You Test the Failures You Anticipated: Envelope Coverage Under a Takt Budget, and Capability as an Early Warning

Architecture, Computational Methods, and a Modelled Scenario for MileSoft AC Performance Test

MileSoft Engineering Research Group

MileSoft Software Technologies

https://milesoft.net/

Corresponding authors: *Deepak Daryani* (deepakdaryani@milesoft.net), *Tanuj Daryani* (tanujdaryani@milesoft.net)

Working paper — Version 1.0 — 24 August 2026

Permanent link: https://milesoft.net/research/products/ac-performance-test

Abstract

The product's material states the problem in one sentence: you test for the failure modes you anticipated, you ship clean for the ones you did not, and the warranty data three quarters later tells you which you missed. An end-of-line functional test is a sample of an operating envelope, and everything that escapes lives where the sample was not taken.

This paper presents the architecture and computational methods of MileSoft AC Performance Test, a station running a multi-condition sequence — refrigerant charge, compressor duty cycle, evaporator and condenser temperature differential at calibrated airflow, sensor health and acoustic signature — inside production takt, with capability tracking on key measurements and automatic escalation on failure.

Three results follow. Test design is a coverage problem under a cycle-time budget rather than a completeness exercise: each candidate condition covers some mass of the failure distribution and costs some seconds of takt, which makes the selection a knapsack with a computable objective. Capability tracking is an early-warning instrument rather than a report, and its value is a lead time equal to the margin between the process mean and the acceptance limit divided by the drift rate — a quantity a station can publish. And no end-of-line test observes a time-dependent degradation, so a component within specification today and outside it in six months passes every condition in the sequence; what addresses that is margin-based acceptance, which is a different intervention from broader coverage.

This station is named in a published Tier-1 deployment as a downstream cross-check, but no result is attributed to it there and none is claimed here. Section 5 is a modelled scenario with every assumption printed, and every figure is an authored schematic because this product has no captures of any kind.

Keywords: AC performance test; End of line functional test; Automotive air conditioning test; Refrigerant charge verification; Compressor duty cycle test; Test envelope coverage; Cp Cpk drift alarm; Field escape prevention; Acoustic signature testing; Takt time test design; Acceptance band configuration; Non-conformance escalation

1 Introduction

End-of-line functional testing is every quality organisation's favourite lever and most lines' largest blind spot. The product's material states why in one sentence worth quoting in substance: you test for the failure modes you anticipated, you ship clean for the ones you did not, and the warranty data three quarters later tells you which ones you missed.

That is not a criticism of testing. It is a structural property of it. A functional test measures a unit under conditions somebody chose, and a unit that works under those conditions is a unit that works under those conditions.

This paper treats the choice of conditions as the design problem, which is where the station's engineering actually is. Broadening the sequence increases coverage and costs cycle time, and a station engineered for a real line takt is making that trade whether or not it is stated.

1.1 Three failure classes, and which testing reaches

It helps to separate what an end-of-line test can catch from what it cannot, because the second category is where the expensive failures live and no amount of test design moves them into the first.

Table 1. Three classes of failure and what reaches each. Only the first is addressed by broader coverage.

Class	Example	What reaches it
Present and in the envelope	Undercharge; a dead sensor; a leak	Testing at the right conditions — Section 4.1
Present and outside the envelope	Noise under sustained high load	Broader coverage, at a takt cost — Section 4.2
Not yet present	A sensor in spec today, out in six months	Nothing in a test; margin-based acceptance — Section 4.4

The third row is the honest one and it is stated here rather than in the limitations, because a paper about end-of-line testing that does not say this up front is selling the wrong instrument.

1.2 Contributions

1. Envelope coverage stated formally, with escape probability as the mass of the failure distribution outside the tested set (Section 4.1).

2. Test selection as a knapsack under a takt budget, making explicit what a cycle-time-engineered station chooses not to test (Section 4.2).
3. Capability tracking as an early warning with a publishable lead time (Section 4.3).
4. Margin-based acceptance as the only response to time-dependent degradation (Section 4.4).
5. An eight-dimension capability reference framework for end-of-line functional test stations (Section 6.3).

2 Background and Related Work

An end-of-line test station is a measurement instrument operating under production constraints, and both halves of that description carry a literature.

2.1 The station is a gauge before it is a test

Every value the station reports is a measurement, and ISO 5725 separates the two ways a measurement can be wrong: trueness, the closeness of a mean to a reference, and precision, the scatter of repeated observations. AIAG's measurement systems analysis supplies the study that quantifies both for a specific gauge.

This is easy to skip in a functional test because the output looks like a verdict rather than a number. A pass verdict is a number compared against a limit, and a comparison inherits the uncertainty of the number.

For a temperature-differential measurement at calibrated airflow, the airflow calibration is itself an uncertainty source with its own standards — ISO 5801 and AMCA 210 govern the duct methods behind it. A differential measured at an airflow known to five per cent carries that five per cent.

