Three Faces of Aeronautical Fatigue

Abraham Brot

A. Brot, Engineering Consultants, abrot99@gmail.com

Abstract: Similar in concept to the billions of human faces that exist, each one being unique, there are many "faces" of aeronautical fatigue, each having unique features. Three such *faces* have been selected for this paper, without any obvious attempt to connect one to the others. The first face is called "We Had Eyes but Could Not See", which deals with a perplexing problem that occurred nearly 50 years' ago. The second face is called, "Weibull or Log-Normal Distributions to Characterize Fatigue Life Scatter: Nearly Identical Twins or Distant Cousins?" This deals with a common misconception about these two statistical models. The third face is called, "Is There an Acceptable Risk for Widespread Fatigue Damage? This, of course, is a rhetorical question.

FACE 1: WE HAD EYES BUT COULD NOT SEE

About three thousand years ago, King David, the second king of Israel, reigned for 33 years in Jerusalem, and there he composed the book of Psalms. In Psalm 115 and again in Psalm 135 he stated: *"They have eyes but cannot see"*, in which he referred to idol worshippers and their idols. I chose to call the "incident", which I will now describe, *"We had eyes but could not see."*

Many consider the fourteen-year period from 1958 to 1972 as the *golden age of commercial aviation*. At the start of this period, jet transports began to be introduced into airline service, and by the end of the period, *seventeen different turbojet aircraft models were in sustainable airline service!*

In the U.S., Boeing introduced the 707, 720, 727, 737 and 747 models into service, all between 1958 to 1970. During the same period, Douglas introduced the DC-8, D-9 and DC-10, while General Dynamics introduced the Convair 880. Near the end of the period, Lockheed introduced the L-1011 into airline service.

In the U.K., Vickers placed the VC-10 into airline service, while the British Aircraft Corporation introduced the BAC-111, and de Havilland put the Comet 4 into service, all during these 14 years. During the same period, Sud Aviation from France introduced the SE 210 Caravelle airliner into service. Russia had the distinction of introducing the *first jet aircraft into sustainable airline service* in 1956, the Tupolev Tu-104, and later the Tu-124. Ilyushin also placed the IL-62 into airline service during this period.

These turbojet or turbofan powered airliners fly higher and faster than the existing propeller driven aircraft, giving the passenger a quicker and more comfortable ride. A turbojet powered airliner flying from New York to London would be in the air for only 6.5 hours, compared to about 13.5 hours for a propeller driven airliner. Previously, a westbound transatlantic propeller driven flight may have needed to refuel in Shannon (Ireland) or Gander (Newfoundland), making a long flight even longer. This scenario is quite rare for turbojet powered transatlantic airliners. For these reasons, there was much demand in the sixties for turbojet powered airliners, as well as for the engines that powered them.

To contrast this "golden age" to the more modern era, Boeing, who put five aircraft models into airline service between 1958 and 1970, only introduced four additional aircraft models (757, 767, 777 and 787) in the *nearly fifty years* since the introduction of the 747.

This was a very good time to be working in the jet transport industry, and in 1961 I started working for a large jet engine manufacturer as a structural engineer. This company designed and manufactured many of the jet engines that powered several of the aircraft that were mentioned above.

In 1969, an incident occurred which I would like to describe. The company received many orders for a specific aircraft engine model. The demand for engines was so high, that the company employed *subcontractors* to manufacture several of the complex engine parts, including the "high compressor duct". This specific duct was manufactured by the subcontractor in parallel with "in-house" manufacturing. Figure 1 shows the high-compressor for this engine.

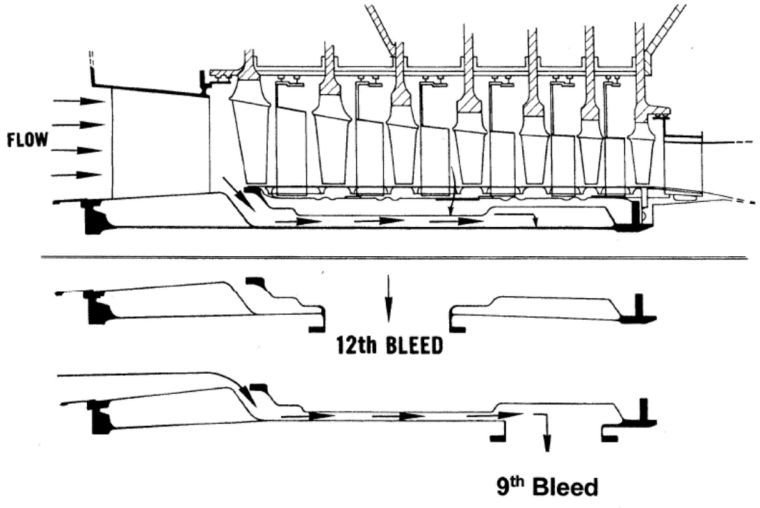

Figure 1. High-compressor configuration.

Figure 1 shows both the static structure (duct) and rotating structure (rotor) for this engine. Since every aircraft requires a source of compressed-air, this engine provided compressed-air both from the 9th stage rotor (9th bleed) and the 12th stage rotor (12th bleed). The subcontractor manufactured the entire high-compressor duct from drawings prepared by the engine manufacturer. The duct was manufactured from several steel parts that were welded together. Ducts produced *in-house*, as well as those produced by the *subcontractor*, were quickly installed on the engines, which were then shipped to the aircraft manufacturer for installation on his aircraft.

Soon after the engines entered service, several airlines reported fatigue cracks and failures of the high-compressor duct. Figure 2 shows a typical in-service failure resulting from high-cycle fatigue. Analysis of these failures disclosed that *nearly all of the fatigue failures occurred on ducts manufactured by the subcontractor.*

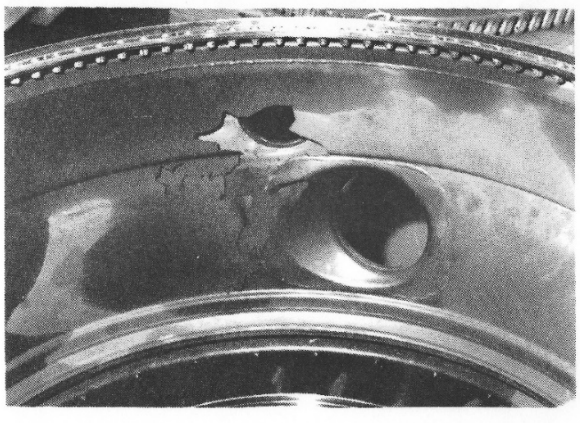

Figure 2. Typical high-cycle fatigue failure of the high-compressor duct at the 9th bleed chamber.

A task-force of three engineers were appointed to study the matter. Two were structural engineers, while the third member was an engine testing specialist. The service data was received and it was decided that the first step of the

investigation was to compare several ducts produced by the subcontractor and those produced in-house, against the design specifications. Among the manufacturing parameters that were checked were: complying with material specifications *(composition as well as heat-treatment),* complying with the dimensional data specified in the drawings and complying with the quality of the welding compared to the specifications. ***It was found that all the ducts produced in-house, and all those produced by the subcontractor met all the requirements.***

It was then decided to perform an extensive series of engine tests, with engines having ducts manufactured in-house and by the subcontractor. This extensive testing took many months, as is described in Figure 3. It became very difficult to meet contractual manufacturing requirements during this period, since only in-house produced high-compressor ducts were being installed on engines.

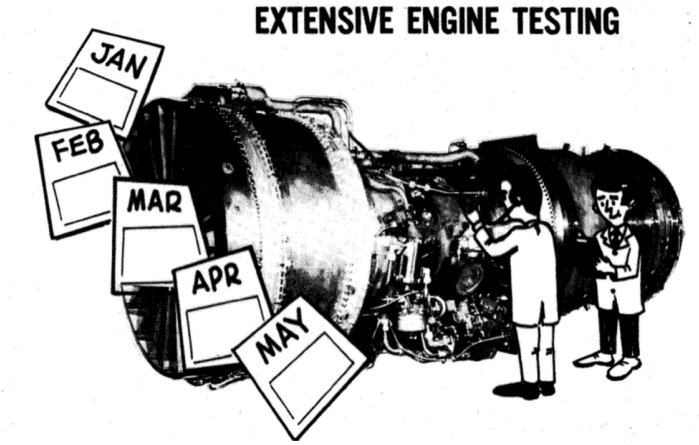

Figure 3. Many months went by while extensive testing was performed**.**

Much important information was acquired from the testing. In these engine tests, an engine was placed on a test-stand and the rotors were slowly accelerated from idle speed to takeoff speed. Figure 4 shows a typical result of engine testing using a high-compressor duct supplied by the subcontractor.

As is shown in Figure 4, the engine was slowly accelerated from idle to takeoff speed *(5,000 RPM at idle to 10,000 RPM at takeoff, for the high rotor)*. The subcontractor supplied high-compressor duct was instrumented to measure stresses on the high-compressor duct, noise levels inside the duct, and vibration levels of the duct.

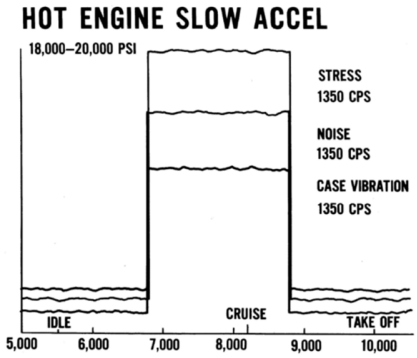

Figure 4. Typical test result acquired by engine testing.

Figure 4 shows that when the high-rotor reached about 6,800 RPM, the pressure-transducers began "singing" at 1350 Hz (cps), while the strain-gages and accelerometers were "dancing" at the same frequency. The peak measured noise levels reached 170 dB, while the measured stress levels on the duct were 18 to 20 ksi, which certainly could explain the failures. As the high-rotor speed reached 8,800 RPM, all the measured parameters suddenly quieted down.

An analysis was performed on the measured noise content inside the duct, and the results clearly show that several frequencies were present, as is shown in Figure 5, but the highest noise level (170 dB) was basically at 1350 Hz.

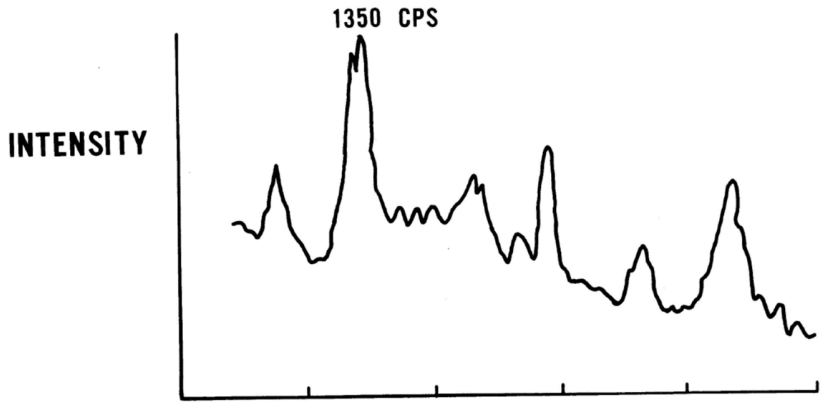

Figure 5. Noise frequency content measured in the high-compressor duct.

An additional discovery was made, as is shown in Figure 6. It was found that while compressed air was bled from the 9th stage bleed port, the engine became quiet, as compared to the noisy engine when air was not bled from the 9th stage. All this information was very pertinent in understanding the problem, and ultimately, helped develop the solution.

A study was performed by inserting several levels of damping material into the 9th stage bleed chamber. Figure 7 shows the results of the study. Figure 7 showed that sufficient damping introduced in the bleed chamber would tend to quiet the engine for various values of engine air-flow. Damping was introduced in the bleed chamber by wrapping wire cables inside the chamber.

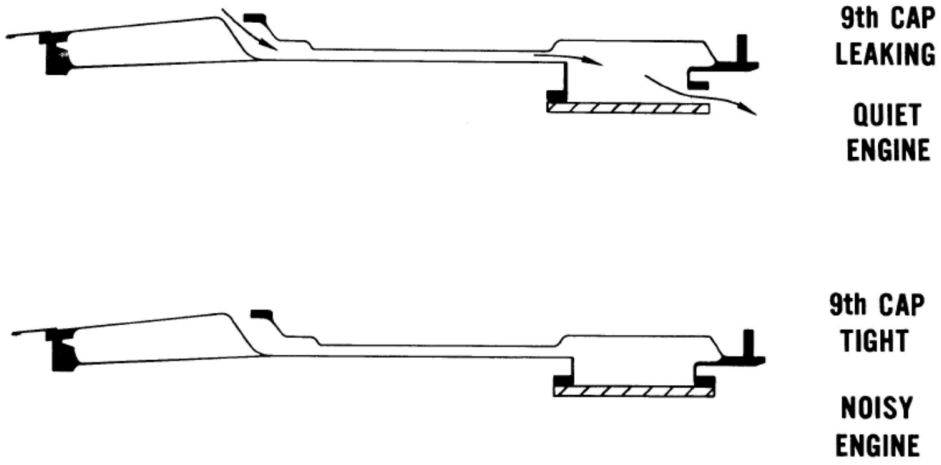

Figure 6: Effect of passing compressed air through the 9th stage bleed chamber.

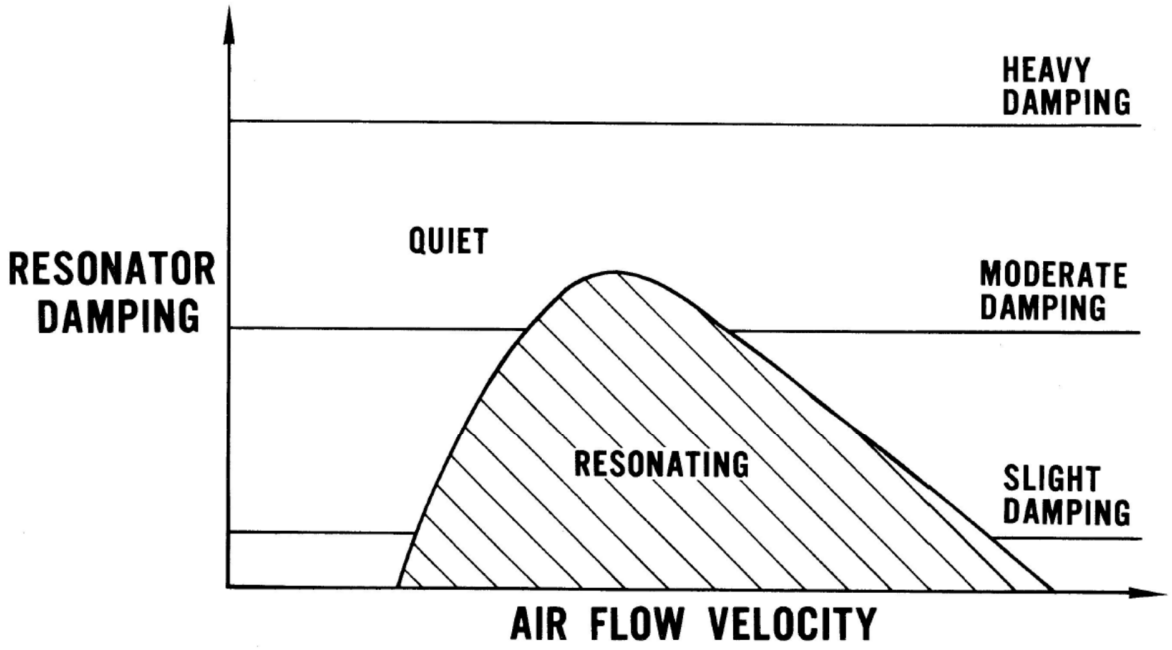

Figure 7. Effect of damping on the duct noise and vibration as a function of engine air-flow velocity.

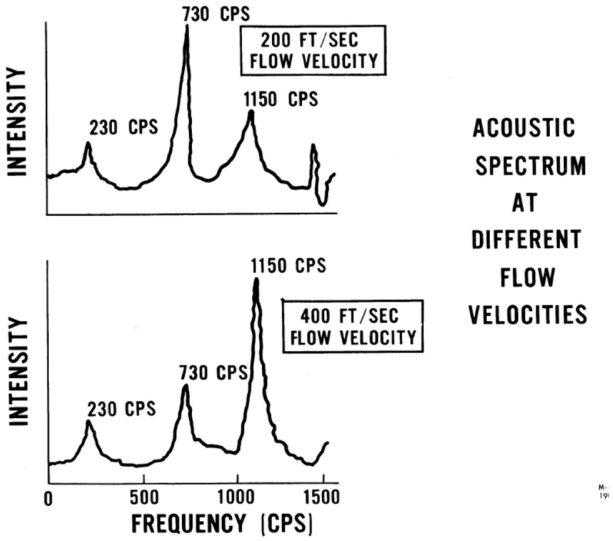

Figure 8. Test rig results used to measure the acoustic spectrum for various levels of air flow.

