SAND2014-0248C

The Use of Degradation Measures to Design Reliability Test Plans

Stephen V. Crowder
Sandia National Laboratories
Albuquerque, NM 87185

Jonathan W. Lane
Sandia National Laboratories
Albuquerque, NM 87185

Abstract

With short production development times, there is an increased need to demonstrate product reliability relatively quickly with minimal testing. In such cases there may be few if any observed failures. Thus it may be difficult to assess reliability using the traditional reliability test plans that measure only time (or cycles) to failure. For many components, degradation measures will contain important information about performance and reliability. These measures can be used to design a minimal test plan, in terms of number of units placed on test and duration of the test, necessary to demonstrate a reliability goal. In this work we present a case study involving an electronic component subject to degradation. The data, consisting of 42 degradation paths of cycles to failure, are first used to estimate a reliability function. Bootstrapping techniques are then used to perform power studies and develop a minimal reliability test plan for future production of this component.

Introduction

Often it is difficult to assess component reliability using traditional methods due to a lack of observed failures. For many such components, degradation measures, recorded over time, will contain important information about performance and reliability. These measures can be used to predict remaining time to failure and to estimate the overall reliability distribution for that component. Degradation measures are inherent in situations where failure occurs due to a process of accumulation of damage. In mechanical systems, the failure (breakage of some part) may be caused by the impact of a peak load. The amount of accumulated wear or fatigue damage plays the role of the degradation measure. There are also failures in electronic parts (such as a short circuit, dielectric breakdown, breaking of a circuit, etc.) which are speeded up by some gradual process of deterioration, e.g. corrosion, mechanical deformation, chemical reactions and so on. Metrics on these processes all serve as degradation measures. Lu and Meeker (1993), Meeker and Escobar (1993, 1998) and Meeker and Hamada (1995) all discuss important aspects of degradation data and reliability estimation using degradation data.

The purpose of this paper is to present a case study in which we design a reliability test plan that relies on a degradation measure rather than time to failure data. We will use methods developed by Lu and Meeker (1993) to estimate a reliability distribution using degradation measures. The methods will be applied to estimating the reliability of an electrical component using historical degradation data. Bootstrapping simulation techniques will then be used to investigate minimum sample size requirements and test duration necessary to demonstrate a reliability goal. It will be shown that test plans designed using this approach may require far less testing than traditional life test plans. The test plans developed can be used to demonstrate the reliability of future production lots of this component.

Degradation Measures

Several assumptions are usually made regarding a degradation measure:

1. The state (health) of the component can be characterized by a randomly changing time-dependent variable that we will denote y_t.
2. A failure of the component is defined as a certain catastrophic event ("hard" failure) whose probability of occurrence depends on the value of the variable y_t, or is defined as the variable y_t entering some critical region ("soft" failure).
3. The variable y_t is accessible for either continuous or discrete observations, i.e. a degradation "path" is obtainable.
4. The probabilistic laws governing the changes in the degradation measure y_t are known (physical model) or can be approximated (empirical model).

Under these assumptions y_t is called a degradation measure. Component tests in which degradation data are collected provide, for each unit tested, data consisting of an observed path of degradation measurements (either discrete or continuous) over time. The observed degradation path is a unit's actual degradation path, a monotonic function of time, plus some measurement error. Note also that "time" could be in units of real time or other units such as number of cycles or amount of usage. The path may be censored if the unit is taken off test before failure occurs or it may include the entire degradation path to the point of failure.

An accurate degradation model can be used to show that failure probability is small when the degradation variable is far from the critical region. With censored data, reliability can then be measured in terms of the distance of the degradation measure from the critical region. This is especially important with high reliability products, because high reliability goals can then be demonstrated with less testing than would be required using just time to failure (including censoring times) data. This advantage of degradation data will be exploited in designing a minimal reliability test plan in which no failures may be observed.

Estimating Reliability with a Degradation Measure

The procedure that we will use to estimate reliability using a degradation measure follows that given by Lu and Meeker (1993) and Meeker and Escobar (1998). The steps are:

1. Fit a general path model to each of the n sample degradation paths. Least squares estimation can be used to estimate the parameters for each path.
2. Determine the statistical distribution (using the estimates from the n sample paths) of each of the random parameters from the general path model.
3. Use the resulting distributions to solve for the time to failure distribution $F_T(t)$ if a close form expression exists.
4. If no closed form expression for $F_T(t)$ exists, use the parameter distributions from (2) to simulate a large number N of random degradation paths.
5. To estimate $F_T(t)$, compute the proportion of random paths generated in (3) that cross a pre-determined critical level (which defines failure) before time t. That proportion is the estimate of $F_T(t)$.

To quantify the uncertainty associated with the estimate of $F_T(t)$ we used a slightly different method than Lu and Meeker (1993). The proposed procedure that we investigated is:

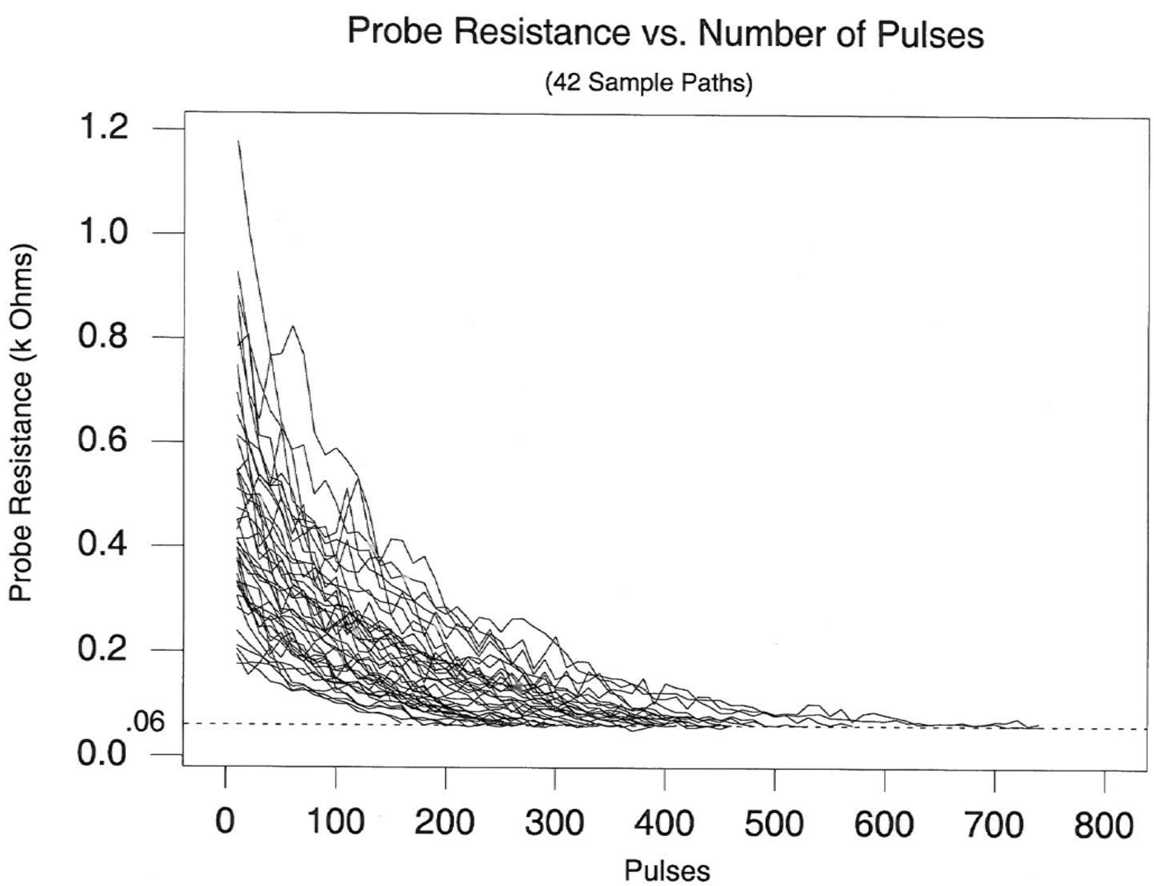
1. Choose a bootstrap sample (sampling with replacement) of n degradation paths from the original n sample paths.
2. Repeat steps 1-4 from the estimation procedure above to obtain an estimate of $F_T(t)$.
3. Repeat steps 1 and 2 here many times (say 1000) to obtain a distribution of values for $F_T(t)$. The central $(100-\alpha)$% of this distribution defines a $(100-\alpha)$% confidence interval for $F_T(t)$.
4. This interval becomes the uncertainty interval for $F_T(t)$.