2.2 Capability, and when an index may be reported

ISO 22514 gives the capability and performance indices and, more usefully here, the condition governing which may be reported: capability indices assume the process is in statistical control, and where it is not, performance indices are the honest statement.

That distinction is the whole basis of Section 4.3. A drift alarm is precisely an out-of-control signal, so a station reporting a healthy capability index on a drifting process is reporting a number whose own standard says it does not apply — and it is doing so at exactly the moment the warning was available.

> A capability index on a drifting process is not a slightly wrong number. It is a statement about a stationary process, computed on a process that is not one.

2.3 Where a defect is caught determines what it costs

The pokayoke literature's detection-cost escalation applies directly: a fault caught at the station costs a rework, at end of line a disassembly, and in the field a warranty claim plus whatever the customer's programme costs.

Shingo's control-versus-warning distinction also applies, in a form specific to a test station. A test is a warning function by nature — it detects rather than prevents — so its value depends entirely on what happens next. The product's material puts a failing unit into an andon alert and an opened non-conformance workflow that holds the line, which is what converts a detection into a control at the line level rather than the unit level.

Nakajima's loss framing supplies the other half of the trade. A test consuming takt is a planned availability loss, so a station's coverage decision is a quality-against-availability decision, and Section 4.2 states it as one.

3 System Overview

The station is a controller-orchestrated test sequence over an instrumented fixture, with acceptance bands held per part number, a per-unit signed report, capability tracking across units, and automatic escalation on failure.

> Every figure in this paper is an authored schematic. This product's page carries no screenshots of any kind, and no image belonging to another product is used to stand in for one.

3.1 What the sequence measures

The product's material names five things a defensible sequence must cover, and each addresses a distinct physical failure mode.

Table 2. The five conditions, the failure each addresses, and what it costs in cycle time. The third column is why Section 4.2 exists.

Condition	Failure it reaches	Takt cost
Refrigerant charge by mass	Under- or overcharge, and leaks	Low — a settled measurement
Compressor duty across the envelope	Behaviour away from the rating point	High — requires dwell at several conditions
Evaporator and condenser differential	Capacity shortfall, restricted flow	Moderate — needs thermal settling
Sensor sweep	Dead or out-of-range thermistors and transducers	Low — electrical, fast
Acoustic signature	Noise complaints, mechanical interference	Low to moderate — a short capture

The second row is the expensive one and the one a catalogue-point test omits. Compressor behaviour at the rating point is a single sample of a duty envelope, and the failures the product's material describes as surfacing in warranty data are concentrated away from that point.

The fifth row is worth noting because it is unusual in an end-of-line test. Acoustic capture addresses a failure class defined by customer perception rather than by specification, which means its acceptance band is derived from complaint data rather than from a drawing.

3.2 Where the limits live

Acceptance bands and sequences are configurable per part number through the station interface, and the station validates which configuration it has loaded before each cycle.

The validation step is the part that matters. A station holding per-variant limits and not checking which set it loaded is a station that will test a variant against another variant's limits, and the result will be a confident pass or a confusing fail with nothing to indicate which.

Section 4.4 makes the reconfiguration argument in coverage terms rather than convenience terms: limits that require an engineering release are limits that lag, and every unit tested during the lag was tested against a specification that no longer applied.

3.3 What happens when a unit fails

A failing unit triggers an andon alert and opens a non-conformance workflow, and the line does not process further units until the fault is resolved. Results are linked to the unit's identity in the traceability record and surfaced in the production record.

Holding the line on a single failure is a strong policy and it deserves examination rather than approval. It is correct when a failure indicates a process condition affecting subsequent units — a charge station drifting, a fixture leaking — and it is expensive when the failure is a genuinely isolated component defect.

The distinction is exactly what capability tracking supplies. A failure occurring while the distribution is centred and stable is probably isolated; a failure occurring while the mean has been walking for two shifts is probably the first of many, and Section 4.3 gives the instrument that tells them apart.

> A hold-the-line policy without a capability signal treats every failure as systemic, which is safe and costly. With one, the policy can be conditioned on evidence rather than on the fact that something failed.

3.4 Calibrating the instrument

The station manages calibration of its own sensors on a workflow the material describes as auditor-recognised, which is the same obligation the companion alignment-station paper identifies and it bites the same way.

Every result the station has produced is relative to its own instruments. A drifted thermocouple or a mis-calibrated mass measurement does not produce obviously wrong numbers; it produces plausible ones, and the population invalidated is bounded by the calibration interval rather than by anything the station notices.

4 Computational Methods

Four computations carry the paper: what coverage is, how to choose it under takt, what capability tracking buys, and what to do about the failures no test can see.

