INTRODUCTION

This paper presents results of Monte Carlo simulation of the Retirement-for-Cause (RFC) engine maintenance system as developed by Pratt and Whitney Aircraft and the U.S. Air Force. The Retirement-for-Cause concept is addressed, conventional Monte Carlo modeling techniques are explained, and an alternative approach developed at Pratt and Whitney is presented. Next, a simplified non-ideal Non-Destructive-Evaluation (NDE) model with fixed probabilities of Type I and Type II errors is described and simulation results obtained using this model are presented and discussed. An appendix presents a survey of various methods used to model NDE.

RETIREMENT-FOR-CAUSE

The fatigue process is often considered conceptually to be comprised of initiation and propagation phases, with the initiation phase exhibiting the greater variability. During initiation, small microcracks develop, grow, and eventually link up to form a crack, which would then propagate subcritically until the combination of service load (stress) and crack size exceeds the material fracture toughness. Catastrophic failure would result had not the component been retired from service. To preclude such failures, rotating components are typically retired at a time when 1 in 1000 (for example) could be expected to have developed a short fatigue crack. By definition then, 99.9 percent of the retired parts still have useful life remaining at the time they are removed from service. Under the Retirement-for-Cause philosophy, each of these components could be inspected and returned to service if no crack was found to
have initiated. The Return-to-Service (RTS) interval is determined from a fracture mechanics calculation of the remaining propagation life from a crack just small enough to have been missed during inspection. This procedure could be repeated until some measurable damage had accrued to the part, at which time it would be retired for that reason (cause). Parts would be retired for having incurred quantifiable damage, rather than because an analytically predetermined minimum design life had been exceeded. The goal here is safe utilization of inherent component life capacity, not life extension.

NDE must provide the means of screening disks with flaws that could cause component failure within an economically feasible RTS interval. NDE capability with acceptable flaw detection resolution has been available for some time, but adequate reliability of flaw detection has been lacking. Complementary inspections and recent improvement in NDE single inspection reliability (by automation), can now provide the required reliability for many gas turbine engine components to utilize economically the RFC maintenance concept.

Since any real inspection will of neccessity result in a certain number of Type I (misses) and Type II (false calls) errors, the effect of these errors must be determined. A Monte Carlo simulator modeling the inspection system and life cycle behavior for engine disks has been developed which allows for statistical variations in component behavior and simulates the effect of various parameters on Life Cycle Cost (LCC). The effects of non-ideal NDE on life cycle cost using a simple delta function model were examined using this simulator.

A conventional Monte Carlo simulator follows each engine disk as it is inspected, returned to service, reinspected, and so on. The Probabalistic Life Analysis Technique (PLAT) simulator developed at Pratt and Whitney Aircraft is of this type. The collective behavior of new disks however, obscured investigation into the behavior of older disks since there was no simple way of determining the relative proportions of disk ages in any given RTS interval. Furthermore, it can be seen that as the entire population progresses through time, an ever increasing proportion of computer time is spent simulating the behavior of new disks, essentially a redundant exercise.

SIMULATION TECHNIQUE

A new technique which models the life cycle without disk replacement has been developed, the PLAT-2 simulator. At first glance, this would appear to defeat the purpose of Retirement-for-Cause entirely, since the concept is based on disk replacement when a specified damage criterion is met. This apparent problem, however, is easily disposed of; the simulator is broken into two func-
tional parts. In Part 1, the specified number of disks is tracked through the several inspection intervals. Since no disks are replaced when they are retired (or fail) the in-service population becomes smaller with each successive RTS interval. This provides the relative frequency or probability of disk rejection, the probability of failure, and the probability of error for disks after 1, 2, ... n service intervals. An additional computational benefit also accrues: as more discs are removed from service, the computer time required decreases since new disks are not being re-simulated. The results from Part I and the laws of probability are then used in Part II to generate the same results obtained by the PLAT simulator which required significantly more computer time. The resolution developed in the distribution tails is greatly increased since a much larger population of older components was sampled from. In addition, the effect of non-constant fleet size or varying RTS interval can be simulated easily.

