Gear Fault Detection Effectiveness as Applied to Tooth Surface Pitting Fatigue Damage

David G. Lewicki
Glenn Research Center, Cleveland, Ohio

Paula J. Dempsey
Glenn Research Center, Cleveland, Ohio

Gregory F. Heath and Perumal Shanthakumaran
The Boeing Company, Mesa, Arizona

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI Program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or cosponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at *http://www.sti.nasa.gov*

- E-mail your question via the Internet to *help@sti.nasa.gov*

- Fax your question to the NASA STI Help Desk at 443–757–5803

- Telephone the NASA STI Help Desk at 443–757–5802

- Write to:
  NASA Center for AeroSpace Information (CASI)
  7115 Standard Drive
  Hanover, MD 21076–1320

NASA/TM—2009-215667

Gear Fault Detection Effectiveness as Applied to Tooth Surface Pitting Fatigue Damage

David G. Lewicki
Glenn Research Center, Cleveland, Ohio

Paula J. Dempsey
Glenn Research Center, Cleveland, Ohio

Gregory F. Heath and Perumal Shanthakumaran
The Boeing Company, Mesa, Arizona

October 2009

Acknowledgments

Level of Review: This material has been technically reviewed by technical management.

Available from

NASA Center for Aerospace Information
7115 Standard Drive
Hanover, MD 21076–1320

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161

Available electronically at http://gltrs.grc.nasa.gov

Gear Fault Detection Effectiveness as Applied
to Tooth Surface Pitting Fatigue Damage

David G. Lewicki
Glenn Research Center
Cleveland, Ohio 44135

Paula J. Dempsey
National Aeronautics and Space Administration
Glenn Research Center
Cleveland, Ohio 44135

Gregory F. Heath and Perumal Shanthakumaran
The Boeing Company
Mesa, Arizona 85212

Abstract

A study was performed to evaluate fault detection effectiveness as applied to gear tooth pitting fatigue damage. Vibration and oil-debris monitoring (ODM) data were gathered from 24 sets of spur pinion and face gears run during a previous endurance evaluation study. Three common condition indicators (RMS, FM4, and NA4) were deduced from the time-averaged vibration data and used with the ODM to evaluate their performance for gear fault detection. The NA4 parameter showed to be a very good condition indicator for the detection of gear tooth surface pitting failures. The FM4 and RMS parameters performed average to below average in detection of gear tooth surface pitting failures. The ODM sensor was successful in detecting a significant amount of debris from all the gear tooth pitting fatigue failures. Excluding outliers, the average cumulative mass at the end of a test was 40 mg.

Introduction

Gears are used extensively in rotorcraft drive systems. Effective gear fault detection is crucial to ensure flight safety. In addition, tremendous economic benefits can result from condition based maintenance practices, for which gear fault detection plays an important role.

Over the past 25 years, much research has been devoted to the development of Health and Usage Monitoring systems for rotorcraft gearbox and drivetrain components. Three classic publications on gear diagnostics are by Stewart (Ref. 1), McFadden (Ref. 2), and Zakrajsek (Ref. 3). Samuel and Pines give a comprehensive review of the state-of-the-art in vibration-based helicopter transmission diagnostics (Ref. 4). Dempsey, et al., presents a summary of current methods to identify gear health, with emphases to FAA and U.S. Army rotorcraft applications (Ref. 5). Recent refinements to vibration-based gear fault detection have been made (Refs. 6 to 8) along with other methods such as vibroacoustics (Ref. 9), acoustic emission (Ref. 10), and impact velocity modeling (Ref. 11). A common theme noticed is that experimental data verifying fault detection algorithms and condition indicator (CI) thresholds are sparse.

In a recent study on face gear endurance (Ref. 12), a number of test sets were instrumented with a gear fault detection system and run until failure. The gears failed from tooth surface fatigue and a large fault detection database was populated. The objective of this study is to use this database and evaluate fault detection effectiveness as applied to gear tooth pitting fatigue damage. A further objective is to evaluate the repeatability of the fault detection methods. Vibration and oil-debris monitoring data were gathered from 24 sets of gears run during the previous endurance evaluation study. The gears were tapered involute spur pinions in mesh with face gears. Three common condition indicators (RMS, FM4, and NA4) were deduced from the vibration data and used to evaluate gear fault detection. Receiver operating characteristic curves were further used on the data to define threshold limits. Lastly, cumulative mass from oil-debris monitoring was used for fault detection.

Apparatus

Test Facility

The experiments reported in this report were tested in the NASA Glenn Research Center (GRC) spiral-bevel-gear/face-gear test facility. An overview sketch of the facility is shown in Figure 1(a) and a schematic of the power loop is shown in Figure 1(b). The facility operates in a closed-loop arrangement. A spur pinion drives a face gear in the test (left) section. The face gear drives a set of helical gears, which in turn, drive a face gear and spur pinion in the slave (right) section. The pinions of the slave and test sections are connected by a cross shaft, thereby closing the loop. Torque is supplied in the loop by physically twisting and locking a torque in the pre-load coupling on the slave section shaft.

Additional torque is applied through a thrust piston (supplied with high pressure nitrogen gas), which exerts an axial force on one of the helical gears. The total desired level of torque is achieved by adjusting the nitrogen supply pressure to the piston. A 100–hp DC drive motor, connected to the loop by V-belts and pulleys, controls the speed as well as provides power to overcome friction. The facility has the capability to operate at 750 hp and 20,000 rpm pinion speed. A torquemeter in the loop on the test side measures torque and speed. The facility is also equipped with thermocouples, oil flow meters, pressure transducers, accelerometers, counters, and shutdown instrumentation to allow 24-hr unattended operation.

Test Gears

The design parameters for the pinions and face gears used in the tests are given in Table I. A photograph of the test specimens is shown in Figure 2. The set was primarily designed to fail in surface pitting fatigue mode. The set had a reduction ratio of 3.842:1. The pinions were slightly tapered, which allows the independent setting of backlash for the multiple pinions and idlers in the split-torque transmission application (Ref. 13). The pinions and face gears were made from carburized and ground vacuum induction melting-vacuum arc remelting (VIM–VAR) Pyrowear 53 steel per AMS 6308 using standard aerospace practices. At 6000 lb per in. face gear torque, the calculated AGMA contact stress index was 250 ksi and the calculated AGMA bending stress index was 72 ksi using approximate spur gear calculations per AGMA (Ref. 14).

Gear Fault Detection Instrumentation

A schematic of the gear fault detection instrumentation is shown in Figure 3. Two high-frequency accelerometers and two photoelectric tachometers were used for vibration monitoring. One accelerometer was installed on the test (left) side pinion housing and the other was installed on the slave (right) side pinion housing and were used to monitor the left and right side meshes, respectively. The accelerometers had integral electronics with a nominal 10 mV/g sensitivity, 70 kHz resonant frequency, and were linear within 10 percent up to 20 kHz. One tachometer was installed on the high-speed pinion shaft and the other was installed on the low-speed face gear shaft. Each produced once-per-shaft-revolution indications and were used for time averaging of the vibration data. The outputs of the accelerometers and tachometers were acquired and digitized by a PC.

Vibration data were acquired once every minute during the tests. The accelerometers and tachometers signals were sampled at 155 kHz sampling rate (each) for 10 sec duration by an in-house developed computer program. The program performed linear interpolation and time synchronous averaging. This produced left and right vibration traces relative to the pinion and gear shafts. For the 10 sec acquisition, approximately 380 averages were achieved for a gear trace and over 1000 averages

for a pinion trace. The traces represented the time-averaged vibration for a period of one revolution of the corresponding shaft using 1024 points for the pinion shaft trace and 2048 points for the gear shaft trace. From these traces, three common condition indicators (CI's) were calculated at each acquisition: RMS, FM4, and NA4. Detailed definitions of the CI's are given in Appendix A.

A commercially available in-line oil-debris monitor (ODM) was used to measure metallic content generated in the lubrication system due to mechanical component fatigue failures (Ref. 15). The ODM sensor element consisted of three coils that surrounded a nonconductive section of tubing. The two outside field coils were oppositely wound and driven by an AC current source. The center coil measured the disturbance to the magnetic fields caused by the passage of metallic particles through the sensor. The disturbance was measured as a sinusoidal voltage where the magnitude of the disturbance was proportional to the size of the particle. The ODM controller continuously monitored the sensor and stored values of the calculated cumulated mass of the debris as well as particle counts assembled in bins of particles sizes. The PC system from above polled the ODM controller through its COM port during each vibration acquisition where it time-stamped and stored the cumulated mass along with the vibration CI's.

The ODM sensor was installed in the gravity-fed scavenge oil line coming from the test hardware (Fig. 3). This line contained oil from the left side mesh, left side pinion support bearing, right side mesh, and right side pinion support bearing. Unfortunately, due to the test rig design, isolation of the oil lines for these components was not possible. However, the ODM data was still used as an indicator of the health of the gears as a whole.

Test Procedure

For each set tested, detailed installation and break-in run procedures as described in (Ref. 12) were followed to produce acceptable contact patterns and backlash. After acceptable installation, the pre-load coupling was adjusted to produce a face-gear torque between 3000 to 5000 lb-in. The gears were then run at required speed and torque for the specific test (torque adjusted using load piston). Facility parameters (speed, torque, oil pressures and flows, temperatures) as well as the previously mentioned vibration and ODM data were collected. During the tests, the gears were inspected at routine intervals (5 to 10 million face gear cycles) or when an abnormal facility shutdown occurred. The gears were run until a surface durability failure occurred or a suspension was defined. A surface durability failure was defined as macro-pitting or spalling of at least 0.1 in. continuous length along the contact area on any tooth of a tested pinion or face gear. Once a test was completed, the failed gears were removed from the facility, cleaned, and photographed for documentation purposes. A replacement set was installed per above and testing continued.

Twenty-four sets of gears were tested. Tests were performed at three load levels: 7200 lb-in. face gear torque (275 ksi calculated AGMA contact stress), 8185 lb-in. face gear torque (292 ksi contact stress), and 9075 lb-in. face gear torque (307 ksi contact stress). Test speeds were 2190 to 3280 rpm face gear speed, depending on the vibration levels of the test.

Results and Discussion

Endurance Test Results

A summary of the results from the endurance tests is given in Table II. Twelve sets were run at 7200 lb-in. 7 sets at 8185 lb-in, and 5 sets at 9075 lb-in. face gear torque. The test speeds were 2190 to 3280 rpm face gear speed. Initial tests were run at higher speeds to produce more cycles per time. However, due to wear of the specimens during test, excessive facility vibration levels were produced and the speeds were reduced to reduce vibration to acceptable levels. During pre-test facility check-out runs, resonant speeds from around 2500 to 3000 rpm were discovered, and thus, avoided during test.

