Executive Overview

Reliable predictive maintenance in safety-critical systems requires more than high retrospective accuracy; it requires an explicit understanding of data structure, uncertainty, and failure risk. In time-series failure modelling, models trained without prior structural diagnosis may appear accurate in controlled evaluation while failing unpredictably in operational settings.

This report documents the exploratory data analysis (EDA) conducted for FusionCore v0, a research and benchmarking initiative designed to de-risk architectural decisions prior to model benchmarking and system scaling. EDA is treated not as a preparatory step for feature engineering, but as a diagnostic process whose purpose is to expose structural properties of the data that materially constrain model behaviour, generalisation, and decision risk.

Two stages of analysis are presented. EDA-01 examines each dataset independently to characterise internal lifecycle structure and sensor behaviour. EDA-02 extends this analysis across datasets and engines to expose generalisation risk, uncertainty, and early-life signal availability. Together, these analyses establish the empirical foundation required for responsible multi-model benchmarking in subsequent FusionCore v0 stages.

1. Role of EDA in Risk-Aware Temporal Modelling

Predictive maintenance differs fundamentally from conventional forecasting. The objective is not merely to minimise error, but to manage asymmetric risk: false negatives may result in asset damage or catastrophic failure, while false positives incur operational cost and disruption.

Exploratory data analysis is therefore framed as a risk identification exercise, answering structural questions that performance metrics alone cannot address.

1.1 Why Benchmarking Without Prior EDA Is Misleading

Common Modelling Assumption	Empirically Violated Because
Stationarity	Sensor dynamics evolve with lifecycle progression
Asset homogeneity	Engines exhibit heterogeneous degradation paths
Metric sufficiency	Error costs are asymmetric and time-dependent
Early signal presence	Degradation often emerges late in life

Without EDA, benchmarking conflates genuine predictive capability with artefacts arising from lifecycle imbalance, sensor dominance, or delayed signal emergence.

2. Dataset Context & Experimental Scope

The analyses are conducted on the NASA CMAPSS FD00x dataset family, treated as related but non-interchangeable operational regimes. The purpose of this section is to define analytical scope, not to provide exhaustive dataset documentation.

2.1 Experimental Contract

Dimension	Decision
Dataset family	CMAPSS FD00x only
Dataset pooling	Explicitly avoided
Model training	None
Feature engineering	None
Causal inference	Not attempted
Regime clustering	Not performed

These constraints are deliberate design choices that preserve interpretability and prevent premature optimisation.

3. EDA-01 — Within-Dataset Structural Diagnostics

EDA-01 examines each dataset independently to determine whether consistent degradation information exists and under what constraints it may be learnable.

3.1 Engine Lifecycle Length Distributions

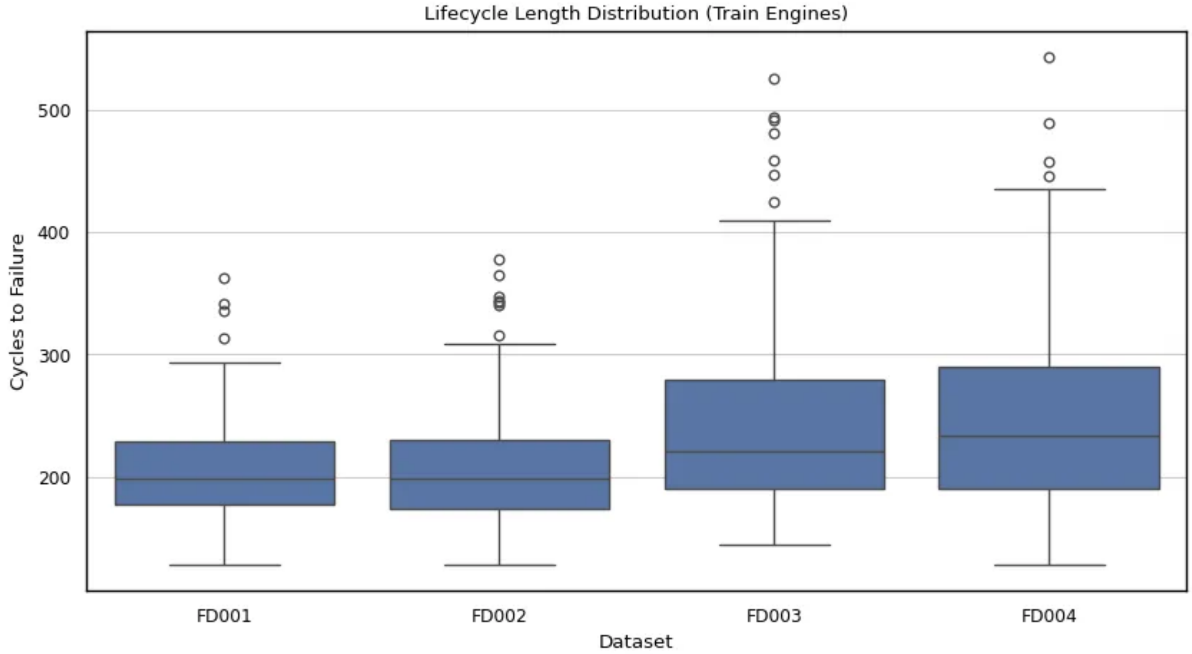
Lifecycle length, defined as the number of operational cycles observed prior to failure, determines the temporal information available for modelling. Distributions were analysed using box plots and empirical cumulative distribution functions (ECDFs).

