Gear Damage Detection Using Oil Debris Analysis

Paula J. Dempsey
Glenn Research Center, Cleveland, Ohio

Prepared for the
14th International Congress and Exhibition on Condition Monitoring
and Diagnostic Engineering Management
cosponsored by The University of Manchester, COMADEM International,
Society for Machinery Failure Prevention Technology, Holroyd Instruments,
University of Tennessee, and STLE
Manchester, United Kingdom, September 4–6, 2001

National Aeronautics and
Space Administration

Glenn Research Center

September 2001
GEAR DAMAGE DETECTION USING OIL DEBRIS ANALYSIS

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

ABSTRACT
The purpose of this paper was to verify, when using an oil debris sensor, that accumulated mass predicts gear pitting damage and to identify a method to set threshold limits for damaged gears. Oil debris data was collected from 8 experiments with no damage and 8 with pitting damage in the NASA Glenn Spur Gear Fatigue Rig. Oil debris feature analysis was performed on this data. Video images of damage progression were also collected from 6 of the experiments with pitting damage. During each test, data from an oil debris sensor was monitored and recorded for the occurrence of pitting damage. The data measured from the oil debris sensor during experiments with damage and with no damage was used to identify membership functions to build a simple fuzzy logic model. Using fuzzy logic techniques and the oil debris data, threshold limits were defined that discriminate between stages of pitting wear. Results indicate accumulated mass combined with fuzzy logic analysis techniques is a good predictor of pitting damage on spur gears.

KEYWORDS
Gears, Transmissions, Oil debris sensor, Damage detection, Health monitoring, Pitting fatigue

INTRODUCTION
One of NASA’s current goals, the National Aviation Safety Goal, is to reduce the aircraft accident rate by a factor of 5 within 10 years, and by a factor of 10 within 25. One of the leading factors in fatal aircraft accidents is loss of control in flight, which can occur due to flying in severe weather conditions, pilot error, and vehicle/system failure. Focusing on helicopters system failures, an investigation in 1989 found that 32 percent of helicopter accidents due to fatigue failures were caused by damaged engine and transmission components (Astride (1989)). In more recent statistics, of the world total of 192 turbine helicopter accidents in 1999, 28 were directly due to mechanical failures with the most common in the drive train of the gearboxes (Learmont (2000)). A study published in July 1998, in support of the National Aviation Safety Goal, recommended areas most likely to reduce rotorcraft fatalities in the next ten years. The study of 1168 fatal and nonfatal accidents, that occurred from 1990 to 1996, found that after human factors related causes of accidents, the next most frequent cause of accidents were due to various system
and structural failures (Aviation Safety and Security Program, the Helicopter Accident Analysis Team (1998)). Loss of power in-flight caused 26 percent of this type of accident and loss of control in-flight caused 18 percent of this type of accident. The technology area recommended by this study for helicopter accident reduction was helicopter Health and Usage Monitoring Systems (HUMS) capable of predicting imminent equipment failure for on-condition maintenance and more advanced systems capable of warning pilots of impending equipment failures.

Helicopter transmission diagnostics are an important part of a helicopter health monitoring system because helicopters depend on the power train for propulsion, lift, and flight maneuvering. In order to predict transmission failures, the diagnostic tools used in the HUMS must provide real-time performance monitoring of aircraft operating parameters and must demonstrate a high level of reliability to minimize false alarms. Various diagnostic tools exist for diagnosing damage in helicopter transmissions, the most common being vibration. Using vibration data collected from gearbox accelerometers, algorithms are developed to detect when gear damage has occurred (Stewart (1977)); Zakrjseck, Townsend, and Decker (1993)). Oil debris is also used to identify abnormal wear related conditions of transmissions. Oil debris monitoring for gearboxes consists mainly of off-line oil analysis, or plug type chip detectors. And, although not commonly used for gear damage detection, many engines have on-line oil debris sensors for detecting the failure of rolling element bearings. These on-line, inductance type, sensors count the number of particles, their approximate size, then calculate an accumulated mass (Hunt (1993)).

The goal of future HUMS is to increase reliability and decrease false alarms. HUMS are not yet capable of real-time, on-line, health monitoring. Current data collected by HUMS is processed after the flight and is plagued with high false alarm rates and undetected faults. The current fault detection rate of commercially available HUMS through vibration analysis is 60 percent. False warning rates average 1 per hundred flight hours (Stewart (1997)). This is due to a variety of reasons. Vibration based systems require extensive interpretation by trained diagnosticians. Operational effects, can adversely impact the performance of vibration diagnostic parameters and result in false alarms (Dempsey and Zakrjseck (2001)); Campbell, Byington, and Lebold (2000)). Oil debris sensors also require expert analysis of data. False alarms of oil debris technologies are often caused by non-failure debris. This debris can bridge the gap of plug type chip detectors. Inductance type oil debris sensors cannot differentiate between fault and no-fault sourced data (Howard and Reintjes (1999)).

