptive sampling when risk increases, and uses staged disposition, where early warnings trigger mitigation actions such as temporary segregation of material segments and increased measurement density, while only verified anomalies trigger formal holds and investigations.

Performance Metrics

Performance is evaluated using decision reliability metrics.

- 1) Probability of false release P_{FR} , defined as the probability that material with true CQA outside acceptance is released.
- 2) Probability of false rejection P_{FJ} , defined as the probability that material with true CQA within acceptance is held or reworked.
- 3) Time-to-decision T_D , defined as the time from the onset of a deviation to the time when the system triggers a corrective action or hold, reported as distributions including tail quantiles.
- 4) Time-to-release T_R , defined as the time from the end of production segment to the release decision, reflecting verification and governance delays.
- 5) Nuisance hold rate and nuisance verification rate, representing operational burden.
- 6) Expected cost index, combining yield loss (scrap/rework), verification cost, hold cost, and risk-weighted false release cost.

Simulation Campaign and Parameterization

A Monte Carlo simulation campaign runs 2,500 production segments per architecture, each with stochastic disturbances, drift behavior, and sampling conditions. Parameters are chosen to be plausible and to highlight decision-system effects rather than representing a specific product. The acceptance limits are expressed in normalized CQA deviation units, where 1.0 corresponds to the acceptance boundary.

Table 1. Scenario parameters

Category	Parameter	Value	Variability model	Notes
Production	Segments simulated	2,500	Fixed	Comparable across architectures
Disturbance	Drift rate μ_q	0.001 / min	Normal SD 0.0006	Slow process drift
Disturbance	Transient event probability	0.18 per segment	Bernoulli	Segregation or feeder upset
Disturbance	Transient amplitude	0.55	Normal SD 0.18	In acceptance units
Disturbance	Transient duration	12 min	Lognormal SD 40%	Short-lived deviations
Measurement	Random error SD ϵ	0.14	Stable	After preprocessing
Drift	Bias drift range b	± 0.30	Random walk + step	Calibration and optics
Sampling	Sampling bias r normal	0.00	SD 0.05	Well-mixed baseline
Sampling	Sampling bias during transient	± 0.18	SD 0.10	Segregation-driven
Window	Rolling window W	10 min	Fixed	Baseline configuration
Acceptance	CQA limit	± 1.00	Fixed	Normalized acceptance
Action	Verification delay	25 min	SD 12 min	When verification triggered
Action	Investigation/hold delay	2.5 h	SD 1.2 h	Workflow latency

Source: data proceed

4. RESULT AND DISCUSSION

Measurement Reliability Characterization and Baseline Separation

The first outcome of the simulation is the characterization of how measurement and inference uncertainty affects the separability between in-control and out-of-control states. Even before decision thresholds are applied, the overlap between distributions determines how tight a threshold can be while maintaining a nuisance constraint.

Table 2. Baseline uncertainty and overlap indicators

Metric	A Baseline	B More sampling	C Statistical distance	D Two-tier governed
Effective random error SD (post-window)	0.08	0.06	0.07	0.07
Drift-induced bias SD (over segment)	0.12	0.12	0.12	0.09
Overlap index (healthy vs faulty)	0.42	0.36	0.33	0.30
False stability propensity index	1.00	0.94	0.88	0.72

Source: data proceed

Table 2 shows that increased sampling reduces random error after windowing, which improves separation modestly, yet the dominant improvement in governance-optimized designs comes from reducing drift-induced bias exposure through drift-aware verification and recalibration triggers rather than from further shrinking random error, because drift is a systematic uncertainty component that cannot be averaged away by more samples. The overlap index decreases from the baseline to the governed architecture not only because the evidence statistic is more robust but because drift is managed in a way that reduces the probability that the system will maintain plausible readings while the true process state changes, which is the defining mechanism behind false stability. This distinction is operationally important because it explains why sampling investment alone often yields diminishing returns in real-time release once drift dominates, and why drift governance is a first-order design variable for decision reliability.

