ASTRASTATE SYSTEMS

G0 Technical White Paper

Version 2 · 24 October 2026

Correction notice. v2 supersedes the September 2026 version. v1 normalised each failure by its own full-life statistics, which a deployed system cannot do. We corrected this and added baselines; the headline moved from 8.79 days (27.9%) to 10.1 days (about 16%).

Executive Summary

Actuator failure on spacecraft — brushless DC motors, harmonic drives, ball screws — is difficult to predict and expensive to get wrong. Unscheduled failure risks mission loss; overly conservative maintenance schedules waste flight hardware and budget that could extend mission life.

G0 is a physics-informed AI system, built by AstraState Systems, that is designed to estimate Remaining Useful Life (RUL) for these actuator types. It is deployed as an air-gapped, on-premise container with no telemetry and no external network dependency — a deployment model built for environments where sensor data cannot leave the customer's control.

This paper reports one primary result: RUL prediction error on real bearing run-to-failure data from the IMS/Rexnord dataset, using a leave-one-out protocol across all four documented real failures, with causal preprocessing and compared against a naive baseline and an RMS-only baseline. It also summarises two further public bearing benchmarks on which G0 showed no reliable advantage over simple baselines. Everything in this paper concerns bearing vibration data; nothing here validates G0 on spacecraft actuators.

1. The Problem

Spacecraft actuators degrade gradually before they fail outright — bearing wear, lubricant breakdown, gear mesh degradation. The physical signal is present in motor current, vibration, temperature, and position error well before failure. The difficulty is threefold:

- Failure data is scarce. Spacecraft actuators rarely run to failure in the field — that is the point of a space program. Real run-to-failure datasets are limited to ground-test rigs and industrial analogues.
- Physics-only models are hard to calibrate across operating regimes; pure black-box ML models generalize poorly outside the exact conditions they were trained on and offer no physical interpretability.
- Any prognostics claim is only as strong as the data it was tested against, and only as honest as the paper is about what that data can and cannot support.

G0 pairs physics-informed modelling with an evaluation method designed to avoid the most common way these numbers get overstated: tuning against the same data used to report performance.

2. Methodology

2.1 Preprocessing and baselines

Features are normalised using only each trajectory’s first 20 snapshots and training-fold statistics, so nothing in preprocessing uses information from the held-out failure’s future (causal preprocessing). The September 2026 version normalised each failure using statistics from its full life, which a deployed system cannot do; that has been corrected and the results restated. Two

baselines are reported: a naive baseline (the mean training RUL, capped at 40 days to match the model's training target) and an RMS-only baseline that uses vibration RMS alone.

2.2 Evaluation Data

The result in this paper is evaluated against the IMS/Rexnord bearing run-to-failure dataset: four documented real bearing failures, evaluated leave-one-out — the model trains on three failures and is tested on the fourth, held out entirely, repeated so each of the four failures serves as the test case once. Only vibration features were used in this evaluation, although the product takes five input signals. Data provenance: IMS bearing data from the Center for Intelligent Maintenance Systems, University of Cincinnati, with Rexnord, hosted in NASA's Prognostics Data Repository. This is not a NASA test and not NASA-endorsed.

2.3 Avoiding Test-Set Leakage

A single, fixed model configuration was used across all four held-out splits, and the reported results are the mean of three training seeds (42, 43, 44). No per-split hyperparameter search was used to select the reported result, because selecting the best configuration per test split would let information about the held-out failure leak into model selection. The number in this paper reflects one configuration, evaluated identically across all four real failures, with no tuning advantage given to any individual split.

3. Result: RUL Prediction Accuracy

Evaluated leave-one-out on the four documented IMS/Rexnord failures, G0 achieved a mean remaining-life error of 10.1 days, about 16% lower than a naive baseline (12.1 days) and 14% lower than an RMS-only baseline (11.8 days). G0 beat the naive baseline on one of four failures (set2_bearing1) and was slightly worse on the other three, so the average gain comes mainly from one failure. It beat the RMS-only baseline on two of four.

