INVESTIGATION OF CONDITION INDICATORS, OPERATIONAL CONDITIONS AND GEAR HEALTH STATE USING DATA MINING TECHNIQUES

Rhea McCaslin and Abdel Bayoumi
University of South Carolina
300 Main Street
Columbia, SC 29208
mccaslir@email.sc.edu

Paula J. Dempsey
NASA Glenn Research Center
Cleveland, Ohio 44135
paula.j.dempsey@nasa.gov

Abstract: Damage progression tests were performed in the NASA Glenn Spiral Bevel Gear Fatigue Rig. Six gear sets with varying levels of tooth damage were tested, and vibration-based gear condition indicators, amount of debris generated, and oil temperatures were measured. The damage state was documented with photographs taken at inspection intervals throughout the test and was quantified with a numerical continuous damage factor. Condition indicator performance was first assessed with traditional methods, and then data-mining methods in the form of a clustering analysis were applied to the operational data and condition indicator data. This analysis was then fed into a decision tree model to predict the gear damage state. Results indicate gear health state can be determined with minimum knowledge of the dataset.

Key words: Health management; spiral bevel gears; data mining

Background: Power train health is a critical part of a rotorcraft health management system since no other air vehicle relies on the propulsion system for propulsion, lift, and maneuverability through a transmission with critical single-load paths. Many rotorcraft are equipped with health and usage monitoring systems (HUMS) that monitor vibration signatures, referred to as condition indicators (CI), to detect transmission dynamic mechanical component health and also monitor operating conditions to track component usage.

Many factors can affect a condition indicator's ability to respond to tooth damage through vibration response. The response of the CI to a specific fault can depend on its method of calculation, operational conditions, and type of failure mode. The fault type (gear or pinion), mode, class, degree, magnitude, how it initiates and progresses, how it changes the signature response at the mesh, and how it interacts with the rig and gear design all affect the measured vibration response. Assessing whether a change in any particular condition indicator response is due solely to a change in damage level, a change in operating condition, or some combination of both can be a challenge.

The objective of this paper is to assess the relationships between gear condition indicator response, operational conditions, and gear health state by applying data-mining techniques to data collected in the NASA Glenn Spiral Bevel Gear Fatigue Test Rig

during initiation and progression of damage to the gear and pinion teeth. Six gear sets, with varying types and amounts of damage, were used for this analysis. Relationships were defined, and patterns were extracted from gear condition indicators, pinion condition indicators, operational parameters, and a damage factor. The operational parameters included amount of generated debris and oil temperatures.

The data-mining analysis stepped through the process from data reduction, assessment of the condition indicators with traditional methods, and application of a clustering analysis to decision tree modeling. Results evaluated the effectiveness and influence of each CI with respect to operational conditions and fault mode.

Test Rig and Instrumentation Description: Tests were performed in the Spiral Bevel Gear Fatigue Test Rig at NASA Glenn Research Center. A detailed description of this test facility is provided in References [1] and [2]. The Spiral Bevel Gear Fatigue Test Rig is illustrated with a cross sectional view in Figure 1. The facility operates as a closed-loop torque regenerative system, where the drive motor only needs enough power to overcome the losses within the system. The load is locked into the loop via a split shaft and a thrust piston that forces a floating helical gear axially into mesh. The 100-hp drive motor supplies the test rig with rotation and overcomes loop losses via v-belts to the axially stationary helical gear.

Two sets of spiral bevel gears, referenced as left and right when facing the gearboxes, are installed in the gearbox. The concave side of the pinion is always in contact with the convex side of the gear on both the left and right side. However, the pinion drives the gear in the normal speed reducer mode on the left side while the pinion acts as a speed increaser on the right side.

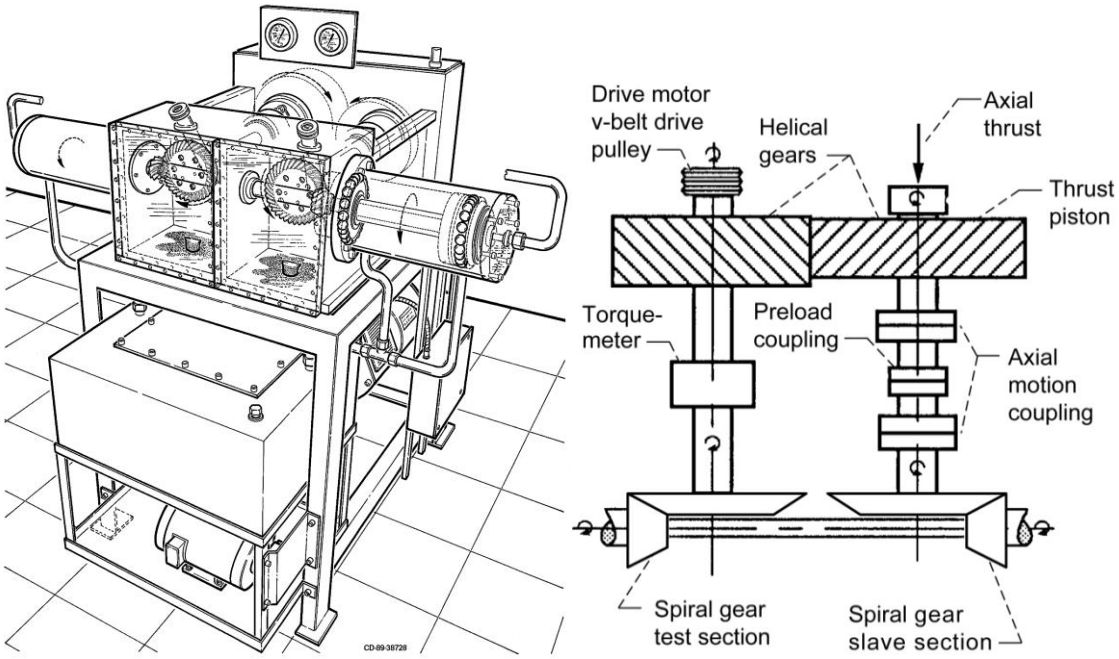


Figure 1: Spiral Bevel Gear Fatigue Test Rig.

Both gear sets are lubricated with oil jets pumped from an oil reservoir using qualified helicopter transmission oil. The oil drains from the gearbox, flows through an inductance type in-line oil debris sensor, then flows past a magnetic chip detector. A strainer and a 3-µm filter capture any debris before returning to the gearbox.

Facility operational parameters, torque, speed, and gearbox oil temperatures were collected every minute with a facility data acquisition (DAQ) system. A commercially available non-contact rotary transformer shaft-mounted torque sensor was used to measure torque during testing. Oil inlet, outlet, and fling-off temperatures were measured with thermocouples. The fling-off temperature is measured inside the gearbox where the oil is flung off of the gears at the out-of-mesh position.

Vibration, oil debris, torque, and speed data were also collected once every minute with the NASA Glenn research DAQ system, the Mechanical Diagnostic System Software (MDSS). The NASA MDSS system acquires, digitizes, and processes the tachometer pulses and accelerometer data. Torque and oil debris sensor data were also recorded every minute with this system. A new experiment is set up when a new gear set is installed on the left side of the test rig.

Oil debris data were collected from an inductance-type oil debris sensor and a magnetic chip detector. The inductance-type oil debris sensor was used to measure the ferrous debris generated during fatigue damage to the gear teeth. The MDSS records the number of particles and their approximate size based on user-defined particle size ranges or bins. Based on the bin configuration, the average particle size for each bin is used to calculate the cumulative mass by assuming the average particle size as the diameter that is spherical in shape and multiplying it by the density of steel. Detailed analysis of the oil debris data generated during testing can be found in Reference [3].

Vibration data were measured with accelerometers installed on the right and left sides of the test rig pinion support housings, radially and vertically with respect to the pinion, as shown in Figure 2. Facing the gearboxes, the left gear set (pinion and gear) and right gear set (pinion and gear) accelerometers were referenced as such in the MDSS system. Speed

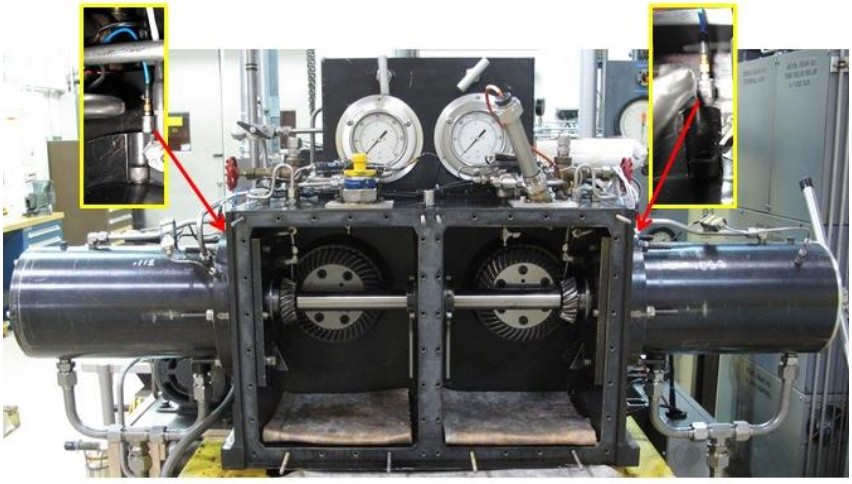


Figure 2: Location of MDSS Accelerometers.

was measured with optical tachometers mounted on the left pinion shaft and left gear shaft to produce a separate once-per-revolution tachometer pulse for the pinion and gears. Additional details on the vibration data collected during these tests can be found in Reference [4].

