POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

An Integrated Model for the Performance Evaluation of Manufacturing Systems with Operational Learning of AI Quality Control Algorithms during Ramp-Up

Laurea Magistrale in Management Engineering - Ingegneria Gestionale

Author: **Alessandro Bordone**
Advisor: **Prof. Maria Chiara Magnanini**
Academic year: **2024-2025**

---

1. Introduction

The production ramp-up phase represents one of the most critical stages in the life cycle of a manufacturing system. During this transition period, production volumes increase while processes and operational procedures have not yet stabilized, leading to temporary productivity losses and operational inefficiencies [1]. Economic impacts are often concentrated in this interval, making ramp-up strategically critical, particularly within increasingly global supply chains.

Ramp-up performance depends on coordinated decision-making under limited operational maturity and strong time-to-market pressure. Early-stage actions can significantly influence subsequent performance trajectories, highlighting the structural sensitivity of this phase [3].

To manage quality variability during ramp-up, artificial intelligence–based inspection systems are increasingly deployed under conditions of limited data availability. In such environments, classification error probabilities evolve as functions of cumulative production experience, following learning curve dynamics widely studied in statistical learning theory [4]. Decision reliability therefore improves progressively but remains structurally data-dependent.

Although the literature investigates learning curves and ramp-up stabilization separately, it does not formalize activation timing as a policy variable within a unified framework linking learning dynamics to cumulative ramp-up exposure. Existing contributions address either learning-dependent reliability or production ramp-up evolution, but do not integrate these dimensions through activation timing within a coherent decision architecture. In particular, the cumulative implications of activation timing have not been explicitly evaluated through volume-integrated risk exposure. The central managerial question therefore shifts from how AI-based inspection systems learn to when adaptive inspection should be activated during ramp-up. This thesis addresses this gap by formalizing activation timing as a governance decision evaluated through cumulative risk exposure.

2. Research Question and Core Objectives

Building upon the industrial and theoretical gaps identified above, the central research question of this thesis can be stated as follows:

"Does the activation timing of an AI-based inspection system during ramp-up structurally affect cumulative risk exposure, even when steady-state learning performance remains unchanged?"

This question explicitly connects three dimensions: learning-dependent reliability, ramp-up dynamics, and cumulative risk exposure. Rather than assessing instantaneous classification accuracy, the focus is on how activation policies shape the accumulation of risk over production volume.

2.1. Modeling and Governance Objectives

To address this question, the work pursues three integrated objectives. First, it formalizes the evolution of learning-dependent error probabilities as functions of cumulative production experience. Second, it embeds these dynamics into a decision-oriented architecture capable of evaluating alternative activation policies. Third, it introduces a governance-oriented evaluation criterion based on risk exposure integrated over cumulative production volume.

The final objective is therefore to develop a decision framework in which activation timing can be evaluated as a policy variable under learning-dependent uncertainty during ramp-up.

3. Methodological Framework

The defined objective leads to the design of an integrated functional architecture, conceived to integrate analysis, forecasting and control modules, predictive models and process data, providing a coherent methodological framework.

The purpose of the architecture is to dynamically represent the interactions between the actual production system, the predictive inspection model and the performance evaluation system. The proposed functional architecture has been developed to integrate two dimensions of the ramp-up phase that are usually treated as separate domains within manufacturing systems: the technological and operational layer - focused on process reconfigurations, bottleneck analysis and efficiency improvement - and the data-driven quality layer - dedicated to the learning and validation of AI-based inspection models. Figure 1 outlines the functional architecture of the AS-IS configuration:

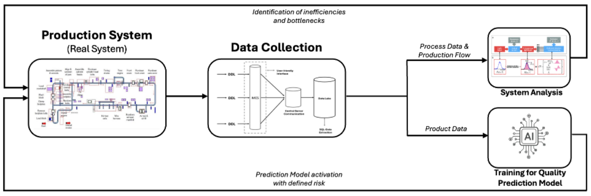

Figure 1: AS-IS configuration - Functional Architecture.

This thesis proposes an evolutionary configuration of the AS-IS setup of the functional architecture, aimed at representing and integrating the interaction between the two traditionally separate dimensions of ramp-up. The proposed TO-BE framework formalizes activation timing as a governance decision embedded within a learning-dependent production environment.

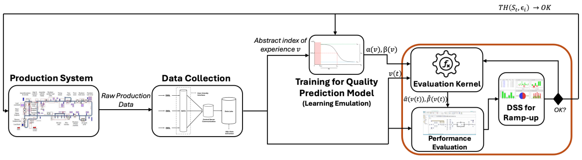

Figure 2: TO-BE Functional architecture of the learning-driven inspection governance framework.

