Prognostics, The Real Issues Involved With Predicting Life Remaining[1]

Stephen J. Engel
Northrop Grumman Corp.
South Oyster Bay Rd
Bethpage, NY 11714
516-346-2893
engelst@mail.northgrum.com

Barbara J. Gilmartin
Northrop Grumman Corp.
South Oyster Bay Rd
Bethpage, NY 11714
516-575-2326
gilmaba@mail.northgrum.com

Kenneth Bongort
Northrop Grumman Corp.
South Oyster Bay Rd
Bethpage, NY 11714
516-346-4138
bongoke@mail.northgrum.com

Andrew Hess
NAWC AD
Propulsion and Power Department
Patuxent River, MD
301-757-0477
HessAJ@navair.navy.mil

Abstract—This paper reviews the fundamentals of prognostics with emphasis on the estimation of remaining life and the interrelationships between accuracy, precision and confidence. A distinction is made between the static view of failure distributions derived from historical data and the dynamic view of remaining life derived from condition. The non-stationary nature of prognoses is illustrated using data from a failing SH-60 helicopter gearbox. A method is demonstrated that measures the accuracy and uncertainty of remaining life estimates using example prognostic features. This method isolates the uncertainty attributable to features and their interpretation from the uncertainty due to the random variables that govern the physics of component failure. Results from the example features support a hypothesized trend of improved accuracy and lower uncertainty as remaining life decreases.

TABLE OF CONTENTS

1. INTRODUCTION

Most efforts in the area of Prognostics and Health Management (PHM) are focussed on the detection and diagnosis of events that mark the early stages of component failure. Early detection and correct interpretation of indications are certainly key factors in the success of PHM for aerospace applications. These aspects, however, generally overshadow and even sidestep the more difficult problem of prognostics. The essence of prognostics is the estimation of remaining life in meaningful terms that have consequence in the maintenance decision process. The extrapolation of trends based on recent observations is a common method for calculating remaining life. This calculation alone, however, does not provide sufficient information to form a decision or corrective action. Without corresponding measures of the uncertainty associated with the calculation, remaining life projections may have little practical value. Admittedly, if all the relevant factors involved in the calculation of remaining life are known for all time (past, present and future), remaining life computation would be deterministic, accurate and precise. Unfortunately, omniscience will never be affordable even if it were possible, so we are reduced to probabilistic estimates of remaining life and bound by the rules governing their interpretation and use.

2. PROGNOSTICS PRIMER

"Prognostics" in aerospace applications is not a new discipline, and yet many of the concepts surrounding it appear misunderstood and not yet fully formed. Discussions continue in several forums simply to establish a universally acceptable definition for the term "prognostics." A working definition is offered as follows: prognostics is the capability to provide early detection of the precursor and/or incipient fault condition (very "small" fault) of a component, and to have the technology and means to manage and predict the progression of this fault condition to component failure. The early-detected fault condition is monitored and safely managed from a small fault as it progresses until it warrants maintenance action and/or replacement. Impacts on other components and secondary damage are also continually monitored and considered during this fault progression process. Through this early detection and the monitoring of fault progression management; the health of the component is known at any point in time and the future failure event can be safely predicted in time to prevent it. That is, useful life remaining can be predicted with some reasonably acceptable degree of confidence.

This paper deals with the estimation of life remaining as the uncertain duration between the present and the point where a component can no longer perform its function. The aspect that prognostics takes place in the present requires that it be a dynamic process that evolves with time from the moment a component is first used until it has failed. As such, prognostics is fundamentally different from a static, a priori estimate of life expectancy (i.e., mean time to failure). In this context, prognostics is a remaining life estimation methodology that is condition-based and dynamic in both accuracy and uncertainty. The following discussion is an attempt to bring some of these issues into focus in understandable terms.

Condition-based assessments, the underpinning of the Condition-Based Maintenance philosophy, have usually emphasized the diagnosis of problems rather than the prediction of remaining life. Prognoses are considerably more difficult to formulate since their accuracy is subject to stochastic processes that have not yet happened. As a consequence of uncertainty, prognostics methods must consider the interrelationships between accuracy, precision, and confidence. For our purposes, *accuracy* is a measure of how close a point estimate of failure time is to the actual failure time. *Precision* is a measure of the narrowness of an interval in which the remaining life falls. The interval is enclosed by upper and lower bounds. *Confidence* is the probability of the actual remaining life falling between the bounds defined by the precision. This is analogous to confidence intervals commonly used in statistical analysis, except that the interval mentioned relates to an outcome of the remaining life random variable rather than a parameter of its (or some other underlying) distribution. We use this interpretation of confidence in the figures and discussions that follow to reflect the generic use of the term. With this in mind, consider the following paradox: the more precise the remaining life estimate, the less probable this estimate will be correct. In order to clarify this assertion, we begin by reviewing the distinction between the following four idealized Probability Density Functions (PDFs) for remaining life:

(A) True a priori PDF at time zero

(B) Modeled PDF estimating the true PDF at time zero

(C) True a posteriori PDF conditioned on observations during component use,

(D) Modeled PDF estimating the true a posteriori PDF during component use.

A notional illustration of all four is shown in Figure 1. The true a priori PDF for an arbitrary component type reflects the actual frequency of lifetimes for all the components ever made (past, present and future). For obvious reasons, we usually settle for a model of this function (PDF B) that can be formed either from a limited number of samples available from real life, or similar experiences derived analytically and/or empirically. The objective of the model is to approach the true distribution so that any differences are inconsequential. PDF B represents our best estimate of the a priori failure PDF at time zero for any new component of the given type (the so-called "bathtub curve"). Experience shows that components have variations in quality, and are subjected to different use and abuse. From the moment a component is first put into use, the a priori PDFs should be updated since the probability of infant mortality gradually diminishes with operation. In the absence of any other information, a reasonable PDF model can be formed at each moment in time by normalizing the original basic shape of the future portion of the distribution to maintain a cumulative probability of 1.0.

Intuition suggests that a better estimate of remaining life for a specific instance can be found during use by knowing the current condition of the component. This gives rise to the conditional remaining-life PDFs (PDF C and PDF D). At a given point in time and for a particular observation of component condition (i.e., the point labeled "present" in Figure 1), there exists an a posteriori PDF for the true remaining life (PDF C), and a prognostic model that approximates this truth (PDF D). Note that these remaining life distributions ideally are narrower (and taller to maintain a total area of 1.0) yielding more precise (less uncertain) estimates. They represent a subset of all possible instances where the distribution has been conditioned on additional information beyond the simple fact that the component is still operating. Note also that the true distributions need not be smooth or symmetrical, whereas the models are often forged into well-understood functions for computational convenience. PDFs A and B reflect the view before the component is used (at time zero), while PDFs C and D reflect an instantaneous view at some point during use after some damage indication is observed. These are drawn on the same figure for comparative purposes. Remaining life estimates can be formed from a priori models (PDF B) prior to component use, normalized a priori models from the moment use begins, or refined PDFs models (i.e., PDF D) if something is known about component condition. Each of these remaining life estimates is covered under the broadest definition of "prognostics." However, given the current state-of-the-art, the meaning of prognostics is moving toward the prediction of life remaining on the basis of the detection and tracking of faults from their inception. The advent of new technologies including sensors, advanced signal processing and early detection techniques are making it possible to perform condition-based prognostics with sufficient lead time to make PHM a reality. With this in mind, the focus of this paper is the estimation of remaining life from the inception of a fault given the capability to observe and interpret the earliest indications of an incipient failure (i.e., PDF D).