Gear Design and Test Procedures: The gears tested were designed to represent a rotorcraft drive system gear mesh. To minimize scuffing and force a failure on the left side gear set, several gear sets were super-finished, a process applied to the gears that improves gear surface and extends gear life, and installed on the right side of the gearbox [5]. Surface roughness was improved by a factor of 4 on average after applying this process. The gear set on the right side was unchanged and undamaged during the six gear tests.

The gear sets were designed to an estimated service life to ensure rapid fatigue failures, within rig operating conditions, to limit overall test time. The design code used for these gear sets estimated the fatigue life to be between 100 to 200 hr at operating conditions. Gear sets were ran at a gear speed of 3500 rpm. A speed sweep was performed that determined the run speed was free of rig resonances. At the start of each test, a break-in was performed for a minimum of 1 hr at 4000 in-lb gear torque/3500 rpm gear speed. For some tests, the gear torque was increased to 8000 in-lb for the remainder of the test. For some tests, the rig ran a minimum of 1 hr at 4000 in-lb gear torque, a minimum of 1 hr at 6000 in-lb gear torque, then at 8000 in-lb gear torque for the remainder of the test.

Gear Set Failure Modes: The planned failure mode to be investigated was surface or contact fatigue that occurs when small pieces of material break off from the gear surface because of "exceeding the endurance limit of the material," producing pits on the contacting surfaces due to "surface and subsurface stressors [6]." The failure mode for these tests, defined by American Gear Manufacturers Association (AGMA) standards, was defined by AGMA class (contact fatigue), general mode (macro pitting), and degree (progressive) in which pits are observed in different shapes and sizes greater than 0.04 in diameter [7]. Gear sets were tested until progressive macropitting was observed on a significant area of two or more gear or pinion tooth surfaces. An unanticipated failure mode, scuffing, was also observed on some teeth during testing. Scuffing is a failure mode that causes transfer of metal from one tooth surface to another without any substantial debris generation.

Summary of Gears Tested: A summary of the failure modes observed on the gear teeth between inspections, are shown for each test in Table 1. Each is shaded based on levels of damage; the damage scale is shown with Table 1. The test number is shown in the first row of each test inspection table. Comparable failure modes are paired together for further analysis. For example, for tests L4545R5050 and L1515R5050, the left pinion was damaged at the end of the test. Columns are labeled as inspection number, timeframe, run-time between inspection intervals in minutes and in hours. Components are identified by gear left (GL) and pinion left (PL). Photographs of damaged teeth, taken during inspection intervals, can be viewed in References [3 and 4]. An example of the inspection photos for test L4545R5050 are shown in Figure 3.

Table 1: Failure Modes Observed During Tests.

Inspection no.	Timeframe (min)	Run-time between inspections		GL	PL
		min	hr		
L4545R5050					
1	1 to 76	76	1.3		
2	76 to 324	248	4.1		
3	324 to 1370	1046	17.4		
4	1370 to 2120	750	12.5		1
5	2120 to 2403	283	4.7		2
6	2403 to 2833	430	7.2		2
L1515R5050					
1	1 to 63	63	1.1		
2	63 to 705	642	10.7		1
3	705 to 1022	317	5.3		2
4	1022 to 1291	269	4.5		2
L3030R5050					
1	1 to 70	70	1.2		
2	70 to 1784	1714	28.6		Micro
3	1784 to 3270	1486	24.8		Micro
4	3270 to 4633	1363	22.7	1	Micro
5	4633 to 5359	726	12.1	1	Micro
6	5359 to 5962	603	10.1	2	1
7	5962 to 6037	75	1.3	2	1
L3535R5050					
1	1 to 178	178	3.0		
2	178 to 636	458	7.6		
3	636 to 6276	5640	94.0		
4	6276 to 6818	542	9.0	2	
5	6818 to 7617	799	13.3	2	
6	7617 to 9358	1741	29.0	2	1
7	9358 to 9578	220	3.7	2	1
L2020R5050					
1	1 to 70	70	1.2		
2	70 to 217	147	2.5	s-all	s-all, pit
L4040R5050					
1	1 to 63	63	1.1		
2	63 to 370	307	5.1	s-all	s-all

Damage Scale
No damage
Micropitting or edge wear
Scuffing
1 tooth macropitting
2 or more teeth macropitting

Test L4545R5050

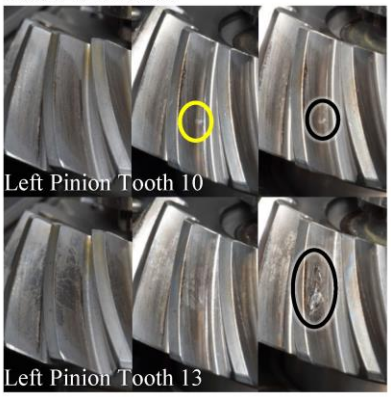


Figure 3: Tests L4545R5050 Tooth Damage Photographs for Left Pinion.

Condition Indicators and Operational Data: Table 2 summarizes run-time, average torque, and average left oil inlet temperature (LOI) of the gear sets tested. Table 3 lists the data available for analysis.

Vibration data were collected at sample rates that provided sufficient vibration data for calculating time-synchronous-averaged data (TSA). TSA refers to techniques for averaging vibration signals over several revolutions of the shaft, in the time domain, to improve the signal-to-noise ratio [8]. From the TSA data, several gear condition indicators were calculated for this analysis: figure of merit 4 (FM4), root mean square (RMS), sideband index (SI) and M8A [8]. FM4, RMS and M8A are common time-domain, statistically based, vibration algorithms used in commercial HUMS [9].

Table 2: Summary of Operational Parameters.

Gear set	Minutes	Hours	Average torque (in-lb)	Average LOI (°F)
L4545R5050	2120	35	6640	250
L1515R5050	705	12	7500	250
L3030R5050	4566	76	7882	250
L3535R5050	6818	114	7782	190
L2020R5050	217	4	6638	250
L4040R5050	370	6	6815	190

Table 3: Data Available for Analysis.

Operational Parameters: Left Gear Box Only
Run Time (min)
Torque (in-lb)
Left oil inlet temperature (LOI)
Left fling-off temperature (LFO)
Left oil outlet temperature (LOO)
Condition Indicators for Gear (GL) and Pinion (PL)
Debris (mg)
GL RMS CI
PL RMS CI
GL FM4 CI
PL FM4 CI
GL SI1 CI
PL SI1 CI
GL SI3 CI
PL SI3 CI
GL M8A CI
PL M8A CI
Damage state
PL or GL damage state/scale

The SI is another CI used to indicate gear tooth damage [10]. SI is a frequency-domain-based CI. The CI value is an average value of sideband amplitudes about the fundamental gear mesh (GM) frequency. All gears generate a dominant GM frequency in the vibration signature due to each tooth impacting the gear it is driving as the pinion and gear mesh. One (SI1) and three (SI3) sidebands were included in the calculation. Since the gear set on the right side was unchanged and undamaged during the six gear tests, only CIs for the GL and GP were evaluated.

Prior to performing a data-mining analysis, an initial data preparation step is used to determine if the number of variables provide redundant information and can be reduced. One approach is to apply a correlation analysis to the variables, to determine if any had strong linear correlations and could possibly be removed. Pearson Correlation Coefficients r were calculated for the variables [11]. A correlation matrix was generated for each test, and those with values of $r \geq 0.8$ were assessed for removal.

Since correlation analysis requires continuous values, damage state was quantified as a "damage factor," with a numerical continuous damage value assigned to the damage scale. Per the damage scale in Table 1, no damage is indicated by 0; micropitting, by 0.25; scuffing, by 0.25; macropitting of one tooth, by 0.5; and macropitting of two or more teeth, by 1. During intervals where the state transitioned, a sloped line was generated from start to end of the inspection interval.