Of the 24 sets of gears tested, 17 sets resulted in spalling/macro-pitting failures. The other 7 sets were suspended with moderate to heavy wear but had no spalling. For all the 17 sets that failed, spalling occurred on the pinion. In some cases, spalling occurred on both the pinion and face gear. There were zero instances of face gear spalling with no pinion spalling. Thus, the remainder of this study will concentrate on pinion results only. The tests sets were classified into four groups: (1) pinion macro-pitting with single or few teeth pitted (this occurred for 5 sets), (2) pinion macro-pitting with multiple/all teeth pitted (this occurred for 12 sets), (3) moderate pinion wear but no macro-pitting (this occurred for 3 sets), and (4) heavy pinion wear but no macro-pitting (this occurred for 4 sets). An example of a pinion with single or few teeth pitted is given in Figure 4(a). An example of a pinion with multiple teeth pitted is given in Figure 4(b). The number of cycles tested per set ranged from 32.7 to 590.9 million pinion cycles.

Vibration and ODM data were continuously collected once every minute during all tests. Three gear fault CI's (RMS, FM4, and NA4) were calculated from the time averaged vibration signal for the pinions. The results for all the tests are given in Appendix B. Plotted are RMS, FM4, and NA4 versus data point, where each data point represents one minute of test. As previously mentioned in the Test Procedure section, test gears were replaced after failure or suspension with new sets and testing continued. The absolute start and end times for the 24 sets were intermixed. For each set shown in Appendix B, the data point number is relative to the specific set in question. Thus, as an example, data point 10000 for set 1 (Fig. B1) does not correspond to the same point in time as data point 10000 for set 2 (Fig. B2).

The plots in Appendix B are divided with two types of separators. The first separator is labeled "Rig shutdown" (dotted lines) and was due to rig shutdowns either for routine inspection or abnormal facility parameter exceedance. In these cases, no changes were made to the test gear set setup or vibration monitoring system. The second separator is labeled "Vib reset" and occurred when the vibration monitoring system was reset. This primarily occurred when the opposite side set was replaced due to failure or suspension. The major significance of a "Vib reset" is the reinitialization of the running average of the variance for the NA4 parameter (see Equation (3), Appendix A). Lastly, portions of the data in Appendix B are also classified as "Healthy" and "Faulty," corresponding to a healthy or faulty pinion condition. This classification will be used for determining thresholds as described in a later section of this study.

The results from Appendix B will be used for analysis of gear fault detection and described in detail in later sections of this study. For now, however, a few general comments can be made. Rig shutdowns and "Vib resets" produced discontinuities in the CI responses. Some discontinuities were significant (the RMS response for data points 4335 to 5742 of Figure B5 as an example). For most cases, a failure of the opposite side set was apparent in the CI responses of a given set. Figure B6 for set 6 is an example where set 5 failed at data point 1128. In general, the magnitude of the RMS CI varied from set to set. FM4 was generally bounded within values of 2 to 5. NA4 was also generally bounded for healthy components, but showed a significant increase during failure. NA4, however, was usually more sensitive to inspections and shutdowns.

Evaluation of Data From Healthy Components

The objective of this section is to investigate the variability of the CI's for known healthy components. The data labeled "Healthy" in Appendix B were assembled and the means and standard deviations of the CI's for these data were determined. For 15 of the 24 sets, the healthy data was selected at the start of the set installation. For the remaining sets, the healthy data was offset due to the influence of the opposite side set failures on the CI results. The mean and standard deviation results are shown in Table III and Figure 5.

RMS had a large variation among sets, ranging in mean values from 2.53 to 10.73 g's. FM4 had a fairly steady value of means, with a total average of 2.75 and a relatively low standard deviation. NA4 had a slightly higher mean than FM4 and significantly larger scatter.

Qualitative Analysis of Gear Fault Detection

For the qualitative analysis, the gear fault detection effectiveness was evaluated based on visual inspection of the CI plots from Appendix B. Each CI was rated for fault detection effectiveness for each set with macro-pitting. Ratings varied from 1 to 5, where 5 was excellent effectiveness and 1 was poor effectiveness. A CI was given a 5 rating for a set if it showed an indisputable increase in value

at the time of failure. An example of this is the NA4 response for set 13 (Fig. B13). In this case, NA4 increased by a factor of 50 at the end of the test. A CI was subjectively rated less effective when it did not show a noticeable increase at time of failure, it decreased with increasing failure progression, it exhibited extraneous jumps or spikes, or it was clouded with noise throughout the test. An example of a 3 rating is given for FM4 for set 17 (Fig. B17). Here, FM4 increased at the start of failure (data point 4500) but decreased as the pitting failure propagated. An example of a 1 rating is given for FM4 for set 4 (Fig. B4). Here, FM4 showed no response to the failure at the end of the test.

Figure 6 depicts the results of the qualitative analysis. For the single/few teeth macro-pitting failures (Fig. 6(a)), NA4 showed an excellent fault detection effectiveness. FM4 showed a slightly above average effectiveness. NA4 and FM4 were primarily developed to detect isolated gear tooth faults, which explains the excellent performance of NA4. FM4 suffered in effectiveness due to noise and the decrease in values with increased fault progression. RMS showed a slightly below average effectiveness, indicating that isolated gear faults did not significantly increase the overall vibration signature.

For the multiple teeth macro-pitting failures (Fig. 6(b)), the fault detection effectiveness of NA4 and FM4 decreased compared to the single/few teeth failure modes. Again, this is not surprising since the parameters were developed to detect isolated tooth faults. The RMS fault detection effectiveness increased due to the increased influence of the multiple teeth faults on the overall vibration signature. In general considering all failures (Fig. 6(c)), NA4 showed a good fault detection effectiveness, FM4 was slightly below average, and RMS was average.

Some general observations were noted. Again, CI discontinuities from the inspections and resets increased the difficulty for successful fault detection. This was especially true in the current test setup where opposite side set failures influenced CI performance. Another general observation was that the vibration spectrum was dominated by the gear meshes. This was deduced from analyzing gear orders in the time-averaged vibration as well as analyzing raw vibration signals (nontime averaged) from facility accelerometers.

Quantitative Analysis of Gear Fault Detection

Receiver operating characteristic (ROC) curves were used to validate the qualitative analysis. ROC curves are used in signal detection theory to identify tradeoffs between failure detection and false alarms. They have been used in the medical fields for health decision making and for assessing the predictive accuracy of the tools used to make these decisions (Refs. 16 and 17). Interpretation of medical tests can vary between diagnosticians. ROC curves have been used as a tool to assess the performance of tests independent of the threshold, providing a common metric for comparison (Ref. 18).

The procedure in using ROC curves is as follows. First, CI data is extracted into healthy and faulty groups corresponding to healthy and faulty components. The means and standard deviations of the groups are then determined. Figure 7 shows probability density functions for sample data with a mean and standard deviation of 3.0 and 0.5, respectively, for the healthy set, and a mean and standard deviation of 5.0 and 1.0, respectively, for the faulty set. Note that normal distributions are used in this example and this assumption was used on all the data in this study. For a given CI value (CI = 3.5 in Figure 7 as an example), the false alarm rate and hit rate are the shaded areas in the figure and can be determined from statistics using the CI value probability distribution to calculate the area under the curve. By sweeping through a range of CI's (usually from the mean of the healthy to the mean of the faulty set), one can tabulate and plot the hit rates versus false alarm rates. This is known as the ROC curve. The ROC curve can be used to evaluate the CI fault detection effectiveness as well as to determine a threshold CI value. The threshold CI value with the best performance is the point corresponding to the upper-left most point on the ROC curve. This maximizes the hit rate while minimizing the false alarm rate. One method to determine the optimum numerical value of the threshold is to determine the CI value for the intersection of the tail edge of the healthy probability density function with the leading edge of the faulty probability density function.

ROC curves are given in Figure 8 for two examples. The first example has considerable overlap between the healthy and faulty groups. The threshold value is 3.62 for this example. The ROC curve is fairly smooth (Fig. 8(a)) and the threshold value has less significance due to poor separation of healthy and faulty data. If actual data performed in this manner, the CI would be a poor fault detection indicator. The second example has a greater spread between the healthy and faulty groups. The ROC curve has a sharp edge (Fig. 8(b)) at the upper-left location and thus, a tangible threshold. The threshold value with the optimum performance is 4.42 for this example. If actual data performed in this manner, the CI would be a good fault detection indicator.

ROC curves for RMS, FM4, and NA4 are given in Figures 9 to 11 for the macro-pitting, single/few teeth failures (pinion condition 1). This was based on the healthy and faulty data of sets 13, 15, 17, 19, and 22. The means and standard deviations of the healthy and faulty data, along with the estimated thresholds from the ROC curve analysis, are given in Table IV. ROC curves for the macro-pitting, multiple teeth failures (pinion condition 2) are given in Figures 12 to 14. The means, standard deviations, and thresholds are given in Table V. Note that analysis for the macro-pitting, multiple teeth failures only included 9 out of the 12 total sets for this failure mode (sets 4, 5, 7, 8, 9, 12, 16, 20, and 23). This was due to difficulty in classifying the faulty data regimes for the excluded sets (sets 6, 11, and 21). The CI plots of Appendix B show the groupings of healthy and faulty data that were used for the ROC curve analysis.

Results of the analysis showed that both RMS and FM4 did not show good separation between healthy and faulty data (Figs. 9, 10, 12, and 13). For RMS, significant variation in values from set to set occurred for both healthy and fault data. This increased the standard deviation of the data and thus, caused poor separation. The RMS ROC curves were rather smooth, making the threshold less significant due to the poor separation between healthy and faulty data. For RMS from Tables IV and V, thresholds of 4.24 and 6.14 g's gave hit rates of 0.74 and 0.59 and false alarm rates of 0.14 and 0.16, indicating rather poor gear fault detection effectiveness in itself.

For FM4, considerably less scatter occurred but the means between healthy and faulty data were relatively close together. One characteristic of FM4 is the decrease in value with increased fault progression. This lowers the mean for the faulty data and decreases the separation between healthy and faulty data. The FM4 ROC curves showed a slight inflection point at the upper-left portion of the curve. However, the hit rates were rather low. From Tables IV and V, FM4 thresholds of 3.29 and 3.04 gave hit rates of 0.61 and 0.77 and false alarm rates of 0.06 and 0.05. Although the false alarm was low, the hit rate was also rather low which hurt the gear fault detection effectiveness of FM4.

The analysis showed that NA4 had very good separation between healthy and faulty data (Figs. 11 and 14). Even though NA4 had a significant amount of scatter (standard deviation), there was an extremely noticeable increase in mean for the faulty data, thus providing good separation. There was a problem, however, with the NA4 analysis. As stated before, normal distributions were used in this study. This was a poor choice for the NA4 faulty data. NA4 values significantly increased with fault progression. Even though this increased the mean for the faulty data, it also significantly increased the standard deviation of the fault data also. Since normal distributions were used, a symmetry scatter about the mean resulted. This caused artificially induced lower hit rates. To help alleviate this problem, NA4 values were constrained to a maximum value of 50 in this study. Figure 13(d) shows hit rates approximately 0.85 for NA4 values of 5 or less. In actuality, these hit rates approach 1.0. A better choice for the probability density distribution would have been a nonsymmetry distribution, such as a three-parameter Weibull distribution. From Tables IV and V, thresholds of 7.14 and 5.52 gave hit rates of 0.99 (correcting the value shown in Table V) and false alarm rates less than 0.01. Thus NA4 showed excellent gear fault detection effectiveness.