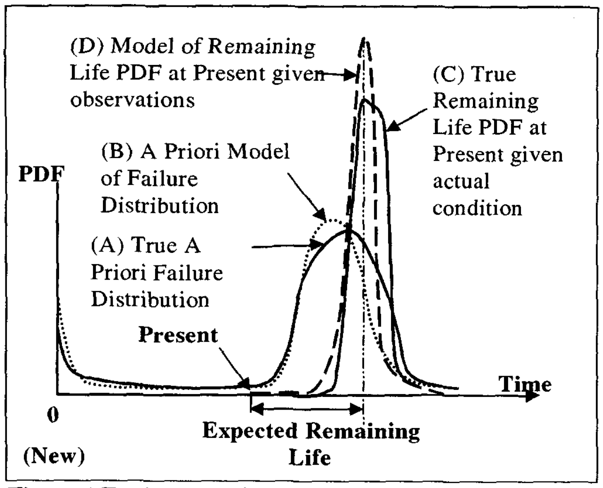


Figure 1 Truth and models of a priori and a posteriori PDFs

Regardless of which PDF is used, the probability of failure is numerically equal to the area under the PDF function between any two points in time. As long as it holds true that the component will eventually fail and has not failed yet, the total area under each of these curves from the present to eternity must equal 1.0. In other words, the probability that a new component will fail by time t is the cumulative probability from time zero to time t. Similarly, the probability that a failure will occur between 1:00 AM and 2:00 AM tomorrow is numerically equal to the area under the PDF curve between those hours. As the interval becomes narrower, the probability (area) approaches zero. For this reason, a remaining life estimate of precisely 3 hours has a probability of precisely zero of being correct (in theory). At the other precision extreme, an estimate of 3 hours plus 100 years and minus 3 hours is likely to be 100% probable while simultaneously being 100% useless. In both examples, the expected value is 3 hours. Of course there are many issues associated with predicting expected remaining life, such as how do we recognize current condition and how can we derive the PDFs without a large database of failures. Some of these will be discussed in subsequent sections.

To the extent that we cannot control the future, the determination of remaining life is a probabilistic computation. Without omniscience, one cannot know exactly when a component will fail because the factors responsible for failure generally have unknown future values. Finding where an extrapolated trend meets a condemnation threshold may provide an expectation of remaining life, but it does not provide sufficient information to make a decision. The probability of failure at this exact moment is essentially zero, and the corresponding confidence interval is unknown. The most informative solution furnishes the estimated PDF, but this is sometimes viewed as too much information. An acceptable solution might provide expected remaining life, and the lower and upper bounds that enclose an area under the PDF that equals the desired confidence. For example, suppose we are required to find the latest point in time for servicing a component that will preclude 95% of the previously experienced failures of components having a specific health condition. Figure 2 illustrates a hypothetical remaining life PDF model for components with the specified condition.

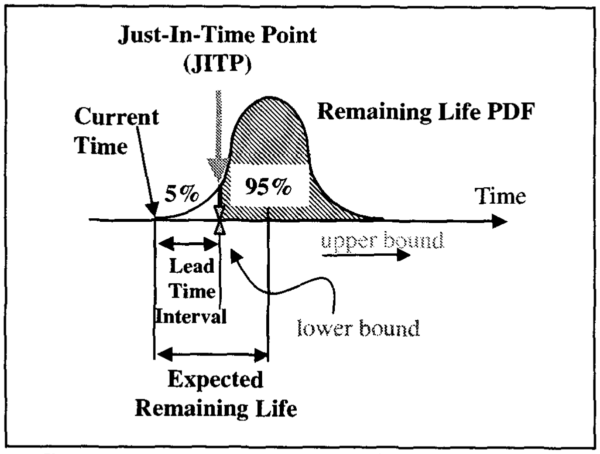


Figure 2 Max Lead-Time Interval for 95% confidence

The expected point of failure is shown as the middle of the distribution, but the decision point is actually earlier and is a function of the shape of the PDF. Note that expected value is one way of describing the 'center' of the distribution. Depending on the nature of the distribution, one might choose to use the median or the point of maximum likelihood. The place labeled Just-In-Time Point (JITP) where 5% of the PDF area has passed, is the point in time where 95% of the failures, estimated from previous experience and/or analysis, have not yet happened. The time interval from the present to the JITP we define as the Lead-Time Interval (LTI). In this example, the upper bound is (technically) plus infinity indicating an estimate with poor precision. Ideally, we would prefer the distribution to be as narrow as possible so that all failures are avoided and no unnecessary maintenance is performed. Unfortunately, the shape of the PDF is not under our control as will be explained later. Assigning values to any two of the three parameters (upper bound, lower bound, and confidence) uniquely determines the remaining parameter. As an example, Figure 3 illustrates three possible JITPs that satisfy the 95% confidence requirement, that is to say, 95% of the anticipated failures occur in the hashed in area (between the lower and upper bounds). In all three cases, the expected remaining life is the same while the LTI varies from 0 to the maximum (as shown in figure 2). The upper and lower bounds in each graph are indicated by the letters U and L respectively. Note that the lower bound corresponds to the JITP by our previous definition. The top graph suggests that servicing be done now. This most conservative philosophy has the side effect of precluding 100% of the possible failures while also having the highest unnecessary maintenance and zero component availability. The bottom graph is the longest we can wait and still ensure that 95% of the anticipated failures will be avoided (least conservative). The perceived problem with the bottom graph is the width of the confidence interval (spread between upper and lower bounds). Having a distant upper bound leads to the perception that unnecessary maintenance will be performed since the component may last for quite some time in the future. Of the three shown, this affords the least precision, and, counter-intuitively, the least unnecessary maintenance for the same 95% confidence. Knowing only that the upper bound is far in the future does not reveal the likelihood of failures far in the future. The middle graph precludes 97.5% of anticipated failures and (in this example) also has the smallest spread between upper and lower bounds and thus offers the best precision.

