

A Comprehensive Analysis of the Performance of Gear Fault Detection Algorithms

Eric Bechhoefer¹, Brent Butterworth²

¹*GPMS Inc, Cornwall, VT, 05753, USA*
eric@GPMS-VT.com

²*Garmin International, Midlothian, TX, 76065, USA*
Brent.butterworth@garmin.com

ABSTRACT

Modern gear fault detection analysis began with algorithms based on the time synchronous average. Over the course of decades, many gear analyses have been proposed, but with no evidence that the analysis was significantly more powerful in terms of fault detection than existing algorithms. This study focuses on a comprehensive comparison of gear fault detection algorithms to evaluate their performance. Using a large, statistically significant set of data from three nominal machines and a damaged machine, the condition indicator (CI) responses of 88 different analysis are compared in terms of their statistical significance to detect a cracked tooth. The comparison includes residual, energy operator (and its variants), the narrowband analysis (with a comparison of bandwidth requirements), the amplitude and frequency modulation analysis, an analysis of variance of the "factor" analysis: crest, shape, impulse, and margin, and other standard gear fault CIs. Further, the effect of CI selection in the establishment of gear component health is evaluated, where given a set of CIs, a gear health indicator is built, showing that CIs with high statistical separability and low correlation have improved fault detection power. This is validated on a third, dissimilar gear fault propagation test.

1. INTRODUCTION

Detection of gear faults is challenging. Undetected gear faults can result in catastrophic gearbox failures, which depending on the criticality of the application, can be life threatening. For example, as a result of a number of North Sea helicopter accidents, many large helicopter operators have installed Health and Usage Monitoring Systems (HUMS). The Civil Aviation Authority (CAA), by issuing ADD 001-05-99,

recognized the benefits of monitoring, mandating gearbox monitoring to protect, in part, against gear fault.

A vast literature on automate gear fault analysis has been generated, beginning with the seminal work of R.M. Stewart's *Some Useful Data Analysis Techniques for Gearbox Diagnostics* (1977). Stewart gave us the use of the time synchronous average (Bechhoefer, 2009), the residual/difference analysis and mesh specific Figures of Merit. Building on Stewart, P.D. McFadden reported Amplitude and Phase Demodulation techniques (1986), while Ma (1995) introduced Teager's energy operator to the condition monitoring community. A number of NASA researchers (Zakrajsek, Lewicki, 1989, 1993 and 2009) investigate the ability of these, and other various algorithms to detect gear faults. However, to date, there has not been a comparison of the power of the algorithms to detect fault. That is to say, which gear analysis algorithms are the best at detecting a gear fault?

Further complicating this issue is that the gear themselves have a number of failure modes (Gopinath, 2010), to include: Scoring/Frosting, Wear (adhesive, abrasive, corrosive), Pitting (flaking, spalling), Plastic Flow (Ridging/Rippling) and tooth breakage.

Critical to determining an algorithms fault detection capability is a metric for measuring the performance of an analysis. Also, of importance is how this information is displayed for a user. For this paper, a statistical method is proposed to measure separability of a number of standard analyses. These results will be used to construct in improved health indicator, which is designed to provide actionable information to a maintainer. This analysis is based on a large (300+ raw data acquisitions) data set from three nominal gearboxes, and 300 raw data acquisitions from a damaged gear (tooth breakage). The results (e.g. best condition indicators) are tested against another, dissimilar gearbox with a known propagating fault. This study is focused on broken/cracked tooth, as it presents the failure mode that is

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

associated with catastrophic failure. It is cautioned that other failure modes, such as scuffing, may require a different set of analysis.

2. QUANTIFYING GEAR ALGORITHMS PERFORMANCE

Detection: the action or process of identifying the presences of something concealed. For vibration monitoring, it is identifying a feature within the background noise of a gearbox which is indicative of fault. For critical system, the risk is a missed detection that allows a mission to proceed when it should be aborted. Otherwise, the risk is performing unnecessary maintenance due to a false alarm. Inherent to both Probability of Detection (PD), and Probability of False Alarm (PFA) is signal to noise: improving SNR increases the PD and floor a given PFA. A linear measure of signal to noise is signal divided by the standard deviation. Form a statistical sense, this is related to Hypothesis testing. When nominally assuming a Gaussian signal, if the normalized signal exceeds the mean by 3 standard deviation, the PD is 0.9987, while the PFA is 0.0013. For this test, a similar metric is used: statistical separability.

2.1. Statistical Separability

Separability defines the statistical distances between two populations. The concept inherently implies that for distinct populations (e.g. a nominal gear vs. a damaged gear) using an algorithm that can detect the damage, the separability will be large. This is quantified using the pooled sampled standard deviation. This test assumes Gaussian PDF, which for some distribution is a poor assumption. The test statistic (see Medenhall, 1990) used is the small-sample test for comparing two population means, where:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{S\sqrt{1/n_1 + 1/n_2}} \quad (1)$$

Where, \bar{Y}_1 is the nominal gear CI mean, \bar{Y}_2 is the damage gear CI mean, n_1 is the number of samples for the nominal gear, n_2 is the number of samples for the damaged gear, and S , the pooled sample variance is, where S_1^2 is the nominal gear sample variance, and S_2^2 is the damage gear sample variance:

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

The population mean and sample variance are calculated from the gear analysis condition indicators (CIs), where the CIs are statistics derived from the gear analysis.

All analysis and comparisons of analysis for this study will be based on T , the separability of an analysis. This allows a convent way to measure the performance of one analysis against another.

