Improving the Performance of Bearing Fault Detection on a Turbo Shaft Engine

Eric Bechhoefer[1], and Joshua Kethan[2]

[1,2]*GPMG International Inc*

eric@gpms-vt.com
josh@gpms-vt.com

ABSTRACT

Health and usage monitoring systems (aka "HUMS") have typically not been associated with CFR Title 14 type 27 normal category rotorcraft (weighing less than 7000 pounds, with a seating capacity of 9 or less) due in part to the cost of such systems relative to the asset value. This paper describes performance improvements to HUMS bearing diagnostics methodologies to enhance functionality and improve the business case for HUMS. From a system perspective, this implementation allowed the detection of a real-world turboshaft bearing fault prior to a catastrophic engine failure. The system requirements used vibration data that was resampled to control for changes in shaft rate to reduce smearing of the spectrum. A spectral estimation algorithm is then performed to improve the measured energy associated with the bearing defect. Bearing energy was calculated using the envelope analysis, where a novel approach is taken to correctly select the best window. Finally, a process for thresholding and alerting was implemented, the results being that the aircraft was grounded at the appropriate time and a repair was effectuated prior to engine failure.

1. INTRODUCTION

HUMS (health and usage monitoring systems) have typically been implemented on larger type 29 category transport helicopters. While the benefits of HUMS for enhanced safety, improved availability, and reduction in operating costs is well documented, there have been relatively few implementations of HUMS in the type 27 market. This lack of market penetration can be explained by the relatively high cost and weight of most HUMS relative to the asset value of the aircraft. Hence, given the lack of a HUMS regulatory mandate, the operator is left with a business decision to install HUMS based on a return on investment. For these operators, implementation of HUMS will likely be based on the value the HUMS brings to their operation.

This suggests that HUMS must be explicitly designed to deliver quantifiable returns on investment. HUMS can do this via Rotor Track and Balance (RTB), Flight Data Monitoring/Operational Exceedance Monitoring, Engine Performance Monitoring, and Drivetrain Monitoring/ Prognostics.

While most of these functions directly benefit by reducing maintenance or improving safety through the support of a Safety Management System, Drivetrain Monitoring is an investment. In the rare event that a propagating drivetrain occurs, Condition Monitoring (CM) and prognostics allow the operator to plan when maintenance is carried out and replace components prior to them, reducing the reliability/safety of the aircraft.

Condition monitoring uses vibration sensors and configuration representing the drivetrain of the helicopter to calculate condition indicators (CIs). These CIs are used to infer the current health of the component and, with a health threshold, estimate the remaining useful life (RUL) of the component. The RUL (e.g., prognostics) allows the operator to better manage the asset by scheduling maintenance opportunistically. The goal, along with increased asset safety, is improved availability and more opportunities for revenue generation.

Along with shafts and gears, the drivetrain has bearings that can degrade over time. While shaft/gear fault detection is usually based on synchronous analysis, bearings are asynchronous because their motion depends on non-Hertzian contact and geometry. Additionally, due to the nature of bearing faults (e.g., measuring the effect of an impact-inducing resonance in the bearing itself), successful fault detection requires careful consideration of parameter inputs necessary (e.g., envelope window) to perform the analysis.

On Oct 31st, 2020, a lightweight HUMS-equipped Bell 407GXi helicopter generated an alert on the M250 engine power turbine shaft. This was the third operation of the day, and after the automated download, the system reported a high 2/Rev. An alert was sent to both HUMS monitoring personnel and the pilot. This alert was not a surprise: on Aug 12th of that year (about 100 flight hours prior), the HUMS had identified damage on the No. 5 bearing. The high 2/Rev resulted from wear/excessive clearance in the No. 5 bearing.

In addition to the HUMS text message alert, the pilot reported a flickering engine oil filter bypass flight. After the alert, the pilot checked the magnetic chip detectors and found heavy deposits of fine metallic "fuzz" (see figure 8). An inspection of the oil filter revealed more metal debris. The alert and findings were reported to Bell Flight, and the aircraft was pulled from service to have the engine replaced.

For a bearing fault to be detected, the design of the signal processing techniques must be capable of extracting the bearing fault feature. To do this effectively, it must:

- Determine the bearing resonance, window bandwidth, acquisition sample rate, and spectrum resolution.
- Determine the component health from the measured condition indicators, and from the current state, determine the remaining useful life.

The following will focus on the processes for determining bearing fault frequencies and the analysis performed to quantify bearing damage as a condition indicator. Thresholding and remaining useful life calculation have been discussed in previous papers (Bechhoefer, Dube 2020).

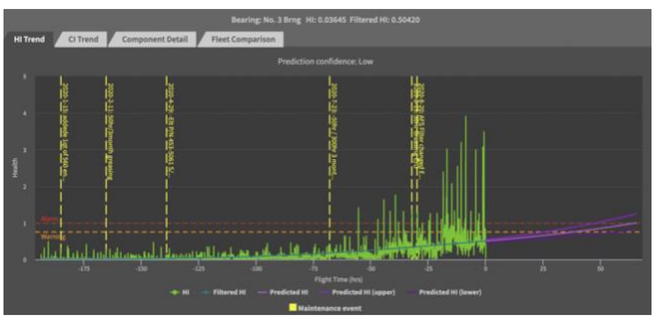

Figure 1. Prognostics of a No. 5 Bearing Fault.