Sensitivity to sample size was determined by varying the size n of the bootstrap sample in (1). Sensitivity to the length of the test (amount of degradation path observed) was determined by truncating the original sample paths at various points and performing the estimation and uncertainty analyses above using the truncated paths.

Component Example

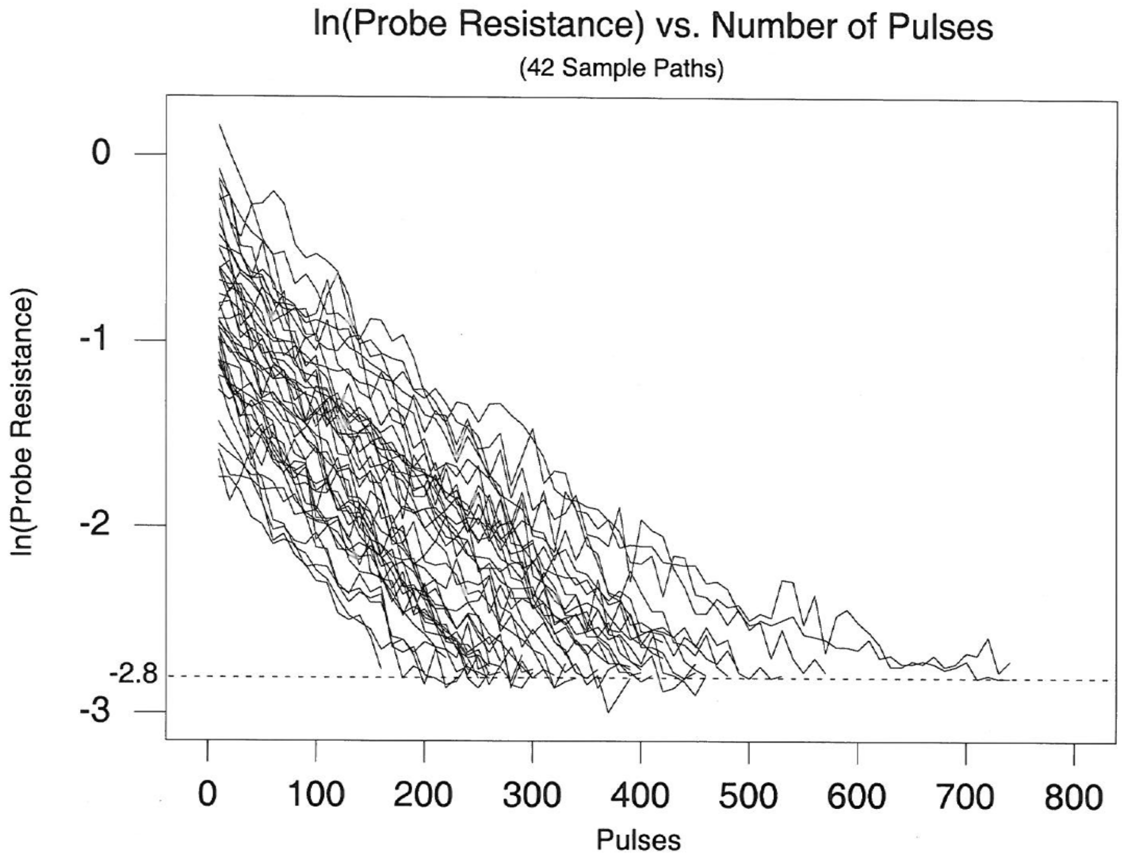
The data used in this case study come from the degradation of an electrical component used to close a high-energy circuit. A low energy pulse causes the component to fire, forming a plasma at the probe, closing the high-energy circuit. The resulting plasma causes a carbon residue to be deposited around the probe. The degradation mechanism is the carbon buildup, and the degradation measure is probe resistance. As the carbon residue builds up due to repeated firings, the probe resistance gradually decreases until a misfire (failure) occurs. The raw probe resistance data for the components are given in Figure 1.

Figure 1. Sample Degradation Paths



The data consist of 42 sample degradation paths of probe resistance from first pulse of the component until failure (misfire). This plot shows that an appropriate critical degradation level is approximately 0.06 kΩ, the level we will use to define "failure" of the component. Because the paths appear to decay exponentially, a logarithmic transformation was applied to the data (see Figure 2). On the log scale the critical degradation level is approximately -2.8.

Figure 2. Sample Degradation Paths on Logarithmic Scale



This plot suggests that a simple linear model should fit the data well, although a constant intercept model will not be appropriate.