The TO-BE functional architecture illustrates the separation between the real production system, the data acquisition layer, the learning surrogate module, the evaluation kernel and the decision support system (DSS). Rather than modeling machine learning training directly, the proposed approach isolates learning dynamics through parametric surrogate functions and integrates them into a governance-oriented evaluation architecture. Therefore, the methodological structure separates three layers: (i) learning-dependent error dynamics, (ii) statistical activation threshold, and (iii) volume-integrated risk evaluation.

3.1. Parametric Representation of Learning Dynamics

Learning-dependent reliability is represented through two distinct parametric generalized logistic functions describing the evolution of decision error probabilities, denoted $\alpha(v)$ and $\beta(v)$, corresponding to false positive and false negative decisions, respectively, as functions of cumulative first-pass inspection volume v, which serves as the abstract index of experience. Each error component $\varepsilon(v) \in \{\alpha(v), \beta(v)\}$ evolves according to:

$$\varepsilon(v) = \varepsilon_\infty + \frac{\varepsilon_0 - \varepsilon_\infty}{1 + e^{k\,(v - v_f)}} \tag{1}$$

where ε_0 and ε_∞ define the initial and asymptotic error regimes, respectively, k determines the slope coefficient and v_f identifies the volume corresponding to the inflection point of the curve. These functions constitute controlled parametric representations of learning-dependent reliability. Rather than empirically fitted training curves, they serve as analytical abstractions introduced to isolate governance effects from model-training variability. The learning representation is defined for $v \geq V_{start}$, reflecting that adaptive inspection becomes active only once statistical sufficiency has been achieved.

3.2. Statistical Activation Threshold

Activation is not treated as an arbitrary decision, but as a statistically grounded threshold. A minimum cumulative observation requirement $N_{min,tot}$is derived from binomial confidence interval considerations under conservative assumptions on defect prevalence, corresponding to worst-case variance conditions. The formulation is grounded in classical exact confidence interval theory for binomial proportions [2], ensuring that activation occurs only once performance estimates reach a predefined level of statistical reliability. The activation threshold V_{start} is then defined as the minimum production volume satisfying this requirement. This distinction separates statistical sufficiency from operational activation, the latter representing the governance decision to deploy adaptive inspection. Activation timing is therefore formalized as a threshold structurally embedded within the decision architecture.

3.3. Composite Risk and Volume Integration

To enable governance-oriented comparison of activation policies, a composite risk function is defined as:

$$R(v) = c_\alpha \alpha(v) + c_\beta \beta(v) \tag{2}$$

where c_α and c_β denote relative severity weights assigned to the false positive and false negative decision components, respectively.
Rather than comparing instantaneous risk levels, cumulative exposure is assessed through integration over production volume:

$$AUC_R = \int_{v_{start}}^{v_{end}} R(v)dv \tag{3}$$

Integration over cumulative first-pass inspection volume, rather than time, reflects exposure per produced unit and captures the structural accumulation of composite risk during ramp-up. This volume-based aggregation constitutes the key mechanism through which transient learning differences become structurally persistent, thereby enabling governance-oriented policy comparison.

4. Industrial Instantiation

The proposed TO-BE framework is instantiated in the Bosch HGI assembly line during its production ramp-up phase. The industrial setting is characterized by progressive volume increase, evolving process stability and the coexistence of full-cycle and reduced-cycle inspection configurations. In this environment, inspection policies must balance throughput efficiency with quality control under limited data availability in early production stages. Operationally, false negative and false positive decisions carry asymmetric consequences. Missed defects may propagate downstream, generating rework and customer risk, whereas false alarms increase inspection effort and cycle time. This asymmetry makes activation timing of the adaptive inspection regime particularly relevant in shaping system behavior during ramp-up. Therefore, the Bosch case provides a realistic context in which learning-dependent reliability, the statistical activation threshold and risk-weighted policy evaluation can be jointly assessed.

4.1. Parameters Calibration Logic

The learning parameters are calibrated using industrially consistent assumptions rather than empirical curve fitting. The initial and asymptotic error levels (ε_0, ε_∞) are selected within benchmark ranges reported in automotive inspection contexts. The slope parameter k is derived from a predefined transition width corresponding to the 20–80% performance improvement interval of the learning curve. To reflect industrial risk prioritization, asymmetric convergence dynamics are imposed on the two decision error components, allowing false negatives to decline over a narrower transition interval than false positives. This design choice reflects the higher severity typically associated with undetected defects. The statistical activation threshold V_{start} is derived from binomial confidence interval theory under conservative assumptions, ensuring that adaptive inspection is activated only once performance estimates attain a predefined reliability level. Importantly, the parametric learning curves used in this framework are not derived from the operational regression model deployed on the Bosch assembly line. Instead, they serve as controlled analytical representations introduced to isolate governance effects while preserving industrial plausibility.