Strong correlations were observed between FM4 and M8A for both the gear and pinion, SI1 and SI3 for just the pinion, debris and run-time, and all three oil temperatures. This indicated that these variables provided similar information and could be removed from the analysis. Based on this analysis, M8A, left oil outlet (LOO) temperature, and run-time were removed from the model. Torque was also removed because torque had little variation during testing.

Representative plots generated during every test for FM4, RMS, and SI3 for the GL and PL, pinion and gear torque, and left oil inlet (LOI), fling-off (LFO) and outlet temperatures (LFO), are shown for test L4545R5050 in Figure 4. Plots for the other five tests can be viewed in Reference [4]. Note that the triangles on the x-axis for all the plots correlate to inspections and are color coded per observed damage during inspections. Mass generated from the oil debris accumulated during each test was measured and is plotted in Figure 5. The rate and amount of debris mass generated varied for each test. Additional information on the oil debris data can be found in Reference [3].

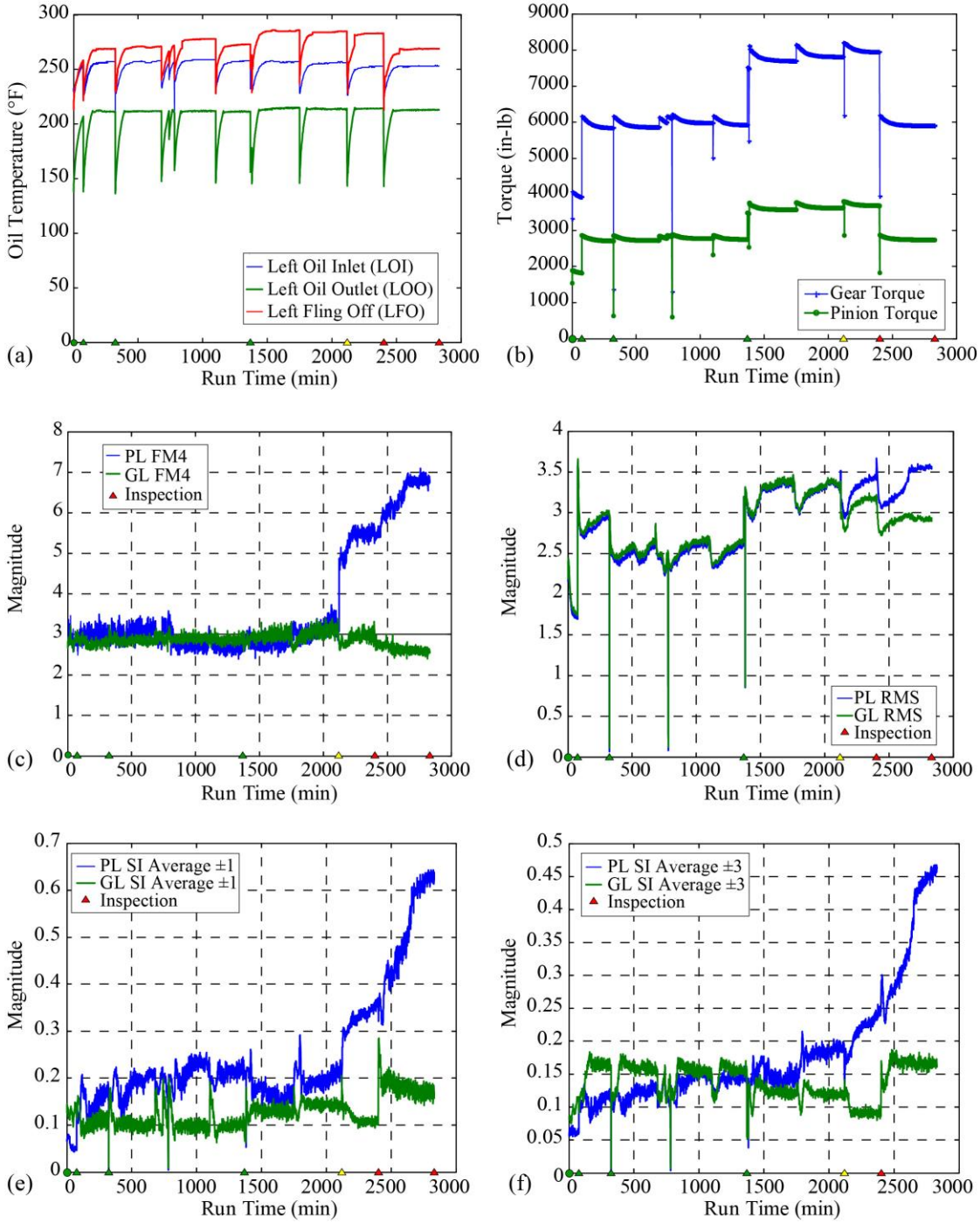


Figure 4: Plots of Operational and Condition Indicators During Test L4545R5050. (a) Left-side oil temperatures; (b) Gear and pinion torque; (c) Left pinion (PL) and left gear (GL) FM4; (d) PL and GL RMS; (e) PL and GL SI1; (f) PL and GL SI3.

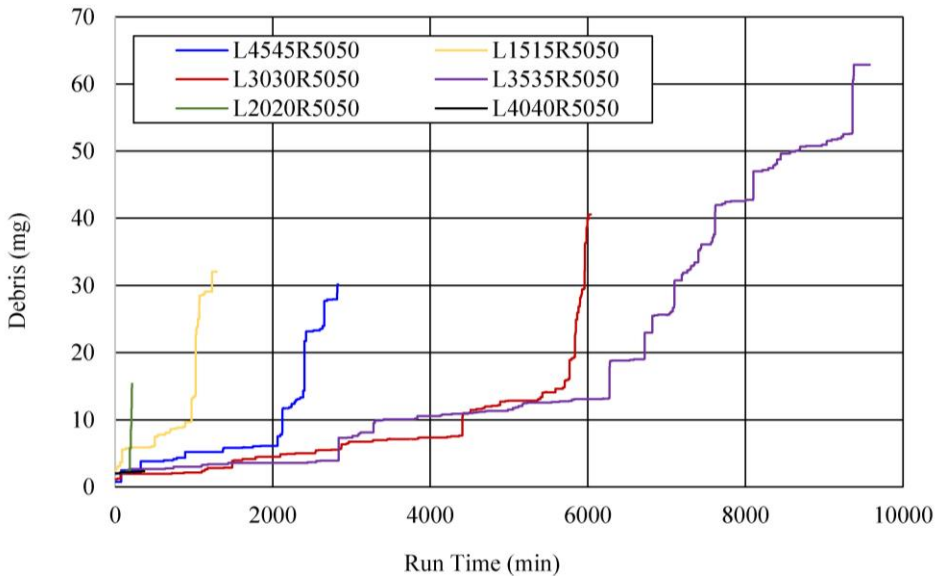


Figure 5: Oil Debris Accumulated Mass Generated During Each Test.

Condition Indicator Performance: Prior to applying data mining to the dataset, the performance of the condition indicators were assessed with traditional methods using discrete thresholds. The true positive rate (TPR) detection, false positive rate (FPR) false alarms and accuracy of the condition indicators were calculated for each CI. Table 4 steps through the process for test L4545R5050. The calculations are limited to the damage mode of macropitting on two or more teeth. The no-damage state data used is in the run time, listed from 1 to 1370, in Table 1. The damage state data used is in the time frame listed in black from 2120 to 2833. The data within the dark gray region will not be used since this is not the damage mode of interest and this region was also the transition region, when the damage state went from white to black. In Table 4, the first column identifies the number of cases, or data points, available within each inspection interval for each gear state. The second column identifies the observed state at that inspection. For this test, only the left pinion teeth had damage. The next four columns identify the number of times the CI values exceeded the thresholds within the inspection interval. The black arrows shown where the data came from to make the calculations for PL RMS.

Reviewing Table 4, the only CI that achieved a detection rate of >90 percent at a threshold of 4.5 was FM4. However, the false-alarm rate was 0 percent for all four CIs. That is why the accuracy is >70 percent for each CI. Although the detection rate was poor, the false alarm rate was good. The detection rate and false alarm rate are equally weighted for these calculations.

Table 5 lists the TPR, FPR, and accuracy of condition indicators RMS, FM4, SI1, and SI3. The false positive rate was <0.10 for all CIs for all tests. CI damage detection and accuracy rate was >0.90 for test L4545R5050 (PL FM4), test L3535R5050 (PL SI1 and PL SI3), and test L2020R5050 (PL RMS and GL RMS). CI damage detection rate and accuracy rate was >0.75 for test L3030R5050 (PL SI1, PL SI3, and GL FM4), test L3535R5050 (GL FM4), and test L2020R5050 (PL SI1). Reviewing Table 5, the CI with the highest detection rate and lowest false alarm rates across all tests was pinion SI1.