Hypothesized Path Model for the Component

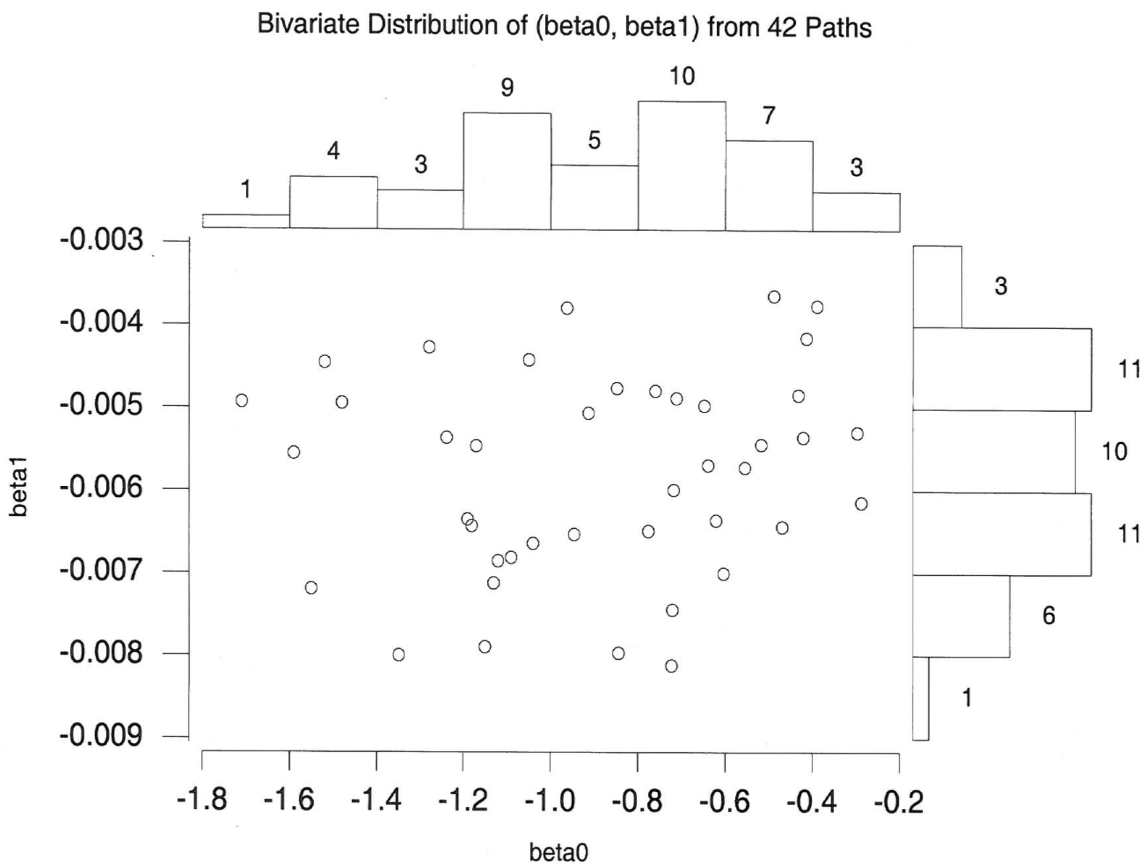
Based on Figures 1 and 2, a plausible path model for the component data is

$$\ln(y_{ij}) = \beta_{0i} + \beta_{1i} t_j + \varepsilon_{ij}, \qquad i = 1, 2, \ldots, 42, \qquad (1)$$

$$\varepsilon_{ij}\text{'s i.i.d.} \sim N(0, \sigma_\varepsilon^2), \qquad j = 1, 2, \ldots, \text{\# Pulses for ith tube,}$$

where y_{ij} is the the jth measured probe resistance of path i, t_j = number of pulses at the jth measurement, ε_{ij} = jth measurement error of path i. Both slope and intercept for this model are assumed to be random. The parameter β_{0i} is the ith unit random intercept and β_{1i} is the ith unit random slope. This model was fit to each of the 42 sample paths, resulting in an empirical distribution of β_{0i}'s and β_{1i}'s. The bivariate plot of the β_{1i}'s vs. the β_{0i}'s in Figure 3 suggests that these parameters are independent and both approximately Normal.

Figure 3. Empirical Bivariate Distribution of the β_{0i}'s and β_{1i}'s



Formal tests of independence and normality showed these assumptions to be reasonable. If we assume $\beta_0 \sim N(\mu_{\beta_0}, \sigma^2_{\beta_0})$ and $\beta_1 \sim N(\mu_{\beta_1}, \sigma^2_{\beta_1})$, it is straightforward to derive a closed form expression for the failure distribution using degradation data.

Failure Distribution Using Degradation Data

The actual degradation of a particular component after t pulses is thus modeled by $D(t) = \beta_0 + \beta_1 t$, where β_0 is the random initial probe resistance and β_1 is the random slope. Both parameters vary from unit to unit according to a Normal distribution. We assume that the probe resistance degrades monotonically and D(t) is a decreasing function, so $\Pr(\beta_1 < 0) = 1$. Given a critical level D_f, we can express the distribution function of T, the random pulses to failure, as

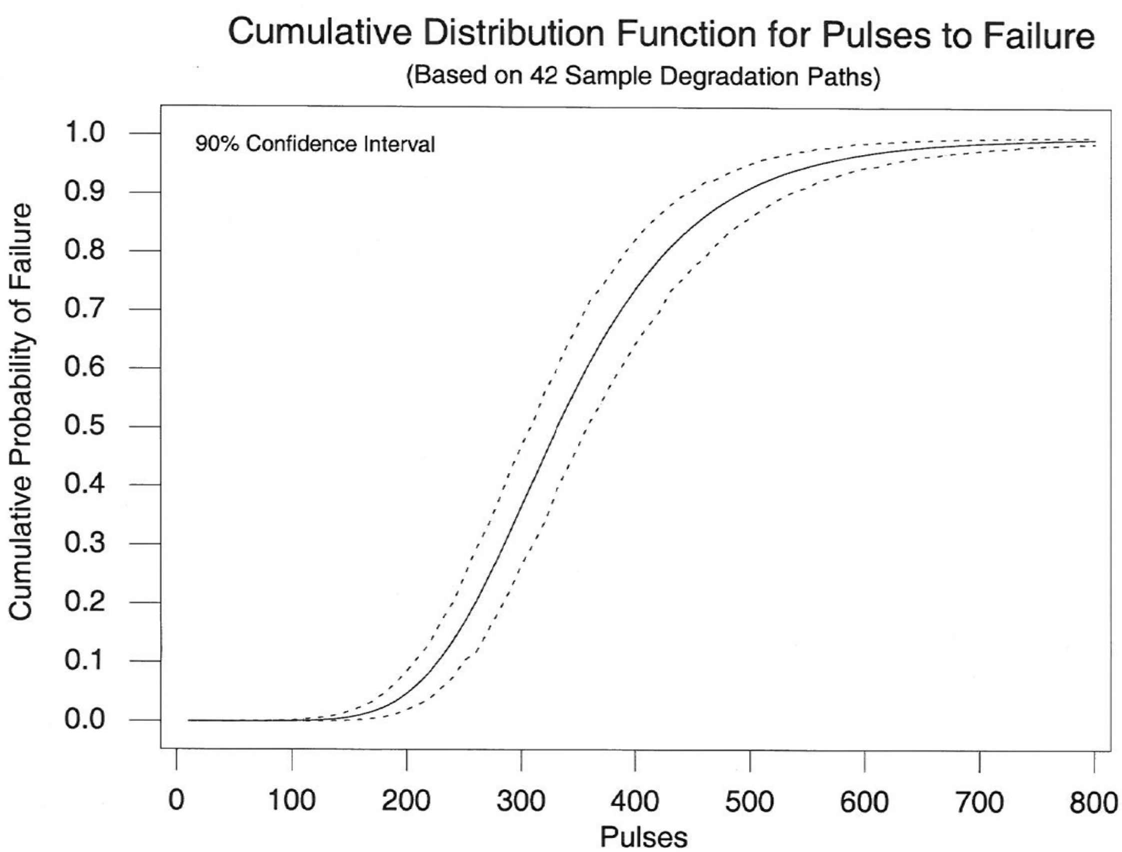
$$F_T(t) = \Pr(T \le t) = \Pr(\beta_0 + \beta_1 t \le D_f)$$

$$= \Pr\left(\frac{\beta_0 + \beta_1 t - (\mu_{\beta_0} + \mu_{\beta_1} t)}{(\sigma^2_{\beta_0} + t^2\sigma^2_{\beta_1})^{1/2}} \le \frac{D_f - (\mu_{\beta_0} + \mu_{\beta_1} t)}{(\sigma^2_{\beta_0} + t^2\sigma^2_{\beta_1})^{1/2}} \right)$$

$$= \Phi\left(\frac{D_f - (\mu_{\beta_0} + \mu_{\beta_1} t)}{(\sigma^2_{\beta_0} + t^2\sigma^2_{\beta_1})^{1/2}} \right), \tag{2}$$

where $\Phi(\cdot)$ is the standard normal distribution function. Figure 4 shows the estimated cumulative distribution function (CDF) for pulses to failure based on degradation data, along with a 90% confidence interval about the estimated function.