2. BEARING ANALYSIS CONSIDERATIONS

The Rolls Royce M250 engine is a complex analysis environment. The power turbine section comprises five shafts, six gears, and 13 bearings. The RPM range of the monitored shafts is 73 to 536 Hz, with gear mesh frequencies from 3900 to 18800 Hz and bearing fault frequencies ranging from 30 to 5800 Hz. When shaft and gear harmonics are added, this results in a busy vibratory signal, making bearing fault detection difficult. While calculating the bearing rates is well established, it is a complex problem in practice.

Bearing Envelope Analysis (BEA) is based on the demodulation of high-frequency resonance associated with bearing element impacts. For rolling element bearings, when the rolling elements strike a local fault on the inner or outer race, or a fault on a rolling element strikes the inner or outer race, an impact is produced. These impacts modulate a signal at the associated bearing pass frequencies, such as that associated with the Cage (FTF, fundamental train frequency),

$$FTF = \frac{S}{2}\left(1 - {}^{Bd}/_{D}\, cos(\phi)\right) \tag{1}$$

or the Ball Pass Frequency Inner Race (BPFI),

$$BPFI = \frac{Nb \times S}{2}\left(1 + {}^{Bd}/_{D}\, cos(\phi)\right) \tag{2}$$

or the Ball Pass Frequency Outer Race (BPFO),

$$BPFO = \frac{Nb \times S}{2}\left(1 - {}^{Bd}/_{D}\, cos(\phi)\right) \tag{3}$$

and the Ball Pass Spin Frequency (BSF),

$$BSF = \frac{D \times S}{Bd}\left(1 - \left({}^{Bd}/_{D}\right)^2 cos(\phi)^2\right) \tag{4}$$

Where:

S is the shaft frequency,
Bd is the ball or roller element diameter
Nb is the number of balls or rollers
D is the pitch diameter and
Φ is the contact angle.

Note that the BSF is usually constructed at half of the rate (divided by two), but because a spall or damage to the roller hits both the outer and inner race, the observed frequency is 2x, or as given in eq 4. Be aware that these rates are not exact due to non-Hertzian contact and because of the change in geometry under thrust (Hamrock, Dowson 1987). This is important as the No. 5 bearing is a thrust bearing.

From Hamrock., it is seen that for rigidly mounted bearings incapable of radial deformation, the contract angle β, due to a thrust load, can be written as:

$$\beta = cos^{-1}\left({}^{D - P_d/2}/_{D + \delta}\right) \tag{5}$$

Where P_d is the diametral clearance, and δ is total elastic deformation or wear, then:

$$\cos(\beta_f) = {}^{D - P_d/2}/_{D} \tag{6}$$

Given a manufacturer contract angle β_f of 14.7 degrees, the P_d is calculated as 0.0183. Rearranging terms give:

$$\delta = cos^{-1}\left({}^{cos(\beta_f)}/_{cos(\beta)}\right) \tag{7}$$

The thrust loads axial deflection, δ_t, given in 7, is then:

$$\delta_t = (D + \delta)sin(\beta) - Dsin(\beta_f) \tag{8}$$

and finally:

$$\delta_t = {}^{D\,sin(\beta - \beta_f)}/_{cos(\beta)} \tag{9}$$

Given that the No. 5 bearing is under thrust, one can hypothesize that the turbine axial load will cause a displacement δ_t, which is both changing the contact angle of the bearing from β_f to β and the pitch diameter from D to $D+\delta_t$. This change in contact angle and apparent pitch diameter *increases* the observed frequency.

This model helps explain that the fault frequency is 6% higher than expected for the calculated fault rates.

Further, in many real-world bearing faults, it is typical to see a 0.5 to 1% slip due to the non-Hertzian contact. This slip usually does not change as a fault propagates. However, for the No. 5 bearing, it was observed that the slip increases non-linearly with damage (Figure 10). The calculated BPFI rate (eq 2) is 7.05 orders. When fault propagation began, it was initially seen that the fault frequency was 7.22 orders or 3.0% high. However, as the damage propagated, the slip increased, and the fault frequency dropped to 6.66 orders or 5.6% low.

To capture these faults on thrust bearings, the HUMS system should be designed to search the peak energy between 4% above and 5% below the calculated fault feature rate.

2.1. Improvements to Spectral Estimation: Resampling

As the fuel control system on the M250 engine has finite bandwidth, the RPM of the power turbine varies slightly over time. In general, this can vary as much as half a percent. Figure 2 shows the power turbine shaft rate change over a 2-second acquisition.

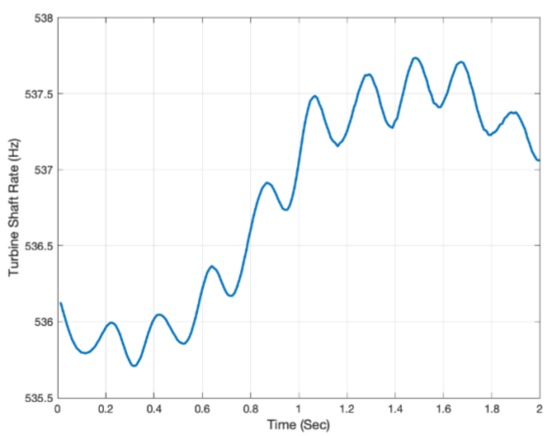

Figure 2. Power Turbine Shaft Rate vs. Time

It is not unusual to see a change in RPM of 0.5%. Consider that for this application, the minimum bandwidth needed to capture the fault feature (No. 5 inner race rate is 5.8kHz) is 6 kHz, then the spectral bin width is, say, 1.5 Hz. Given a change in shaft speed of 0.25%, the spectral energy would be spread over ten bins: 5800 *0.0025 / 1.5 = 10. This suggests that when the spectrum is calculated, the measured energy is smeared over 9 to 10 bins and underestimates the energy associated with the bearing fault by up to 60%.