Table 4.—Calculating TPR, FPR, and Accuracy for Rig Test L4545R5050.

L4545R5050	Thresholds	3.5	4.5	0.5	0.35
Cases	Pinion Left CI	PL RMS CI	PL FM4 CI	PL SI1 CI	PL SI3 CI
76	No damage	0	0	0	0
248	No damage	4	0	0	0
1046	No damage	0	0	0	0
750		0	0	0	0
283	≥2 teeth macropitting	3	279	0	0
430	≥2 teeth macropitting	182	430	215	206
		PL RMS CI	PL FM4 CI	PL SI1 CI	PL SI3 CI
a	True positive (TP)	185	709	215	206
a+c	Number of faults	713	713	713	713
a/(a+c)	True positive rate (TPR)	0.26	0.99	0.30	0.29
b	False positive (FP)	4	0	0	0
b+d	Number of healthy	1370	1370	1370	1370
b/(b+d)	False positive rate (FPR)	0.00	0.00	0.00	0.00
d	True negative (TN)	1366	1370	1370	1370
a+d		1551	2079	1585	1576
a+b+c+d	Total number of cases	2083	2083	2083	2083
(a+d)/(a+b+c+d)	**Accuracy**	0.74	**1.00**	0.76	0.76

Table 5.—TPR, FPR, and Accuracy for Rig Tests.

Test		PL RMS CI	PL FM4 CI	PL SI1 CI	PL SI3 CI	GL RMS CI	GL FM4 CI	GL SI1 CI	GL SI3 CI
L4545R5050	TPR	0.26	0.99	0.30	0.29				
	FPR	0.00	0.00	0.00	0.00				
	Accuracy	0.74	1.00	0.76	0.76				
L1515R5050	TPR	0.00	0.43	0.00	0.00				
	FPR	0.00	0.00	0.00	0.00				
	Accuracy	0.10	0.49	0.10	0.10				
L3030R5050	TPR	0.35	0.00	0.87	0.83	0.05	0.76	0.00	0.00
	FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Accuracy	0.41	0.09	0.88	0.85	0.14	0.78	0.09	0.09
L3535R5050	TPR	0.08	0.00	0.97	0.99	0.08	0.78	0.05	0.00
	FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Accuracy	0.72	0.69	0.99	0.99	0.72	0.93	0.71	0.69
L2020R5050	TPR	0.97	0.00	0.86	0.00	0.97	0.00	0.00	0.00
	FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Accuracy	0.98	0.32	0.91	0.32	0.98	0.32	0.32	0.32
L4040R5050	TPR	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.00
	FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Accuracy	0.20	0.17	0.17	0.17	0.20	0.17	0.17	0.17

Data mining analysis: Data mining is a process that extracts patterns, correlations, and other useful information from datasets using methods taken from statistics and machine learning [12]. The results are often used to make predictions about future outcomes or

responses. The approach can be supervised or unsupervised. Supervised learning uses existing labeled data as inputs and responses to build, train, and test the model. In this type of learning, the inputs into the model are seen as a pair made up of the existing data and a target output value. Unsupervised learning does not require that the data be labeled and have a target value [13]. The objective of this data mining analysis is to build a model to predict the damage state by identifying informative parameters and finding latent relationships in the data. The inputs to the model are listed in Table 3.

The initial approach was to apply a clustering analysis, where groups of similar data are separated into clusters based on some measure of similarity within the cluster, to find a new pattern within this dataset. The expectation was that the response of the inputs during different gear health states would respond in a similar fashion; that is, different clusters could be related to the varying gear health states.

The data were separated into five clusters using the k-means algorithm, where k identifies the number of clusters. The k-means clustering is performed by minimizing the sum-of-squares distances between data and the corresponding cluster centroid. Data with the same features are grouped into the same cluster, with the minimum amount of spread within the cluster and no overlap between clusters. The clustering is performed by minimizing the sum of the squared Euclidean distance between the data and its cluster centroid.

One of the disadvantages of the k-means algorithm is that k is not given, and predicting the optimal number of clusters can be difficult. When choosing the number of clusters, there needs to be balance between the accuracy of a cluster and the information gained from the addition of that cluster. One method is to plot and look at the variance of using different cluster numbers. This was done for the six tests. In Figure 6, a representative plot of the "within sum of squares value" for 10 clusters is shown for test L2020R5050. The within sum of squares value is the sum of squared distances of each data point in the cluster to the cluster mean. The first few clusters give the highest amount of information which is shown in the plot. Five was chosen for the number of clusters because at this point the information gained had leveled out.

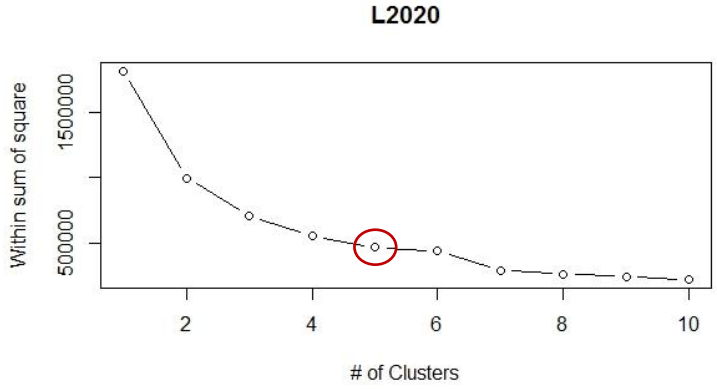


Figure 6: Within Sum of Squares Plot for Clusters.

The k-means type of clustering algorithm separates the data into a definite cluster, with no overlap, based on their location and distance with respect to other data points. The k-means clustering algorithm employs an iterative process, and each iteration contains two steps. The first step is to randomly generate the initial means of the clusters. Then, in step

2, the clusters are formed by associating the mean of each observation with nearest cluster mean. After this, in step 3, the centroid of each cluster becomes the new cluster mean. Steps 2 and 3 are then repeated until no further change in the clusters occurs [13]. For the k-means analysis, the data were split into a testing and training set. The data were separated sequentially because of the time dependency of the dataset, and 1/3 was set aside for testing and 2/3 was used for training the model.

The k-means algorithm is based on means, and the inputs can be ranked in order of importance relative to a specific target, which in this analysis is the gear damage state. The k-means analysis was applied to the data for the six tests [14]. Tables 6 through 8 are tables of the clustered data, from the test dataset, with inputs ranked in order of importance. The size indicates the number of points in each cluster. Their importance is quantified from 0 to 1 in the first column, with 1 being the most important input. The mean value of the inputs within each cluster are also listed. This information illustrates the transitions between clusters and the means of the parameters.

For the two tests with only pinion damage (L4545 and L1515), debris and pinion FM4 were the best predictors, defined by predictor performance values ≥0.8 that were common to both tests. This is shown in Table 6. For the two tests with pinion and gear damage (L3030 and L3535), debris, gear FM4, and gear and pinion SI1 and SI3 were the best predictors of damage. This is shown in Table 7.

Table 6: Tests L4545R5050 and L1515R5050 Clusters.