2.1.1. Analysis of Variance/Ranking

Further performance metric will be based on Analysis of Variance (ANOVA). For example, while the T represents the mean separability, one has to ask if it is statically different from another analysis. For a number of similar analysis (energy operator and its related analysis), ANOVA can determine if there is any statistical different between the CI populations. This will also be done for the “factor” analysis: Crest, Shape, Impact and Margin.

2.2. Building Actionable Information: Gear Health

In order to reduce maintainer workload, and improve fault detection, the concept of the health indicator was introduced in (Bechhoefer, 2011). This allows the maintainer a common paradigm for determining when a maintenance action needs to be carried out. In general, the Health Indicator (HI) is a function of distributions of n CIs, where the HI is scaled such that:

- HI ranges from 0 to a positive value, where the probability of exceeding an HI of 0.35 is the PFA (probability of false alarm),
- A warning (yellow) alert is generated when the HI is greater than or equal to 0.75. 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. Continued operations could cause collateral damage.

Note that this nomenclature does not define a probability of failure for the component, or that the component fails when the HI is 1.0. Instead, it suggests a change in operator behavior to a proactive maintenance policy: perform maintenance before the generations of collateral or cascading faults. For example, by performing maintenance on a bearing prior the bearing shedding extensive material, costly gearbox replacement can be avoided.

Hence, from a maintainer perspective, this is a stoplight-based threshold setting/alerting system: when a component is yellow, plan maintenance, and when the component turns red: do maintenance

2.3. Importance of Controlling Correlation Between CIs

All CIs have a probability distribution (PDF). Any operation on the CI to form a health index (HI) is then a function of distributions. The HI function is defined as the norm of n CIs (energy):

$$HI = 0.35 /_{crit} \sqrt{\mathbf{Y}^T \mathbf{Y}} \quad (3)$$

where \mathbf{Y} is the whitened, normalized array of CIs, and $crit$, is the critical value (see Bechhoefer et al, 2011).

The function is valid if and only if the distribution (e.g., CIs) are independent and identical (e.g., IID). As an example, for

Gaussian distribution, subtracting the mean and dividing by the standard deviation will give identical Z distributions. The issue of ensuring independence is much more difficult. In general, the correlation between CIs is non-zero. For instance, **Error! Reference source not found.** shows the correlation coefficients for 6 CIs used for gear fault analysis: most are statically significant.

Table 1: Correlation Coefficients for the Six CIs Used in the Study

ρ_{ij}	CI 1	CI 2	CI 3	CI 4	CI 5	CI 6
CI 1	1	0.84	0.79	0.66	-0.47	0.74
CI 2		1	0.46	0.27	-0.59	0.36
CI 3			1	0.96	-0.03	0.97
CI 4				1	0.11	0.98
CI 5					1	0.05
CI 6						1

This correlation between CIs implies that for a given function of distributions to have a threshold that operationally meets the design PFA, the CIs must be whitened (e.g., decorrelated). A whitening solution can be found using Cholesky decomposition.

The Cholesky decomposition of Hermitian, positive definite matrix results in $\mathbf{A} = \mathbf{L}\mathbf{L}^*$, where \mathbf{L} is a lower triangular, and \mathbf{L}^* is its conjugate transpose. By definition, the inverse covariance is positive definite Hermitian. It then follows that if (Bechhoefer, et al, 2011):

$$\mathbf{L}\mathbf{L}^* = \mathbf{\Sigma}^{-1}, \text{ then } \mathbf{Y} = \mathbf{L} \times \mathbf{CI}^T \quad (4)$$

The vector \mathbf{CI} is the correlated CIs used for the HI calculation, and \mathbf{Y} is 1 to n independent CIs with unit variance (one CI representing the trivial case). The Cholesky decomposition, in effect, creates the square root of the inverse covariance. This, in turn, is analogous to dividing the CI by its standard deviation (the trivial case of one CI). It can be shown that $\mathbf{Y} = \mathbf{L} \times \mathbf{CI}^T$ then creates the necessary independent and identical distributions required to calculate the critical values for a function of distributions.

2.4. Determining the Critical Value for Scaling Gear Health

The critical value is taken from the inverse cumulative distribution function for the HI. The CIs used have Rayleigh like PDFs (e.g., heavily tailed). For magnitude-based CIs, it can be shown that for the nominal case, the CI probability distribution function (PDF) is Rayleigh (Bechhoefer, 2005). For Gear CIs and Bearing CIs (where magnitudes which are biased by RMS), a transform is used by make the CI more Rayleigh like. The transform “left shifts” the CI. For example, a shift such that the .05 CDF (cumulative distribution function) is assigned to 0.0.

Consequently, the HI function is based using the Rayleigh distribution. The PDF for the Rayleigh distribution uses a

single parameter, β , defining the mean $\mu = \beta^*(\pi/2)^{0.5}$, and variance $\sigma^2 = (2 - \pi/2) * \beta^2$. The PDF of the Rayleigh is: $x/\beta^2 \exp(-x/2\beta^2)$. Note that when applying these operations to the whitening process, the value for β for each CI will then be: $\sigma^2 = 1$, such that: $\beta = \sigma^2 / (2 - \pi/2)^{0.5} = 1.5264$.

For the HI equation in (4), the normalized energy of the CIs, it can be shown that the function defines a Nakagami PDF (Bechhoefer, 2011). The statistics for the Nakagami are $\eta = n$, and $\omega = 1/(2-\pi/2)*2*n$, where n is the number IID CIs used in the HI calculation.