4.1 Coverage, and the escape it leaves

Let the operating envelope be the set of conditions a unit will experience in service, with f a density over it describing where failures actually occur.

the covered set is the union of the neighbourhoods each test condition exercises; coverage is the failure density integrated over it; escape probability is what is left outside (coverage)

K is the set of test conditions run, B(k) the region of the envelope condition k meaningfully exercises, and f the density of failure occurrence over the envelope — which is estimated from warranty and field data, not assumed uniform.

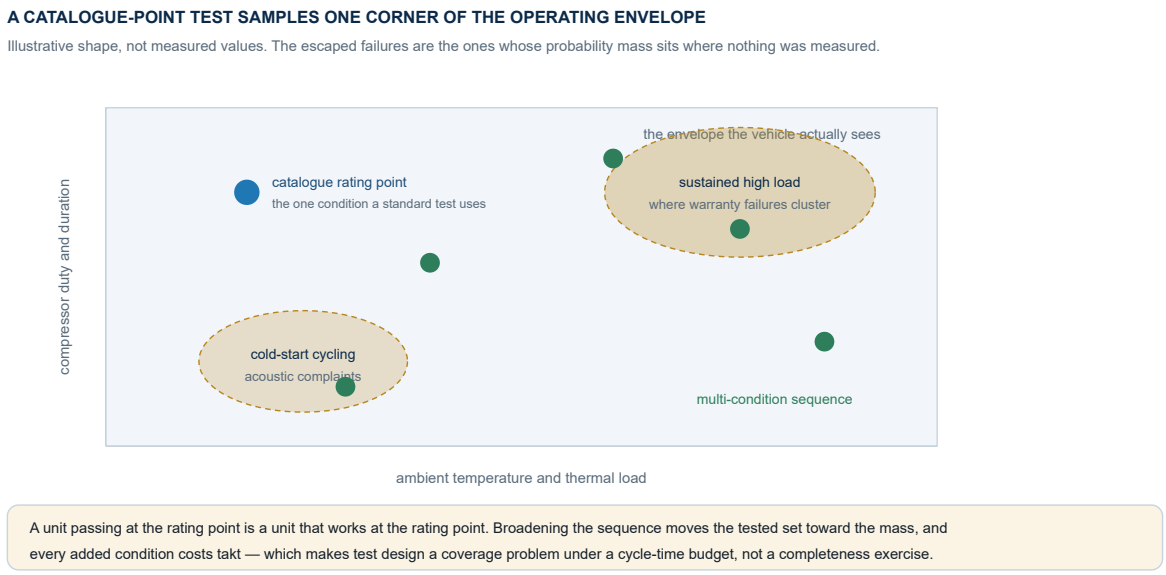

Figure 1. *Schematic.* Schematic (illustrative shape, not measured values): the operating envelope, the single catalogue rating point a standard test uses, the broader set a multi-condition sequence exercises, and the two regions where warranty failures cluster. A unit passing at the rating point is a unit that works at the rating point.

Two properties of f decide everything. It is not uniform — failures concentrate in specific regions, typically sustained high load and thermal cycling. And it is estimable: the warranty data the product's material describes as arriving three quarters later is a sample from f, which means a plant with warranty history can compute its own coverage rather than guessing.

> The recommendation follows immediately and is unusually cheap. Take last year's warranty claims, locate each in the operating envelope, and ask which of your current test conditions would have exercised it. The answer is your coverage figure.

4.2 Choosing conditions under a takt budget

Every condition costs cycle time and the station must fit takt. That makes the selection a bounded optimisation rather than a wish list.

$$\textit{choose the set of conditions maximising covered failure mass, subject to their total duration fitting inside takt less handling time} \qquad \text{(knapsack)}$$

t(k) is the dwell and settling time condition k requires, T(takt) the line's cycle time, and t(handle) the fixture load, unload and identification overhead. Coverage is submodular because regions overlap, so marginal value falls as conditions are added.

The submodularity matters practically. Two conditions covering overlapping regions of the envelope deliver less together than the sum of their individual coverage, so the correct ranking is by marginal coverage per second rather than by absolute coverage — and a condition that looked valuable in isolation can be worth almost nothing once another is in the set.

Table (conditions) shows why this is not academic. Compressor duty across the envelope is simultaneously the highest-coverage condition and the most expensive in takt, so it is the one under permanent pressure — and the one whose removal is hardest to detect, because the units it would have caught pass everything else.

The framing also gives a defensible answer to a question test engineers are often asked without one: what would you add if takt allowed. The answer is the condition with the highest marginal coverage per second, and it is computable from the same data as Section 4.1.

4.3 What a capability alarm actually buys

The product's material states that drift alarms fire typically before a single unit would have failed acceptance. That is the right claim and it can be given a magnitude.

$$\textit{the capability index is the distance from the mean to the nearer limit over three standard deviations; the lead time to the first failing unit is the remaining margin beyond three sigma divided by the drift rate} \qquad \text{(leadtime)}$$

USL and LSL are the acceptance limits, mu the process mean, sigma its spread, and the derivative the rate at which the mean is moving. The alarm threshold is a chosen capability index; the lead time is what that choice buys.