Oil Debris Monitoring

The results from the oil-debris monitoring (ODM) system is given in Figure 15. Data from all 17 failed sets are included. Shown is the calculated cumulative mass per data point (one data point every minute). The ODM responded to all 17 failures. Some sets had definitive inflection points, indicating increased gear tooth pitting (Fig. 15(a), set 22, at data point

4900, as an example). Others had a steady increase in debris (Fig. 15(a), set 13). Three sets were outliers with a larger amount of debris (sets 4, 5, and to some degree, set 22). There did not appear to be significantly more tooth damage (or bearing failures) to correlate with the larger amount of debris, so its cause is unknown. Excluding the three outliers, the results were fairly consistent among sets with an average value of about 40 mg cumulative mass at the end of test.

As stated before, there were difficulties in the facility setup with the ODM. A single sensor was used for both the left and right test sides. Thus, it was not possible to separate the results per side. This posed two problems. First, the measured results included the debris from both sides. Second, the failure of the opposite side set during a test of a given set produced a significant amount of debris. Therefore, the ODM was reset to zero after each failure, thus producing an offset for some sets. Fortunately, no failures occurred at the same time for the left and right sides, leaving enough separation in the results to give meaningful data.

Conclusions

The objective of this study was to evaluate fault detection effectiveness as applied to gear tooth pitting fatigue damage. Vibration and oil-debris monitoring (ODM) data were gathered from 24 sets of gears run during an endurance evaluation study. Three common condition indicators (RMS, FM4, and NA4) were deduced from the time-averaged vibration data and used with the ODM to evaluate gear fault detection. The following conclusions were obtained:

1. The NA4 parameter showed to be a very good condition indicator for the detection of gear tooth surface pitting failures. Very good separation between healthy and faulty data occurred with NA4.
2. The FM4 and RMS parameters performed average to below average in detection of gear tooth surface pitting failures. FM4 had low scatter in results but had a relatively small separation in mean values of healthy and fault data. For RMS, significant variation in values from set to set occurred.
3. The ODM sensor was successful in detecting a significant amount of debris from all the gear tooth pitting fatigue failures. Excluding outliers, the average cumulative mass at the end of a test was 40 mg.

References

1. Stewart, R.M., "Some Useful Data Analysis Techniques for Gearbox Diagnostics," Jul. 1977.
2. McFadden, P.D., "Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration," Journal of Vibration Acoustics Stress and Reliability in Design-Transactions of the ASME, Vol. 108, no. 2, Apr. 1986, pp. 165–170.

3. Zakrajsek, J.J., "An Investigation of Gear Mesh Failure Prediction Techniques," NASA-TM-102340, 1989.

4. Samuel, P.D., and Pines, D.J., "A Review of Vibration-Based Techniques for Helicopter Transmission Diagnostics," Journal of Sound and Vibration, vol. 282, no. 1–2, 2005, pp. 475–508.

5. Dempsey, P.J., Lewicki, D.G., and Le, D.D., "Investigation of Current Methods to Identify Helicopter Gear Health," Proceedings of the 2007 IEEE Aerospace Conference, Big Sky, MT, Mar. 3–10, 2007.

6. Hochmann, D., and Bechhoefer, E., "Gear Tooth Crack Signals and Their Detection Via the FM4 Measure in Application for a Helicopter HUMS (Health Usage and Management System)," Aerospace Conference, 2003. Proceedings. 2003 IEEE, Big Sky, MT, Mar. 8–15, 2003, pp. 3313–3326.

7. Belsak, A., and Flasker, J., "Method for Detecting Fatigue Crack in Gears," Theoretical and Applied Fracture Mechanics, vol. 46, no. 2, 2006, pp. 105–113.

8. Roan, M.J., Erling, J.G., and Sibul, L.H., "A New, Non-Linear, Adaptive, Blind Source Separation Approach to Gear Tooth Failure Detection and Analysis," Mechanical Systems and Signal Processing, vol. 16, no. 5, 2002, pp. 719–740.

9. Hood, A., Samuel, P., and Pines, D., "A Vibro-Acoustic Methodology for the Detection and Characterization of Spur Gear Tooth Damage," Proceedings of American Helicopter Society 61st Annual Forum, Grapevine, TX, Jun. 1–3, 2005.

10. Tan, C.K., Irving, P., and Mba, D., "A Comparative Experimental Study on the Diagnostic and Prognostic Capabilities of Acoustics Emission, Vibration and Spectrometric Oil Analysis for Spur Gears," Mechanical Systems and Signal Processing, vol. 21, no. 1, 2007, pp. 208–233.

11. Parey, A., and Tandon, N., "Impact Velocity Modelling and Signal Processing of Spur Gear Vibration for the Estimation of Defect Size," Mechanical Systems and Signal Processing, vol. 21, no. 1, 2007, pp. 234–243.

12. Lewicki, D.G., Heath, G.F., Filler, R.R., Slaughter, S.C., and Fetty, J., "Face-Gear Surface Durability Investigations," Journal of the American Helicopter Society, vol. 53, no. 3, Jul. 2008, pp. 282–289.

13. Heath, G.F., Filler, R.R., and Tan, J., "Development of Face Gear Technology for Industrial and Aerospace Power Transmission," NASA/CR—2002–211320, Army Research Laboratory Contractor Report ARL–CR–0485, The Boeing Company, Cooperative Agreement NCC3–356, May 2002.

14. ANSI/AGMA 2001–C95, "Fundamental Rating Factors and Calculation Methods for Involute Spur and Helical Gear Teeth," The American Gear Manufacturers Association, 1995.

15. Dubowski, D., and Witwer, D., "Application of Metalscan Oil Debris Monitor to a Fleet of Sea King Helicopters," American Helicopter Society 60th Annual Forum, Baltimore, MD, Jun. 7–10, 2004.

16. Fawcett, T., "An Introduction to ROC Analysis," Pattern Recognition Letters, vol. 27, 2006, pp. 861–874.

17. Gonen, M., "Receiver Operating Characteristic Curves," Proceedings of the 31st Annual SAS Users Group International Conference, San Francisco, CA, Mar. 26–29, 2006.

18. Park, S.H., Goo, J.M., and Jo, C.H., "Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists," Korean Journal of Radiology, vol. 5, no. 1, Mar. 2004, pp. 1–18.

19. Zakrajsek, J.J., Townsend, D.P., and Decker, H.J., "An Analysis of Gear Fault Detection Methods as Applied to Pitting Fatigue Failure Data," MFPT Mtg. no. 47, 1993.

TABLE I.—TEST GEAR DESIGN DATA

AGMA quality	12
Number of teeth; pinion, gear	19, 73
Diametral pitch (teeth/in.)	10.6
Pressure angle (deg)	27.5
Shaft angle (deg)	90
Face width (in.); pinion, gear	0.8, 0.6
Hardness (Rc); case, core	62, 38
RMS surface finish (min)	16
Material	X53 steel

TABLE II.—RESULTS OF ENDURANCE TESTS

Set no.	Side	Face gear speed, rpm	Face gear torque, lb-in.	M Pin cycles	Pinion condition
1	Right	2200 to 3280	7200	361.5	4
2	Left	2880 to 3280	7200	590.9	4
3	Right	2880 to 3280	7200	559.8	4
4	Left	2300	9075	77.2	2
5	Right	2300	9075	88.0	2
6	Left	2300	9075	38.4	2
7	Right	2300	9075	41.9	2
8	Left	2300	9075	32.7	2
9	Right	2300	8185	37.7	2
10	Left	2200 to 2300	7200	461.8	4
11	Right	2300	7200	65.7	2
12	Right	2300	7200	66.1	2
13	Left	2280	8185	126.0	1
14	Right	2300	8185	202.9	3
15	Left	2300	8185	102.6	1
16	Right	2300	8185	212.9	2
17	Left	2300	8185	42.6	1
18	Left	2300	8185	144.5	3
19	Left	2300	7200	35.7	1
20	Right	2190 to 2300	7200	45.3	2
21	Right	2190 to 2300	7200	99.1	2
22	Left	2300	7200	60.7	1
23	Left	2190 to 2300	7200	161.0	2
24	Right	2200	7200	113.0	3

Pinion condition:
1 = Macro-pitting, single/few teeth.
2 = Macro-pitting, multiple teeth.
3 = Moderate wear
4 = Heavy wear

TABLE III.—MEAN AND STANDARD DEVIATION STATISTICS
FOR ALL SETS, HEALTHY STATE CONDITION

Set no.	No. points	RMS		FM4		NA4	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
1	8 782	7.82	0.61	2.85	.23	1.83	0.25
2	10 000	3.93	.52	2.87	.51	4.52	2.22
3	10 000	6.14	1.58	3.25	.55	5.32	2.85
4	2 000	2.90	0.10	3.16	.24	3.25	0.92
5	4 000	4.95	.65	2.26	.10	1.56	.50
6	873	4.21	.15	3.21	.25	4.38	1.43
7	647	7.92	.22	2.42	.05	2.46	0.49
8	532	3.00	.24	2.68	.11	3.21	.47
9	54	5.95	.27	2.55	.03	2.35	.12
10	15 440	4.74	1.30	2.81	.20	3.72	.69
11	6 510	5.04	1.06	2.57	.32	6.94	2.92
12	1 000	3.42	0.10	2.83	.09	2.29	0.16
13	13 155	3.64	.61	2.99	.26	4.08	1.30
14	13 155	9.34	1.85	2.31	.31	1.67	0.43
15	8 960	3.10	0.45	2.89	.12	3.11	.54
16	4 242	6.31	.11	2.14	.04	3.83	.67
17	4 000	2.53	.22	2.97	.18	2.73	1.04
18	12 918	3.07	.30	2.59	.14	4.57	1.04
19	2 000	5.24	.39	2.85	.24	2.85	0.38
20	237	5.03	.32	3.02	.09	3.43	.25
21	3 889	10.73	.68	2.08	.13	2.66	.60
22	2 000	2.98	.18	2.59	.13	3.25	.62
23	3 309	3.12	.15	2.40	.13	2.62	.36
24	8 768	6.13	1.03	3.13	.17	4.23	1.07
All	136 471	5.23	2.39	2.75	.42	3.65	1.91

TABLE IV.—DATA SUMMARY FOR MACRO-PITTING,
SINGLE/FEW TEETH FAILURE MODE
[Pinion condition 1 of Table II.]

Condition indicator	Healthy		Faulty		Threshold		
	Mean	Std. dev.	Mean	Std. dev.	Value	Hit rate	False rate
RMS	3.39	0.79	4.97	1.14	4.24	0.74	0.14
FM4	2.92	.23	3.50	0.78	3.29	.61	.06
NA4	3.47	1.14	38.46	13.53	7.14	.99	.00

TABLE V.—DATA SUMMARY FOR MACRO-PITTING, MULTIPLE
TEETH FAILURE MODE
[Pinion condition 2 of Table II.]