2.5. Advantages of the HI vs. the CI for the Maintainer

The use of a common threshold for damage, as per the HI, simplifies the user experience. It does offer two additional benefits. First, the HI allows better control of the probability of false alarm as it whitens/removes correlation between CIs. As an example of the importance of correlation, consider a simple HI function $HI = CI1 + CI2$. The CIs will be normally distributed with mean 0 and standard deviation of 1. The standard deviation of this HI is:

$$\sigma_{HI} = \sqrt{\sigma_{CI1}^2 + \sigma_{CI2}^2 + 2\rho_{CI1,CI2}\sigma_{ci1}\sigma_{ci2}} \quad (5)$$

Where $\rho_{CI1, CI2}$ is the correlation between CI1 and CI2. If one assumes $\rho_{CI1,CI2}$ is 0.0, then $\sigma_{HI} = \sqrt{2}$. For a PFA of 10e-6, the threshold is then 6.722. Consider the case in which the observed correlation is closer to 1 (e.g., $\rho_{CI1, CI2}$ is 1.0), then the observed $\sigma_{HI} = 2$. For a threshold of 6.722, the operational PFA is 4×10^{-4} . This is 390 times greater than the designed PFA. This illustrates the deleterious effect of correlation on the threshold setting.

Second, using an HI with multiple CI give a process gain and improves sensitivity. Consider a simple example where one sets up a case where there are four CIs that have similar performance (e.g. for a damaged component, the normalized CI value is 1). One can ask the question, at what level would the CI have to be for the HI to be one? That is:

$$1 = \sqrt{CI_1^2 + CI_2^2 + CI_3^2 + CI_4^2} = \sqrt{4CI_x^2} = 2CI \quad (6)$$

Hence, the CI values would be 1/2: the HI using four CI would have a process gain of 3dB. In this study, the HI algorithm will use six CIs, potentially a gain of 3.9 dB.

3. GEAR FAULT CONDITION INDICATORS

It should be observed that there is analysis (e.g. the Time Synchronous Average), and then statistics of the analysis, which define the condition indicator. Generally speaking, some papers (McFadden, 86, Ma, 1995) describe an analysis, but not specific condition indicators, whereas other papers, (Zakrajsek, 1989, and 1993, or Lewicki, 2009) refer the condition indicator (CI) which is a statistic of an analysis (such as the kurtosis of the Narrowband Analysis). Finally,

Stewart (1977) describe CIs which are derived as gear mesh specific statistics, usually ratio of analysis statistics.

For this analysis, the statistics for CIs applied to analysis are:

- RMS: Root Mean Square
- Skewness
- Kurtosis
- Peak to peak (max – min value of a waveform)
- Crest Factor
- Shape Factor
- Impulse Factor and
- Margin Factor

Whereas the analysis performed were based on the:

- Time Synchronous Average
- Residual Analysis
- Energy Operator
- Narrowband Analysis
- Amplitude Modulation Analysis and
- Frequency Modulation Analysis

The flow of the processing is given Figure 1

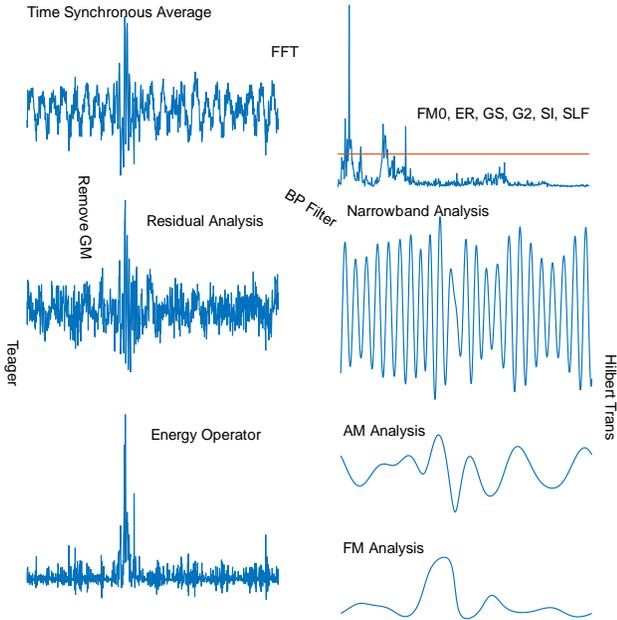


Figure 1 Gear Analysis Process Flow

Note that the starting point for all analysis is the time synchronous average. Observed that the AM and FM analysis start with the Narrowband analysis.

3.1. The “Factor Analysis” Statistic

The “factor” analysis is a family of related, non-dimensional, time domain features which have often proven successful in fault detection (Caesarendra, 2017). The crest factor (CF) is a measure of the impact occurring over one revolution of the TSA and is defined as:

$$\frac{\max(\text{TSA}) - \min(\text{TSA})}{\sqrt{1/n \sum_{i=1}^n \text{TSA}_i^2}} \quad (7)$$

This is the TSA peak to peak divided by the TSA RMS. The shape factor (SF) is affected by the TSA shape, but is independent of its dimensions:

$$\frac{\sqrt{1/n \sum_{i=1}^n \text{TSA}_i^2}}{1/n \sum_{i=1}^n |\text{TSA}_i|} \quad (8)$$

It is the TSA RMS divided by the mean absolute value of the TSA. Both the impact factor (IF) and margin factor (MF) = are similar to the CF, as both are a measure of impact. IF is defined as:

$$\frac{\max(\text{TSA}) - \min(\text{TSA})}{1/n \sum_{i=1}^n |\text{TSA}_i|} \quad (9)$$