Held-out failure	G0 (days)	Naive (days)	RMS-only (days)	G0 vs naive
set1_bearing3	10.28	9.74	9.73	−5.5%
set1_bearing4	10.04	9.74	8.47	−3.1%
set2_bearing1	5.82	15.12	8.78	+61.5%
set3_bearing3	14.37	13.84	20.18	−3.8%
Mean	**10.13**	**12.11**	**11.79**	**+16%**

Table 1. Mean of three training seeds (42, 43, 44), one fixed configuration, no per-failure tuning. Features are normalised using only each trajectory's first 20 snapshots and training-fold statistics, so nothing in preprocessing uses information from the held-out failure's future. The naive baseline is the mean training RUL capped at 40 days, matching the model's training target. Positive percentages mean G0's error is lower.

This result answers the question a maintenance planner needs answered: given a real degrading actuator, how close is the predicted remaining life to the true remaining life? It does not address alarm-timing false-positive rate, which is not reported in this version of the paper.

3.1 Other public bearing benchmarks

We also tested two further public bearing datasets: University of Ferrara (6 failures) and FEMTO/PRONOSTIA (17 failures). On these, G0 showed no reliable advantage over simple baselines. Protocols and per-failure results are available on request.

4. Deployment Model

G0 ships as an air-gapped Docker container with a hardware-locked annual license, designed for environments — defense, aerospace, and other regulated sectors — where sensor data cannot leave customer premises under any circumstances.

- No telemetry, no phone-home, no logs leaving the deployment environment at any point in normal or emergency operation.
- Pooled-token licensing across a customer's rig fleet, with a graceful degrade-to-lockout sequence on connectivity loss rather than an abrupt hard stop, and a signed, out-of-band emergency-unlock procedure for extended outages that requires no telemetry or remote access to use.
- The shipped model artifact includes trained inference components only (context encoder, energy model, physics head); training code and raw model weights are not included in the deployed container.

5. Scope and Limitations

This paper reports RUL prediction accuracy on bearing vibration data only. It does not report an alarm false-positive rate, on either real or synthetic data, nor interval calibration (how often the uncertainty intervals contain the true value), nor results using signals other than vibration features, nor results on spacecraft actuators, motors, harmonic drives or ball screws. Readers evaluating G0 for a maintenance program should treat alarm-timing performance as a separate question, and should ask any prognostics vendor — including us — for the specific evaluation protocol and dataset behind any alarm-calibration number before relying on it.

The IMS/Rexnord dataset provides four real bearing failures. This is the standard public dataset for this kind of evaluation, and leave-one-out across all four is a conservative protocol, but four failures is a small sample, and the average gain comes mainly from one of them. Table 1 shows the per-failure spread rather than only the mean.

6. Conclusion

On public bearing data, G0 showed a modest reduction in mean RUL error against a naive baseline (about 10.1 against 12.1 days), driven mainly by one of four failures, and no reliable advantage over simple baselines on two further public benchmarks. This is early-stage evidence on bearing vibration only. It does not establish performance on spacecraft actuators; for that we propose blind pilot validation on partners’ own run-to-failure data.

For a technical evaluation or deployment discussion, contact AstraState Systems.

7. Glossary

- RUL (Remaining Useful Life): the estimated time remaining before a component is expected to fail, expressed in days.
- Leave-one-out evaluation: a validation method where the model is tested on one real failure it has never seen, having been trained only on the others — repeated so every real failure gets a turn as the held-out test case.
- RUL-MAE: mean absolute error between predicted and true remaining useful life, in days.
- Naive baseline: a simple non-learned predictor (mean RUL of the training failures, capped at 40 days to match the model’s training target) used as a reference point for how much a learned model actually improves on the simplest reasonable guess.
- RMS-only baseline: a simple predictor that uses vibration RMS alone, reported alongside the naive baseline.

- Causal preprocessing: normalisation that uses only data available up to the time of each prediction, with no statistics from the future or from the failure’s whole life.
- Interval calibration: how often a stated uncertainty interval actually contains the true value. Not yet reported for G0.
- Held-out data: data never used during training, calibration, or model selection — used only once, at the end, purely to measure performance.
- Air-gapped: a deployment with no network connection to the outside world, used where data cannot be permitted to leave a secure facility under any circumstances.