This change in shaft rate violates the stationarity required for spectral analysis. Implementing a resampling algorithm can control this and reduce spectral smearing. The model for vibration in a shaft in a gearbox was given by McFadden, 1987 as:

$$x(t) = \sum_{i=1}^{k} X_i \times \left(1 + a_i(t)\right) \times cos(2\pi i\, FM(t) + \Phi_i) + b\,(t) \tag{10}$$

Where:

- X_i is the amplitude of the k^{th} mesh harmonic
- *FM(t)* is the average frequency of interest
- $a_i(t)$ is the amplitude modulation function of the *i*th feature harmonic.
- Φ_i is the initial phase of harmonic k, and
- *b(t)* is additive background noise.

The measured frequency is a function of the shaft rotational speed: $FM = Bf(t)$, where B is the bearing fault feature rate, and *f(t)* is the shaft speed as a function of time. As noted, because of the finite bandwidth of the feedback control, there is some wonder in the turbine power shaft speed. This change in speed will result in the smearing of amplitude energy in the frequency domain.

Suppose a tachometer signal is present (such as a key phasor), and the ratio from the key phasor to the shaft under analysis is known. In that case, the vibration data can be resampled so that the number of data points between one revolution and the next is the same. In the case of time-synchronous averaging (TSA), the ensemble of (eq 10) is calculated by summing each revolution resampled data, then dividing by the number of revolutions during the acquisition.

Since the radix-2 FFT is most used, the number of data points in one shaft revolution (r_n) are interpolated into *m* number of data points, such that:

- For all shaft revolutions *n*, *m* is larger than *r*, and
- $m = 2^{\text{ceiling}\,(\log2\,(r))}$ (again assuming Radix 2 FFT)

However, this is not the case for an asynchronous resample algorithm. Since bearing envelop analysis will be done on the resampled signal, a radix-2 length is unnecessary. Instead, the largest time between key phasor zero crossing is used (corresponds to the lowest shaft rate during an acquisition) to calculate the resample data length, *l*. Then, for each revolution, the current set of data points over one revolution are resampled to length *l*.

For example, say the sample rate was 1000 samples per second, and the lowest shaft rate was 10 Hz for a .5 second acquisition. The resample length, *l*, is set to 100. The number of data points between each key phasor (for example) is 87, 92, 100, 95, 89, and 37. For each shaft revolution, the data is resampled to length *l*.:

Rev 1: 87 resamples to 100.
Rev 2: 92 resamples to 100,
Rev 3: 100 resamples to100,
Rev 4: 95 resamples to 100,
Rev 5: 89 resamples to 100.

Note that for half of a second of data, there are 500 data points: the remaining 37 data were in the next incomplete revolution, so the last 37 data points are dropped.

Because of interpolation, the sample rate for each revolution is now changed. An apparent sample rate is needed to determine the frequency associated with an FFT bin. The apparent sample rate is the original sample rate × length of the resampled data/length of the original data: 1000 × 500/463, or 1082.

For shaft and gear analysis, existing TSA algorithms control for changes in shaft speed. For a bearing, the TSA is inappropriate for three reasons: Frist, bearings are quasi-stationary – there is always some slippage such that even with the correct geometry, the rates are not exact. This will make the bearing component non-synchronous with the TSA algorithm and may separate the bearing signal out of the TSA. Second, a bearing has rates for each component: cage, ball, inner and outer race. This would require the TSA to be run four separate times for each bearing. While this may not be a problem for offline analysis, it may exceed the resources of an online analysis system. Consider that any given shaft is supported by 2 to 3 bearings, which would require 8 to 12 TSA analyses.

Finally, in evaluating bearing health, it is essential to see the relationship between the shaft, cage, ball, inner and outer race fault features. For example, an inner race fault is modulated by shaft (e.g., sidebands that are spaced at one order from the inner race fault) can be used to validate the fault. That is, it would be expected to find sidebands on inner race or ball faults because they should be modulated by the load. The TSA cannot capture that because the other components in the bearing would be asynchronous to it for any given component.

2.2. Improvements to Spectral Estimation: FFT

While not obvious, the spectrum is exact only for integer value frequencies. This means that the spectrum peak does not represent the total energy at that frequency for arbitrary signals. For length *n* sampled sinusoid signals, such as used in the FFT, orthogonality only holds for harmonics of the sample rate divided by *n*, or specifically, for these frequencies:

$$f_k = k\,{}^{f_s}\!/_{N}, \;\; k = 0,1,2,\dots,N-1. \tag{11}$$

Because of this, the FFT is defined only for frequencies that are exact integers of the sample rate (f_s), i.e., $\omega_k = 2\pi k f_s/N$. For frequencies that are not exact integers of the sample rate, it can be shown that the FFT acts as a digital filter, where the frequency response for some unknown frequency, ω_k, is:

$$|X(\omega_x)| = \left| {}^{sin\{(\omega_x-\omega_k)\,NT/2\}}\!\big/_{sin\{(\omega_x-\omega_k)\,T/2\}} \right| \tag{12}$$

Which causes spectral leakage and error in estimating the amplitude for frequency ω_k. Note that spectral leakage is not minimized by increasing the length of the FFT.