A test-rig was constructed to study the effect of engine air-flow, totally independent of the engine testing. The results shown in Figure 8 clearly indicate that the acoustic resonant frequencies of the chamber remained the same for different levels of engine flow velocity, but the intensity of each resonant peak was clearly a function of the air-flow velocity. As is shown in Figure 8, for a flow velocity of 200 ft/sec, the 730 Hz peak is dominant while for a flow velocity of 400 ft/sec, the 1150 Hz peak becomes dominant. This helps explain why the engine became quieter or noisier as a function of the engine air-flow velocity, as was shown in Figure 7.

At this point it was clear that we were dealing with a *self-destructing acoustic resonator* in the 9th bleed chamber, but we were no closer to understanding why the ducts produced by the subcontractor were "noisy" while the in-house produced ducts were "quiet".

One day, as the two structural engineers of the task-force were discussing recent test results, both simultaneously blurted out the first step towards the solution of the mystery. They both said, ***"we compared the in-house and subcontracted ducts against the specifications, but we never compared them to each other!"***

It was then decided to perform comparative dimensional checks of several ducts that were representative of the two manufacturers. This proved to be the "breakthrough". The significant results are shown in Figure 9.

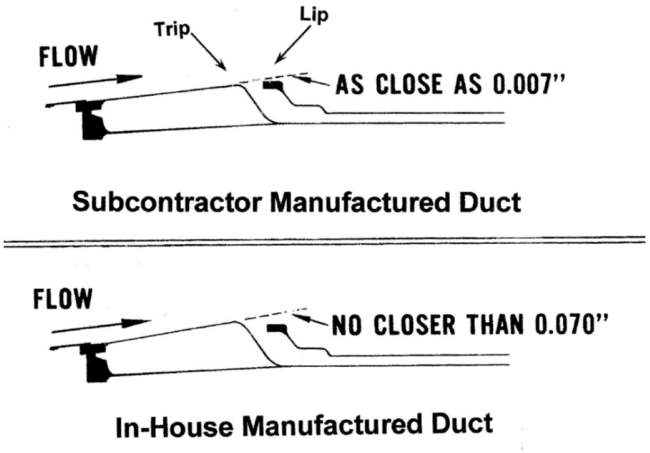

Figure 9. The significant difference between the subcontractor and in-house manufactured ducts.

Figure 9 shows the relative position of the "trip" and the "lip" for the two manufactured ducts. For the subcontractor manufactured ducts, the distance from the trip-to-lip was *as close as 0.007 inch*. While for the in-house manufactured ducts, the corresponding distance from the trip-to-lip was *no closer than 0.070 inch.*

It should be noted that the high-compressor duct design was such that the lip structure was "floating" and it connected up with the main duct aft of the 9th bleed port, as is shown in Figure 12. In view of this, the critical distance shown in Figure 9 was not "called out" on the drawing. As a result, there were very *large cumulative tolerances* between the trip and the lip, as is shown in Figure 9. The significance is that both types of ducts *(subcontractor and in-house manufactured)* met the required specifications, in spite of the differences shown in Figure 9.

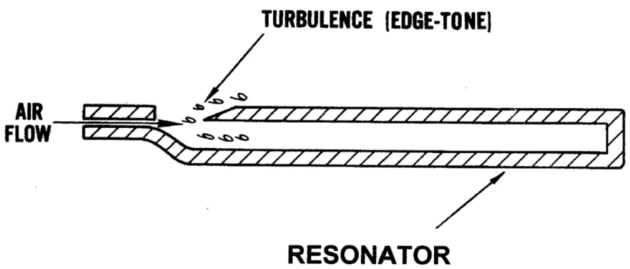

Figure 10. The excitation mechanism causing resonance of the chamber.

The excitation mechanism was now clear, as is shown in Figure 10. Engine air-flow passing from the trip to the lip produced turbulence *(edge tone)* which excited the acoustic frequencies of the 9th stage bleed chamber, with the chamber acting as an "organ pipe". The configuration of "trip-to-lip" for the in-house manufactured ducts resulted in no significant excitation, therefore, the chamber remained quiet. The configuration of "trip-to-lip" for the subcontractor's ducts resulted in a very significant excitation, and the chamber becomes very noisy, resulting in a violent near-resonance response of the 9th bleed chamber, which caused the fatigue cracking and failure of its structure.

At this point, ***"we had eyes and <u>now</u> we could see"!***

Further engine testing demonstrated that a non-linear "feedback mechanism", which caused the engine air-flow to "lock-in" with the reflected wave in the chamber, could greatly magnify the resonance, as is described in Figure 11. It was determined by analysis that the 1350 Hz measured resonant frequency corresponds to one-and-a-quarter wave lengths along the length of the 9th bleed chamber, which is a typical resonant mode for chamber having one closed end and one open end.

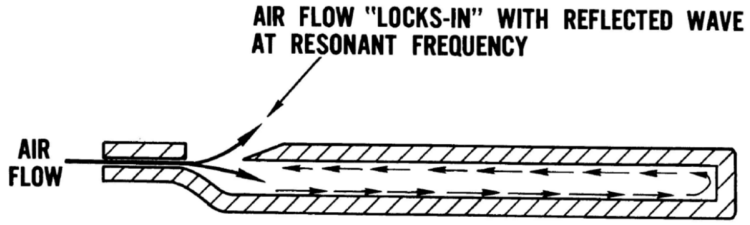

Figure 11: Non-linear feedback mechanism which enhanced the resonance.

Figure 12 summarizes the various parameters which contributed to the failures, and led to the solution, as was discovered by engine and rig testing. As is shown, the resonance phenomenon required a critical engine air-flow velocity. In addition, the excitation mechanism described in Figure 10 *needed to exist* in order to cause the resonance. It was shown by testing that bleeding compressed air through the 9th bleed chamber can effectively quiet the resonance. Sufficient acoustic damping (wire cables) was also shown to be effective in quieting the resonance.

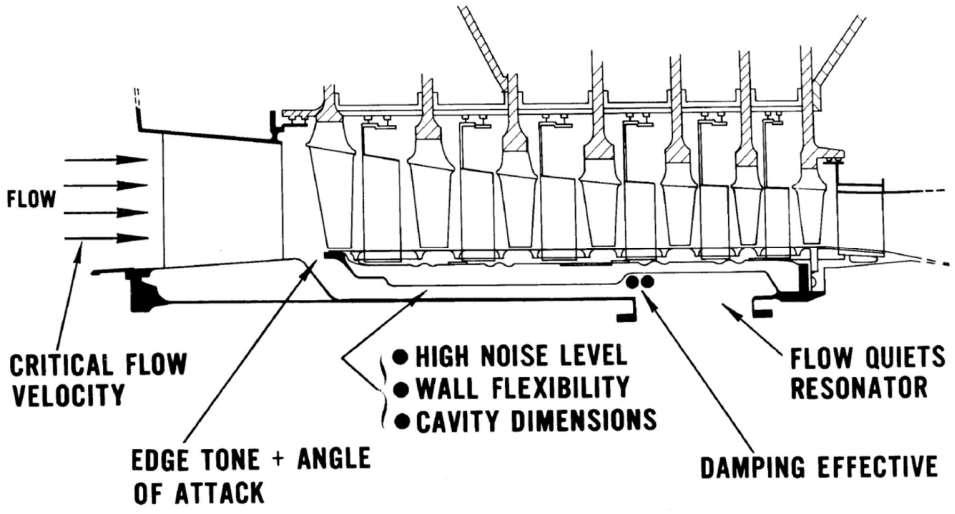

Figure 12. The significant parameters that caused the resonant phenomenon and led to the solution.

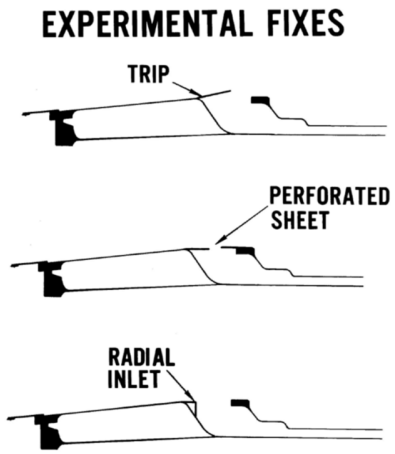

Figure 13: Experimental fixes that were found during the engine testing.

Figure 13 shows several configurations that effectively quieted the resonance during engine testing. The trip angle was enhanced by adding sheet-metal. Perforated sheets and a "radial input" were used to better direct the flow, as is shown in Figure 13.

Production fixes were selected based on the test results, as is described in Figure 14. These include making modifications to the lip, as well as providing enhanced acoustic damping by wrapping wire cables around the chamber. These fixes were installed in existing engines. Of course, the high-compressor duct drawings were modified to control the trip-to-lip gap during production, as was shown in Figure 9.

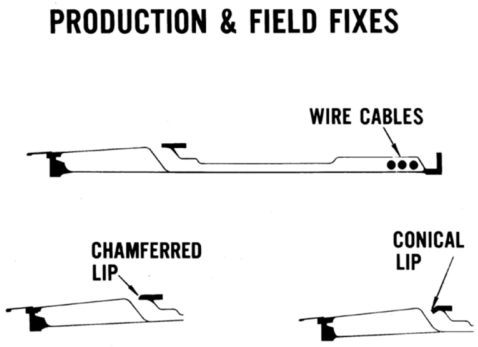

Figure 14. Production fixes used to quiet the resonator in existing engines

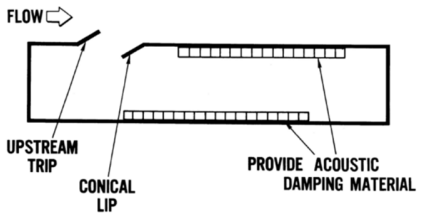

Figure 15. Lessons learned aid in designing a chamber that will not be a self-destructing acoustical resonator.

Figure 15 summarizes some of the lessons learned during this experience. Carefully designing the trip-to-lip configuration, and/or providing sufficient acoustic damping, can effectively eliminate the *self-destructive resonator concept.*

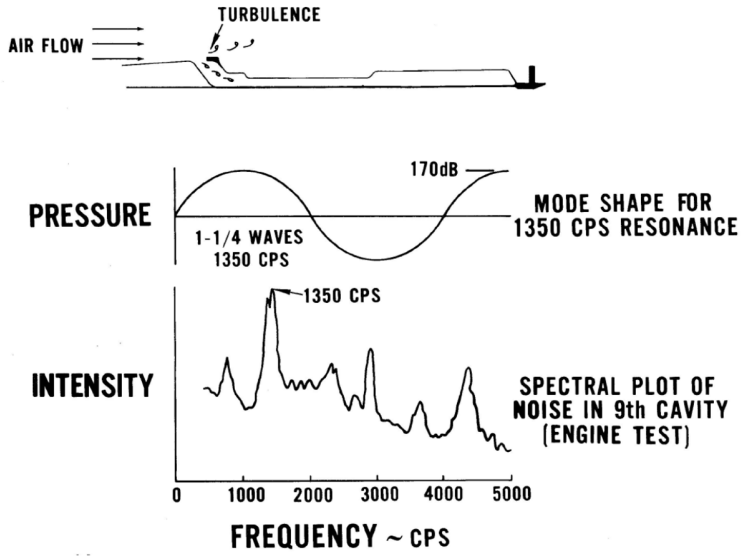

Figure 16. Excitation mechanism, pressure mode shape in the 9th bleed chamber and the resulting noise spectrum.

Figure 16 summarized the *self-destructive acoustic resonator* phenomenon, including a description of the excitation mechanism, the one-and-a quarter pressure waves that existed in the 9th bleed chamber, and the relative intensity of the noise spectrum, which reached 170 dB.

Until this problem was solved, more than sixty high-compressor ducts failed in service. This problem took about ***eight months*** to solve, at a cost of about $700,000 (1969 dollars), which today is equivalent to approximately ***$5,000,000.***

FACE 2: WEIBULL OR LOG-NORMAL DISTRIBUTIONS TO CHARACTERIZE FATIGUE LIFE SCATTER: NEARLY IDENTICAL TWINS OR DISTANT COUSINS?

Introduction

Over the last half-century, organizations that dealt with structural reliability related to fatigue life variability of aircraft structures, *generally* performed their reliability analyses by using the Weibull distribution. The Log-Normal distribution was considered to be more complicated, but *nearly identical to the Weibull distribution*, so it was less often utilized.

Damage-tolerance methodology has been used successfully for approximately forty years to certify both military and commercial aircraft. Nevertheless, there are many aircraft structures and situations where *fatigue life methodology* is used instead, to certify structural components. *(The term "fatigue life" is used in this paper to describe the process of computing the number of flights or flight-hours in service up to crack-initiation or nucleation.)* Several metallic structures or situations where *fatigue life methodology* is used for certification purposes are shown below:

1. Aircraft structures <u>not</u> classified as "Principal Structural Elements" (PSEs) nor "Critical Parts" on civilian or military aircraft.

2. Military aircraft that are designed for use *only by the U.S. Navy or other services that do not require damage-tolerance certification.*

3. Landing gear and other structural elements where certification to damage-tolerance methodology is *considered impractical*.

4. Civilian aircraft certified to FAR 23 *(up to nine passengers, maximum takeoff weight of 12,500 pounds).*

5. Metallic structures on *unconventional* aircraft, such as certain types of UAVs.

It should be noted that fatigue life certification generally requires some experimental confirmation of the results. One of the downsides of using fatigue life methodology is that considerable scatter exists in determining the life to crack-initiation. This needs to be accounted for in every analysis or test.

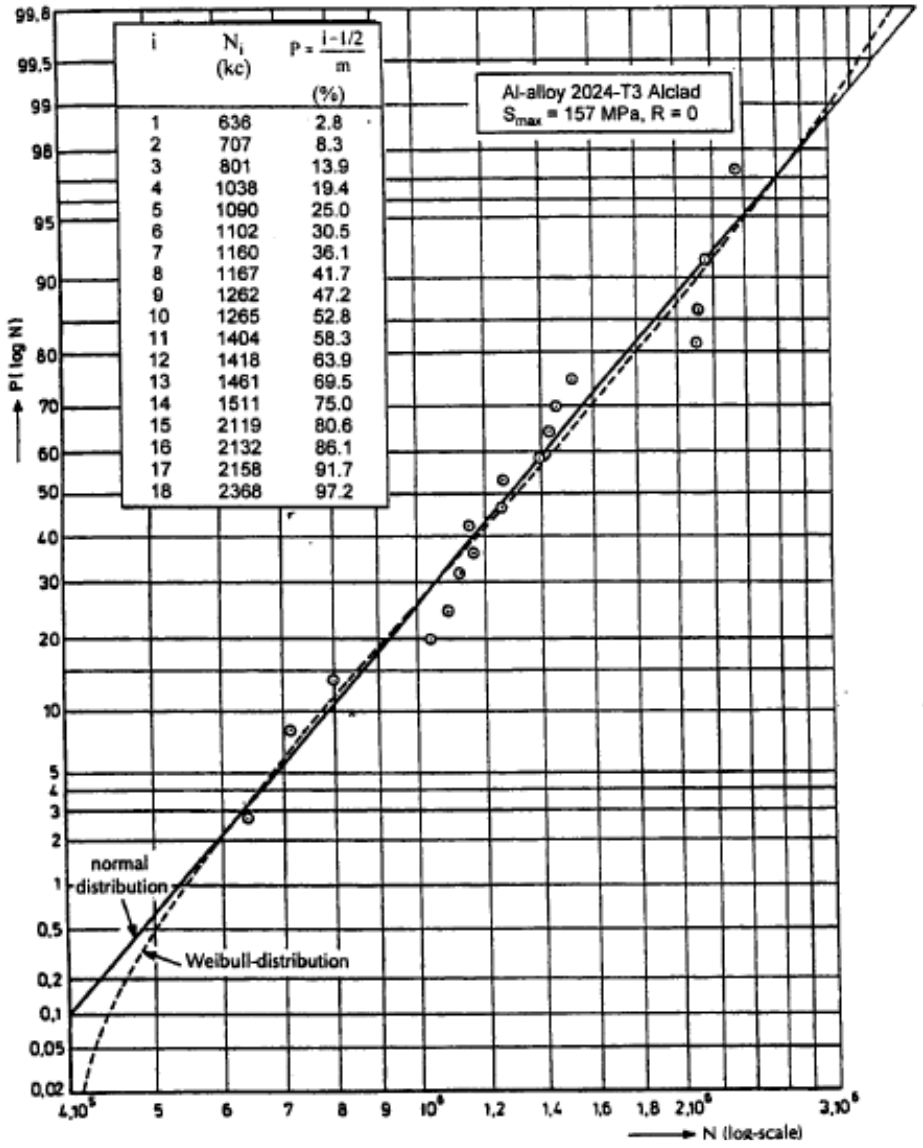

Figure 17. Fatigue test results for eighteen 2024-T3 aluminum specimens [1, 6].
Horizontal axis: Life to Failure (Cycles). Vertical axis: Cumulative Probability of Failure (%)

Both the Weibull and Log-Normal probabilistic distributions have been used to characterize fatigue life scatter. Figure 17 which was published by Schijve [1, 6], shows the results of a fatigue test that was performed in 1955 on eighteen identical 2024-T3 aluminum specimens. *(Please note that the fatigue life test results are stated in kilocycles, and are plotted against the ranking of each specimen, as is shown on the plot.)*

It should be noted that Figure 17 is plotted on "Log-Normal axes". This means that results that are described by the Log-Normal distribution will plot as straight lines. Therefore, the straight line on Figure 17, which passes through the data, defines a Log-Normal distribution. On the other hand, the dashed line on Figure 17, which also passes through the data, represents a Weibull distribution. Without a doubt, both distributions appear to adequately represent the test data.

