Fault Detection and Remaining Useful Life Estimation of a Scuffed Gear

Eric Bechhoefer[1], Changik Cho[2], and Lotfi Saidi[3]

[1,2]*GPMS International, Inc, Waterbury, VT, 05753, USA*
eric@gpms-vt.com
changik@gpms-vt.com

[3]*ESSTHS, University of Sousse, Tunisia, Sousse, Tunisia*
lotfi.saidi@esstt.rnu.tn

Abstract

In helicopter drivetrains, vibration-based gear fault detection is critical to safety, reliability, and operational availability. Although there are numerous gear fault detection algorithms have been proposed, there is comparatively little work on gear wear.

This study presents a quantitative comparison of time synchronous average (TSA) based gear fault detection algorithms for identifying scuffing and wear in a turboshaft engine high-speed turbine shaft pinion. A large field dataset from a degraded engine is compared with data from a nominal replacement component. Condition indicator (CI) responses are evaluated using statistical separability to determine each algorithm's ability to distinguish the damaged gear from the nominal baseline.

The comparison includes residual and difference-signal analysis, energy-operator methods and variants, narrowband analysis with bandwidth sensitivity evaluation, amplitude and frequency modulation analysis, and standard gear fault condition indicators. The statistical significance and robustness of each method are quantified.

The best-performing indicators are then used to construct a component health index (HI), which is applied to assess gear condition and demonstrate remaining useful life (RUL) estimation in a maintenance-relevant framework.

1. Introduction

Helicopter drivetrain design requires a tradeoff between performance and safety. Reduced drivetrain weight improves aircraft performance, but lighter designs are more susceptible to high-cycle fatigue, wear, and damage. Gear wear is commonly classified as adhesive, abrasive, corrosive, or fatigue related. Under sustained high-load conditions, progressive wear can lead to cracked or broken teeth.

During helicopter type certification, the manufacturer establishes maintenance requirements, including time between overhaul (TBO) intervals for life-limited components. For example, the Bell 407 M250-C47 engine has a TBO of 3,200 hours.

To meet reliability requirements and reduce the risk of catastrophic failure, magnetic chip detectors are installed in the engine lubrication system to detect gear wear. Normal engine operation can liberate ferrous particles, which are carried by oil and captured by the magnetic chip detector (MCD). On the Bell 407, the MCD is an active detector: once sufficient metallic debris bridges the detector's permanent magnets and closes the electrical circuit, an indication is sent to the cockpit, and a maintenance inspection is required.

Health and Usage Monitoring Systems (HUMS) are sensor-based systems that measure the health and performance of mission-critical aircraft components. HUMS provides actionable information that enables maintainers to make data-informed decisions about component condition (see HeliOffshore, 2023). HUMS vibration processing techniques have improved operational readiness, reduced maintenance cost, and enhanced safety by supporting a proactive maintenance paradigm.

Many gear fault detection algorithms are based on the time synchronous average (TSA). Although the Fourier transform (FT) is a powerful tool for vibration analysis, its utility for detecting gear faults is limited when rotational speed varies. In helicopter gearboxes, the FT assumption of stationarity is violated. In addition, many gear fault features are not sinusoidal; they are impulsive, such as the response from a soft or cracked tooth or from Hertzian contact associated with gear scuffing.

Some of these features appear in the frequency domain as gear mesh sidebands, but the TSA improves signal-to-noise ratio by enforcing stationarity through resampling to a shaft key phasor. Resampling averages the vibration data over each shaft revolution. In effect, the TSA acts as a DC finite impulse response filter, where attenuation of non-synchronous gearbox frequencies depends on the number of shaft revolutions included in the average (Stewart, 1977).

The TSA is a useful tool for orthogonalizing time-domain data. For example, with 30 revolutions, the TSA bandwidth is approximately 4%. For a shaft operating at 100 Hz, one second of data contains 100 revolutions; source frequencies above 101.2 Hz or below 98.8 Hz have more than 50% of their energy attenuated. Signals above 110 Hz or below 90 Hz are attenuated by more than 99.3%. This orthogonalization allows statistics such as peak-to-peak value, RMS, and kurtosis to reflect changes in the time-domain waveform associated with tooth or gear damage.

Vibration-based techniques such as Figure of Merit 0 (FM0) from Stewart (Stewart, 1977) and amplitude modulation (AM) and phase/frequency modulation (FM) techniques from McFadden (1996) provide established analysis tools for detecting fatigue cracks in gears. FM0 is the TSA peak-to-peak value divided by the sum of the TSA gear-mesh energy. The AM operation is the magnitude of the Hilbert transform of the bandpass-filtered TSA, whereas the FM operation is the pseudo-derivative of the Hilbert-transform phase of the bandpass-filtered TSA.

Sharma & Parey (2016), tested Condition Indicators (CI) and used them to quantify the level of vibration generated by the defected gears. A comprehensive comparison of various CIs, i.e., RMS, kurtosis, crest factor, FM0, FM4, M6A, NB4, energy ratio, NA4, energy operator and two new proposed condition indicators (PS-I and PS-II), This comparative analysis shows the responsiveness of indicators towards crack detection, but did not quantify CI performance.

In Koukoura et. al (2019) compare several gear vibration analysis methods on the same population of real wind-turbine gearbox failures, then use the extracted features as inputs to a supervised classifier. Their stated goal was to compare diagnostic methods based on classification accuracy, using historical vibration data that progress from healthy operation toward failure. The focus was on gear tooth crack, not scuffing. Koukoura evaluated: side band analysis, SMFL, cepstrum, TSA statistics, and spectral kurtosis.

