An End-to-End PHM Framework for Industrial Compressor Gearbox Fault Detection and RUL Prediction

Changik Cho[1], Eric Bechhoefer[2]

[1,2]*GPMS International, Waterbury, VT, 05676, USA*

changik@gpms-vt.com
eric@gpms-vt.com

ABSTRACT

Gearbox health monitoring in industrial machinery is essential for maintaining production efficiency, reducing maintenance costs, and ensuring reliable operation. However, the validation of complete prognostic frameworks on real-world industrial gearboxes remains limited compared to studies conducted on controlled laboratory test rigs. This study presents a full end-to-end Prognostics and Health Management (PHM) framework validated on a run-to-failure dataset collected from an operational industrial spur gearbox used in compressor machinery. Vibration data is acquired using a radially mounted accelerometer while shaft speed is derived from a 1/rev tachometer signal. From the processed signals, condition indicators (CIs) are extracted across multiple vibration analysis domains, including residual signal analysis, amplitude and frequency demodulation, energy operator techniques, and sideband modulation analysis. From these algorithms, features sensitivity to gear fault modes such as tooth cracking, pitting, scuffing, and misalignment are calculated. These condition indicators are fused into a single scalar health index (HI) which can determine when maintenance should be performed. A Remaining Useful Life (RUL) is estimated using a high cycle fatigue crack growth model without requiring detailed material-specific parameters. The RUL provides the operator with actionable information as to when best to schedule maintenance. Results show that the proposed framework enables fault detection and provides physically interpretable RUL predictions, supporting condition-based maintenance decisions and improving operational readiness in industrial gearbox systems.

1. INTRODUCTION

Gearboxes are critical components in industrial rotating machinery and are widely used in compressors, turbines, pumps, and manufacturing systems (Jardine, 2006). Unexpected gearbox failures can lead to costly downtime, secondary equipment damage, and reduced production efficiency. Consequently, Prognostics and Health Management has become an important maintenance strategy for improving equipment reliability while reducing operational costs through condition-based maintenance (CBM).

Vibration analysis is the most widely adopted technique for gearbox condition monitoring because localized defects such as tooth cracking, pitting, scuffing, and misalignment in shafts, bearings, and gears produce characteristic dynamic signatures in vibration signals. Numerous signal processing methods have been developed for gearbox diagnostics, based on the Time Synchronous Averaging (TSA), including residual signal analysis, amplitude and frequency demodulation, and sideband analysis, all of which have demonstrated effectiveness in detecting gear faults (McFadden, 1987, Randall, 2011 and Antoni, J. 2007). In the aerospace sector, particularly for helicopter Health and Usage Monitoring Systems (HUMS), Bechhoefer (2019) and Cho (2025) present several run-to-failure and seeded-fault datasets collected under representative operating conditions. These datasets enable the development and validation of practical PHM algorithms across varying torque levels, and mission profiles. However, comparable datasets from industrial rotating machinery are rarely available due to proprietary restrictions, long degradation periods, and the difficulty of collecting complete run-to-failure data from production equipment. Consequently, most published industrial gearbox studies have been validated using laboratory test rigs under controlled operating conditions, while comparatively few demonstrate complete prognostic frameworks using real industrial run-to-failure datasets.

In addition to fault detection, practical PHM systems require a reliable assessment of machine health and an estimate of Remaining Useful Life (RUL). Although data-driven methods have recently shown promising prediction performance, they generally require large-labeled datasets including faults. It is difficult to define every fault in various machinery components and often data driven techniques lack physical interpretability (Zhao et al., 2019). Physics-based approaches offer a more explainable alternative by incorporating degradation mechanisms into the prediction process. Many existing methods rely on material-specific parameters that are difficult to obtain in industrial applications.

Recent studies have expanded rotating machinery PHM toward deep-learning-based fault diagnosis, advanced Health Index construction, and hybrid physics/data-driven prognostic approaches. Ahmad et al. (2024) provide a systematic review of recent deep-learning approaches for planetary gearbox fault diagnosis and highlight both their diagnostic capabilities and the challenges associated with their practical industrial implementation. Health Index construction has similarly evolved from individual statistical features toward signal-processing and machine-learning-based feature fusion approaches for representing machinery degradation (Zhou et al., 2022). For RUL prediction, recent physics-informed approaches have attempted to combine physical degradation knowledge with data-driven models to improve generalization when representative run-to-failure data are limited (Xiong et al., 2023). These developments highlight the continuing need for prognostic frameworks that combine interpretable condition indicators, robust health assessment, and degradation models while remaining practical for deployment on real industrial machinery.

This paper presents an end-to-end PHM framework for an industrial compressor gearbox using a real run-to-failure dataset. The proposed framework combines order-domain vibration processing, multi-domain condition indicator extraction, probabilistic HI construction using Nakagami statistical modeling, and physics-guided RUL prediction based on Paris' fatigue crack growth model. The proposed approach enables early fault detection and provides physically interpretable RUL estimates suitable for practical condition-based maintenance.

The main contributions of this work are:

- Development of a complete PHM framework from raw vibration acquisition to RUL prediction.
- Validation using a real industrial run-to-failure compressor gearbox dataset.
- Probabilistic Health Index construction using multiple condition indicators and Nakagami modeling.
- Physics-guided RUL prediction without requiring material-specific fatigue parameters.

2. END-TO-END PHM FRAMEWORK

Figure 1 shows a gear tooth contact surface with micropitting and scuffing tracks from the spur gear after 350 hours of degraded operations in industrial compressor. These scuffing tracks are adhesive wear marks formed when the oil film breaks down and metal-to-metal contact occurs repeatedly at the same tooth positions.