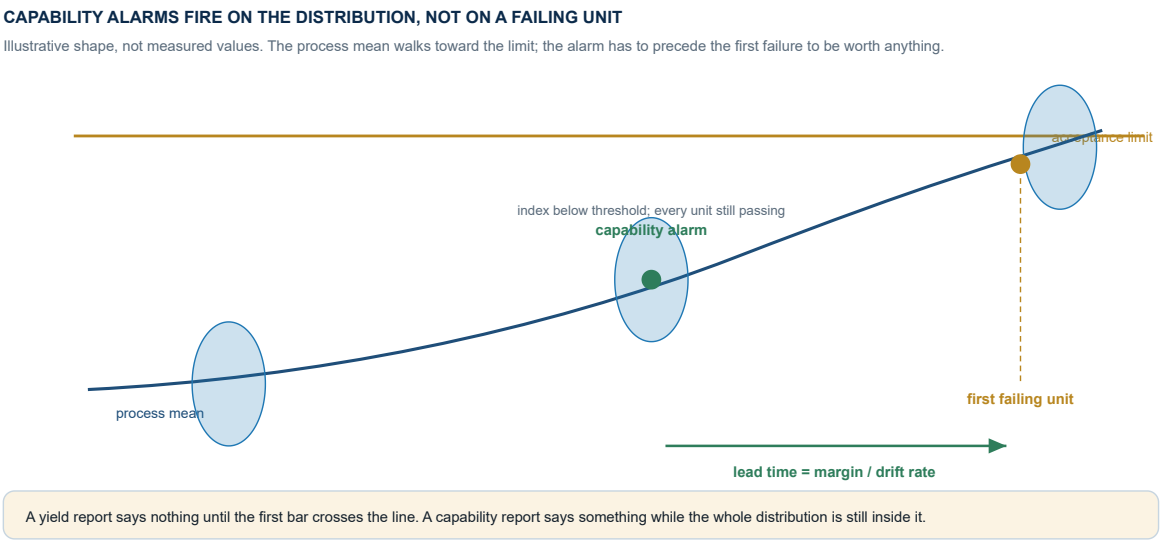

Figure 2. *Schematic.* Schematic (illustrative shape, not measured values): the process mean walking toward an acceptance limit. The capability alarm fires while every unit is still passing; the first failing unit arrives later. The interval between them is the margin divided by the drift rate, and it is the whole value of tracking the distribution rather than the yield.

Two design consequences follow. The alarm threshold is a lead-time decision rather than a quality one: a higher threshold fires earlier and produces more investigations that turn out to be nothing, and the right setting depends on how fast the plant can act. And the lead time is proportional to margin, so a process running comfortably inside its limits gets a long warning while one running near them gets almost none — which means the plants most in need of the warning receive the least of it.

The second consequence deserves emphasis because it is counter-intuitive. Tightening acceptance limits improves outgoing quality and shortens the warning; widening them lengthens the warning and admits worse units. That is a real trade and the station should expose both sides of it rather than presenting the alarm as free.

> This is also why capability must be computed per shift and per line rather than pooled. A drift on one line averaged with three stable ones produces an index that moves too slowly to warn about anything.

4.4 The failures no sequence reaches

The product's own material identifies the hardest class directly: sensor drift that is within specification on day one and outside it six months later. No test condition observes it, because at test time the unit conforms.

$$\textit{accept only if the measured value plus the expected in-service degradation plus a measurement allowance stays inside the limit; the margin a unit ships with is the distance between its measurement and the limit} \qquad \text{(margin)}$$

x(0) is the value measured at end of line, Delta(life) the expected drift over the warranty period, u the measurement uncertainty and k its coverage factor. A unit passing on x(0) alone ships with margin tau but no statement about whether tau survives the warranty.

This is a different intervention from broader coverage and it is worth being clear about the distinction. Coverage moves the tested set toward where failures occur; margin-based acceptance rejects units that would pass today and fail later. The first addresses class two in Table (classes); only the second touches class three.

It is also more expensive than it looks. Rejecting units that conform to specification requires an engineering position that the specification is insufficient for the warranty period, and it produces scrap that a supplier's customer may not accept as chargeable. The honest statement is that this is a design and contractual decision the test station can enforce but cannot make.

What the station can contribute is the data to make it. Recording x(0) per unit, and joining it to warranty returns through the traceability record, produces exactly the sample needed to estimate the degradation term — which is the mechanism by which a plant learns its own value for Delta(life) rather than assuming one.

> The margin a unit shipped with is on its report. A plant that joins those margins to its warranty returns can compute, for the first time, whether its acceptance limits are the right ones — and that analysis needs no new instrument, only the join.

5 Modelled Scenario

> No published outcome is attributed to this station. It is named in a Tier-1 deployment as a downstream cross-check on a sister product family, but the four results that deployment reports belong to the kitting and alignment stations. Everything below is a design statement or a modelled scenario with its assumptions printed.