However, research on algorithms for gear wear detection remains limited. While comparative studies exist for cracks and pitting, but there is limited quantitative comparison of gear-wear/scuffing algorithms using real field data and a common statistical performance metric. Daman et al. (2026) note that vibration-based techniques are not commonly used for wear monitoring, likely because relatively few methods can diagnose moderate to severe wear.

Zakrajsek et al. (1993) investigated fault detection for pitting fatigue failure, a wear-related mechanism, and introduced NA4 as an improvement over FM4. FM4 is the kurtosis of the residual signal; in this paper, FM4 analysis is referred to as Residual Kurtosis. The residual is the TSA with regular meshing components removed, including shaft frequency and harmonics, primary meshing frequency and harmonics, and first-order sidebands. NA4 is similar to FM4, except that the first-order sidebands remain in the residual. In this paper, that analysis is referred to as Difference Residual analysis.

As such, there has been limited quantitative evidence comparing the effectiveness of FM4, NA4, and other gear wear detection analyses. Because this study uses data from a real-world gear wear, it provides an opportunity to evaluate these methods quantitatively and determine which gear analyses provide the best wear detection performance.

2. STATISTICAL SEPARABILITY OF GEAR ALGORITHMS

In vibration monitoring, fault detection is the process of identifying damage-related features in the presence of noise. For critical systems, operational risk occurs when a system allows a mission to proceed when maintenance should be performed. Conversely, a false alarm occurs when unnecessary maintenance is triggered. Both probability of detection (PD) and probability of false alarm (PFA) depend on signal-to-noise ratio (SNR). A useful linear measure of SNR is the signal magnitude divided by its standard deviation; statistically, this quantity can be used to measure separability.

Separability defines the statistical distance between two populations. For two distinct populations, such as a nominal gear and a damaged gear, an effective detection algorithm should produce a large separation. In this study, separability is quantified using the pooled sample standard deviation and the small-sample test statistic for comparing two population means.

$$T = \left. \bar{Y_1} - \bar{Y_2} \middle/ S\sqrt{1/n_1 + 1/n_2} \right. \quad (1)$$

where n_1 is the number of samples and $\bar{Y}_1$ is the mean value for the nominal gear, n_2 is the number of samples and $\bar{Y}_2$ for the damaged gear, and S is the pooled sample variance:

$$S = \sqrt{\left. (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 \middle/ n_1 + n_2 - 2 \right.} \quad (2)$$

The population means and sample variances are calculated from gear-analysis condition indicators (CIs), where each CI is a statistic derived from a specific gear analysis.

All algorithm comparisons in this study are based on T, the separability statistic. This provides a convenient way to compare one analysis with another: a larger positive separability indicates stronger fault detection performance.

2.1. Dataset

The dataset was collected from a Bell 407 equipped with an M250-C47 engine and HUMS. Data were collected every 3 to 4 minutes during a stable operating regime, producing 18 condition indicators over more than 350 hours. Raw vibration data were also collected once per flight. Each raw-data acquisition was sampled at 93,750 samples per second for one second. These raw data enabled algorithm testing on both the damaged gear and the nominal replacement gear, allowing the separability of a large set of gear fault algorithms to be evaluated quantitatively.

The gear under analysis is the N2 Pinion (35 tooth), which is driven by the engine power turbine. The turbine RPM is approximately 32,200. The engine was replaced due to an alert of the engine chip detector and confirmed with HUMS. The damage due to wear on the gear is seen in Figure 1.

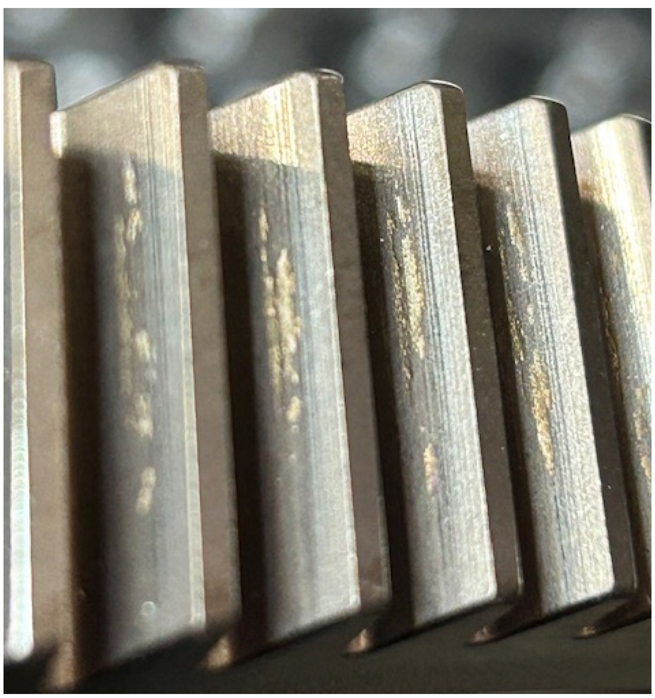

Figure 1 Wear found on the N2 Pinion

2.2. Gear Wear Analysis Tested

In the review papers, and previous work on a known cracked-tooth fault, the best performing gear fault analysis were based on the TSA. As such, all analysis were statics of the TSA, or statics based on operations performed on the TSA. For this study, the statistics evaluated were RMS, kurtosis, peak-to-peak value (maximum minus minimum), and crest factor, defined as:

$$\max(\text{TSA}) - min(\text{TSA}) \Big/ \sqrt{1/n \sum_{i=1}^{n} TSA_i^2} \tag{3}$$

2.2.1. Residual Based Analysis

The residual, or difference signal, follows Stewart (1977). The residual signal is defined as the synchronous averaged signal with the gear mesh frequency and harmonics, driveshaft frequency, and second driveshaft harmonic removed. Two additional residual variants were also tested:

- First-order sidebands about the gear mesh frequencies are filtered out, producing the difference signal.
- A threshold-based search algorithm removes statistically significant frequencies. This approach does not require a gear tooth count and can be applied to planetary gearboxes where the gear mesh frequency may differ from the gear tooth count.