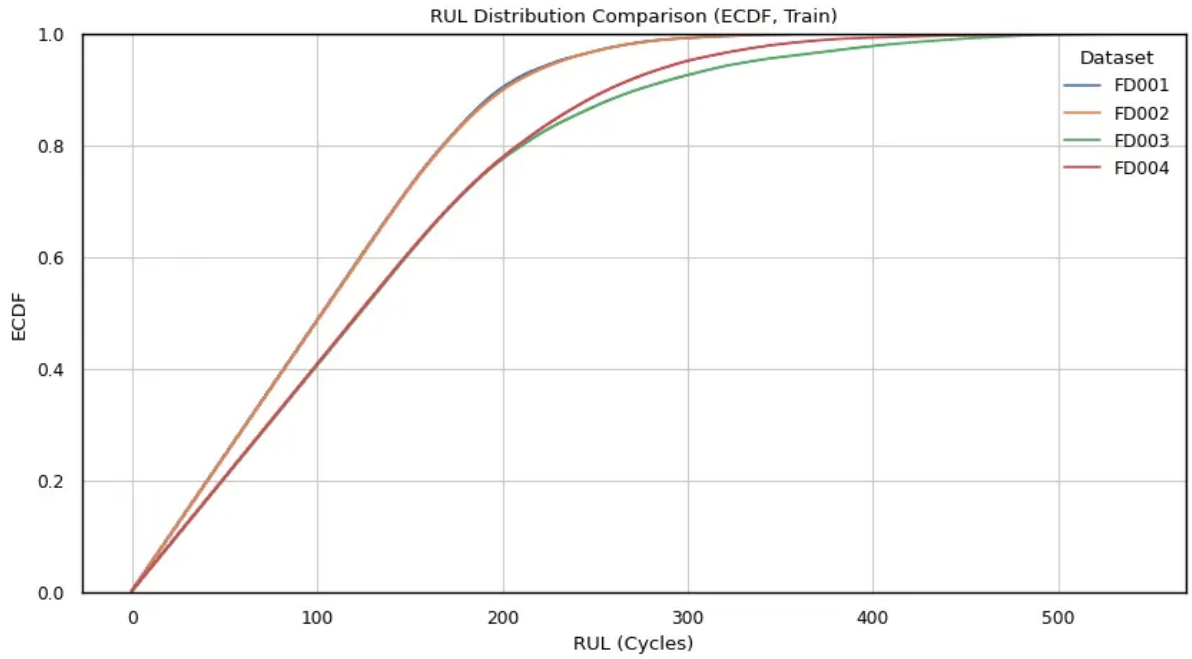
An ECDF shows, for any given value x, the proportion of observations that are less than or equal to x.

$$\mathrm{ECDF}(x) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq x) \qquad \text{Eq. 1}$$

Where:

- $X_1, \ldots, X_n$ are observed values (here: lifecycle lengths or RUL values)
- $\mathbf{1}(\cdot)$ is an indicator function (1 if true, 0 otherwise)





Key structural observations:

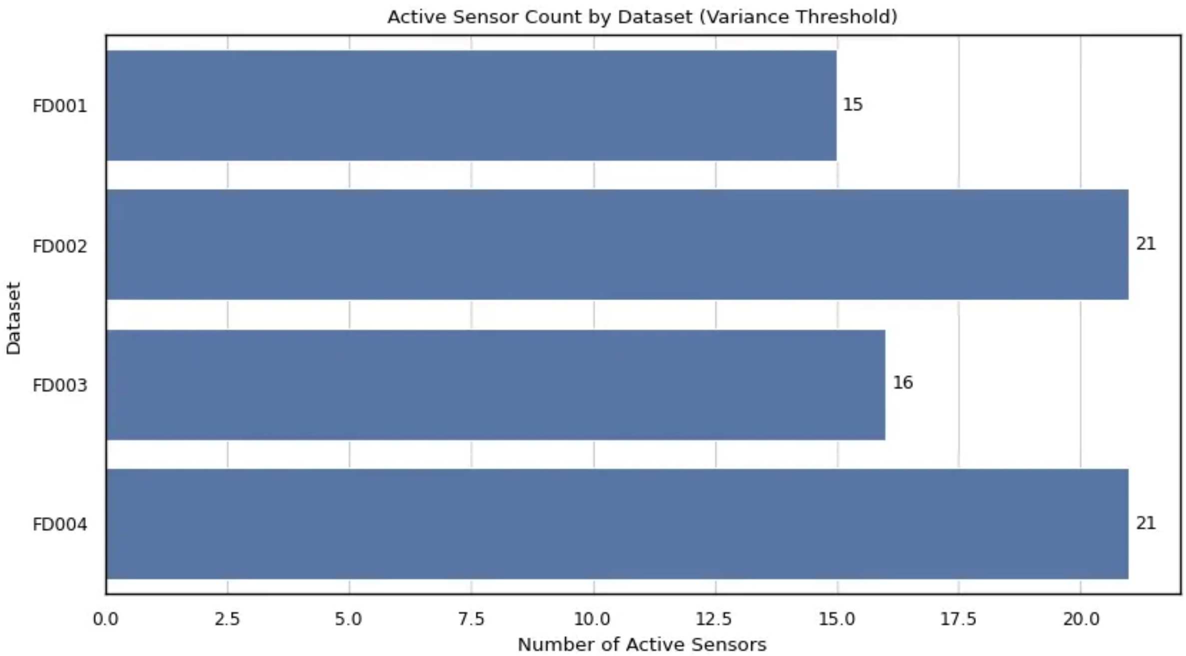
Observation	Modelling Constraint
Wide lifecycle dispersion	Models must handle variable temporal horizons
Long distribution tails	Early-life data imbalance
Dataset-specific medians	Regime-specific calibration required

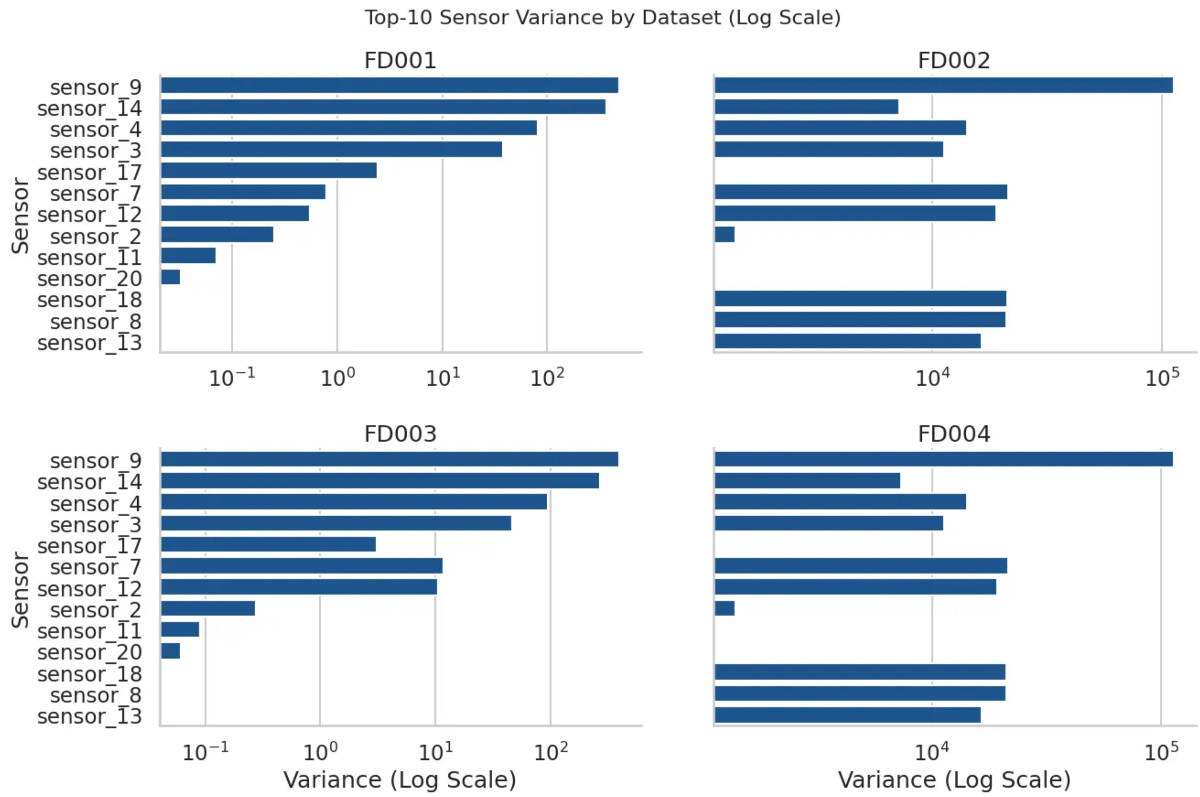
These findings invalidate assumptions of uniform degradation timing within datasets.