5.1 The deployment this station appears in

A Tier-1 automotive supplier deployed three pokayoke stations across two plants. The published account describes this station being added downstream as a cross-check on the same line for a sister product family, with all three stations streaming events into one dashboard for shift leads and audit playback.

> No result is attributed. The deployment's published outcomes — zero wrong-part picks, a 92% containment reduction, five-day ramp-up and an 18% throughput gain — belong to the kitting and alignment stations or to the programme, and this paper converts none of them into a claim for this one.

What the account does establish is the escalation coupling of Section 3.3. A station whose events reach the same dashboard as the upstream stations is one where a failure can be correlated with what happened before it, which is the difference between a test result and a diagnosis.

5.2 Design statements published for this station

Table 3. Statements published about this station, separated by kind. Every design statement is checkable by demonstration; none is a measurement.

Statement	What it asserts	Kind
Multi-condition sequence	charge, duty cycle, differential, sensors, acoustics	Design
Cycle-time engineered for takt	the sequence fits the line's actual cycle time	Design
Per-unit signed report	all measured values, exportable to customer formats	Design
Capability drift alarms	per shift, per line, on key measurements	Design
Configurable acceptance bands	per part number, without a station rebuild	Design
Automatic escalation	andon alert and non-conformance workflow on failure	Design
Alarms before the first failure	the alarm typically precedes a failing unit	Claim

The last row is the only claim in the set and Section 4.3 gives it a magnitude it does not carry on its own. Typically before a failing unit is a direction; margin divided by drift rate is a number, and a station could report it.

5.3 Modelled field escapes avoided

The published return model prices the saving as field escapes avoided. Its assumptions are printed here so a reader can substitute their own.

Table 4. Modelled scenario — not a deployment result. Assumptions are the published defaults of the return-on-investment model for this product.

Assumption	Value
Units tested per month	8,000 (96,000 per year)
Baseline field-escape rate	2% (1,920 per year)
Reduction attributed to the station	75%
Modelled escapes avoided	1,440 per year
Cost per escape	site-specific; the published default illustrates only

The 75% is the assumption to interrogate, and Section 4.1 says exactly what it is a claim about: the fraction of the failure density the broadened sequence covers. That is estimable from a plant's own warranty history rather than adoptable from a vendor, and the estimate is the analysis this paper most recommends running before any procurement.

The model also carries a structural optimism worth naming. It prices escapes as a rate reduction, and Table (classes) shows the third failure class — present nowhere at test time — is untouched by coverage. A plant whose escapes are dominated by in-service degradation should expect substantially less than 75%, and Section 4.4 says what would help instead.

6 Discussion

6.1 The warranty file is the specification for the test

Section 4.1 makes coverage depend on a failure density that has to come from somewhere, and the answer changes how a test station should be specified.

Test sequences are conventionally derived from design intent: the engineering team lists what could go wrong and the station checks for it. That is the process the product's material criticises, and correctly — it produces coverage of anticipated modes and nothing else.

The alternative is to derive the sequence from the warranty file. Every claim is an observation of where in the envelope a failure occurred, so a year of claims is a sample from f, and the conditions that maximise coverage against that sample are computable rather than debatable.

This inverts the usual relationship between quality functions. Warranty data is normally the scoreboard; here it is the input, and a plant that reviews its test sequence annually against last year's claims has a station that improves rather than one that ages.

> The uncomfortable corollary: a station specified once at line installation and never revised is covering the failure modes somebody anticipated years ago, which is precisely the condition the product's material identifies as the problem.

6.2 The trade nobody states

Section 4.3 produces a result worth dwelling on because it runs against instinct: tightening acceptance limits shortens the drift warning.

The mechanism is simple. Lead time is the margin between the process mean and the limit divided by the drift rate, so moving the limit inward reduces the numerator. A plant that tightens limits to improve outgoing quality has, in the same action, reduced the interval in which a developing problem can be caught before it produces failures.

Neither direction is wrong and the point is that it is a choice. Tight limits with a short warning suit a process that is stable and monitored closely; wider limits with a long warning suit one that drifts and is watched less. A station should present both figures — outgoing quality and lead time — so the decision is made rather than inherited.

There is a third option that dominates both where it is available: reduce sigma. Halving the process spread widens the margin at unchanged limits, improving outgoing quality and lengthening the warning simultaneously — which is the case for treating capability improvement as an alternative to limit-setting rather than a consequence of it.

6.3 A capability reference framework for end-of-line functional test stations

Table 5. Capability reference framework. Each dimension separates a functional test station from a pass-fail rig, and each is answerable by demonstration on a live station rather than by reading a specification.