Two additional residual-based analyses were also tested. These were based on the squared residual envelope, following Wang (2001). This type of analysis can detect cyclostationary behavior that may be associated with gear wear.

$$amRES = |Hilbert(RES^2)| \tag{4}$$

The $amRESD$ indicator is a low-pass-filtered version of $amRES$. The "D" denotes decimation; in this study, the bandwidth was 0.5, corresponding to decimation by 2.

2.2.2. Narrowband, AM and FM Analysis

Sideband modulation of the TSA does not always make the severity of a gear-tooth defect clear. For early faults, such as fatigue cracks, it may be more effective to bandwidth-limit the signal to improve SNR. McFadden (1986) showed that bandpass filtering to remove gear mesh harmonics, followed by enveloping, can directly measure amplitude and phase/frequency modulation in the original TSA. McFadden hypothesized that, as a soft or cracked tooth enters the load zone, the reduced stiffness transfers load to surrounding teeth, affecting both displacement (AM modulation) and rotational speed (FM modulation).

2.2.3. Teager's Energy Operator

For an amplitude- and phase-modulated signal, the energy operator (EO) can estimate the product of instantaneous amplitude and instantaneous frequency. In practice, EO-based CIs can be sensitive to torque, which is why kurtosis and crest-factor are often used; these statistics tend to normalize changes in magnitude caused by torque variation.

In Bechhoefer et al (2019) three variants of the EO were tested and found to produce similar results. For this study, the EO was implemented as:

$$\Psi_{EO}(TSA_n) = TSA_n^2 - TSA_{n+1} \times TSA_{n-1} \tag{5}$$

2.2.4. Fourier-Based Analyses

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

$$FM0 = TSA\ peak\text{-}to\text{-}peak \,/\, (GM1 + GM2 + GM3) \qquad (6)$$

where GM is the gear mesh harmonic taken from the FFT of the TSA. The peak-to-peak feature is a time-domain phenomenon associated with tooth impact or a breathing crack, whereas gear mesh energy is calculated from the FFT and is less sensitive to soft-tooth features. Therefore, as tooth damage increases, FM0 is expected to increase.

The energy ratio is the residual RMS divided by the TSA RMS. The concept is that, as a gear fault progresses, the residual RMS approaches the TSA RMS because the TSA becomes increasingly dominated by fault-related gear mesh signatures. If the residual signal is r_i and the value of the TSA at index i, is tsa_i:

$$er = \sqrt{\frac{\sum_{i=1}^{n}(r_i-\bar{r})^2}{n}} \Bigg/ \sqrt{\frac{\sum_{i=1}^{n}(tsa_i-\overline{tsa})^2}{n}} \qquad (7)$$

The sideband level factor is defined as the sum of the first-order sideband amplitudes about the gear mesh divided by the TSA RMS:

$$SLF = TSA_{gm-1} + TSA_{gm+1} \Bigg/ \sqrt{\frac{\sum_{i=1}^{n}(tsa_i-\overline{tsa})^2}{n}} \qquad (8)$$

SMLF (sideband modulation level factor is:

$$SMLF = TSA_{gm-1} + TSA_{gm+1} \Big/ TSA_{gm} \qquad (10)$$

The G2 analysis is the ratio of the second gear mesh harmonic energy to the first gear mesh harmonic energy.

2.3. Gear Wear CI Results and Discussion

All residual-based analyses performed well when RMS and peak-to-peak statistics were used. Surprisingly, and contrary to the observations of Zakrajsek et al. (1992) and Lewicki et al., (2009), kurtosis-based residual analyses did not perform well. In addition, methods developed primarily for cracked or soft-tooth detection, including Stewart's FM0 and McFadden's narrowband, AM, and FM analyses, generally did not detect this wear fault effectively. Crest factor also performed poorly as a statistic (Table 1).

Several factors may explain the poor performance of kurtosis in this dataset. First, many prior studies used test-stand data. The loads required to propagate a fault on a test stand may be high relative to the design load of the M250 gearbox, producing low-cycle fatigue rather than high-cycle fatigue degradation. Second, test-stand architecture, often based on back-to-back rigs, differs from the M250 configuration, in which the N2 input pinion has 35 teeth and drives a 107-tooth torque gear. Third, under constant-load operation, kurtosis may perform better; however, in helicopter operation, load commonly varies from approximately 60% to 90%.