3.2 Sensor Variance and Stability

Sensor variance was computed per dataset and ranked to expose dominance patterns. Variance is displayed on a logarithmic scale to accommodate multi-order magnitude differences. Sensors were classified as active using a fixed variance threshold τ.

$$\tau = 1 \times 10^{-6} \quad \text{Eq. 2}$$





Structural synthesis:

Finding	Architectural Risk
Few sensors dominate variance	Noise-driven learning
Active sensor sets differ by dataset	Feature instability
Threshold-dependent counts	Non-transferable heuristics

Importantly, sensitivity analysis shows that while absolute counts vary with $τ$, the relative dominance structure remains stable, indicating intrinsic dataset properties rather than analytical artefacts.

3.3 EDA-01 Summary

Within-dataset diagnostics establish that:

- Lifecycle heterogeneity is structural, not noise-driven.
- Sensor relevance is sparse and unevenly distributed.
- Distributional properties convey more information than averages.

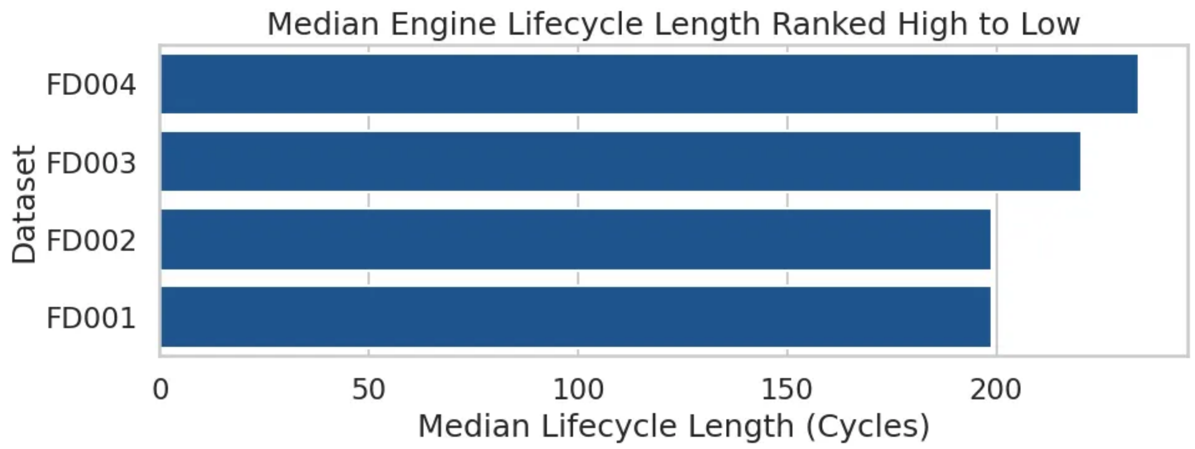
These constraints motivate cross-dataset analysis.

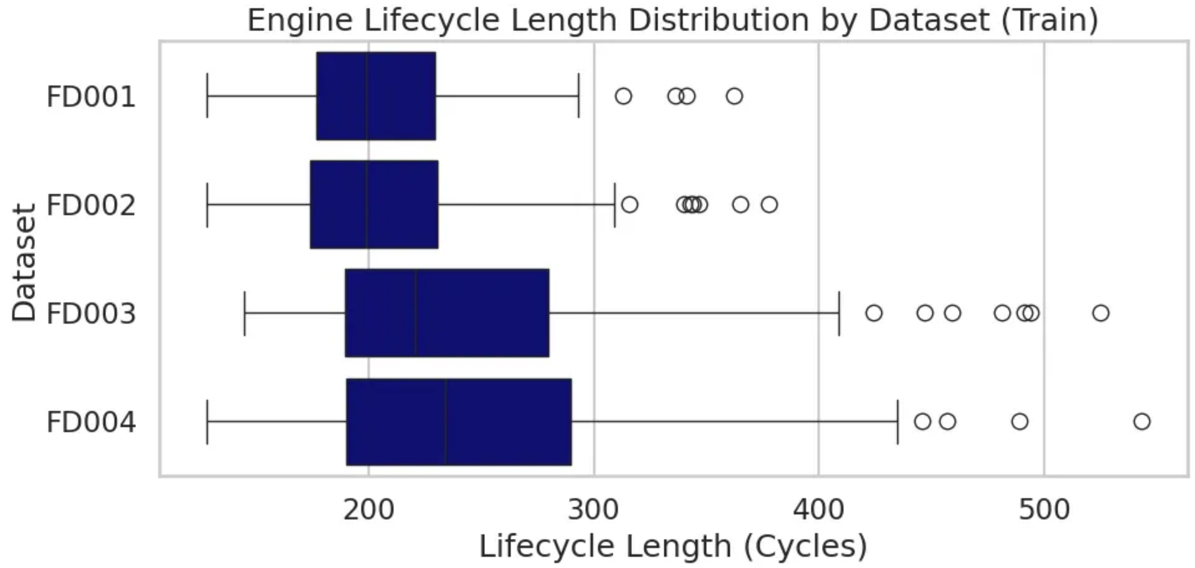
4. EDA-02 — Cross-Dataset & Cross-Engine Diagnostics