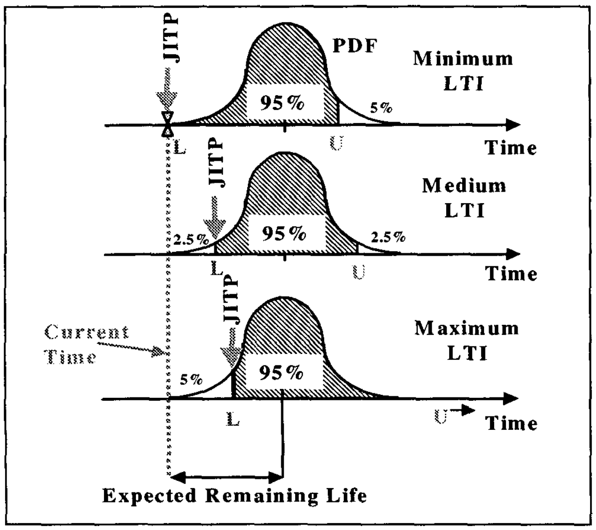


Figure 3 Prognosis With 95% Confidence and Various LTIs

The fact that the expected remaining life is midway between the upper and lower bounds in this example is simply due to the PDF symmetry and is not necessarily true in general. These examples show that having the tightest tolerance (least uncertainty) is not always desirable from a PHM point of view given criteria based on 95% confidence. In reality, as time marches along, the PDFs are updated making the LTI a moving target. There are many other criteria that are involved in the maintenance decision that are beyond the scope of this paper.

Putting the mechanics of prognostic methods aside for the moment, if 1,000 components having identical (imperfect) conditions at a given time t1 were allowed to run to failure in a realistic environment, we should not be surprised to have 1,000 different failure times. This is especially true if the damage condition at time t1 is only slight. The PDF formed by this experiment might be used to model the remaining life for the given condition. If this model were statistically accurate, it would reflect the theoretical best that any prognostics model and attendant algorithms could achieve at time t1. It is essential to realize that there is a theoretical limit to the accuracy and precision of any condition-based prognostic method regardless of our ability to know component condition, sensor data quality, preprocessing, feature extraction methods, remaining life projection and associated algorithms.

3. INDICATIONS AND INTERPRETATIONS

In keeping with the intuition that condition-based remaining life estimates are better than a priori estimates, we turn to the assessment of condition. The true damage condition of a component typically exacerbates nonlinearly with time. Consider the hypothetical example in Figure 4. The lowest curve in the figure represents the true condition of a component (not to be confused with the indications thereof). For simplicity, the condition depicted is monotonic. In reality, this may not always be true (counter examples include initial break-in, lubrication changes…). Similar graphs depicting the indications of impending failure are typically noisy and may be non-monotonic, making simple threshold comparisons unreliable for remaining life estimations. The top graph in the figure depicts the PDF from the a priori failure model (B). This provides the traditional mean time to failure estimate of remaining life. From the moment the first indication is observed until ultimate failure, a conditional PDF (D) should be used to better estimate remaining life (middle graph in figure). The former is an initial view of remaining life that is only valid at time zero. As previously discussed, this distribution can be reinterpreted as time progresses by truncating the part of the PDF to the left of the present time and re-normalizing the right side. Generally speaking, the shape to the right remains approximately the same (with an area always equal to one). The middle graph is a snapshot of the remaining life PDF, frozen at the time labeled 'current time' in the figure. Unlike a normalized version of the top graph, the middle graph is based upon an indication(s) of damage.

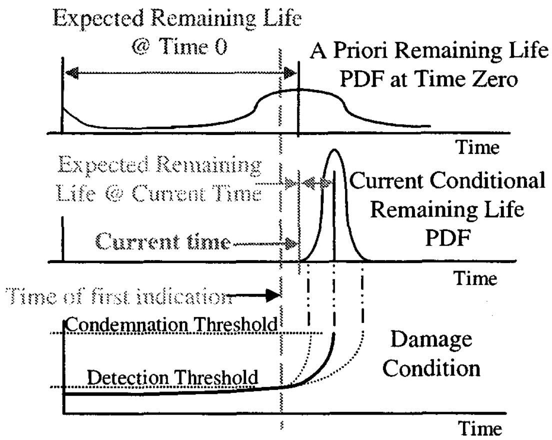


Figure 4 Notional View of Actual Damage Condition, Detection Threshold and Remaining Life Estimates Before and After First Detection

In practice, failure indications typically become more pronounced and easier to interpret as remaining life decreases. In general, the true remaining life PDFs should become narrower (less uncertain) and more stable (localization of expected failure time) as a damage condition progresses toward failure. This notion is illustrated in Figure 5.

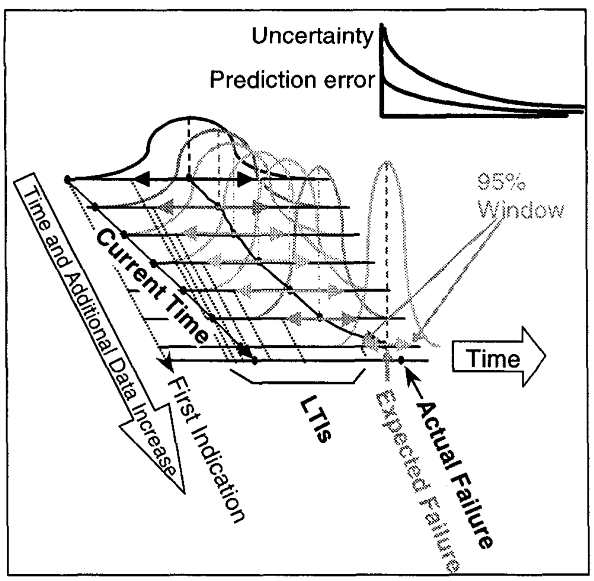


Figure 5 Idealized Condition-Based Remaining Life Estimates Typically Become More Accurate and Precise as the Time to Failure Decreases

This figure shows the broadest PDF distribution and largest prediction error (prediction error = actual – expected failure times) in the beginning when the first indication is noticed (graph farthest back in the figure). As the current time moves closer to the actual failure, the PDF distributions become narrower (less uncertain) and the prediction error becomes progressively smaller. The JITPs also tend to move in time at a decreasing rate until they converge just prior to failure. Of course this is an idealized portrait of what is expected to happen. In reality, the shape of the PDFs and their evolution in time is affected by several random factors that can make this progression noticeably irregular.

4. DERIVATION OF PDFs

The previous discussion might imply that it is necessary to have available enough failure data to estimate the distribution parameters with the desired precision and accuracy in order to predict remaining life. Clearly, most situations will not afford the luxury of an available database of examples of components with various damage conditions run to failure. Even if such a database existed, it would not provide statistics about problems not yet seen. How then is it possible to select an appropriate family of PDFs as described above? If it were possible to know all the parameters involved in the wear of a component (i.e., condition of the lubricant, smoothness of the contacting elements, hardness of materials used, composition and purity of the materials, static and dynamic loads in all axes for the past, present, and future, etc.), and we could monitor these parameters during use, we could produce a more deterministic answer with reasonable confidence even without prior failure examples. Unfortunately, this is not likely to be practical either. There may not be a perfect answer to this question; however, there are methods that can contribute towards a workable solution. Keys to the success of these methods are an understanding of the mechanics involved in the failure progression, the indications thereof, the recognition of current conditions and the exploitation of available data. Rather than discussing the development of these specific methods, we turn now to a discussion of how to evaluate prognostic methods in general.