Table 1 Measure of CI Separability by Analysis

Analysis	T	Analysis	T
Residual RMS	3.54	**FM RMS**	0.871
Sideband Index	3.527	**FM P2P**	0.684
Residual P2P	3.366	**FM0**	0.449
SLF	2.956	**Diff CF**	0.378
amDif P2P	2.641	**Diff Kurt**	0.251
Diff P2P	2.532	**NB Kurt**	0.207
AM P2P	2.525	**amDif CF**	0.097
amDif RMS	2.516	**amDif Kurt**	0.045
TSA P2P	2.475	**EO Kurt**	0.005
EO RMS	2.382	**AM CF**	-0.082
Diff RMS	2.372	**G2 Analysis**	-0.109
Thr Res RMS	2.365	**damRes CF**	-0.189
AM RMS	2.16	**damDif CF**	-0.189
EO P2P	2.049	**amRes CF**	-0.236
NB P2P	1.923	**energy ratio**	-0.248
Thr Res P2P	1.858	**AM Kurt**	-0.412
TSA RMS	1.826	**Thr Res CF**	-0.44
Sideband LF 1 h	1.797	**EO CF**	-0.445
amRes RMS	1.723	**NB CF**	-0.445
NB RMS	1.687	**Thr Res Kurt**	-0.507
damRes RMS	1.599	**amRes Kurt**	-0.513
damDif RMS	1.599	**damRes Kurt**	-0.6
SMLF	1.597	**damDif Kurt**	-0.6
damRes P2P	1.5	**Residual CF**	-0.743
damDif P2P	1.5	**Residual Kurt**	-0.879
amRes P2P	1.357	**FM Kurt**	-1.419
Gear Mesh	1.282	**FM CF**	-1.695

Regardless of these differences, a HUMS should be sensitive to both gear wear and gear fault conditions. Based on this study and prior cracked-tooth results, the following CIs were selected for monitoring the gear:

- Threshold Residual RMS. This CI was selected instead of FM4/Residual RMS because it performs better for planetary gearboxes and has shown good performance for both gear wear and gear damage detection.
- Energy Operator Kurtosis. This CI has shown good performance for cracked-tooth detection.
- Sideband Level Factor. This CI has shown good performance for both wear and fault detection.
- Amplitude Modulation (AM) RMS. This CI has shown good performance for both wear and fault detection.
- Frequency Modulation (FM) Kurtosis. Although its wear detection capability is poor in this dataset, it has shown good performance for gear fault detection.

3. DECISION TO PERFORM MAINTENANCE

The purpose of the health index (HI) is to combine multiple condition indicators (CIs) into a single statistically interpretable measure of component condition. Individual CIs may respond to different characteristics of a fault and may exhibit substantial scatter due to operating condition, load, measurement noise, and other sources of variability. The HI provides a means of combining this information while controlling the probability of false alarm and providing a common scale for maintenance decision making.

The underlying problem is most appropriately formulated as a hypothesis test. Let

$$H_0 = \text{component is nominal}$$

and

$$H_1 = \text{component is not nominal.}$$

This formulation is particularly appropriate for helicopter drivetrain monitoring because large populations of nominal data are generally available, whereas statistically representative populations of failed components are rare.

Antoni et al. (2024) approached the same general problem by defining an optimal health indicator in the Neyman-Pearson sense: the probability of detection is maximized for a specified probability of false alarm. The ideal detector is based on the likelihood ratio

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)}. \tag{9}$$

When the parameters describing the nominal and faulty distributions are unknown, Antoni et al. use a generalized likelihood ratio test (GLRT),

$$I(\mathbf{x}) = \left\langle \ln p_1\left(x_n; \widehat{\boldsymbol{\theta}}_1\right) - \ln p_0\left(x_n; \widehat{\boldsymbol{\theta}}_0\right)\right\rangle, \tag{10}$$

where the unknown model parameters are replaced by their maximum-likelihood estimates.

For early fault detection, the faulty probability density function is modeled as a small perturbation of the nominal distribution,

$$p_1(x; \theta; \nu) = p_0(x; \theta)\, \phi(x; \theta; \nu), \tag{11}$$

where $\nu = 0$ corresponds to the nominal condition and increasing ν describes a departure from that condition.

For sufficiently small departures from nominal, the resulting optimal health indicator is approximately proportional to the square of the estimated perturbation,

$$I(\mathbf{x}) \approx \frac{\mu}{2} \hat{\nu}^{\,2}. \tag{12}$$

Under H_0, the OHI asymptotically follows

$$I(\mathbf{x}) \sim \frac{\chi_1^2}{2L}, \tag{13}$$

which permits a detection threshold to be selected directly for a prescribed probability of false alarm.

An especially useful form of Antoni's result expresses the OHI in terms of an existing scalar subindicator $\hat{s}$:

$$I(\mathbf{x}) \approx \frac{1}{2L}\left[\frac{\hat{s}(\mathbf{x}) - E\{\hat{s}|H_0\}}{\sqrt{\mathrm{Var}\{\hat{s}|H_0\}}}\right]^2. \tag{14}$$

Thus, for a scalar CI, the OHI is fundamentally a normalized statistical distance from the nominal population.

3.1. Multivariate Health Index

The health-index formulation described by Bechhoefer and Dube (2020) extends this statistical-distance concept to multiple condition indicators. Let the CI vector be

$$\boldsymbol{CI} = [CI_1 \quad CI_2 \quad \cdots \quad CI_\eta]^T, \tag{15}$$

where $\eta$is the number of CIs included in the HI.

Because condition indicators derived from the same vibration signal are generally correlated, simply combining their magnitudes would give excessive weight to redundant information. The covariance structure of the nominal CI population is therefore estimated as

$$\Sigma = \mathrm{Cov}(\mathbf{CI}). \tag{16}$$

A whitening transformation is constructed from the Cholesky decomposition of the inverse covariance matrix:

$$LL^T = \Sigma^{-1}. \tag{17}$$

The corresponding multivariate distance is

$$D = \sqrt{\mathbf{CI}^T LL^T \mathbf{CI}} = \sqrt{\mathbf{CI}^T \Sigma^{-1} \mathbf{CI}}. \tag{18}$$

This quantity is the Mahalanobis magnitude of the CI vector relative to the nominal population. It accounts for both the natural variance of each CI and the covariance between CIs.

