



www.inl.gov



Quality Control Using Inferential Statistics In Weibull Based Reliability Analyses

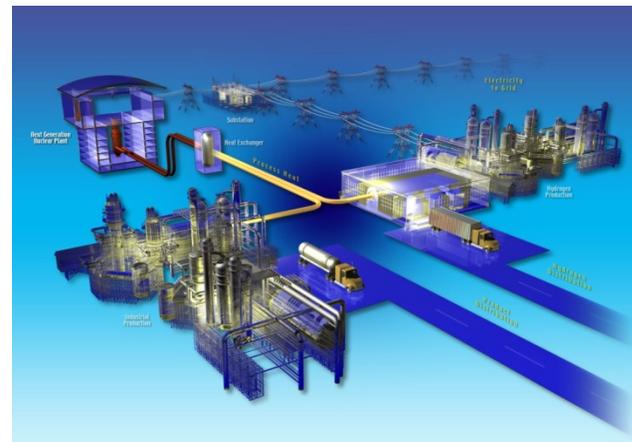
S. F. Duffy¹ and A. Parikh²

¹Cleveland State University

²N & R Engineering

*ASTM Symposium on
Graphite Testing for Nuclear Applications:
The Significance of Test Specimen Volume and
Geometry and the Statistical Significance of Test
Specimen Population*

September 19-20, 2013, Seattle, WA, USA



MOTIVATION

Recently a standard practice on “Reporting Uniaxial Strength Data and Estimating Weibull Distribution Parameters for Advanced Graphites” (ASTM D 7486) survived the ASTM voting process and has been accepted as a Society Standard Practice.

Efforts within the ASME Boiler and Pressure Vessel committees have produced an extensive treatment on designing reactor components from graphite materials.

The presentation today is a small attempt to continue efforts to bridge the ASTM graphite testing community and the ASME graphite design community. The seemingly simple question of how many test specimens should be tested is addressed.

INTRODUCTION

Simple procedures have been proposed for accepting a graphite material based on strength. An enhanced approach that incorporates the inherent variability associated with parameter estimation as well as component reliability is presented.

Tensile strength is a random variable characterized by a two parameter Weibull distribution. The Weibull distribution is an extreme value distribution and this facet makes it a preferred distribution to characterize the “minimum strength” of graphite.

In ASTM Standard Practice D 7486 Weibull distribution parameters are estimated using maximum likelihood estimators. Point estimates computed from data are approximate values of the true distribution parameter values.

*The point estimates are dependent on the number of strength tests conducted and the question of whether the estimated values are of sufficiently high quality is directly related to the fundamental question of “**how many samples must be tested?**” However, the more appropriate question is “**how many samples must be tested to establish a given confidence level for a stipulated component reliability?**” We wish to address the latter question properly by utilizing interval estimates along with hypothesis testing.*

Confidence intervals are used to indicate the quality of estimated parameters. Confidence intervals on parameter estimates represent the range of values for the distribution parameters that are both reasonable and plausible. Within the interval one can expect the true value of a distribution parameter with a quantifiable degree of confidence.

When confidence intervals are combined with hypothesis testing a rejection region is established, and the notion of a likelihood ratio can be developed wherein the estimates should reside, and moreover, where the true distribution parameter pair can be found.

Hypothesis Testing

Hypothesis testing is applied to estimates of the Weibull distribution parameters (σ_θ , m). Achieving statistical significance is equivalent to accepting that the observed results (the point estimates) are plausible and a null hypothesis (H_0) is not rejected. If properly crafted, the null hypothesis helps to derive a rejection region in a distribution parameter space.

To do this a test statistic is developed that aids in our decision to reject the null hypothesis. For our use the statistic, based on a ratio of likelihood functions, helps in defining the relative size of the rejection region.

The objective is to define the process relative to strength testing of graphite materials and establish a robust material acceptance criteria. The process has been discussed within the ASME nuclear design code committee for graphite.

