

Evaluation of Gear Condition Indicator Performance on Rotorcraft Fleet

Lance J. Antolick and Jeremy S. Branning
RMC Inc.

U.S. Army Aviation Engineering Directorate
Redstone Arsenal, Alabama

lance.antolick@us.army.mil and jeremy.branning@us.army.mil

Daniel R. Wade

U.S. Army Aviation Engineering Directorate
Redstone Arsenal, Alabama

Paula J. Dempsey

National Aeronautics and Space Administration
Glenn Research Center

Cleveland, Ohio

ABSTRACT

The U.S. Army is currently expanding its fleet of Health Usage Monitoring Systems (HUMS) equipped aircraft at significant rates, to now include over 1,000 rotorcraft. Two different on-board HUMS, the Honeywell Modern Signal Processing Unit (MSPU) and the Goodrich Integrated Vehicle Health Management System (IVHMS), are collecting vibration health data on aircraft that include the Apache, Blackhawk, Chinook, and Kiowa Warrior.

The objective of this paper is to recommend the most effective gear condition indicators for fleet use based on both a theoretical foundation and field data. Gear diagnostics with better performance will be recommended based on both a theoretical foundation and results of in-fleet use. In order to evaluate the gear condition indicator performance on rotorcraft fleets, results of more than five years of health monitoring for gear faults in the entire HUMS equipped Army helicopter fleet will be presented. More than ten examples of gear faults indicated by the gear CI have been compiled and each reviewed for accuracy. False alarms indications will also be discussed. Performance data from test rigs and seeded fault tests will also be presented.

The results of the fleet analysis will be discussed, and a performance metric assigned to each of the competing algorithms. Gear fault diagnostic algorithms that are compliant with ADS-79A will be recommended for future use and development.

The performance of gear algorithms used in the commercial units and the effectiveness of the gear CI as a fault identifier will be assessed using the criteria outlined in the standards in ADS-79A-HDBK, an Army handbook that outlines the conversion from Reliability Centered Maintenance to the *On-Condition* status of Condition Based Maintenance.

INTRODUCTION

To date, the U.S. Army Condition Based Maintenance (CBM) program has equipped approximately 1,000 H-64, H-60, and H-47 series aircraft with HUMS, referred to as digital source collectors (DSCs), and suites of accelerometers. All drive train gear sets, including fixed-axis and epicyclic, are monitored by these systems. A variety of signal processing algorithms (Condition Indicators or CIs) have been developed to detect incipient faults in the gear sets by various government agencies and commercial vendors.

The traditional methods of calculating gear vibration based condition indicators use a statistical measurement of vibration energy. Differences in condition indicators are based on which of the characteristic frequencies are included, excluded, or used as a reference. Many of these different condition indicators have been evaluated experimentally by the US Navy and NASA using gearbox test stands and gear seeded faults. The various gear

condition indicators used on Army aircraft vary by model and are based on previous test stand successful detections. These test stands have been the only source of ground proof data that directly associate faulted gears to changes in CIs.

The Army has amassed nearly a decade of DSC CI data that includes thousands of hours of healthy and faulted gear vibration signatures. The vibration data and CI trends have been cross-referenced with historical information about the removal of gearboxes and their associated tear down analyses. As part of the CBM program, it is desired to continuously improve the functionality of the gear monitoring functions in the DSCs. The Aviation Engineering Directorate therefore is tasked by the Army to review the effectiveness of gear CIs according to Aeronautical Design Standard Handbook (ADS-79A-HDBK) for Condition Based Maintenance Systems for US Army Aircraft Systems. The ADS-79A prescribes required detection rates for CIs and other vibration health measurements in order to extend gear life and meet CBM objectives. The ADS-79A refers to True

Positives (correctly identifying a fault), True Negatives (correctly identifying a no fault condition), False Positives (incorrectly identifying a no fault condition), and False Negatives (incorrectly identifying a fault). FPs and FNs impact the airworthiness of the CBM system and therefore must be minimized.

This paper is a compilation of gear failures that have been recorded by two Army installed DSCs, the Modern Signal Processing Unit (MSPU) and the Integrated Vehicle Health Monitoring System (IVHMS). The former is installed on the Apache and the latter is installed on the Blackhawk. Algorithms that are automatically calculated by the on board systems are presented, as are algorithms that have been developed more recently. These additional algorithms have been *post processed* using the recorded data. Special attention has been paid to an algorithm, used for gear diagnostics, known widely as FM4, an algorithm which has

8 or more variants. The next section will define the FM4 variants that are most widely used on Army aircraft and test stands.

THEORETICAL EXPLANATION OF FM4

The definition of FM4 has changed since it was first developed and since then has taken on many different incarnations. Of course, the Army has two distinct varieties as installed in the IVHMS and MSPU. In the IVHMS, the calculation that is most similar to the traditional FM4 is named Residual Kurtosis. In the MSPU, the calculation that is most similar to the traditional FM4 is named FM4. For the purposes of this paper, the traditional definition of FM4 will be taken as follows in Table 1. Per the table, FM4 is calculated from time synchronous averages (TSA) and filtering is performed in the frequency domain on the time synchronous averaged signals.