If our purpose was only to find a mathematical function that closely represents the test data, either of the above distributions would be acceptable. Alternately, a polynomial could be defined to adequately represent the data, as is often performed by *Excel, in order to calculate a "trendline"*.

A close look at Figure 17 shows that at the lower left-hand corner of the plot as well as at the upper right-hand corner of the plot, the Weibull and Log-Normal distributions begin to diverge. Since we often use fatigue test results to determine *acceptable service lives* that need to have a *very low probability of failure,* this divergence between the two models *hints* that they may not give consistent results when the results are extrapolated downward.

Comparison of the Weibull and Log-Normal distributions for several fatigue tests that were performed
SuperSMITH Software [2] has been developed by *Fulton Findings* in order to analyze various statistical distributions, including obtaining their statistical parameters and plotting their results. It is based on the theoretical methods described in the *New Weibull Handbook* [3].

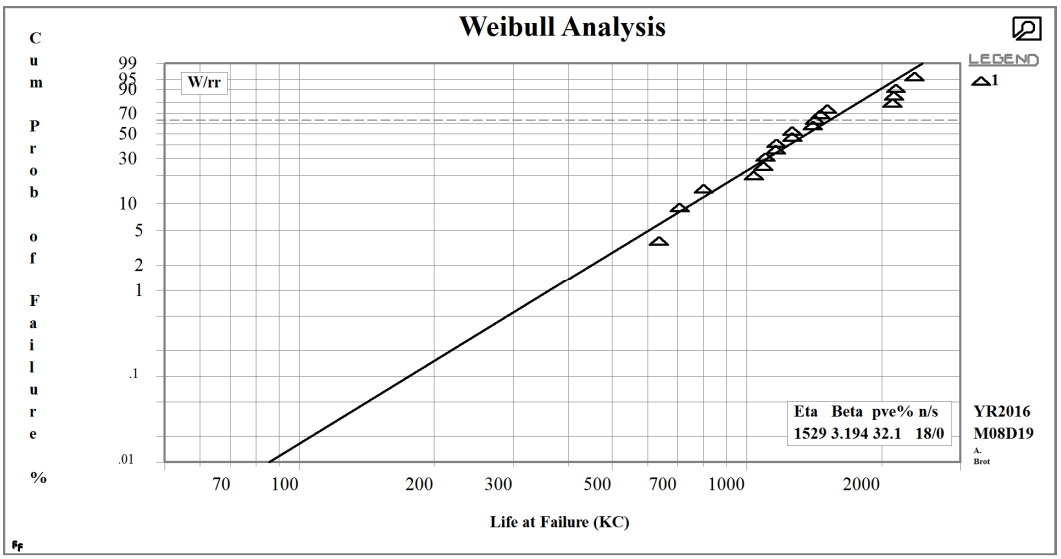

Figure 18. Weibull plot for the eighteen-specimen test *(including the calculated Weibull parameters)*.

The data for the above eighteen specimens, as were shown in Figure 17, was input into the SuperSMITH software and both Weibull and Log-Normal plots were obtained, as are shown in Figures 18 and 19. *(Please note that all the SuperSMITH runs shown in this paper are based on the "median rank" procedure that assigns a cumulative probability of failure to each failed specimen. Median rank methodology determines the best fit by least-square regression, and is usually recommended for analyses [3].)*

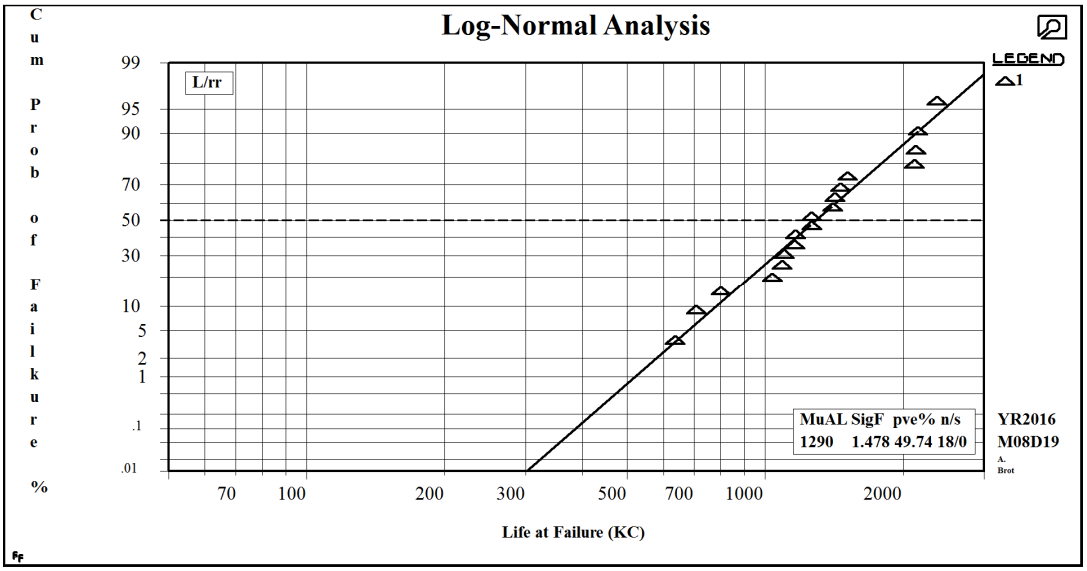

Figure 19. Log-Normal plot for the eighteen-specimen test *(including the calculated Log-Normal parameters).*

Comparing Figure 18 (Weibull) to Figure 19 (Log-Normal) shows *very large differences* for the fatigue life for the lower values of the probability of failure. For example: for a probability of failure of 0.01% the Weibull life is 86 kc while the Log-Normal life is about 300 kc. It should be noted that long-range statistical extrapolations will always be problematic. However, extrapolation differences as large as 350%, when using different distributions, are clearly not acceptable.

It is clear from both Figures 18 and 19 that the *four highest lives* seem not to be consistent with the remaining results. The data used to construct Figures 18 and 19 were rerun with SuperSMITH with the highest four tests *(outliers)* removed. The results are shown in Figures 20 and 21.

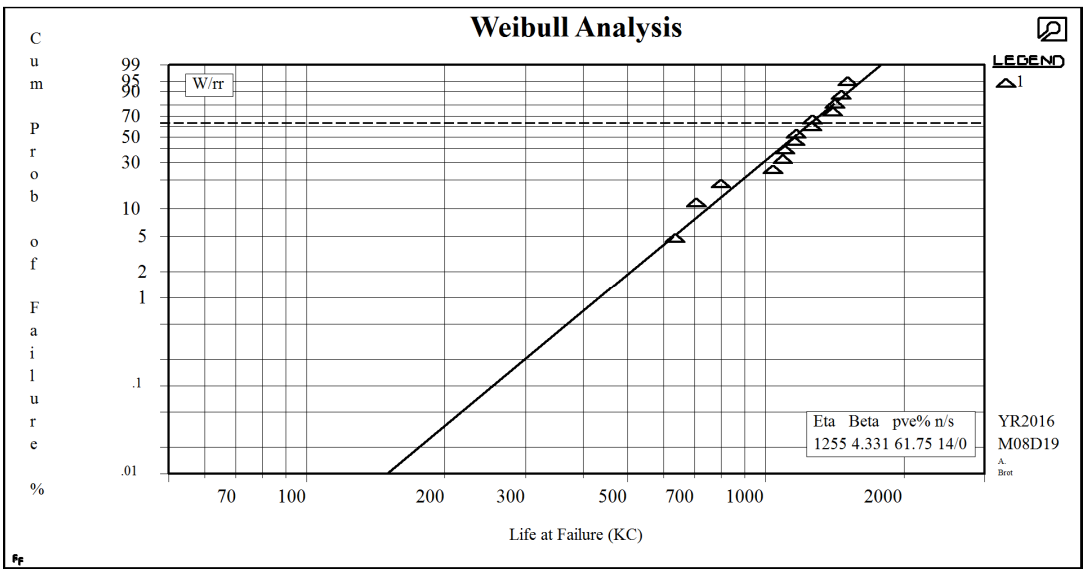

Figure 20. Weibull plot for the fourteen-specimen test *(including the calculated Weibull parameters).*

Comparing Figure 20 (Weibull) to Figure 21 (Log-Normal) still shows large differences for the fatigue life for the lower values of probability of failure. For example: for a probability of failure of 0.01% the Weibull life is 160 kc while the Log-Normal life is 400 kc. Clearly, the discrepancy in lives between the Weibull and Log-Normal runs *were not due to the four outlier results that were eliminated.* The results were recalculated and plotted log-log scales, as is shown in Figures 22 and 23. It is seen that the difference between the Weibull and Log-Normal results are increased as the allowable probability of failure is reduced.

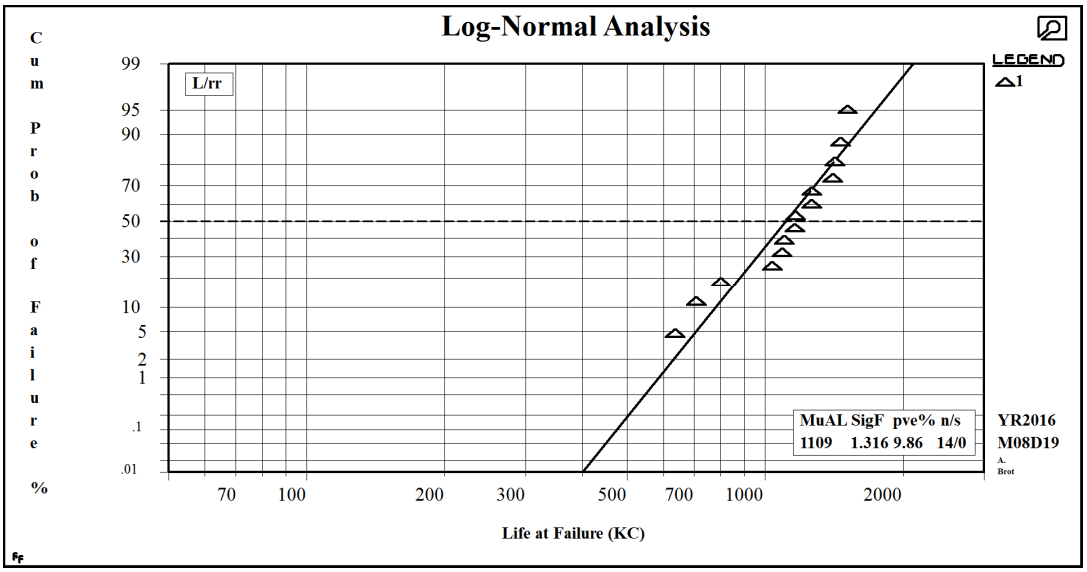

Figure 21. Log-Normal plot for the fourteen-specimen test *(including the calculated Log-Normal parameters).*

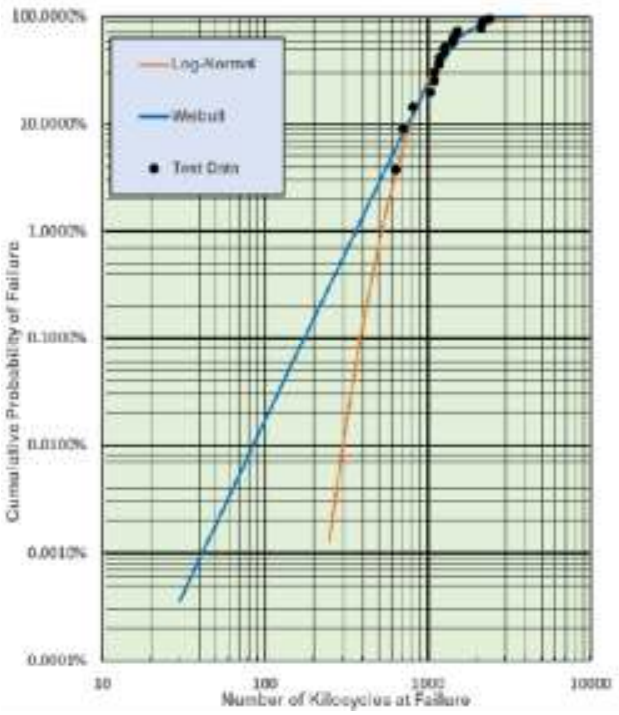

Figure 22. Comparison of the Weibull and Log-Normal results for the 18-specimen test.
(*Test results are plotted against "median rank")*

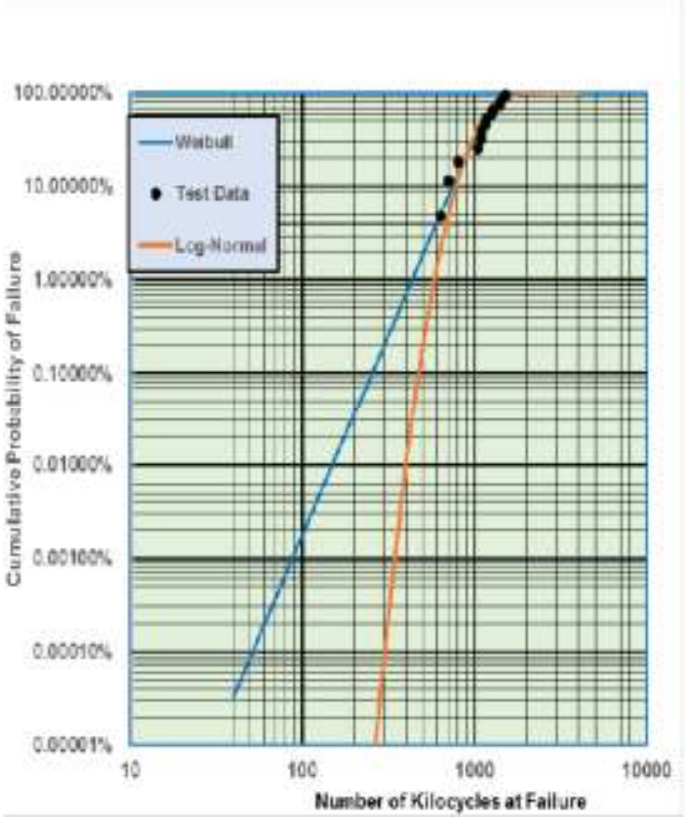

Figure 23. Comparison of the Weibull and Log-Normal results for the 14-specimen test.
(*Test results are plotted against "median rank")*

In another example, fatigue testing was performed in 2011 at Israel Aerospace Industries on ten fatigue specimens manufactured from 7075-T73 aluminum. The surface-finish for all these specimens, including the interior of their holes, was *very closely controlled* to ensure that each of the specimens was of very high quality.

Analysis, using SuperSMITH software, was performed, and the results are shown in Figure 24 and 25. The comparative results were recalculated and plotted on log-log scales on Figure 26.

The results are very similar to what was seen in Figures 22 and 23. Sizable differences in the fatigue lives, corresponding to the lower probabilities of failure, were again seen when comparing the Weibull and Log-Normal results. Note the steepness of the lines in Figure 26 resulting from the low scatter due to the very high quality of the specimens resulting from the enhanced surface-finish of the specimens.

Figures 18 –26 all show that that there are very sizable differences between the Weibull and Log-Normal distributions in setting allowable lives to minimize the probability of failure. *In all cases, the Weibull results are much more conservative than the Log-Normal results.*

However, from these results it is impossible to say which distribution is the more accurate. In actuality, the true result could be to the left of the Weibull result, between the Weibull result and the Log-Normal result or even to the right of the Log-Normal result.