5. Results - Structural Findings

The simulation-based evaluation compares alternative activation thresholds under identical steady-state learning performance assumptions. The results reveal structural differences that arise exclusively during the ramp-up phase and become magnified through volume-integrated cumulative exposure.

5.1. Learning Dynamics

Figure 3 illustrates the evolution of the two decision error components as functions of cumulative first-pass inspection volume, together with the statistical activation threshold. The trajectories highlight asymmetric learning dynamics: the false negative component declines more rapidly than the false positive component, reflecting its higher operational severity.

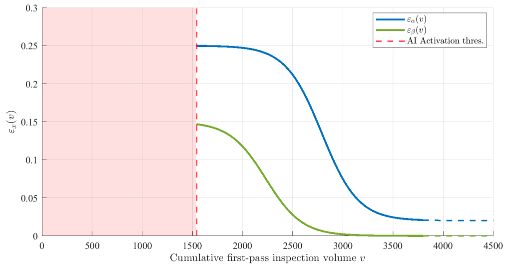

Figure 3: Learning-Dependent Decision Error Dynamics.

5.2. Instantaneous Risk Trajectory

The instantaneous composite risk $R(v)$, under alternative activation thresholds, is reported in Figure 4.

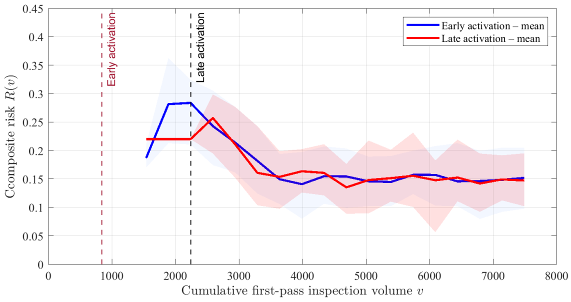

Figure 4: Composite risk for early and late activation policies.

Early activation enables the adaptive decision mechanism at lower cumulative experience, exposing the system to learning-driven decisions under higher initial uncertainty. Performance indicators are computed and reported at batch level; consequently, the plotted trajectories start at the first reported measurement volume rather than at the exact activation threshold. The divergence between early and delayed activation is confined to the transitional window defined by the two activation thresholds. Within this interval, early activation produces higher short-term composite risk, whereas delayed activation prolongs the conservative regime and postpones the onset of learning-dependent dynamics. Once both policies operate within the adaptive regime, instantaneous composite risk converges toward comparable steady-state levels. These findings confirm that activation timing reshapes the risk trajectory during ramp-up without altering steady-state system performance.

5.3. Cumulative Risk Exposure

Although instantaneous composite risk differences are confined to the transitional window, integration over cumulative production volume transforms these temporary asymmetries into persistent divergences in cumulative risk exposure. Even when steady-state risk levels converge, the accumulated exposure reflects the trajectory followed during ramp-up. Figure 5 compares cumulative risk exposure, measured through the volume-integrated metric AUC_R, under early and delayed activation thresholds.

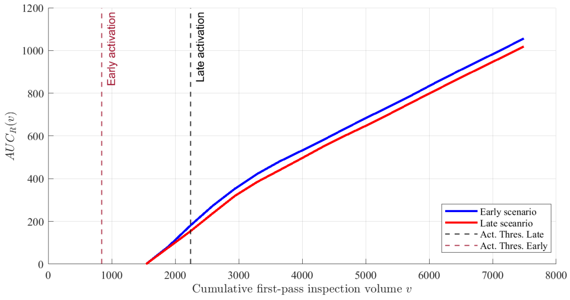

Figure 5: Volume-integrated cumulative risk exposure.

Under the baseline composite risk weighting, delayed activation yields lower cumulative exposure; however, this ordering depends on the underlying risk prioritization. A sensitivity analysis on the relative weighting of false positive and false negative decision errors demonstrates that the policy ranking depends on the underlying risk structure. As the weight assigned to false negatives increases, delayed activation becomes preferable; conversely, greater emphasis on false positives favors early activation in terms of cumulative risk exposure. A break-even weighting point exists at which the cumulative risk difference reverses sign. This confirms that activation timing cannot be assessed independently of the underlying risk structure and must be interpreted as a governance choice rather than a universally optimal policy.
Importantly, the observed differences in cumulative risk are not accompanied by comparable differences in operational throughput performance (Figure 6). Throughput trajectories under early and delayed activation remain nearly identical and converge toward equivalent steady-state levels, with only marginal deviations during the transitional phase.

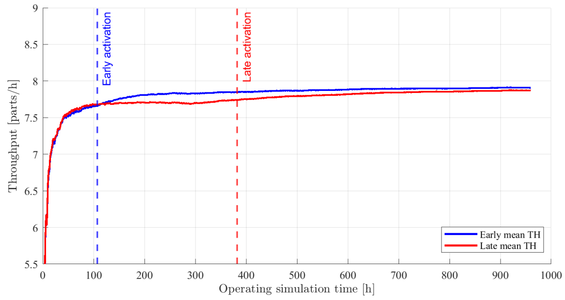

Figure 6: Throughput evolution under early and delayed activation policies.