Table 1: FM4 Condition Indicator Calculation (Ref. 1)

CI	Filter	Formula	Numerator	Denominator	Nominal Value	Threshold Value
FM4	Remove gear mesh & harmonics, 1 st order side bands, shaft 1/rev, 2/rev	$\frac{\frac{1}{N} \sum_{n=1}^N (d_n - \bar{d})^4}{\left[\frac{1}{N} \sum_{n=1}^N (d_n - \bar{d})^2 \right]^2}$	4 th moment about mean of difference signal	square of variance of difference signal	3	5-7

FM4 is the kurtosis of the residual or difference signal normalized by the square of the variance (Ref. 2) and kurtosis is the fourth statistical moment of a signal about its mean (Ref. 3). The residual or difference signal is most commonly obtained by taking the Fast Fourier Transformation (FFT) of the signal average and subtracting the gear meshing frequency (GMF), its harmonics, and first order shaft-modulated sidebands around the fundamental GMF, then performing the inverse FFT to put the signal back in the time domain (Refs. 2, 3, and 4). Variations of this general method may involve removing different numbers of harmonics, or including more than one pair of sidebands around each harmonic (Ref. 4).

Another method of obtaining the residual signal is performed by converting the signal average into the frequency domain and removing all frequency content that is above a pre-defined value, then converting back to the time domain. The value used for removing frequency content is relative to the energy of the signal, set statistically. An example might be two standard deviations above the average in the frequency domain.

The purpose of taking the kurtosis of any signal is to identify the presence of peaks. When normalized by the squared variance, the kurtosis becomes a measure of both the number and amplitude of the peaks (Ref. 6). The normalized kurtosis of a sine wave is 1.5, while a normal

distribution is 3.0 (Ref. 7). Normalized kurtosis values greater than 3.0 indicate peaks in the distribution. FM4 applies the normalized kurtosis to the residual signal, where peaks would indicate an irregularity with the gear, since all expected frequencies have been removed. FM4 should rise with the presence of a local fault, but fall as the fault distributes because the rising energy of the signal means the normalization factor, the squared variance, is increasing.

In both the MSPU and IVHMS, the built in FM4 algorithms are trusted indicators of gear health and the traditional threshold is typically between 5 and 7. In fact, FM4 has given good results on test stands when single tooth faults are introduced (Ref. 5)

The Army has experienced a very high FP rate with the use of FM4 on both the Blackhawk and Apache. FM4 has demonstrated that it is sensitive to unusual mesh patterns discovered during gearbox tear down analysis. These gearboxes were then subject to engineering scrutiny and it was determined that the perceived problem did not impact the airworthiness of the gearbox.

While a high FP rate does not impact the airworthiness of the CBM system, it does reduce crew members' trust of the DSC readings and negatively impact any cost savings associated with the installation of a DSC. The goal of CBM is to reduce maintenance time and cost, so FPs must be avoided.

The following sections therefore detail the path forward for Army gear diagnostics. The authors will show the FM4 variant diagnostics that are capable of identifying TPs and TNs and are correctly implemented such that they do not result in any FPs or FNs. The data used in these sections is taken directly from the Army database of *on aircraft ground truth data*.

The authors will further show the gear diagnostics other than FM4 that have successfully identified, based on the standards set forth in the ADS-79A, the condition of gears removed from Apache nose gearboxes. It is important to note here that there have not been any monitored/recorded Blackhawk gear failures and therefore IVHMS algorithms were evaluated using the Apache data.

METHODOLOGY AND RESULTS

FPs erode confidence in the DSC and increase maintenance time and costs. A number of FPs have been identified from both the Blackhawk and Apache fleets.

These FP cases were used to analyze known variants of the FM4 algorithm to determine if and how the variants perform differently.

The FM4 variants differ both in the way that the difference signal is calculated and in the way that the difference signal is reconstructed after the desired tones are removed. Table 2 illustrates the naming convention and techniques used to calculate FM4 variants.

Time synchronous averages were used to calculate all known variants of FM4 for confirmed FP cases. Figure 1 shows an example FP from aircraft 07-05514. The variants behave similarly, with the variant that removes all tones greater than the mean of the time synchronous averages plus two standard deviations (known as Residual Kurtosis) performing slightly better. Table A1 in the appendix summarizes how the variants perform for all the FP cases. Although this algorithm performs significantly better than the others, it does not comply with ADS-79A requirements, because of its 69% false positive rate for gear components monitored on 8 different helicopters.

Table 2. FM4 Variants and Definitions

CI	Filter	Difference Signal Reconstruction Method
FM4 Trad. Sine	Remove gear mesh & harmonics, 1 st order side bands, shaft 1/rev, 2/rev	Sine wave interpolation
FM4 Trad. FFT	Remove gear mesh & harmonics, 1 st order side bands, shaft 1/rev, 2/rev	Inverse FFT
FM4 >2STD (Residual Kurtosis)	Remove all tones greater than the mean(FFT(TSA)) plus 2*standard deviation(abs(FFT(TSA)))	Inverse FFT
FM4 CoIn. Rem.	Remove gear mesh & harmonics, 1 st order side bands from both gear of interest and attached gears, shaft 1/rev, 2/rev	Sine wave interpolation