Condition indicator	Healthy		Faulty		Threshold		
	Mean	Std. dev.	Mean	Std. dev.	Value	Hit rate	False rate
RMS	4.64	1.54	6.67	2.41	6.14	0.59	0.16
FM4	2.44	0.36	3.89	1.13	3.04	.77	.05
NA4	2.76	1.03	28.45	22.23	5.52	.85[a]	.00

[a]Artificially low due to normal distribution
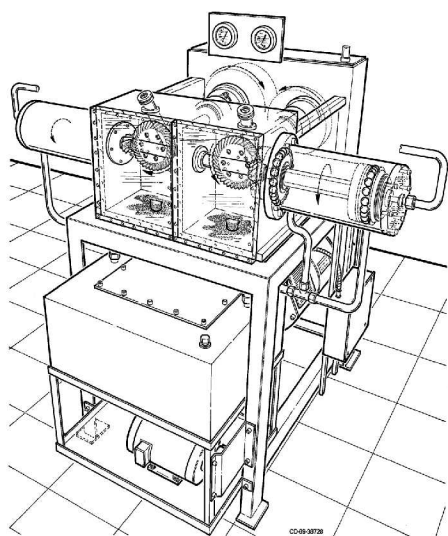
(a) Overview of facility.
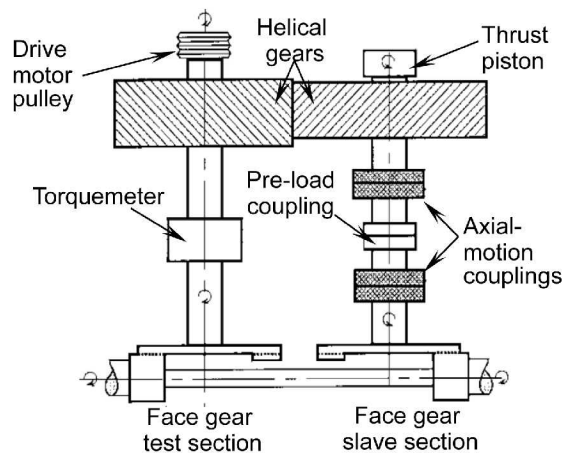
(b) Schematic view.

Figure 1.—NASA Glenn spiral-bevel-gear, face-gear test facility.
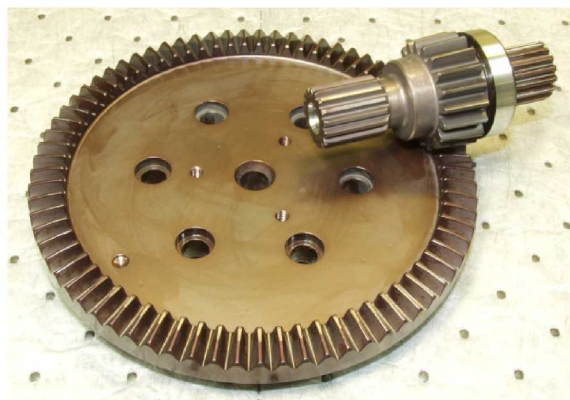
Figure 2.—Test gears.
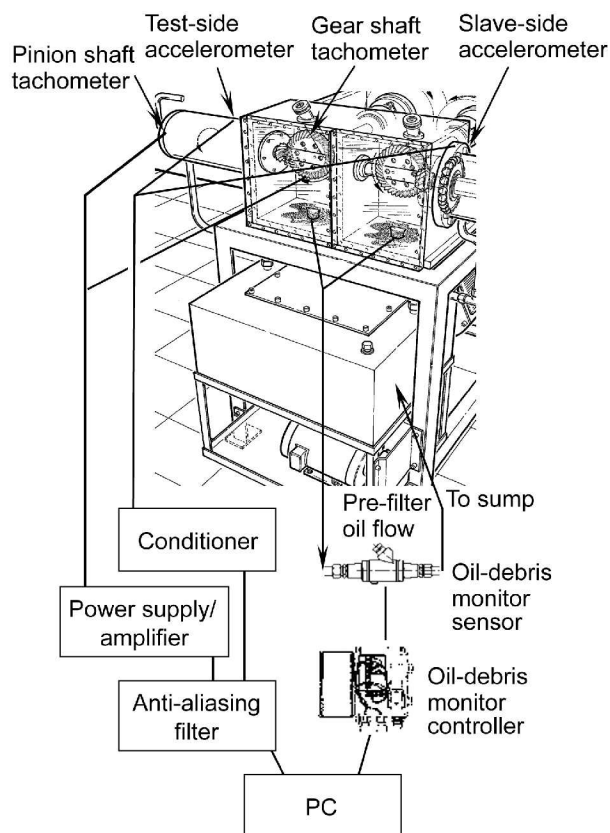
Figure 3.—Gear fault detection instrumentation.

(a) Single/few teeth macro-pitting failure.

(b) Multiple teeth macro-pitting failure.

Figure 4.—Typical macro-pitting pinion tooth surface fatigue failures.
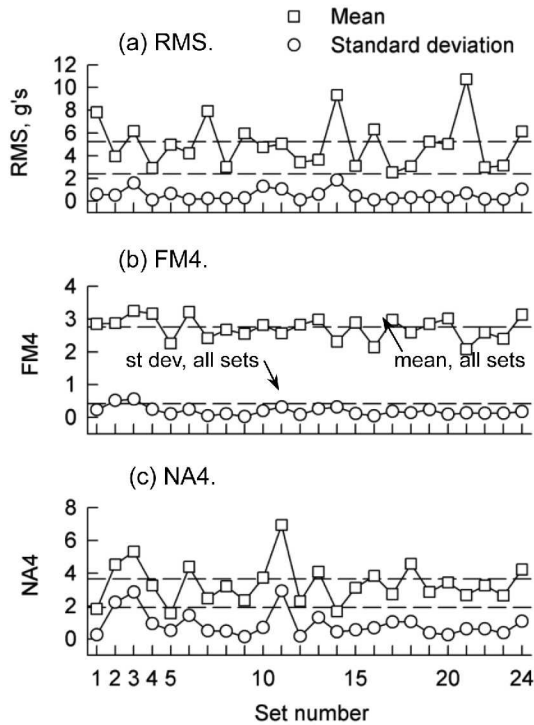
Figure 5.—Mean and standard deviation statistics for all sets, healthy state condition.

Figure 7.—Sample probability density functions.
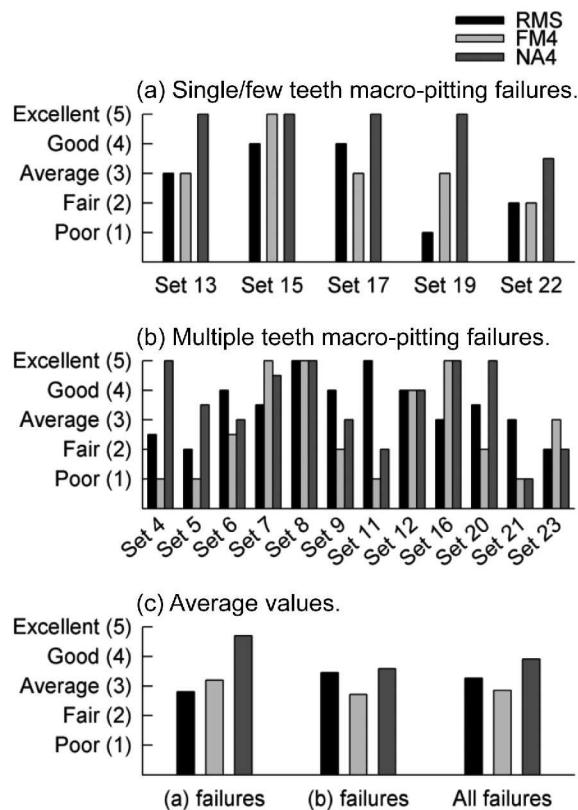
Figure 6.—Qualitative analysis of condition indicator fault detection effectiveness.
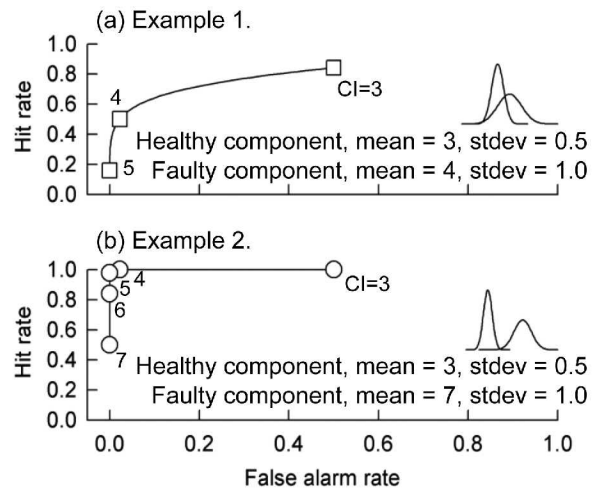
Figure 8.—Sample receiver operating characteristic (ROC) curves.

(a) Healthy data.

(a) Healthy data.

(b) Faulty data.

(b) Faulty data.

(c) Probability density functions.

(c) Probability density functions.

(d) Receiver operating characteristic curve.