In reality, the prognostic model reflects the uncertainty from the random mechanisms affecting the failure, compounded by the error in the prognostics process. The former is governed by the physics of the failure, while the latter involves our ability to predict the future from the information at hand. This includes the sensor data quality, signal processing, features and their interpretation, and the probabilistic treatment of remaining life. The discussion below is focussed on the prediction process and a methodology to evaluate the uncertainty introduced therein. For illustration purposes, we use an example from the intermediate gearbox of an SH-60 helicopter. By applying various predictors to this one failure, we have isolated the statistics of the prediction process from the randomness of multiple failure instances.

SH-60 Helicopter Example

The evolution of automated diagnostics for helicopter mechanical systems has been aided by a Navy program [1] of systematic testing of drive train components having known anomalies (seeded faults) while simultaneously executing a suite of diagnostic techniques to identify and classify the mechanical anomalies. This program, called the Helicopter Integrated Diagnostic System (HIDS), has been carried out using an iron bird test stand (SH-60) at NAWC - Trenton, and SH-60B/F flight vehicles at NAWC - Patuxent River. The SH-60 HIDS program has been the Navy's cornerstone effort to develop, demonstrate, and justify integrated mechanical diagnostic system capabilities for its helicopter fleets. A critical part of the HIDS program is to demonstrate the detection of catastrophic gear faults. The most serious of which are root bending fatigue failures. Depending upon gear design, this type of crack can either propagate through the gear tooth causing tooth loss, or through the web causing catastrophic gear failure and possible loss of aircraft.

A means used in the helicopter community to promulgate this type of investigation is to weaken the tooth by implanting an Electronic Discharge Machine (EDM) notch in the gear tooth root. This action creates a localized stress concentration at the tooth root that eventually leads to the formation and growth of a crack. To ensure a failure, two EDM notches (.25" Length x 006" Width x .040" Depth) were implanted along the length of the intermediate gearbox (IGB) gear tooth root by PH Tool of New Britain, PA. The location of the notches is critical as they were implanted in the input pinion where the gear tooth root bending stress is greatest. A test was then run at 100% tail power for a total of 2 million cycles. Testing was terminated prior to gearbox failure when a gross change in the raw FFT spectra was observed on an HP36650 Spectrum Analyzer. During this test, data were recorded at 100 kHz (using 16-bit integer precision) from two nearby accelerometers. A total of 36 recordings were made (most acquisitions were spaced about 15 minutes apart), covering a run-time duration of 548 minutes. Each recording contained 30 seconds of data, of which 5 seconds were used for our analysis. This data set provided a rare and nearly continuous view of gear damage progression from healthy to condemnable conditions.

During postmortem analysis, a crack was found that extended from the tooth root, through the gear web, and stopped at the bearing support diameter. There also was a void at the toe end of the notched tooth where a large section of the tooth broke off. No indication from the gearbox chip detector was observed during the test.

5. ASSESSMENT OF SELECTED PROGNOSTIC FEATURES

Diagnostics and prognostics for mechanical systems typically begin with the processing of sensor data to draw out aspects that are relevant to the purpose at hand. The enhanced representations from raw data, broadly referred to as features, may be elemental or higher-order (features of features). After feature extraction, there are generally two approaches that may be used separately or in combination. The first involves an estimation of current condition based upon the recognition of indications (symptoms). Keep in mind that there need not be a failure in the making in order to estimate remaining life. The absence of failure indications also contains useful information. The second approach involves characterizing baseline features from a healthy machine and assessing the trend away from this healthy baseline. This approach may also include some measure of distance and the notion of a state-space trajectory toward a failure condition. The first approach requires a priori knowledge of the expected failures and their indications, while the second requires an ability to discriminate between detrimental and benign conditions as evidenced by non-nominal indications. The presence of non-nominal feature values does not necessarily indicate a problem. A normal change in load, speed, temperature, operating modes, etc. can produce anomalous values in features that should not be misinterpreted as problems. Predicting remaining life is further confounded when multiple failures are present. Without features based upon the mechanics of failures, or a method of discriminating benign from detrimental novelty, the simplistic detection of deviations from some baseline is likely to produce many false alarms and inaccurate remaining life projections. This is one area where the integration of prognostic analyses with advanced diagnostic methods, such as model-based reasoning (MBR), is most advantageous. MBR algorithms can differentiate normal operational changes from detrimental novelty. This assessment, as well as its capability to analyze multiple failure conditions, will make prognostic estimations more robust.

Regardless of the approach, the repeatability and temporal behavior of the features used for prognostics must be analyzed to determine their contribution to the error and uncertainty in the remaining life estimate. For illustrative purposes, we have chosen a set of four features that have been traditionally used for gear fault detection. We examine these as prognostic features by characterizing their evolution as damage progresses. These were intentionally chosen, not necessarily as the best features known for gear prognostics, but rather to best serve the illustration. As an initial preprocessing step, accelerometer data were time-synchronously averaged to reduce noise and incoherent vibration from unrelated sources. Sixteen complete gear revolutions were used to form the average in the following analysis. Normally occurring frequencies (i.e., primary gear mesh frequencies and their harmonics) were then removed to create a residue. Some methods also remove first order sidebands but this was not done in our experiments. The signal resulting from this initial step, referred to as the Time Synchronous Average Residue (TSAR), should be low in amplitude and primarily composed of Gaussian noise if the gear is operating properly. The four chosen features are simply different ways of characterizing the residue signal. The first feature is the normalized kurtosis (fourth statistical moment), which provides a measure of how outlier-prone (spikes) the distribution is. This feature has been referred to as FM4 as credited to Stewart [2]. The second feature is the RMS value of the TSAR signal, which intuitively provides a measure of the magnitude of the defect indication. The third feature is the variance, a fundamental parameter for characterizing distributions. The fourth feature is proportional energy, which is similar to the metric used by Swansson [3], but redesigned for localized problems as opposed to heavy uniform wear. The reader is reminded to consider the following analysis as a methodology for evaluating prognostic features rather than as a general solution for predicting gear failures.