EDA-02 exposes generalisation risk by comparing lifecycle structure, degradation behaviour, uncertainty, and early-life signal availability across datasets and engines.

4.1 Cross-Dataset Lifecycle Comparison

Median lifecycles and full distributions were compared using box plots and ECDFs. Results reveal a clear hierarchy of regimes, ranging from relatively homogeneous lifecycles to highly dispersed, long-tailed structures. Lifecycle structure is therefore dataset-specific and cannot be assumed to generalise.



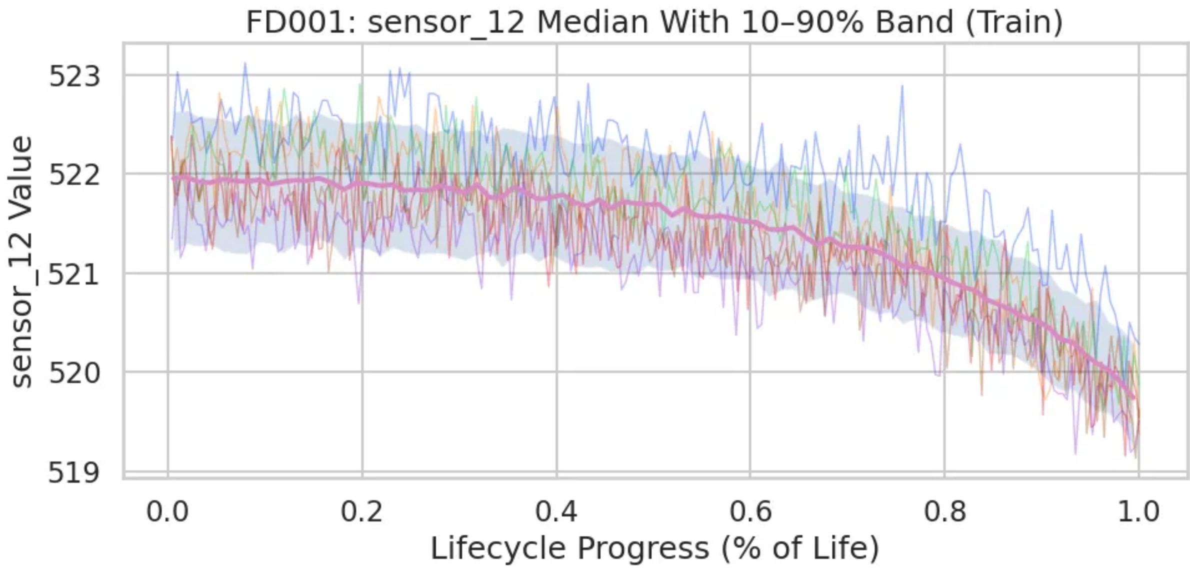


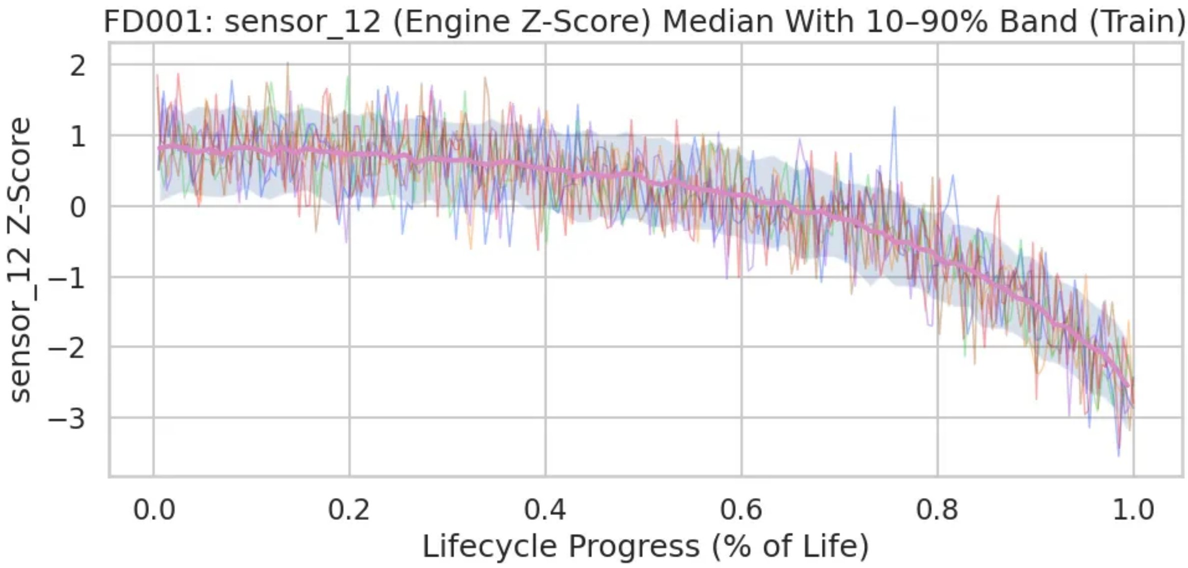
Similar to 3.1 Observations and Outcomes: "Lifecycle Lengths Distribution (Train Engines)"

4.2 Engine-Normalised Degradation Trajectories

Raw sensor trajectories obscure degradation shape due to engine-specific baselines and operating conditions. To enable comparison:

- Time was normalised as a percentage of total lifecycle.
- Sensor values were z-score normalised per engine.