The HUMS should implement spectral interpolation to accurately measure the bearing energies of a fault. Interpolation can allow for an estimate of the maximum energy value and the frequency at which it occurs to be by assuming a continuous function or distribution. A polynomial quadratic, which is a smoothing function, can be used to interpolate the maximum energy through a triplet of points. A polynomial quadratic has a general formula that can be expressed as:

$$y(x) = a(x-p)^2 + b \tag{13}$$

Where *a* is the curvature and depends in this context on the window used in the FFT; *p* is the center point and gives the interpolated peak location, and *b* is the amplitude that here equals the peak amplitude of the FFT spectrum. The three samples nearest the peak can be defined as:

$$y(-1) = \alpha,\; y(0) = \beta,\; \text{and}\; y(1) = \lambda,$$

Where the bins about those three peaks are defined at [-1 0 1]. Substituting the bin values for x and simplifying results in the following expressions for α, β, and λ:

$$\alpha = ap^2 + 2ap + a + b \tag{14}$$

$$\beta = ap^2 + b \tag{15}$$

$$\lambda = ap^2 - 2ap + a + b \tag{16}$$

By combining equations, substituting, and rearranging terms, the following relations can be derived:

$$\alpha - \lambda = 4ap \tag{17}$$

$$p = {}^{\alpha-\lambda}\!/_{4a} \tag{18}$$

$$\alpha = ap^2 + \left({}^{\alpha-\lambda}\!/_{2}\right) + a + (\beta - ap^2) \tag{19}$$

$$a = {}^{1}\!/_{2}\,(\alpha - 2\beta + \lambda) \tag{20}$$

Then, the interpolated peak location, *p*, in bins, can be expressed as:

$$p = {}^{1}\!/_{2}\,\frac{\alpha-\lambda}{\alpha-2\beta+\lambda} \tag{21}$$

If *k* is the index of the maximum spectral value, then the interpolated frequency (*frq*) would be determined as follows:

$$frq = (k+p)\,{}^{SR}\!/_{window\ length} \tag{22}$$

The interpolated magnitude (*mag*) is determined from the following:

$$mag = y(x) = \beta - {}^{1}\!/_{4}\,(\alpha-\lambda)p \tag{23}$$

In the example data from figure 3., the three points at 157.346 Hz, 158.776 Hz, and 160.21 Hz, with energies of 0.2895 G, 0.9447 G, and 0.7184 G, respectively, are used to interpolate the maximum energy.

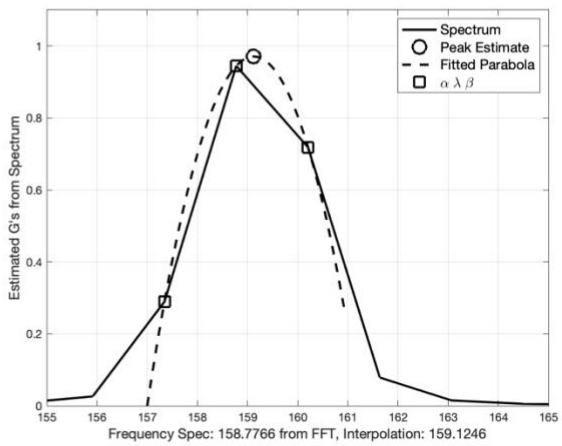

Figure 3. Polynomial Estimation of 0.971 Gs and a Frequency of 159.12 Hz.

These three points serve as the points through which the polynomial is fitted and are used to find values for α, β, and λ for the data. With no further transformation of this data, the above interpolation technique has a peak value of 0.9708 G with a frequency of 159.1246 Hz.

This interpolation minimizes the error of spectral estimation, reducing the maximum error of the FFT spectrum from 14.16% to 0.06%, while the mean magnitude error is reduced from 5% to 0.0009%. Combining the resampling algorithm with spectral interpolation helps ensure the most accurate measurement of the spectral energy associated with a bearing fault.

2.3. Bearing Envelope Analysis: Window Selection

The bearing fault phenomenology has been modeled extensively (example: McFadden 84). Typically, it is described as a Dirac or impulse at the bearing fault feature rate (FTF, BPFI, BFPO, or BSF), convolved with the resonate mode of the bearing structure. This can be thought of as an amplitude modulation of the bearing rate (*a*) with the high-frequency carrier signal (resonant frequency (*b*)).

Often it is not easy to distinguish the exact frequency of the resonance. The resonance frequency is critical in the ability of the envelope analysis to extract a fault feature. Demodulation techniques typically do not need to know the exact frequency, but it must be close. With the resonance frequency, the vibration signal is then convolved with that frequency, made analytic, low pass filtered, and enveloped (e.g., take the absolute value of the analytic signal).

The envelope can be formed by taking the FFT (fast Fourier transform) of the time domain signal and calculate the frequency of each bin (e.g., sample rate/length of time series). It is necessary to calculate the “window” indexes (e.g., the range of frequencies in which the bearing is resonating, say 18 to 25 kHz). The Fourier coefficients from the desired window are then copied into the baseband (e.g., 0 to 7 kHz) indexes (this is a Heterodyne operation). Then zeros are written into the remaining Fourier coefficients (Fourier filtering and Hilbert transform, recalling that the Hilbert transform is defined at the real frequency of the Fourier domain, Randall, 2011). Finally, take the absolute value of inverse FFT Fourier coefficients. See appendix for example Matlab© code.

A successful bearing analysis is a function of determining the resonance of the bearing structure. Bearing components have several vibration modes, which will correspondingly generate resonance at various frequencies throughout the spectrum. The frequency range selection used to demodulate the bearing rate signal (e.g., the window center frequency) should consider some issues.