The features above were analyzed from both repeatability and temporal viewpoints. For repeatability analysis, a five-second sample from each of the thirty six recordings was used to generate twenty one values (using non-overlapping windows) for each feature. Assuming that the condition of the IGB is essentially constant for the 5-second duration, all twenty one values should be the same, within some acceptable spread. If the values are not close to each other, the feature may be too brittle to extrapolate without further processing. The degree of spread combined with the uniqueness of the range of values provides a measure of the robustness of each feature for this data set. The characteristics of these values are shown using a box plot diagram. For temporal analysis, the sequence of values from healthy to failure for each feature was modeled as a polynomial. A sub-sequence of four consecutive values were then fit to the polynomial model (in a least squares sense) to derive an expected remaining life. Using the bootstrap statistical method, a series of remaining life estimates at each recording time were generated, from which errors and error distributions were calculated. As such, the temporal analysis reflects the consequences of static repeatability and the uniqueness of a sequence of values on the ability to use the given feature to estimate remaining life.

Figure 6 is a legend explaining the box plot symbology (from the MATLAB statistics toolbox) for the figures that follow. The top and bottom of the boxes (polygons) mark the interquartile range. The black line near the middle indicates the mean value. The asymmetry shows the degree of skewness. The whiskers (the horizontal lines above and below the boxes) show the maximum and minimum samples excluding outliers. An outlier was defined as a value that is more than 1.5 times the interquartile range away from the top or bottom of the box. Box plots for each of the four features are shown below in Figures 7 to 10.

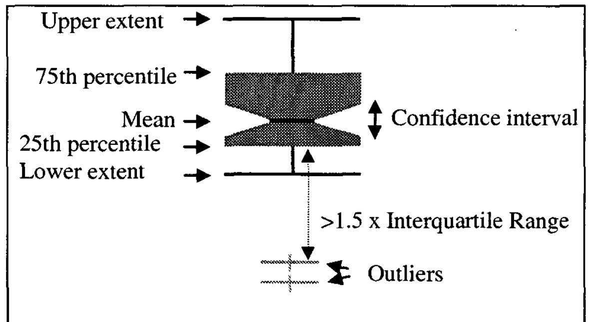


Figure 6 Box Plot Legend (MATLAB[2] conventions)

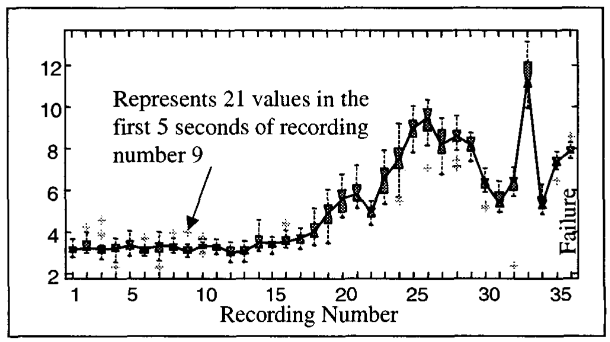


Figure 7 Kurtosis of TSAR (Healthy to Failure Conditions)

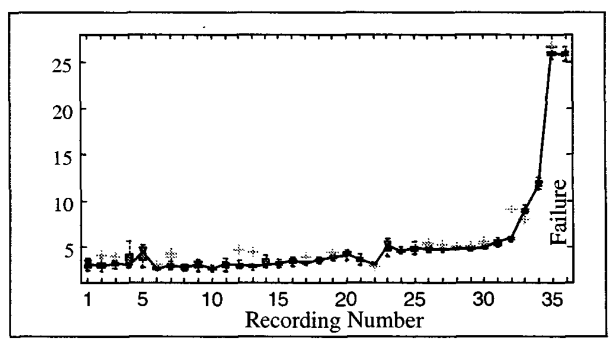


Figure 8 RMS of TSAR (Healthy to Failure Conditions)

[2] Box plots were produced using the MATLAB. MATLAB is a registered trademark of The MathWorks, Inc.

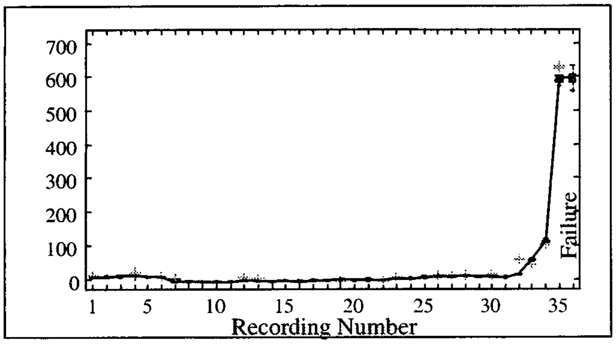


Figure 9 Variance of TSAR (Healthy to Failure Condition)

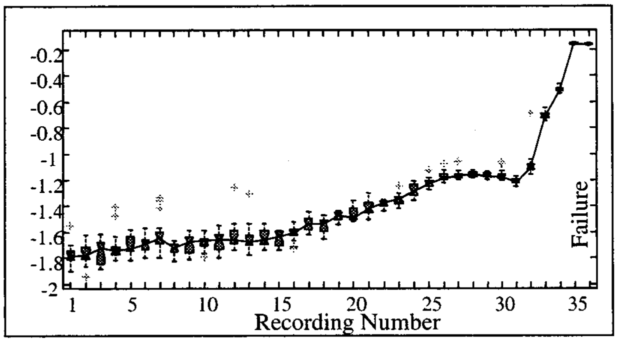


Figure 10 Proportional Energy (Healthy to Failure Conditions)

There are several things to consider when interpreting Figures 7 through 10. Features that are non-monotonic and/or flat (zero average slopes) are difficult to interpret for the purpose of estimating remaining life. If the variance in these features is also large, it may be meaningless to extrapolate. TSAR Kurtosis (see Figure 7) for example, is non-monotonic after recording 26, so it may not be sufficiently well behaved to be used as a predictor near the end of life. Although it provides a strong indication of a developing problem in the intermediate stages (recordings in the early twenties), the variance is probably too large to be used for prediction through simple extrapolation. TSAR Kurtosis is probably best used as an early indicator of a problem. Early detection could then invoke a predictive model using a better-behaved feature near the end of life.
In contrast, many of the features that are well behaved during the rapid decline stage are largely flat in the earlier stages. For example, values of TSAR Variance are virtually the same for the first 30 recordings (see Figure 9). These values do not even reveal a problem until the last four recordings. However, if the values for the last six recordings of TSAR Variance are modeled (as an exponential or polynomial), an extrapolation might prove useful. A good compromise would be a combination of the two features where TSAR Kurtosis is used for initial detection, and TSAR Variance to estimate the rapid decline near the end of life. Figure 8 shows the quality of TSAR RMS, a feature closely related to variance but less well behaved. Figure 10 shows that the Proportional Energy feature is reasonably well behaved, largely monotonic, and has a favorable first derivative throughout the progression of damage. We believe that it is relatively immune to non-stationary noise (since it is based on a time-synchronous average), and invariant to changes in operating environment, and multiple failures. Due to the original experiment design, it is not possible to test these assertions with the data available.