Several companies manufacturer on-line inductance type oil debris sensors that measure debris size and count particles (Hunt (1993)). New oil debris sensors are also being developed that measure debris shape in addition to debris size in which the shape is used to classify the failure mechanism (Howard, et al (1998)). The oil debris sensor used in this analysis was selected for several reasons. The first three reasons were sensor capabilities, availability and researcher experience with this sensor. Results from preliminary research indicate the debris mass measured by the oil debris sensor showed a significant increase when pitting damage began to occur (Dempsey (2000)). This sensor has also been used in aerospace applications for detecting bearing failures in aerospace turbine engines. From the manufacturers experience with rolling element bearing failures, an equation was developed to set warning and alarm threshold limits for damaged bearings based on accumulated mass. Regarding its use in helicopter transmissions, a modified version of this sensor has been developed and installed in an engine nose gearbox and is currently being evaluated for an operational AH-64 (Howe and Muir (1998)). Due to limited access to oil debris data collected by this type of sensor from gear failures, no such equation is available that defines oil debris threshold limits for damaged gears.

The objective of the work reported herein is to first identify the best feature for detecting gear pitting damage from a commercially available on-line oil debris sensor. Then, once the feature is defined,
identify a method to set threshold limits for different levels of damage to gears. The oil debris data analysis will be performed on gear damage data collected from an oil debris monitor in the NASA Glenn Spur Gear Fatigue Rig.

TEST PROCEDURE

Experimental data was recorded from tests performed in the Spur Gear Fatigue Test Rig at NASA Glenn Research Center (Scibbe, Townsend, and Aron (1984)). This rig is capable of loading gears, then running gears until pitting failure is detected. A sketch of the test rig is shown in Figure 1. Torque is applied by a hydraulic loading mechanism that twists 1 slave gear relative to its shaft. The power required to drive the system is only enough to overcome friction losses in the system (Lynwander (1983)). The test gears are standard spur gears having 28 teeth, 8.89 cm pitch diameter, and 0.64 cm face width. The test gears are run offset to provide a narrow effective face width to maximize gear contact stress while maintaining an acceptable bending stress. Offset testing also allows four tests on one pair of gears. Two filters are located downstream of the oil debris monitor to capture the debris after it is measured by the sensor.

Figure 1.—Spur gear fatigue test rig.
Fatigue tests were run in a manner that allows damage to be correlated to the oil debris sensor data. For these tests, run speed was 10 000 rpm and applied torque was 72 and 96 N-m. Prior to collecting test data, the gears were run-in for 1 hr at a torque of 14 N-m. The data measured during this run-in was stored, then the oil debris sensor was reset to zero at the start of the loaded test. Test gears were inspected periodically for damage either manually or using a micro camera connected to a VCR and monitor. The video inspection did not require gearbox cover removal. When damage was found, the damage was documented and correlated to the test data based on a reading number. Reading number is equivalent to minutes and can also be interpreted as mesh cycles equal to reading number times $10^4$. In order to document tooth damage, reference marks were made on the driver and driven gears during installation to identify tooth 1. The mating teeth numbers on the driver and driven gears were then numbered from this reference. Figure 2 identifies the driver and driven gear with the gearbox cover removed.

Data was collected once per minute from oil debris, speed and pressure sensors installed on the test rig using the program ALBERT, Ames-Lewis Basic Experimentation in Real Time, co-developed by NASA Glenn and NASA Ames. Oil debris data was collected using a commercially available oil debris sensor that measures the change in a magnetic field caused by passage of a metal particle where the amplitude of the sensor output signal is proportional to the particle mass. The sensor measures the number of particles, their approximate size (125 to 1000 μm) and calculates an accumulated mass (Howe and Muir (1998)). Shaft speed was measured by an optical sensor once per each shaft revolution. Load pressure was measured using a capacitance pressure transducer.