In particular, the M250 engine gearbox spectrum contains many high frequencies from shaft and gear harmonics, which would prevent analysis at lower bearing envelope frequencies. It is essential to ensure that the other rotating sources do not mask the demodulated bearing frequencies, such as shaft and gear mesh, present at FTF, BPFO, BPFI, and BSF rates. Be aware that shaft order amplitudes of 0.1 G's and gear mesh amplitudes of 10s of G's are typical. Damaged bearing amplitudes are 0.01 G's and can be easily concealed by these other sources.

Several techniques have been suggested to identify an appropriate envelope window. Spectral Kurtosis (McFadden, 1984, Randal, 2011) has been reported as a powerful method for determining the best window.

Kurtosis is a non-dimensional quantity that measures the relative “peaked-ness" of a distribution relative to the Gaussian distribution. Spectral kurtosis (SK) is a statistical parameter indicating how the impulsiveness of a signal varies with frequency. Faults associated with rolling element bearings give rise to short impulses. The SK will be significant in frequency bands where the fault signal is dominant and small where stationary signals dominate the spectrum. Antoni, Randall 2006, developed the kurtogram, which is a map indicating the optimum center frequency and bandwidth combination (Figure 4).

While spectral kurtosis works well in a simple gearbox system, it can return poor results in complex gearboxes. The kurtogram in figure 4 has a maximum kurtosis at a center frequency of 41.5 kHz, with a bandwidth of 200 Hz. Using the envelop window would entirely miss the bearing fault. The next best window has a center frequency of 35.2 kHz, with a bandwidth of 23.4kHz. This would capture the bearing fault, but the spectrum bin width is too large to enable a specific bearing classification and may not allow detection because of signal contamination. From figure 5, it is seen that there is broad spectral content from DC to 50 kHz, with gear mesh present at up to 18.8 kHz (Power Turbine Input Pinion) and spectral content at 41.5 kHz, as seen in both figures 5 and 6. This is generated by the compressor itself (flow tones) and is not a bearing. Hence, something other than spectral kurtosis is needed to inform the bearing envelop window frequencies.

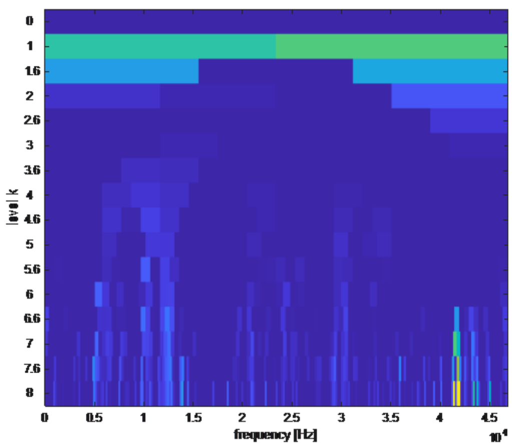

Figure 4. Fast Kurtogram from the M250 engine.

Lord Rayleigh (1894) equated kinetic energy at the mean position of a beam to strain energy at the maximum displacement on a ring with equal nodal configuration. When evaluated, this equation seemed to underestimate the natural frequency of the bearing when tested. Timoshenko (1940.) further developed the concept of Rayleigh. Timoshenko teaches that for a ring with uniform mass, the exact shape of the mode of vibration consists of a curve which is a sinusoid on the developed circumference of the ring.

The natural frequencies are then:

$$\omega_s = n(n^2 - 1)/\sqrt{n^2 + 1} \quad \sqrt{EI/\mu R^4} \qquad (24)$$

where:

μ is the mass per unit length,

EI is the bending stiffness (Youngs Modulus x Inertia)

R is the radius.

Taking measurements from the bearing for diameters of the inner and outer race, the width of the bearing, density of steel, and the natural frequencies for the second through fifth modes are:

- 6.1, 17.2, 32.9, and 53.2 kHz.

Note that the analysis requires a bandwidth of at least 6 kHz for this application to capture the inner race. As we wish to try and avoid gear mesh frequencies and that the accelerometer itself is bandwidth limited to about 43 kHz, analysis was performed with two configurations: 17 to 23 kHz and 32 to 38 kHz.

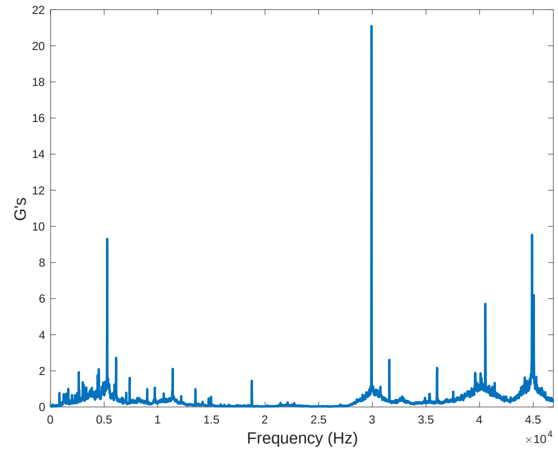

Figure 5. Raw Spectrum from the M250 Engine.

As can be seen, using the 17 to 23kHz window (Figures 6 and 7), the environment is quite complex. Note that the BPFI is 3% high early in the fault propagation, and one month later, the BPFI rate has dropped to just under the calculated 7.05 value (figure 6, 7).

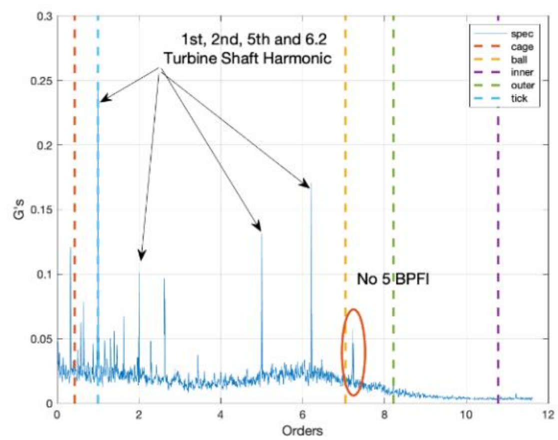

Figure 6. Initiation of the bearing fault.