The results of the temporal analysis are shown in Figures 11 to 14. These figures demonstrate a progression toward lower prediction error and uncertainty similar to the idealized illustration in Figure 5. For the sake of an uncluttered representation, Figures 11 to 14 are 90 degree rotated versions of Figure 5 showing only the expected failure point and the 95% confidence interval. In these figures, the Just-In-Time Point (using the maximum LTI) for each feature is shown as a function of time (labeled Just-In-Time Line). Also shown (as a diagonal line from top left to bottom right) is the actual remaining life. Expected remaining life and 95% confidence intervals were computed as a function of time by considering the PDF distributions attributable to predictions generated by each of the given features. Each figure shows that as data accumulates and indications become easier to interpret (over time), the expected remaining life and lead-time estimates approach the actual remaining life.

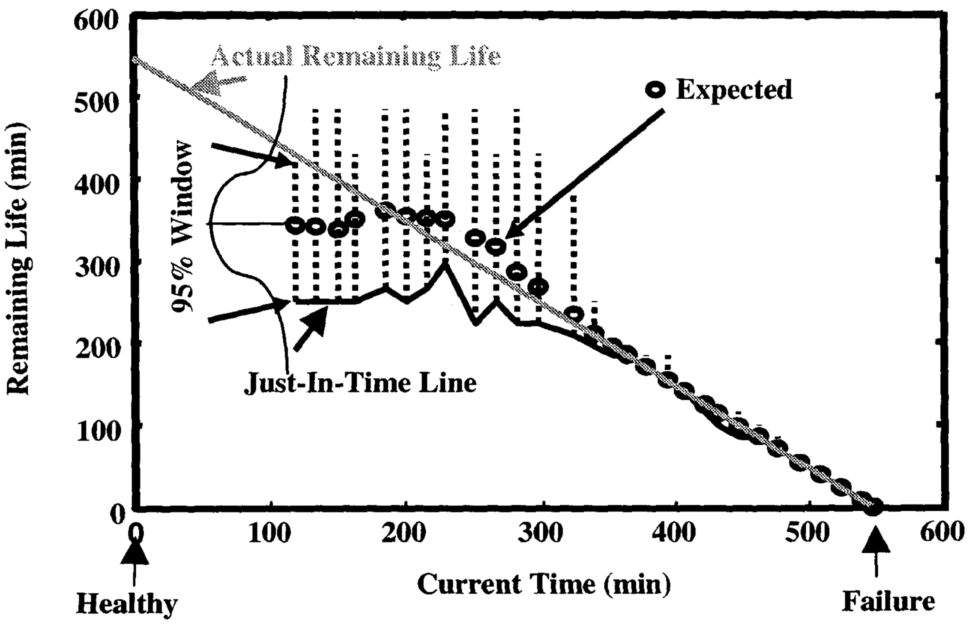


Figure 11 Just-In-Time Line for TSAR Kurtosis

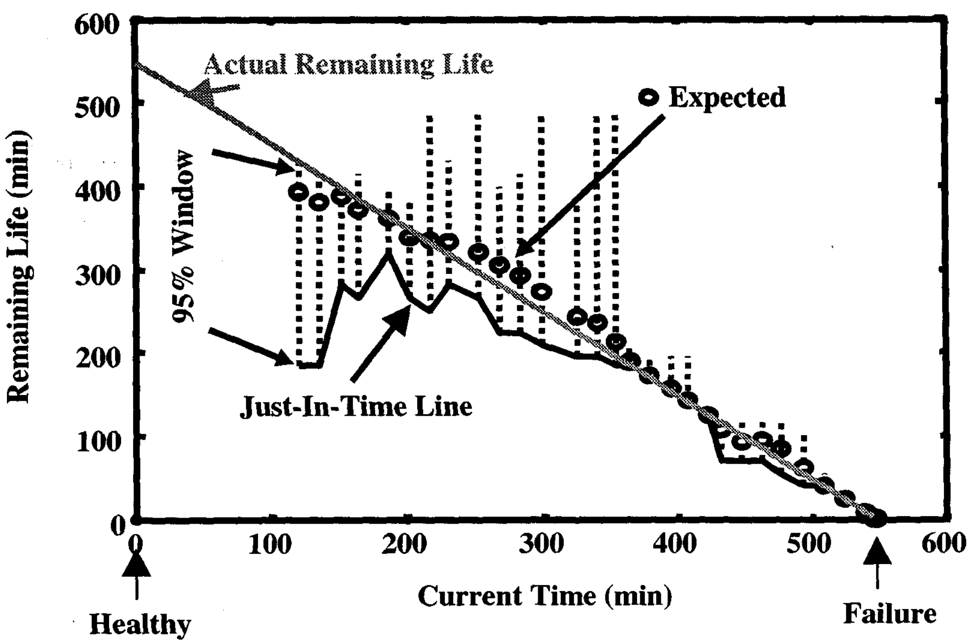


Figure 12 Just-In-Time Line for TSAR RMS

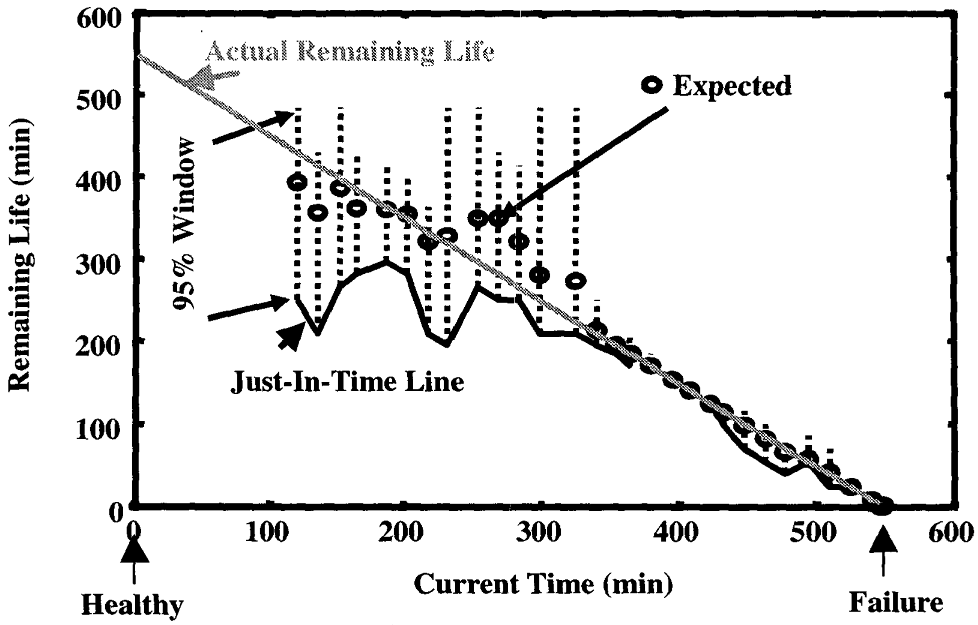


Figure 13 Just-In-Time Line for TSAR Variance

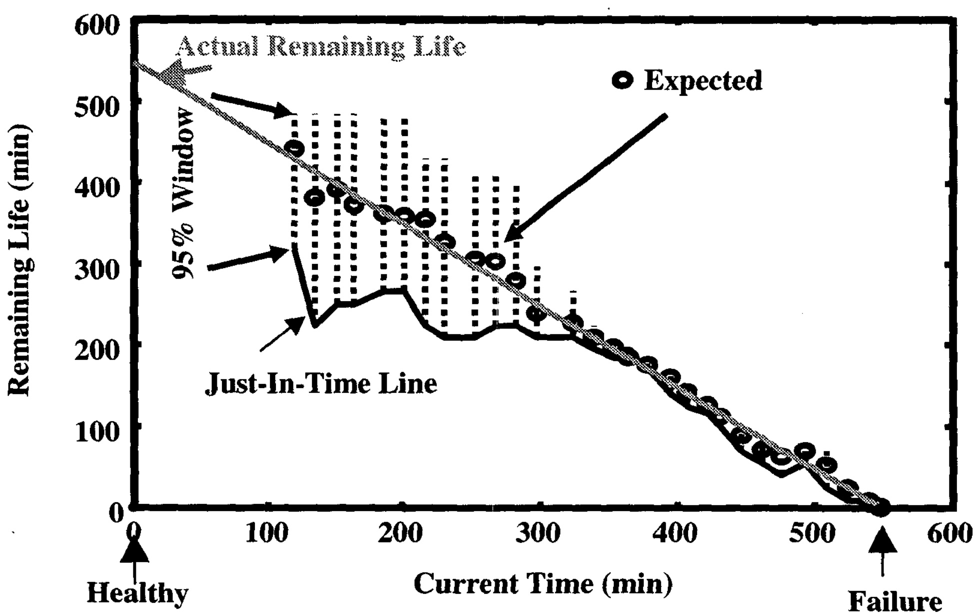


Figure 14 Just-In-Time Line for TSAR Proportional Energy

Predictions derived from each of the chosen features show the anticipated tendency to become accurate and less uncertain toward the end of component life. The Just-In-Time line derived from the kurtosis feature (Figure 11) exhibits the most uniform initial approach toward the truth while also having the most conservative initial estimates of life remaining. Also of interest is the observation that kurtosis remains certain in the region around 100 minutes prior to failure while the other features afford less precision. Recall that kurtosis became non-monotonic near the end of life (Figure 7) making its absolute value an unreliable predictor. Considering the temporal sequence of its values, however, suggests that the erratic behavior of this feature is a sure sign that the end is near. The RMS feature (Figure 12) has good overall expected remaining life estimates. Unfortunately, the uncertainty is large and the skew in the distribution goes from one extreme to the other. The variance feature (Figure 13) in this example is similar in character to the RMS feature. One difference is the expected values in the time period between 250 and 300 minutes where the variance prediction has the largest remaining life error (on the dangerous side) of all the features shown. The proportional energy feature (Figure 14) is the best general performer in this example. It has the smallest average error from the beginning, most uniformity in convergence of uncertainty, and the best average PDF symmetry.