(d) Receiver operating characteristic curve.
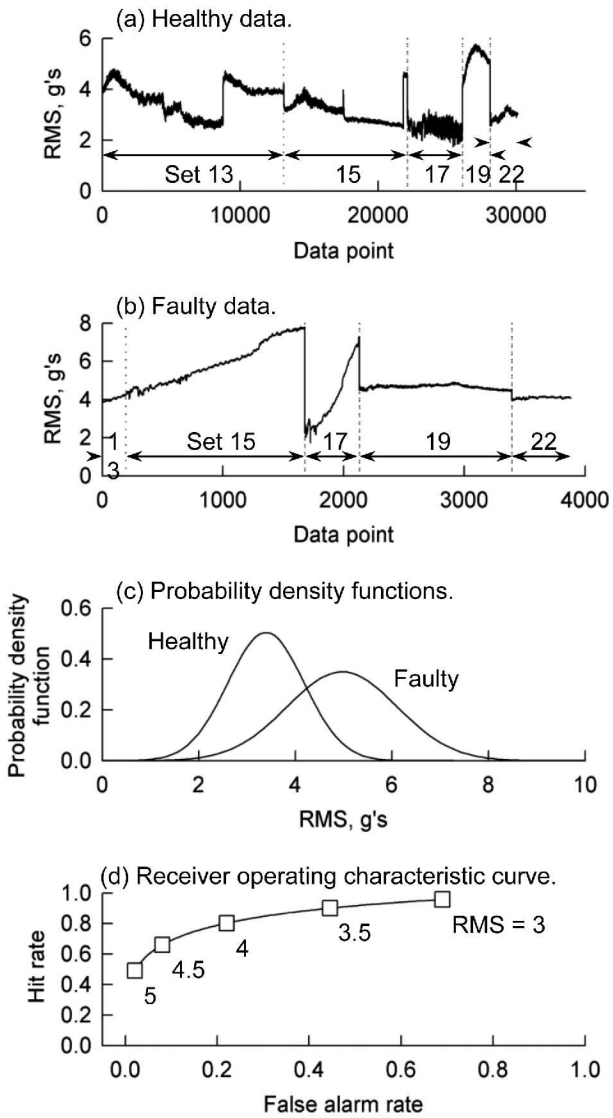
Figure 9.—Summary results for RMS condition indicator for macro-pitting, single/few teeth failures.
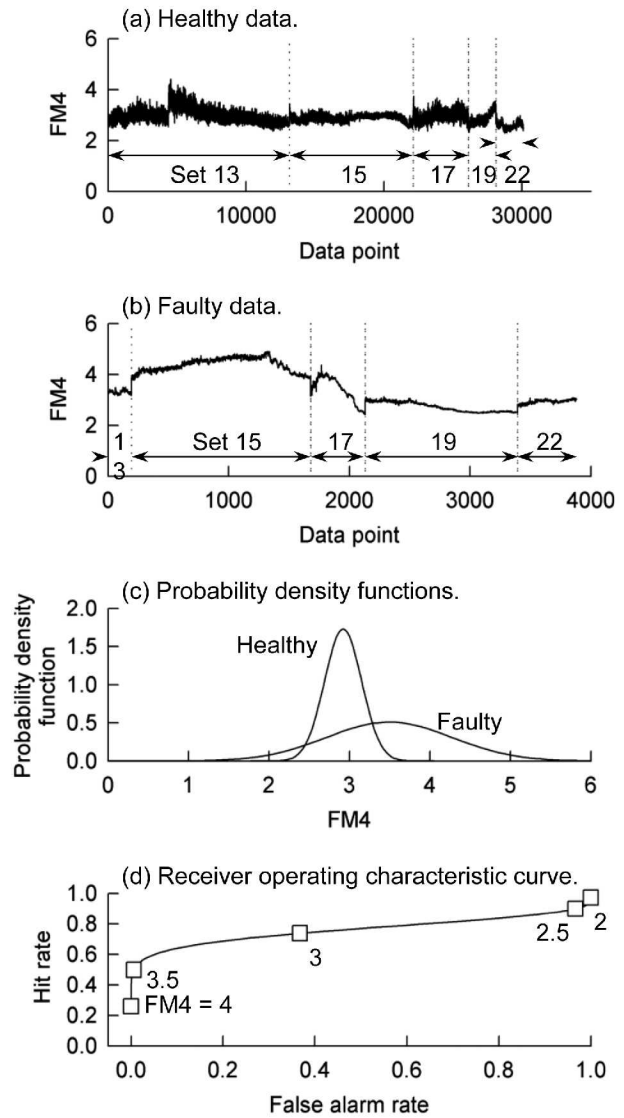
Figure 10.—Summary results for FM4 condition indicator for macro-pitting, single/few teeth failures.
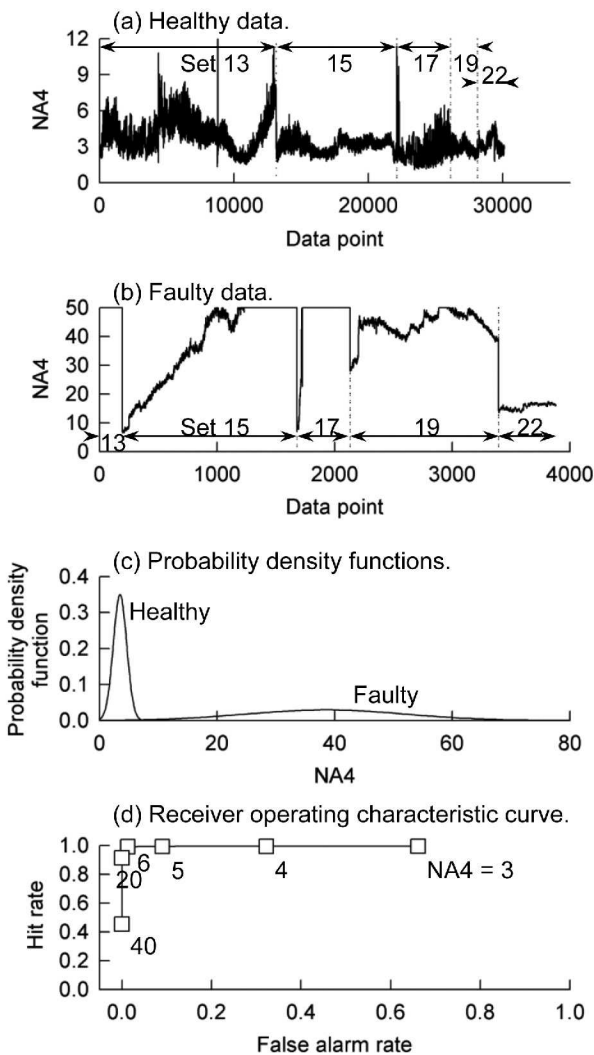
Figure 11.—Summary results for NA4 condition indicator for macro-pitting, single/few teeth failures.
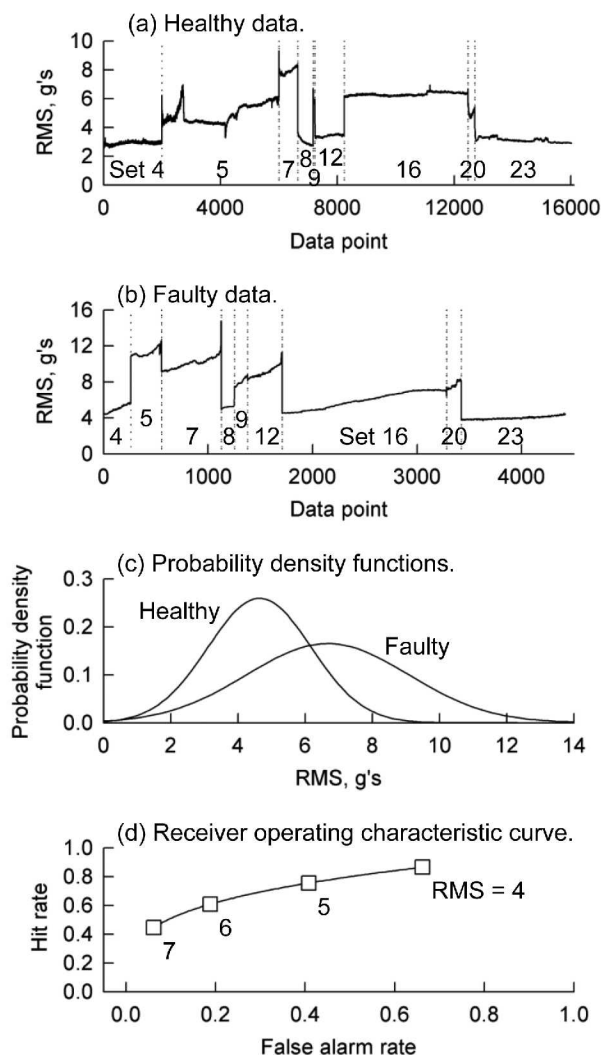
Figure 12.—Summary results for RMS condition indicator for macro-pitting, multiple teeth failures.
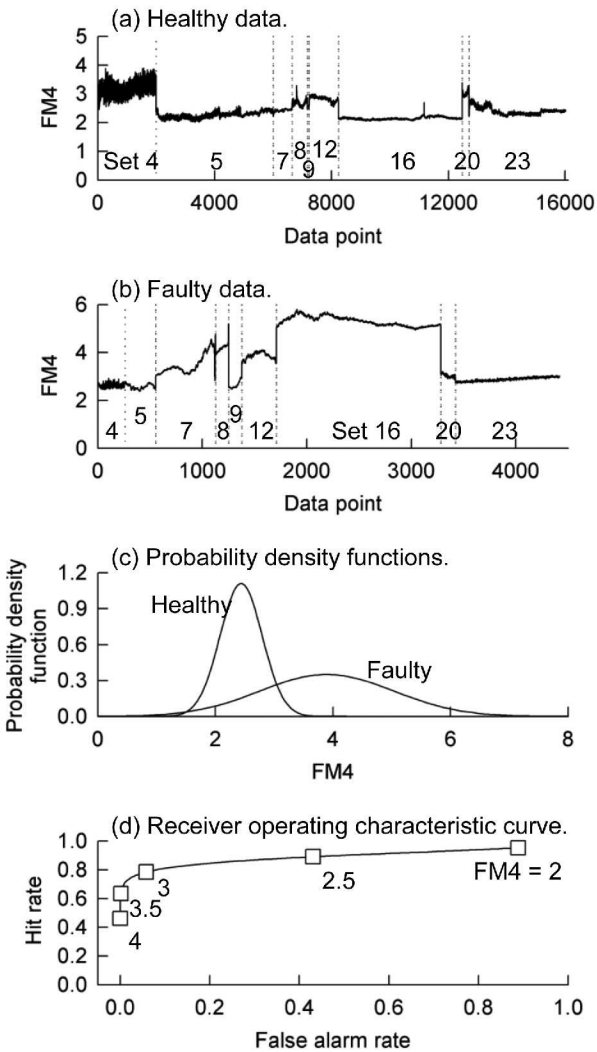
Figure 13.—Summary results for FM4 condition indicator for macro-pitting, multiple teeth failures.
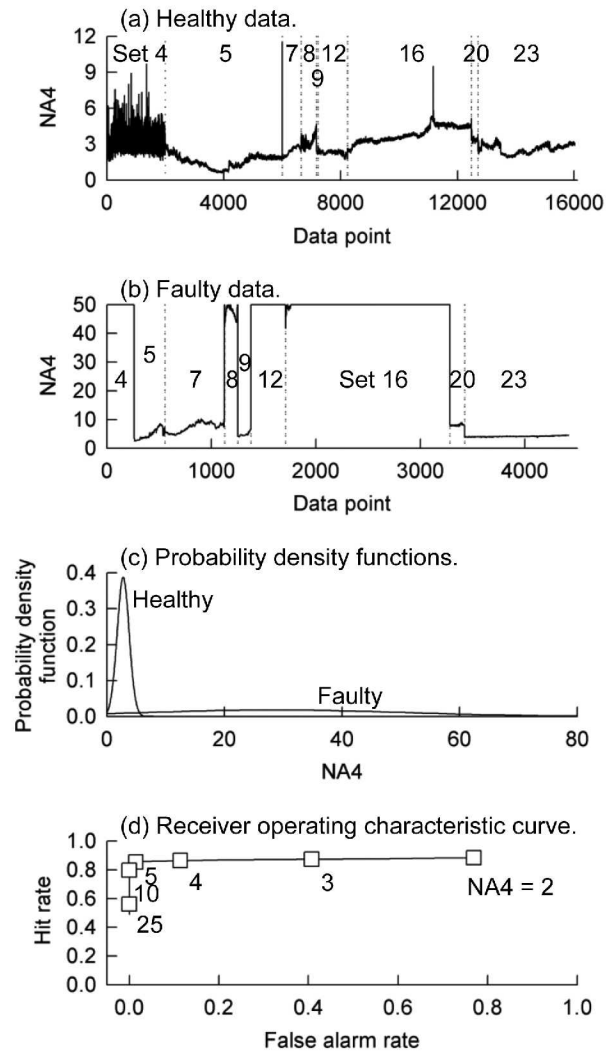
Figure 14.—Summary results for NA4 condition indicator for macro-pitting, multiple teeth failures.

(a) Macro-pitting, single/few teeth.