L4545R5050						**L1515R5050**					
	Cluster						Cluster				
	5	3	1	2	4		5	2	1	4	3
Size	37.8% (356)	32.1% (303)	13.1% (124)	13.1% (124)	3.8% (36)	Size	64.1% (261)	20.9% (85)	12.0% (49)	2.7% (11)	0.2% (1)
Inputs 1.00	Debris (mg) 4.20	Debris (mg) 7.76	Debris (mg) 3.70	Debris (mg) 25.63	Debris (mg) 17.78	Inputs 1.00	Debris (mg) 7.66	Debris (mg) 29.04	Debris (mg) 6.72	Debris (mg) 6.82	Debris (mg) 5.87
1.00	LFO (°F) 271.60	LFO (°F) 283.36	LFO (°F) 251.53	LFO (°F) 267.74	LFO (°F) 252.31	0.88	GL FM4 CI 2.69	GL FM4 CI 2.67	GL FM4 CI 2.86	GL FM4 CI 3.02	GL FM4 CI 8.61
1.00	PL SI1 CI 0.19	PL SI1 CI 0.22	PL SI1 CI 0.16	PL SI1 CI 0.53	PL SI1 CI 0.33	0.86	PL FM4 CI 2.92	PL FM4 CI 5.29	PL FM4 CI 3.24	PL FM4 CI 3.09	PL FM4 CI 3.68
1.00	PL SI3 CI 0.13	PL SI3 CI 0.18	PL SI3 CI 0.11	PL SI3 CI 0.37	PL SI3 CI 0.22	0.67	GL SI3 CI 0.18	GL SI3 CI 0.18	GL SI3 CI 0.14	GL SI3 CI 0.05	GL SI3 CI 0.03
0.96	PL FM4 CI 2.88	PL FM4 CI 3.57	PL FM4 CI 2.97	PL FM4 CI 6.49	PL FM4 CI 5.24	0.65	LFO (°F) 276.37	LFO (°F) 272.85	LFO (°F) 198.95	LFO (°F) 157.72	LFO (°F) 276.20
0.94	GL SI3 CI 0.16	GL SI3 CI 0.11	GL SI3 CI 0.11	GL SI3 CI 0.17	GL SI3 CI 0.13	0.63	PL SI3 CI 0.17	PL SI3 CI 0.22	PL SI3 CI 0.13	PL SI3 CI 0.09	PL SI3 CI 0.03
0.86	LOI (°F) 256.55	LOI (°F) 254.45	LOI (°F) 240.36	LOI (°F) 252.67	LOI (°F) 239.06	0.60	LOI (°F) 255.61	LOI (°F) 251.66	LOI (°F) 172.48	LOI (°F) 133.45	LOI (°F) 257.00
0.85	PL RMS CI 2.56	PL RMS CI 3.27	PL RMS CI 2.55	PL RMS CI 3.37	PL RMS CI 3.13	0.41	PL RMS CI 1.79	PL RMS CI 2.00	PL RMS CI 1.32	PL RMS CI 2.35	PL RMS CI 0.77
0.77	GL SI1 CI 0.10	GL SI1 CI 0.13	GL SI1 CI 0.13	GL SI1 CI 0.18	GL SI1 CI 0.18	0.40	PL SI1 CI 0.34	PL SI1 CI 0.37	PL SI1 CI 0.25	PL SI1 CI 0.19	PL SI1 CI 0.02
0.65	GL RMS CI 2.62	GL RMS CI 3.23	GL RMS CI 2.58	GL RMS CI 2.92	GL RMS CI 2.85	0.35	GL RMS CI 1.83	GL RMS CI 1.91	GL RMS CI 1.40	GL RMS CI 2.39	GL RMS CI 2.31
0.45	GL FM4 CI 2.85	GL FM4 CI 2.99	GL FM4 CI 2.87	GL FM4 CI 2.63	GL FM4 CI 2.77	0.31	GL SI1 CI 0.19	GL SI1 CI 0.20	GL SI1 CI 0.24	GL SI1 CI 0.09	GL SI1 CI 0.02

Table 7: Tests L3030R5050 and L3535R5050 Clusters.

L3030R5050						**L3535R5050**					
Cluster						Cluster					
	5	3	1	2	4		5	2	1	4	3
Size	67.4% (1356)	20.2% (407)	6.5% (131)	4.6% (92)	1.2% (25)	Size	36.8% (1171)	22.9% (729)	16.6% (529)	13.1% (416)	10.6% (337)
Inputs 1.00	Debris (mg) 5.12	Debris (mg) 14.91	Debris (mg) 3.51	Debris (mg) 9.11	Debris (mg) 39.01	Inputs 1.00	Debris (mg) 7.06	Debris (mg) 15.17	Debris (mg) 51.26	Debris (mg) 2.93	Debris (mg) 35.96
1.00	GL SI3 CI 0.15	GL SI3 CI 0.18	GL SI3 CI 0.12	GL SI3 CI 0.14	GL SI3 CI 0.12	1.00	GL FM4 CI 2.69	GL FM4 CI 2.69	GL FM4 CI 5.92	GL FM4 CI 2.85	GL FM4 CI 4.77
1.00	LFO (°F) 279.03	LFO (°F) 284.78	LFO (°F) 255.13	LFO (°F) 257.09	LFO (°F) 179.37	1.00	GL RMS CI 2.01	GL RMS CI 2.08	GL RMS CI 3.28	GL RMS CI 1.96	GL RMS CI 2.67
1.00	LOI (°F) 253.93	LOI (°F) 254.59	LOI (°F) 240.12	LOI (°F) 235.89	LOI (°F) 140.34	1.00	GL SI1 CI 0.21	GL SI1 CI 0.35	GL SI1 CI 0.39	GL SI1 CI 0.13	GL SI1 CI 0.38
1.00	PL SI1 CI 0.28	PL SI1 CI 0.48	PL SI1 CI 0.19	PL SI1 CI 0.43	PL SI1 CI 0.41	1.00	GL SI3 CI 0.14	GL SI3 CI 0.21	GL SI3 CI 0.25	GL SI3 CI 0.12	GL SI3 CI 0.25
1.00	PL SI3 CI 0.19	PL SI3 CI 0.34	PL SI3 CI 0.14	PL SI3 CI 0.32	PL SI3 CI 0.28	1.00	PL FM4 CI 2.94	PL FM4 CI 3.21	PL FM4 CI 3.43	PL FM4 CI 2.94	PL FM4 CI 3.39
0.99	GL SI1 CI 0.12	GL SI1 CI 0.18	GL SI1 CI 0.17	GL SI1 CI 0.22	GL SI1 CI 0.18	1.00	PL RMS CI 2.03	PL RMS CI 2.22	PL RMS CI 3.28	PL RMS CI 1.90	PL RMS CI 2.77
0.96	GL FM4 CI 2.92	GL FM4 CI 4.35	GL FM4 CI 2.92	GL FM4 CI 3.60	GL FM4 CI 6.96	1.00	PL SI1 CI 0.27	PL SI1 CI 0.42	PL SI1 CI 0.68	PL SI1 CI 0.12	PL SI1 CI 0.57
0.44	PL FM4 CI 2.77	PL FM4 CI 2.50	PL FM4 CI 3.04	PL FM4 CI 2.52	PL FM4 CI 2.55	1.00	PL SI3 CI 0.15	PL SI3 CI 0.31	PL SI3 CI 0.48	PL SI3 CI 0.08	PL SI3 CI 0.41
0.41	GL RMS CI 3.39	GL RMS CI 3.23	GL RMS CI 3.00	GL RMS CI 3.02	GL RMS CI 3.02	0.36	LFO (°F) 233.20	LFO (°F) 236.53	LFO (°F) 238.47	LFO (°F) 222.12	LFO (°F) 235.00
0.36	PL RMS CI 3.42	PL RMS CI 3.48	PL RMS CI 3.00	PL RMS CI 3.25	PL RMS CI 3.40	0.12	LOI (°F) 189.13	LOI (°F) 189.17	LOI (°F) 189.80	LOI (°F) 184.67	LOI (°F) 189.64

Table 8: Tests L2020R5050 and L4040R5050 Clusters.