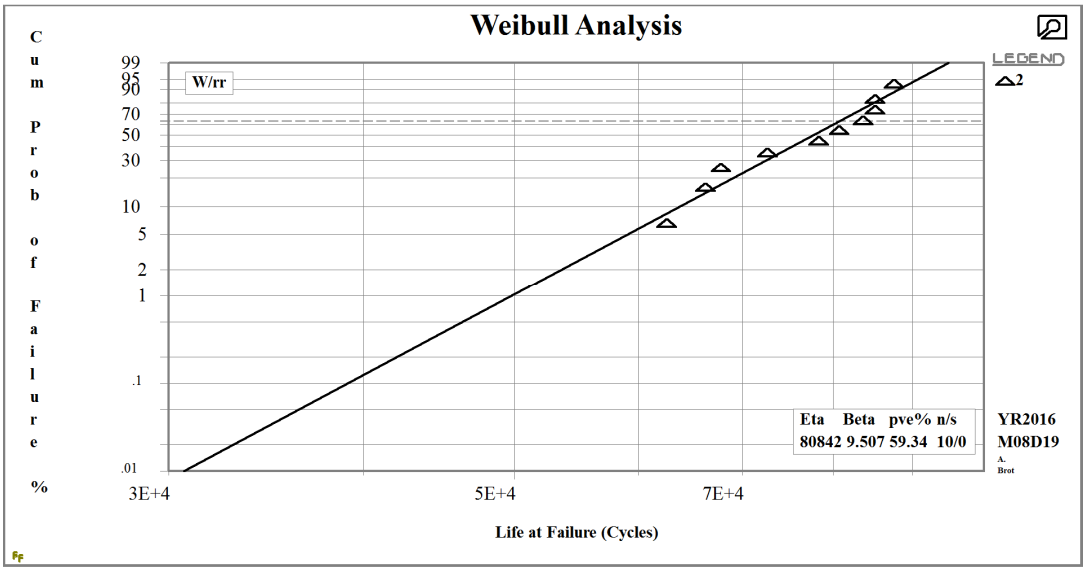

Figure 24. Weibull plot for the ten-specimen test *(including the calculated Weibull parameters).*

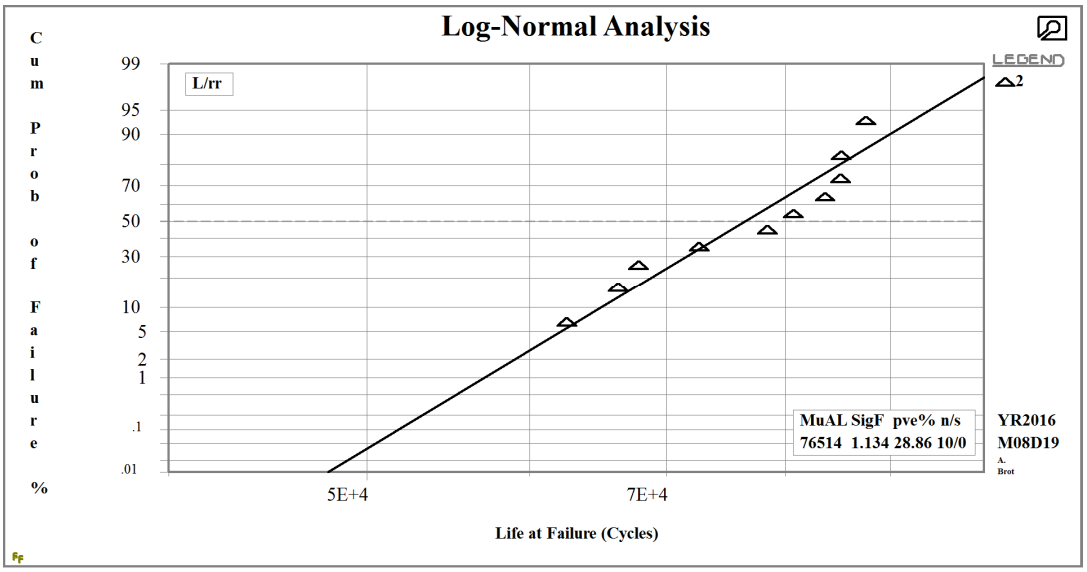

Figure 25. Log-Normal plot for the ten-specimen test *(including the calculated Log-Normal parameters).*

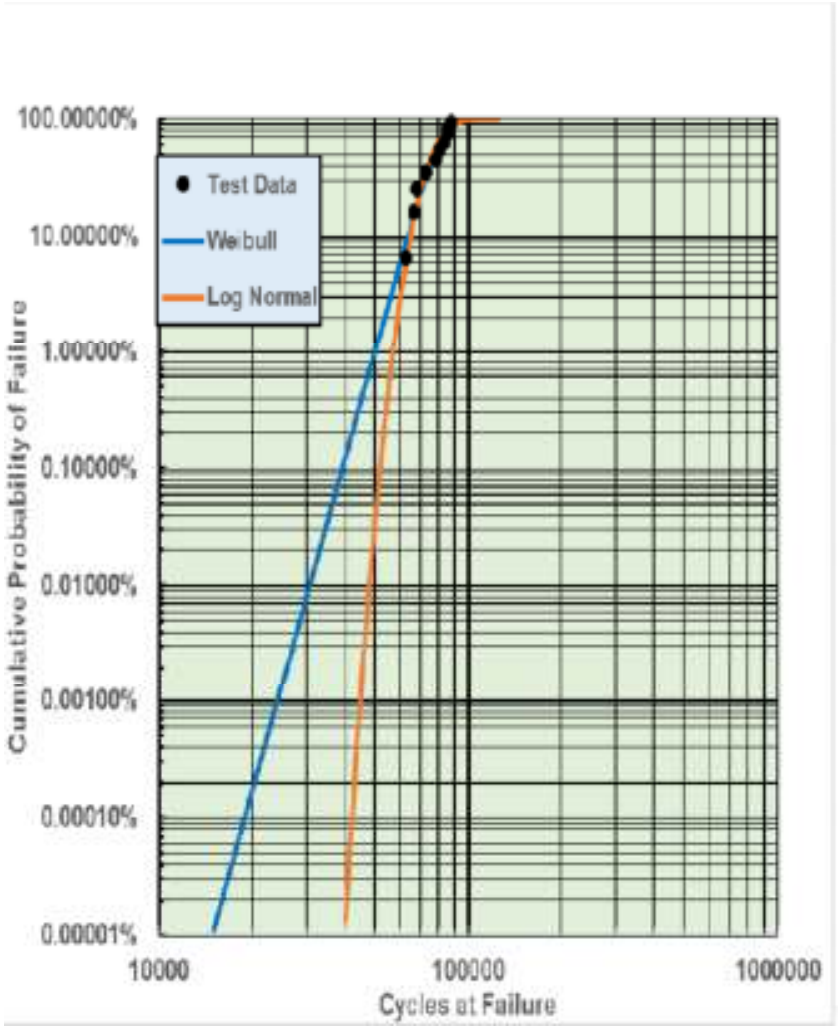

Figure 26. Comparison of the Weibull and Log-Normal results for the ten-specimen test.
(*Test results are plotted against "median rank")*

Fatigue Life Substantiation of a Steel Landing Gear

It has been FAA and EASA practice to exempt landing gears from a damage-tolerance certification due to the impracticality of such a substantiation. Instead, a five-lifetime fatigue test is generally employed to certify the landing gear. After the landing gear survives the five-lifetime fatigue test, performed to a realistic spectrum, it is certified as a *"safe-life structure",* as described in FAR 25.571(c). The implications are, that after completing one design lifetime in actual service, the landing gear *must be retired from any future service*.

Landing gears are usually designed to have many critical parts manufactured from steel, which has a larger degree of scatter than aluminum [5]. A typical main landing gear, having steel parts, has been selected for this study. The landing gear has been designed to provide an allowable service life of 20,000 flights. Therefore, its fatigue test is scheduled to reach 100,000 flights. The landing gear is characterized the by a Weibull distribution having a characteristic life of 100,000 flights and a shape-factor of 3.0, which is typical for steel parts [5]. Figure 27 shows the Weibull plot for the landing gear. This approach implies, with the absence of any other data, that the landing gear test would have failed during flight 100,001. As a result of the successful completion of the five-lifetime fatigue test, Figure 27 shows that a 0.79% probability of failure can be expected during the 20,000-flight service life, according to the selected Weibull parameters. If we include a lower-confidence-level of 95%, that accounts for the single test result, the expected probability of failure could increase to 3.8%, as is shown in Figure 27.

The corresponding Log-Normal plot is shown in Figure 28. The Log-Normal parameters have been automatically computed by the SuperSMITH software. Figure 28 shows that only a 0.0005% probability of failure can be expected during the 20,000-flight service life, according to the calculated Log-Normal parameters. If we include a lower-confidence-level of 95%, that accounts for the single test result, the expected probability of failure could increase to only 0.056%. The Log-Normal distribution shows a much lower probability of failure than the Weibull distribution.

The comparative results *(Weibull and Log-Normal)* were then recalculated and plotted on Log-Log scales on Figure 29. Clearly, the Log-Normal distribution predicts a *much lower probability of failure* during its 20,000-flight service life.

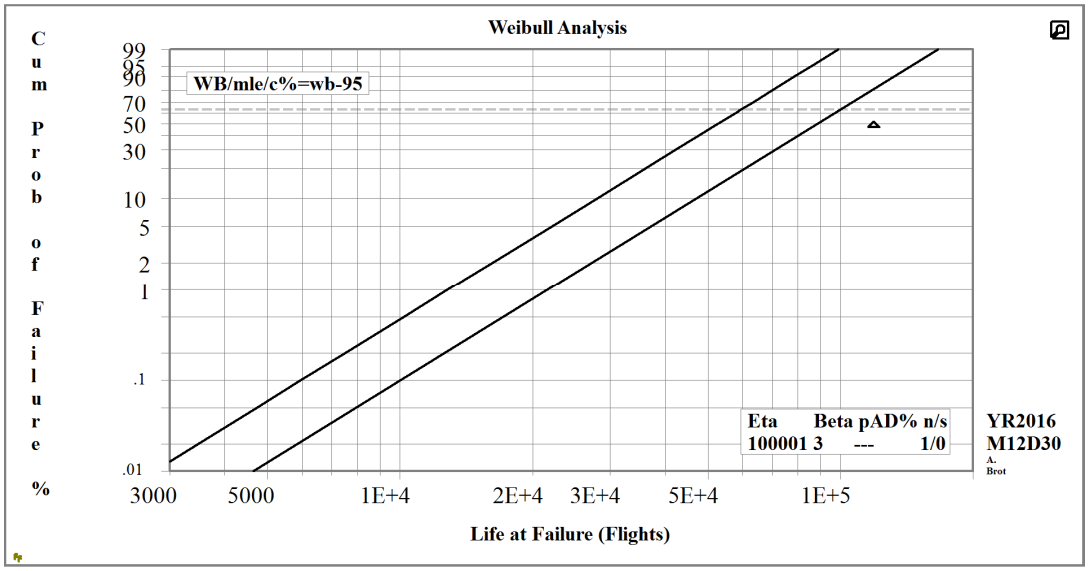

Figure 27. Weibull plot for the steel landing gear test. The upper curve accounts for a 95% lower-confidence-level.

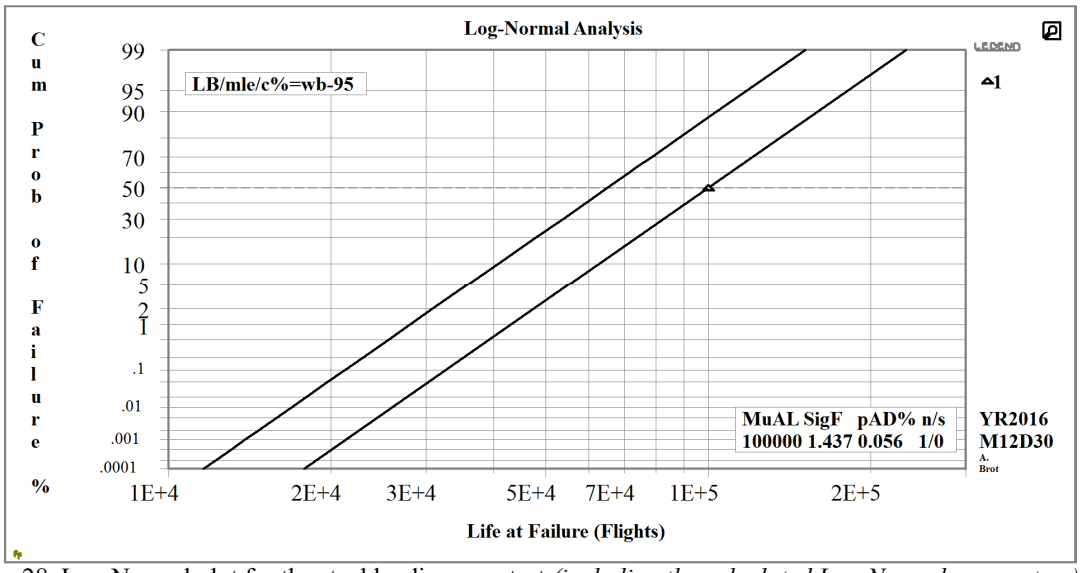

Figure 28. Log-Normal plot for the steel landing gear test *(including the calculated Log-Normal parameters).* The upper curve accounts for a 95% lower-confidence-level.

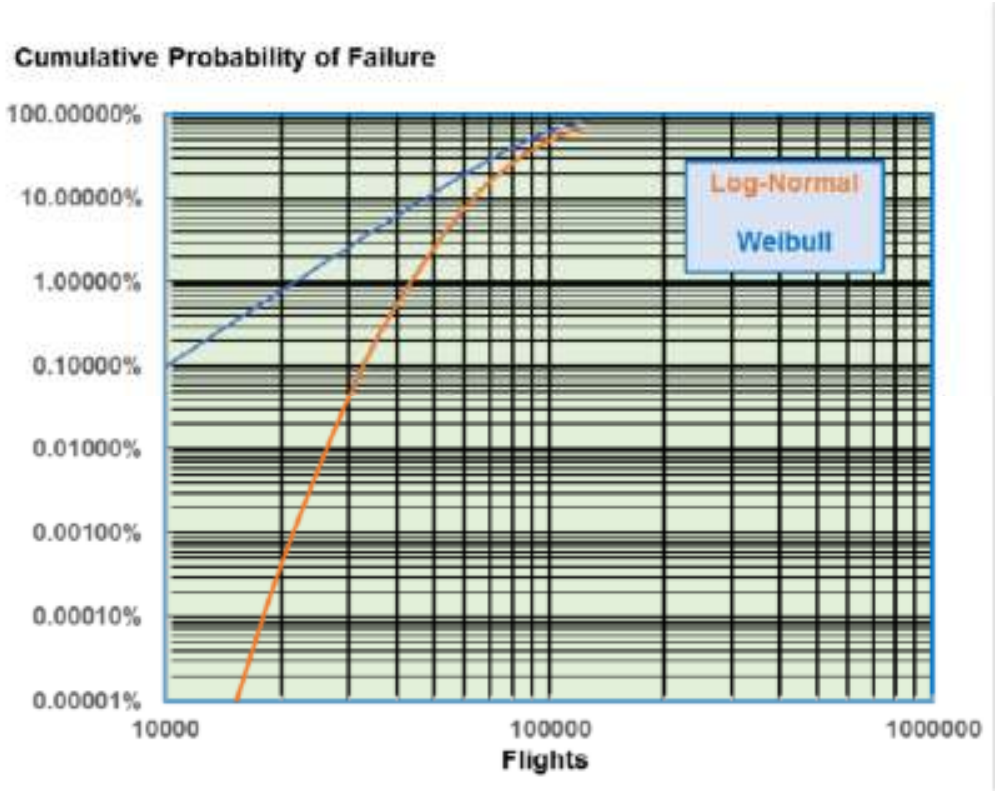

Figure 29. Comparison of the Weibull and Log-Normal results for the steel landing gear test.

The Weibull results present a somewhat alarming situation. If we include the 95% lower-confidence-level analysis, the cumulative probability of failure, within the service life of 20,000 flights, is 3.8%, which is equivalent to a structural reliability of *only 96.2%.* The corresponding structural reliability according to the Log-Normal distribution is *greater than 99.9%,* which is much more reassuring.

The examples shown above used two important terms: "life to failure" and "life to crack-initiation" somewhat interchangeably. In actual circumstances, there may be considerable differences between the two terms, since life to failure also includes some crack growth before the failure. For structures where crack growth is only a *small fraction* of the failure life, no action needs to be taken. For structures where extensive cracking can be expected before failure, it is best to install crack wires or crack gages on the test specimen in order to separate the crack-initiation and crack growth phases.

For the "ten specimen fatigue test" described previously, crack wires were installed to separate between the crack-initiation and crack growth phases. The results were very satisfactory.

Effect of Selecting a "Lower-Confidence-Level" of 95%

Fatigue life scatter analysis usually includes selecting a lower-confidence-level of 95%, as was discussed in the previous section. This insures that if a test program based on a very large number of specimens *would have been performed*, there is 95% confidence that the results would not be lower than the actual computed values including the lower-confidence-level.

Figure 30 shows the results of a five-specimen test that was analyzed using the Weibull distribution. The service life criterion was 99.5% reliability *(allowable failure rate of 0.5%)* with a lower-confidence-level of 95%. Figure 30 shows that the allowable service life would be 3,680 cycles, which represents an overall scatter-factor of 5.4 compared to the mean result of the five test specimens (19,900 cycles).

Figure 31 shows the results of the same five-specimen test that was analyzed using the Log-Normal distribution. The service life criterion was 99.5% reliability *(allowable failure rate of 0.5%)* with a lower-confidence-level of 95%. Figure 31 shows that the allowable service life would be 7,400 cycles, which represents an overall scatter-factor of only 2.7 compared to the mean result of the five test specimens (19,900 cycles).

This again demonstrates the large differences obtained between the Weibull and Log-Normal distributions, with the Weibull distribution always being much more conservative than the Log-Normal distribution.