The component health index is then defined as

$$HI = \frac{\text{scale}}{\text{critical}} \sqrt{\mathbf{CI}^T LL^T \mathbf{CI}}. \tag{19}$$

The value critical is determined from the statistical distribution of the nominal HI. For the CI transformations used in this implementation, the constituent indicators are treated as approximately Rayleigh distributed and the magnitude of the combined CI vector is modeled by a Nakagami distribution. The critical value is therefore obtained from the inverse cumulative probability function for a specified probability of false alarm,

$$\text{critical} = F_{Nakagami}^{-1}(1; -; P_{FA}\eta\omega), \tag{20}$$

where

$$\omega = \frac{2\eta}{2 - \pi/2}. \tag{21}$$

A probability of false alarm of 10e-6 while the scale factor converts the statistical detection threshold into an engineering health scale. In the present implementation, scales = 0.5.

Consequently, $HI = 0.5$ represents the point at which the nominal hypothesis is rejected at the designed false-alarm probability. The higher value $HI = 1$is separately defined as the maintenance-action threshold.

This distinction is important. The statistical threshold determines when there is sufficient evidence that the component is no longer nominal; it does not necessarily imply that immediate maintenance is required.

For this application, the operating interpretation is therefore:

- $HI < 0.5$: component behavior is consistent with the nominal population.
- $HI \geq 0.5$: the nominal hypothesis is rejected at the prescribed P_{FA}.
- $HI \geq 0.75$: Warning condition; damage is considered highly probable and maintenance planning should begin.
- $HI \geq 1$: Alarm condition; maintenance should be initiated.

The HI plot (Figure 2) shows substantial scatter in the unfiltered HI, attributable in part to variations in operating condition. The filtered HI is therefore used for maintenance assessment. The time at which the filtered HI reaches 1 is defined as zero RUL. In the example shown in Figure 2, magnetic-chip detection occurred approximately 40 (HI value 1.4) operating hours after the filtered HI reached 1.

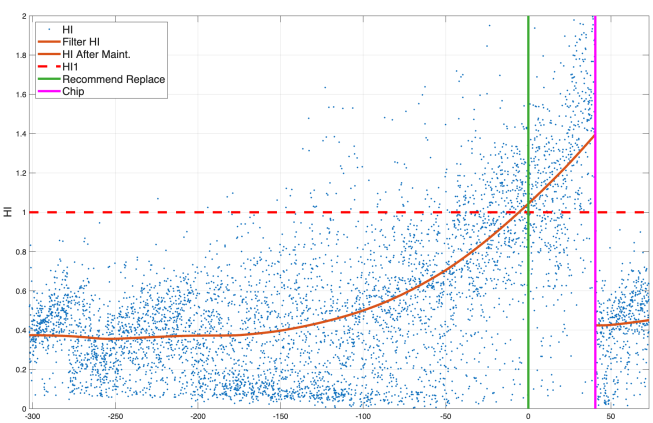

Figure 2 N2 Pinion HI. Chips Were Detected Approximately 40 Hours After Filtered HI = 1.

Importantly, $HI = 1$ does not represent predicted catastrophic failure. It represents a maintenance threshold selected to permit replacement or repair while restoring the drivetrain to its intended reliability level before the damage significantly affects operational readiness.

3.2. Relationship to the Optimal Health Indicator

The Antoni (2024) and Bechhoefer (2020) formulations address the same fundamental detection problem but begin from different practical assumptions.

For a scalar condition indicator, Antoni's small-perturbation OHI can be written in the form

$$I_A \propto \left(\frac{s-\mu_0}{\sigma_0}\right)^2, \tag{22}$$

where $\mu_0$and $\sigma_0$characterize the nominal population.

For multiple statistically independent CIs, the analogous standardized distance is

$$D^2 = \sum_{i=1}^{\eta} \left(\frac{CI_i-\mu_i}{\sigma_i}\right)^2. \tag{23}$$

When the CIs are correlated, the natural multivariate generalization is

$$D^2 = (\mathbf{CI} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{CI} - \boldsymbol{\mu}). \tag{24}$$

Equation (24) has essentially the same statistical-distance structure as the Bechhoefer HI.

The principal difference lies in how the health statistic is developed. Antoni begins with probability-density models for the nominal and faulty states and derives the statistic that is optimal for detecting the specified perturbation. Consequently, the optimal indicator depends on the statistical characteristics of both $H_0$and H_1.

The Bechhoefer formulation instead begins with existing physics-based condition indicators that have already been demonstrated to respond to relevant gear-fault mechanisms. These CIs are normalized using the nominal population and then combined using their covariance structure. The required statistical model is therefore primarily

$$p(\mathbf{CI} \mid H_0), \tag{25}$$

and the detection question becomes

$$P(HI \geq HI_{obs} \mid H_0) \leq P_{FA}. \tag{26}$$

This distinction is particularly important for helicopter HUMS. Drivetrain components are highly reliable, and large nominal populations are available, while statistically representative populations of each failure mode are generally not. A supervised classifier requiring a sufficiently large population of both healthy and failed components is therefore difficult to develop and validate.

In the Bechhoefer approach, the physics of fault detection is incorporated through the selection of the individual CIs, while the HI provides the statistical decision layer:

"Physics-based CIs→nominal statistical normalization →fusion →threshold"

The resulting HI can therefore be viewed as a multivariate extension of the standardized-departure concept underlying Antoni's OHI, which was shown to be optimal.