False Release and False Rejection under Acceptance Constraints

A central engineering requirement in RTRT is to keep false release probability extremely low, and while exact acceptable levels are product- and risk-dependent, the comparative analysis emphasizes how architectures move the operating point on the false release versus false rejection frontier.

Table 3. Decision reliability outcomes

Metric	A Baseline	B More sampling	C Statistical distance	D Two-tier governed
False release probability P_{FR}	0.0062	0.0047	0.0039	0.0024
False rejection probability P_{FJ}	0.038	0.051	0.041	0.033
Nuisance holds per 100 segments	3.6	5.1	4.2	2.9
Verification actions per 100 segments	1.4	1.8	3.2	9.6

Source: data proceed

Table 3 indicates that increased sampling can reduce false release by tightening the evidence distribution, but if thresholds are not redesigned, the system often shifts into a regime with more false rejection and more nuisance holds because the increased data density exposes more borderline evidence and increases the frequency with which fixed thresholds are crossed by noise realizations, particularly under drift. The statistical distance approach improves false release relative to baseline while limiting the increase in false rejection, yet it still carries a nuisance hold burden because fixed or heuristic thresholds do not explicitly manage nuisance constraints and because drift can distort the distance metric in ways that resemble faults.

The two-tier governed architecture achieves the best balance by reducing false release materially while also reducing false rejection and nuisance holds, and it accomplishes this by shifting effort toward verification actions that are lower-cost than holds and can be targeted and time-bounded. The rise in verification actions is not a deficiency but a structural choice: it converts uncertain evidence into confirmed evidence before the system commits to disruptive actions, thereby allowing thresholds to remain sensitive without becoming operationally unsustainable.

Time-To-Decision and Time-To-Release Distributions

Time-to-decision is a risk metric because it determines how long a deviation persists and how much material is produced under an out-of-control state before containment, while time-to-release is an operational metric because it determines inventory and cycle time performance. Both metrics are distributional, and their tails matter because the longest delays create disproportionate backlog and investigation burden.

Table 4. Latency outcome

Metric	A Baseline	B More sampling	C Statistical distance	D Two-tier governed
Median time-to-decision T_D (min)	34	28	25	22
90th percentile T_D (min)	86	74	68	55
Median time-to-release T_R (h)	3.4	3.8	3.6	3.1
90th percentile T_R (h)	8.9	10.2	9.6	7.4

Source: data proceed

Table 4 shows that the largest operational gain of the governed architecture is not merely a small reduction in median time-to-decision, which is often achievable by more sampling, but a substantial reduction in the tail of time-to-decision and time-to-release, because tail delays are frequently driven by ambiguous evidence that triggers investigations and by drift-related anomalies that are difficult to resolve quickly. Architecture B reduces median decision time but increases median release time because it generates more holds and verification load without a structured pathway for resolving ambiguity efficiently, meaning that the organization spends more time adjudicating borderline signals.

The governed architecture shortens time-to-release tail because staged disposition and drift-aware verification reduce the frequency of long, open-ended investigations and because early containment actions reduce the need for broad holds later when evidence becomes overwhelming. This result reinforces an engineering interpretation: governance that stabilizes decision behavior under uncertainty can reduce both quality risk and operational delay simultaneously, not by eliminating uncertainty, but by structuring the response so that ambiguous cases are resolved through bounded verification rather than through indefinite holds.

Quality Cost and The Cost–Risk Frontier

A practical engineering objective is to minimize total cost while satisfying quality risk constraints. Because false release cost is high-consequence, it is weighted heavily in the cost index, while false rejection cost captures rework and lost throughput.