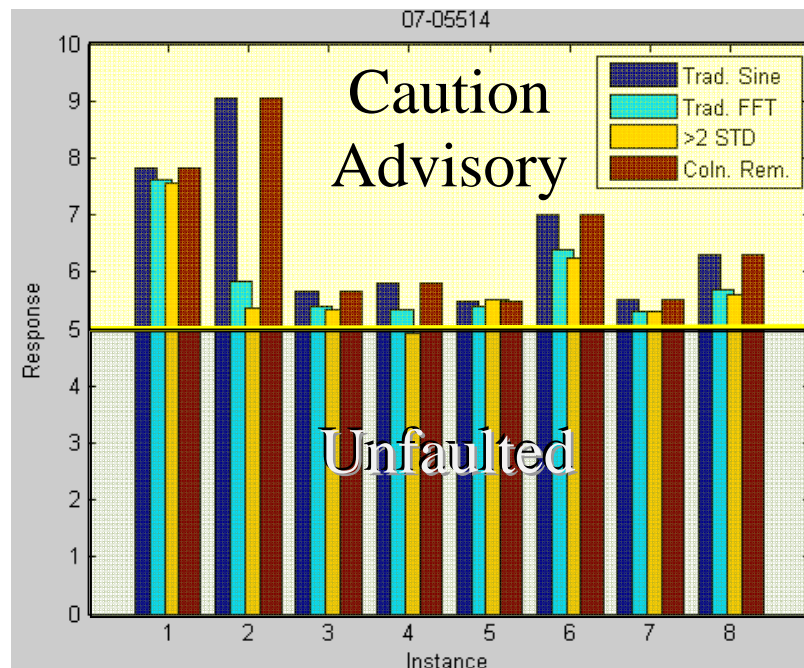


Figure 1. FM4 Variants calculated for FP Cases from 07-05514.

The next step taken was to evaluate the effectiveness of all FM4 variants from confirmed damaged gears. Unfortunately, only Apache nose gearboxes populate the Army's list of confirmed gear damage cases through tear down analysis. Twelve separate nose gearboxes (NGB) were identified and included in the analysis. The damage sustained by the gears varies from gearbox to gearbox, but the typical failure modes are spalling, pitting, and excessive wear, examples of which are shown in Figures 2 and 3. In addition to the 12 helicopter gearboxes, data from damaged gears recorded from the University of South Carolina Tail Rotor Drive Train test stand and the NASA Glenn Spur Gear Fatigue Rig is also included in the table.

The FM4 variants were calculated for all the faulted cases and the results are summarized in Table 3. At the end of the table, the ADS-79A *batting average* is calculated to show overall on aircraft effectiveness of the CI. Batting average is the total number of TNs and TPs divided by the number of data points. Any N/A results shown in this section indicate that there was insufficient data to calculate the CI. A *Y* indicates that the CI responded to the gear damage and likewise *N* indicates no response. All FM4 variants performed poorly.

Four other CIs were calculated using the gear fault data. The FM4 algorithm derivatives are FM4*, NA4, NA4*, and NA4reset. These algorithms were developed to correct the shortcomings of FM4 and are defined in Table 4. Their results are displayed in Table 5. FM4* and NA4 performed the best with a true positive rate of 67%.



Figure 2. Example of typical gear damage found in the Apache nose gearbox.



Figure 3. Additional examples of typical gear damage found in the Apache nose gearbox.

Table 3. Summary of faulted gear cases and FM4 variant calculations

Tail	Component	FM4 Sine	FM4 FFT	FM4 >2STD	FM4 CoIn Rem.
01-05277	#1NGB	Y	Y	N	Y
03-05363	#1NGB	N	N	N	N
07-05508	#1NGB	N	N	N	N
96-05010	#1NGB	N	N	N	N
99-05144	#1NGB	N	N	N	N
00-05187	#2NGB	N	N	N	N
02-05303	#2NGB	Y	Y	N	Y
00-05195	#2NGB	N	N	N	N
00-05206	#2NGB	N/A	N/A	N/A	N/A
00-05208	#2NGB	N	N	N	N
02-05331	#2NGB	N	N	N	N
90-00337	#2NGB	N	N	N	N
USC Test Stand	TGB	N	N	N	N
NASA Test Stand	Bevel Gear	N/A	N/A	N/A	N
Batting Average		0.167	0.167	0	0.153

Table 4: CI Calculations

CI	Filter	Formula	Numerator	Denominator
NA4	Remove gear mesh and harmonics, shaft 1/rev, 2/rev	$\frac{\frac{1}{N} \sum_{n=1}^N (r_n - \bar{r})^4}{\left[\frac{1}{M} \sum_{m=1}^M \frac{1}{N} \sum_{n=1}^N (r_{n,m} - \bar{r}_m)^2 \right]^2}$	4 th moment about mean of residual signal	Square of average variance of all residual signals up to current time (running average)
NA4*	Same as NA4	Same as NA4	Same as NA4	Each new den. is tested via a hypothesis test. If it falls within the limit of a healthy dataset it is included in the den.
NA4 Reset	Same as NA4	Same as NA4	Same as NA4	Reset when load varies x%
FM4*	Same as FM4	Same as FM4 except running average in denominator	Same as FM4	Each new den. is tested via a hypothesis test. If it falls within the limit of a healthy dataset it is included in the den.