$$z_{i,t} = \frac{x_{i,t} - \mu_i}{\sigma_i} \qquad \text{Eq. 3}$$

where:

- $x_{i,t}$ is the sensor value for engine i at time t,
- μ_i and σ_i are the mean and standard deviation of that sensor for engine i.

Following normalisation, median trajectories and uncertainty envelopes reveal that engines within the same dataset may follow qualitatively different degradation paths, ranging from gradual decline to abrupt late-stage changes. This heterogeneity challenges models that assume a single canonical failure trajectory.

4.3 Quantile Bands and Uncertainty Envelopes

Uncertainty was characterised using quantile bands (10th–90th percentiles) around median trajectories.

$$Q_p(X) = \inf\{x: F_X(x) \geq p\} \quad \text{Eq. 4}$$

Core finding:

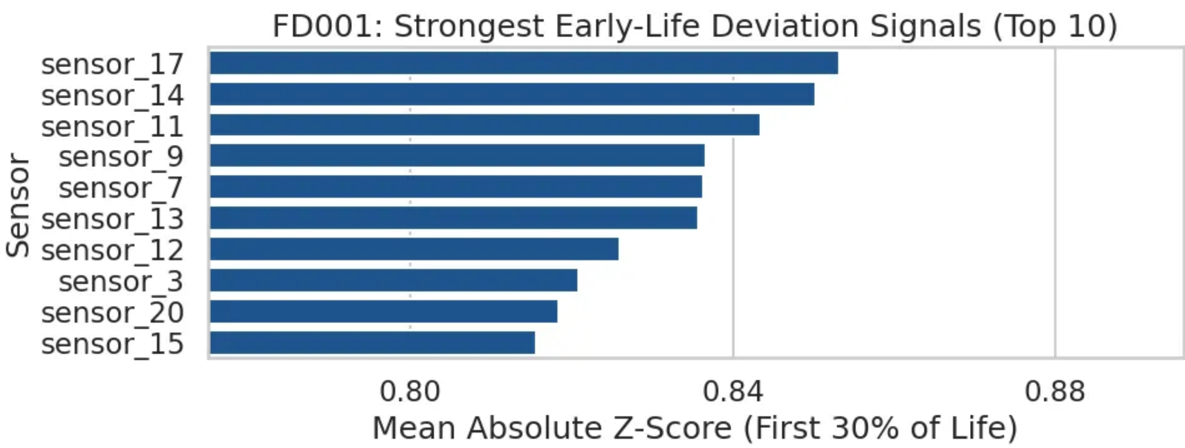
Uncertainty widens as failure approaches—precisely when accurate decision-making is most critical—justifying probabilistic or ensemble modelling approaches.