Dimension	Question the system must answer by demonstration
D1 Stated coverage	Which regions of the operating envelope does the sequence exercise, and which does it not?
D2 Sequence provenance	Was the sequence derived from design intent, or from warranty claims?
D3 Marginal value	What would you add if takt allowed, and why that condition?
D4 Stated uncertainty	What is the measurement uncertainty on each key value, and how was it established?
D5 Capability, not yield	Does the station alarm on a distribution shift while all units still pass?
D6 Published lead time	How long, in units or hours, between that alarm and the first expected failure?
D7 Configuration validated	Does the station verify which acceptance band it loaded before each cycle?
D8 Margin recorded	Is each unit's distance from the limit on its report, and joinable to warranty returns?

D2 is the question that predicts whether the station will still be catching the right failures in three years. D8 is the one that costs nothing and enables the only analysis capable of answering whether the limits are right.

6.4 Generalisability

The coverage-under-budget formulation generalises to every end-of-line test in every industry, and so does the observation that the highest-coverage condition is usually the most expensive in takt and therefore permanently under pressure.

The lead-time result generalises to any statistical process control on a drifting mean, and the trade it exposes — tighter limits, shorter warning — applies wherever acceptance limits and monitoring coexist, which is most regulated manufacturing.

What does not generalise is the modelled 75% escape reduction. It is a claim about how much of one plant's failure density a broadened sequence covers, and Section 4.1 gives the method for computing a local figure instead.

7 Threats to Validity and Limitations

1. No outcome is attributed to this station. It is named in a deployment as a downstream cross-check and the results reported there belong to other stations, so this paper reports no field evidence at all.
2. The modelled 75% escape reduction is an assumption. Section 5.3 argues it is a claim about coverage of a local failure density and should be replaced by a figure computed from the plant's own warranty history.
3. No test observes a failure that has not happened. Table (classes) states this at the outset: a component conforming today and drifting later passes every condition, and coverage does not address it.
4. The failure density f is estimated from claims, which are a biased sample. Warranty data over-represents failures customers notice and report, and under-represents those they tolerate or misattribute.
5. Coverage regions are not sharply bounded. Treating each condition as exercising a region of the envelope is a modelling convenience; in practice a test's diagnostic reach fades rather than stopping.
6. No measurement uncertainty is published. Equation (margin) needs it and no figure for the station's temperature, pressure, mass or acoustic uncertainty appears in the product's material.
7. The lead-time result assumes a linear drift. A step change in the process produces no warning at all, and Equation (leadtime) is silent about it.
8. No figure in this paper is a product capture. This product's page carries no screenshots, so nothing here demonstrates the described station exists in the form modelled.

The seventh limitation bounds the paper's main positive result. Capability alarms warn about processes that drift; they say nothing about a fixture that is changed incorrectly between shifts, and that failure mode needs a different instrument entirely.

8 Future Work

- Coverage computed and published per station. Locating a year of warranty claims in the operating envelope and reporting which current conditions would have exercised each turns coverage from an argument into a number.
- Marginal coverage per second of takt, so the question of what to add if cycle time allowed has a computed answer rather than an opinion.
- Lead time published alongside the alarm. Section 4.3 shows the interval is computable from margin and drift rate; reporting it would let a plant judge whether the warning is long enough to act on.
- Margin joined to warranty returns. Recording each unit's distance from the limit and joining it through the traceability record to field failures is the only analysis capable of saying whether the acceptance limits are correct.
- Step-change detection alongside drift detection, since Section 7 notes a capability alarm gives no warning of a fixture changed incorrectly between shifts.

9 Conclusion

An end-of-line test is a sample of an operating envelope, and everything that escapes lives where the sample was not taken. That makes test design a coverage problem rather than a completeness exercise, and it makes a station engineered for takt one that is deliberately choosing what not to test.

Three results follow. Coverage is computable: the failure density over the envelope is estimable from warranty claims, so a plant can measure what its current sequence reaches rather than assuming. Selecting conditions is a knapsack under a cycle-time budget with submodular value, which means conditions should be ranked by marginal coverage per second and that the highest-coverage condition is usually the most expensive and therefore permanently under pressure. And capability tracking is an early-warning instrument whose value is a lead time — margin divided by drift rate — with the uncomfortable consequence that tightening acceptance limits shortens the warning.

The fourth result is the one this paper puts first rather than last. No test observes a failure that has not happened yet, so a component conforming today and drifting in six months passes every condition in any sequence. Broader coverage does not touch it; margin-based acceptance does, and that is a design and contractual decision the station can enforce but cannot make.

The framework of Section 6.3 is offered as the durable contribution, and its second dimension predicts whether a station will still be catching the right failures in three years: was the sequence derived from design intent, or from what actually came back?

Appendix A Nomenclature

Table 6. Symbols and abbreviations used in this paper.