The results above are intended to illustrate the analysis method rather than to serve as a general critique of the four features shown. This method may be repeated while varying algorithmic parameters such as the number of gear revolutions used in the average, the number of points used in the curve fitting sequence, the sampling frequency, and so on. Prognostic sensitivity to each variable can thus be quantified. Although a wide assortment of features have been proposed in open literature for detection and diagnosis of impending mechanical failures, very few have been analyzed for the purpose of estimating remaining life. Where one may be beneficial for early detection, another may better serve near the end of component life. It is not necessarily true that the best feature for early detection of a particular failure is also the best feature for estimating the remaining life as that failure progresses.

Perhaps the best way to develop an overall solution is to fuse the information from a variety of methods together into a composite solution. Given that remaining life estimates are probabilistic, the fusion method can become rather complex. Bayesian methods (i.e., Bayes Nets) have become increasingly popular for similar fusion applications, but simpler methods may also be effective in many cases. One of the simplest methods is "best of breed." For our purposes, the best of breed solution simply selects the prediction method that is known to have the best solution quality under the circumstances. The measure of quality can be prediction error, uncertainty, or some combination of both. An example of the fused results using least prediction error is shown in Figure 15.

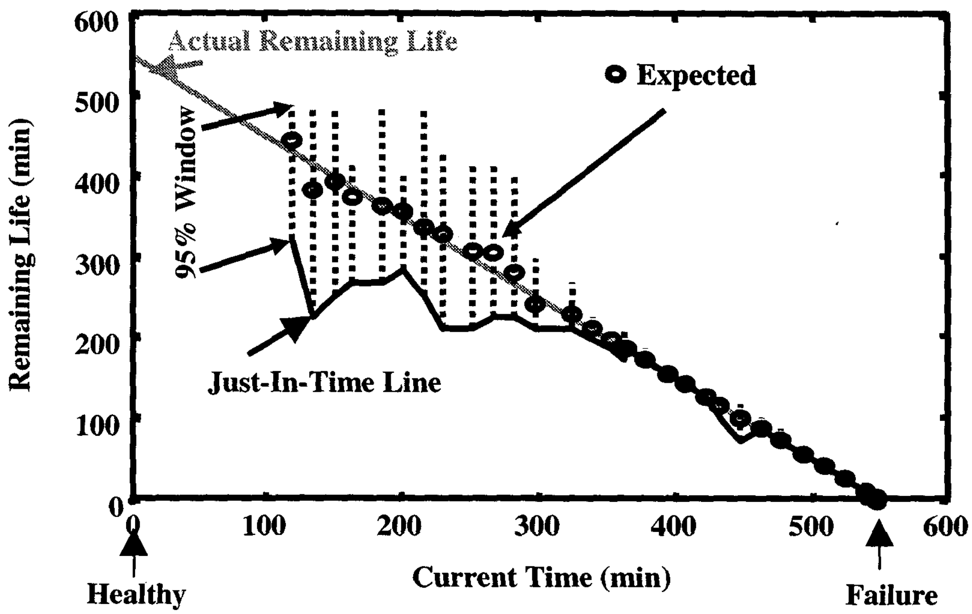


Figure 15 Just-In-Time Line for Fused Predictions

Note the similarity between the fused result (Figure 15) and the result of the proportional energy feature alone (Figure 14). This suggests that the proportional energy feature is probably sufficient if best of breed and minimum error are the criteria for fusion. If a Bayesian methodology were used to fuse predictors, the width of the fused confidence interval would have also been shortened.