Table 5. Summary of effectiveness of FM4 derivatives calculated from faulted gear cases

Tail	Component	FM4*	NA4	NA4*	NA4 Reset
01-05277	#1NGB	N/A	Y	N/A	Y
03-05363	#1NGB	Y	N	N	N
07-05508	#1NGB	Y	Y	Y	N
96-05010	#1NGB	N/A	Y	N/A	N
99-05144	#1NGB	N/A	N	N/A	N
00-05187	#2NGB	Y	Y	Y	N
02-05303	#2NGB	N/A	Y	N/A	Y
00-05195	#2NGB	N	N	N	N
00-05206	#2NGB	N/A	N/A	N/A	N/A
00-05208	#2NGB	N	N	N	N
02-05331	#2NGB	N/A	Y	N/A	N
90-00337	#2NGB	N/A	Y	N/A	N
USC Test Stand	TGB	Y	Y	Y	N/A
NASA Test Stand	Bevel Gear	N/A	N/A	N/A	Y
Batting Average		0.667	0.667	0.5	0.25

Since the Army database of on-aircraft faulted gear cases only included Apache nose gearboxes, it was deemed necessary to compare results to other types of gears. The NASA Spiral Bevel Gear Fatigue rig and the University of South Carolina AH-64 Tail Rotor Test Stand were identified as having the gear failure results necessary to compare to the on-aircraft results. Several plots of different CIs calculated for faulted cases on the test stand are shown in Appendix A.

FM4 is currently used as a primary condition indicator for gear health on US Army HUMS-equipped helicopters. Because of FM4's unacceptable performance, replacements

for it must be identified. The gear damage cases that were used to evaluate the effectiveness of FM4 were used this time to identify the condition indicators that did react to the gear damage. Sideband Index (SI), Data Algorithm one (DA1), and Sideband Level Factor (SLF) responded well to the gear damage shown above. These algorithms are defined in Table 6. Additional details can be found in Appendix B. The case by case results of the condition indicators are shown in Table 7. SI performed very well with a true positive rate of 93%.

Table 6: CI Calculations

CI	Description	Formula
SI	The average of the first order sidebands of the fundamental gear meshing frequency	$SI = \frac{R_{l,-1}^{esb}(x) + R_{l,+1}^{esb}(x)}{2}$
SLF	Sum of the first order sideband amplitudes of the fundamental gear meshing frequency normalized by the RMS of the synchronous time average	$SLF = \frac{R_{l,-1}^{esb}(x) + R_{l,+1}^{esb}(x)}{RMS(x)}$
DA1	RMS of the STA subtracted from the mean of the synchronous time average.	$DA1 = RMS(STA - \overline{STA})$

Table 7. Results of best performing CIs for damaged gears.

Tail	Component	DA1	SLF	SI
01-05277	#1NGB	Y	N	Y
03-05363	#1NGB	N	Y	Y
07-05508	#1NGB	Y	Y	Y
96-05010	#1NGB	Y	N	Y
99-05144	#1NGB	Y	N	Y
00-05187	#2NGB	Y	N	Y
02-05303	#2NGB	Y	N	Y
00-05195	#2NGB	N/A	N	Y
00-05206	#2NGB	Y	N	Y
00-05208	#2NGB	N	Y	N
02-05331	#2NGB	N	Y	Y
90-00337	#2NGB	Y	Y	Y
USC Test Stand	TGB	Y	N	Y
NASA Test Stand	Bevel Gear	Y	N	Y
Batting Average		0.769	0.357	0.929

SUMMARY

False Positives

Blackhawk and Apache helicopter FM4 FP cases were gathered. All known FM4 algorithms were calculated for each FP case obtained. There were four separate FM4 algorithms calculated. The purpose of this analysis was to determine if one of the algorithms reduced the number of false positives. The results have shown that one algorithm produced slightly fewer FPs, but did not reduce FPs to a rate acceptable according to the ADS-79A. To calculate the comprehensive ineffectiveness of FM4, its TP rate was analyzed.

FM4 Derivatives and Faulted Cases

Some of the algorithms that were created to correct the shortcomings of FM4 had success in identifying the gear faults. Notably FM4* and NA4 reacted to gear damage in over half of the applicable cases. However, these algorithms present difficulties regarding implementation in the Army fleet. FM4* requires 30 previous TSAs to calculate its denominator. This poses problems for use in a maintenance environment where components are continuously being changed and moved between aircraft. There will not be a CI value for FM4* for newly equipped helicopters until at least 30 data points are collected, and the on-board system does not know when a gearbox is replaced. Furthermore, a new

nominal value would need to be calculated for a *good* gearbox but there is no way of knowing that information currently. FM4* requires either data to be stored on the on-board system to calculate the new CI value or it would have to be calculated on a ground-based station. If the latter option is selected, FM4* cannot be used for *real-time* in-flight diagnostics. NA4 poses identical problems. Since its denominator calculates the average variance of all previous TSAs, all of the above is also applicable to NA4. Also, NA4 is known to be very sensitive to changes in torque (Ref. 8). Therefore it would only be feasible to calculate FM4* and NA4 on a ground-based system.