And MF is defined as:

$$\frac{\max(\text{TSA}) - \min(\text{TSA})}{(1/n \sum_{i=1}^n \sqrt{|\text{TSA}_i|})^2} \quad (10)$$

3.2. Gear Analysis: The Time Synchronous Average

All of the analyses start with the TSA (Figure 1). A solution to the TSA starts as a model of vibration from rotating equipment, as given in (Bechhoefer, 2009):

$$x(t) = \sum_{k=1}^K X_k [1 + a_k(t)] \times \cos[2\pi k f_m(t) + \phi_k] + b(t) \quad (11)$$

where:

- X_k is the amplitude of the k th mesh harmonic
- $f_m(t)$ is the average mesh frequency
- $a_k(t)$ is the amplitude modulation function of the k th mesh harmonic.
- $\phi_k(t)$ is the phase modulation function of the k th mesh harmonic.
- Φ_k is the initial phase of harmonic k , and
- $b(t)$ is additive background noise.

The mesh frequency is a function of the shaft rotational speed: $f_m = Nf$, where N is the number of teeth on the gear and f is the shaft speed. The vibration model in (11) assumes that that f is constant. This is never the case: there is some wander in the shaft speed due to changes in load or feedback delay in the control system. This change in speed will result in smearing of amplitude energy in the frequency domain. Hence, the TSA algorithm (Appendix A), controls this by interpreting the zero cross time (for ratios that are not 1 and

where the zero cross time is generated from a shaft with more than one pulse per revolution for the target), and by resampling the vibration so that, from revolution to revolution, the $x(t)$ in (11) has the same radian position on the shaft. That is, a one per revolution imbalance represents cosine with period of one revolution. This resampling of the data mitigates the smearing effect, and reduction of non-synchronous noise by $1/\sqrt{\text{rev}}$, where rev is the number of revolutions in the TSA.

Note that when the Fourier transform (usually implemented at an FFT with radix 2 length), bin 0 is DC, bin 1 is one per rev, and the bin for gear mesh is the number of teeth (e.g. 31 tooth gear would be bin 31).

3.3. Gear Analysis: The Residual Analysis

The residual, or different signal, is given by Stewart (1977). A residual signal is defined as a synchronous averaged signal without the gear mesh frequency, its harmonics, driveshaft frequency and its second harmonics. If the first order sidebands about the gear mesh frequencies are filtered out, a differential signal is created. Conceptually, the TSA (a time domain signal) will measure the periodic signal associated a soft/cracked tooth. This fault feature is not sinusoidal and is not associated with $1/\text{Rev}$ or gear mesh, which are removed in the residual signal, leaving the ‘‘impact’’ as is clearly seen in Figure 1.

This study compared four types of residual signal. The ‘‘classical’’ residual were the gear mesh and sidebands are explicitly removed, and the ‘‘Threshold Residual’’, were a search algorithm removes frequencies that are statistically significant (were no gear tooth count is needed).

The last two residual are based on the envelope of the residual, square, as given inn Wang (2001). This is an improvement to McFadden’s (1987) paper. In this test, the ‘‘*amRES*’’, is the envelop of the threshold residual, squared:

$$amRES = |\text{Hilbert}(RES^2)| \quad (12)$$

Whereas *amRES_D* is a low pass filtered version of *amRES*, (the *D* is for decimated, as the bandwidth was 0.5, or decimation by 2).

An implementation of the residual is given in Appendix B.

3.4. Gear Analysis: The Narrowband/AM/FM Analysis

McFadden (1986) understood that in complex gearboxes, it is was not always clear from the sideband modulations, how sever a gear tooth defect might be. Further, for early faults, such as a fatigue cracks, more effective analysis may be needed. McFadden shows that by band pass filtering to remove gear mesh harmonics (e.g. the narrow band analysis, where the bandpass is approximately 25% of the gear tooth count), and then enveloping, one can directly measuring the amplitude and phase (frequency) modulation present in the original time synchronous average. McFadden hypothesised

that as the soft/cracked tooth enters the load zone, the reduced stiffness of the damaged tooth transfers load to the surrounding teeth. This both affects the displacement (AM modulation), and the rotational speed of the gear (FM modulation).

It should be noted that the performance of the narrowband analysis is based on the bandwidth of the bandpass filter used. As such a separate analysis was conducted to evaluate the performance of these analysis based on the varying the bandwidth.

This is a very powerful technique, given in Appendix C.

3.5. Gear Analysis: Teagers’ Energy Operator Analysis

The energy operator is nonlinear tracker which, given an amplitude and phase modulated signal, can measure the product of their instantaneous frequency. It has been observed that in practice the analysis is quite sensitive to torque, which is why these CIs often use kurtosis and crest factor (which tends to normalize the change in magnitude due to torque effects).

Mayo et al. (2016) reported variants of the energy operator (EO) that should have improved performance. Whereas the EO is given as:

$$\Psi_{EO}(TSA_n) = TSA_n^2 - TSA_{n+1} \times TSA_{n-1} \quad (13)$$

The computational enhanced energy operator (CEEO) is given as:

$$\Psi_{CEEO}(TSA_n) = TSA_n^2 - TSA_{n+2} \times TSA_{n-2} \quad (14)$$

Mayo then presents a Taylor series expansion of (12) where the second order term is kept, giving EO123:

$$\Psi_{EO123}(TSA_n) = TSA_{n+1}^2 - TSA_{n+2} \times TSA_n - \frac{1}{3}[TSA_n^2 - TSA_{n+1} \times TSA_{n-1}] + \frac{1}{12}[TSA_n^2 - TSA_{n+2} \times TSA_{n-2}] \quad (15)$$

The pseudo code is given in Appendix D.