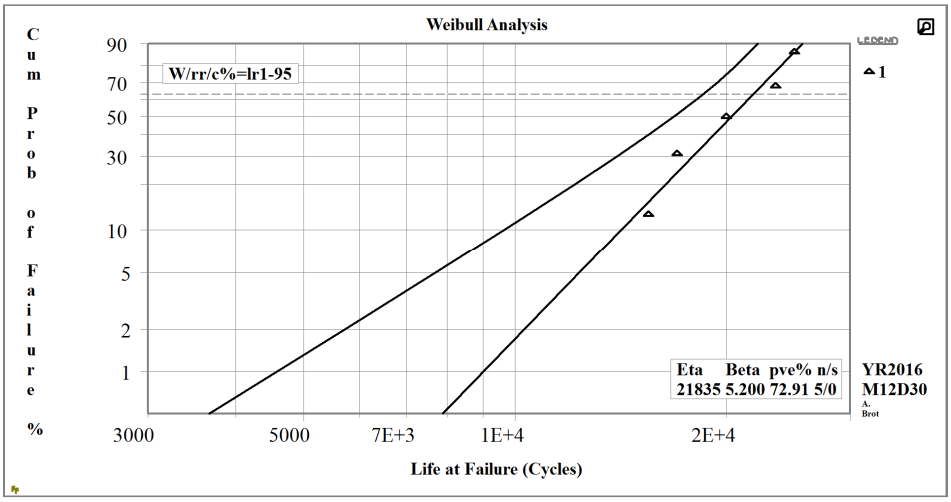

Figure 30. Weibull analysis of a five-specimen test.
(The curve to the left represents the lower-confidence-level of 95%.)

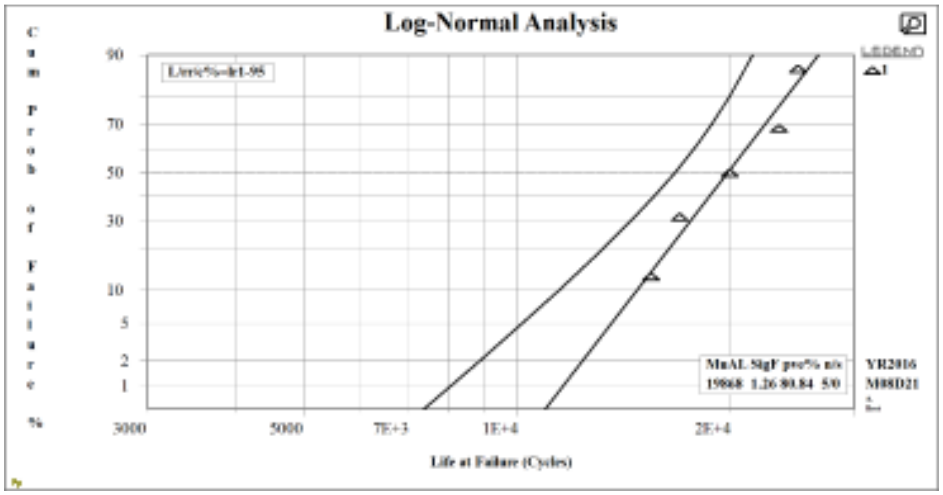

Figure 31. Log-Normal analysis of a five-specimen test.
(The curve to the left represents the lower-confidence-level of 95%)

Research Performed by Boeing on Fatigue Life Reliability

In 1969 the Boeing Company published the results of a study that was performed on fatigue life variability of aluminum aircraft structures [4]. This study developed the methodology to determine fatigue life variability (scatter) using the two-parameter Weibull distribution. The results of this report became the basis of Boeing's methodology to determine required scatter-factors for safe-life aircraft structures [7]. In this Boeing report, a brief comparison between

the Weibull and Log-Normal distributions was made. It was concluded that the Weibull distribution generally results in a lower estimate of the safe-life than the Log-Normal distribution.

In 1972, Boeing issued a follow-up report [5], which extended the results shown in [4] to titanium and steel alloys. In this Boeing report [5], additional comparisons between the Weibull and Log-Normal distributions was made for aluminum, titanium and steel alloys.

The report states, *"Because of the lack of large samples of data suitable for definition of the basic fatigue variability distribution …attention is focused on the possible use of many groups of data with only a very small number of details in each group"* [5]. The report made many comparisons between the two distributions using the above methodology. A typical comparison is shown in Figure 32, showing basic agreement between the test data and the Weibull distribution. Based on these comparisons, the report [5] concluded:

1. Both the Log-Normal and the Weibull distributions are capable of describing the fatigue data between 5% and 95% cumulative probability of failure.
2. Below 5% cumulative probability of failure, the Log-Normal model produces an *optimistic assessment* of the fatigue data distribution.
3. Below 5% cumulative probability of failure, the Weibull model produces an *acceptable assessment* of the fatigue data distribution.

It is not very clear what are the implications of, *"many groups of data with only a very small number of details in each group".* As a result, it is difficult to accept the above conclusions without further investigation.

Based on these studies [4] and [5], Boeing developed the methodology to determine fatigue life variability of safe-life structures using the Weibull distribution [7]. On the other hand, it was reported that "the Log-Normal distribution is used extensively by Airbus for fatigue life variability analyses" [28].

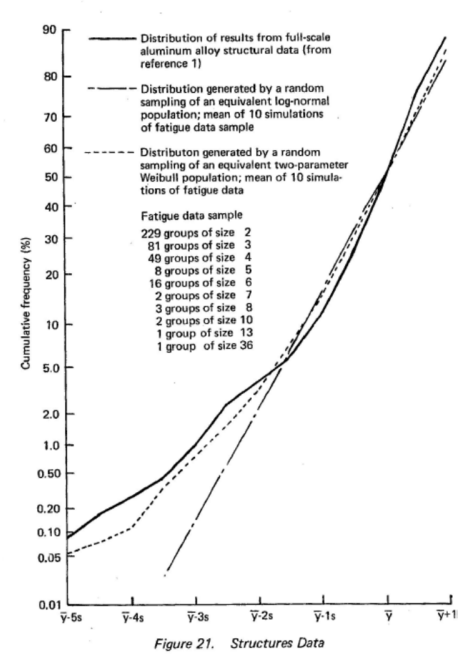

Figure 32. Comparison of the Boeing data *(as described in the figure)* to Log-Normal and Weibull distributions, produced by random sampling [5].

A Proposed Test Program to Better Evaluate the Fit of the Statistical Distributions and the Test Data
A sixty-specimen fatigue test program is proposed here in order to better evaluate the fit of the Weibull and Log-Normal distributions to the test data. It is recommended to use *simple* aluminum test specimens for this investigation.

1. The specimens are to be manufactured to *"typical aircraft standards"* for dimensions and surface-finish. The aim is that all sixty specimens be *carefully manufactured* in order to meet the selected standard. (If needed, crack-wires should be mounted to the specimens, in order to differentiate between crack-initiation and crack growth.)

2 All sixty specimens are to be tested under *well-controlled*, constant-amplitude loading. The lives at the start of cracking and at failure are to be recorded for each specimen, and the results are to be ranked in *ascending order* of their crack-initiation lives.

3 The above results are to be assigned "median rank" cumulative probability of failure values. *(Median rank methodology determines the best fit by least-square regression and is usually recommended for analysis of test results* [3].)

4 Using SuperSMITH software, the test results are to be fitted to both Weibull and Log-Normal distributions, as is demonstrated by Figures 33-36 for two sets of *synthetic* test specimens.

5 From the fatigue test results, it can be determined whether the test results fit a Weibull distribution, a Log-Normal distribution, or neither, as is demonstrated by Figures 33-36. It is likely that sixty specimens will be sufficient to differentiate between the two distributions.

6 If it is found that sixty specimens were not sufficient, an additional ten or more specimens should be prepared *in advance*, to extend the test to seventy or more specimens.

7 If the test results show that neither the Weibull nor Log-Normal distributions adequately describe adequately the results at the lower tail, a new statistical distribution needs to be developed, based on these test results.

8 The above test program can later be extended to titanium and steel alloys.

For test planning purposes, two *synthetic* sets of test results were prepared:

A. "Set-W", which contains sixty simulated failures, based on Weibull distribution characteristics.

B. "Set-LN", which also contains sixty simulated failures, based on Log-Normal distribution characteristics.

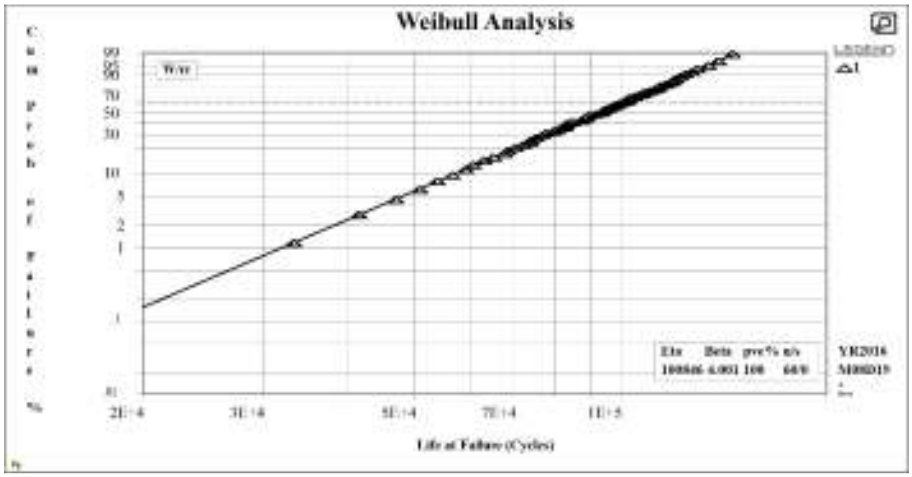

Figure 33. "Set-W" data analyzed for a Weibull distribution, showing *excellent correlation*

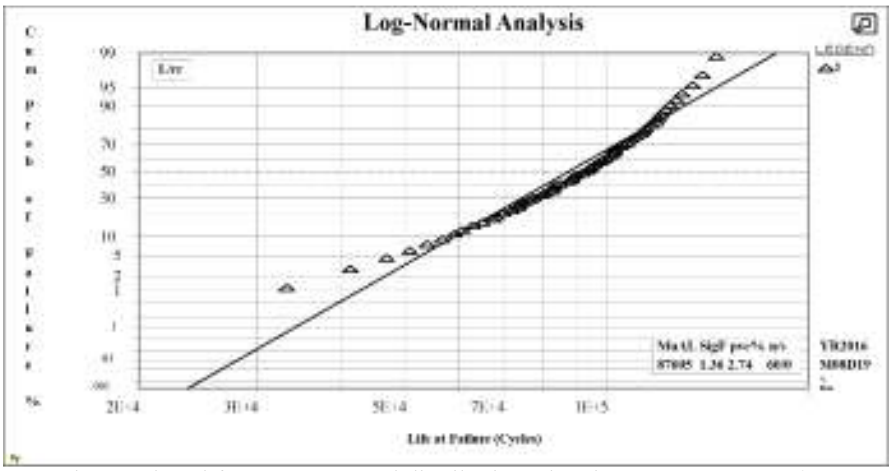

Figure 34. "Set-W" data analyzed for a Log-Normal distribution, showing *very poor correlation at the lower tail.*

Each of the above calculated results were analyzed using SuperSMITH software for both Weibull and Log-Normal distributions. Figure 33 and 34 show the "Set-W" data analyzed by both Weibull and Log-Normal distribution, respectively. Figure 35 and 36 show the "Set-LN" data analyzed by both Weibull and Log-Normal distribution, respectively.

From the results shown in Figures 33 to 36, it is clear that the SuperSMITH software is capable of distinguishing which distribution is more suitable for a specific data set.

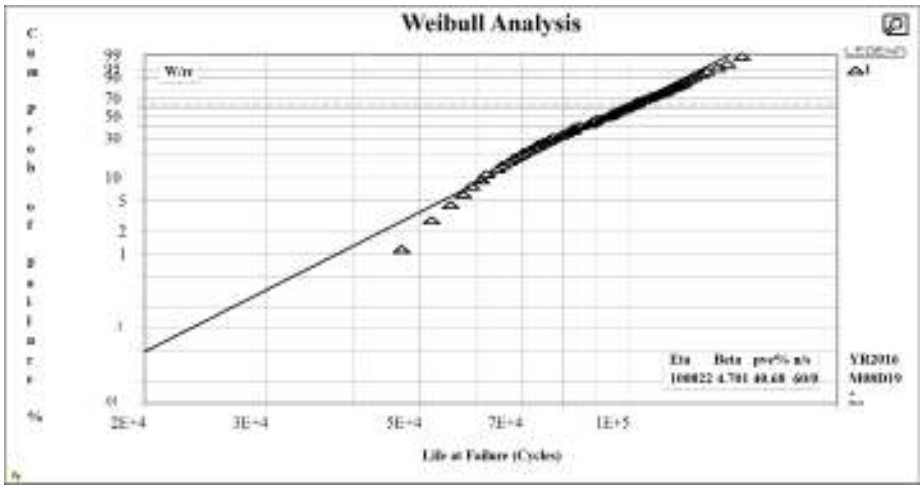

Figure 35. "Set-LN" data analyzed for a Weibull distribution, showing *poor correlation at the lower tail*

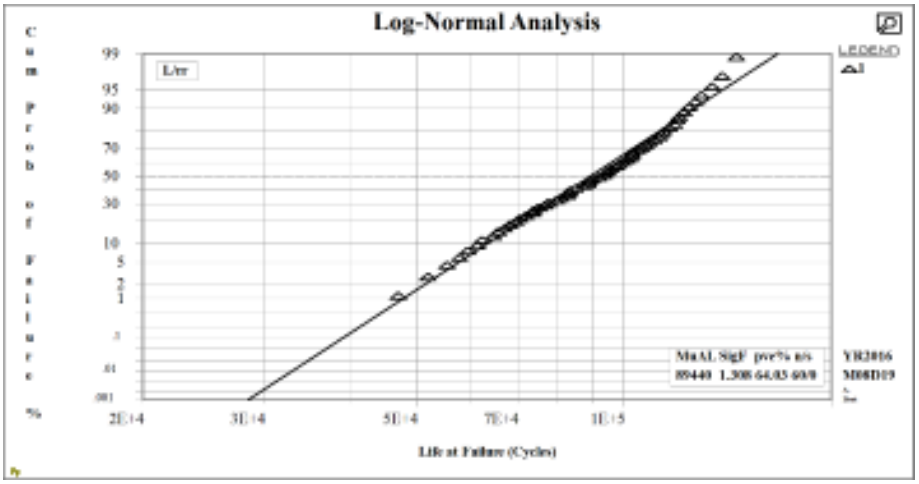

Figure 36. "Set-LN" data analyzed for a Log-Normal distribution, showing *good correlation, especially at the lower tail*

Figure 37 shows the two sets of test data. The data on the left is a *typical result* that would be produced by a Weibull distribution. The data on the right is a *typical result* that would be produced by a Log-Normal distribution. The calculated Weibull and Log-Normal distributions are also shown on Figure 37. It is clear that there is *excellent correlation* between the typical Weibull data and its distribution, as well as between the typical Log-Normal data and its distribution.

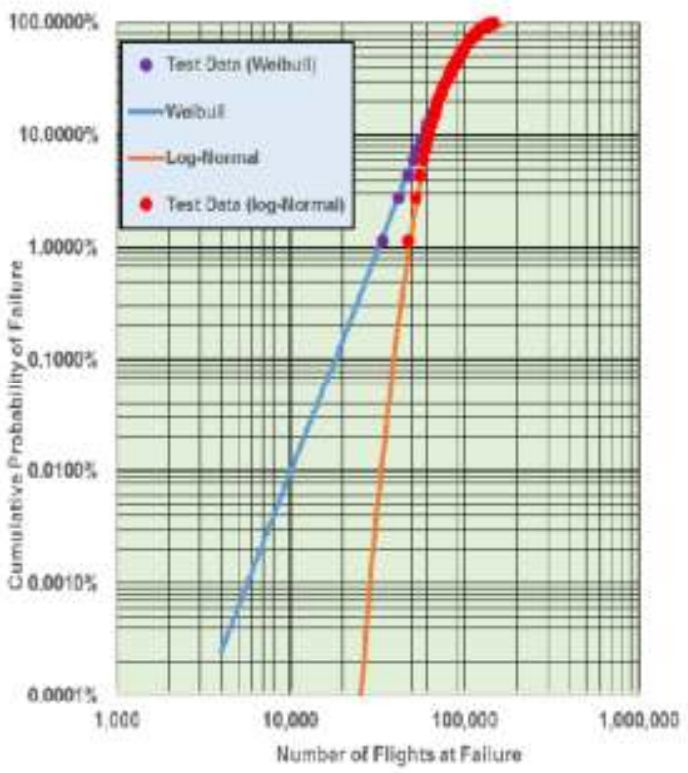

Figure 37. Simulated results for the proposed sixty specimen test.
(Test results are plotted against "median rank")

Summary and Conclusions

1. Over the last half-century, organizations that dealt with structural reliability related to fatigue life variability of aircraft structures, *generally* performed their reliability analyses by using the Weibull distribution. The Log-Normal distribution was considered to be more complicated, but *nearly identical* to the Weibull distribution, so it was less often utilized.