Parameters for a distribution are identified generically by the vector

$$\theta_i = (\theta_1, \theta_2, \theta_3, \dots)$$

Since the tensile strength for graphite is assumed characterized by a two parameter Weibull distribution, then a vector of distribution parameters whose components are the MLE parameter estimates can be identified as

$$(\hat{\theta}_1, \hat{\theta}_2) = (\hat{m}, \hat{\sigma}_\theta)$$

A null hypothesis is stipulated such that

$$H_0 : (\theta_1, \theta_2) \in \Theta_0 = (\hat{m}, \hat{\sigma}_\theta)$$

that is the components of the vector Θ_0 , i.e., the MLE parameter estimates, are equal to the true distribution parameter. The alternative hypothesis is the vector components are not equal

$$H_1 : (\theta_1, \theta_2) \in \Theta_0^c$$

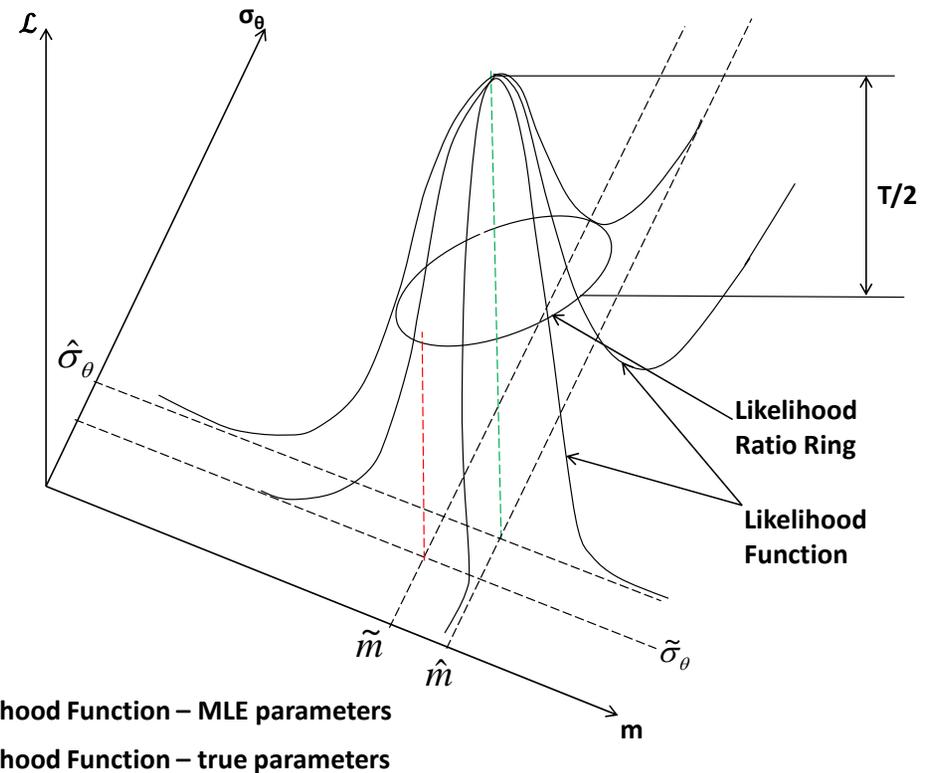
Constructing the Test Statistic

For the two parameter Weibull distribution define a log likelihood function as

$$\mathcal{L} = \ln \left\{ \prod_{i=1}^n f(x_i | m, \sigma_\theta) \right\}$$

Given a specific data set the likelihood function can be constructed in the $m - \sigma_\theta$ parameter space.

The MLE parameter values define the location of the peak of this function. The functional value of the log likelihood function at the maximum is represented in green.



If we knew the true population distribution parameters, the functional value of the log likelihood function for the same data set would be represented by the red line.