(b) Macro-pitting, multiple teeth.
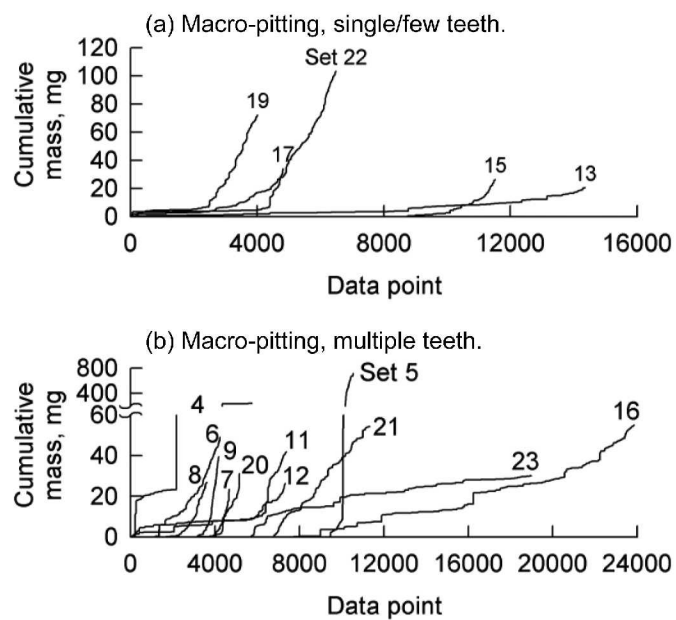
Figure 15.—Oil-debris monitor results.

Appendix A.—CI Definitions

Root Mean Square

The root mean square (RMS) is defined as the square root of the average of the sum of the squares of the time-averaged vibration trace (Eq. (1)). For a simple sine wave, the RMS value is approximately 0.707 times the amplitude of the signal.

$$\text{RMS} = \sqrt{\frac{1}{N}\left[\sum_{i=1}^{N} S_i^2\right]} \tag{1}$$

where
S time-averaged vibration trace
i data point number in vibration trace
N total number of data points in vibration trace

FM4

The FM4 parameter (Eq. (2)) was developed to detect changes in vibration pattern resulting from damage to a single gear tooth (Ref. 1). The metric is calculated by dividing the fourth statistical moment (kurtosis) of the difference signal by the square of the variance of the difference signal. The difference signal is defined as the time-averaged vibration trace, S, minus the gear mesh frequencies and shaft orders. The metric is nondimensional with a nominal value of 3 for Gaussian noise (assumed for a healthy component).

$$\text{FM4} = \frac{N\sum_{i=1}^{N}(d_i - \bar{d})^4}{\left[\sum_{i=1}^{N}(d_i - \bar{d})^2\right]^2} \tag{2}$$

where
d difference signal
$\bar{d}$ mean value of difference signal
i data point number in difference signal
N total number of data points in difference signal

NA4

The NA4 metric (Eq. (3)) was developed to overcome a shortcoming of the FM4 metric (Ref. 19). As the occurrences of damage progresses in both number and severity, FM4 becomes less sensitive to the new damage. Two changes were made to the FM4 metric to develop the NA4 metric as one that is more sensitive to progressing damage. One change is that FM4 is calculated from the difference signal while NA4 is calculated from the residual signal. The residual signal includes the first order sidebands that were removed from the difference signal. The second change is that trending was incorporated into the NA4 metric. While FM4 is calculated as the ratio of the kurtosis of the data record divided by the square of the variance of the same data record, NA4 is calculated as the ratio of the kurtosis of the data record divided by the square of the average variance. The average variance is the mean value of the variance of all previous data records in the run ensemble. These two changes make the NA4 metric a more sensitive and robust metric. The NA4 metric is calculated by

$$NA4 = \frac{N \sum\limits_{i=1}^{N} (r_i - \bar{r})^4}{\frac{1}{M} \sum\limits_{j=1}^{M} \left[\sum\limits_{i=1}^{N} (r_{ij} - \bar{r_j})^2 \right]^2} \qquad (3)$$

where

r	residual signal
$\bar{r}$	mean value of residual signal
i	data point number in residual signal
N	total number of points in residual signal
j	time record number in run ensemble
M	current time record in run ensemble

Appendix B.—CI Traces
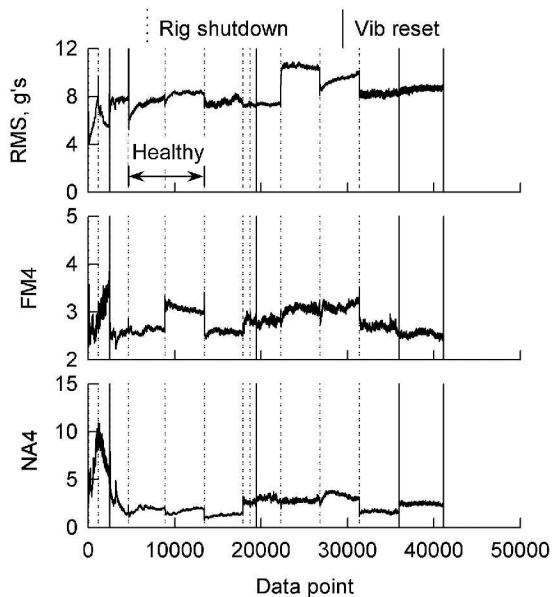
Figure B1.—Set 1 vibration fault detection data.
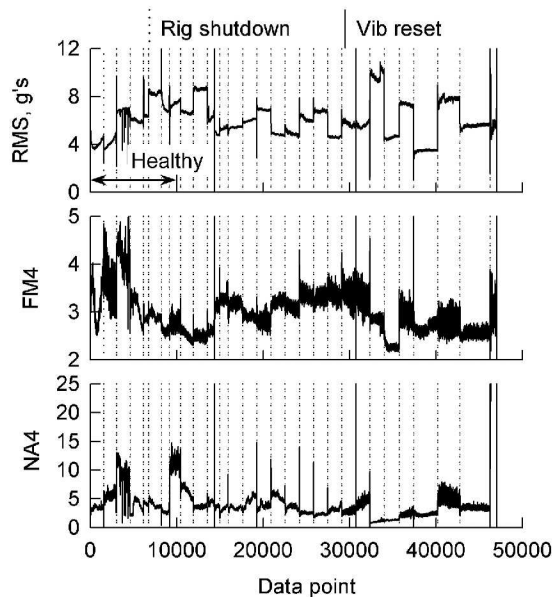
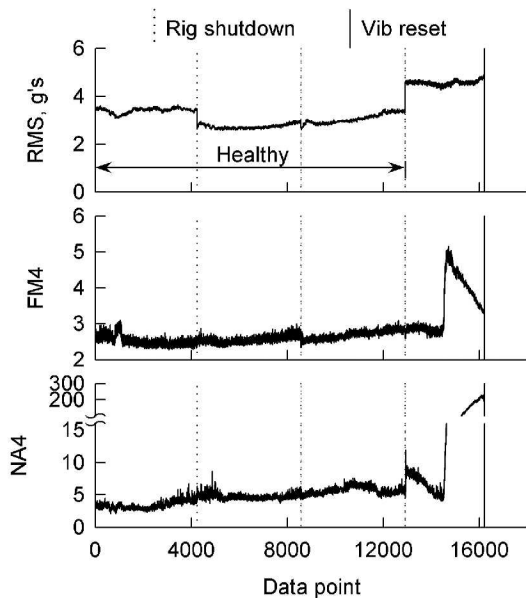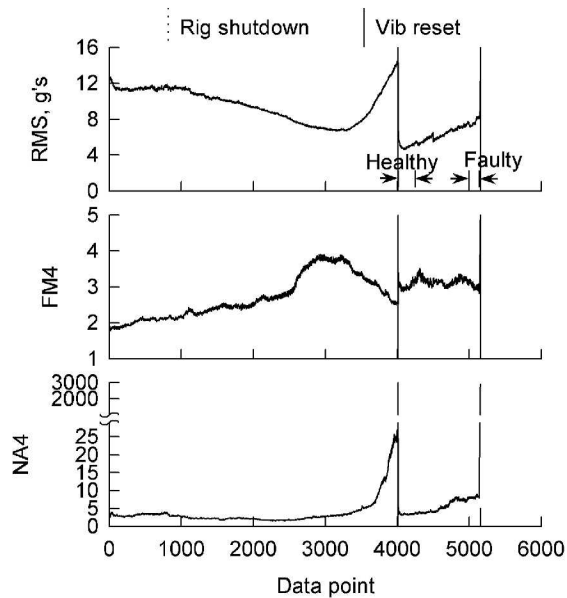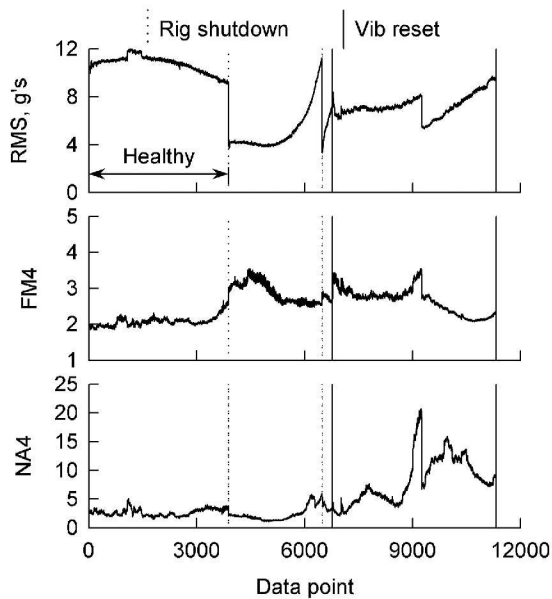