2. It has been shown in this paper, using several actual examples, that there are very sizable differences between the Weibull and Log-Normal distributions for setting allowable service lives to *minimize* the probability of failure. *In all cases, the Weibull results are much more conservative than the Log-Normal results for setting allowable service lives.*

3. From these results, it is difficult to say which distribution more accurately describes the fatigue life variation.

4. An analysis was performed to determine the allowable service life of a steel landing gear. The results clearly indicate that the Log-Normal distribution predicts a *much lower probability of failure* during its service life than the Weibull distribution.

5. A comparative analysis was performed by selecting a lower-confidence-level of 95% for a five-specimen fatigue test and calculating the allowable service life with 99.5% reliability and 95% confidence. The results again clearly indicate that the Weibull distribution produces far more conservative results than the Log-Normal distribution.

6. In 1969 and in 1972, the Boeing Company published the results of a study that was performed on fatigue life variability of aluminum, titanium and steel aircraft structures [4, 5]. Based on these studies, Boeing concluded that, below a 5% cumulative probability of failure, *the Log-Normal model produces an optimistic assessment of the fatigue data distribution, while the Weibull model produces an acceptable assessment of the fatigue data distribution.* On the other hand, it was reported that "the Log-Normal distribution is used extensively by Airbus for fatigue life variability analyses" [28].

7. The author of this paper found it difficult to accept Boeing's conclusions without further investigation.

8. A sixty-specimen fatigue test program was proposed by the author in order to better evaluate the fit of the Weibull and Log-Normal distributions *to actual fatigue test data.*

(All references are listed on the last two pages of this paper.)

FACE 3: IS THERE AN ACCEPTABLE RISK FOR WIDESPREAD FATIGUE DAMAGE?

Introduction

Since the accumulation of fatigue damage is a process that entails a large degree of scatter, it is not possible to determine an *absolute value* of the life to failure. Therefore, a small probability of failure is always present during the aircraft service life. The aircraft manufacturer needs to take every possible step to reduce this probability of failure to an "acceptable value".

There have been several aircraft failures attributed to multi-site damage (MSD) and widespread fatigue damage (WFD) over the last sixty years. *Multi-site damage (MSD) is a situation where small fatigue cracks develop at each of many repeating details, such as fastener holes. In time, these cracks may combine to form a large crack, as is shown in*

Figure 38. Widespread fatigue damage (WFD) is a situation where the extent of MSD damage may threaten the structural integrity of the aircraft. An example of an impending WFD failure is the Aloha Airlines extensive failure shown in Figure 39. *(One fatality occurred during the Aloha Airlines incident.)*

Among these aircraft failures are the two Comet disasters (1954), the Japan Airlines catastrophe of a Boeing 747 (1985), the Aloha Airlines failure of a Boeing 737 (1988) and the Southwest Airline incident, also occurring on a Boeing 737 (2011). These failures are described in detail in [9], and typical photos are shown as Figures 38 and 39.

The aim of this paper is to propose an ***acceptable risk criterion, and a method to calculate the risk,*** that will ensure a very high level of safety for a fleet of commercial aircraft, but will take into account that absolute safety is unattainable. Since the fatigue related accidents shown above were all related to the MSD and/or WFD type of failure, the analysis shown will be performed for a structure that is susceptible to the MSD and WFD failure mechanism.

A damage-tolerance analysis of a typical fuselage lap-splice joint will be presented in this paper. This type of structure is prone to *Multi-Site Damage* (MSD), which can lead to a *Widespread Fatigue Damage* (WFD) catastrophic failure. The infamous *"Aloha Airlines Disaster"*, which occurred in 1988, was the result of multi-site damage at a fuselage lap-splice joint of the Boeing 737 aircraft, as is shown in Figure 39.

A typical fuselage lap-splice joint will be analyzed, and inspection intervals will be determined using the methodology described in the paper. Figure 40 shows a typical three-row fuselage lap-splice joint where the outer side usually has countersunk fastener heads. The critical location for fatigue is at the countersunk fastener head at the first row.

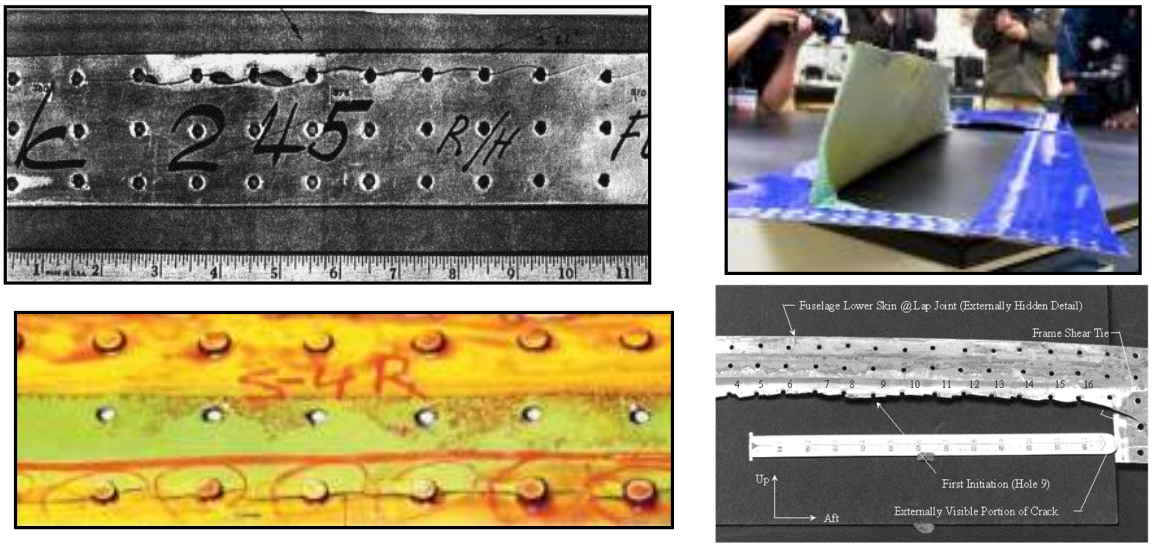

Figure 38. Typical MSD failures at fuselage lap-splice joints of Boeing 727 and 737 aircraft.

Figure 39. The Aloha Airlines disaster (Boeing 737) of 1988.

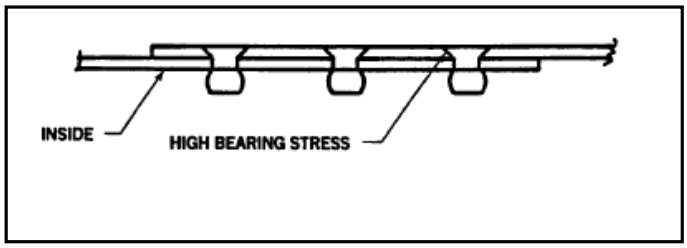

Figure 40. Typical fuselage lap-splice joint.

Description of INSIM Software

Several researchers have developed analysis methodology for widespread fatigue damage of a lap-splice joint using a deterministic analysis; an example is shown in [10]. My colleagues and I have presented several papers on WFD of a lap-joint that used probabilistic software called INSIM [8, 11-14]. It is our view that more realistic WFD analyses can be performed using the probabilistic approach.

The INSIM (INspection SIMulation) software has been developed in order to simulate the entire fatigue environment that a structure must withstand. INSIM simulates, in a probabilistic manner, service life variation, and service load severity, time to crack-initiation, crack growth history and NDI detection capability, as is described in [8].

There are three, mutually exclusive, outcomes of the fatigue process:

(1) The aircraft may reach the end of its operational life and be retired from service.
(2) A crack may be detected during scheduled maintenance operations. The affected part is usually repaired or replaced.
(3) A crack reaches its critical size undetected and the structure fails in service.

Crack-initiation, crack growth and crack detection characteristics of a specific structural location are input into INSIM, which performs the simulation of the fatigue process for the entire virtual fleet. Cracks initiate at various times and grow at variable rates in each aircraft. Inspections are performed according to a predetermined schedule. Cracks are detected during these inspections according to the statistical expectation of detection. As the simulation proceeds from aircraft to aircraft, cracks are detected, aircraft are retired from service or failures occur. In order to provide statistically significant results, a large number of simulations must be performed. Based on these simulations, INSIM calculates the probability that failure has occurred.

Figure 41 describes the "triple timeline" that is coordinated by INSIM. The first timeline represents the service life to retirement. Based on the options input into INSIM, the retirement life of each individual aircraft is assumed. However, no aircraft will be permitted to continue to fly after the Limit of Validity (LOV) is reached.

The second timeline represents inspections for cracks. Starting from the inspection threshold, the chosen NDI inspections are performed, according to the specified interval. The moment a crack is detected by INSIM, the aircraft is considered to have been grounded. *(Although some detected cracks can be repaired, and the aircraft may be re-introduced into service, this option was not included in INSIM.)*

The third timeline represents fatigue damage. The cracks initiate and grow, according to the probabilistic models that were selected. Fracture occurs when the crack reaches the critical size under spectrum loading.

If the first timeline reaches its destination before the others reach their destination, the aircraft is considered to have *retired*. If the second timeline reaches its destination before the others reach their destination, the aircraft is considered to be *grounded* because of the detection of a crack. If the third timeline reaches its destination before the others reach their destination, the aircraft is considered to have had a *catastrophic failure*. This process is repeated for all the aircraft in the *virtual fleet*, and the statistics are compiled. Figure 42 shows a typical output summary from an INSIM analysis consisting of *four million simulations.*

A description of the various probabilistic models used in INSIM are discussed in detail in [8].

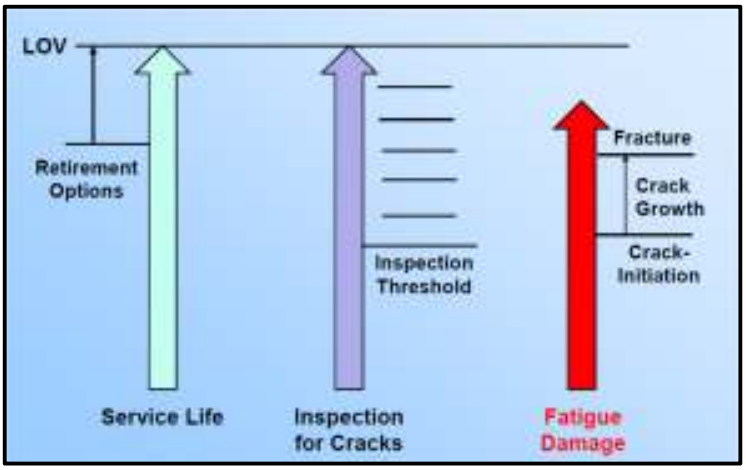

Figure 41. Probabilistic timelines used in INSIM.

```
                    SUMMARY & STATISTICS

Fleet size:  4000000   Threshold:   12000   Interval(s):  3400

Upper Limit for NDI Detection Probability:  0.980
Mean service life:       75000              High-time aircraft:       75000
Min. crack initiation:    5499              Max. crack initiation:   778276
Min. crack growth:        9681              Max. crack growth:        31017

Inspections performed:    1.677722E+07      Inspections per aircraft:  4.19

                                  Number of Aircraft       Percent

Aircraft retired uncracked:            3969597              99.24%

Aircraft retired cracked:                14887               0.37%

Cracks < Ap detected:                    15455               0.39%

Cracks > Ap detected:                       16               0.0004%

Failures:                                   45               0.00112%

Indicate number of additional aircraft to inspect (RETURN for zero):
```

Figure 42. A typical INSIM output summary for four million simulations.

Fatigue and Damage-Tolerance Analysis of a Lap-Splice Joint

The example chosen is a typical lap-splice joint for the fuselage of a narrow-body jet airliner. *(A narrow-body airliner was selected since it usually accumulates many more flight cycles in its service life than a wide-body aircraft.)* The splice material was taken as 2024-T3 aluminum sheet having a material thickness of 0.050 inches. Aluminum rivets of 0.1875-inch diameter with countersunk heads were taken, installed in three rows with a pitch of 1.0 inch between fasteners. A *nominal* pressure differential of 8.6 psi at cruise altitude was assumed. The fuselage radius was taken as 74 inches, as being typical for narrow-body jetliners. The aircraft was assumed to have a design service goal (DSG) of 60,000 flights. The narrow-body jet airliner was assumed to have a typical flight length of 1.5 hours. The details of calculating the loads, stresses, and the stress-concentration factor are described in [8].

A crack-initiation analysis was performed using S-N curves provided in MMPDS-7 [19] for 2024-T3 alloy, resulting in a mean life to crack-initiation of 265,400 flights. This result accounts for the "size effect" factor of two, between the actual splice and the test specimens used to generate the S-N diagram, as was recommended by Safarian [10].

A crack growth analysis was performed using NASGRO ver. 8.0 software. The analysis assumed a 0.010-inch crack at *each side of each fastener hole*. The analysis showed that the crack will fail under spectrum loading after 17,400 flights with a crack size of 0.283 inches. Is should be noted that the failure was due to "net-section yield" which is a typical mode for MSD failures.

Figure 43 shows the calculated crack growth curve. Periodic inspections are used to detect cracks during the service life of each aircraft, using high-frequency eddy-current (HFEC) methodology. It should be noted, that approximately 0.041 inches of the crack will be blocked by the countersunk fastener head. Therefore, the effective crack growth curve will be used, which accounts for the partial blockage of the HFEC probe, as is shown in Figure 43.

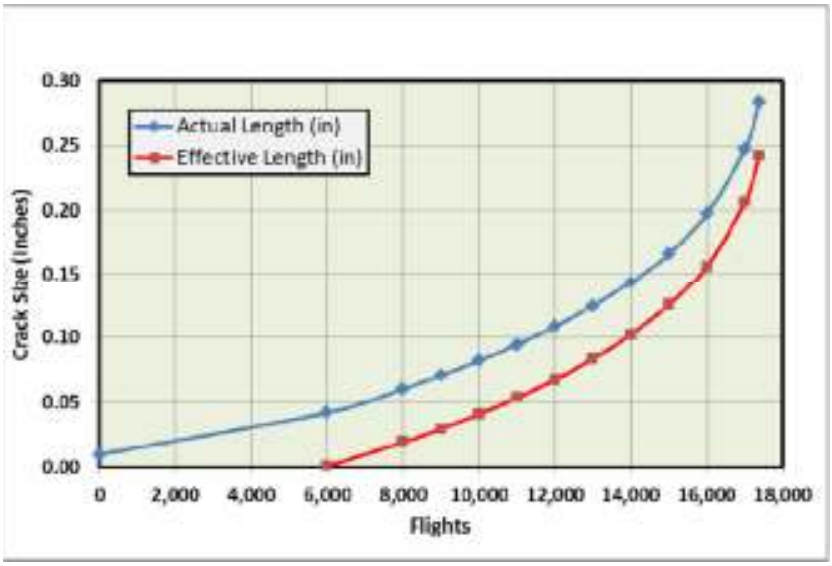

Figure 43. Calculated actual and effective crack growth at the lap-joint.

The Effect of Multiple Cracks on Crack Detection

The MSD problem has been studied by many researchers; recently by A. Skorupa and M. Skorupa in 2012 [20]. It has been shown, by analysis and test, that the hoop-stress in an aircraft fuselage is not constant along the length of the fuselage, but varies significantly [20]. The highest hoop-stresses can be found between the circumferential frames, while the lowest hoop-stresses are found at the frames. Therefore, it is very likely that the first cracks will develop between the frames, and much later adjacent to the frames. In support of this assumption, an inspection of a riveted lap-joint on an aging Boeing 727 aircraft disclosed that the longest cracks were indeed detected midway between frames while no cracks were found at the frame locations [20].

Table 1 shows the three-parameter Weibull parameters for HFEC type NDI for the detection of *at least one crack* [11]. The characteristic size (λ) is the crack size having a 63.2% probability of detection (POD).

The design of our narrow-body airliner has a fuselage frame spacing of 20 inches. With a pitch distance of one inch between fasteners, there are 20 lap-splice fasteners per bay. Since each fastener hole can have two cracks, there is a potential for 40 cracks in each bay. In order to be realistic, but somewhat conservative, it will be assumed that, from the standpoint of *crack detection*, six cracks exist, and the value of λ (characteristic size) is 0.0509 inches, as was shown in Table I. This value will be used for all crack detection probability calculations. Under this assumption, a crack size of 0.048 inch will be detected 50% of the time; a crack size of 0.060 inch will be detected 90% of the time; a crack size of 0.069 inch will be detected 98% of the time.

Table I. Characteristic crack sizes (POD = 63.2%) for detecting *at least one crack* for multiple cracks using HFEC methodology [17]. *(The additional Weibull parameters are: $\alpha = 1.78$ and $a_0 = 0.035$ inch)*

Number of Cracks	λ - Characteristic Size (in)
1	0.0780
2	0.0641
3	0.0582
5	0.0524
6	**0.0509**
10	0.0468

Determining Inspection Intervals and Threshold

Two criteria were used to estimate the required inspection interval when using the High-Frequency Eddy-Current (HFEC) inspection method:

(1) From previous experience [8, 11,13], it was found that a Cumulative Probability of Failure (CPOF) of 0.001% will yield an *acceptable risk* in operation of this type of aircraft. (See Figure 45.)