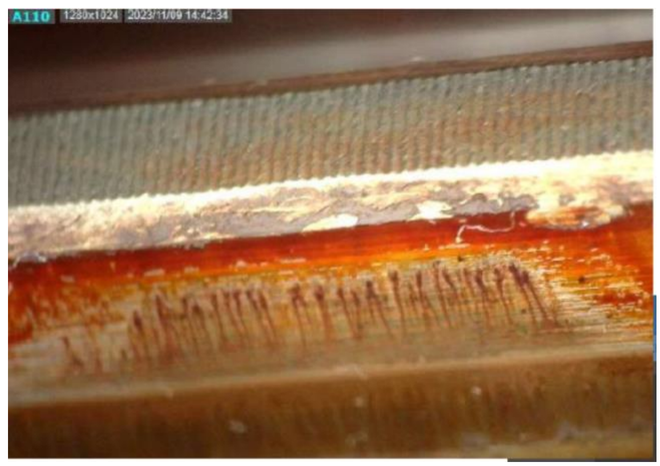


Figure 1 Scuffed Gear in Industrial Compressor

Figure 2 shows the overall workflow of the proposed PHM framework. During operation, vibration and tachometer signals are acquired every six minutes from the industrial

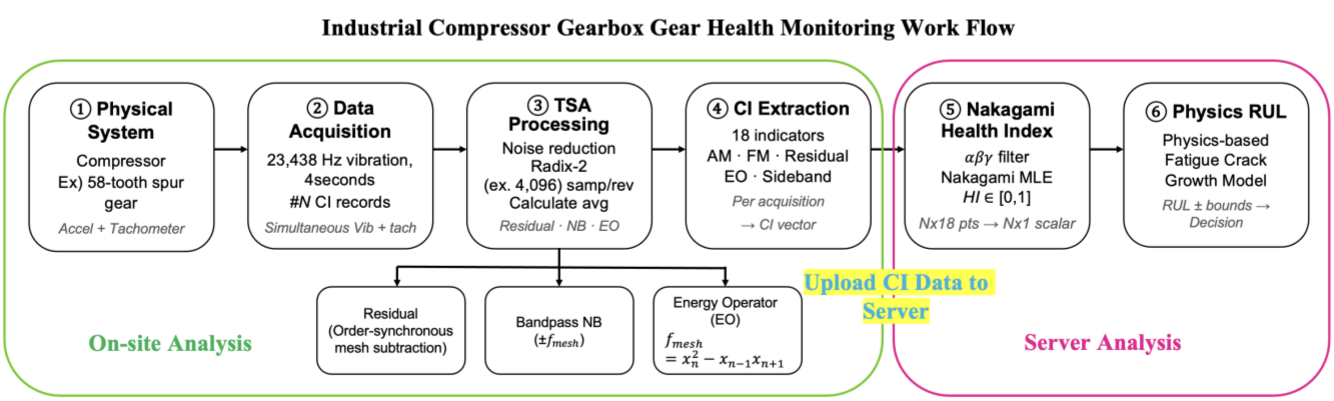


Figure 2 PHM End-to-End Workflow

compressor gearbox. The vibration signals are first converted into the order domain through angular resampling and processed using TSA to suppress non-synchronous vibration components and improve the signal-to-noise ratio (Bechhoefer, 2011). Multiple condition indicators are subsequently extracted from complementary signal processing techniques, including residual analysis, narrowband filtering, energy operator analysis, and amplitude/frequency demodulation.

To improve the practicality of continuous industrial monitoring, the proposed framework adopts a two-stage architecture consisting of on-site edge processing and server-side analysis. Rather than continuously storing and transmitting all high-frequency raw vibration signals, the on-site processor extracts a compact CI vector from each acquisition and periodically uploads only the condition indicators to the server. Since the CI representation is significantly smaller than the original vibration waveform, this approach substantially reduces communication bandwidth and long-term storage requirements while preserving the information required for health assessment.

Although retaining every raw vibration record would provide maximum diagnostic information, it is generally impractical for long-term industrial monitoring because of the large storage requirements associated with high sampling rates and continuous data acquisition. The proposed framework archives only a small subset of representative raw vibration records (approximately one out of every ten acquisitions). These archived signals allow maintenance engineers to perform detailed offline investigations, such as FFT analysis, TSA verification, and waveform inspection, whenever additional fault diagnosis is required.

The uploaded CI data are processed on the server, where storage and computational resources are less constrained. The server performs probabilistic HI construction using the Nakagami distribution model, followed by the physics-guided RUL prediction derived from Paris' fatigue crack growth law. By separating computationally intensive prognostic calculations from the on-site monitoring hardware, the proposed architecture provides an efficient and scalable solution for practical industrial gearbox health monitoring while maintaining access to representative raw vibration data for detailed engineering analysis.

2.1. Vibration Sensor and Tachometer

Figure 3 shows a representative vibration signal and RPM measured from the tachometer, respectively. The upper plot shows a 4-second vibration signal measured from the compressor gearbox using a radially mounted accelerometer sampled at 23,438 Hz. The lower plot shows the corresponding shaft speed measured by a 1/rev tachometer after tachometer jitter correction, providing a stable rotational reference for angular resampling and TSA.