Symbol / term	Meaning
omega	A point in the operating envelope — a combination of ambient, load and duty
f(omega)	Density of failure occurrence over the envelope, estimated from field data
B(k)	The region of the envelope test condition k meaningfully exercises
C (script)	The covered set — the union of those regions
gamma	Coverage — the failure density integrated over the covered set
t(k)	Dwell and settling time condition k requires
T(takt), t(handle)	Line cycle time, and fixture load, unload and identification overhead
USL, LSL	Upper and lower acceptance limits
mu, sigma	Process mean and standard deviation of a measured quantity
Cpk	Capability index against the nearer limit
t(lead)	Interval from a capability alarm to the first expected failing unit
x(0)	The value measured at end of line
Delta(life)	Expected in-service drift of that value over the warranty period
u, k	Measurement uncertainty and its coverage factor
tau(margin)	The distance between a unit's measurement and the limit it shipped with
NCR	Non-conformance report — the workflow a failure opens

Appendix B Worked Numerical Examples

Appendix B.1 Computing coverage from last year's claims

A plant reviews 240 warranty claims and locates each in the operating envelope. 96 cluster at sustained high load, 62 at cold-start cycling, 44 at the rating point or near it, and 38 are scattered.

Its current sequence tests only at the rating point. Applying Equation (coverage), the covered region contains the 44 rating-point claims plus perhaps 10 of the scattered ones, so gamma is roughly 54 of 240, or 0.23 — and the escape probability is 0.77.

Adding a sustained-high-load condition brings in 96 more, taking gamma to 0.63. Adding cold-start cycling brings 62 more, taking it to 0.88.

The plant now has a defensible statement it did not have before: its current test reaches under a quarter of where its failures actually occur, and two additional conditions would reach seven-eighths. Neither number required a new instrument — only the join between claims and envelope positions.

> Note that the rating point covered 44 claims and was where all the test effort went. It is not that the catalogue condition is useless; it is that it was the only one, and the failures do not live there.

Appendix B.2 What fits inside takt

Line takt is 96 seconds and fixture handling consumes 22, leaving 74 seconds for testing. Candidate conditions, with duration and marginal coverage from Appendix B.1: charge by mass 8 s for 0.05; sensor sweep 6 s for 0.04; differential 18 s for 0.10; acoustic capture 12 s for 0.24; sustained high load 46 s for 0.40; cold-start cycling 34 s for 0.25.

Ranking by marginal coverage per second: acoustic 0.0200, cold-start 0.0074, sustained load 0.0087, differential 0.0056, charge 0.0063, sensors 0.0067.

Applying Equation (knapsack) greedily within 74 seconds: acoustic (12 s, 0.24), sustained load (46 s, 0.64 cumulative), sensors (6 s, 0.68), charge (8 s, 0.73). Total 72 seconds, coverage 0.73. Cold-start cycling and the differential do not fit.

The result is uncomfortable and useful. Acoustic capture — the cheapest condition and the one most likely to be dismissed as a nicety — has by far the best coverage per second and belongs first. The differential test, which feels like the core of an air-conditioning test, is the worst value in the set and is the one to drop.

A plant reaching this conclusion should check its coverage estimates before acting on it, which is the point: the framework makes the assumptions explicit enough to argue about.

Appendix B.3 How long the alarm actually buys

A charge measurement has an upper acceptance limit of 640 g. The process mean is 592 g with a standard deviation of 9 g, and a worn filling valve is causing the mean to rise at 1.4 g per shift.

Applying Equation (leadtime): the current capability index against the upper limit is (640 — 592) / (3 x 9) = 1.78. The first unit at three sigma beyond the mean crosses the limit when the mean reaches 640 — 27 = 613 g, which is (613 — 592) / 1.4 = 15 shifts away.

An alarm threshold at Cpk = 1.33 fires when the mean reaches 640 — 3 x 9 x 1.33 = 604 g, which is 8.6 shifts from now — leaving 6.4 shifts of warning before the first failure. A threshold at 1.67 fires at 595 g, in 2 shifts, leaving 13 shifts of warning.

Now tighten the acceptance limit to 620 g to improve outgoing quality. The first failure arrives when the mean reaches 593 g — less than one shift away — and a 1.33 threshold has already been breached. Tightening the limit by 3% converted a fifteen-shift problem into an immediate one, which is Section 6.2's

trade in numbers.

Data provenance

Every quantitative claim in this paper resolves to one of the entries below. Figures marked as modelled estimates are not deployment measurements; their assumptions are stated.

