

Silence Is Not Safety: Minute-Level Forensics of Spur Gear Spall Progression from Low-Cost Condition Indicators

Sang Bong Song

Song Industry Research, Seoul, Republic of Korea

sangbong908@gmail.com

Sang Bong Song. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

EXTENDED ABSTRACT

Extended abstract for the PHM North America 2026 Data Challenge; validation entry second of nineteen valid submissions from thirty-five registered teams. The full paper follows after the conference, accompanied by the extraction scripts and the pre-registered gate protocol that make the analysis reproducible end to end.

1. PRIOR WORK AND POSITION

The sensitivity loss of classical condition indicators during damage progression has been on record since the early NASA gear test programs: FM4 peaks early in pitting tests and returns toward its nominal value as damage spreads (Zakrajsek et al., 1993; Decker et al., 1994), and Dempsey (2000) drew the practical conclusion that vibration indicators alone are an unreliable basis for damage assessment, arguing for fusion with oil-debris measurement. Lewicki et al. (2009) rated FM4 and RMS below average for detecting surface pitting on a helicopter-scale test program. That the indicator families are not interchangeable is also established: Bechhoefer and Butterworth (2019), ranking eighty-eight analyses by separability on a cracked tooth, retained the energy operator despite its weak showing there on the reasoning that no single indicator serves every gear fault, and closed by cautioning that a pitting fault would call for a different set. What the present work adds is an axis: their comparison runs across failure modes and across acquisitions, whereas the separation reported here is temporal -- within a single progression of a single mode, resolving inside one six-hour run.

2. THE DATA AND ITS DIFFICULTIES

Four properties of the challenge data shaped the submission, and each was met with an explicit rule rather than a modeling choice. First, every run opens with a 10-20 minute transient in the energy-operator channels regardless of damage state -- the signature of inspection, reassembly, and load settling, indistinguishable in magnitude from initiation signal. The first 22 minutes of every run were excluded; the

window was set by cross-experiment observation, pre-registered, and is protocol-specific, to be re-estimated on any other installation. Second, the damage labels bound run totals only: photographs exist at run boundaries, so the within-run timing of any advance is not observed. Claims were restricted accordingly; the paper adopts the conventional reading that mass loss is concurrent with the higher-moment response, and states the alternative it excludes. Third, the noise question was answered by declaration rather than by default: within each minute the upper decile of second-level values is treated as signal, an assumption admissible here because speed and load are fixed, and stated as the condition under which every result should be read. Fourth, the scored metric is photo-derived surface area: it saturates after each mass-loss event and is not a life proxy, so persistent indicator activity was never read as proportional further damage.

3. APPROACH, AND WHAT IS NEW IN IT

The dataset stores one file per operating minute. The submission analyzes at that native resolution instead of the customary one-summary-per-run, carrying one value per indicator per minute as a p90 ratio to a healthy baseline locked from each experiment's opening runs. Read this way, the classical families separate into sequential roles: the energy-operator statistics rise first and fall back to baseline exactly as the higher moments (M8A*) take over -- in Experiment D, run 7, the handoff falls at minute 38 (Figure 1). The ordering is preserved wherever both phases are visible, while the durations vary by factors of 10-30 across gears; the ordering, not the durations, is the portable claim. The submitted trajectory for the validation gear was assembled from this grammar plus two anchors calibrated on the public leaderboard, giving explosion onset at minute 223 of run 16 and a terminal value of 9.70 within an admissible band of 9.3-9.7.