MODELING NDE WITHIN CONTEXT OF RFC

Perfect NDE consists of an ideal step function with the step at the inspection crack size, a-insp, as shown in Figure 1. All cracks smaller than a-insp will be passed by the inspection process, while the cracks larger than a-insp will be rejected. Unfortunately, this is never the case in reality.

There are various approaches to modeling non-ideal inspection. Three major types of models have been used, viz: Probability-of-detection (POD) vs. crack size (a); indicated crack size (a-hat) vs. actual crack size; and a probability distribution of normalized crack size (a-hat/a). These models are treated more fully in the Appendix.

The approach taken in the investigations described here utilizes a step function, but no longer assumes either zero or 100% POD. By making approximation that a constant percentage of all cracks smaller than a-insp will be detected (and the part rejected, resulting in a Type II error) and a constant percentage of all cracks larger than a-insp will be missed (and the part passed,

---

Besuner and Rau of Failure Analysis Associates, Palo Alto, CA, correctly point out that Probability of Detection is a misnomer. A more correct term would be Probability of Rejection, since many surface anomalies (eg: scratches, wear, machining marks) can be "detected" without their leading to a component's rejection. POD, however, appears to be the more common term and will be used here with the understanding that detection subsequently requires component rejection.
Fig. 1. Perfect Nondestructive Evaluation Models

Fig. 2. Step Function Nondestructive Evaluation Models
resulting in a Type I error) provides a very simple model which describes non-ideal inspection results. An additional advantage of this description is that it eliminates the need to model the shape of the POD curve. In modeling a non-ideal inspection the shape of the middle of the curve can vary greatly; an infinite number of curves may be drawn through any three points selected, while the step function model allows only one curve to be drawn through the three points. Using this approach avoids a costly parametric evaluation of rise time, the effects of which depend greatly upon the underlying distribution of crack sizes being inspected. The curves used in this investigation are shown in Figure 2.

CONSEQUENCES OF IMPERFECT NDE

For the purposes of this paper, several ground rules were adopted. Simulation of RFC situations would result in very few interesting occurrences (disk failures, disk removals, errors) since the purpose of RFC is to prevent failures by removing those few which might be candidates. Conditions were therefore intentionally biased to increase the number of these salient occurrences. Therefore, a short propagation life (relative to the return-to-service interval), a low initiation life (relative to total disk life) and a low propagation margin were selected. (Propagation margin is the ratio of propagation life to RTS interval.)

The results presented here are NOT, therefore, representative of actual engine conditions, but were selected to demonstrate the influence of NDE fallibility on Retirement-for-Cause.

Two conclusions are drawn concerning the effect of Type I errors during the disk lifetime. First, Type I errors increase with time as shown in Figures 3 through 6. The incidence of error increases even though the probability of error, given that a crack exists, is constant. This is because the mode of the cracksize distribution increases over time as the fatigue process progresses. These figures are comparison plots over the eight inspections simulated showing percent Type I errors occurring within each interval; three levels of probability of Type I errors (0%, 10%, and 20%) at a given level of probability of Type II errors (eg: 5% in Figure 3.) are presented. It can also be seen from a comparison of the four figures that the probability of Type II error has little effect on Type I error. Only when the probability of any removal is very high does the incidence of failures due to Type I error diminish. Figures 7 through 10 show the effect of varying Type I and Type II probability on disk failure rate. As expected, increased Type I probability results in increased disk failures, or alternatively, requires shorter RTS intervals to maintain a constant failure rate.
Fig. 3  Type I errors increase with time.

Fig. 4  Type I errors increase with time (cont.).

Fig. 5  Type I errors increase with time (cont.).

Fig. 6  Type I errors increase with time (cont.).
Fig. 7 Disk failure rate increases with time and Type I errors.

Fig. 8 Disk failure rate increases with time and Type I errors (cont.)