3.6. Gear Analysis: Miscellaneous Analysis

The Figure of Merit 0 is a well know analysis derived from Stewart (1977), and is generally calculated as:

$$fm0 = \frac{tsa \text{ peak to peak}}{\sum_{i=1}^3 GM_i} \quad (16)$$

Where *GM* is the gear mesh harmonic taken from the FFT of the TSA. The peak to peak features is a time domain phenomenon (from tooth impact, breathing crack), whereas the gear mesh in calculated with the FFT with is not sensitive to soft toot feature. Hence, as tooth damage increases, so too does FM0.

The energy ratio is the residual RMS divided by the TSA rms. The concept is that as the gear fault progresses, the residual RMS will approach the TSA RMS (e.g. the TSA will be

dominated by gear mesh signatures associated with the gear fault). If the residual signal is defined r_i , and the TSA is tsa_i :

$$er = \frac{\sqrt{\sum_{i=1}^n (r_i - \bar{r})^2}}{n} / \frac{\sqrt{\sum_{i=1}^n (tsa_i - \bar{tsa})^2}}{n} \quad (17)$$

The sideband level factor is defined as the sum of the first order sideband amplitudes about the gear mesh, divided by the TSA rms (Kellar, 2003):

$$SLF = \frac{TSA_{gm-1} + TSA_{gm+1}}{\sqrt{\sum_{i=1}^n (tsa_i - \bar{tsa})^2}} \quad (18)$$

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

4. RESULTS: BANDWIDTH/AM/FM ANALYSIS

The initial study was to evaluate the effect of narrowband analysis bandwidth selection on the separability of the CIs generated by the narrowband, AM and FM analysis. Recall that for each analysis, there are eight condition indicators: RMS, Skewness, Kurtosis, Peak to Peak, Crest, Shape, Impact and Margin factor. The bandwidth selected was based on a percentage of the number of teeth on the damage pinion (which had 23 teeth). The bandwidth was varied from 3 (e.g. bandpass from 23 +/-3) to 9 bins, which represents a bandpass of 13% to 39% of the gear tooth. McFadden (1986) reported that 25% was optimal, and in prior studies (Bechhoefer 2015) found this to be a good value.

As such, the test space for this experiment was 7 x 24 trails. Representative results are given in Table 2, where “Mean All” is the separability for all CIs (24),

Table 2 Bandwidth vs. Separability

BW	NB RMS	NB CF	AM RMS	AM KRT	FM RMS	FM KRT	MEAN ALL
13%	1.7	7.3	5.9	-4	13.2	7.1	5.8
17%	1.92	7.2	6.3	1.6	5.4	6.4	5.7
22%	2	6.5	6.1	2.3	4.6	6.6	5.6
26%	2	6	6.0	2.1	4.3	6.2	5.5
30%	2.1	4.5	6.2	1.9	3.8	5.2	4.4
35%	3	6.6	8.3	-8	3.7	-3	4.1
39%	3.2	7.8	8.3	-3	6.2	-8	5.6

It is clear from these results that the separability is a function of both the CI algorithm/Analysis and the bandwidth. That said, a separability greater than 3 should be considered significant. As there is some risk of over training a result - McFadden’s recommendation of 25% should be considered.

5. RESULTS: ENERGY OPERATOR AND ITS VARIANTS

As the EO and EO123 are numerically similar (essentially a time shaft), it was found that there is no statistical different between them, as based on ANOVA. However, it was found

that the CEEO was statically different and in all case, had greater separability (Table 3)

Table 3 EO and Variant Separability

CI	EO	CEE0	EO123
RMS	0.51	0.75	0.51
SKEW	2.39	2.91	2.06
KURT	1.76	2.09	1.52
P2P	0.58	0.68	0.59
CF	1.49	2.27	1.51
SF	2.1	3.6	1.81
IF	2.0	2.6	1.82
MF	2.0	2.3	1.8

Note that while for this test, the power of the EO and its variants may be low compared to other analysis (less than 3). That said, in other studies looking at scuffing, EO was found to be a good indicator.

6. RESULTS: FACTOR ANALYSIS

The ANOVA analysis indicated that the difference between the factor analysis were statically significant. In table 4, the result for CF, SF, IF and MF are compared with each analysis.

Table 4 Factor Analysis Separability

ANAL	CF	SF	IF	MF
THRES	5.9	5.6	6.2	6.2
RES	5.0	3.2	4.4	4.2
AMRES	2.3	6.5	3.7	3.8
AMRESD	2.4	7.2	4.8	4.9
EO	1.5	2.1	2.0	2.0
CEE0	2.3	3.7	2.6	2.3
EO123	1.5	1.5	1.5	1.5
NB	5.7	6.1	5.8	8.2
AM	3.4	11.3	9.8	9.7
FM	2.1	2.2	2.6	2.6
MEAN	3.2	4.9	4.3	4.54

From the results of Table 4. Shape Factor is the more powerful analysis.

7. RESULTS: ALL ANALYSIS AND ALGORITHMS

The overall test results are presented in Appendix E. In general, these results can be compressed as the performance of similar analysis do not diverge significantly. Table 5 gives the overall results for the Residual, Computational Enhanced Energy Operator, the Narrowband analysis (using a bandwidth of 6, or 26%), Amplitude Modulation, Frequency Modulation and Amplitude Modulated Residual.