Figure 4. CDF for Pulses to Failure



The estimate of the CDF in Figure 4 was based upon evaluation of expression (2). The usual sample averages and variances, based on n=42 fits of model (1), were used to estimate $\mu_{\beta_0}, \sigma^2_{\beta_0}$, μ_{β_1}, and $\sigma^2_{\beta_1}$. Confidence intervals were constructed using the bootstrap percentile method discussed in Efron and Tibshirani (1993). The approach used here was to randomly choose 1000 bootstrap samples of size n=42 from the original 42 sample paths. Those samples were then used with expression (2) to develop a bootstrap distribution at each evaluation point of the CDF. A (100-2α)% confidence interval is then defined by the αth and (100-α)th percentiles of the resulting bootstrap distribution.

Designing a Reliability Test Plan for Components

The initial objective of the testing was to demonstrate that this component design can be pulsed 100 times without failure. A reliability goal of 0.995 (0.005 unreliability) at 100 pulses was established for this component, meaning that the desired probability of surviving at least 100 pulses is 0.995 or greater. Figure 5 is shows that this goal can be easily demonstrated (at a 95% confidence level) based on the test of 42 components pulsed to failure.

Figure 5. CDF of Pulses to Failure with 95% Confidence Interval

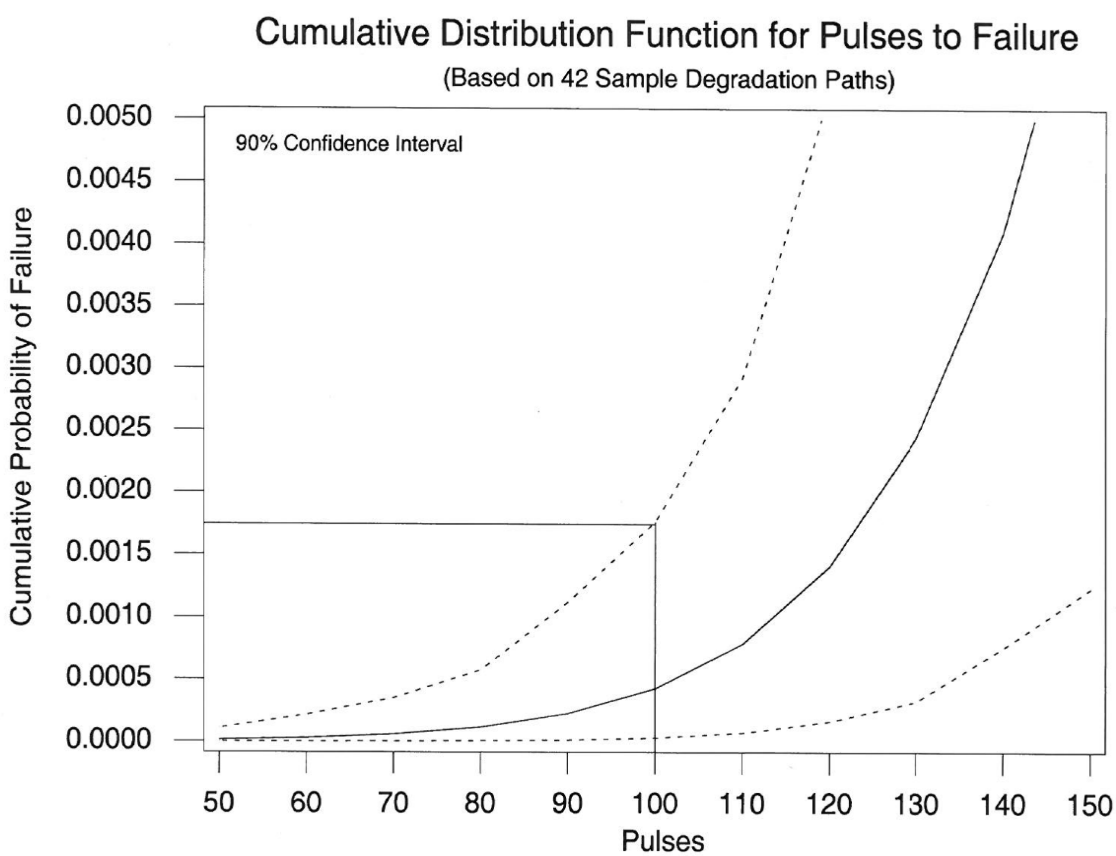


Figure 5 shows the CDF for pulses to failure around 100 pulses, the region of greatest interest. At 100 pulses, a 95% upper confidence bound on *unreliability* is approximately 0.002, so that a 95% lower confidence bound on *reliability* is approximately 0.998. Thus the reliability goal of 0.995 is demonstrated.

The major cost associated with this test is the cost of the components, which are destroyed by testing to failure. Individual components may be expensive to manufacture, so any reduction of sample size results in significant savings. Additional costs are associated with the manpower and equipment required to pulse the components, so reducing the duration (number of pulses) of the test would also result in savings. To demonstrate the reliability goal of the component in future production lots, it is thus of interest to reduce both the number of components tested and the duration of the test. Reducing the number of pulses in this test corresponds to reducing "time on test" in the traditional component reliability study, an important consideration in product development.