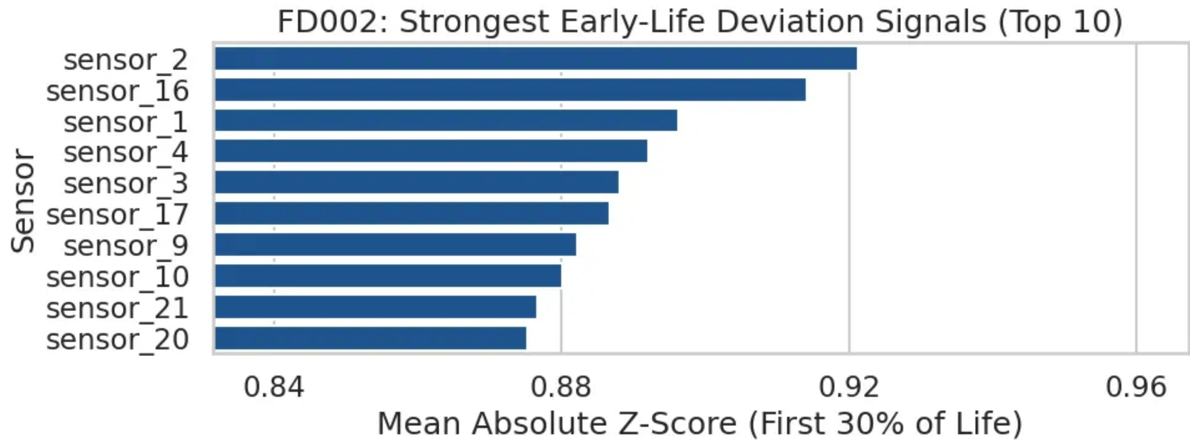
4.4 Early-Life Signal Detectability

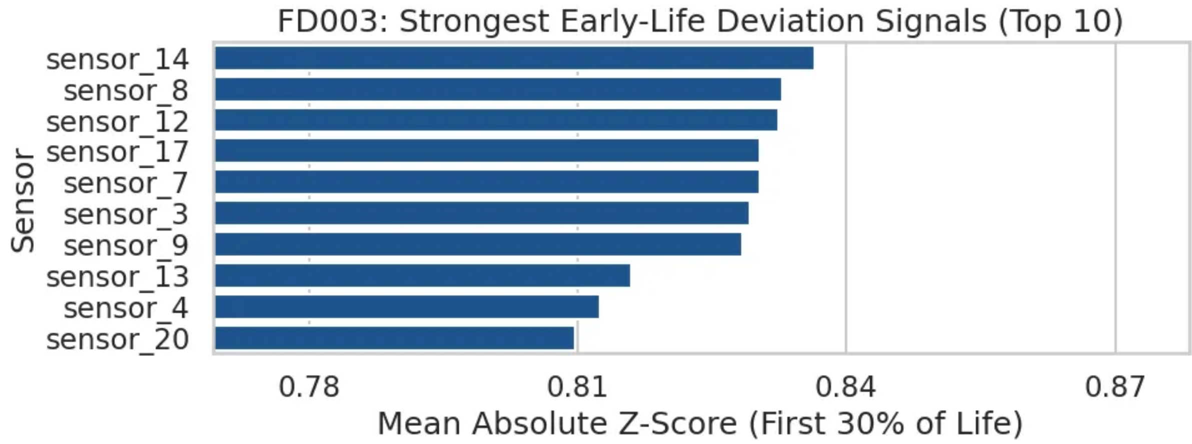
Early-life behaviour was analysed over the first 30% of lifecycle using mean absolute z-score deviation.

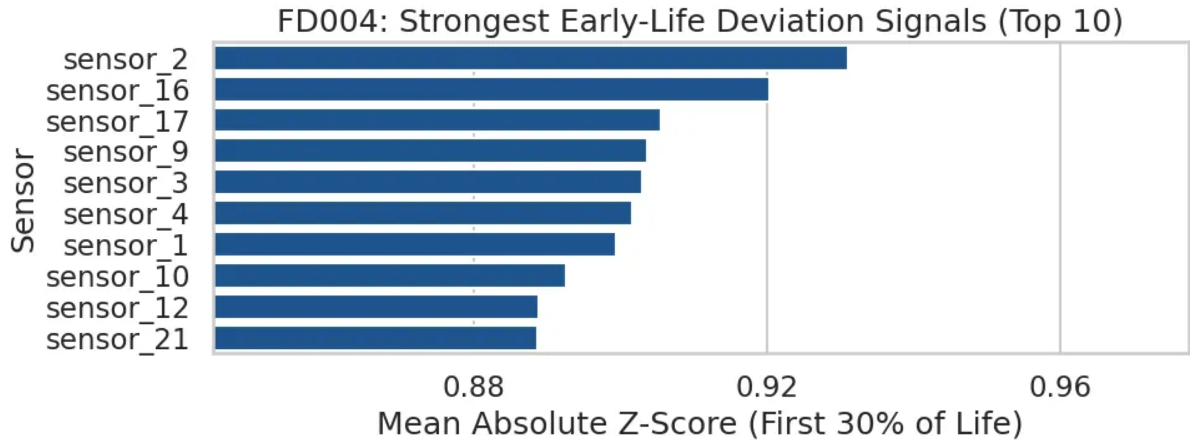
$$S = \frac{1}{T}\sum_{t=1}^{T}|z_t| \quad \text{Eq. 5}$$