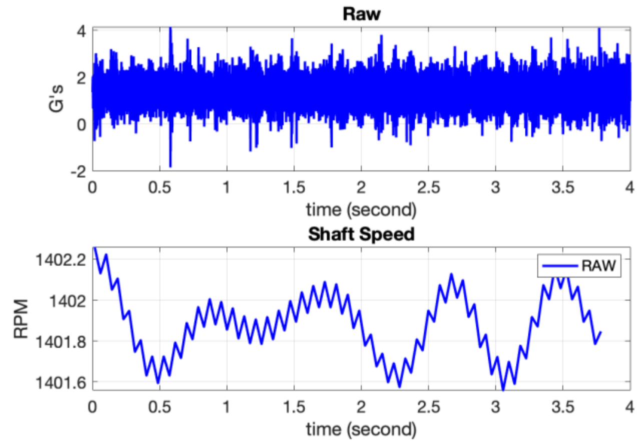


Figure 3 Example of the Measured Vibration and Shaft Speed

Tachometer signals are often affected by jitter, introducing noise that can lead to inaccurate analysis. To address this, a jitter-free algorithm developed by Bechhoefer and He (2016) is applied to the raw VR or optical sensor signal. TSA is then used to suppress noise unrelated to the shaft harmonics and gear mesh frequencies with shaft rate. After TSA processing, a Fast Fourier Transform (FFT) is applied to convert the vibration signal into the frequency domain. TSA length is chosen with radix-2 number for faster computation in Eq (1).

$$TSA\ length = 2\text{^}ceil(\log_2 sps/freq) \tag{1}$$

Where sps is sample rate, and $freq$ is frequency of shaft rate.

The complete run-to-failure dataset consists of 3,524 vibration acquisitions, each collected every 6 minutes, while 384 representative raw vibration records were selected for visualization and signal processing examples presented throughout this paper.

2.2. Time Synchronous Averaging (TSA)

Figure 4 shows TSA waveforms plotted over one shaft revolution. The three plots represent the beginning (First), middle (Middle), and end (Last) stages of the run-to-failure experiment. As gearbox degradation progresses, the TSA waveform shows increasing dynamic amplitude and localized impact-like CIs, indicating the development of gear tooth damage. The final acquisition shows pronounced impulsive behavior and waveform distortion, consistent with severe gear deterioration prior to failure.

Figure 5 shows the comparison between the frequency spectrum of the raw vibration signal and the corresponding TSA signal. The raw spectrum contains significant broadband noise and non-synchronous vibration components originating from the surrounding machinery and measurement environment. After TSA, these components are substantially attenuated, resulting in a lower spectral noise floor while preserving the gear mesh frequencies and their harmonics. The improved signal-to-noise ratio enables more reliable extraction of fault-sensitive condition indicators from the subsequent residual, narrowband, and demodulation analyses.