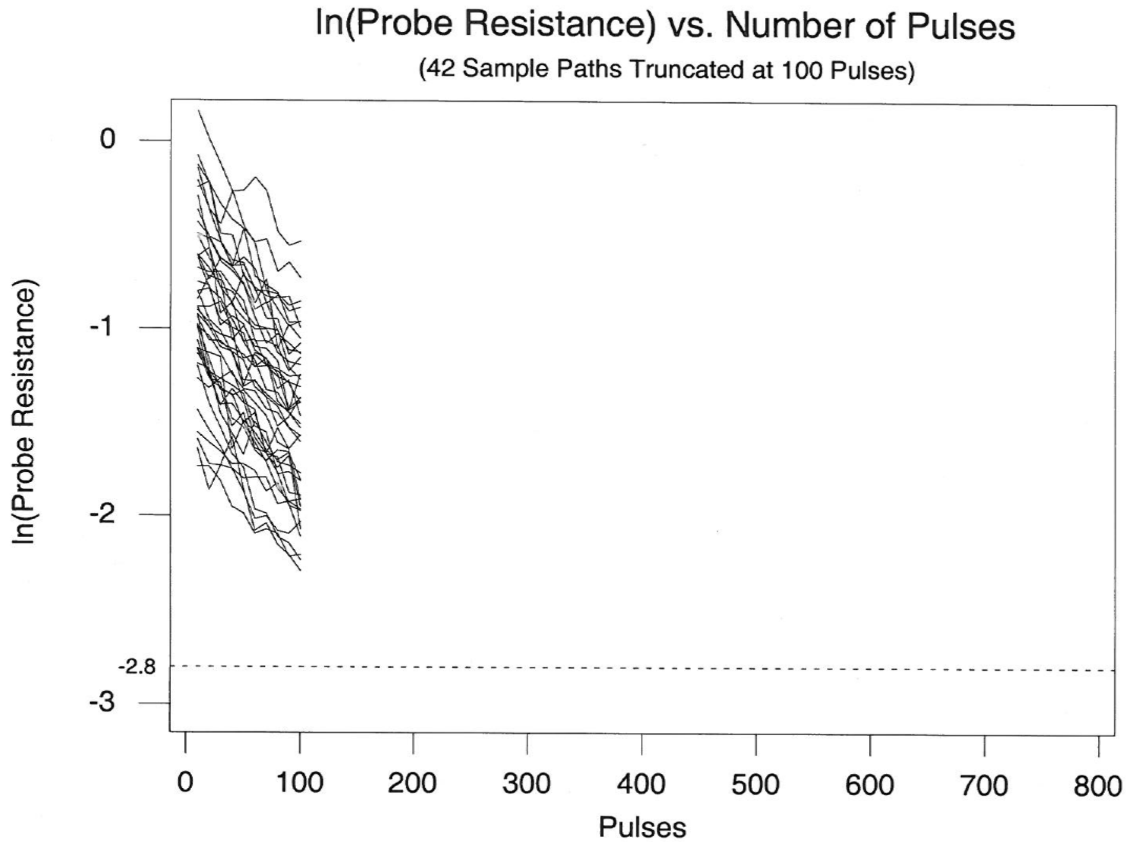
For the purpose of designing a more efficient reliability test plan for components, we next address the issues of test duration (number of pulses per component) and sample size (number of components) necessary to demonstrate the desired reliability goal with high confidence. In particular, we will show how to identify the minimum test duration and minimum sample size requirements to demonstrate the reliability goal using degradation data.

Minimum Test Plan Requirements

To design a minimum reliability test plan we generated, using the same approach as above, reliability functions for the cases in which the test duration was 100, 150, or 200 pulses, or to the point of failure. We also investigated the relative power of samples of size 10, 20, 30, and 42 by analyzing results for bootstrap *sub-samples* of size 10, 20, 30, and 42 from the original 42 sample paths.

Figure 6 shows what the outcome of the test would have been if the test had been truncated after 100 pulses of each component.

Figure 6. Sample Degradation Paths Truncated at 100 Pulses

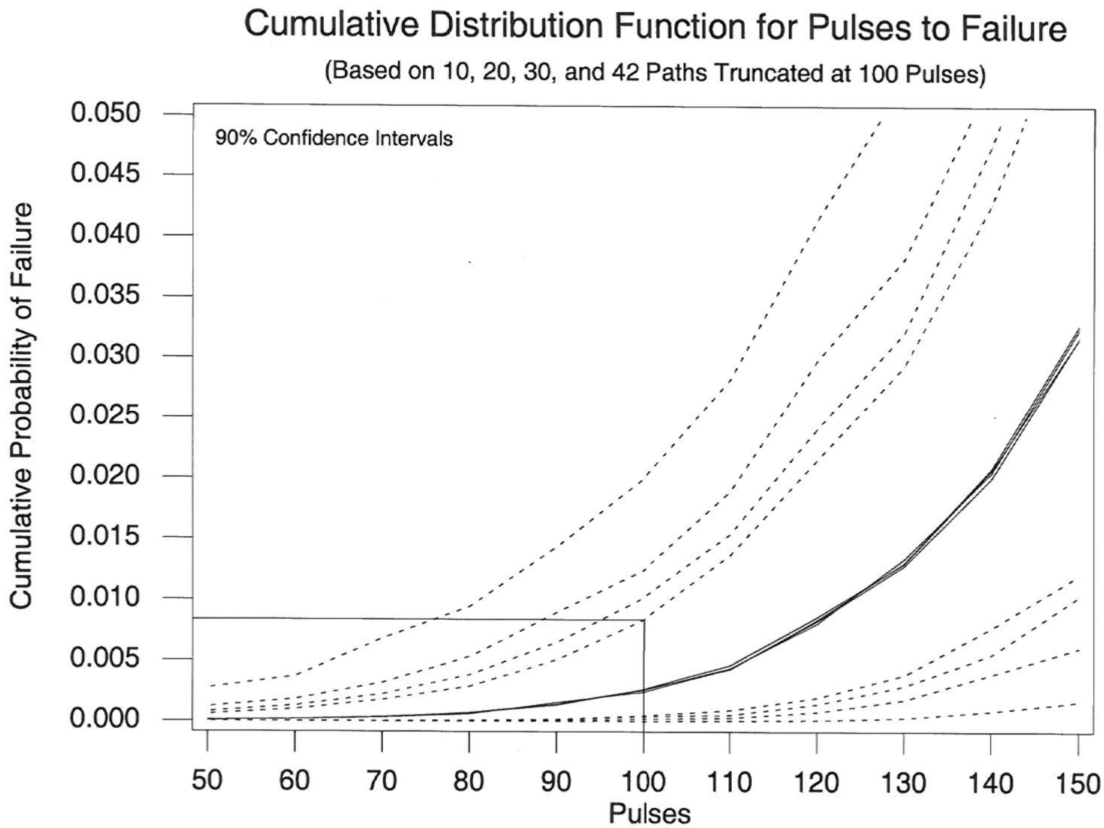


This figure illustrates why there may be great value in recording the probe resistance rather than simply counting the number of pulses. After only 100 pulses we can see from the degradation data that some components are much better than others, based on the different intercepts and slopes of the truncated sample paths. By computing the intercepts and slopes of the truncated paths we can extrapolate to a number of pulses at failure. Without this data, we have no way to extrapolate and important information is lost.

Using the same approach as above, with various numbers of sample paths and various degrees of truncation, we constructed reliability functions and confidence intervals to determine if the goal would have been demonstrated with much less testing. These results provide guidelines for future test recommendations.

Figure 7 shows the reliability functions in the region of interest for 10, 20, 30, and 42 sample paths truncated at 100 pulses.