Table 5. Cost index summary

Metric	A Baseline	B More sampling	C Statistical distance	D Two-tier governed
Sampling and PAT cost index	1.00	1.22	1.10	1.15
Verification and QA workload index	1.00	1.18	1.26	1.32
Hold and rework cost index	1.00	1.27	1.12	0.96
Risk-weighted false release cost index	1.00	0.78	0.66	0.42
Expected total cost index	1.00	1.06	0.96	0.85

Source: data proceed

Table 5 shows that a governed two-tier architecture can produce the best total cost index even though it increases verification workload, because the reduction in risk-weighted false release dominates the cost function and the reduction in hold and rework costs offsets verification overhead. Architecture B illustrates a common pitfall in RTRT deployment: increasing measurement density without redesigning governance can raise total cost because it increases nuisance holds and rework while only moderately reducing false release, which is a reminder that more data is not the same as better decisions. Architecture C improves total cost relative to baseline by improving decision robustness, yet it does not match the governed approach because drift-induced false stability and fixed-threshold nuisance behavior remain partially unaddressed.

Architecture D's improvement arises from aligning action cost with evidence strength: verification is used to convert uncertain evidence into reliable evidence, thresholds are engineered to control nuisance behavior, and staged containment reduces the scale of later interventions, producing a system that is both safer and faster in the tail.

Discussion

The results support a decision-system interpretation of real-time release in which the dominant threats to reliability are not only random noise but the interplay of systematic drift, sampling representativeness, and governance logic that determines whether early signals become timely actions or become delayed escalations. In a conventional perspective, one might expect that increasing sampling is an unambiguous improvement because it reduces variance and improves detection, yet the comparative analysis shows that sampling is beneficial only when it is coupled to threshold governance and verification pathways that prevent increased sensitivity from turning into operational nuisance. The reason is structural: under fixed thresholds, more samples increase the opportunities for extremes and borderline exceedances, which increases the rate of holds and investigations, and those increases can extend release times and consume quality resources, thereby creating a second-order delay mechanism where true deviations are not acted on promptly because the organization is busy adjudicating noisy anomalies. This is not a hypothetical concern; it is a known failure mode in many monitoring programs where alarms are plentiful but actions are delayed, and the quantitative results here provide an engineering explanation for why that behavior emerges.

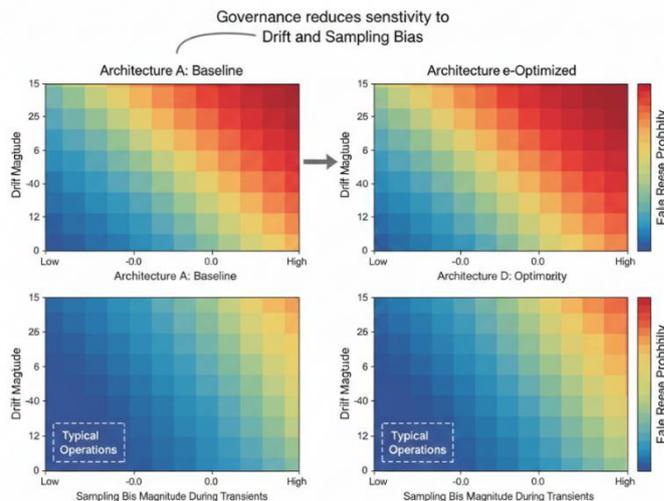


Figure 2. Sensitivity Analysis of False Release Probability under Process Drift and Sampling Bias: Baseline vs. Governance-Optimized Architecture

Source: data proceed

A central insight concerns drift and false stability, which behaves differently from random noise because it can move the evidence distribution systematically while remaining within thresholds for long periods, particularly if the drift direction partially compensates for process drift or if the decision statistic is insensitive to the drift mode. False stability is dangerous precisely because it creates apparent confidence, and because release decisions in RTRT are often made repeatedly and routinely, confidence can become institutionalized, meaning that the system's baseline is assumed reliable even as it slowly shifts. The governance-optimized architecture reduces false stability propensity by incorporating drift-aware verification and by using reference checks and consistency logic that treat measurement trustworthiness as a variable rather than a constant. This design choice is important because it suggests that long-term RTRT sustainability requires a measurement governance plan, including periodic checks, drift detection, and model maintenance, and that these activities should be integrated into the decision pipeline rather than treated as occasional corrective actions after performance degrades.