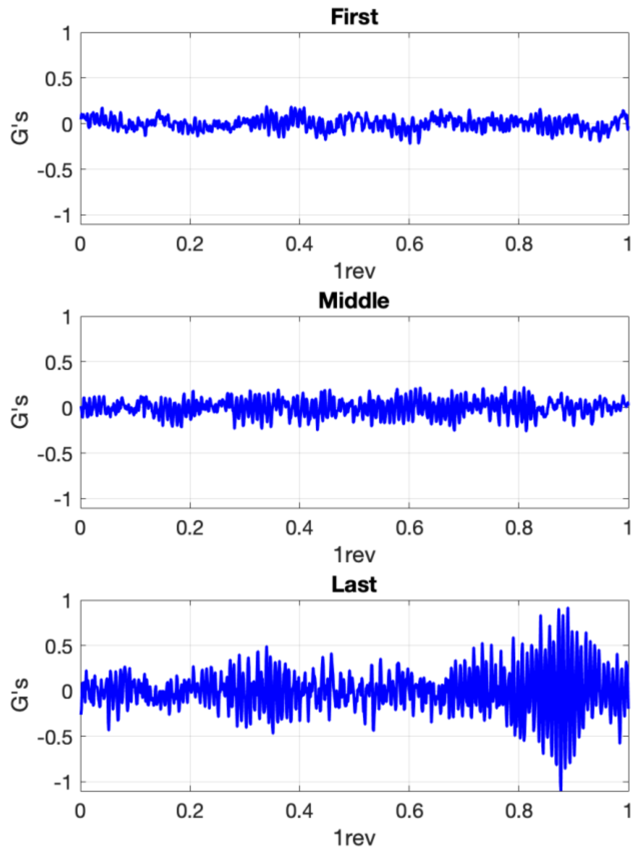


Figure 4 TSA Vibration Signals with Compressor Gear Degradation

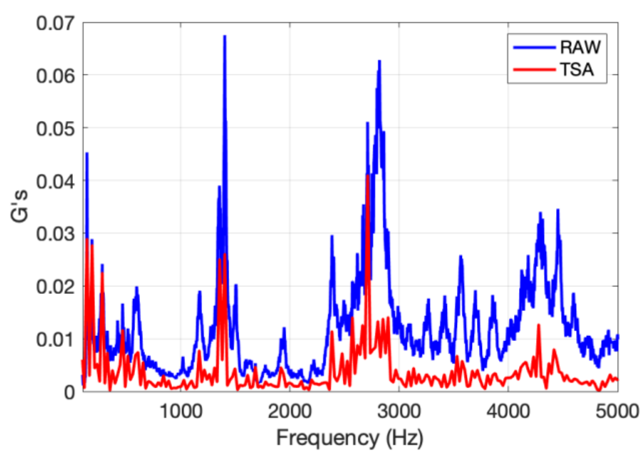


Figure 5 FFT of the Raw Vibration Signal and TSA signal

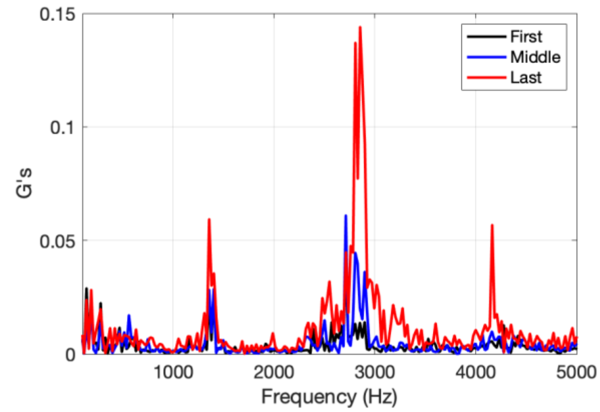


Figure 6 FFT of the TSA signals with Compressor Gear Degradation

Figure 6 compares the frequency spectra of the TSA signals at three representative stages of the run-to-failure experiment. During the early stage of operation, the gear mesh frequency (GMF) and its harmonics remain relatively stable with low sideband energy. As the gearbox degrades, increased modulation around the GMF results in higher harmonic amplitudes and more pronounced sideband components. Near failure, substantial energy growth is observed at the GMF, higher-order harmonics, and adjacent modulation frequencies, indicating severe localized gear tooth damage. These progressive spectral changes are subsequently quantified through multiple condition indicators derived from residual analysis, narrowband processing, amplitude and frequency demodulation, and sideband analysis. As degradation progresses, the second gear mesh harmonic (2×GMF) and its surrounding modulation sidebands become increasingly dominant, reflecting the development of localized tooth damage.

3. CONDITION INDICATOR (CI)

A total of 18 condition indicators were extracted from multiple vibration analysis domains. Rather than relying on a single CI, the proposed framework combines indicators derived from residual analysis, narrowband filtering, amplitude modulation (AM), frequency modulation (FM), energy operator (EO), and sideband analysis.

3.1. Residual Analysis

Stewart (1977) presented the residual (or differential) signal analysis. It is obtained by removing the gear mesh frequency, its harmonics, the shaft rotational frequency, and its harmonics from TSA signal. When the first-order sidebands around the gear mesh frequencies are additionally suppressed, the resulting signal is referred to as the differential signal. The dominant periodic components associated with shaft rotation and gear meshing are removed. The residual signal primarily retains the impulsive vibration generated by localized gear

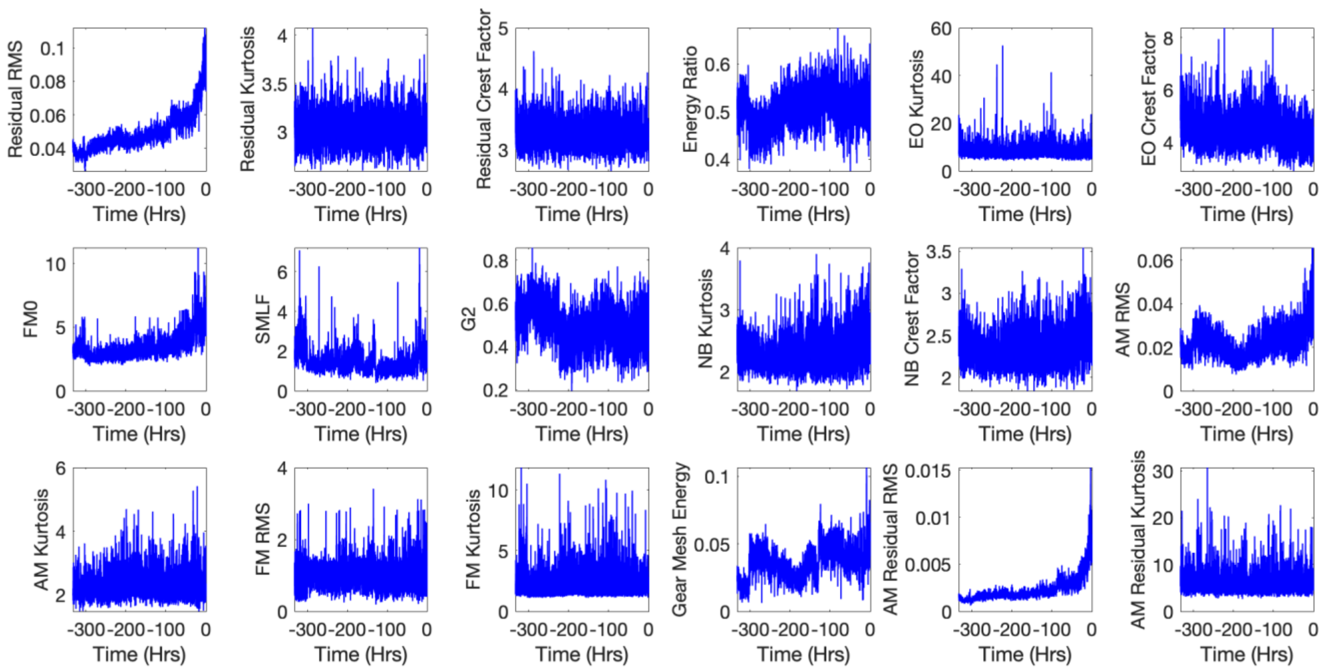


Figure 7 Condition Indicators

defects, such as a cracked or softened tooth. These defect-induced transients are non-sinusoidal in nature and become more prominent after the deterministic components have been eliminated.

In this study, a threshold residual signal is calculated by identifying and removing statistically significant deterministic frequency components from the TSA signal using an adaptive search algorithm. This approach eliminates the need for prior knowledge of the gear tooth count. RMS, kurtosis, crest factor, and energy ratio are calculated from the resulting threshold residual and used as condition indicators. The energy ratio is defined as the residual RMS normalized by the standard deviation of the original TSA signal.