3.3. Calculating RUL

The health index provides a measure of current component condition. Maintenance planning additionally requires an estimate of the time remaining before the HI reaches the maintenance threshold.

For the M250 drivetrain considered here, a simple linear-elastic crack-propagation model is used as the basis for the RUL formulation. The rate of change of crack half-length $a$with loading cycles $N$is represented by

$$\frac{da}{dN} = D(\Delta K)^m, \tag{27}$$

where D is a material-dependent constant, m is the crack-growth exponent, and ΔK represents the cyclic stress-intensity term.

For the assumed geometry,

$$\Delta K = \alpha\, \Delta\sigma\sqrt{\pi a}, \tag{28}$$

where $\alpha$is a geometry factor and $\Delta\sigma$is the cyclic stress range.

Substitution into Eq. (27) gives

$$\frac{da}{dN} = D\left(\alpha\Delta\sigma\sqrt{\pi a}\right)^m. \tag{29}$$

Collecting the terms independent of crack length into the constant $C = D(\alpha\Delta\sigma)^m \pi^{m/2}$, gives

$$\frac{da}{dN} = Ca^{m/2}. \tag{30}$$

For the present implementation, the crack-growth exponent is assumed to be $m = 2$. Equation (30) therefore becomes

$$\frac{da}{dN} = Ca, \tag{31}$$

and its inverse is

$$\frac{dN}{da} = \frac{1}{Ca}. \tag{32}$$

The remaining number of cycles from the current crack half-length $a_0$to a maintenance limit $a_f$is

$$N_{RUL} = \int_{a_0}^{a_f} \frac{1}{Ca}\, da, \tag{33}$$

which gives

$$N_{RUL} = \frac{1}{C}\ln\left(\frac{a_f}{a_0}\right). \tag{34}$$

The unknown constant C can be eliminated using the presently observed damage-growth rate. At $a = a_0$,

$$\left.\frac{dN}{da}\right|_{a_0} = \frac{1}{Ca_0}, \tag{35}$$

and therefore

$$\frac{1}{C} = a_0 \left.\frac{dN}{da}\right|_{a_0}. \tag{36}$$

Substitution into Eq. (34) gives

$$N_{RUL} = a_0 \left.\frac{dN}{da}\right|_{a_0} \ln\left(\frac{a_f}{a_0}\right). \tag{37}$$

If the slope is represented using the decreasing remaining-life convention, the equivalent positive RUL expression is

$$N_{RUL} = -\frac{dN}{da}\, a_0 \ln\left(\frac{a_f}{a_0}\right). \tag{38}$$

As crack/scuff length cannot be directly observed during normal aircraft operation, the filtered HI is used as a surrogate damage-state variable. The maintenance limit normalized to $HI_f = 1$. Making the substitutions $a_0 \to HI$ and $a_f \to 1$ gives the time-domain RUL estimate

$$RUL = \frac{HI}{\dot{HI}}\ln\left(\frac{1}{HI}\right) \tag{39}$$

Where $\dot{HI} = \frac{dHI}{dt} > 0$.

This formulation eliminates the unknown material, stress, and proportionality constants by using the locally observed HI growth rate.

The filtered HI and $\dot{HI}$are estimated using a state observer. Because physical damage is assumed not to reverse, $\dot{HI}$is constrained to be non-negative. A lower bound is also imposed for numerical stability. In the present example, the minimum growth rate is taken as the inverse of a 20,000-hour mean time between failure, representative of the design life of the gearbox (Figure 3)

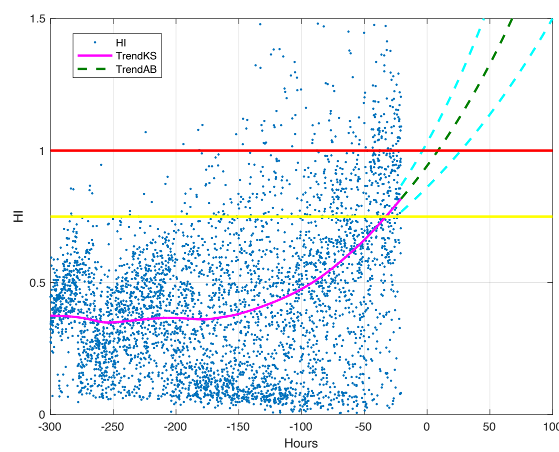

Figure 3 HI Trend with Predicted RUL

The use of a linear-elastic propagation model is an approximation. Pitting is a cumulative fatigue process and is reasonably consistent with a progressive damage model. Scuffing, however, is strongly influenced by lubrication condition and contact limits; consequently, the linear-elastic model should be interpreted as a local propagation model rather than a complete physical description of scuffing.

A useful measure of RUL-model consistency is $\frac{d(RUL)}{dt}$.

As, if one hour of operation consumes one hour of remaining life, a correctly calibrated RUL model should approach

$$\frac{d(RUL)}{dt} = -1 \tag{40}$$

In the presented data, the estimated $\dot{HI}$ begins to depart from its minimum value approximately 180 operating hours before the maintenance threshold, indicating the onset of measurable propagation. The calculated $d(RUL)/dt$ subsequently approaches -1approximately 40 hours before $HI = 1$. This suggests that the assumed propagation model becomes increasingly representative as the gear approaches the maintenance limit.