(2) Cumulative Probability of Detection (CPOD) Method: CPOD is the probability that a crack will be detected in *any one* of the multiple inspections that will be performed throughout the operating lifetime of the aircraft. A CPOD of 99.9% is generally used for a pressurized fuselage. Boeing uses the "DTR Method", and specifies DTR=10 for a critical cabin pressurization application. A DTR of 10 is equivalent to a CPOD of 99.9% [8].

INSIM runs showed that the first criterion can be met with an inspection interval of 3,400 flights when using HFEC methodology, and employing the detection parameters established by FAA sponsored testing [17].

The inspection interval for the "CPOD Method" was determined using INSPOD ver. 4 software. This software calculates CPOD using the following relationship: CPOD = 1 – Q1 * Q2 * Q3*… Qn, where Q is the probability of not detecting the crack in any one of the "n" inspections. Due to the "explosive potential" of an MSD failure of the fuselage lap-splice joint (See Figure 39.), a cumulative probability of detection of 99.9% was taken. INSPOD runs demonstrated that the inspection interval ***should also be set at 3,400 flights*** for maintaining a CPOD of 99.9%.

INSIM runs have shown that selecting an *inspection threshold* in the range of 10,000 flights to 15,000 flights has only a minor effect on the CPOF. Therefore, an alternative criterion, that the inspection threshold should not exceed 70% of the crack growth life, was used. On this basis, the inspection threshold was set as 12,000 flights.

Probabilistic Considerations

As was stated previously, several probabilistic models are used in the INSIM software.

Figure 44 plots the two-parameter Weibull distribution which was selected to represent crack-initiation at the lap-joint. *(The first parameter is η which is called the "shape-factor", and was given a value of four, which is*

characteristic of aluminum structures [7]. The second parameter is β, which is called the "characteristic life", was given the calculated typical fatigue life of 265,400 flights for the lap-joint, as was previously described.)

Figure 44 shows that a very wide range of fatigue lives are possible (5,000 to 500,000 flights) when going from a probability of failure of 0.00001% to 99.99%. These results were obtained by using SuperSMITH Software [2], where the special scales of the axes of Figure 44 are set to produce a straight line when dealing with a two-parameter Weibull distribution.

The Log-Normal model can also be used to describe crack-initiation distributions. While both the *two-parameter Weibull distribution* and the *Log-Normal distribution* can adequately model observed fatigue test results, their treatment of the upper and lower "statistical tails" are quite different, as was described in "Face 2" of this paper.

Typical crack growth scatter was taken as a *maximum of ±20% in life*, as per data presented by Schijve [1]. Spectrum variation was taken as a *maximum of ±12% in stresses* to account for variations in cabin differential pressure at the highest altitude reached during each flight.

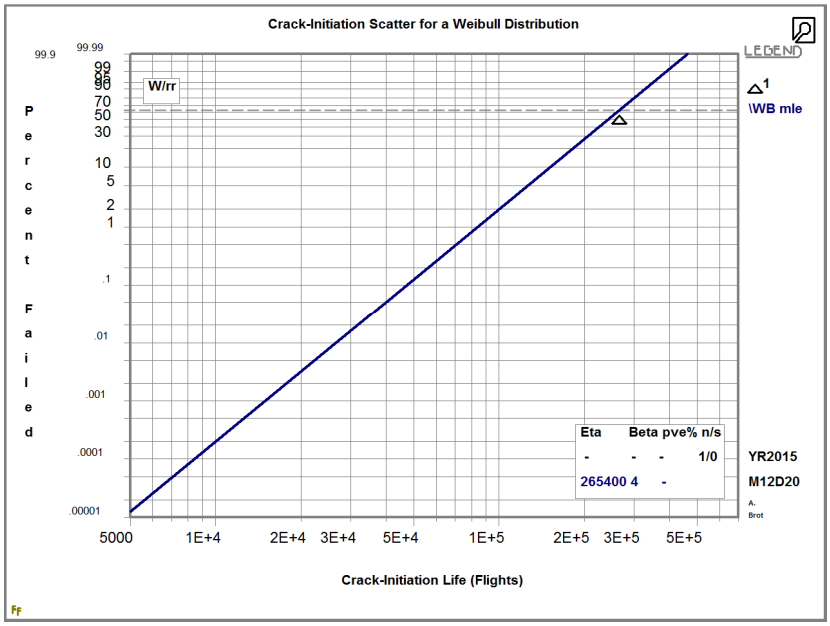

Figure 44: Weibull plot for crack-initiation at the lap-joint

Typical INSIM runs were performed using two million simulations each, which can be performed in only a few minutes. Three results will be shown.

Table II assumes that *all the aircraft retire at 90,000 flights*, simulating a lack of an *effective* limit-of-validity (LOV). Table II shows that the cumulative probability of aircraft failure is 0.0033%, which may be considered to be excessive.

Table III takes the results of Table II, but imposes a *tentative* LOV of 75,000 flights. This reduces the probability of failure from 0.0033% to 0.0009%, which is a reduction factor of 3.67. Since only 0.29% of the aircraft have had cracks detected by the HFEC NDI during the eighteen inspections that are scheduled from 12,000 flights to 75,000 flights, it is tempting to say that, perhaps, no inspections need to be implemented.

However, if the inspection program *was not implemented*, aircraft failures would rise to 0.21%, *which is totally unacceptable*. This is shown in Table IV.

Aircraft failure, as is defined in Tables II-IV, assumes the failure of *one fuselage bay*. This is a very severe event since it will result in rapid decompression of the fuselage and *may* lead to a widespread fatigue damage (WFD) scenario.

Table II. Basic results from the INSIM run *(2,000,000 program simulations with an inspection interval of 3,400 flights). All aircraft are assumed to retire at 90,000 flights.*

Expected Outcome	**Probability of Occurrence**
Aircraft retires *(cracked or not cracked)*	99.15%
Cracks are detected by HFEC inspections and the aircraft is removed from service.	0.85%
Aircraft failure occurs	0.0033%

Table III. Basic results from the INSIM run *(2,000,000 program simulations with an inspection interval of 3,400 flights). All aircraft are assumed to retire no later than 75,000 flights.*

Expected Outcome	**Probability of Occurrence**
Aircraft retires *(cracked or not cracked)*	99.71%
Cracks are detected by HFEC inspections and the aircraft is removed from service.	0.29%
Aircraft failure occurs	0.0009%

Table IV. Basic results from the INSIM run *(2,000,000 program simulations with **no inspections**). All aircraft are assumed to retire no later than 75,000 flights.*

Expected Outcome	**Probability of Occurrence**
Aircraft retires *(cracked or not cracked)*	99.79%
Cracks are detected by HFEC inspections and the aircraft is removed from service.	0 %
Aircraft failure occurs	0.21%

Acceptable Values of CPOF and Hazard-Rate per USAF and FAA Criteria

INSIM was run with a variety of inspection intervals, all employing high-frequency eddy-current (HFEC) crack detection methodology. In all cases, the inspection threshold was kept at 12,000 flights. Figure 45 shows the results of this analysis. If we take 0.001% *as a tentative acceptable value of CPOF*, then an inspection interval of approximately *3,400 flights is appropriate,* as was previously described.

Since the CPOF parameter deals with the *overall probability of failure*, it does not have the ability to pinpoint specific potentially dangerous situations that occur as an aircraft *approaches its limit-of-validity,* just before it is required to

be retired from service. Therefore, the hazard-rate is considered to be a more appropriate parameter. *(The hazard-rate is defined as the probability of failure occurring during the next flight.)*

Figure 46 shows the values of the hazard-rate as a function of the aircraft service life, as was determined by INSIM. The hazard-rate was *closely approximated* by using the following procedure:

1. INSIM was run for a specific case, using a large number of program simulations *(four million for our example)*.
2. All failures were identified, together with the *number of flights* at which they occurred.
3. The range of flights where failures occurred *(40,900 to 80,000 flights for our example)* were broken down into bands, each of which has the width of a portion of the inspection interval.
4. Each failure is assigned to the band that corresponds to its time of failure.
5. The hazard-rate for each band was *closely approximated* by dividing the number of failures in its band by the number of program simulations performed and by the number of flights represented by the band-width.

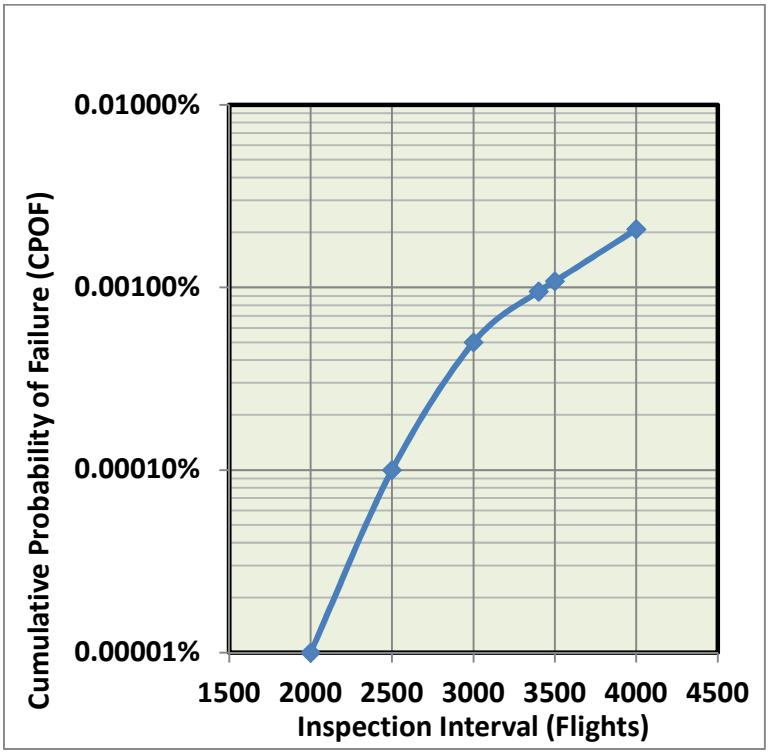

Figure 45. Cumulative probability of failure (CPOF) as a function of the inspection interval.

The resulting data was plotted in Figure 46, with the peaks representing the hazard-rate at the time just before the inspection, while the valleys representing the hazard-rate at the time between inspections.

Figure 46 clearly shows the increase in hazard-rate as the aircraft service life is increased. It also demonstrates the peaks and valleys of the hazard-rate parameter, with a peak occurring *just before* an inspection is scheduled to take place, and a valley occurring *after* the inspection, when all aircraft with detected cracks have been removed from service.

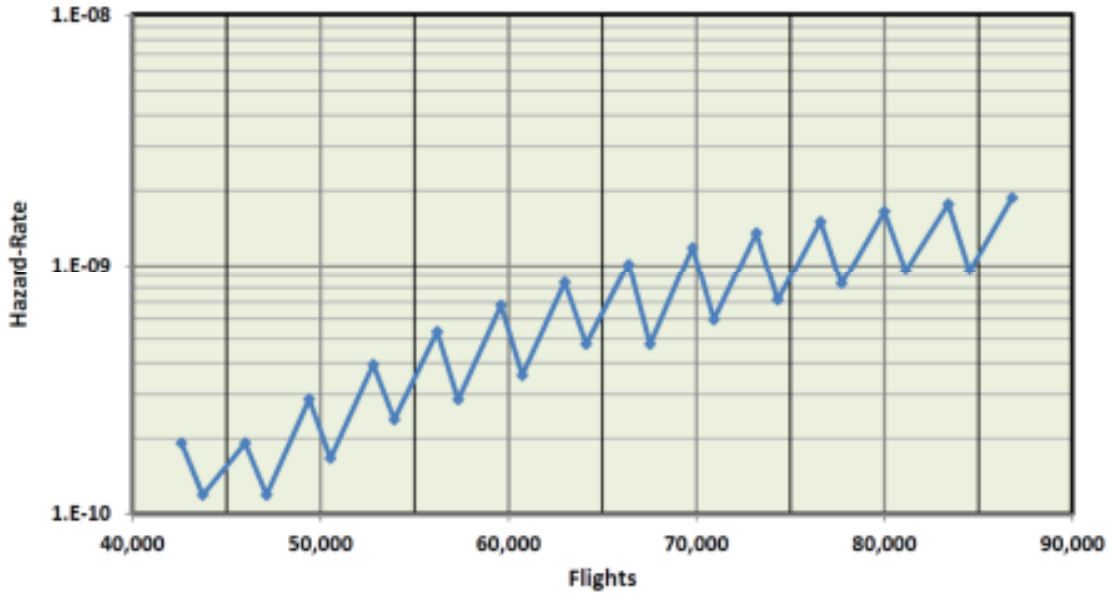

Figure 46. Values of the calculated hazard-rate as a function of service life for an inspection interval of 3,400 flights. *(Hazard-rate is defined as the probability of failure occurring during the next flight.)*

Figure 47 defines the USAF criterion for acceptable hazard-rate, as was described by Gallagher [15]. Any hazard-rate less than 10^{-7} failures per flight is defined as acceptable *(green area of Figure 47).* Any hazard-rate above 10^{-5} failures per flight is defined as unacceptable *(red area of Figure 47).* Between these limits *(yellow area of Figure 47),* the hazard-rate may be acceptable if the "exposure" is limited by "mitigating risks through inspections, repair or applying operational restrictions" [15, 16]. Since our type of structure may go for thousands of flights between inspections, the hazard-rate should not exceed 10^{-7} failures per flight, according to the USAF standards. *(See Figure 48.)*

It should be noted that the data shown in Figure 47 is supported by MIL-STD-1530C *(Aircraft Structural Integrity Program)* which is the standard used by the USAF [16].

Figure 48 shows a typical output of "PROF Software for Aging Aircraft Risk Analysis" [22], which was developed by the University of Dayton Research Institute (UDRI) to the USAF standards for risk analysis. Figure 48 clearly shows the sequence of hazard-rate peaks *(at inspection times)* show larger and larger hazard-rates, very similar in concept to what is shown in Figure 46.

A more difficult and important question is, what is an acceptable value of the hazard-rate according to FAA criteria? The FAA Federal Aviation Regulations Part 25.571 [23] state, "… must show that catastrophic failure due to fatigue, corrosion, manufacturing defects, or accidental damage, *will be* ***avoided*** *throughout the operational life of the airplane."*

The advisory circular (AC) for this paragraph in the FAA regulations [24] states in paragraph 6b, "a probabilistic analysis may be used if it can be shown that … loss of the aircraft is *extremely improbable*…"

Figure 47. USAF criterion for acceptable hazard-rate [15, 16, 22].

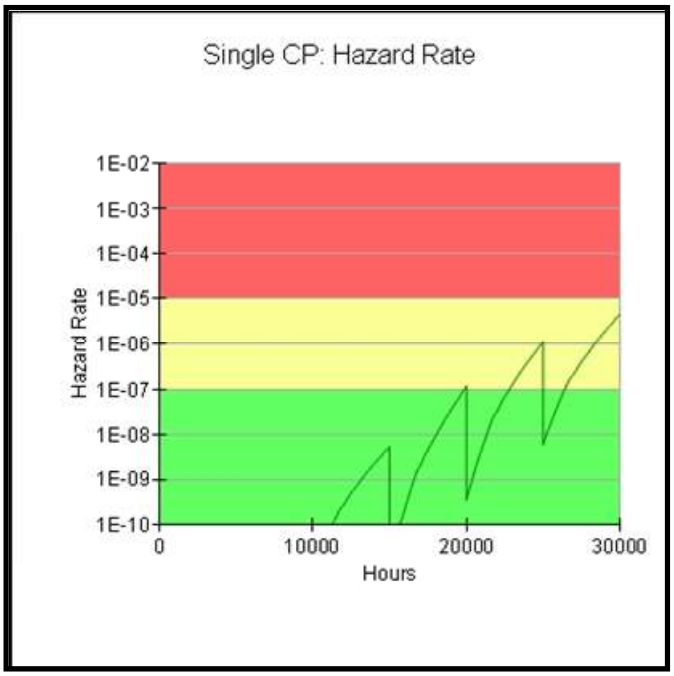

Figure 48. An example of USAF acceptable hazard-rates as calculated by PROF software [22].

The FAA Systems Safety Handbook [25] and FAA AC-25-19 Certification Maintenance Requirements [26] both define "extremely improbable" as "the probability of failure per *operational hour* is less than 10^{-9} ". Azevedo [28] also concurs with this value.

It can be seen that the *actual* FAA regulations [23] only require that: "catastrophic failure … will be avoided throughout the operational life of the airplane" while the FAA advisory documents [24-26, 28] recommend that "the probability of failure per *operational hour* be less than 10^{-9} ".

Using the *recommendations* stated above [24-26, 28], and accounting for the fact that the representative aircraft selected in this study has a typical flight length of 1.5 hours, the acceptable hazard rate for commercial operation is 1.5×10^{-9} failures per flight. Figure 46 shows that at the last inspection before 75,000 flights is reached, the hazard rate is equal to 1.24×10^{-9} failures per flight. This reinforces the selection of 75,000 flights as a suitable LOV of this aircraft, as was stated previously. This means that the selected value of LOV complies with the *severe interpretation* of acceptable hazard-rate, as is defined by the FAA regulations [24 - 26].