Figure 7. CDFs Based on Degradation Paths Truncated at 100 Pulses



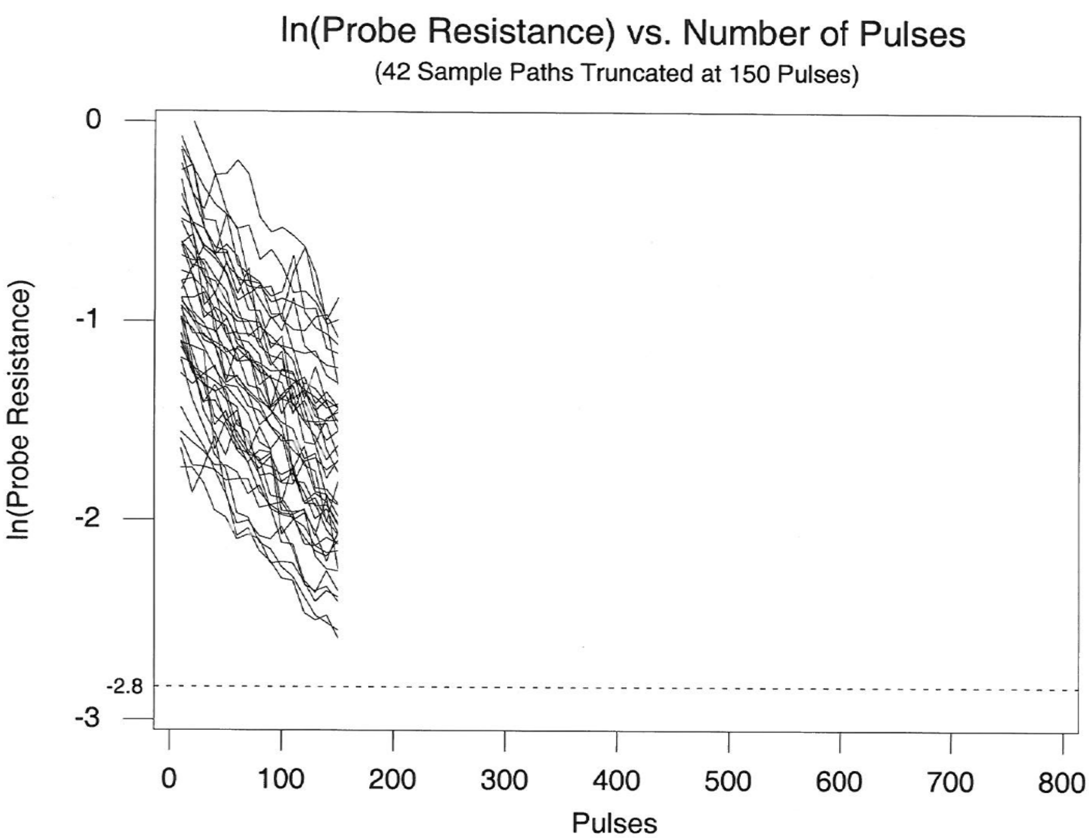
These functions were constructed as described above, with estimates of the β_{0i}'s and β_{1i}'s based on the truncated paths rather than the complete paths. The confidence intervals were again constructed using the bootstrap percentile method, choosing 1000 sub-samples of size 10, 20, 30, and 42 from the original 42 sample paths. The wider confidence intervals in Figure 7 are associated with the smaller sample sizes. The confidence interval constructed for sample size 10 is roughly twice as wide as the confidence interval constructed for sample size 42, suggesting an approximate $n^{1/2}$ relationship between interval width and sample size.

It should be noted that truncation in this case introduces a slight bias, since the complete sample paths had slight curvature near their failure points that is not present in the truncated paths (see Figure 2). Because the bias introduced is in the conservative direction (predicted pulses to failure are biased low), the reliability demonstrated with truncated paths will be somewhat less than could be demonstrated with the complete sample paths. Examination of Figure 7 shows that even with 42 sample paths, a reliability of 0.995 at 100 pulses cannot be demonstrated with sample paths truncated at 100 pulses. The 95% upper bound on unreliability is approximately 0.008, slightly too high.

Using degradation data and the analysis in Figure 7, we can demonstrate a reliability of 0.992 with 42 components and test duration 100 pulses. We can also demonstrate a reliability of 0.980 with just 10 components (see outer confidence interval in Figure 7) and test duration 100 pulses. By comparison, if we ignored the degradation data and used binomial calculations to compute 95% upper confidence bounds on reliability, we could demonstrate a reliability of only 0.93 with 42 components tested to 100 pulses without failure. We could also demonstrate a reliability of only 0.74 with 10 components tested to 100 pulses without failure. Thus much is gained in terms of reliability demonstration by using the degradation data.

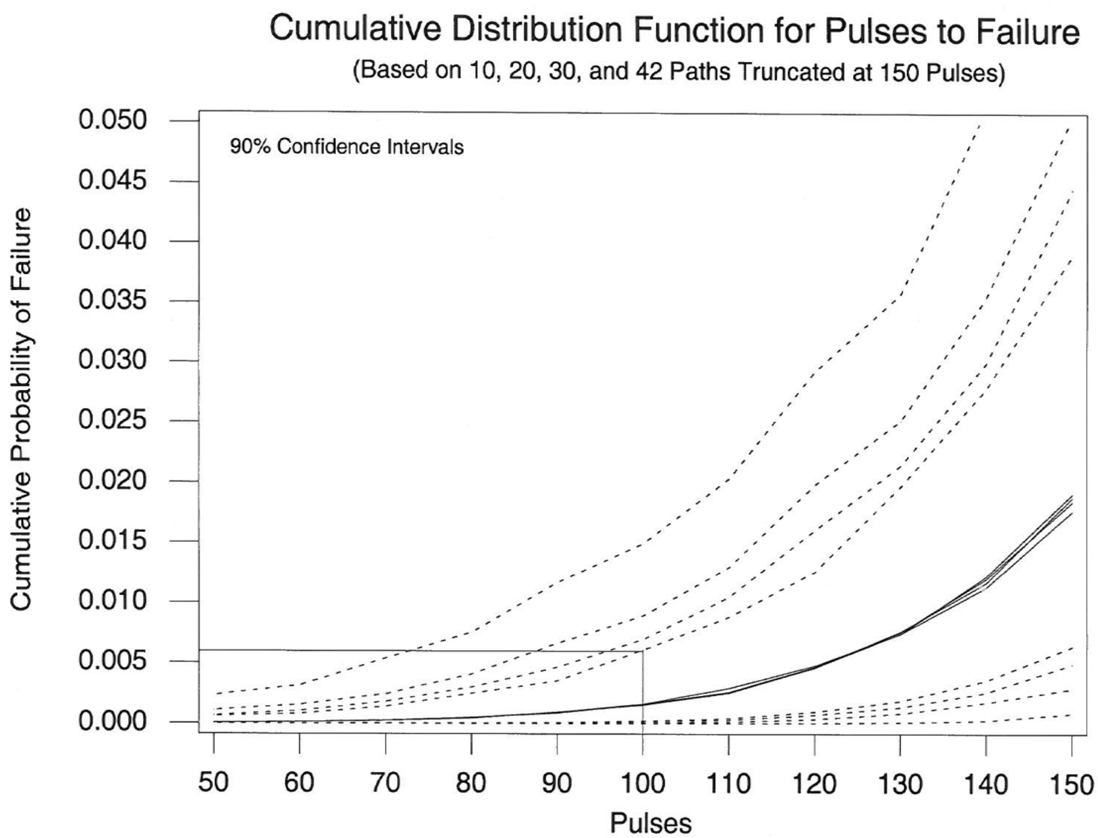
The next step in the investigation was to consider sample paths truncated after 150 pulses. Figure 8 shows what the outcome of the test would have been if the test had been truncated after 150 pulses of each component.

Figure 8. Sample Degradation Paths Truncated at 150 Pulses



Using the same approach as above with 10, 20, 30 and 42 sample paths, we constructed reliability functions and confidence intervals to determine if the goal could have been demonstrated with a test truncated after 150 pulses. Figure 9 shows the reliability functions in the region of interest for 10, 20, 30, and 42 sample paths truncated at 150 pulses.