L2020R5050						**L4040R5050**					
Cluster						Cluster					
	5	3	1	2	4		5	2	1	4	3
Size	41.7% (30)	22.2% (16)	18.1% (13)	9.7% (7)	8.3% (6)	Size	43.1% (53)	39.8% (49)	12.2% (15)	4.1% (5)	0.8% (1)
Inputs 1.00	PL RMS CI 6.32	PL RMS CI 2.04	PL RMS CI 4.28	PL RMS CI 2.71	PL RMS CI 5.83	Inputs 1.00	LFO (°F) 232.99	LFO (°F) 220.78	LFO (°F) 198.04	LFO (°F) 184.71	LFO (°F) 158.17
0.98	PL SI1 CI 0.58	PL SI1 CI 0.21	PL SI1 CI 0.49	PL SI1 CI 0.20	PL SI1 CI 0.57	0.96	LOI (°F) 188.32	LOI (°F) 172.73	LOI (°F) 170.43	LOI (°F) 155.83	LOI (°F) 158.17
0.98	GL RMS CI 6.22	GL RMS CI 2.07	GL RMS CI 4.20	GL RMS CI 2.76	GL RMS CI 5.75	0.93	PL SI1 CI 0.33	PL SI1 CI 0.34	PL SI1 CI 0.14	PL SI1 CI 0.11	PL SI1 CI 0.18
0.90	PL SI3 CI 0.30	PL SI3 CI 0.10	PL SI3 CI 0.25	PL SI3 CI 0.10	PL SI3 CI 0.30	0.74	PL SI3 CI 0.18	PL SI3 CI 0.17	PL SI3 CI 0.11	PL SI3 CI 0.09	PL SI3 CI 0.11
0.58	GL SI1 CI 0.22	GL SI1 CI 0.02	GL SI1 CI 0.23	GL SI1 CI 0.04	GL SI1 CI 0.30	0.74	GL SI1 CI 0.11	GL SI1 CI 0.13	GL SI1 CI 0.29	GL SI1 CI 0.39	GL SI1 CI 0.10
0.55	LOI (°F) 255.39	LOI (°F) 247.64	LOI (°F) 242.46	LOI (°F) 234.09	LOI (°F) 253.68	0.65	GL SI3 CI 0.07	GL SI3 CI 0.07	GL SI3 CI 0.14	GL SI3 CI 0.17	GL SI3 CI 0.05
0.37	LFO (°F) 277.47	LFO (°F) 259.01	LFO (°F) 274.84	LFO (°F) 242.34	LFO (°F) 276.02	0.50	PL RMS CI 3.09	PL RMS CI 2.37	PL RMS CI 1.97	PL RMS CI 1.53	PL RMS CI 1.20
0.32	GL SI3 CI 0.10	GL SI3 CI 0.08	GL SI3 CI 0.10	GL SI3 CI 0.04	GL SI3 CI 0.20	0.49	GL RMS CI 3.10	GL RMS CI 2.38	GL RMS CI 2.00	GL RMS CI 1.60	GL RMS CI 1.29
0.29	GL FM4 CI 3.07	GL FM4 CI 2.90	GL FM4 CI 3.42	GL FM4 CI 2.90	GL FM4 CI 3.07	0.49	Debris (mg) 2.25	Debris (mg) 2.19	Debris (mg) 2.03	Debris (mg) 2.03	Debris (mg) 2.14
0.12	PL FM4 CI 2.96	PL FM4 CI 3.08	PL FM4 CI 2.81	PL FM4 CI 3.31	PL FM4 CI 3.20	0.35	PL FM4 CI 3.20	PL FM4 CI 2.79	PL FM4 CI 2.81	PL FM4 CI 2.68	PL FM4 CI 2.73
0.05	Debris (mg) 4.27	Debris (mg) 2.00	Debris (mg) 2.00	Debris (mg) 2.00	Debris (mg) 2.00	0.17	GL FM4 CI 3.35	GL FM4 CI 3.11	GL FM4 CI 3.16	GL FM4 CI 2.81	GL FM4 CI 2.72

For the tests with scuffing damage, L2020 had scuffing and pitting, while L4040 had scuffing only damage. For these two tests, the importance of debris as a predictor of tooth damage decreased, as shown in Table 8. For both tests, pinion SI1 was a good predictor. For the test with scuffing only, L4040, the fling-off oil temperature was an important predictor. This aligns with previous experimental work that has shown that a spike in fling-off temperature is a good predictor of scuffing damage and that debris is not a good predictor of scuffing damage [3, 4].

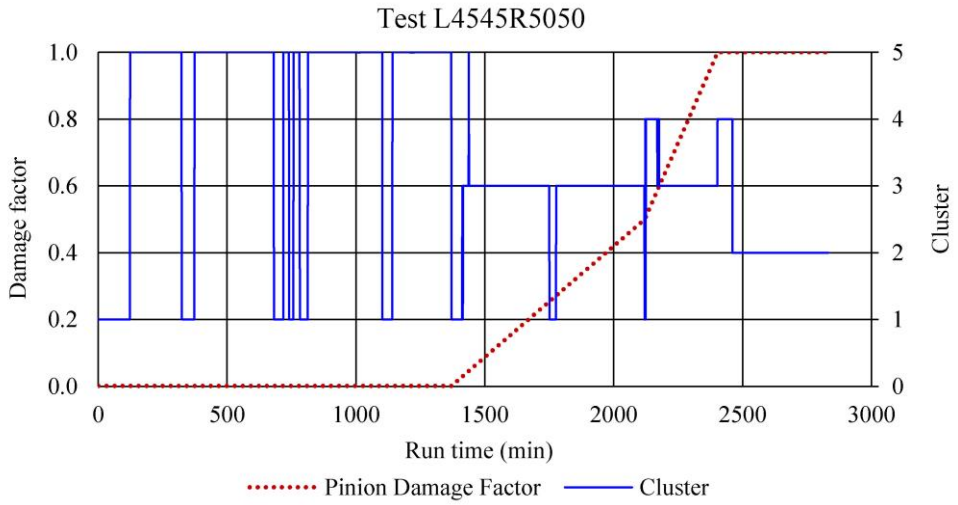
To illustrate this further, tables were created that summarize the clusters and values associated with the best predictors across the damage state. Plots were also generated, for each test, that compare the damage factor during test progression to how those data were clustered. The tables and figures are combined and shown together in Figures 7 through 9. The damage factor is the numerical continuous damage value assigned to the damage observed during inspections. Note that at rig starts, the clusters typically oscillate between two cluster groups.

For the two tests with only pinion damage (L4545 and L1515), Figure 7 summarizes the clusters and values associated with the best predictors across the damage state. These results align well with the performance metrics calculated in Table 5. Cluster 2 was a good indicator for the damage state of two or more pinion teeth.

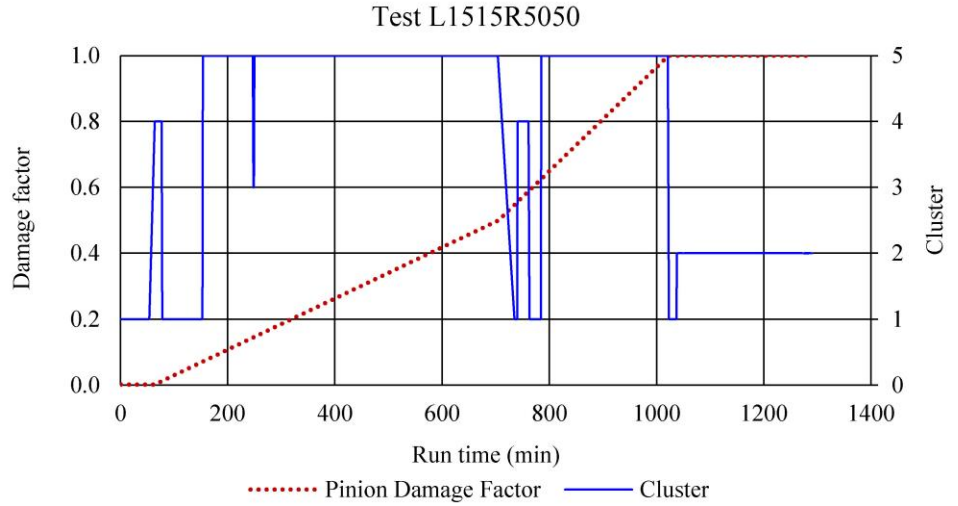
For the two tests with pinion and gear damage (L3030 and L3535), Figure 8 summarizes the clusters and values associated with the best predictors across the damage state. These results also align well with the performance metrics calculated in Table 5. Cluster 3 was a good indicator for the two or more gear teeth damage state with combined pinion damage for test L3030. Cluster 2 and 5 were good indicators for the two or more gear teeth damage state with combined pinion damage for test L3030.

In test L2020, the damage state can be divided into two states (no damage and scuffing/pitting). Figure 9 summarizes the clusters and mean cluster values associated with the best predictors across the damage states. Clusters 1 and 4 describe the state with no damage, and clusters 2, 3, and 5 describe the scuffing/pitting state. In test L4040, the damage state can be divided into two states (no damage and scuffing). Figure 10 summarizes the clusters and mean cluster values associated with the best predictors across the damage states. The clusters identified in this test can be assigned to a damage state. Clusters 1 and 5 describe the no-damage state, and clusters 2, 3, and 4 describe the scuffing-only state.

What this analysis shows is that with very little knowledge of the dataset, the parameters appeared to cluster differently based on gear health state. The results of the clustering analysis can be used to feed into a supervised learning classification model. Using this approach, observed data and known responses can be used to build the predictive model, or classifier, with discrete response variables [15].