Figure B3.—Set 3 vibration fault detection data.
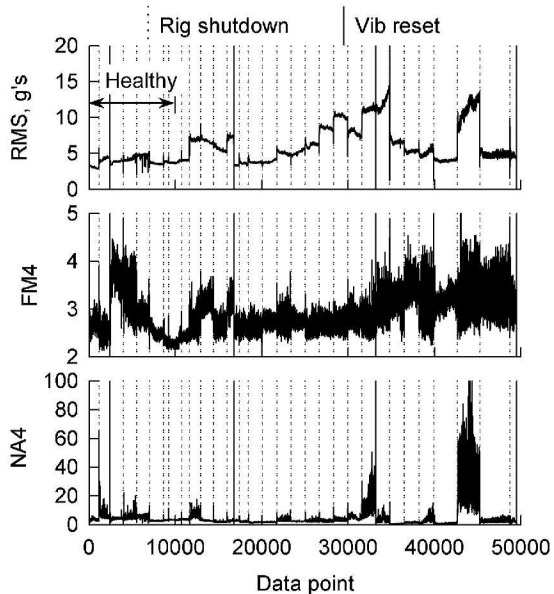
Figure B2.—Set 2 vibration fault detection data.
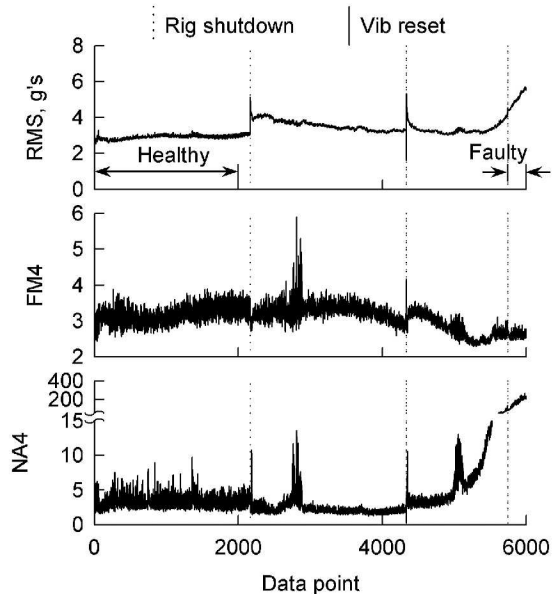
Figure B4.—Set 4 vibration fault detection data.
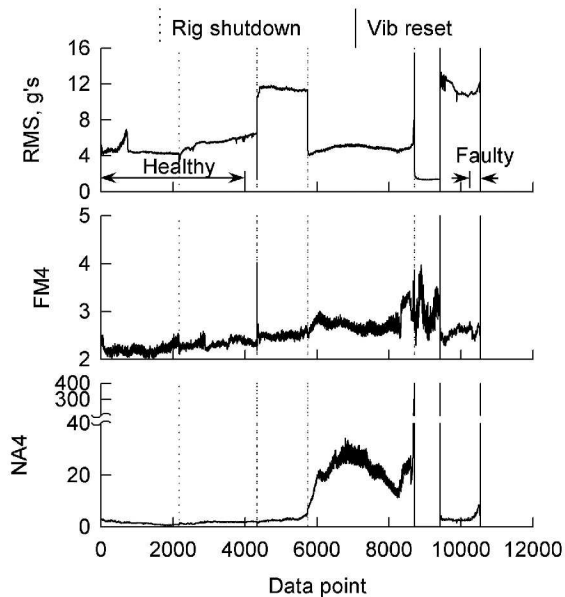
Figure B5.—Set 5 vibration fault detection data.
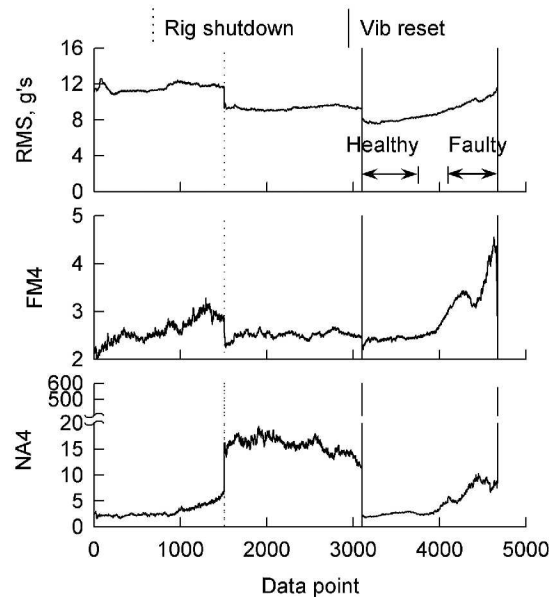
Figure B7.—Set 7 vibration fault detection data.
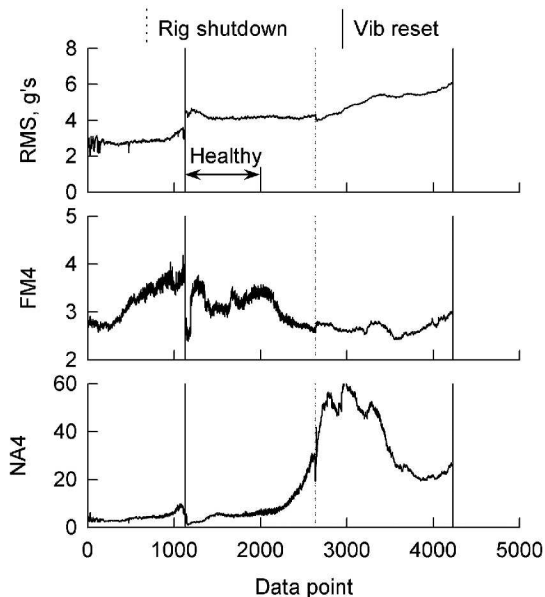
Figure B6.—Set 6 vibration fault detection data.
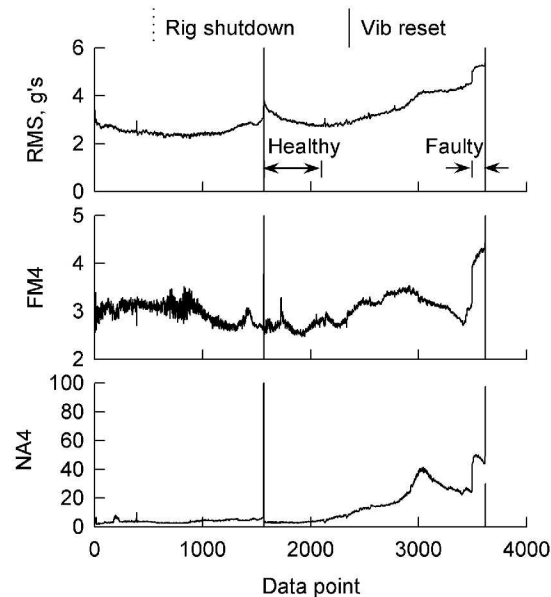
Figure B8.—Set 8 vibration fault detection data.
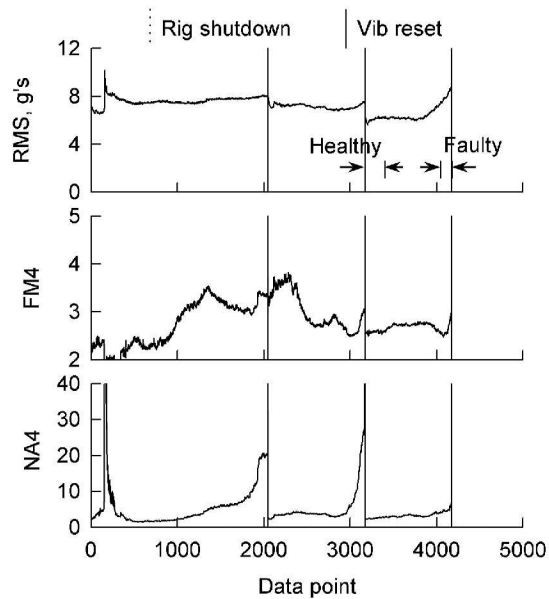
Figure B9.—Set 9 vibration fault detection data.
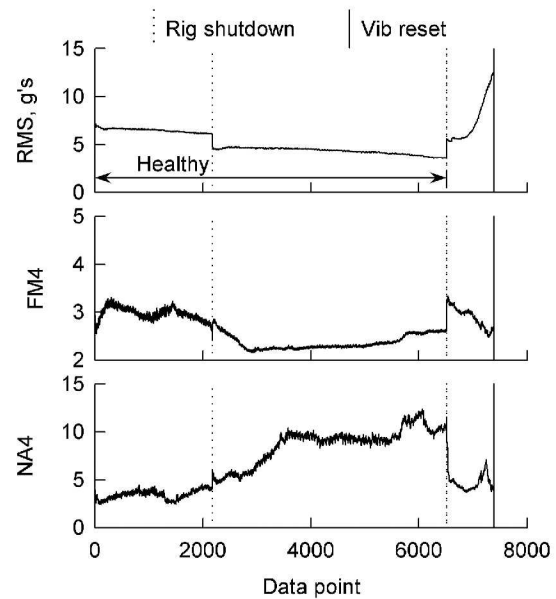
Figure B11.—Set 11 vibration fault detection data.
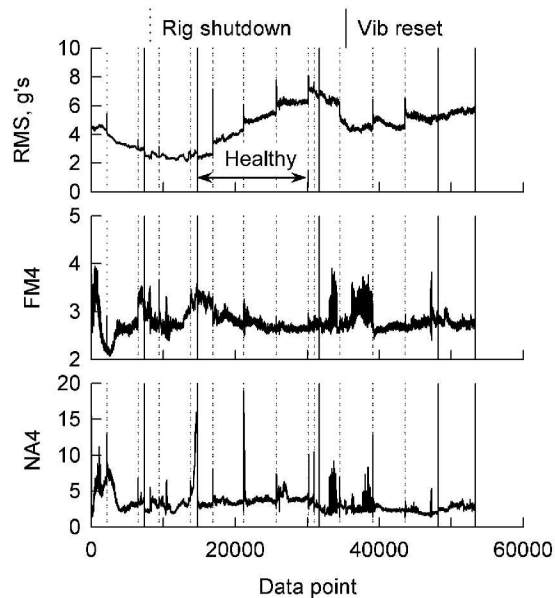
Figure B10.—Set 10 vibration fault detection data.
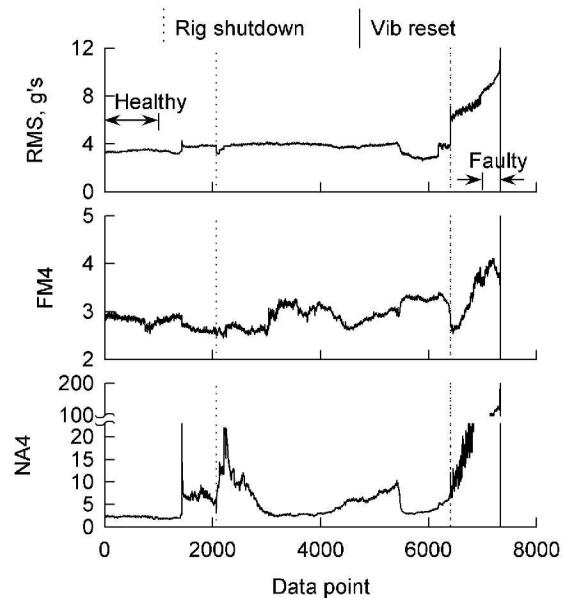
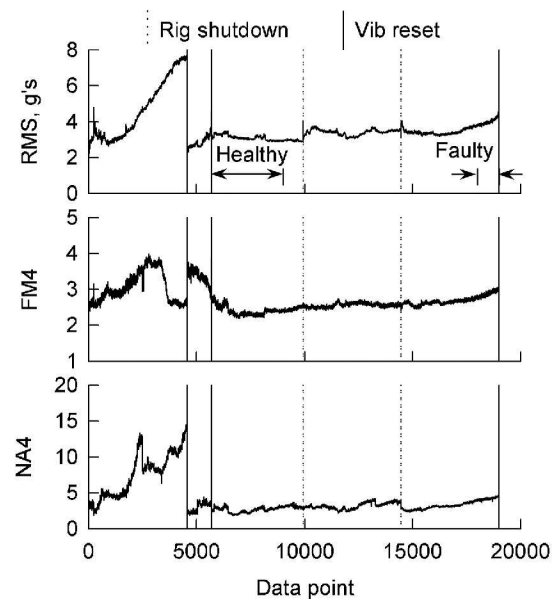Figure B12.—Set 12 vibration fault detection data.
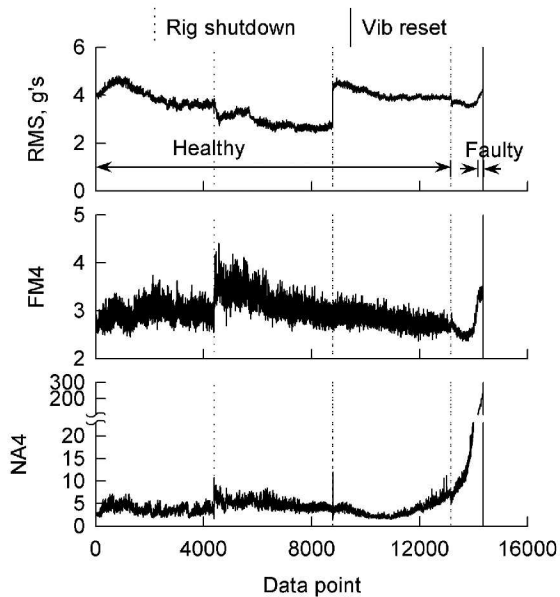
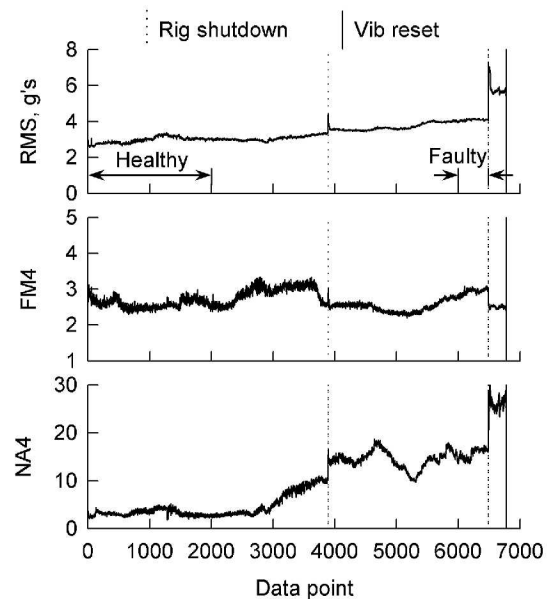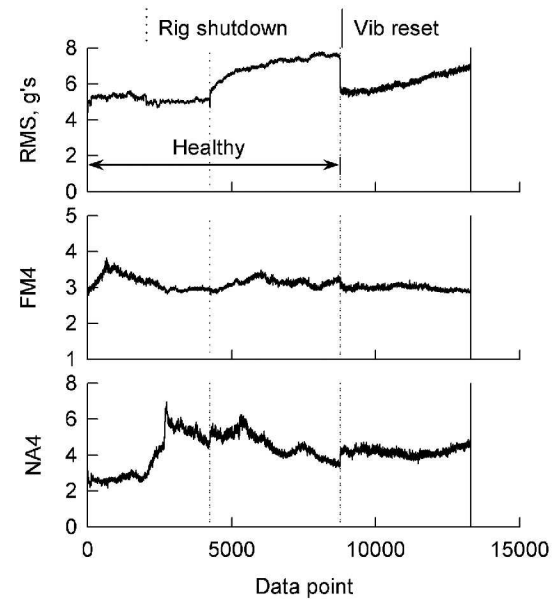Figure B13.—Set 13 vibration fault detection data.
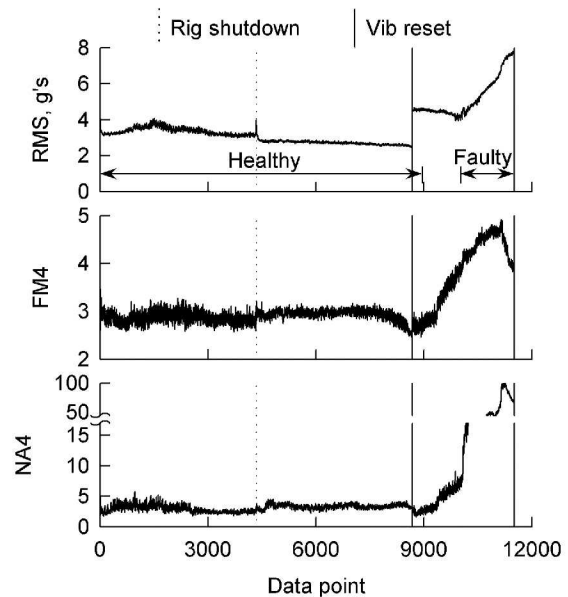
Figure B15.—Set 15 vibration fault detection data.
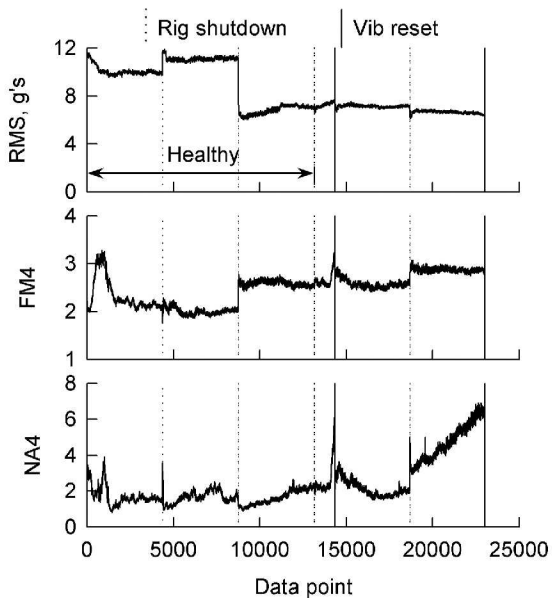
Figure B14.—Set 14 vibration fault detection data.
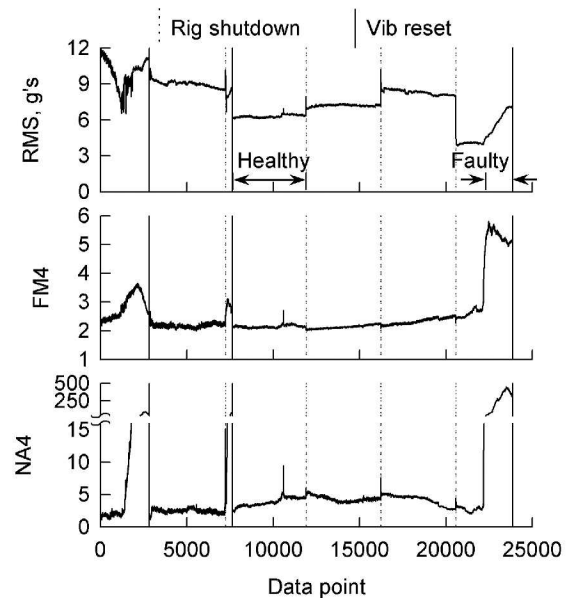
Figure B16.—Set 16 vibration fault detection data.
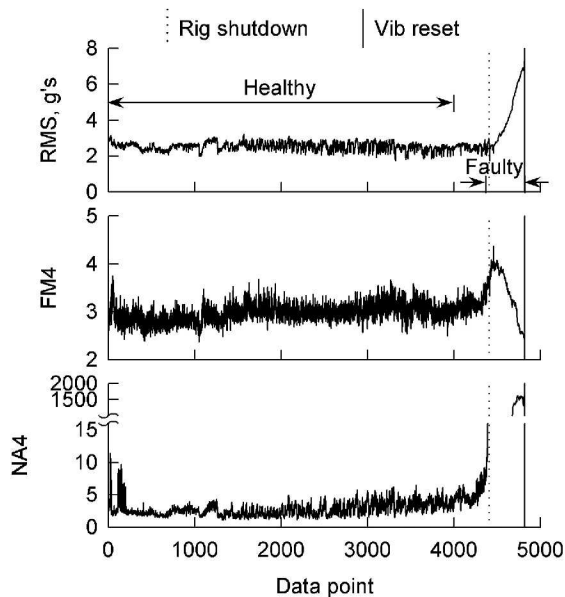
Figure B17.—Set 17 vibration fault detection data.
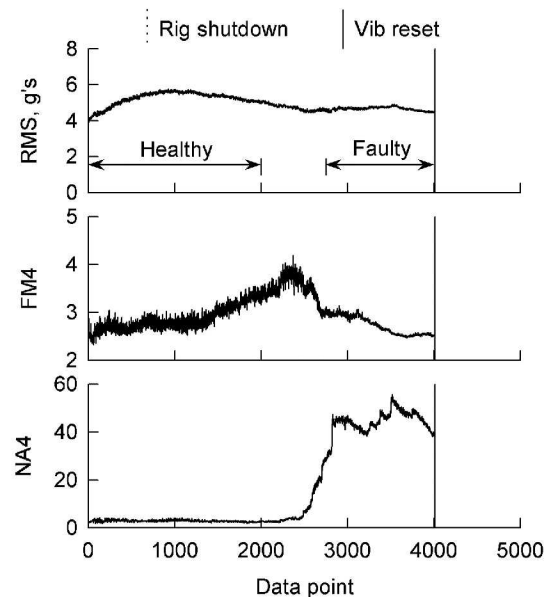
Figure B19.—Set 19 vibration fault detection data.