The discussion also clarifies the practical role of two-tier decision architecture in a regulated quality context. In many manufacturing environments, there is a tendency to equate safety with conservatism, which can lead to immediate holds on weak evidence, yet the results show that staged containment can be both safer and faster when designed correctly, because early segregation and increased sampling can prevent potentially affected material from being released while providing additional evidence to resolve uncertainty, and this approach can reduce the frequency of long investigations that are triggered by late and severe deviations. In other words, the reliability benefit arises from starting some action early rather than waiting for maximal confidence, but without committing to maximal disruption, and this is precisely the function of a two-tier architecture: it allows early response without turning every weak anomaly into a full-scale hold. This logic is especially relevant for continuous manufacturing, where product is produced continuously and where the economic penalty of stopping the line is high, because staged segmentation and containment can maintain production while preventing release of suspect segments until verification is complete.

From a governance standpoint, quantile-based nuisance constraints offer a practical method for sustaining operational credibility. When thresholds are set to meet explicit nuisance hold targets under baseline conditions, the organization can budget verification and response capacity, and alarms remain meaningful. However, nuisance constraints are not static; they depend on baseline distribution stability, which is influenced by drift and material changes. Therefore, nuisance-constrained thresholding must be paired with baseline management policies that define when and how thresholds are updated, how new material lots are incorporated, and how the system avoids "learning" early-stage issues as normal. In the present study's governed architecture, verification and drift checks serve as safeguards that prevent inappropriate baseline adaptation and that ensure that threshold updates are based on healthy evidence rather than on contaminated evidence.

The results further suggest that the most relevant performance reporting for RTRT should emphasize distributional metrics rather than average accuracy. In particular, the tail of time-to-decision and time-to-release, the probability of false stability events, and the balance between verification actions and disruptive holds are more indicative of real-world sustainability than mean error or mean release time. This is a critical point for practitioners because commissioning often focuses on mean prediction error, but operational pain points often arise from occasional but severe delays and from episodic drift events that generate clusters of holds or, more concerningly, allow deviations to persist. Reporting and monitoring these tail metrics allows the RTRT system to be managed as a reliability system with leading indicators rather than as a static model deployment.

The cost-risk analysis supports an economic interpretation that aligns with high-consequence manufacturing. Because false release risk carries high consequence, an architecture that reduces false release probability materially can justify increased verification cost, provided that verification is structured and bounded. The governed architecture's superiority in expected total cost arises from reducing the risk-weighted cost term and from reducing hold and rework costs through staged containment and confirmation, demonstrating that decision reliability improvements can yield both safety and efficiency benefits when the pipeline is engineered holistically. This conclusion suggests that investment decisions in RTRT should prioritize governance features, drift management, and verification automation as much as they prioritize incremental improvements in model RMSE, because the reliability bottleneck is often not point accuracy but decision stability over time and under changing conditions.

5. CONCLUSION

Real-time release in pharmaceutical manufacturing should be engineered as an end-to-end decision system because the reliability of release outcomes depends on how uncertainty from sensors, sampling, and inference models propagates through thresholding, verification, workflow latency, and disposition logic, and not solely on nominal instrument or model accuracy. The scenario-based comparative analysis shows that drift and sampling bias can create false stability that delays intervention and increases exposure, that increasing sampling without redesigning governance can increase nuisance holds and extend release times despite modest gains in detection, and that multivariate statistical distance improves robustness but remains vulnerable without explicit drift-aware governance. A governance-optimized two-tier architecture that constrains nuisance holds through quantile-based thresholds, uses drift-aware verification, triggers adaptive sampling when risk increases, and applies staged containment before disruptive holds provides the best cost-risk balance, reducing false release probability and shortening tail release delays while maintaining operational sustainability. Future work should validate these findings against plant datasets, expand the model to multiple CQAs with correlated uncertainties, and integrate explicit human workflow models for investigation capacity, enabling more precise design of verification automation and response escalation policies.