Table 5 Mean Separability of Analysis for RMS, Kurtosis and Shape Factor

RES	CEE0	NB	AM	FM	AM RES
-----	------	----	----	----	-----------

SEP	6.4	1.6	4.4	6.7	4.1	3.7
------------	-----	-----	-----	-----	-----	-----

The reduction from 90 values (Appendix E) to six value in Table 5 was: that the analysis suggested that skewness and peak to peak were in general not powerful statistics. Similarly, from Table 4, the shape factor was found to be the highest performance factor analysis. Hence, Table 5 represents the mean separability response for RMS, Kurtosis and Shape Factor.

From Figure 2, the relationship between the classes of residual analysis are compared. Clearly, the fault at 0.4 revolution is seen. While the amplitude modulation residual is easily seen as a large change in subplot 3, its performance from Appendix E suggest that it is more sensitive to noise, hence the lower separability.

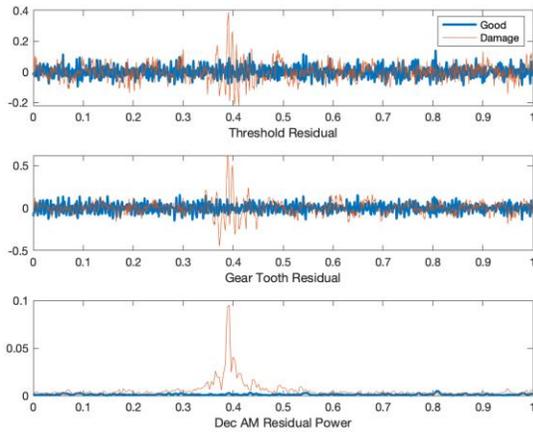


Figure 2 Comparison of Residual based analysis for good to bad pinion

As noted in Table 3, the performance of the energy operator and its variants are all similar. Again, it is clear that the lower separability score that the EO, CEEO and EO123 are typically a noisier analysis (Figure 3).

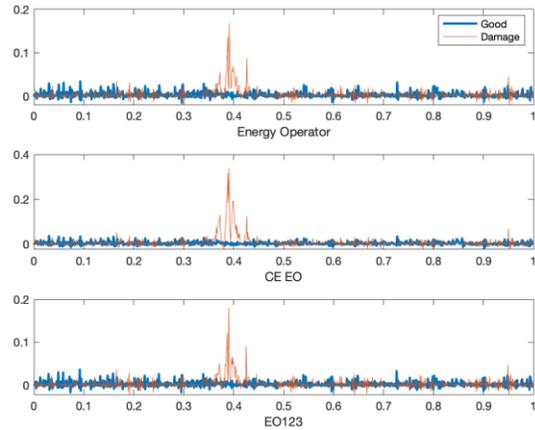


Figure 3 Comparison of EO and its variants for good and bad pinion

The narrowband, amplitude modulation and frequency modulation analysis are consistently powerful. While it is not as easy to visually identify the gear fault as the residual and energy operator, the reduced bandwidth (e.g. noise) of the analysis greatly improve the separability score (Figure 4).

It should be noted that these test results were performed on machines running at similar, constant RPM. As acceleration is the second derivative of displacement, the sensor response is a function of RPM, squared. Hence amplitude-based statistics are more sensitive to changes in RPM, than say, shape-based statistics. For example, if the machine is to operate under varying RPM condition, condition indicators based on kurtosis would be less effected by RPM than RMS, or Peak to Peak.

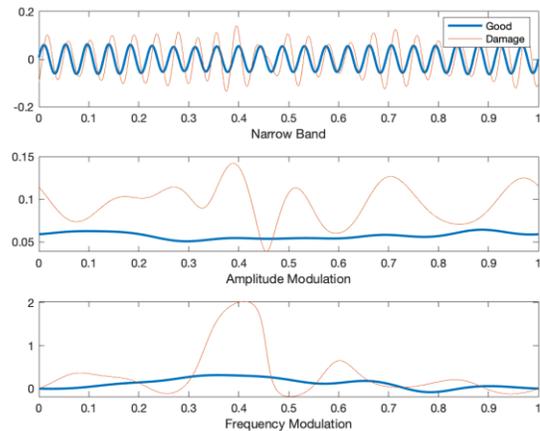


Figure 4 Comparison of the narrowband, amplitude and frequency modulation analysis for good and bad pinion

7.1. Validation of Condition Indicator Selection

Data was collected on a single stage, spiral bevel gearbox, which was run under high load. The instrumentation was

installed on a nominal gearbox, with the intent of propagating a fault. After approximately 24 days, the spiral bevel pinion showed indications of fatigue cracking at the root of a tooth. For the test, the health algorithm (eq 3), the gear CI used were index: Residual Kurtosis, Residual Crest Factor, Energy Operator Kurtosis, Figure of Merit 0, Amplitude Modulation RMS and Frequency Modulation RMS (index 2, 3, 5, 7 and 12, Figure 5).

These CIs based on a study (Bechhoefer 2011), with a consideration of attempting of algorithm diversity. That is, it was hypothesized that because there is no single CI that works for every gear fault, it would be wise to use at least one CI from each type of analysis. It should also be noted that for (ref 12), the damage was scuffing, not a cracked tooth.