Figure B18.—Set 18 vibration fault detection data.

Figure B20.—Set 20 vibration fault detection data.

Figure B21.—Set 21 vibration fault detection data.

Figure B22.—Set 22 vibration fault detection data.

Figure B23.—Set 23 vibration fault detection data.

Figure B24.—Set 24 vibration fault detection data.

1. REPORT DATE (DD-MM-YYYY) 01-10-2009	2. REPORT TYPE Technical Memorandum	3. DATES COVERED (From - To)

4. TITLE AND SUBTITLE Gear Fault Detection Effectiveness as Applied to Tooth Surface Pitting Fatigue Damage	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Lewicki, David, G.; Dempsey, Paula, J.; Heath, Gregory, F.; Shanthakumaran, Perumal	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER WBS 877868.02.07.03.01.01.02

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration John H. Glenn Research Center at Lewis Field Cleveland, Ohio 44135-3191	8. PERFORMING ORGANIZATION REPORT NUMBER E-17009

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001	10. SPONSORING/MONITOR'S ACRONYM(S) NASA
	11. SPONSORING/MONITORING REPORT NUMBER NASA/TM-2009-215667

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified-Unlimited
Subject Category: 37
Available electronically at http://gltrs.grc.nasa.gov
This publication is available from the NASA Center for AeroSpace Information, 443-757-5802

13. SUPPLEMENTARY NOTES

14. ABSTRACT
A study was performed to evaluate fault detection effectiveness as applied to gear tooth pitting fatigue damage. Vibration and oil-debris monitoring (ODM) data were gathered from 24 sets of spur pinion and face gears run during a previous endurance evaluation study. Three common condition indicators (RMS, FM4, and NA4) were deduced from the time-averaged vibration data and used with the ODM to evaluate their performance for gear fault detection. The NA4 parameter showed to be a very good condition indicator for the detection of gear tooth surface pitting failures. The FM4 and RMS parameters performed average to below average in detection of gear tooth surface pitting failures. The ODM sensor was successful in detecting a significant amount of debris from all the gear tooth pitting fatigue failures. Excluding outliers, the average cumulative mass at the end of a test was 40 mg.

15. SUBJECT TERMS
Gears; Fault detection; Vibration; Oil debris; Receiver operating characteristic curve

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON STI Help Desk (email:help@sti.nasa.gov)
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU	26	19b. TELEPHONE NUMBER (include area code) 443-757-5802