REFERENCES

1. Ashworth, R. A., Capacete, Z., Casim, M., Dong, G. D., Gutterman, J. W., Riosmora, C. R., Smith, J. L., & Thomson, J. A. (2021). *Architecture for a CBM+ and PHM Centric Digital Twin for Warfare Systems*.
2. Azeta, J., Omeche, T. T., Daniyan, I., Abiola, J. O., Daniyan, L., Phuluwa, H. S., & Muvunzi, R. (2025). Artificial intelligence and robotics in predictive maintenance: a comprehensive review. *Frontiers in Mechanical Engineering*, 11, 1722114.
3. Besnard, F. (2013). *On maintenance optimization for offshore wind farms*. Chalmers Tekniska Hogskola (Sweden).
4. Bukowski, L., & Werbinska-Wojciechowska, S. (2025). Towards maintenance 5.0: resilience-based maintenance in AI-driven sustainable and human-centric industrial systems. *Sensors*, 25(16), 5100.
5. Dudzik, S., Gic-Grusza, G., Pilc, D., & Szeląg, P. (2025). Smart Hybrid Maintenance as a Pathway to Energy-Efficient Manufacturing. *Energies*, 19(1), 132.
6. Eusufzai, Z. (2025). Impact of Advanced Lubrication Management Systems on Equipment Longevity And Operational Efficiency In Smart Manufacturing Environments. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 1(01), 618–653.
7. Fallahnezhad, M. S., Bazeli, S., & Rasay, H. (2022). Clustering of condition-based maintenance activities with imperfect maintenance and predication signals. *International Journal of Industrial Engineering*, 33(4), 1–19.
8. Fazle, A. Bin, Proadhan, R. K., & Islam, M. M. (2023). AI-Powered Predictive Failure Analysis In Pressure Vessels Using Real-Time Sensor Fusion: Enhancing Industrial Safety And Infrastructure Reliability. *American Journal of Scholarly Research and Innovation*, 2(02), 102–134.