The observed chip indication occurred approximately 40 hours after the filtered HI reached 1. Thus, the HI trend provided approximately 220 hours of advance indication from the beginning of measurable propagation and approximately 40 hours of margin between the selected maintenance threshold and the subsequent chip-detector indication (Figure 4, Figure 5).

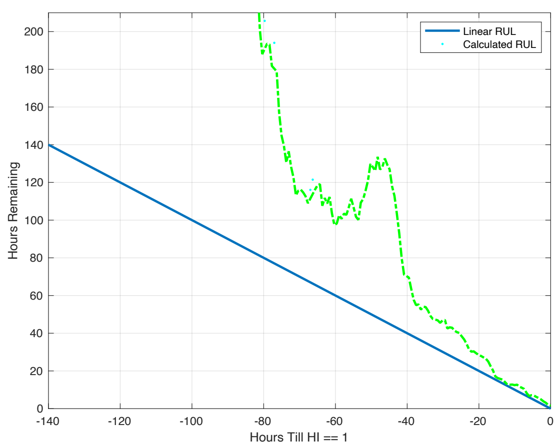

Figure 4 Linear vs. Calculated RUL

Figure 5 is a plot of *dRUL/dt* vs RUL. Two things are evident. At approximately -180 hours, the *dHI/dt* becomes something greater that 1/MTBF, and the RUL decreases shapely, as indicated by the large negative *dRUL/dt* value. This is a good indicator that there is propagating wear and the start of the trend (see Figure 3). The second observation is in the zoomed plot, where the *dRUL* /dt starts to approach -1, starting at -40 hours. This suggest that wear is approaching the linear elastic model towards the end of its RUL. Again, note that the time of failure was 40 hours after the HI of 1, giving a trend detection of 220 hours (four and a half months of operational time).

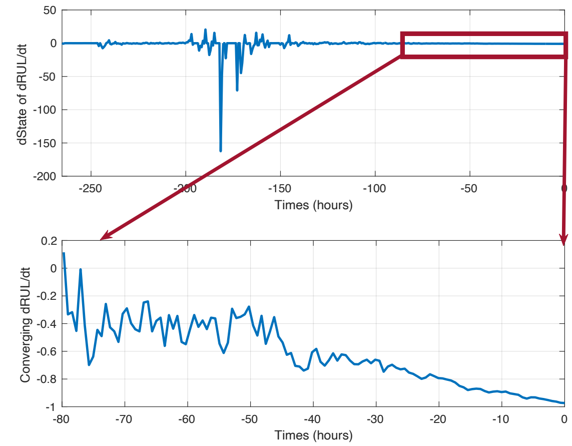

Figure 5 Derivative of the RUL, *dRUL/dt* vs RUL

4. CONCLUSION

The goal of the study was, in an end-to-end manner determine when maintenance should be performed on a helicopter drivetrain. That time is when there is a high probability of damage, but prior to mechanical failure that would affect operational readiness. The end-to-end approach requires:

- Fault feature extraction (e.g. condition indicators, CIs),
- When maintenance (e.g. threshold) should be performed, and,
- An estimate (remaining useful life) to allow scheduling of the maintenance event.

In this paper, we quantify the performance of various TSA based gear wear algorithms from data on the Rolls Royce M250C47, input pinion gear of a Bell407 aircraft. This evaluation was done by evaluating the statistical separability between data from when the gear was damage, and from nominal data after the repair. It was observed that the algorithms that detected wear tended to have sensitivity to cyclostationarity. A future study well evaluation algorithm specific to cyclostationarity, such as the envelop analysis which is typically used for gear fault detection.

As both gear wear (pitting/scuffing) and damage (cracked/soft tooth) need to be detected, six condition indicators were selected and used to from a health index (HI). It was shown that using a hypothesis testing approach, one can control the probability of false alarm, and be assured that the when the HI is greater than 1, the gear is no longer nominal.

The remaining useful life (RUL) of the gear is then the time from the current HI to an HI of 1. For this study, a linear elastic model was used. It was shown that the fit of an RUL model can be quantified by evaluating its derivative. A good model should have an *dRUL* /dt of -1. Using this, it was observed that the damage propagated linear elastic model fit was good for 40 hours prior to recommending maintenance, which would give the operator almost of month to schedule maintenance.

REFERENCES

HeliOffshore. Health and Usage Monitoring Systems: HeliOffshore Recommended Practice Guidance. OGP Report 690, June 2023.

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.

Daman, A., Peng, Z., Chin, Z. Y., & Smith, W. A. A multi-technique approach to gear wear monitoring and fault diagnostics. Tribology International, 214, 111187.

Zakrajsek, J. J., Townsend, D. P., & Decker, H. J. (1993). An Analysis of Gear Fault Detection Method Applied to Pitting Fatigue Failure Data. 47th Meeting of the Society for Machinery Failure Prevention Technology.

Lewicki, D., Dempsey, P., Heath, G., & Shanthakumaran, P. (2009). Gear fault detection effectiveness as applied to tooth surface pitting fatigue damage. American Helicopter Society 65th Annual Forum, Grapevine, Texas, May 27-29, 2009.

Sharma & Parey (2016), “Gear crack detection using modified TSA and proposed fault indicators for fluctuating speed conditions,” Measurement.

Koukoura, S., Carroll, J., McDonald, A., & Weiss, S. (2019). "Comparison of wind turbine gearbox vibration analysis algorithms based on feature extraction and classification," *IET Renewable Power Generation*, 13(11), 1857–1864

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

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.

Antoni, J., Kestel, K., Peeters, C., Leclère, Q., Girardin, F., Ooijevaar, T., & Helsen, J. (2024). On the design of optimal health indicators for early fault detection and their statistical thresholds. *Mechanical Systems and Signal Processing*, 218, 111518.