Because the of the result (Appendix E) indicated that the Residual RMS, Narrowband Kurtosis, FM Kurtosis performance (e.g. separability) was improved, and that FMO did not perform well, the CIs for the health algorithm were changed to Residual RMS, Energy Operator Kurtosis, Narrowband Kurtosis, Amplitude Modulation and Frequency Modulation Kurtosis, or index: 1, 5, 8, 10, 12, and 15, Figure 5. Note that the final HI value changed from approximately 0.45 to 1.5, or almost 3x improvement in fault detection. It is interesting to note that the separability of Residual RMS (10.18) is only 41% improved over the Residual CF (5.94). The energy operator, was not a particularly powerful (the separability was 1.55), was left in because it may improve the HI performance with other gear fault modes.

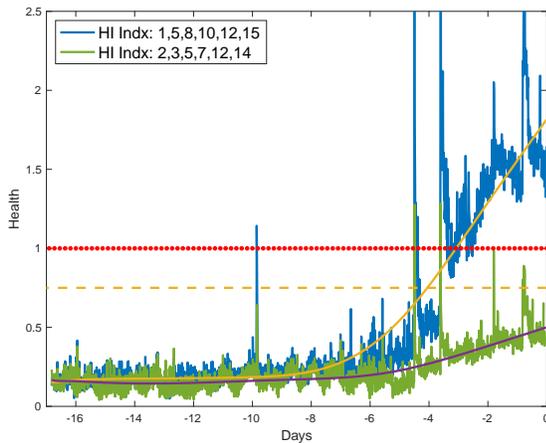


Figure 5 Comparison of the gear health performance based on CI selection

One question to ask would be: why not use a greater number of the CIs with positive separability? This was tried, but after normalizing, there was no increase in performance. That is, the HI using 15 CI indexes did have a greater HI value. However, its starting (nominal HI) value was 0.275 vs. 0.174. When the new HI was scaled by 0.174/0.275 or 0.64, there was no observable change in the HI. It is suspected that an orthogonal transformation, such as principal component

analysis, would indicate that using 15 CI was over dimensioned, as using both Residual RMS, Kurtosis and CF (as an example) are highly correlated (Figure 6).

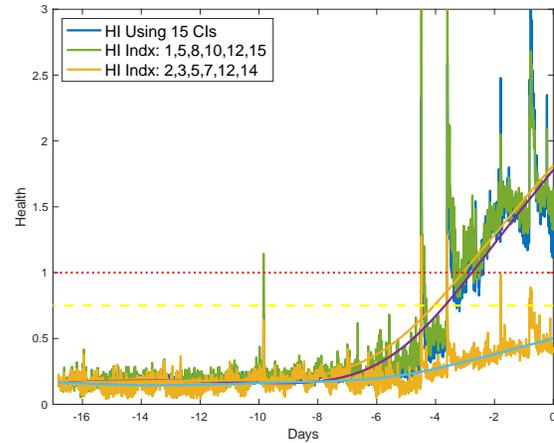


Figure 6 Comparison of HIs using 15 CIs vs. 6 CIs

8. CONCLUSION

Few studies have asked the question: how good is a particular gear analysis algorithm? Generally, the analysis is given on a small set of data, with no comparison to other analysis. This study, using a statistically significant data set, compared the performance of gear CI analysis between a good and damage population. The performance metric used was statistical separability. Separability is a natural metric as condition monitoring is in effect a decision support tool (e.g. hypothesis testing), the decision being to bring an asset out of service for inspection, or not. CI algorithms with a large separability, will reduce the probability of a false alarm.

For a decision support tool, this study uses the concept of the health indicator (HI). The HI is a mapping of n CIs into a common, threshold value which allow the maintainer to make decision as to when to perform maintenance. When the component HI is greater than 1, then it is appropriate to do maintenance. Using the CIs with the largest separability increased the sensitivity of the HI by 3x, or a 4.5 dB improvement. Further, it was shown that using the HI (eq 3) can give a 3 to 4.5 dB gain over using CIs alone.

Caution should be taken in that, while the validation (e.g. HI improvement) was gained using a different gearbox type than the original data set used for measuring separability (Table 2), the fault mode was similar: a fatigue crack on a tooth. It is likely that for scuffing/pitting type of faults, the optimal set of CIs would be different.

REFERENCES

- 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 July.
- Bechhoefer, E., Kingsley, M., "A Review of Time Synchronous Average Algorithms", The Annual Conference of the Prognostics Health Management Society, 2009.
- McFadden, P.D., (1986), Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration, *J. Vib. Acoust. Stress Reliab. Design*, vol 10, no 2, Apr. pp 165-170.
- Ma, J., (1995). Energy Operator and Other Demodulation Approaches to Gear Defect Detection. Proceeding of the MFPT 49, pp. 127-140.
- Zakrajsek, J., "An Investigation of Gear Mesh Failure Prediction Techniques", NASA Technical Memorandum 102340, 1989.
- Zakrajsek, J., Decker, H., "An Analysis of Gear Fault Detection Methods as Applied to Pitting Fatigue Failure Data", NASA Technical Memorandum, 105950, 1993.
- Lewicki, D., Dempsey, P., Heath, G and Shanthakumaran, P., "Gear Fault Detection Effectiveness as Applied to Tooth Surface Pitting Fatigue Damage" NASA/TM-2009-215667, 2009.
- Gopinath, K, Mayuran., M., Machine Design II, Gear Faults, Indian Institute of Technology Madras, <https://nptel.ac.in/courses/112106137/9>, 2010.
- Mendenhall, W., Wackerly, D., Scheaffer, R., Mathematical Statistics with Applications, Boston, PWS-Kent Publishing, 1990, pp 457
- Bechhoefer, E., He, D., Dempsey, P., "Gear Health Threshold Setting Based On a Probability of False Alarm," Annual Conference of the Prognostics and Health Management Society, 2011.
- Kellar, J., Grabill, P., "Vibration Monitoring of a HU-60A Main Transmission Planetary Carrier Fault", American Helicopter Society 59th Annual Forum, Phoenix, AZ, 2003
- Bechhoefer, E., He, D., Dempsey, P., "Gear Health Threshold Setting Based On a Probability of False Alarm", Annual Conference of the Prognostics and Health Management Society, 2011
- Bechhoefer E., Bernhard, A., "Use of Non-Gaussian Distribution for Analysis of Shaft Components", IEEE Aerospace Conference, 2005.
- Caesarendra, W., Tjahjowidodo T., "A Review of Feature Extraction Methods in Vibration-Based Monitoring and its Application for Degradation Trend Estimation of Low-Speed Slew Bearing", *Machine* 2017, 5, 21.
- Mayo, E., Liang, M., Baddour, Na., "Gear Fault Detection with the Energy Operator and its Variants", ASME 2016 International Design Engineering Technical Conference and Computes and Information in Engineering Conference, Aug 21-24, 2016.
- Bechhoefer, E., Zhang, X., Improved Fault Detection by Appropriate Control of Signal Bandwidth of the TSA,