It is interesting to note that the harmonics of the turbine (e.g., 1/Rev impact, perhaps indicating some mechanical looseness) is always present. It was found that the 6.2 order peak (which may be an interaction with the 5th turbine shaft harmonics and the 31-tooth turbine wheel) was synchronous to the turbine shaft and always present.

3. USER ALERTS AND DISPLAY

In general, helicopters have inspections every 50 hours of flight time, with heavier maintenance conducted at 100 and 300 hours. Aircraft also have annual inspections. Typically, the number of hours flown per month is dependent on the operator's mission, and it is not surprising to see fleets average 300 to 500 hours per annum. Of course, for operators conducting aerial inspections (inventorying power poles and examining power lines for encroachment) or other seasonal missions (firefighting, by dropping water, or delivering man/material to a fire), these aircraft can fly as much as 25 to 40 hours per week.

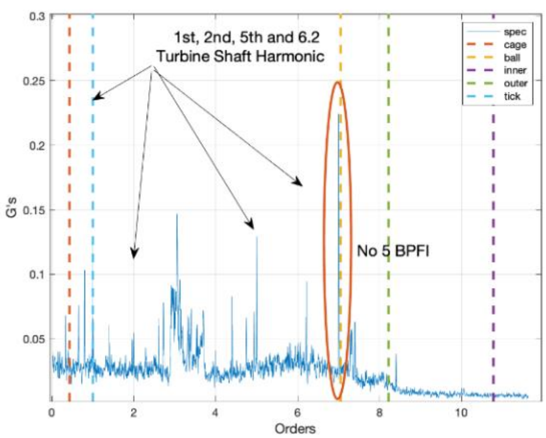

Figure 7. Bearing fault one month after fault initiation.

An accurate remaining useful life (RUL) calculation is essential to support and supplement the already established maintenance practices. HUMS with an RUL capability allows the fleet operation manager to order parts and schedule the right personnel to perform maintenance. This turns an intrusive unscheduled maintenance action into scheduled maintenance with adequate preparation.

For the normal category helicopter market (part 27) or cases where the aircraft has no extended overwater/over hazardous terrain flights, HUMS provides logistic support to improve availability (e.g., allowing the generation of more revenue flights) more so than a safety of flight requirement. Having been type certificated and adequately maintained, these aircraft are inherently safe. In most cases, the worst possible outcome for the bearing fault in figure 1 would be that the pilot sees a chip light annunciation (indicating metal debris in the gearbox) and is forced to land. As noted, a goal of HUMS is to generate a maintenance action before a chip light. This maintenance is done opportunistically while the aircraft is already down for some required scheduled inspection.

Using a fracture mechanics model, a prognostic requires four inputs to calculate an RUL.

- An estimate of the current component health.
- An estimate of when it is appropriate to do maintenance, e.g., the threshold.
- An estimate of the future component load.
- A component degradation process model takes the current component health and the estimated future load and calculates the time/cycles to when it is appropriate to do maintenance.

The HUMS calculates condition indicators (CI) representative of bearing health. Ideally, these condition indicators are also proportional to the extent of the component damage. This allows HUMS to provide a virtual inspection of the drivetrain components.

The estimate of when it is appropriate to do maintenance is a threshold-setting problem. A hypothesis testing approach has been adopted for this paper (Bechhoefer et al, 2011). In this paradigm, the measured set of condition indicators provides evidence that the component is no longer “good ."The CIs are used to reject the Null Hypothesis that the component is nominal. If the component is not nominal, it is appropriate to perform maintenance.

In a hypothesis test, it is observed that all condition indicators (CIs) have a PDF. Any operation on the CI to define a health index (HI) is then a function of distributions. The HI function in the application is the weighted norm of *n* CIs (e.g., the normalized energy of *n* CIs), where the weights are determined by the Jacobian (the inverse covariance):

$$HI = {}^{0.35}/_{critical}\sqrt{\boldsymbol{Y}^T\boldsymbol{Y}} \tag{25}$$

Where ***Y*** is the whitened, normalized array of CIs, and *critical* is the critical value of the test. The critical value is calculated from the inverse cumulative distribution function (ICDF) for a given probability of false alarm in a hypothesis test. For Eq. 25, the ICDF is the Nakagami where η is the number of CIs in the array and = *n*, and ω = η/(2-π/2)*2.

A normalized HI > 0.35 for a component indicates that the Null Hypothesis is rejected. That is, the component is no longer nominal. Note; however that maintenance is not recommended until the HI > 1. These threshold values have been tested on numerous helicopters, wind turbines, and seeded fault testing on 60+ gearboxes. The level of damage for an HI of 1.00 is typically moderate visible damage

3.1. RUL Calculation using the Linear Elastic Model

RUL is taken as the time when it is appropriate to do maintenance (moderate visible damage) and *not* the time until the component fails. Maintenance is a process of restoring the equipment to the original design reliability for aviation applications. Worn or damaged parts have reduced reliability, and maintenance repairs or replacing those parts restore the system's design reliability. The concept that an HI exceeding 1 triggers a maintenance event is complementary to existing maintenance practices as it is designed to restore the system to the manufactures design requirements.