Metric	Reported value	Provenance
Conditions exercised by the multi-condition test sequence	refrigerant charge, compressor duty cycle, evaporator and condenser delta-T, sensor health, acoustic signature	**Product specification** /products/ac-performance-test
When a statistical drift alarm fires relative to the first failing unit	typically before a single unit would have failed acceptance	**Product specification** /products/ac-performance-test — Stated on the product FAQs as the purpose of per-shift capability tracking; no lead-time figure is published.
How acceptance limits change for a new variant	configurable per part number, without a hardware or software rebuild	**Product specification** /products/ac-performance-test
What a failing unit triggers automatically	an andon alert and an opened non-conformance workflow, holding the line	**Product specification** /products/ac-performance-test
Field escapes avoided per year	1,440 of 1,920 (modelled)	**Modelled estimate** 8,000 units tested per month, 96,000 per year; Baseline field-escape rate of 2%; 75% reduction attributed to the station; Model and defaults published in src/data/roiModels.ts

References

[1] International Organization for Standardization (2014). ISO 22514-1: Statistical methods in process management - Capability and performance - Part 1: General principles and concepts. ISO, Geneva. https://www.iso.org/standard/64135.html

[2] International Organization for Standardization (2017). ISO 22514-2: Statistical methods in process management - Capability and performance - Part 2: Process capability and performance of time-dependent process models. ISO, Geneva. https://www.iso.org/standard/71617.html

[3] International Organization for Standardization (2023). ISO 5725-1: Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions. ISO, Geneva (superseding ISO 5725-1:1994).

[4] International Organization for Standardization (2019). ISO 5725-2: Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility. ISO, Geneva.

[5] Automotive Industry Action Group (Chrysler, Ford, General Motors) (2010). Measurement Systems Analysis (MSA) Reference Manual. AIAG, 4th edition, MSA-4.

[6] International Automotive Task Force (2016). Quality management system requirements for automotive production and relevant service parts organizations. IATF 16949:2016, first edition, 1 October 2016 (superseding ISO/TS 16949).

[7] International Organization for Standardization (2017). ISO 5801: Fans - Performance testing using standardized airways. ISO, Geneva; Amendment 1 published January 2025. https://www.iso.org/standard/56517.html

[8] Air Movement and Control Association / ASHRAE (2016). ANSI/AMCA Standard 210 (ANSI/ASHRAE Standard 51): Laboratory Methods of Testing Fans for Certified Aerodynamic Performance Rating. AMCA International, Arlington Heights, IL; determines airflow rate, pressure developed, power, density, speed and efficiency for rating purposes.

[9] European Parliament and Council (2024). Regulation (EU) 2024/573 on fluorinated greenhouse gases, amending Directive (EU) 2019/1937 and repealing Regulation (EU) No 517/2014. Official Journal of the European Union; in force 11 March 2024, with labelling and quota provisions applying from 1 January 2025. https://eur-lex.europa.eu/eli/reg/2024/573/oj/eng

[10] American Society of Heating, Refrigerating and Air-Conditioning Engineers (2020). ANSI/ASHRAE Standard 84: Method of Testing Air-to-Air Heat/Energy Exchangers. ASHRAE, Atlanta, GA; covers regenerative wheels, heat pipes, thermosiphons, run-around loops and fixed-plate exchangers, in laboratory and field tests. A 2024 edition has since been issued.

[11] International Organization for Standardization (2013). ISO 16358-1: Air-cooled air conditioners and air-to-air heat pumps - Testing and calculating methods for seasonal performance factors - Part 1: Cooling seasonal performance factor. ISO, Geneva; Amendment 1 (2019) and Amendment 2 (2024). Covers equipment within the scope of ISO 5151, ISO 13253 and ISO 15042. https://www.iso.org/standard/56467.html

[12] Shingo, S. (translated by A. P. Dillon) (1986). Zero Quality Control: Source Inspection and the Poka-Yoke System. Productivity Press, Cambridge, MA. Originally published in Japanese, 1985.

[13] Nakajima, S. (1988). Introduction to TPM: Total Productive Maintenance. Productivity Press, Cambridge, MA.

[14] International Electrotechnical Commission / International Society of Automation (2025). Enterprise-control system integration - Part 1: Models and terminology. IEC 62264-1; ANSI/ISA-95.00.01-2025 (IEC 62264-1 Mod).

[15] International Society of Automation (2013). Enterprise-control system integration - Part 3: Activity models of manufacturing operations management. ANSI/ISA-95.00.03-2013 (IEC 62264-3 Modified).

[16] International Electrotechnical Commission (2025). IEC 62541-1: OPC unified architecture - Part 1: Overview and concepts. IEC, Geneva. https://webstore.iec.ch/en/publication/81513

[17] International Organization for Standardization (2014). ISO 22400-1: Automation systems and integration - Key performance indicators (KPIs) for manufacturing operations management - Part 1: Overview, concepts and terminology. ISO, Geneva. https://www.iso.org/standard/56847.html

[18] International Organization for Standardization (2014). ISO 22400-2: Automation systems and integration - Key performance indicators (KPIs) for manufacturing operations management - Part 2: Definitions and descriptions. ISO, Geneva; defines OEE as availability x effectiveness x quality ratio. https://www.iso.org/standard/54497.html

[19] International Organization for Standardization / International Electrotechnical Commission (2024). ISO/IEC 19987: Information technology - EPC Information Services (EPCIS) Standard, version 2.0. ISO/IEC, Geneva. https://www.iso.org/standard/85557.html

[20] Lee, J., Bagheri, B., and Kao, H.-A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters, 3, 18-23.