where T denotes the number of time steps in the early-life window.









Dataset-level synthesis:

Dataset	Early Signal Strength	Operational Risk
FD001	Strong	Low detection latency
FD002	Weak	High false-negative risk
FD003	Moderate	Calibration sensitivity
FD004	Moderate	Elevated uncertainty

Signal availability is therefore regime-dependent, with direct implications for early-warning capability.

5. Architectural Risk Synthesis

The combined EDA findings expose structural risks that benchmarking must explicitly test.

Structural Property	Failure Mode
Lifecycle mismatch	RUL miscalibration
Sensor dominance	Dataset-specific overfitting
Heterogeneous trajectories	Average-case collapse
Late-emerging signals	False negatives

Benchmarking that ignores these properties risks selecting brittle architectures.

6. Why This EDA Enables FusionCore v0

Benchmarking without prior structural diagnosis is not merely incomplete; it is misleading. FusionCore v0 is therefore positioned as a filtering layer, not a deployable product. Its role is to identify which modelling approaches are structurally compatible with the data and which are likely to fail under realistic conditions.

The EDA establishes non-negotiable constraints that any viable architecture must satisfy, justifying a multi-model benchmarking strategy driven by data structure rather than algorithmic preference.

7. Limitations and Assumptions

- Analyses are limited to CMAPSS FD00x datasets.
- Findings are descriptive, not causal.
- Sensor interactions are not explicitly modelled.
- Regime clustering is not formally explored.
- Threshold-based definitions are analytical tools, not intrinsic truths.

These limitations define the scope of valid inference.

8. Transition to Benchmarking

Questions Answered by EDA

- How lifecycle structure varies across regimes
- Where uncertainty concentrates
- When degradation signals emerge
- Which structural risks dominate

Questions Requiring Benchmarking

- Which architectures tolerate lifecycle variability
- How models trade off early detection vs false positives
- Sensitivity to sensor dominance
- Stability under regime shift

The transition to benchmarking is therefore logical, constrained, and evidence-driven.

9. EDA Findings → Benchmarking Hypotheses

EDA Finding	Benchmarking Hypothesis	Primary Failure Mode if Violated	Architectures Most Stressed	Key Evaluation Metrics
Lifecycle lengths are heterogeneous within and across datasets	Models with fixed temporal assumptions will miscalibrate RUL under variable lifecycles	Systematic RUL bias (short-life overestimation, long-life underestimation)	Fixed-window regressors, short-context temporal models	RUL error stratified by lifecycle percentile; late-life bias
Sensor variance is dominated by a small, dataset-specific subset	Models implicitly weighting features by variance will overfit dataset artefacts	Apparent in-dataset performance with poor cross-dataset generalisation	Linear models, unregularised deep models	Cross-dataset performance delta; feature attribution stability
Degradation trajectories are heterogeneous across engines	Models optimised for average behaviour will fail on minority failure modes	Low mean error but catastrophic engine-level failures	Single-path models, overly smoothed temporal models	Worst-engine error; tail (90th percentile) error
Prediction uncertainty widens near failure	Point-estimate-only models will be overconfident at critical decision points	Unsafe confidence near failure, poor risk calibration	Deterministic regressors, single-model DL	Calibration error vs lifecycle stage; prediction interval coverage
Early-life signal availability varies by dataset	Models optimised for final accuracy will exhibit unacceptable detection latency	Late detection and high false-negative rates	Full-lifecycle loss models without temporal weighting	Time-to-detection; early-life false-negative rate
FD00x datasets represent distinct operational regimes	No single model will dominate across all regimes without adaptation	Overfitting to “easy” regimes; misleading global rankings	Highly specialised or over-tuned models	Per-dataset rank stability; performance variance across regimes

Benchmarking in FusionCore v0 is treated as risk evaluation, not model selection.

Models are assessed on *where*, *when*, and *how* they fail under known structural constraints, not solely on aggregate accuracy.