Taken together, the results demonstrate that activation timing operates as a governance lever. While steady-state throughput remains largely unchanged, cumulative risk exposure varies structurally depending on when adaptive inspection is activated. The preferred activation threshold therefore reflects risk prioritization rather than productivity maximization.

6. Structural Implications of Learning-Dependent Activation

The results of the thesis directly validate the following three structural claims.

Claim 1

The activation timing of learning-driven inspection changes the regime transition point, not the steady-state behavior.

Activation policies defined by different deployment thresholds converge toward comparable asymptotic learning performance. However, activation timing determines when the system switches from a conservative to an adaptive regime. The effect is therefore transitional (temporal) rather than structural in terms of steady-state capacity.

Claim 2

Activation-induced transient error asymmetries produce path-dependent cumulative risk differences when integrated over production volume.

Although activation-related differences in error probabilities are confined to the transitional window, their integration over cumulative production volume produces persistent divergences in cumulative risk exposure. Volume-based integration thus transforms temporary asymmetries into structurally persistent outcomes.

Claim 3

Activation timing operates as a risk-governance lever rather than a productivity optimization variable.

Steady-state throughput remains largely unchanged under alternative activation policies.

Instead, the preferred activation threshold depends on the weighting assigned to error components within the composite risk function. Activation timing therefore reflects risk prioritization rather than steady-state performance maximization. Beyond this descriptive result, the findings reveal deeper structural implications for learning-dependent governance during ramp-up.

6.1. Structural Interpretation of the Results

First, activation timing alters the regime transition point without affecting steady-state learning performance. Early and delayed activation policies converge toward comparable steady-state error levels and throughput, indicating that the effect of activation timing is confined to the transitional ramp-up phase rather than shaping long-run system performance.

Second, although instantaneous composite risk differences are localized within the activation window, their integration over cumulative production volume generates persistent divergences in cumulative risk exposure. The volume-integrated metric AUC_R reveals structural differences that remain invisible when comparing instantaneous risk values alone, demonstrating that transient learning dynamics can produce structurally persistent cumulative effects during ramp-up.

Finally, the preference between early and delayed activation depends on the relative weighting of false positive and false negative decision errors. A break-even weighting point exists at which the cumulative risk difference reverses sign, confirming that activation timing is not intrinsically optimal in isolation. Throughput trajectories remain largely unchanged under alternative activation policies, indicating that the decision does not materially affect steady-state operational performance. Taken together, these findings demonstrate that activation timing is not a tactical productivity parameter but a structural governance lever reflecting risk prioritization.

7. Conclusions

This work formalizes learning-dependent reliability during production ramp-up by introducing a structured framework for governing the activation timing of AI-based inspection systems. The study moves beyond instantaneous accuracy evaluation to examine how learning dynamics interact with cumulative production exposure. Methodologically, its primary contribution lies in combining parametric learning representations with a volume-integrated composite risk metric. By defining a composite risk function and integrating it over cumulative first-pass inspection volume, the framework captures the structural impact of transient error dynamics during ramp-up. This formulation enables activation policies to be evaluated in terms of cumulative exposure rather than instantaneous performance.

The results demonstrate that activation timing does not materially alter steady-state operational capability but instead shapes how risk accumulates during the transitional ramp-up phase. The relative preference between early and delayed activation depends on the weighting assigned to the underlying error components, confirming that activation timing must be interpreted as a governance decision rather than as a productivity optimization parameter.

Although developed and calibrated on the Bosch HGI assembly line, the framework is generalizable to other production environments in which learning-dependent decision systems operate under dynamic operational conditions. By explicitly separating statistical sufficiency, operational activation, and cumulative risk evaluation, the model provides a structured tool for risk-aware planning of AI-based inspection integration during ramp-up.

The proposed framework shifts the evaluation of AI-based inspection systems from instantaneous accuracy assessment toward cumulative risk governance during production ramp-up.

References

[1] Ramón Bultó, Elisabeth Viles, and Ricardo Mateo. Overview of ramp-up curves: A literature review and new challenges. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 232(5):755–765, 2018.

[2] Charles J Clopper and Egon S Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934.

[3] Maria Chiara Magnanini, Khaled Medini, and Bogdan I Epureanu. Decision making for fast productivity ramp-up of manufacturing systems. In *CIRP Novel Topics in Production Engineering: Volume 1*, pages 235–266. Springer, 2024.

[4] Tom Viering and Marco Loog. The shape of learning curves: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):7799–7819, 2022.