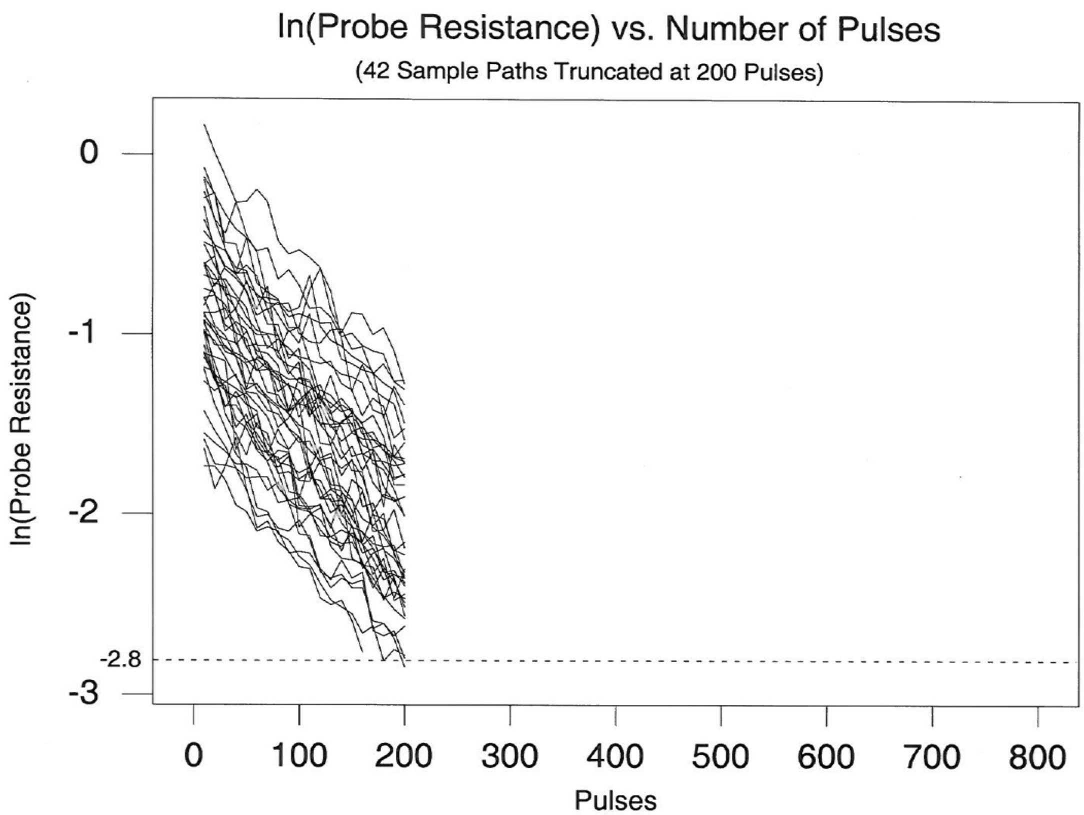
Figure 9. CDFs Based on Degradation Paths Truncated at 150 Pulses



Using degradation data and the analysis in Figure 9, we can demonstrate a reliability of 0.994 with 42 components and test duration 150 pulses (with a 95% upper confidence limit). We can also demonstrate a reliability of 0.985 with just 10 components (see outer confidence interval in Figure 9) and test duration 150 pulses. By comparison, if we ignored the degradation data and used binomial calculations to compute 95% upper confidence bounds on reliability, we could again demonstrate a reliability of only 0.931 with 42 components and a reliability of only 0.741 with 10 components tested 150 pulses without failure. Note that increasing the duration of the test to 150 pulses does not change the reliability estimates when the degradation data is not used.

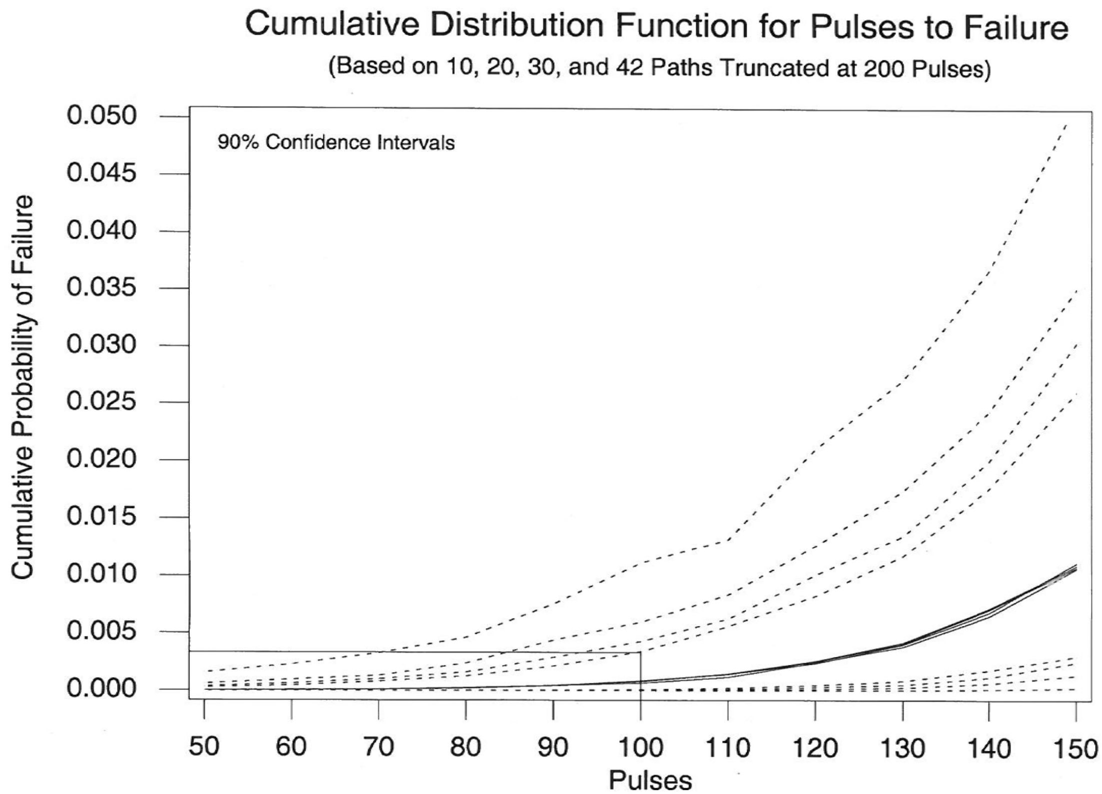
The next step in the investigation was to consider sample paths truncated after 200 pulses. Figure 10 shows what the outcome of the test would have been if the test had been truncated after 200 pulses of each component.