Test Stands

Comparative test stand results were found to be similar to that of the on-aircraft results. The FM4 algorithms, as defined by several sources, do not comply with the ADS-79A detection requirements or FP rates. FM4*, NA4, and NA4* responds to the gear damage simulated on the USC test stand, and NA4 reset responds well to the NASA test stand gear damage.

SI, SLF, and DA1

These gear diagnostic algorithms responded particularly well to the gear damage, notably SI and DA1. These two algorithms also responded to the test stand cases. Results can be seen in the Appendix A.

CONCLUSIONS

No variant of FM4 consistently diagnoses gear faults on Army aircraft. While FM4* and NA4 are usually responsive to faults, they are incompatible with ADS-79A because of their burdensome data requirements, false positives, and the need for more detailed component tracking. All cases presented are spiral bevel gear sets, which comprise most but not all critical drivetrain gear sets in Army aircraft. FM4

may still be acceptable for use on other gear types, where the contact ratio is significantly different than on spiral bevel gear sets.

Three algorithms have proven reliable on most spiral bevel gear damage cases examined: Sideband Index, Sideband Level Factor, and DA1. These algorithms warrant further research and enhancement.

APPENDIX A: ADDITIONAL RESULTS

Table A1. Results of FM4 Variants Calculated from FP Cases

Tail	Component	Number of Instances	FM4 Sine	FM4 FFT	FM4 >2STD	FM4 CoIn Rem.
00-05189	IGB	1	1	1	1	1
02-05302	IGB	11	10	11	9	10
02-05307	TGB	12	12	12	9	12
04-05425	TGB	15	15	15	13	15
07-05514	TGB	8	8	8	7	8
02-05329	TGB	28	28	28	9	28
04-05466	TGB	18	17	18	15	17
H60 585	MBG	3	3	3	3	3
Totals		96	94	96	66	94
Percent of Total			98%	100%	69%	98%

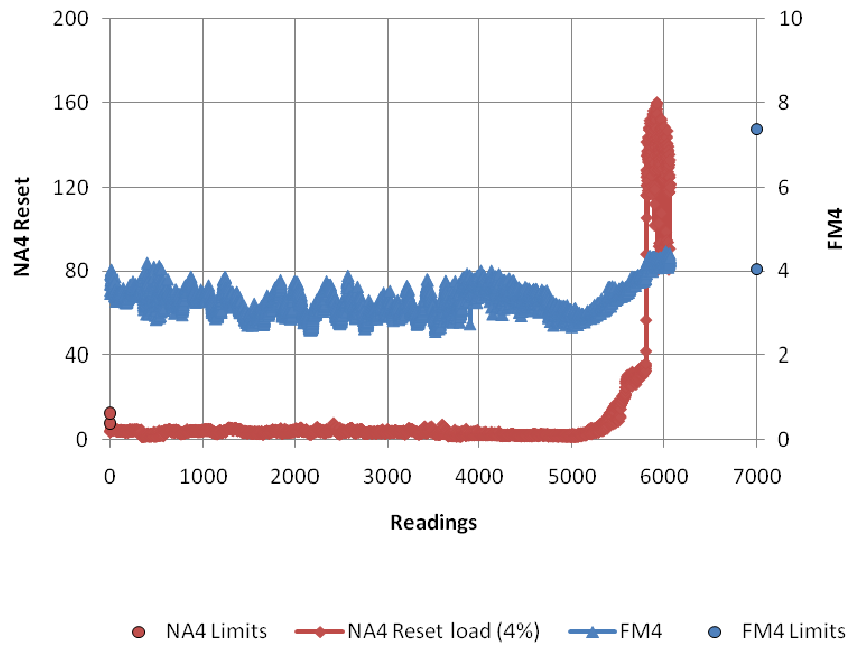


Figure A1: NASA Spiral Bevel Gear Rig NA4 reset and FM4 Response to Gear Damage.

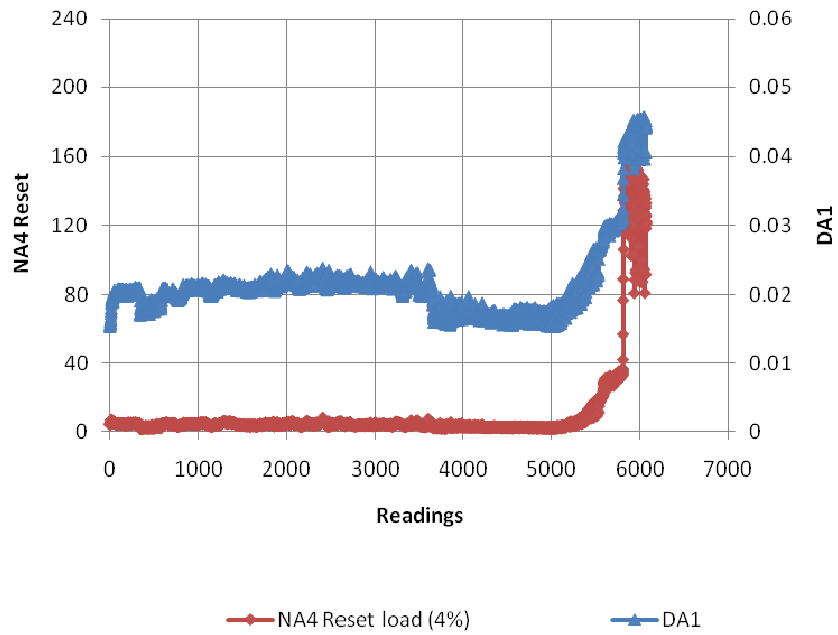


Figure A2: NASA Spiral Bevel Gear Rig NA4 reset and DA1 Response to Gear Damage.