6. SUMMARY AND CONCLUSIONS

The management of uncertainty is an important and often overlooked aspect in the estimation of remaining component life. We have shown that precise estimates of remaining life have a very low probability of being correct. We have also explained why a prediction with low uncertainty may not be desirable from a maintenance point of view. In fact, the remaining life estimate with the widest confidence interval (lowest precision) may offer the least unnecessary maintenance.

Uncertainty comes about in two fundamental ways: first, by random variables that govern the condition, deterioration and use/abuse of components, and second, by the prognostic process itself that is used to estimate the former on the basis of observations. By investigating a singular failure with knowledge of true remaining life, we have isolated these two sources so that the latter can be readily analyzed.

We have shown evidence that supports the conjecture that the prognostic process generally becomes more accurate and confident as remaining life decreases. We believe that this is due in part to the tendency of indications to become more pronounced and easily interpreted as conditions worsen. We also recognize that the random elements that govern the failure mechanism cannot be modeled on the basis of one example. Consequently, the success of a prognostic method largely depends on the judicious choice of features and interpretations thereof that are based on reasonable and expected phenomena.

We further postulate that the convergence of true life remaining is similarly a function of condition, that is to say, the variance of actual remaining life decreases as damage increases. If the rate of this convergence is sufficiently rapid, then at some point near the end of life, the error due to the uncertainty in the conditional distributions of remaining life may be dominated by the quality of the prognostic process. The tendency of the damage progression to accelerate exponentially leads to the conclusion that accuracy and uncertainty fundamentally improve near the end of life. In other words, as the slope of the true condition curve (see the bottom curve in Figure 4) increases, its projection on the time axis becomes increasingly narrow. As a consequence, the prognostic process must become increasingly attentive.

7. AREAS FOR CONTINUED RESEARCH

We have demonstrated a methodology that measures the quality of prognostic features as a function of actual remaining life. For practical reasons, we may never have sufficient data to completely characterize the true remaining life PDF of a given component empirically in all possible damage conditions (PDF C explained in the primer). Consequently, prognostic PDF models (PDF D) may be flawed without an opportunity for validation. We also recognize that some level of PDF inaccuracy can be tolerated and, that maintenance decisions can still be effective if the uncertainty can at least be bounded. With this in mind, research in the following areas should provide more effective prognostics for future applications.

- Creation of features specifically designed for remaining life estimation. Efforts to date have focussed on the search for features suitable for early detection and diagnosis. These may or may not be best suited for predicting the temporal signature of damage progression.

- Creation of methods that can bound the uncertainty in remaining life predictions in lieu of well-characterized PDFs. Decision points are generally based on the bounds of the confidence interval rather than on the expected remaining life itself.

- Studies involving the convergence rates of accuracy and uncertainty as a function of life remaining. One of the best measures of remaining life may be the confluence of uncertainty in predictions formed using various features rather than the estimates produced by features themselves.

- Development of a fusion testbed for prognostic methods. There are many methods described in open literature for the detection, diagnosis and prognosis of component failure. It is doubtful that any one method will be the best in all situations. The objective of this effort is to exploit the strengths of various methods so that they may be applied to the appropriate situation and that their result can be fused together to form a synergistic solution.

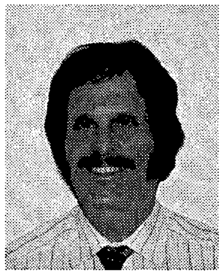
8. REFERENCES

[1] Hardman, W., Hess, A., and Sheaffer, J.: "SH-60 Helicopter Integrated Diagnostic System (HIDS) Program – Diagnostic and Prognostic Development Experience." 1999 IEEE Aerospace Conference ....

[2] Stewart, R. M.: "Some Useful Data Analysis Techniques for Gearbox Diagnostics." Report MHM/R/10/77, Machine Health Monitoring Group, Institute of Sound and Vibration Research, University of Southhampton, July 1977

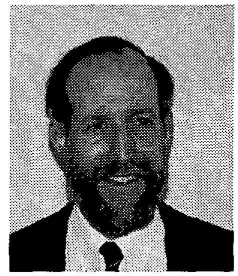
[3] Swansson, N. S.: "Application of Vibration Signal Analysis Techniques to Signal Monitoring." Conference on Friction and Wear in Engineering 1980, Institute of Engineers, Australia. Barton, Australia 1980, pp. 262-267.

9. BIOGRAPHIES



Stephen J. Engel *is a senior technical specialist in Soft-Computing Technologies for the Integrated Systems and Aerostructures (ISA) Sector of the Northrop Grumman Corporation. His primary focus is the design and development of neural networks, fuzzy systems, and probabilistic reasoning methodologies for a variety of military and commercial applications. He was the engineering manager for "Just-In-Time Maintenance" (DARPA) and "Network Early Warning System" (US Air Force) R&D programs, both involving soft-computing technologies for prognostics. He is currently supporting PHM efforts for the JSF Lockheed/Northrop Grumman/BAe team. Mr. Engel received his BSEE from Worcester Polytechnic Institute and MSCS from Polytechnic Institute of New York, and holds several patents in related technologies.*

Dr. Barbara J. Gilmartin *is a principal engineer in Intelligent Systems for Northrop Grumman Corporation's ISA Sector. Currently, she is technical lead on several PHM programs as part of the JSF Lockheed / Northrop Grumman / BAe team. She has also served in this capacity on PHM programs for other manned aircraft, UAVs, ships, and submarines. Her research interests include model-based diagnostics and their integration with intelligent system management / resource allocation methods. Before joining Northrop Grumman, she was director of R&D at a university-based research institute specializing in alternate fuel development. Dr. Gilmartin has her M.S. and Ph.D. degrees in Physics from Adelphi University, Garden City, NY.*



Kenneth Bongort *is an information systems technologist with Northrop Grumman's Logicon Inc. subsidiary. He is currently assigned to Intelligent Systems in Northrop Grumman's ISA Sector. His main interests have been in developing algorithms and software systems that perform applied engineering analyses. These include systems that are used onboard in the offshore oil industry, aircraft fatigue analysis and monitoring systems, and, most recently, the user interface and analysis software for DARPA's "Just-IN-Time Maintenance" project. Mr. Bongort has a BA in Psychology/Mathematics, an MA in Statistics and a BSE in Naval Architecture and Marine Engineering - all from the University of Michigan.*

Andrew Hess *U. S. Navy Propulsion and Power Diagnostics Andy is world renown for his work in fixed and rotary wing health monitoring and is recognized as the father of naval aviation diagnostics. Beginning with the A-7 aircraft engine health monitoring program of the early 80's, Andy has been a leading advocate for health monitoring in the Navy. His efforts have largely led to the successful transition of a development program (HIDS) into production (IMD HUMS) for all H-60, H-53, and H-1 aircraft. Currently, Andy is hard at work integrating health and usage for the Joint Strike Fighter program and frequent participant in the international technical community.*