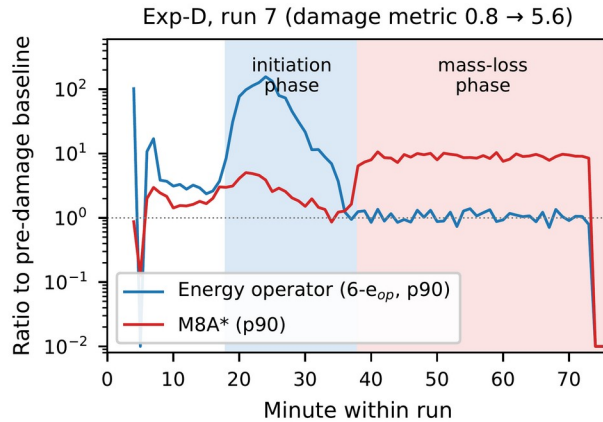


Figure 1. Minute-resolved condition-indicator response during the mass-loss transition of Experiment D, run 7 (damage metric 0.8 → 5.6). Both indicators are shown as p90 ratios to a pre-damage baseline (run 5 median). The energy-operator family responds first (initiation phase, min ~18-38) and returns toward baseline as M8A* activates (mass-loss phase, min 38 onward). The two indicators answer different questions and are sequential, not redundant. The innovation claimed is deliberately not a new statistic. It is a resolution shift plus a reading discipline: minute-level forensics of indicators three decades old, under interpretation rules locked before the data they judge were opened. One quantity does appear to be new at this resolution: a run that ends approximately +7.0 higher on the damage metric shows no sustained indicator response until the final nine of its 372 minutes.

4. HOW THE APPROACH WAS ASSESSED

Assessment was designed before it was needed. Interpretation gates were locked prior to opening the validation gear, and the record they produced includes one falsification: an early assumption that run 15 contained the mass-loss phase was contradicted by the minute-level record and corrected, and that failure is reported rather than smoothed. No feedback of any kind was taken from the validation gear -- zero submissions were scored on it before the final entry -- so the validation score (MSE 4.259, second among the nineteen valid entries) is the only external check on the trajectory and was unavailable during its construction. Three discarded approaches are documented in the full paper, including values produced by the author's own prior framework that conflicted with the sensor evidence and were dropped. Sixteen of thirty-five teams did not submit a valid validation entry, which we read as evidence of the problem's difficulty under the two-stage design (cf. Blum and Hardt, 2015) rather than of disengagement.

The operational maxim the data support is asymmetric: indicator activity is informative; indicator silence is not evidence of health -- both conditional on knowing which excursions the statistics should count, a question about the machine rather than the statistic.

DECLARATION OF AI USE

The author directed this investigation, decided which lines of analysis to pursue and which to discard, and is responsible for the findings. Analysis code, figures, and drafting were produced with Anthropic's Claude working under the author's direction; the author verified all outputs.

REFERENCES

- Bechhoefer, E., & Butterworth, B. (2019). A comprehensive analysis of the performance of gear fault detection algorithms. Annual Conference of the PHM Society, 11(1). <https://doi.org/10.36001/phmconf.2019.v11i1.823>
- Blum, A., & Hardt, M. (2015). The Ladder: A reliable leaderboard for machine learning competitions. Proceedings of the 32nd International Conference on Machine Learning, PMLR 37, 1006-1014.
- Decker, H. J., Handschuh, R. F., & Zakrajsek, J. J. (1994). An enhancement to the NA4 gear vibration diagnostic parameter (NASA TM-106553). NASA Lewis Research Center.
- Dempsey, P. J. (2000). A comparison of vibration and oil debris gear damage detection methods applied to pitting damage (NASA/TM-2000-210371). NASA Glenn Research Center.
- Lewicki, D. G., Dempsey, P. J., Heath, G. F., & Shanthakumaran, P. (2009). Gear fault detection effectiveness as applied to tooth surface pitting fatigue damage (NASA/TM-2009-215667). NASA Glenn Research Center.
- Zakrajsek, J. J., Townsend, D. P., & Decker, H. J. (1993). An analysis of gear fault detection methods as applied to pitting fatigue failure data (NASA TM-105950). NASA Lewis Research Center.
- PHM Society. (2026). PHM North America 2026 Conference Data Challenge. <https://data.phmsociety.org/phm-north-america-2026-conference-data-challenge/>