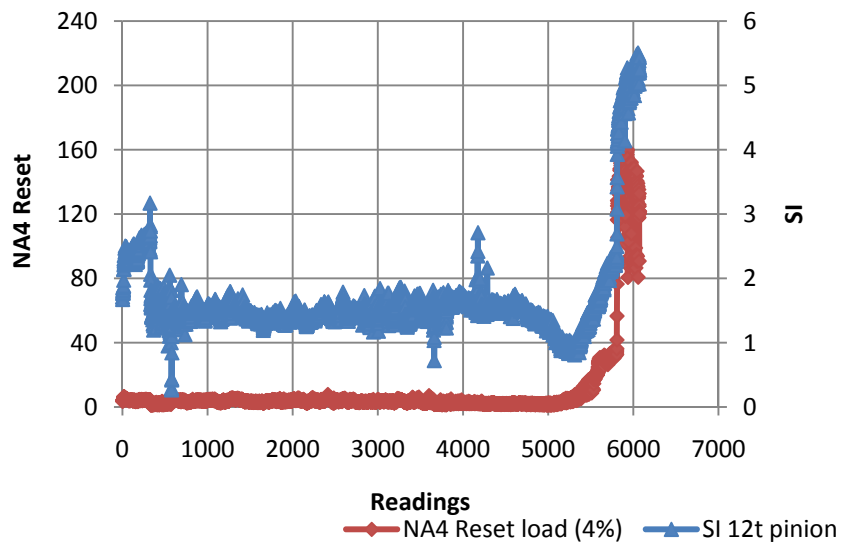


Figure A3: NASA Spiral Bevel Gear Rig NA4 reset and SI Response to Gear Damage.



Figure A4: NASA Spiral Bevel Gear Rig Gear Damage.

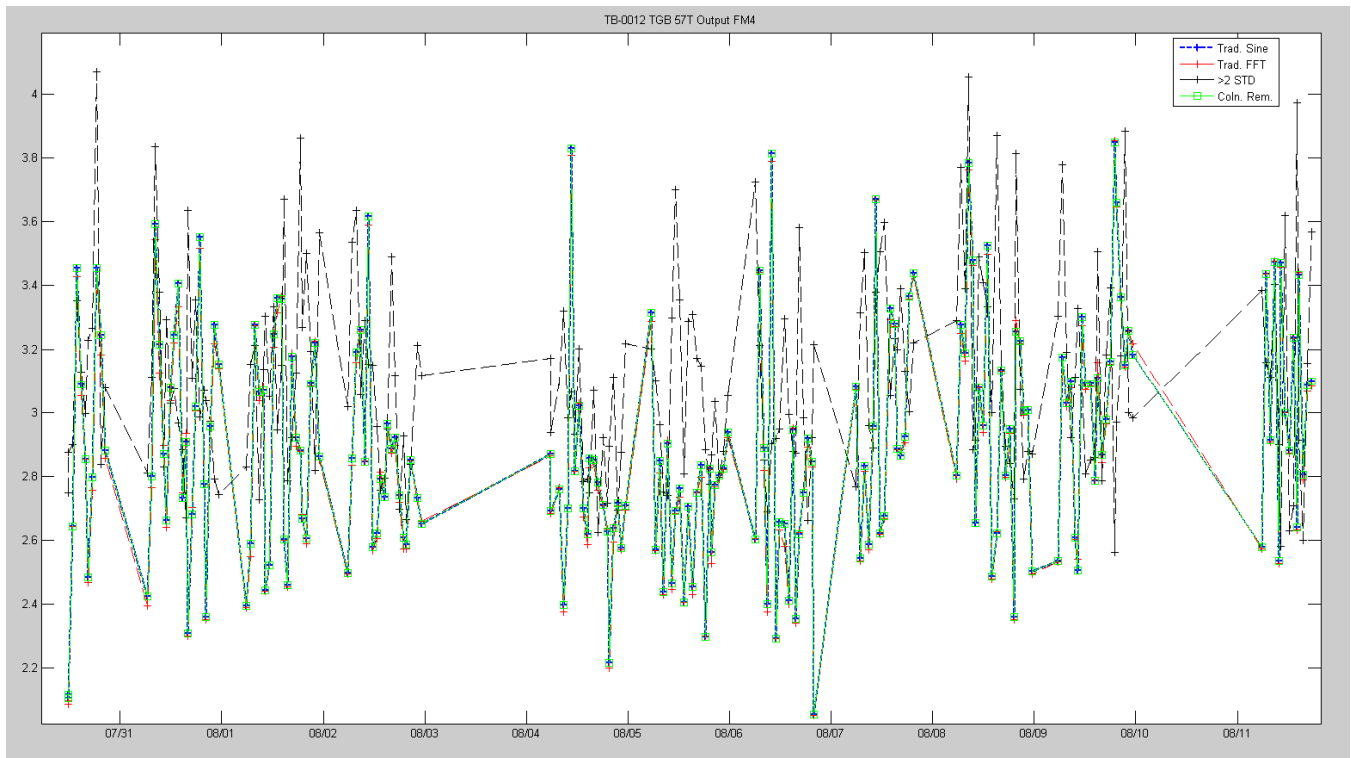


Figure A5: USC Test Stand FM4 Variants' Response to Gear Damage.