Bechhoefer, E., & Dube, M. (2020). Contending remaining useful life algorithms. Annual Conference of the PHM Society, 12(1), 9.

BIOGRAPHIES

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.

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, PHM Conferences, and the AIAA Journal of Guidance, Control and Dynamics.

APPENDIX

Example Time Synchronous Average

```
Function [tsadata, navgs,rpm] = tsaLinearInterp(
data, zct, sr, ratio, ppr)
%[tsadata,
navgs,rpm]=tsaLinearInterp(data,zct,sr,ratio,ppr,n
avgs)
%Inputs:
%   data:       time domain data in g's
%   zct:        zero cross time
%   sr:         sample rate
%   ratio:      gear ratio/pulse per revolution on
the tach
%   ppr:        pulse per rev
%Output:
%   tsadata:    time synchronous average data
%   navgs:      the number of averages in the TSA
%   rpm:        mean shaft rpm

%data = data - mean(data);

ndata = length(data);
dt = ndata/sr;  %sample length
rev = 0;
```

```
i = 1+ppr;

while zct(i) < dt && i < length(zct)-1
    rev = rev + 1;
    i = i + ppr;
end

% Define the number of averages to perform
navgs = floor(rev * ratio);
trev = zct(navgs*ppr) - zct(1);
rpm =  navgs/trev*60*ratio;

% Determine radix 2 number where # of points in
resampled TSA
% is at sample rate just greater than fsample
N=(2^(ceil(log2(60/rpm*sr))));

% now calculate times for each rev (1/ratio teeth
pass by)
% resample vibe data using zero crossing times to
interpolate the vibe
yy = zeros(1,N); %data to accumulate the resampled
signal once per rev
ya = yy; %ya is the resample signal once per rev
iN = 1/N; %resample N points per rev
ir = 1/(ratio/ppr);  %inverse ratio - how much to
advance zct
tidx = 1;                %start of zct index

while (floor(zct(tidx)*sr) == 0)
     tidx = tidx + 1;
end

zct1 = zct(tidx);%start zct time;

for k = 1:navgs
    tidx = tidx + ir;         %get the zct for the
shaft

    stidx  =  floor(tidx)-1;     %start  idx  for
interpolation
    dx = tidx - stidx;
    yo = zct(stidx);
    dy = zct(stidx+1)-yo;
    zcti = yo + dx*dy;      %interpolated ZCT
    dtrev = zcti - zct1;    %time of 1 rev
    dtic = dtrev*iN;      %time between each sample
    zct1c = zct1;
    for j = 1:N
        cidx = floor(zct1c*sr);
        yo = data(cidx); y1 = data(cidx+1);
        x1 = zct1c*sr;
        xo = floor(x1);
        dx = x1-xo;
        dy = y1-yo;
        yaj = yo + dx*dy;     %simple linear interp
        ya(j) = yaj;
        zct1c  =  zct1  +  j*dtic;  %increment  to  the
next sample
    end

    zct1 = zcti;

    yy = yy + ya;     %accumulate the tsa per reve
end

tsadata = yy/navgs;           % compute the average
```

Example Residual Analysis

```
function [xres] = residualSignal(x, geartooth)
%[xres] = residualSignal(x, geartooth)
%Inputs:
%  x        :input TSA signal
%geartooth  :array with number of teeth on a gear
%from Vercer
x = x(:)';

n = length(x);
n2 = n/2;
nHarmonics = 3;
X = rfft(x);                        %real fft - no
conjugate
X(1) = 0;                           % DC is removed
X(2) = 0;                           % SO1 is removed
X(3) = 0;                           % SO2 is removed
nGears = length(geartooth);
for j = 1:nGears
    crtGear = geartooth(j);
    for i = 1:nHarmonics
        indx = 1+crtGear*i;
        if indx < n2                %projection
against running over the array
            X(indx) = 0;            %gear tooth meash
are removed
        end
    end
end

xres = irfft(X);          % residual signal from
the inverses real fft

function [nb,am,fm] = narrowband(x, gt, BW)
%[nb,am,fm] = narrowband(x, gt, BW)
% x is the TSA
% gt is the number of gear teeth and
% BW is bandwidth, usually 25% of gt.
%Output:
%    nb: narrow band signal
%    am: amplitude modulated signal
%    fm: phase modulated signal

X = rfft(x);

lw = gt-BW;
%calculate the band pass indexes
hi = gt+BW + 2;

X(1:lw) = 0;
%idealized filter
X(hi:end) = 0;

nb = irfft(X);

n = length(nb);
n2 = n/2;
X = fft(x);
%take the Hilbert Transform
X(1:n2) = X(1:n2) * 2;
X(n2:end) = 0;
h = ifft(X);
%Analytic Signal

%   Amplitude Modulation signal - am
am = abs(h);

%   Phase Modulation signal - fm
arg = unwrap(angle(h));
%take the argument
fm = arg - (arg(end)-arg(1))*linspace(0,1,n);
%take the derivate

function eo = energyOperator(x)
```

```
%eo = energyOperator(x);
%Inputs:
%    x:  the TSA
%Outputs:
%    eo: the energy operator

n = length(x);
data = x(:) - ones(n,1)*mean(x);

%Generate the x(n-1) and x(n+1) sequences
dataminus1=[data(end); data(1:end-1,:)];
dataplus1=[data(2:end); data(1)];
%Calculate the E
eo=data.*data-dataminus1.*dataplus1;
```