For an example of a critical system, the design reliability is typical "six-nines" (e.g., the probability of failure of the part under design loads is less than 10^{-6} per hour). For the damaged part, the reliability may be reduced to three-nines or a probability of failure of 10^{-3}. Thus, the appropriateness of repairing the faulty component can be seen as an action to restore the designed reliability of the system. From a maintainer perspective, then:

- HI reflects the current components damage, where the probability of exceeding an HI of 0.35 is the PFA.
- A warning (yellow) alert is generated when the HI is greater than or equal to 0.75. Therefore, maintenance should be planned by estimating the RUL until the HI is 1.0.
- An alarm (red) alert is generated when the HI is greater than or equal to 1.0. The propagation rate becomes unpredictable, and continued operations could cause collateral damage.

This threshold setting model ensures that the probability of a false alarm is exceedingly small when the HI reaches 1. However, from numerous installations and seeded fault tests in practice, a bearing at HI 1 has easily seen physical damage.

A component with an HI value does not define the probability of failure for the component, nor does it indicate that the component may fail when the HI reaches 1.0. Instead, defining maintenance at an HI of 1 initiates a proactive policy to change operator behavior. The desire is to reduce the cost and time associated with component failure by performing maintenance prior to generating collateral or cascading faults. For example, by performing maintenance on a bearing before the bearing sheds extensive material, costly gearbox replacement can be avoided, and the reliability of the gearbox can be restored to its design requirements.

As such: RUL is defined as the time from the current HI until the HI is greater than or equal to 1.

For many materials, such as steel used in gears and bearings, which are subject to tensile loading cycle, the fatigue crack growth is Mode 1 and can be expressed as:

$$\frac{da}{dN} = D\left(2\sigma(\pi)^{1/2}\alpha\right)^m a^{m/2} \tag{26}$$

Where:

- da/dN is the rate of change in the half crack length per cycle, where crack length is a
- D is a material constant
- m is the crack growth exponent for steel is 4.
- σ ιs strain (e.g., torque would be a surrogate)
- α is a shape factor

Inverting and integrating gives N, the number of cycles gives. For constant shaft rate machines, such as a helicopter, N is the RUL. Using the HI as a surrogate for crack length, then it can be shown that an estimate for RUL is:

$$RUL = {}^{-dt}/_{HI} \times HI_0 \times ln\left({}^{1}/_{HI_0}\right) \tag{27}$$

For more details (Bechhoefer, Dube, 2020).

4. DISCUSSION

The No. 5 bearing fault was identified, with an RUL of approximately 120 flight hours in mid-August of 2020. The HUMS sent a notification to both the aircraft operator and the HUMS manufacturer. The bearing was put on a watch list. Based on the aircraft usage, it was estimated that the bearing HI would be in alarm in early November of 2020. Over the next 120 hours, with each download, the status of the bearing was updated. On Oct 31 of 2020, the HUMS alerted that there was an increase in the 2/Rev of the power turbine shaft. This suggested increased mechanical looseness. The No. 5 Bearing HI was 1.2. There were no other secondary indications of fault. At this point, the operator inspected the upper and lower engine chip detectors (figure 8).

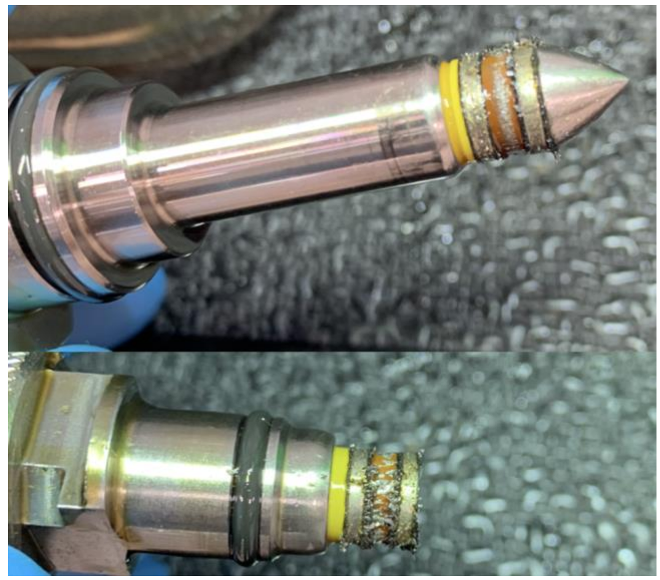

Figure 8. Upper and Lower M250 Chip detectors.

It was clear that there was metal due to some mechanical failure, and the engine was pulled from service. A teardown analysis indicated both ball, inner and outer race damage (figure 9).

Figure 9. No. 5 Bearing outer race damage.

A damaged No. 5 bearing has been known to result in engine shutdown in flight. This can be a dangerous event and lead to loss of life and the asset. The ability to notify the operator of the bearing damage and alert prior to a “chips light” reduced the risk of mishap and facilitated the operator's ability to better manage part replacement. Both the engine and aircraft manufacturer had been notified three months prior to the engine being pulled from service. HUMS provided information that potentially saved both the pilot and the asset.

Bearing failures in aviation applications are rare. The design of a system that periodically captures raw data provides opportunities for continued learning and system maturation. Two things were learned from this successful bearing fault that could be improved. Initially, only one envelop analysis was performed using a window of 17 to 23 kHz. While bearing trend data for the ball and outer race were observed, no trend data was observed on the inner race. This is somewhat puzzling given the amount of metal discovered on the chip detector and teardown. Subsequently, a second envelop analysis was added at the higher 32 to 38 kHz was added to the fleet, and features consistent with inner race faults were detected. That is, it was later seen that the inner race feature was present at the 32 to 38 kHz envelope window.