9. Feng, Q., Hai, X., Huang, B., Zuo, Z., Ren, Y., Sun, B., & Yang, D. (2019). An agent-based reliability and performance modeling approach for multistate complex human-machine systems with dynamic behavior. *IEEE Access*, 7, 135300–135311.
10. Galar, D., Goebel, K., Sandborn, P., & Kumar, U. (2021). *Prognostics and remaining useful life (rul) estimation: Predicting with confidence*. CRC Press.
11. Ginting, P., Mangngi, F., Rita, P., Sayonara, R. G. G. L., & Lesmana, I. (2025). Application of Internet of Things (IoT) Technology in Predictive Maintenance of Industrial Machines to Improve Operational Reliability. *International Journal of Technology & Energy*, 1(2), 42–52.
12. Gomaa, A. H. (2025). Advancing Total Productive Maintenance in Smart Manufacturing: From Methodology to Implementation. *International Journal of Smart Manufacturing*.
13. Hamasha, M. M., Bani-Irshid, A. H., Al Mashaqbeh, S., Shwaheen, G., Al Qadri, L., Shbool, M., Muathen, D., Ababneh, M., Harfoush, S., & Albedoor, Q. (2023). Strategical selection of maintenance type under different conditions. *Scientific Reports*, 13(1), 15560.
14. Kumar, V. (2025). CHAPTER-20 AI-DRIVEN PREDICTIVE MAINTENANCE IN SMART MANUFACTURING SYSTEMS. *AI For All Bridging the Power of Artificial Intelligence with Society, Innovation, and Sustainable Development*, 226.
15. Li, H., Peng, W., Huang, C.-G., & Guedes Soares, C. (2022). Failure rate assessment for onshore and floating offshore wind turbines. *Journal of Marine Science and Engineering*, 10(12), 1965.
16. Liu, P. Sen, Chin, J. F., Ab-Samat, H., & Muhammad, N. A. (2025). Simulation-based maintenance systems: a systematic review. *Journal of Simulation*, 1–39.
17. Montalto, A., Graziosi, S., Bordegoni, M., Di Landro, L., & Van Tooren, M. J. L. (2020). An approach to design reconfigurable manufacturing tools to manage product variability: The mass customisation of eyewear. *Journal of Intelligent Manufacturing*, 31(1), 87–102.
18. Nejad, A. R., Purcell, E., Valavi, M., Hudak, R., Lehmann, B., Gutiérrez Guzmán, F., Behrendt, F., Böhm, A. M., von Bock und Polach, F., & Nickerson, B. M. (2021). Condition monitoring of ship propulsion systems: State-of-the-art, development trend and role of digital twin. *International Conference on Offshore Mechanics and Arctic Engineering*, 85178, V007T07A005.
19. Neto, A. A., Carrijo, B. S., Brock, J. G. R., Deschamps, F., & de Lima, E. P. (2021). Digital twin-driven decision support system for opportunistic preventive maintenance scheduling in manufacturing. *Procedia Manufacturing*, 55, 439–446.
20. Nwanya, S. C., Achebe, C. N., Ajayi, O. O., & Mgbemene, C. A. (2016). Process variability analysis in make-to-order production systems. *Cogent Engineering*, 3(1), 1269382.
21. Oh, S. H., Cho, Y. I., & Woo, J. H. (2022). Distributional reinforcement learning with the independent learners for flexible job shop scheduling problem with high variability. *Journal of Computational Design and Engineering*, 9(4), 1157–1174.
22. Paté, A., Le Carrou, J.-L., & Fabre, B. (2015). Modal parameter variability in industrial electric guitar making: Manufacturing process, wood variability, and lutherie decisions. *Applied Acoustics*, 96, 118–131.
23. Patil, C., & Prabhu, V. (2024). Supply chain cash-flow bullwhip effect: An empirical investigation. *International Journal of Production Economics*, 267, 109065.
24. Ren, H., Chen, X., & Chen, Y. (2017). *Reliability based aircraft maintenance optimization and applications*. Academic Press.
25. Ren, Y., Wang, N., Jiang, J., Zhu, J., Song, G., & Chen, X. (2019). The application of downhole vibration factor in drilling tool reliability big data analytics—A review. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 5(1), 010801.
26. Rodríguez, B., & Aydın, G. (2015). Pricing and assortment decisions for a manufacturer selling through dual channels. *European Journal of Operational Research*, 242(3), 901–909.
27. Rohrig, K., Berkhout, V., Callies, D., Durstewitz, M., Faulstich, S., Hahn, B., Jung, M., Pauscher, L., Seibel, A., & Shan, M. (2019). Powering the 21st century by wind energy—Options, facts, figures. *Applied Physics Reviews*, 6(3), 031303.
28. Silva, D. A., & Rupasinghe, T. D. (2017). A Decision Support System for demand planning: A case study from manufacturing industry. *2017 Moratuwa Engineering Research Conference (MERCon)*, 147–152.
29. Suslu, B., Ali, F., & Jennions, I. K. (2023). Understanding the role of sensor optimisation in complex systems. *Sensors*, 23(18), 7819.
30. Vandawaker, R. M. (2015). *Integrated Systems Health Management as an Enabler for Condition Based Maintenance and Autonomic Logistics*.

31. Vespoli, S., Mattera, G., Marchesano, M. G., Nele, L., & Guizzi, G. (2025). Adaptive manufacturing control with Deep Reinforcement Learning for dynamic WIP management in industry 4.0. *Computers & Industrial Engineering*, 202, 110966.
32. Vieira, H., & Castro, R. (2025). A Forward-Looking Assessment of Robotized Operation and Maintenance Practices for Offshore Wind Farms. *Energies (19961073)*, 18(6).
33. Vogl, G. W., Weiss, B. A., & Helu, M. (2019). A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing*, 30(1), 79–95.
34. Werbinska-Wojciechowska, S., & Rogowski, R. (2025). Proactive Maintenance of Pump Systems Operating in the Mining Industry—A Systematic Review. *Sensors*, 25(8), 2365.