Assuming a usage rate of 2,500 flights per year, an LOV of 75,000 flights will allow the aircraft to be in service for 30 years. The inspection threshold of 12,000 flights will allow the aircraft to fly nearly 5 years before any crack detection inspections are needed. The inspection interval of 3,400 flights will allow approximately 16 months to pass between crack detection inspections.

The Role of the Full-Scale Fatigue Test (FSFT) in Confirming the Adequacy of the Design to Withstand Widespread Fatigue Damage

Paragraph 25.571(b) of the Federal Aircraft Regulations states: *"It must be demonstrated with sufficient full-scale fatigue test evidence that widespread fatigue damage will not occur within the design service goal of the airplane"* [23]. This requirement is usually implemented by extending the FSFT to *more than two design lifetimes* and then performing a teardown inspection of the critical areas.

The Boeing 787, which has a reported design service goal (DSG) of 44,000 flights and an LOV of 66,000 flights, has recently completed its FSFT which included 165,000 simulated flights, which represents 2.5 times the LOV and 3.75 times the DSG.

Our representative aircraft, which has an LOV of 75,000 flights, had an INSIM simulation performed for a FSFT of a duration of 187,500 flights, which represents 2.5 times the LOV and about 3.1 times the DSG of the aircraft, which was set as 60,000 flights.

The results of the INSIM simulations are shown in Table V. The results are based on the three mutually exclusive outcomes shown in the table, and accounts for *both lap-splices* (left and right) that will be present on the fatigue test-article.

The results indicate that a premature failure of the FSFT is quite unlikely (0.15%) but it is possible that cracks will be detected on either of the two lap-splices (35.18%) during the FSFT. These cracks will may to be repaired in order to allow the test to safely continue to its goal of 187,500 test flights. It is also likely that additional cracks will be discovered during teardown inspection following the conclusion of the FSFT.

If the probability of a premature lap-splice failure (0.15%) is considered to be too risky, the inspection interval can be reduced during the latter stages of the FSFT.

Table V. Results from an INSIM run *(2,000,000 simulations with an inspection interval of 3,400 flights)* for a full-scale fatigue test performed for 187,500 flights (2.5 x LOV)

Expected Outcome	**Probability of Occurrence**
Test aircraft reaches the end of the FSFT without any cracks detected on *either of the two lap-splices.*	64.67%
Test aircraft reaches the end of the FSFT with cracks detected on *either of the two lap-splices.*	35.18%
A premature lap-splice failure occurs on *either of the two lap-splices.*	0.15%

An Alternate Treatment of Widespread Fatigue Damage

In October, 2011, The FAA released AC No. 120-104, which is an advisory circular dealing with "Establishing and Implementing Limits-of-Validity to Prevent Widespread Fatigue Damage" [27].

Figure 49 describes schematically the proposed method to deal with widespread fatigue damage (WFD), as is defined by AC No. 120-104. WFD $_{\text{(average behavior)}}$ is defined as the service life at which 50% of the fleet would have failed due to WFD, in the absence of inspections.

The SMP is the structural modification point, where the structure needs to be modified to avoid failure. It is defined in the advisory circular [27] as 50% x WFD $_{\text{(average behavior)}}$. In the event that the manufacturer elects not to modify the structure, the Limit-of-Validity (LOV) will be fixed as 50% x WFD $_{\text{(average behavior)}}$.

The ISP is the point where a special inspection schedule is introduced. It is defined in the advisory circular as 33% x WFD $_{\text{(average behavior)}}$ [27].

Using the above criteria, and taking the crack-initiation life of our representative narrow-body aircraft as 265,400 flights and the typical crack growth life of 17,400 flights, as was described previously, we are able to calculate the SMP and ISP, as was described above.

The value of WFD $_{\text{(average behavior)}}$ can be *estimated* by running INSIM without inspections, while varying the required service life. The result of these INSIM runs show that a 50% Cumulative Probability of Failure (CPOF) will occur for a service life of 257,700 flights. Therefore WFD $_{\text{(average behavior)}}$ is taken as 257,700 flights. As a result, the SMP will be taken as 129,000 flights, and the ISP as 85,000 flights. *(Both these values have been rounded-off, for convenience.)*

It will be assumed that no structural modifications will be performed on our representative aircraft, so the LOV of this aircraft will be taken as 129,000 flights.

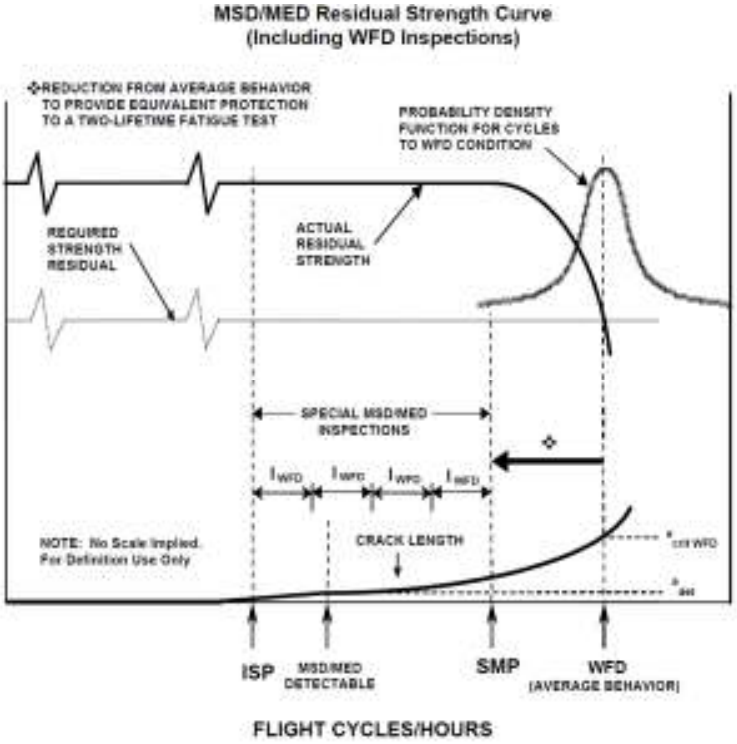

Figure 49. Schematic representation of managing widespread fatigue damage per AC No. 120-104 [27].

The author of this paper has several reservations about using this procedure to manage widespread fatigue damage:

(1) Since WFD $_{(\text{average behavior})}$ is primarily based on the crack-initiation life of the structure, this is basically a return to the "safe-life concept", which has been shown many years' ago to be an *unreliable method to ensure safety of fatigue critical structures.*

(2) Setting the LOV at 50% of the "safe life" appears to be inadequate, even though periodic inspections will take place. Scatter-factors of *five* or more are normally used to provide sufficient safety of safe-life structures. A scatter-factor of approximately *three* seems to be more appropriate for this application.

(3) This procedure does not specify how to determine the inspection interval for the "special inspections" that begin at the ISP point (85,000 flights), as is shown in Figure 49.

(4) The proposed method appears to be more like a "template" than a well-defined procedure.

In order to evaluate this procedure, certain assumptions need to be made. Safarian, who performed such an analysis in 2013 [10], selected an inspection interval of 1/3 of the number of flights between the 90% POD crack size to the critical crack size, to be applied before reaching the ISP and 1/4 of the number of flights between the 90% POD crack size to the critical crack size, to be applied after reaching the ISP and up to the LOV. The 90% crack-size for HFEC was taken as 0.105 inches [17], which, according to Figure 43 occurs at about14,100 flights *(using the effective curve)* while the critical crack size occurs at 17,000 flights. This means that the "inspection window" will be about 2,900 flights. According to the Safarian method [10], the inspection interval will be 970 flights up to 85,000 flights and 725 flights between 85,000 flights and 129,000 flights.

Assuming a usage rate of 2,500 flights per year, an LOV of 129,000 flights will allow the aircraft to be in service for *nearly 52 years*. The inspection intervals of 970 and 725 flights will allow approximately *4.5 months* to pass between crack detection inspections before reaching 85,000 flights, and about *3.5 months to pass* between crack detection inspections between 85,000 flights and 129,000 flights. The aircraft will probably be safe from a WFD standpoint, but the inspection schedule does not appear to be very practical.

A set of INSIM runs were performed to try to find a *more practical solution* to the method described in Figure 49. It was decided to use a *single inspection* interval of 2,700 flights from 12,000 to 129,000 flights, since no real benefit was found by using two different intervals.

The resulting data was plotted in Figure 50, with the peaks representing the hazard-rate at the time just before the inspection, while the valleys representing the hazard-rate at the time between inspections.

At the last inspection before 129,000 flights *(128,100 flights)* the hazard-rate reached 1.18×10^{-9} failures per flight hour, which is *slightly below* the allowable hazard-rate criterion that was previously selected.

Assuming a usage rate of 2,500 flights per year, An LOV of 129,000 flights will allow the aircraft to be in service for *nearly 52 years*. The inspection threshold of 12,000 flights will allow the aircraft to fly nearly *5 years* before any crack detection inspections are needed. The inspection interval of 2,700 flights will allow approximately 13 months to pass between crack detection inspections. These values appear much more practical than those that were calculated using AC No. 120-104 and the assumptions that were made.

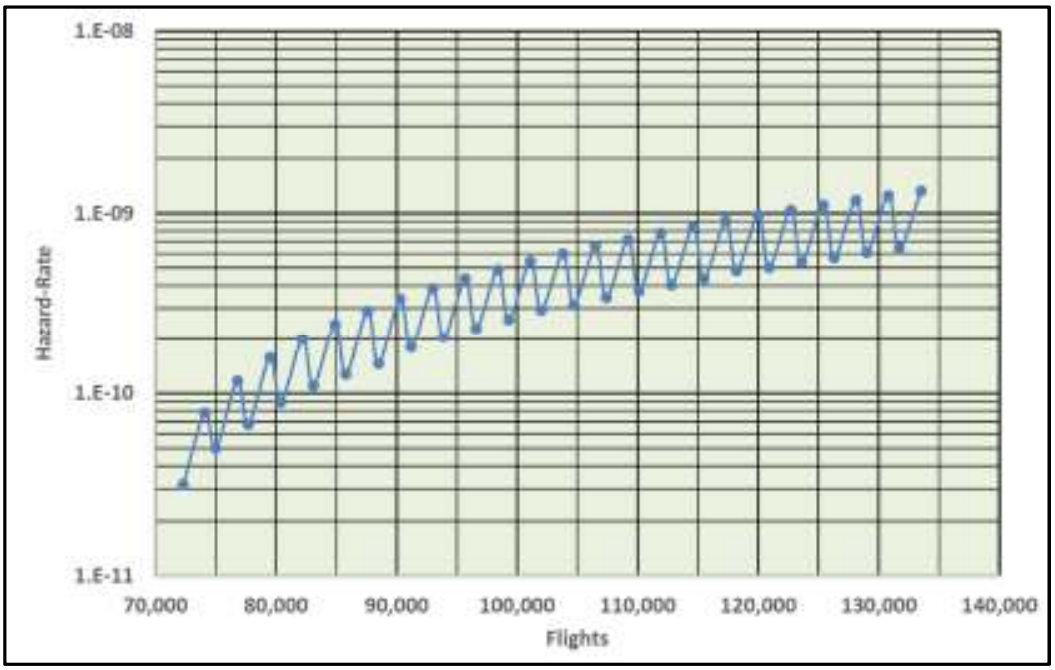

Figure 50. Values of the calculated hazard-rate as a function of service life for an LOV of 129,000 flights and an inspection interval of 2,700 flights. *(Hazard-rate is defined as the probability of failure occurring during the next flight.)*

Summary and Conclusions

1. IMSIM Software was described and was applied to a representative narrow-body airliner, whose properties were selected for this study, in order to determine its expected hazard-rate for a widespread fatigue damage (WFD) failure. Based on several INSIM runs, it was found that an inspection threshold of 12,000 flights and an inspection interval of 3,400 flights, and using HFEC methodology, would give acceptable results.

2. Allowable hazard-rates for WFD, per USAF and FAA standards were investigated and were presented. A method was developed to apply INSIM software to calculate the peak hazard-rate as a function of elapsed flights. The results showed that the peak hazard-rate will not exceed 1.5×10^{-9} failures per flight if the Limit-of-Validity (LOV) of the aircraft does not exceed 75,000 flights, and the inspection threshold and interval do not exceed 12,000 flights and 3,400 flights respectively.

3. An alternate procedure, based on FAA issued AC No. 120-104, was applied to the representative narrow-body airliner properties that was developed for this study. The results showed that the LOV can be extended to 129,000 flights but the required inspection intervals will range from 725 flights to 970 flights. It was concluded that this procedure did not provide practical results for this application.

4. The INSIM based method was then applied to the above task for providing an LOV of 129,000 flights. The results showed that, with an inspection threshold of 12,000 flights and with an inspection interval of 2,700 flights, a much more practical result can be obtained without the peak hazard-rate exceeding 1.5×10^{-9} failures per flight, per FAA regulations.

REFERENCES

[1] J. Schijve, (2009) *Fatigue of Structures and Materials*, 2nd Edition, Springer.

[2] SuperSMITH Software (2016), release 5.0-CY, developed by Fulton Findings.

[3] R. B. Abernethy (2009), *The New Weibull Handbook*, 5th edition, Abernethy.

[4] I. C. Whittaker and P. M. Besuner (1969), *A Reliability Analysis Approach to Fatigue Life Variability of Aircraft Structures*, The Boeing Company, (unlimited distribution).

[5] I. C. Whittaker (1972), "Development of Titanium and Steel Fatigue Variability Model for Application of Reliability Analysis Approach to Aircraft Structures", The Boeing Company, (unlimited distribution).

[6] J. Schijve and F. A. Jacobs, (1955), *Fatigue Tests on Notched and Unnotched Clad 24 S-T Sheet Specimens to Verify the Cumulative Damage Hypothesis*, NLR Report M-982.

[7] U. G. Goranson, (1997), *Fatigue Issues in Aircraft Maintenance and Repairs*, International Journal of Fatigue, Vol. 20, No. 6.

[8] A. Brot, (2016), *Selection of an Acceptable Risk Criterion for Avoiding a Major Fatigue Failure on a Commercial Aircraft*, Proceedings of the 56th Israel Annual Conference on Aerospace Sciences.

[9] A. Brot, (2012), *Developing Strategies to Combat Threats against the Structural Integrity of Aircraft,* Proceedings of the 52nd Israel Annual Conference on Aerospace Sciences.

[10] P. Safarian, (2013), *Widespread Fatigue Damage Evaluation Method*, Proceedings of the 27th ICAF Symposium, Jerusalem.

[11] A. Brot, (2013), *Optimizing the Inspection Intervals for Aircraft Structures Prone to Multi-Site Damage,* Proceedings of the 53rd Israel Annual Conference on Aerospace Sciences.

[12] D. Elmalich, A. Brot, Y. Freed, et al (2015), *A Risk Analysis Based Methodology for Widespread Fatigue Damage Assessment*, Proceedings of the 55th Israel Annual Conference on Aerospace Sciences.

[13] Y. Freed, D. Elmalich, A. Brot, et al (2015), *A Risk Analysis Based Methodology for Widespread Fatigue Damage Evaluation,* Proceedings of the 28th ICAF Symposium, Helsinki.

[14] A. Brot (2010), *Optimization Strategies for Minimizing Fatigue Failures in Metallic Structures*, Presented to the 31st Israel Conference on Mechanical Engineering.

[15] J. Gallagher, (2007), *A Review of the Philosophies, Processes, Methods and Approaches that Protect In-Service Aircraft from the Scourge of Fatigue Failures,* 21st Plantema Memorial Lecture, Proceedings of the 24th ICAF Symposium, Naples.

[16] Mil-STD-1530 C, (2005), *Airplane Structural Integrity Program (ASIP).*

[17] F. W. Spencer, (1994), *Eddy-Current Inspection Reliability at Airline Inspection Facilities*, Proceedings of the 1994 USAF Aircraft Structural Integrity Program Conference, San Antonio, TX, USA.

[18] M. Niu, (1990), *Aircraft Structural Design*, Conmilit Press Ltd.

[19] *Metallic Material Properties Development and Standardization (MMPDS) Handbook*, (2012), Issue 7, Federal Aviation Administration.

[20] A. Skorupa, and M. Skorupa, (2012), *Riveted Lap Joints – Design, Analysis and Properties*, Springer.

[21] U. G. Goranson, (1993), *Damage Tolerance, Facts and Fiction,* 14th Plantema Memorial Lecture, Proceedings of the 17th ICAF Symposium, Stockholm.

[22] *Probability of Fracture (PROF) Software for Aging Aircraft Risk Analysis, (2005) version 3,* University of Dayton Research Institute.

[23] FAA Federal Aviation Regulations Part 25.571, (2010), *Damage Tolerance and Fatigue Evaluation of Structure.*

[24] FAA AC No. 25.571-1D, (2011), *Damage Tolerance and Fatigue Evaluation of Structure.*

[25] FAA System Safety Handbook, (2000), Chapter 3.

[26] FAA AC No. 25.19, (1994), *Certification Maintenance Requirements*.

[27] FAA document AC No. 120-104, (2011), *Establishing and Implementing Limits-of-Validity to Prevent Widespread Fatigue Damage.*

[28] A. Azevedo, (no date specified), *FAA Risk Analysis Course Manual.*