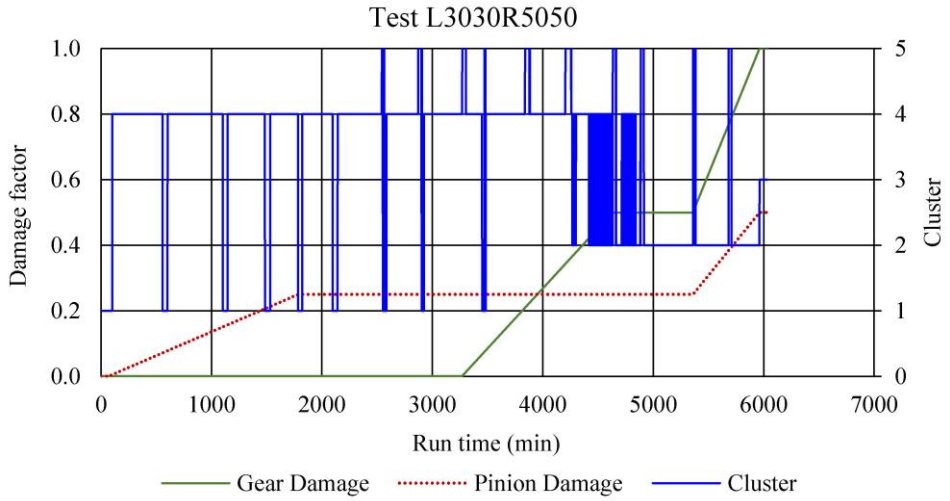


Damage state	Debris mean average	Debris mean range	Pinion FM4 mean average	Pinion FM4 mean range
No damage (clusters 1, 5)	3.95	3.70 to 4.20	2.93	2.88 to 2.97
1 tooth (clusters 3, 4)	12.77	7.76 to 17.78	4.41	3.57 to 5.24
2 or more teeth (cluster 2)	25.63	25.63 to 25.63	6.49	6.49 to 6.49

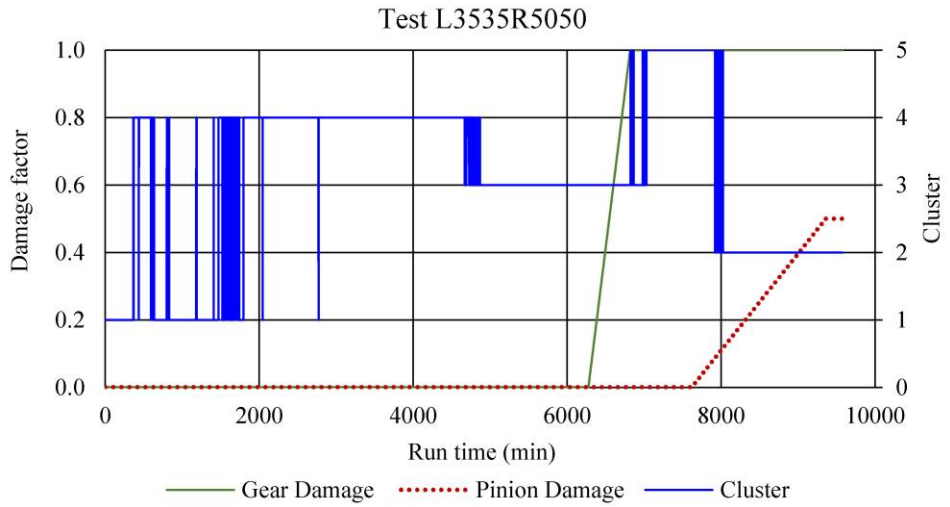


Damage state	Debris mean average	Debris mean range	Pinion FM4 mean average	Pinion FM4 mean range
No damage (clusters 1, 4)	6.77	6.72 to 6.82	3.17	3.09 to 3.24
1 tooth (cluster 5)	7.66	7.66 to 7.66	2.92	2.92 to 2.92
2 or more teeth (cluster 2)	29.04	29.04 to 29.04	5.29	5.29 to 5.29

Figure 7: Comparing Damage Factor, Clusters, and CI Values for Tests L4545 and L1515.

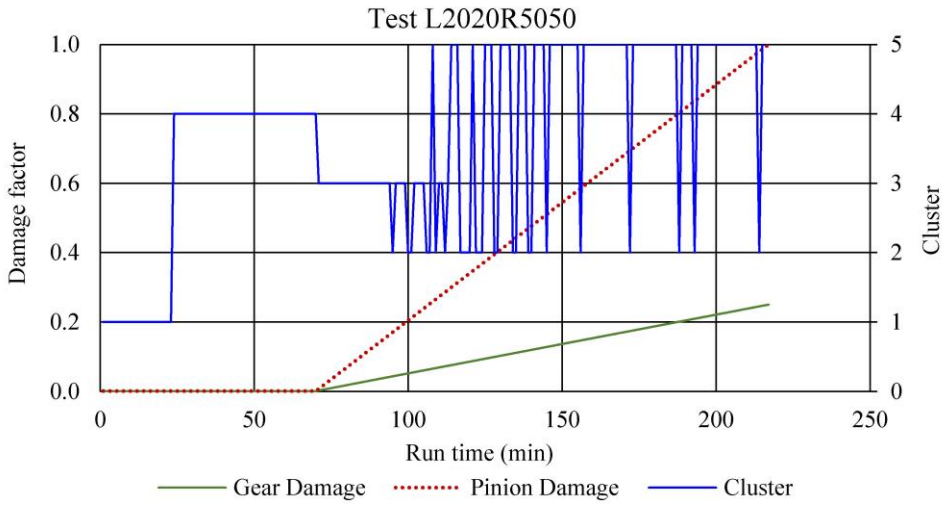


Damage state	Debris mean average	Gear SI3 mean average	Gear SI1 mean average	Pinion SI3 mean average	Pinion SI1 mean average
No damage (clusters 1, 4)	4.32	0.14	0.15	0.17	0.24
1 tooth (cluster 2)	14.91	0.18	0.18	0.28	0.48
2 or more teeth (cluster 3)	39.01	0.12	0.18	0.34	0.41

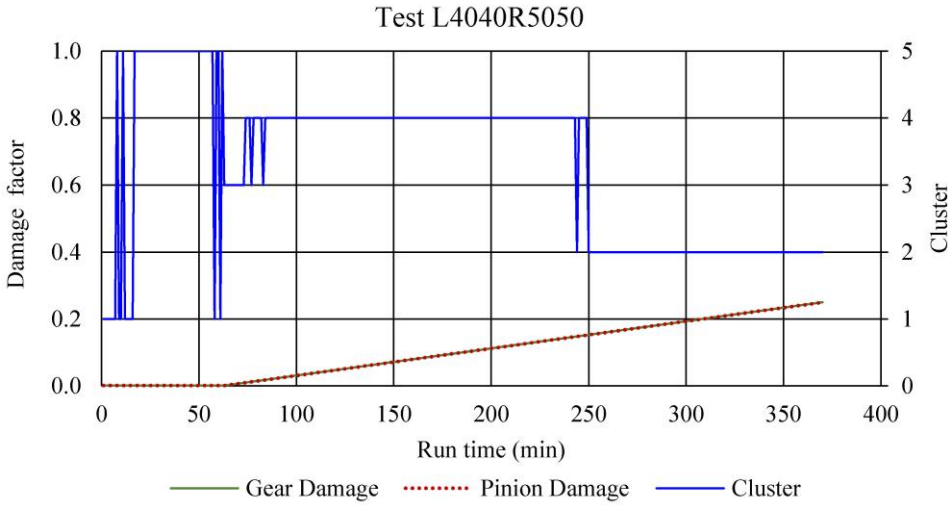


Damage state	Debris mean average	Gear FM4 mean average	Gear SI1 mean average	Gear SI3 mean average	Pinion SI1 mean average	Pinion SI3 mean average
No damage (clusters 1, 3, 4)	8.34	2.74	0.23	0.16	0.27	0.18
Two or more teeth (clusters 2, 5)	43.61	5.35	0.39	0.25	0.63	0.45

Figure 8: Comparing Damage Factor, Clusters, and CI Values for Tests L3030 and L3535.



Damage state	Pinion SI1 mean average	Pinion SI1 mean range	Pinion SI3 mean average	Pinion SI3 mean range
No damage (clusters 1, 4)	0.20	0.20 to 0.21	0.10	0.10 to 0.10
Scuffing/pitting (clusters 2, 3, 5)	0.55	0.49 to 0.58	0.28	0.25 to 0.30



Damage state	Pinion SI1 mean average	Pinion SI1 mean range	Pinion SI3 mean average	Pinion SI3 mean range	LFO mean average	LFO mean range
No damage (clusters 1, 5)	0.15	0.11 to 0.18	0.10	0.09 to 0.11	195.16	184.71 to 205.61
Scuffing (clusters 2, 3, 4)	0.27	0.14 to 0.34	0.15	0.11 to 0.18	217.27	198.04 to 220.78

Figure 9: Comparing Damage Factor, Clusters, and CI Values for Tests L2020 and L4040.

Decision Tree Model: A decision tree was selected to demonstrate the use of clustering analysis to develop a predictive model. A decision tree uses rules to separate the data into branches and nodes based on the decisions from a set of rules. The root node is the start of the tree before any decisions have been made and contains all possible damage states.