Second, in observing many bearing faults, it is rarely seen that the bearing slip from non-Hertzian contact increase with damage. Typically, the HUMS collects data once every three minutes, with raw data once per flight. In the case of the Rolls Royce M250 engine No. 5 bearing, from the period starting in June until the engine was removed from service, raw vibration data was collected 53 times. This allowed for more in-depth analysis and allowed for the slip to be calculated over the degradation period (figure 10).

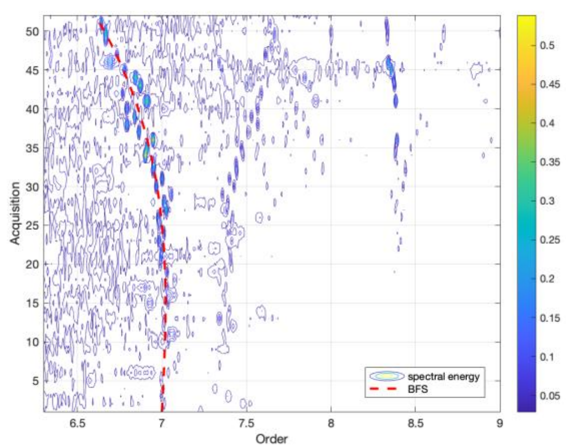

Figure 10. Change of bearing fault frequency over time.

Note: that the x-axis is in under of orders. An order is the ratio of the fault frequency to the shaft rate. This display enhances analysis for asynchronous components such as bearing, as the fault frequency will not be an integer order value.

The slip in bearing fault frequency can be seen in that the ball spin frequency decreased for 7.2 to 6.5 orders or a change of –8%. This caused some confusion with the No. 6 bearing on the engine. The No. 6 bearing has a ball spin frequency of 6.55, which was confounded with the No. 5 bearing data. The analysis algorithm detected energy at 6.5 order, generated by the No. 5 bearing, and applied it to the No 6. bearing health. As the bearing fault frequency dropped below the frequency limit applied for the No. 5 bearing, the health trend of the No. 5 bearing decreased, while the No. 6 bearing health increased. Subsequently, the search range for No. 5 bearing was increased to account for this condition.

5. CONCLUSION

The development of a HUMS for normal category helicopters requires a value-based approach. The system design must ensure that the operator/maintainer will receive a return on investment, and the system must be shown to work. The Foresight HUMS demonstrated the ability to detect and recommend replacing a No. 5 bearing on the engine long before the fault affected the reliability of the aircraft. The customer and airframe manufacturer were notified of the fault three months prior This allowed activation of logistics support such that when it was recommended that the aircraft be pulled from service, replacement parts were available.

To achieve this level of performance in this complex environment, the system's design required several innovations to extract the fault feature and alert the customer. Those design features included:

- A method of system configuration to account for the excursions in the observed bearing fault features due to changes in the bearing dynamics from thrust and non-Hertzian contract.
- Resampling of the vibratory data to account for changes in operating RPM to reduce spectral smearing of the data.
- An improved method of spectral estimation to measure the envelope spectrum more accurately.
- A model-based method is needed to determine the best window frequencies for the bearing envelope analysis.
- Thresholding and reporting of remaining useful life (RUL) to allow the operator to better manage their fleet.

These system-level features allow HUMS to function better. The ability for HUMS to provide value to a customer improves the likelihood of adoption in normal category (type 27) rotorcraft.

REFERENCES

Hamrock, B., Dowson, D., "Ball Bearing Mechanics," NASA Technical Memorandum 81691, 1981

McFadden, P., (McFadden, 1987) "A revised model for the extraction of periodic waveforms by time-domain averaging," Mechanical Systems and Signal Processing 1 (1) 1987, pages 83-95

P. D. McFadden and J. D. Smith, "Model for the vibration produced by a single point defect in a rolling element bearing," Journal of Sound and Vibration, vol. 96, no. 1, pp. 69–82, 1984.

Randal, R. B., Vibration-based Condition Monitoring, John Wiley & Sons, 2011.

Antoni, J., Randall, RB., "The spectral Kurtosis: application to the vibratory surveillance and diagnostics of rotating machines," Mechanical systems and signal processing 20(2), 2006

Rayleigh, Lord, 1894, Theory of Round, 2nd Ed, London, Macmillian

Timoshenko, S., Theory of plates and shells, McGraw Hill, New York, 1940.

Bechhoefer, E., He, D., Dempsey, P., "Gear Health Threshold Setting Based on a Probability of False Alarm," Conference of the Prognostics and Health Management Society, 2011.

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

APPENDIX

Matlab® example for Envelope Analysis.

```
function [env,dty] = envelope( data, dt, lowf,
highf)
%Inputs:
% data: vector, time domain
% dt  :sampling time interval
%lowf: low frequency limit of bandpass filter
%highD:high frequency limit of bandpass filter
%Outputs:
% env       :Envelope of data
% dty       :decimated sample rate
n = length(data);
dfq = 1/dt/n; %freq per bin
idxLow = floor(lowf/dfq);%get the bin low
idxHi = ceil(highf/dfq); %get the bin high
D = fft(data);
idx = idxHi-idxLow + 1;

D(1:idx)  = 2*D(idxLow:idxHi); % this is the
definition of the
D(idx+1:end) = 0;% Hilbert + low pass filter
data = abs(ifft(D)); % The Envelop
bw = highf - lowf;
r = fix(1/(bw*2*dt)); % Decimation rate
env =  data(1:r:n);   % Decimate
dty = dt*r;% New sample rate after decimation
```