The principal focus of this research is detection of pitting damage on spur gears. Pitting is a fatigue failure caused by exceeding the surface fatigue limit of the gear material. Pitting occurs when small pieces of material break off from the gear surface, producing pits on the contacting surfaces (Townsend (1991)). Gears are run until pitting occurs on several teeth. Pitting was detected by visual observation through periodic inspections on 2 of the experiments with damage. Pitting was detected by a video inspection.
system on 6 of the experiments with pitting damage. Two levels of pitting were monitored, initial and destructive pitting. Initial pitting is defined as pits less than 0.04 cm diameter and cover less than 25 percent of tooth contact area. Destructive pitting is more severe and defined as pits greater than 0.04 cm diameter and cover greater than 25 percent of tooth contact area. If not detected in time, destructive pitting can lead to a catastrophic transmission failure if the gear teeth crack.

**DISCUSSION OF RESULTS**

The analysis discussed in this section is based on oil debris data collected during 16 experiments, 8 of which pitting damage occurred. The oil debris sensor records counts of particles in bins set at particle size ranges measured in microns. The particle size ranges and average particle size are shown in Table 1. The average particle size for each bin is used to calculate the cumulative mass of debris for the experiment. The shape of the average particle is assumed to be a sphere with a density of approximately 7922 kg/m³.

**TABLE 1**

<table>
<thead>
<tr>
<th>Bin</th>
<th>Bin range, μm</th>
<th>Average, μm</th>
<th>Bin</th>
<th>Bin range, μm</th>
<th>Average, μm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>125–175</td>
<td>150</td>
<td>9</td>
<td>525–575</td>
<td>550</td>
</tr>
<tr>
<td>2</td>
<td>175–225</td>
<td>200</td>
<td>10</td>
<td>575–625</td>
<td>600</td>
</tr>
<tr>
<td>3</td>
<td>225–275</td>
<td>250</td>
<td>11</td>
<td>625–675</td>
<td>650</td>
</tr>
<tr>
<td>4</td>
<td>275–325</td>
<td>300</td>
<td>12</td>
<td>675–725</td>
<td>700</td>
</tr>
<tr>
<td>5</td>
<td>325–375</td>
<td>350</td>
<td>13</td>
<td>725–775</td>
<td>750</td>
</tr>
<tr>
<td>6</td>
<td>375–425</td>
<td>400</td>
<td>14</td>
<td>775–825</td>
<td>800</td>
</tr>
<tr>
<td>7</td>
<td>425–475</td>
<td>450</td>
<td>15</td>
<td>825–900</td>
<td>862.5</td>
</tr>
<tr>
<td>8</td>
<td>475–525</td>
<td>500</td>
<td>16</td>
<td>900–1016</td>
<td>958</td>
</tr>
</tbody>
</table>

**TABLE 2**

Experiments with video inspection

<table>
<thead>
<tr>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
<th>Experiment 4</th>
<th>Experiment 5</th>
<th>Experiment 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rdg#</td>
<td>Mass, mg</td>
<td>Rdg#</td>
<td>Mass, mg</td>
<td>Rdg#</td>
<td>Mass, mg</td>
</tr>
<tr>
<td>60</td>
<td>1.003</td>
<td>1573</td>
<td>3.285</td>
<td>58</td>
<td>0</td>
</tr>
<tr>
<td>120</td>
<td>1.418</td>
<td>2199</td>
<td>8.934</td>
<td>2669</td>
<td>8.69</td>
</tr>
<tr>
<td>1581</td>
<td>5.113</td>
<td>2296</td>
<td>16.267</td>
<td>2857</td>
<td>11.889</td>
</tr>
<tr>
<td>14369</td>
<td>15.475</td>
<td>2065</td>
<td>12.132</td>
<td>9328</td>
<td>11.692</td>
</tr>
<tr>
<td>14430</td>
<td>22.468</td>
<td>2366</td>
<td>13.977</td>
<td>12061</td>
<td>14.365</td>
</tr>
<tr>
<td>14512</td>
<td>24.586</td>
<td>3671</td>
<td>17.361</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14688</td>
<td>28.451</td>
<td>4655</td>
<td>23.12</td>
<td>12368</td>
<td>22.851</td>
</tr>
<tr>
<td>14846</td>
<td>30.686</td>
<td>4863</td>
<td>26.227</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15136</td>
<td>36.108</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Highlighted cells identify reading and mass when destructive pitting was first observed.