Fig. 9 Disk failure rate increases with time and Type I errors (cont.)

Fig. 10 Disk failure rate increases with time and Type I errors (cont.)
The principal effect of Type II errors is to increase life cycle cost. Since Type II errors result in disk removals, the number of replacement disks must increase as a result of these errors. This effect can be seen in Figures 11 through 14 which show percent removals in each interval for the given NDE capabilities. Reference to Figure 7 through 10 shows that Type II errors also have a minimal effect on disk failure rate.

Notice that the figures do not include the effect of disk replacements over the life cycle. Previous efforts to show the effects of Type I and Type II errors on an entire population which included other effects created by disk replacement were largely unsuccessful; the replacement disks obscured the results. The figures showing the effect of inspection correlation referenced in the following discussion do however, include the effect of disk replacement during the life cycle since this was the area of concern.

The correlation between successive inspections at the same disk location was also considered. Successive inspections of the same disk location are not independent; there is some degree of correlation between the probability of finding a crack in succeeding inspections. A recent paper (Yang, 1982) investigating the effect of totally independent inspections concludes that there is little probability of a disk surviving an inspection due to the cumulative effect of independent Type II errors. Were this actually the case, throughput at overhaul centers would approach zero. This has not been seen. Figures 15 through 17 show the effect of varying the correlation between subsequent inspections. Assuming the inspections are totally uncorrelated greatly increases the replacement rate over cases assuming no correlation, or cases assuming a 50% correlation. (Remember, this correlation is for subsequent inspections of the same disk — with the POD vs. a relationship treated as representing an entire disk — and not single, independent inspections of multiple locations on a single disk.)

SUMMARY, CONCLUSIONS, RECOMMENDATIONS

In summary, several points can be made:

1. It is possible to tolerate (at the price of shorter Return-to-Service intervals and increased inspection cost) less than ideal NDE and still have a viable RFC procedure:

2. The effect of changing NDE reliability is more clearly seen when the results are not obscured by disk replacements.

3. Increasing probability of Type I error requires shorter RTS interval to maintain constant failure rate; incidence of Type I error increases with time for constant error probability.
Fig. 11 Removals increase with Type II errors.

Fig. 12 Removals increase with Type II errors (cont.).

Fig. 13 Removals increase with Type II errors (cont.).

Fig. 14 Removals increase with Type II errors (cont.).
Fig. 15 Decreasing correlation between inspections increases removals.

Fig. 16 Decreasing correlation between inspections increases removals (cont.).

Fig. 17 Decreasing correlation between inspections increases removals (cont.).
4. Increasing probability of Type II error diminishes the chance of a part's being returned to service, increases overall cost of inspection, but has minimal effect on failure rate.

5. RFC results are influenced by assumptions about the correlation between subsequent inspections of the same part; increasing correlation decreases the overall incidence of Type II error and removal rate.

Areas recommended for future work include:

1. Research into the exact degree of correlation between successive inspections at the same disk location;
2. Further quantification of this effect on the RFC process;
3. Increased use of binary data analysis techniques for coefficient evaluation in modeling the POD vs. a relationship, including an investigation of the influence of the functional model used.

APPENDIX

Because RFC is greatly influenced by factors other than NDE (e.g., uncertainties in mission usage, fatigue behavior, stress analysis and life prediction) and because imperfect NDE is a convenient scapegoat for possible RFC difficulties, perfect NDE has been used, until recently, in the simulation of the Retirement-for-Cause system. Cracks larger than the inspection crack size will be rejected; cracks smaller than inspection crack size will be passed. Various NDE models were considered for this investigation.

Perhaps the oldest method of describing NDE reliability is to quote a crack size for which there is a 90% probability of detection with a statistical confidence level of 95%, a-90/95. Since this approach does not describe the NDE behavior at crack sizes smaller (or larger) than a-90/95, it cannot be used for RFC decisions. Furthermore, recent scrutiny of the mathematical statistics used in determining this parameter from a finite amount of data indicates that its expected variability makes it of little practical value in providing even a qualitative assessment of NDE reliability (Berens, 1982).