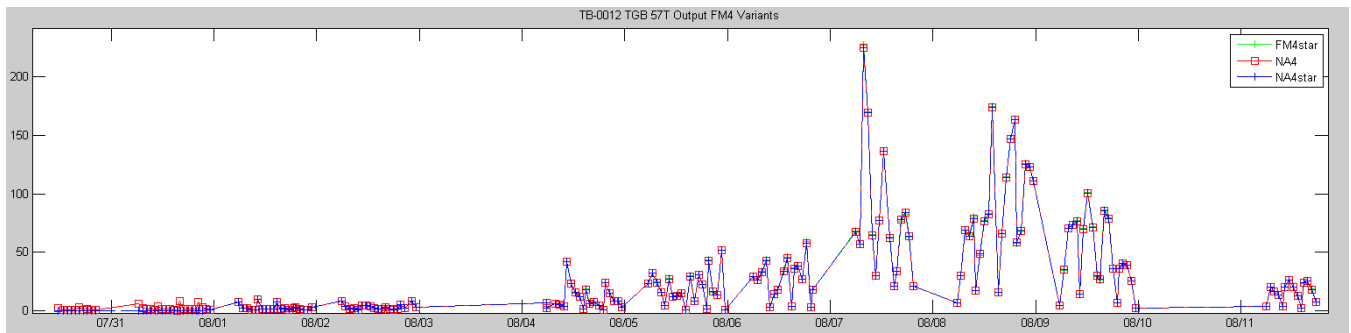


Figure A6: USC Test Stand FM4 Derivatives' Response to Gear Damage.

CI Across Time

USC-64D-TR:TB-0012 for all times

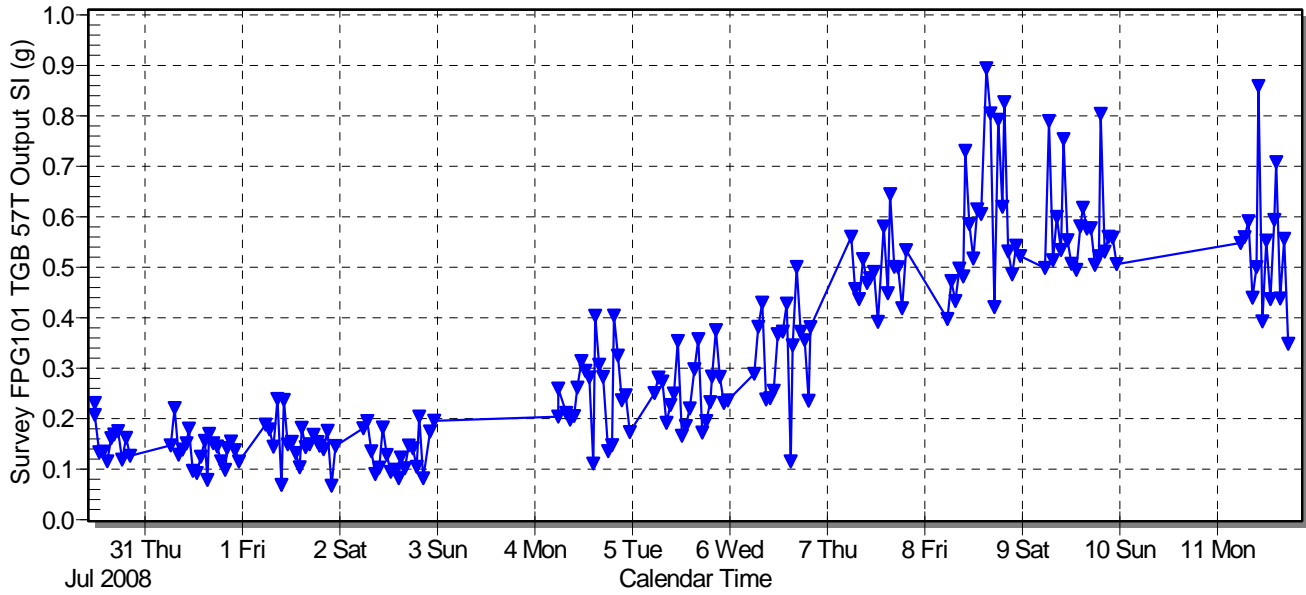


Figure A7: USC Test Stand SI Response to Gear Damage.