Experiments 1 to 6 were performed with the video inspection system installed on the rig. Table 2 lists the reading numbers when inspection was performed and the measured oil debris mass at this reading. The highlighted cells for each experiment identify the reading number and the mass measured when destructive pitting was first observed on one or more teeth. As can be seen from this table, the amount of
mass varied significantly for each experiment. A representative sample of the images obtained from the video damage progression system is shown in Figure 3. The damage progression of tooth 6 on the driver and driven gear for experiment 1 for selected readings is shown in this figure. The damage is only shown on less than half of the tooth because the test gears are run offset to provide a narrow effective face width to maximize gear contact stress.

![Image of gear damage progression]

Figure 3.—Damage progression of driver/driven tooth 6 for experiment 1.

<table>
<thead>
<tr>
<th>Experiment 7</th>
<th>Experiment 8</th>
<th>Pitting Damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rdg#</td>
<td>Mass, mg</td>
<td>Rdg#</td>
</tr>
<tr>
<td>13716</td>
<td>3.381</td>
<td>5181</td>
</tr>
<tr>
<td>5314</td>
<td>19.101</td>
<td></td>
</tr>
</tbody>
</table>

Experiments 7 and 8 were performed with visual inspection. Table 3 lists the reading number when inspection was performed and the measured oil debris mass at this reading. Only initial pitting occurred during experiment 7. During experiment 8, initial pitting was observed at reading 5181 and destructive pitting at reading 5314.

No gear damage occurred during experiments 9 to 16. Oil debris mass measured at test completion is listed in Table 4. At the completion of experiment 10, 5.453 mg of debris was measured, yet no damage occurred. This is more then the debris measured during experiment 7 (3.381 mg) when initial pitting was observed. This and observations made from the data collected during experiments when damage occurred made it obvious that simple linear correlations could not be used to obtain the features for damage levels from the oil debris data.
Prior to discussing methods for feature extraction, it may be beneficial for the reader to get a feel for the amount of debris measured by the oil debris sensor and the amount of damage to one tooth. Applying the definition of destructive pitting, 25 percent of tooth surface contact area for one tooth for these experiments is approximately 0.043 cm². A 0.04 cm diameter pit, assumed spherical in size is equivalent to 0.26 mg oil debris mass. This mass is calculated based on the density used by the sensor software to calculate mass. If 0.04 cm diameter pits densely covered 25 percent of the surface area of 1 tooth, it would be equivalent to approximately 9 mg. Unfortunately, damage distribution is not always densely distributed on 25 percent of a single tooth, but is distributed across many making accurate measures of material removed per tooth extremely difficult.

Several predictive analysis techniques were reviewed to obtain the best feature to predict damage levels from the oil debris sensor. One technique for detecting wear conditions in gear systems is by applying statistical distribution methods to particles collected from lubrication systems (Roylance (1989)). In this reference, mean particle size, variance, kurtosis, and skewness distribution characteristics were calculated from oil debris data collected off-line. The wear activity was determined by the calculated size distribution characteristics. In order to apply this data to on-line oil debris data, calculations were made for each reading number for each bin (Table 1) using the average particle size and the number of particles for each of the sixteen bins. Mean particle size, relative kurtosis, and relative skewness were calculated for each reading for 6 of the experiments with pitting damage. It was not possible, however, to extract a consistent feature that increased in value from the data for all experiments. This may be due to the random nonlinear distribution of the damage progression across all 56 teeth. For this reason a more intelligent feature extraction system was analyzed and will be discussed in the following paragraphs.

When defining an intelligent feature extraction system, the gear states one plans to predict must be defined. Due to the overlap of the accumulated mass features, 3 primary states of the gears were identified: O.K (no gear damage); Inspect (initial pitting); Damage (destructive pitting). The data from Table 2 was plotted in Figure 4. Each plot is labeled with experiment numbers 1 to 6. The triangles on each plot identify the inspection reading number. The triangles circled indicate the reading number when destructive pitting was first observed. The background color indicates the O.K., inspect and damage states. The overlap between the states is also identified with a different background color. The changes in state for each color were defined based on data shown in Tables 2 to 4. The minimum and maximum debris measured during experiments 1 to 6 when destructive pitting was first observed was used to define the upper limit of the inspect scale and the lower limit of the damage scale. The maximum amount of debris measured when no damage occurred (experiment 10) was above the minimum amount of debris measured when initial pitting occurred (experiment 7). This was used as the lower limit of the inspect..