In addition, an amplitude-modulated residual (*amRES*) is obtained by extending the envelope analysis proposed by McFadden (1987) to the threshold residual signal. Specifically, the analytic representation of the threshold residual is obtained using the Hilbert transform, and its magnitude is used as the *amRES* signal. RMS and kurtosis of the resulting *amRES* signal are used as two additional condition indicators.

$$amRES = |Hilbert(Res)| \tag{2}$$

3.2. Narrowband, Amplitude and Frequency Modulation

Sideband modulation in the TSA signal does not always provide a clear indication of gear-tooth defect severity, particularly during the early stages of damage such as fatigue crack initiation. In these cases, bandpass filtering can be applied to isolate the frequency band of interest and improve the signal-to-noise ratio (SNR). McFadden (1986) demonstrated that bandpass filtering around the gear mesh frequency, followed by envelope analysis, enables direct measurement of the amplitude modulation and frequency modulation components present in the original TSA signal. He proposed that as a softened or cracked tooth enters the load zone, the local reduction in mesh stiffness redistributes the transmitted load to adjacent teeth, producing both amplitude variations due to changes in displacement and frequency variations associated with instantaneous rotational speed fluctuations.

In this study, the narrowband (NB) signal is obtained by filtering the TSA signal around the selected gear mesh frequency using a predefined bandwidth. Kurtosis and crest factor of the resulting NB signal are used as condition indicators.

Amplitude modulation (AM) and frequency modulation (FM) features are subsequently obtained from the analytic representation of the narrowband signal using the Hilbert transform. The AM signal is defined as the magnitude of the analytic signal, from which RMS and kurtosis are calculated. The FM signal is obtained from the unwrapped phase of the analytic signal after removing its linear phase trend. RMS and kurtosis of the resulting FM signal are then calculated as additional condition indicators.

3.3. Energy Operator

For an amplitude- and phase-modulated signal, the EO provides an estimate of the product of the instantaneous amplitude and instantaneous frequency. In practice, EO-based CIs are often sensitive to load or torque variations. To reduce this sensitivity, statistical measures such as kurtosis and crest factor are commonly applied, as they are less affected by changes in signal magnitude while remaining sensitive to impulsive fault characteristics. In this study, the EO was implemented as:

$$\Psi_{EO}(TSA_n) = TSA_n^2 - TSA_{n+1} \times TSA_{n-1} \quad (3)$$

3.4. Figure of Merit 0

Figure of Merit 0 (FM0) is a well-known analysis derived from Stewart (1977) and is generally calculated as:

$$FM0 = TSA\ peak\text{-}to\text{-}peak\ /\ (GM1 + GM2 + GM3) \quad (4)$$

where (*GM*) represents the gear mesh harmonic extracted from the FFT of the TSA signal. The peak-to-peak CI is a time-domain indicator that captures impulsive events caused by localized defects, such as tooth impacts or breathing cracks. In contrast, the gear mesh energy is derived from the frequency-domain representation of the TSA and is therefore less sensitive to these localized fault signatures. Consequently, as tooth damage progresses, the *FM0* indicator is expected to increase.

Two additional sideband-based condition indicators, SMLF and G2, are calculated from the spectral amplitudes surrounding the first gear mesh frequency. The sideband level is calculated by summing the amplitudes of the six spectral components on each side of the first gear mesh frequency. SMLF is defined as the resulting sideband level normalized by the amplitude of the first gear mesh harmonic, whereas G2 is defined as the same sideband level normalized by the RMS amplitude of the TSA signal. These normalized indicators quantify the growth of modulation sidebands relative to the dominant gear mesh response and the overall vibration level, respectively.

As shown in Figure 7, CIs are calculated based on the statistics including RMS, kurtosis, peak-to-peak value (max - min), and crest factor and each CI represents the degradation process differently.

Several CIs exhibit a gradual increase as damage accumulates, whereas others remain relatively insensitive during early degradation and become more responsive near failure. Some indicators also show higher variability due to changes in operating conditions and measurement noise. This complementary behavior suggests that no single CI is sufficient for robust health assessment, motivating the probabilistic fusion approach presented in the following section. Therefore, instead of selecting a single condition indicator, all extracted CIs are integrated using a probabilistic Nakagami-based Health Index to improve robustness and provide a single monotonic representation of gearbox health.

4. HEALTH INDEX

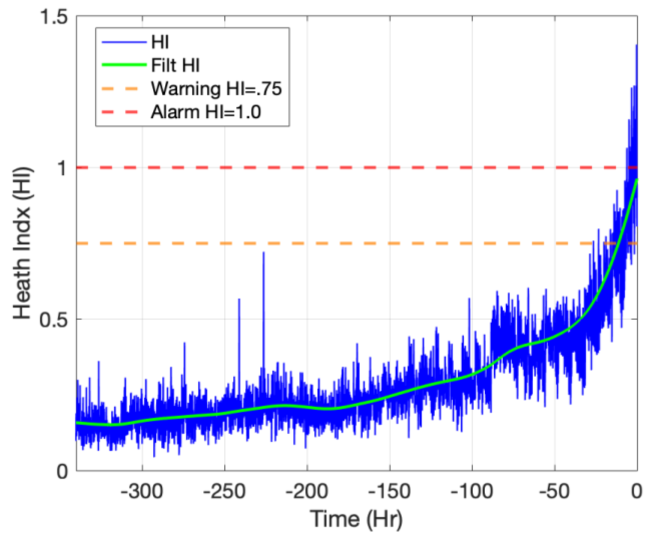


Figure 8 Health Index of the Combination of Multiple Condition Indicators

To calculate a single representation of gearbox health, the extracted condition indicators were chosen to use a Nakagami distribution-based probabilistic model. Unlike individual CIs, which exhibit different sensitivities and noise characteristics, the resulting Health Index provides a unified measure of degradation that evolves more consistently over the equipment lifetime.