CI Across Time

USC-64D-TR:TB-0012 for all times

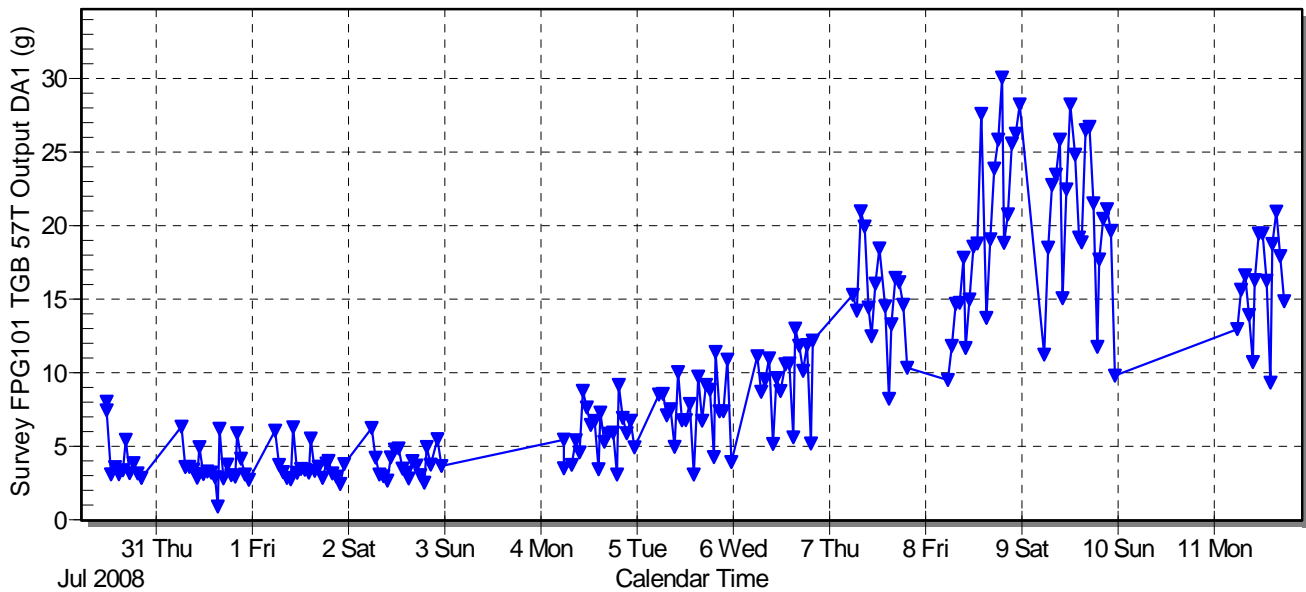


Figure A8: USC Test Stand DA1 Response to Gear Damage

APPENDIX B: CI DEFINITIONS

Sideband index (SI) is a measure of local gear faults and is defined as the average of the order sidebands of the fundamental gear meshing frequency R (Ref. 6). An increase in magnitude of the sidebands of the fundamental gear meshing frequency drives this CI and indicates a local fault.

$$SI = \frac{R_{I-1}^{esb}(x) + R_{I+1}^{esb}(x)}{2}$$

Sideband level factor (SLF) is a coarse indicator of single tooth damage or gear shaft damage. It is defined as the sum of the first order sideband amplitudes of the fundamental gear meshing frequency normalized by the RMS of the synchronous time average (Ref. 6). It differs from SI by the normalization by RMS which makes the CI less sensitive to overall increases in energy.

$$SLF = \frac{R_{I-1}^{esb}(x) + R_{I+1}^{esb}(x)}{RMS(x)}$$

DA1 is the root mean square (RMS) of the TSA subtracted from the mean of the time synchronous average. DA1 detects an overall energy increase in the signal indicating a distributed gear fault.

$$DA1 = RMS(TSA - \overline{TSA})$$

REFERENCES

1. Mosher, M., Pryor, A.H. and Huff, E.M.: Evaluation Of Standard Gear Metrics In Helicopter Flight Operation. Proceedings of the 56th Mechanical Failure Prevention Technology Conference, Virginia Beach, VA, April 15-19, 2002.
2. Inman, D.J. et al, Damage Prognosis for Aerospace, Civil, and Mechanical Systems, 1st Edition, John Wiley & Sons, Ltd, Chichester, West Sussex, 2005.
3. Zakrajsek, J.J. "An Investigation of Gear Mesh Failure Prediction Techniques", NASA Technical Memorandum 102340, November 1989.
4. Stewart, R.M., "Some Useful Data Analysis Techniques for Gearbox Diagnostics", Report MHM/R/10/77, Machine Health Monitoring Group, Institute of Sound and Vibration Research, University of Southampton, July 1977.
5. Mimmagh, M.L. and Neubert, C.G., "SH-60 Helicopter Intermediate Gearbox Gear Tooth Failure Diagnostic Test Results", Interim Report, Naval Air Systems Command Headquarters, Arlington VA, 6 August 1997.
6. Keller, J.A., "A Dictionary of Terms for Machinery Diagnostics of Gearboxes", US Army RDECOM, Aviation Engineering Directorate Aeromechanics Division, Redstone Arsenal, AL.
7. McFadden, P.D., "Analysis of the Vibration of the Input Bevel Pinion in RAN Wessex Helicopter Main Rotor Gearbox WAK143 Prior to Failure", DoD DSTO Aeronautical Research Laboratories AR-004-049, Commonwealth of Australia, 1985.
8. Dempsey, Paula J. and Zakrajsek, J.J. "Minimizing Load Effects on NA4 Gear Vibration Diagnostic Parameter." NASA Glenn Research Center, February 2001.