Perhaps the most common approach to quantitative NDE modeling is the POD - a relationship. Conventionally, data are presented as the fraction of cracks within a specific size range which were detected, POD, plotted against crack size, a. Various algebraic models are then used to describe this relationship using regression techniques to evaluate the model coefficients. Often some linearizing transformation is made to simplify the regression analysis.
When data are available in copious quantities, this approach is acceptable, and is relatively easy to use. However, when there is a paucity of data the scheme of grouping data according to crack size range becomes untenable. For example, consider 30 individual inspections of differing crack sizes. Grouping into three size ranges would provide ten percentage points of resolution on the POD axis, but results in gross size resolution. Any trade-off to improve resolution of crack size is at the expense of resolution in POD. Worst of all, in any event there are only three POD, a woefully insufficient number for even a two-parameter (eg: linear) model.

An alternative approach to determining model coefficients is to use binary data analysis and the maximum likelihood method. POD data are treated as either 0 (component accepted) or 1 (component rejected). Any model describing the POD vs. a relationship can be postulated; usually a two parameter model is adequate. The model coefficients are then evaluated using the maximum likelihood method for determining coefficients which maximize the probability (likelihood) of obtaining the observed POD data pairs. Various functional forms of the POD model have been evaluated using this technique, which shows great promise.

Under Retirement-for-Cause, the decision to return a part to service will be based not on how large a crack actually is, but on the size it appears to be. A second method for NDE modeling directly addresses this sizing problem. By plotting the indicated crack size on the y-axis vs. actual crack size on the x-axis as shown in Figure 18, the variability in this decision parameter becomes apparent. This procedure does not avoid the necessity of modeling, however, since some relation between apparent and actual sizes must be determined. Often a linear relationship suffices. It should be remembered that other methods of modeling NDE do not avoid the sizing problem but treat it implicitly. For example, the POD vs. a relationship considers detections which require subsequent rejection.

A third method is to "normalize" the indicated crack size by dividing by actual crack size. The resulting non-dimensional representation can be illustrated by a probability plot as shown in Figure 19. Symbol sizes in this figure represent the relationship of actual crack sizes to each other. By sampling from this distribution a POD vs. a curve may be generated. This method has been used to model mean eddy current capability in the Retirement for Cause simulator with some success, since the simulator keeps track of actual crack size. Multiplying the ratio obtained from sampling the distribution by actual crack size gives "indicated crack size" for simulation purposes.
Fig. 18 Actual crack size (x-axis) vs. indicated crack size (y-axis)

REFERENCES


Fig. 19 Probability plot showing ratio of indicated to actual crack size vs. cumulative percent. Symbol sizes are proportional to actual crack size.


DISCUSSION

S. Bush (Battelle, Pacific Northwest Laboratory): I'd like to make an analogy of what we just heard. In regard to aircraft gas turbines, I might comment that the same problem currently exists on somewhat larger turbines, namely large steam turbines. One of the aspects of this is the hypothetical example that's not supposed to happen for aircraft gas turbines--namely, a very short initial period in some instances and a variable crack growth rate. One aspect then, of course, is that you can vary the time interval for your nondestructive examinations, and in your case, you were apparently holding them constant. A time interval change can have some obvious impacts, too. Have you investigated this particular one with regard to the effect on type 1 or type 2 errors?

C.G. Annis (Pratt & Whitney Aircraft): You are right, the interval itself can have a big influence, especially as a part gets older. For the sake of simplicity today, I've kept that constant to see how it affected the other variables in which we were interested.

R.B. Thompson (Ames Laboratory): In your conclusions, you said that the survival probability approach to a constant after several inspection intervals is independent of the NDE--

C.G. Annis: I didn't mean to say independent of the NDE. It is independent for a given value of NDE that approaches the statewide, not necessarily the same place.