The false alarm rate of the proposed Health Index can be controlled by exploiting the statistical relationship between the HI and the distributions of the underlying CIs. This relationship can be derived using the method of moments and is valid when the CIs satisfy the assumptions of being independent and identically distributed. Since many of the extracted CIs approximately follow a Rayleigh distribution, or can be transformed to one, an additional preprocessing step is applied to enforce the assumption.

The process of removing statistical correlation among the CIs is referred to as whitening. Whitening transforms the correlated CI vector into a set of uncorrelated variables with normalized covariance. This decorrelation is used to improve consistency between the realized and designed false alarm rates of the monitoring system.

The whitening transformation is constructed from the Cholesky decomposition of the inverse covariance matrix, which effectively provides the square root of the covariance normalization matrix. Applying this transformation to the measured CI vector yields decorrelated variables with normalized covariance. Whitening does not, in general, guarantee statistical independence; therefore, the independence assumption used in the Nakagami-based HI formulation should be regarded as an approximation for the present application. The whitening transformation is defined as:

$$\boldsymbol{LL^T} = \boldsymbol{\Sigma^{-1}} \tag{5}$$

where Σ is the sample covariance matrix of the CIs used in the HI calculation. Then, HI is

$$HI = scale/critical\sqrt{\boldsymbol{CI^T LL^T CI}} \tag{6}$$

HI defined in Eq (6) follows a Nakagami distribution under the assumption that the condition indicators are independent and identically distributed. A decision threshold is established using the inverse cumulative probability function (ICPF), which requires a user-defined probability of false alarm (PFA). False alarms should occur infrequently in condition monitoring applications, the PFA was set to 10^(-6). The ICPF additionally requires the Nakagami distribution parameters (η) and (ω), where (η) corresponds to the number of condition indicators included in the HI calculation and (ω) is the corresponding scale parameter.

$$\omega = 2n/(2 - \pi/2) \tag{7}$$

In Eq. (6), *scale* denotes an empirical HI normalization factor and should not be confused with the Nakagami scale parameter defined in Eq. (7). In this study, the HI normalization factor was set to 0.5, following the convention adopted for the monitoring system.

As shown in Figure 8, the HI remains relatively stable during the early stages of operation, gradually increases as degradation progresses, and rises rapidly during the final stage of the run-to-failure experiment. The alarm threshold (HI = 1.0) defines the point at which the gearbox is considered to have reached an unacceptable health condition and serves as the transition point for Remaining Useful Life prediction. Once the HI reaches the predefined alarm threshold, the degradation trend is used as the input to the physics-guided prognostic model to estimate the RUL and its associated uncertainty.

5. RUL CALCULATION

Once the Health Index exceeds the predefined alarm threshold, the degradation trend is used to initialize the physics-guided prognostic model. In this study, a fatigue crack growth model is used. The crack growth rate, expressed in terms of the crack half-length (a), is modeled using a linear elastic fracture mechanics model proposed by Bechhoefer & Dube (2020). According to the linear elastic fracture mechanics model, the rate of crack propagation is given by:

$$\frac{da}{dN} = D(\Delta K)^m \tag{8}$$

where

- *da/dN* is the rate of change of crack half-length per cycle,
- *D* is a material constant,
- ΔK is the gross strain term (sigma corrected by a shape factor alpha), and
- *m* is the crack growth exponent

Substituting ΔK into Eq (8) gives:

$$\frac{da}{dN} = D\left(2\sigma(\pi)^{1/2}\alpha\right)^m a^{m/2} \tag{9}$$

Inverting and integrating gives the number of cycles, N:

$$N = \int_{a_0}^{a_t} a^{-m/2} \Big/ D\left(2\sigma(\pi)^{1/2}\alpha\right)^m da \tag{10}$$

Taking crack length, *a* as the current measured value a_0 and estimating the inverse crack growth rate, *dN/da*, causes the constants to cancel, leaving the number of cycles:

$$N = -dN/_{da}\; a_0 - a_f\left(a_0/a_f\right)^{m/2} \Big/ {}^{m}/_{2} - 1 \tag{11}$$

Substituting a crack growth exponent of 2, which is common for steel, gives the RUL cycle estimate:

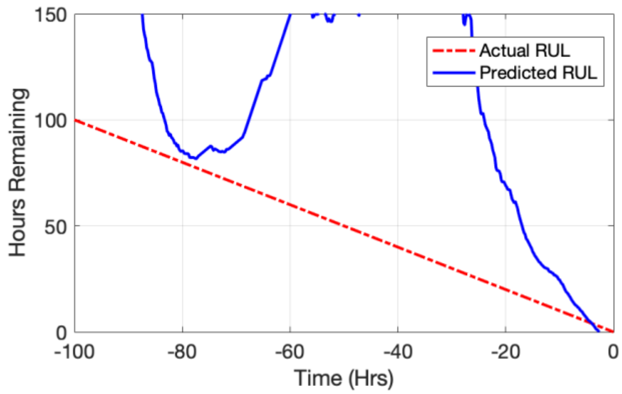
$$N = -dN/_{da} \times a_0 \times ln\left({}^{a_f}/_{a_0}\right) \tag{12}$$