The next node defines a node based on an input and decision rule. The benefit of decision trees is that it enables multivariable analysis, instead of single cause-and-effect relationships, of quantitative and qualitative data [16]. The optimal decision tree would achieve perfect classification with the smallest number of decisions, but this is not always possible because of inconsistencies in the data. In order to cut down on the number of decisions the model would need to make, observations made in the clustering analysis were used to decrease the number of inputs. The following inputs were chosen because they were identified as being good indicators of damage states: debris, pinion FM4, gear FM4, gear SI1, gear SI3, pinion SI1, pinion SI3, and LFO.

Four damage states were classified based on the damage scale shown in Table 1. These states were no damage, PL or GL 1 tooth, PL or GL ≥2 teeth, and scuffing. All six gear sets are partitioned separately into a dataset for training, used to fit the model, and another for testing, to assess how the model will perform on a new dataset. The training dataset was 2/3 of the data set, while test data composed 1/3 of the data set. Because of the time dependency of the dataset—run time affects damage progression—the data were separated into each group by stepping through the dataset sequentially in three steps.

Figure 10 shows part of a decision tree for one test. The root of the tree contains all the damage states at the start, and no decisions have been made. The nodes shaded in blue represent the terminal or leaf nodes of the decision tree, and each leaf node is assigned to one damage state. The structure of the tree is determined by the input variable chosen at each node and the value of the split. In this analysis, a greedy strategy was used, which results in the best partition being used first. The best partition produces the highest amount of information being gained based on that split.

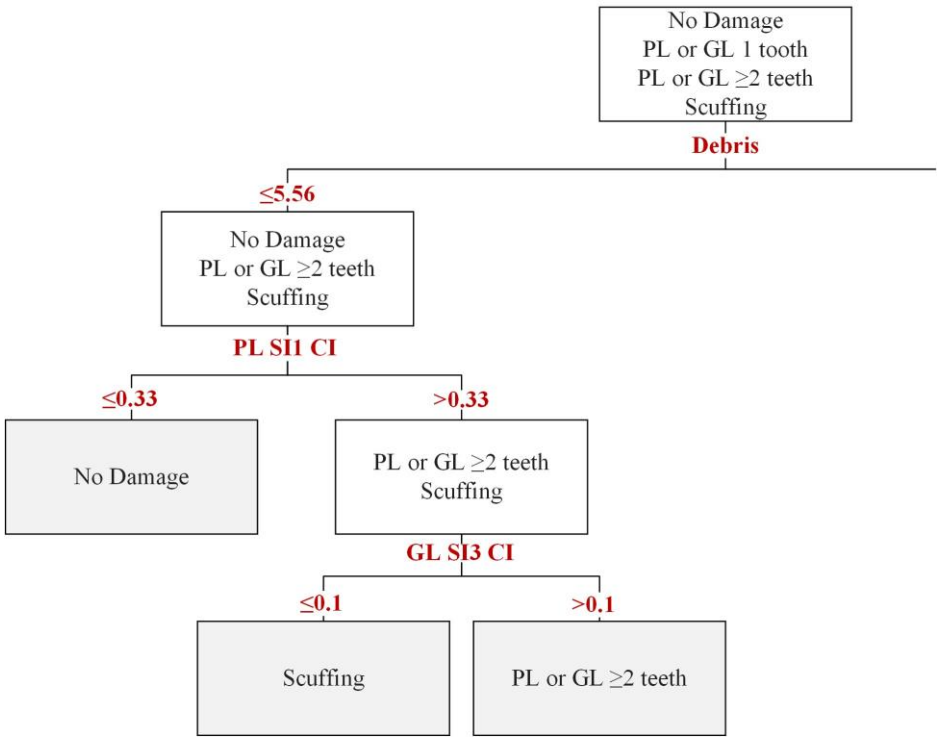


Figure 10: Decision Tree for Classifying Damage.

The first split from the root node looks at the amount of debris. If the amount is ≤5.56, then the node contains data with possible damage states of no damage, PL or GL ≥2 teeth, or scuffing. If the debris amount is >5.56, then the node contains data with possible damage states of no damage, PL or GL 1 tooth, or PL or GL ≥2 teeth. Continuing with the left node, the next split decision is based on the pinion SI1 value. If it is ≤0.33, then the node contains data with the damage state of no damage and becomes a terminal node. If the pinion SI1 value is >0.33, then the node contains data with possible damage states of PL or GL ≥2 teeth or scuffing. The split decision for this node looks at the gear SI3 value. If it is ≤0.1, then the node contains data with the damage state of scuffing. If the value is >0.1, then the node contains data with the damage state of PL or GL ≥2 teeth.

From this analysis of the left side of the tree, the following rules can be derived:

1. If Debris ≤ 5.56 and Pinion SI1 ≤ 0.33, then ***no damage***
2. If Debris ≤ 5.56, Pinion SI1 > 0.33, and Gear SI3 ≤ 0.1, then ***scuffing***
3. If Debris ≤ 5.56, Pinion SI1 > 0.33, and Gear SI3 > 0.1, then ***PL or GL ≥2 teeth***

The analysis of the right side of the tree can be done by returning to the root node and looking at the right node that was produced from the first split based on debris, but its discussion is outside the scope of this paper.

The decision tree example illustrates the benefits of combining clustering with decision trees when developing predictive models. Decision trees enable multivariable analysis on varying measurements that include qualitative and quantitative data. Future studies are planned combining data-mining analysis methods with a decision tree model.

Conclusions: Tests were performed in the NASA Glenn Spiral Bevel Gear Fatigue Rig from damage initiation to progression on the gear and/or pinion teeth. Six gear sets were tested with varying levels of tooth damage, and vibration-based gear condition indicators, amount of debris generated, and oil temperatures were measured and damage progression was documented with photographs. The state of the gear tooth was quantified with a numerical continuous damage factor. Condition indicator performance was assessed with traditional methods and data-mining methods that included clustering and a decision tree model. The condition indicators and operational parameters that were good predictors of health state found from the data mining investigation aligned with those observed during testing, with minimal knowledge of the system. Lessons learned during this investigation will be used to define additional analysis methods to be incorporated into future data-mining models.

References:

[1] R F Handschuh, Thermal Behavior of Spiral Bevel Gears, NASA TM–106518, 1995.

[2] R F Handschuh, Testing of Face-Milled Spiral Bevel Gears at High-Speed and Load, NASA/TM—2001-210743, 2001.

[3] P J Dempsey, and R F Handschuh, Detection of Spiral Bevel Gear Damage Modes Using Oil Debris Particle Distributions, Instrumentation Symposium. International. 61st 2015. (Joint Conference With MFPT 2015) Technology Evolution: Sensors to Systems for Failure Prevention, ISA Volume 507, Huntsville AL, 2015.

[4] P J Dempsey, Investigation of Spiral Bevel Gear Condition Indicator Validation via AC-29-2C Using Test Rig Damage Progression Test Data, NASA/TM—2014-218384, 2014.
[5] P Niskanen, B Hansenand, and L Winkelmann, Scuffing Resistance of Isotropic Superfinished Precision Gears, 2008, Gear Solutions, Solon OH. Retrieved from http://www.gearsolutions.com/article/detail/5810/scuffing-resistance-of-isotropic-superfinished-precision-gears.
[6] D P Townsend, Dudley's Gear Handbook, 1991, McGraw-Hill, New York, NY.
[7] AGMA Nomenclature Committee, Appearance of Gear Teeth—Terminology of Wear and Failure, ANSI/AGMA 1010-E95, 2014.
[8] H R Martin, Statistical Moment Analysis as a Means of Surface Damage Detection, 1989, Proceedings of the 7th International Modal Analysis Conference, Society for Experimental Mechanics, Schenectady NY, Pages 1016-1021.
[9] R M Stewart, Some Useful Data Analysis Techniques for Gearbox Diagnostics, Machine Health Monitoring Group, Institute of Sound and Vibration Research, 1977, University of Southampton, Report MHM/R/10/77.
[10] L Antolick, J Branning, P Dempsey, and D Wade, Evaluation of Gear Condition Indicator Performance on Rotorcraft Fleet, 2010, American Helicopter Society International 66th Annual Forum 2010, Phoenix AZ.
[11] M Triola, Elementary Statistics, 6th edition, 1995, Addison-Wesley, Boston MA.
[12] N Goodman, and A. Bayoumi, Fault Class Identification Through Applied Data Mining of AH-64 Condition Indicators, 2010, AHS 66th Annual Forum, Phoenix, AZ.
[13] C Bishop, Pattern Recognition and Machine Learning, 2009, New York NY.
[14] IBM Corp, IBM SPSS Modeler for Windows, Version 17.0. 2014, Armonk NY.
[15] Mathworks, Inc., Machine Learning with MATLAB, 2015.
[16] B DeVille, Decision Trees for Business Intelligence and Data Mining: Using SAS Enterprise Miner, 2006, SAS Institute, Cary NC.