### Table 4

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Rdg#</th>
<th>Mass, mg</th>
<th>Experiment</th>
<th>Rdg#</th>
<th>Mass, mg</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>29866</td>
<td>2.359</td>
<td>13</td>
<td>25259</td>
<td>3.159</td>
</tr>
<tr>
<td>10</td>
<td>20452</td>
<td>5.453</td>
<td>14</td>
<td>5322</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>204</td>
<td>0.418</td>
<td>15</td>
<td>21016</td>
<td>0.125</td>
</tr>
<tr>
<td>12</td>
<td>15654</td>
<td>2.276</td>
<td>16</td>
<td>21446</td>
<td>0.163</td>
</tr>
</tbody>
</table>
state. The next largest mass measured when no damage occurred (experiment 13) was used as the upper limit of the O.K. scale.

Fuzzy logic was used to extract an intelligent feature from the accumulated mass measured by the oil debris sensor. Fuzzy logic was chosen based on the results of several studies to compare the capability of production rules, fuzzy logic and neural nets. One study found fuzzy logic the most robust when monitoring transitional failure data on a gearbox (Hall, Garga, and Stover (1999)). Another study comparing automated reasoning techniques for condition-based maintenance found fuzzy logic more flexible than standard logic by making allowances for unanticipated behavior (McGonigal (1997)). Fuzzy logic applies fuzzy set theory to data, where fuzzy set theory is a theory of classes with unsharp boundaries and the data belongs in a set based on its degree of membership (Zadeh (1992)). The degree of membership can be any value between 0 and 1.

Defining the fuzzy logic model requires inputs (damage detection features), outputs (state of gear), and rules. Inputs are the levels of damage, and outputs are the states of the gears. Membership values were based on the accumulated mass and the amount of damage observed during inspection. Membership values are defined for the 3 levels of damage: damage low, damage medium, and damage high. Using the Mean of the Maximum (MOM) fuzzy logic defuzzification method, the oil debris mass measured during the 6 experiments with pitting damage was input into a simple fuzzy logic model created using commercially available software (Fuzzy Logic Toolbox (1998)). The output of this model is shown on Figure 5. Threshold limits for the accumulated mass are identified for future tests in the Spur Gear Fatigue Test Rig. Results indicate accumulated mass is a good predictor of pitting damage on spur gears and fuzzy logic is a good technique for setting threshold limits that discriminates between states of pitting wear.
CONCLUSIONS

The purpose of this research was to first verify, when using an inductance type, on-line, oil debris sensor, that accumulated mass predicts gear pitting damage. Then, using accumulated mass as the damage feature, identify a method to set threshold limits for damaged gears that discriminates between different levels of pitting damage. In this process, the membership functions for each feature state were defined based on level of damage. From this data, and a simple fuzzy logic model, accumulated mass measured by an oil debris sensor combined with fuzzy logic analysis techniques can be used to predict transmission health. Applying fuzzy logic incorporates decision making into the diagnostic process that improves fault detection and decreases false alarms.

This approach has several benefits over using the accumulated mass and an arbitrary threshold limit for determining if damage has occurred. One is that it eliminates the need for an expert diagnostician to analyze and interpret the data, since the output would be one of 3 states, O.K., Inspect, and Shutdown. Since benign debris may be introduced into the system, due to periodic inspections, setting the lower limit to above this debris level will minimize false alarms. In addition to this, a more advanced system can be designed with logic built-in to minimize these operational effects. Future tests are planned to collect data from gears with initial pitting to better define the inspect region of the model and the severity of gear damage. Tests are planned for gears of different sizes to determine if a relationship can be developed between damage levels and tooth surface contact area, to minimize the need for extensive tests to develop the membership functions for the threshold levels.

REFERENCES


## Gear Damage Detection Using Oil Debris Analysis

### Authors
Paula J. Dempsey

### Performing Organization
National Aeronautics and Space Administration
John H. Glenn Research Center at Lewis Field
Cleveland, Ohio 44135–3191

### Abstract
The purpose of this paper was to verify, when using an oil debris sensor, that accumulated mass predicts gear pitting damage and to identify a method to set threshold limits for damaged gears. Oil debris data was collected from 8 experiments with no damage and 8 with pitting damage in the NASA Glenn Spur Gear Fatigue Rig. Oil debris feature analysis was performed on this data. Video images of damage progression were also collected from 6 of the experiments with pitting damage. During each test, data from an oil debris sensor was monitored and recorded for the occurrence of pitting damage. The data measured from the oil debris sensor during experiments with damage and with no damage was used to identify membership functions to build a simple fuzzy logic model. Using fuzzy logic techniques and the oil debris data, threshold limits were defined that discriminate between stages of pitting wear. Results indicate accumulated mass combined with fuzzy logic analysis techniques is a good predictor of pitting damage on spur gears.