In most applications, crack length is not directly measured. In the present framework, the HI is therefore used as a normalized surrogate for accumulated damage rather than as a direct measurement of physical crack length. The substitution assumes that the evolution of the HI reflects the progression of the dominant degradation mechanism. Accordingly, the resulting RUL formulation should be interpreted as a physics-guided degradation model rather than a direct crack-length prediction model. Accordingly, a_0 is replaced by the current HI and a_f by the normalized maintenance threshold, HI=1, resulting in:

$$RUL = -dt/_{dHI} \times HI \times ln\left({}^{1}/_{HI}\right) \tag{13}$$

Figure 9 shows the evolution of the predicted Remaining Useful Life during the final degradation stage. The predicted RUL converged to the true RUL within approximately 30 hours prior to failure. At the beginning of the prediction window, the estimated RUL is intentionally conservative because only a limited portion of the degradation process has been observed. As additional Health Index measurements become available, the degradation rate is continuously updated, allowing the predicted RUL to converge toward the true remaining life. Near failure, the prediction closely follows the actual RUL, demonstrating that the proposed framework provides increasingly accurate maintenance forecasts as the fault progresses.

After the Health Index reaches the predefined alarm threshold, the measured degradation trend is incorporated into the proposed physics-guided prognostic model. The model continuously updates the estimated degradation rate as new Health Index measurements become available, enabling real-time RUL prediction.



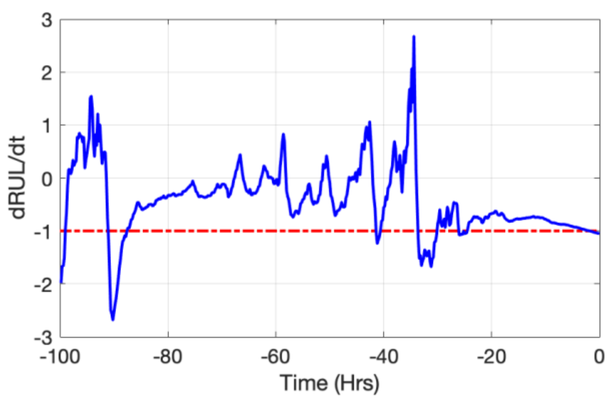


Figure 9 RUL and dRUL/dt

The proposed prognostic model is derived from linear elastic fracture mechanics. The underlying assumption is that damage evolution is governed primarily by crack propagation. However, the dominant degradation mechanism in the early stage of gear scuffing is localized surface wear and material degradation induced by Hertzian contact stress rather than macroscopic crack growth. Some discrepancy between the predicted and actual RUL evolution can be observed during the initial degradation stage. As surface damage accumulates and transitions toward fatigue crack initiation and propagation, the degradation mechanism becomes increasingly consistent with the assumptions of the fracture mechanics model. Accordingly, the predicted RUL converges toward the actual RUL, and the RUL derivative rate approaches the ideal value of $dRUL/dt = -1$ during the later stage of degradation.

Although the model does not explicitly capture the early wear-dominated regime, it provides increasingly accurate prognostic behavior once crack-dominated damage becomes the primary failure mechanism. This explains the improved agreement between the predicted and actual RUL observed in the later stage of the experiments.

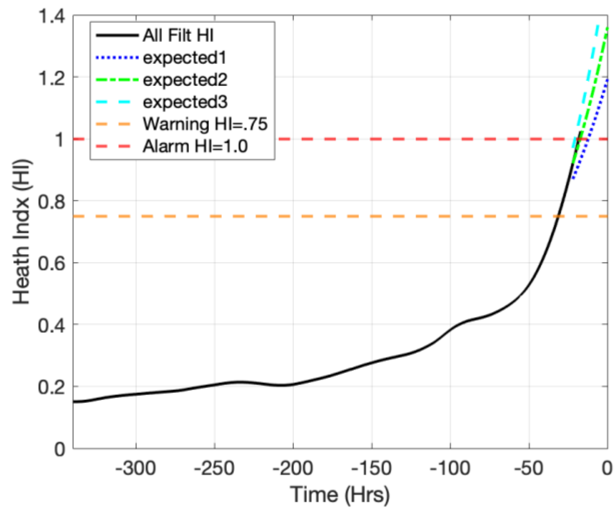


Figure 10 Predicted Remaining Useful Life

Figure 10 shows the prediction capability of the proposed degradation model in the Health Index domain. Once sufficient degradation information has been accumulated, the model extrapolates the future Health Index evolution using the estimated degradation rate. Three representative prediction trajectories are shown to demonstrate the expected degradation behavior under different model assumptions. As the predicted Health Index approaches the predefined alarm threshold (HI = 1.0), the corresponding Remaining Useful Life is calculated from the estimated time-to-threshold. The warning threshold (HI = 0.75) provides an earlier indication for maintenance planning, allowing operators to schedule maintenance before the gearbox reaches its alarm condition. Although comparison with alternative fusion methods including PCA- and autoencoder-based approaches is beyond the scope of this study, Nakagami provides a statistically interpretable HI with controllable false alarm probability.

6. CONCLUSION

This paper presented an end-to-end PHM framework for industrial compressor gearbox monitoring using a real run-to-failure dataset collected from an operating compressor system. The proposed framework integrates vibration acquisition, TSA, multi-domain condition indicator extraction, probabilistic Health Index construction, and physics-guided RUL prediction into a practical workflow for industrial applications.

Multiple condition indicators extracted from complementary vibration analysis techniques were fused using a Nakagami distribution-based model to generate a robust Health Index representing gearbox degradation. The resulting HI was then incorporated into a Paris law-based prognostic model to estimate the Remaining Useful Life without requiring detailed material-specific fatigue parameters. Validation using the industrial dataset demonstrated that the proposed framework successfully tracked gearbox degradation and provided stable, physically interpretable RUL predictions.