Figure 10. Sample Degradation Paths Truncated at 200 Pulses



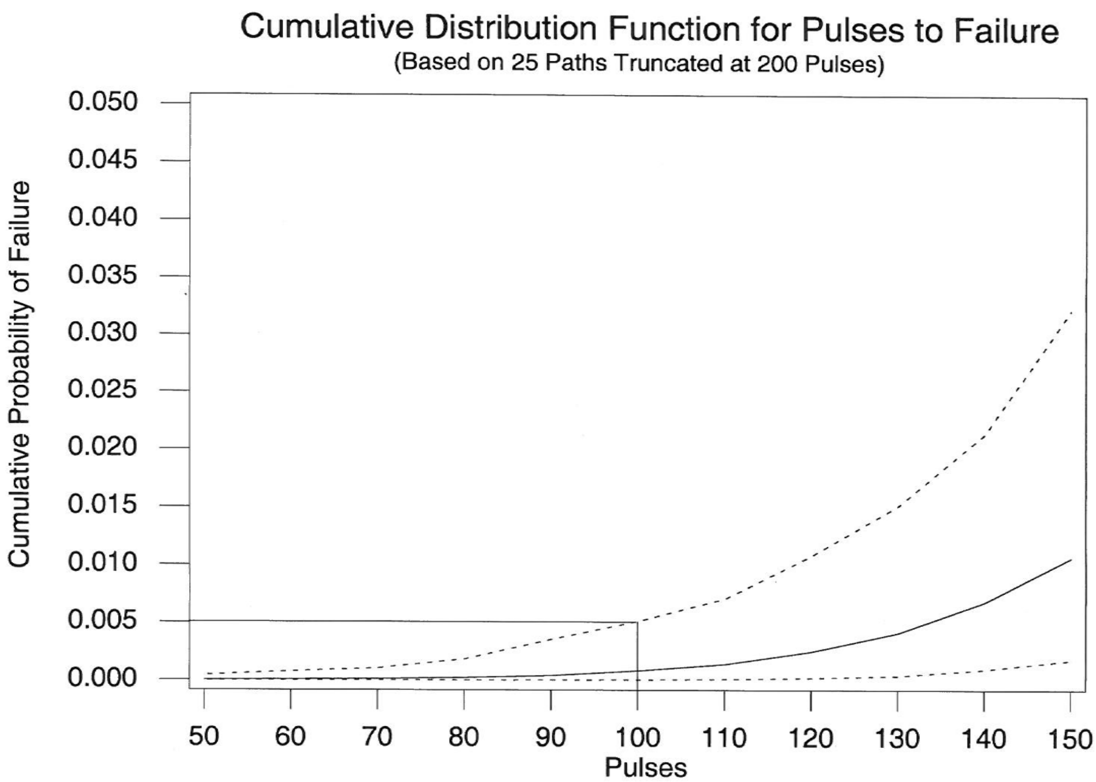
Using the same approach as above with 10, 20, 30 and 42 sample paths, we constructed reliability functions and confidence intervals to determine if the goal could have been demonstrated with a test truncated after 200 pulses. Figure 11 shows the reliability functions in the region of interest for 10, 20, 30, and 42 sample paths truncated at 200 pulses.

Figure 11. CDFs Based on Degradation Paths Truncated at 200 Pulses



Using degradation data and the analysis in Figure 11 we can demonstrate a reliability of 0.996 with 42 components and test duration 200 pulses (with 95% confidence). It also appears that the reliability goal of 0.995 could have been demonstrated with approximately 25 components and test duration 200 pulses. In Figure 12 this is confirmed by examining results for the special case of 25 sample paths with a test duration of 200 pulses.

Figure 12. CDFs Based on 25 Degradation Paths Truncated at 200 Pulses



This analysis indicated that the original test plan consisting of 42 components tested to failure was a much larger test than needed to demonstrate the reliability goal. A test plan consisting of 25 components tested to 200 pulses would have been adequate to demonstrate the reliability goal with the use of the degradation measure. By comparison, to demonstrate 0.995 reliability with 95% confidence using only success/failure data and binomial calculations, approximately 600 components tested to 100 pulses without failure would be required. This example illustrates that tremendous savings can be achieved by using degradation data to demonstrate reliability. The reduction in number of components tested corresponds to savings in cost of product destructively tested, while the reduction in number of pulses corresponds to reduction in time on test. Reducing time on test could be extremely important in terms of reducing product development time.

Summary and Conclusions

We have demonstrated, with a case study involving the reliability of an electrical component, an important advantage to using degradation data and analysis. This approach can be used to develop minimal reliability test plans for demonstrating specific reliability goals. Degradation analysis will often provide considerably more reliability information than traditional failure-time analysis, especially when there are few if any failures over the duration of the test. When data are heavily censored, the degradation approach provides a way of reasonable extrapolation beyond a censored time. In the component example, this ability to extrapolate provided huge savings in number of units tested and test duration required to demonstrate the reliability goal. In general, with heavily censored data, degradation analysis will provide a much tighter confidence bound on the failure distribution in the lower tail of the distribution, important in reliability studies such as presented here. This is because the degradation rates observed early in the test provide much more information about the component's health than simply knowing that a component has survived a given amount of testing.

In practice, it may be that little or no degradation data is initially available. If physical models of degradation can be developed using first principles, then simulation techniques along with bootstrapping can be used to design minimal reliability test plans for actual product as described above. If the model of failure is well understood, it may also be possible in the design stage to reduce the degradation rate or make the component more robust to degradation. In the component example, probe resistance decreased due to a buildup of carbon residue around the probe. This observation could lead to a re-design of the probe, or a change in materials used, so that residue buildup is lessened or does not greatly affect component performance.

Understanding the degradation process can also aid in the establishment of "condition-based" preventive maintenance or replacement plans. Knowing what degradation variable to monitor and having an appropriate degradation model provides a way of determining when a component should be replaced. Degradation measures should be identified during the design and development phases so that methods for accurately measuring the degradation can be built into the product. Downstream analysis of product reliability may result in degradation measures that require destructive inspection or disruptive measurement. Attention to this detail in the design stage will greatly improve the ability to optimize preventive maintenance schemes. Crowder (1997) discusses using degradation measures to establish preventive maintenance plans.

Although there are many advantages to using degradation measures, some cautions must also be stated. In particular, the degradation measure may not in reality be closely related to failure. There should be a physical explanation for the use of the measure. In the component example the physical explanation for a decrease in probe resistance was the buildup of carbon residue after each pulse of the component. Another potential problem is that the model relating the value of the degradation measure to remaining time to failure may be hard to identify, and over-simplified models could introduce bias. In the component example a simple empirical model was developed for probe resistance. A simple linear model with random intercept and random slope fit the data rather well, although a slight bias was introduced by truncating the sample paths. Because the bias introduced was in the conservative direction (the predicted pulses to failure were biased on the low side), the reliability goals could still be demonstrated with truncated paths. Another caution is that degradation measurements may be contaminated by measurement error, although Lu, et al (1993) show that the measurement error must be severe for the standard failure time analysis with censored data to be superior.

References

Crowder, S. V. (1997). Statistical Case Studies for Industrial Process Improvement. V. Czitrom and P. Spagon, editors. Society for Industrial and Applied Mathematics.

Efron, B. E. and R. J. Tibshirani (1993). An Introduction to the Bootstrap. Chapman and Hall, NY.

Lu, J. L. and W. Q. Meeker (1993). "Using Degradation Measures to Estimate a Time-to-Failure Distribution." Technometrics, v. 35, pp. 161-173.

Lu, J. L. , Meeker, W. Q., and L. A. Escobar (1993). "A Comparison of Degradation and Failure-Time Analysis Methods of Estimating a Time-to-Failure Distribution." Preprint Number 93-37, Department of Statistics, Iowa State University.

Meeker, W. Q. and M. Hamada (1995). "Statistical Tools for the Rapid Development and Evaluation of High-Reliability Products." IEEE Transactions on Reliability, v. 44, pp. 187-198.

Meeker, W. Q. and L. A. Escobar (1998). Statistical Methods for Reliability Data. Wiley, NY.