Annual Conference of the Prognostics and Health Management Society, 2015

- Wang, W., "Early Detection of gear Tooth Cracking Using the Resonance Demodulation Technique", *Mechanical Systems and Signal Processing* (2001), 15(5), 887-903
- McFadden, P.D., "Examination of a Techniques 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, (1987)

BIOGRAPHIES

Eric Bechhoefer received his B.S. in Biology from the University of Michigan, his M.S. in Operations Research from the Naval Postgraduate School, and a Ph.D. in General Engineering from Kennedy Western University. His 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.

Brent Butterworth has been in Aviation Rotorcraft Maintenance and Engineering for 25+ years. He is a veteran of 10 years in the USMC and USAF. He held various leadership roles within Airbus Helicopters Technical Support and Engineering group. He is currently a Systems Engineer at Garmin in the G5000H architectures. He received his B.S. in Professional Aeronautics from Embry Riddle Aeronautical. He holds an FAA A&P license, FCC GROL with Radar Endorsement and Amateur Radio License. He is currently pursuing a B.S. in Data Management and Analytics.

APPENDIX A

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

zctl = 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 - zctl; %time of 1 rev
    dtic = dtrev*iN; %time between each sample
    zctlc = zctl;
    for j = 1:N
        cidx = floor(zctlc*sr);
        yo = data(cidx); y1 = data(cidx+1);
        x1 = zctlc*sr;
        xo = floor(x1);
        dx = x1-xo;
        dy = y1-yo;
        yaj = yo + dx*dy; %simple linear interp
        ya(j) = yaj;
        zctlc = zctl + j*dtic; %increment to the
next sample
    end

    zctl = zcti;

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

tsadata = yy/navgs; % compute the average

```

APPENDIX B

```

function [xres] = residualSignal(x, geartooth)
%[xres] = residualSignal(x, geartooth)
%Inputs:
% x :input TSA signal
%geartooth :array with number of teeth on a gear
%from Vercher
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 mesh
            are removed
        end
    end
end

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

```

APPENDIX C

```

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

```

APPENDIX D

```

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;
    
```

APPENDIX E

Overall Separability Analysis Results

Analysis	Sep	Analysis	Sep
AM P2P	15.564	CEEO IF	2.604
SBLF	14.938	FM IF	2.592
AM SF	11.35	FM MF	2.586
RES RMS	10.185	amRes Krt	2.498
AM IF	9.861	amResd CF	2.432
AM MF	9.68	EO Skew	2.394
TRES RMS	8.982	CEEO MF	2.336
TRES P2P	8.487	amRes CF	2.275
NB MF	8.237	CEEO CF	2.273
amResd SF	7.207	FM SF	2.199
amRes SF	6.417	CEEO Krt	2.089
TRES MF	6.202	FM CF	2.088
TRES IF	6.188	EO SF	2.08

NB SF	6.109	EO123 SF	2.062
FM Krt	5.966	EO IF	2.025
TRES CF	5.939	EO MF	2.008
FM P2P	5.926	NB RMS	1.978
AM RMS	5.914	EO123 IF	1.818
NB IF	5.791	EO123 SF	1.815
NB CF	5.747	EO123 MF	1.803
TRES SF	5.59	EO Krt	1.759
FM Skw	5.49	RES P2P	1.698
S Index	5.222	Gs RMS	1.643
NB Krt	5.053	EO123 Krt	1.527
RES CF	4.997	EO123 CF	1.514
amResd MF	4.865	TRES Skw	1.503
SBLF - 2	4.827	EO CF	1.492
amResd IF	4.804	FM0	1.348
TRES Krt	4.721	amResdP2P	1.218
amResdSkw	4.623	Gear Mesh	1.149
RES IF	4.401	amRes P2P	1.038
RES MF	4.193	amResdRMS	0.976
amRes Skw	4.188	amRes RMS	0.87
FM RMS	4	CEEO P2P	0.68
amRes MF	3.805	GM2/GM1	0.642
amRes IF	3.668	EO123 P2P	0.594
CEEO SF	3.661	EO P2P	0.583
AM CF	3.373	CEEO RMS	0.571
RES SF	3.179	EO RMS	0.509
SBLF	3.168	EO123 RMS	0.508
NB P2P	3.15	Eng Ops	0.373