Several limitations of the present study should be acknowledged. The proposed framework was validated using a single industrial run-to-failure gearbox dataset, which limits the ability to assess statistical repeatability and generalizability across different gearbox units and failure modes. In addition, the degradation data were collected over a relatively narrow range of operating conditions, and further validation under broader speed and load conditions would be beneficial. The HI is also intended to represent a normalized degradation surrogate rather than a direct measurement of physical crack length. Further studies incorporating additional run-to-failure datasets, broader operating regimes, and directly measured damage progression would help establish the generalizability of the framework and strengthen the physical interpretation of the proposed physics-guided RUL model.

Overall, the proposed approach demonstrates that combining established vibration diagnostics with probabilistic health assessment and physics-guided prognostics provides an effective and interpretable solution for industrial gearbox condition monitoring and supports practical condition-based maintenance decisions. The proposed method is directly deployable to edge-based industrial monitoring systems where bandwidth and storage are limited.

REFERENCES

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A Review on Machinery Diagnostics and Prognostics Implementing Condition-based Maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012

McFadden, P. D. (1987). Examination of a Technique for the Early Detection of Failure in Gears by Signal Processing of the Time Domain Average of the Meshing Vibration. *Mechanical Systems and Signal Processing*, 1(2), 173–183. https://doi.org/10.1016/0888-3270(87)90020-0

Randall, R. B. (2011). Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications. John Wiley & Sons. https://doi.org/10.1002/9780470977668

Antoni, J. (2007). Fast Computation of the Kurtogram for the Detection of Transient Faults. *Mechanical Systems and Signal Processing*, 21(1), 108–124. https://doi.org/10.1016/j.ymssp.2005.12.002

Bechhoefer, E. (2019). A Comprehensive Analysis of the Performance of Gear Fault Detection Algorithms. *Proceedings of the Annual Conference of the PHM Society,* 11(1). doi.org/10.36001/phmconf.2019.v11i1.823

Cho, C., & Bechhoefer, E. (2025). Gearbox Casing Crack Detection Based on Vibration Signals in Frequency and Time Domain. Vertical Flight Society 81st Annual Forum & Technology Display.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep Learning and its Applications to Machine Health Monitoring. *Mechanical Systems and Signal Processing*, 115, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050

Ahmad, H., Cheng, W., Xing, J., Wang, W., Du, S., Li, L., Zhang, R., Chen, X., & Lu, J. (2024). Deep learning-based fault diagnosis of planetary gearbox: A systematic review. Journal of Manufacturing Systems, 77, 730–745. https://doi.org/10.1016/j.jmsy.2024.10.004

Zhou, H., Huang, X., Wen, G., Lei, Z., Dong, S., Zhang, P., & Chen, X. (2022). Construction of health indicators for condition monitoring of rotating machinery: A review of the research. Expert Systems with Applications, 203, 117297. https://doi.org/10.1016/j.eswa.2022.117297

Xiong, J., Fink, O., Zhou, J., & Ma, Y. (2023). Controlled physics-informed data generation for deep learning-based remaining useful life prediction under unseen operation conditions. Mechanical Systems and Signal Processing, 197, 110359. https://doi.org/10.1016/j.ymssp.2023.110359

Bechhoefer, E., & He, D. (2016). Reducing Tachometer Jitter to Improve Gear Fault Detection. Annual Conference of the PHM Society 8 (1), Denver, Colorado

Bechhoefer, E., Kingsley, M., & He, D. (2011). Processing for Improved Time Synchronous Averaging of Gearbox Vibration Data. Proceedings of the Annual Conference of the PHM Society, 3(1).

Stewart, R. M. (1977). Some Useful Data Analysis Techniques for Gearbox Diagnostics. Machine Health Monitoring Group, Institute of Sound and Vibration Research, University of Southampton, Report MHM/R/10/77.

McFadden, P. D. (1986). Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration. *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, 108(2), 165-170.

Bechhoefer, E., & Dube, M. (2020). Contending Remaining Useful Life Algorithms. Annual Conference of the PHM Society, 12(1), 9.

BIOGRAPHIES

Changik Cho received his B.S. and M.S. degrees in Mechanical Engineering from Sungkyunkwan University in South Korea, and earned his Ph.D. in Mechanical Engineering from Pennsylvania State University, where his research focused on active vibration control for electrical vertically take-off and landing (eVTOL) aircraft. His work spans vibration control, rotor dynamics, and battery modeling, with contributions to both analytical methods and experimental validation. He has collaborated with industry partners such as Siemens on Li-ion battery optimization and with an eVTOL company on full-scale tiltrotor aircraft vibration analysis during his Ph.D. Dr. Cho is currently a Systems Engineer at GPMS, where he focuses on the development of Health and Usage Monitoring Systems (HUMS) for rotorcraft. He has published his work in leading venues including AIAA SciTech, the VFS Forum, ASME IDETC, and the AIAA Journal of Guidance, Control and Dynamics.

Eric Bechhoefer received his BS in Biology from the University of Michigan, his MS in Operations Research from the Naval Postgraduate School, and a Ph.D. in General Engineering from Kennedy Western University. He is a former Naval Aviator who has worked extensively on condition-based maintenance, rotor track and balance, vibration analysis of rotating machinery, and fault detection in electronic systems. Dr. Bechhoefer is a Fellow of the Prognostics Health Management Society, a Fellow of the Society for Machinery Fault Prevention Technology, and a senior member of the IEEE Reliability Society. Additionally, Dr. Bechhoefer is also a member of the SAE committee covering Integrated Vehicle Health Management, and a member of the MSG-3, Rotorcraft Maintenance Programs Industry Group.