The value of the likelihood function associated with a specific data set that is functionally evaluated at the MLE parameter estimates is expressed as

$$\hat{L} = \prod_{i=1}^n f(x_i | \hat{m}, \hat{\sigma}_\theta)$$

A second likelihood function for the same data set is functionally evaluated at the true distribution parameters, i.e.,

$$\tilde{L} = \prod_{i=1}^n f(x_i | \tilde{m}, \tilde{\sigma}_\theta)$$

A test statistic now is introduced that is defined as the natural log of the ratio of the likelihood functions, i.e.,

$$T = -2 \ln \left(\frac{\tilde{L}}{\hat{L}} \right)$$

The Neyman-Pearson lemma (1933) states that this likelihood ratio test is the most powerful test statistic available for testing the null hypothesis stipulated earlier. Wilks (1939) showed that as n increases the test statistic T becomes asymptotically χ^2 -distributed.

One can either compute the likelihood ratio T and compare $-2\ln(T)$ to χ^2 values and determine the corresponding significance level, or define a rejection region by assuming a significance level, calculating the correspond χ^2 value (one degree of freedom), computing T , and finding parameter pairs that satisfy the value of T . This is outlined in the next section.

The test statistic is designed in such a way that the probability of a Type I error does not exceed the constant α , a value that we control. Thus the probability of a Type I error is fixed at an acceptable low value.

The ratio of the two likelihood functions defined previously should be low in the optimal critical region – a result of minimizing α and maximizing $(1 - \beta)$. The ratio is high in the complementary region

Likelihood Ratio Ring (Region of Acceptance)

The likelihood ratio confidence bound is based on the equation

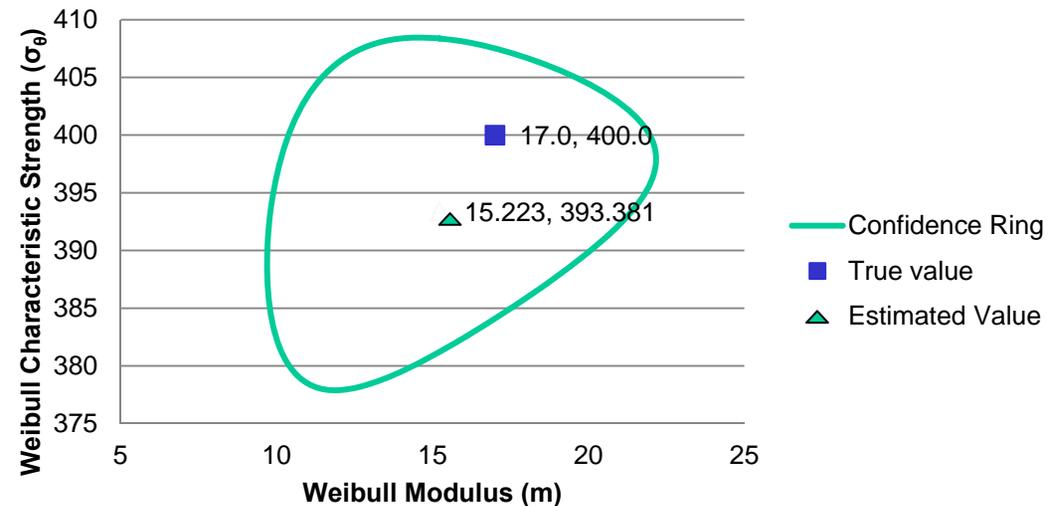
$$-2 \cdot \ln \left(\frac{L(\theta)}{L(\hat{\theta})} \right) \geq \chi_{\alpha;1}^2$$

For a two parameter Weibull distribution this equation is expressed as

$$L(m, \sigma_{\theta}) - L(\hat{m}, \hat{\sigma}_{\theta}) \cdot e^{\frac{\chi_{\alpha;1}^2}{2}} = 0$$

Confidence Ring

$\alpha = 10\%$



Above a virtual data set was generated using Monte Carlo simulation. The sample size was $n=10$. Here the true distribution parameters are known (17, 400).

In real life we do not know the values of the true distribution parameters. However, the confidence bounds provide us with a degree of comfort as to where to expect to find them. That measure of comfort is quantified by α .

The four data sets in the figure to the right ranged in size from 83 to 253.

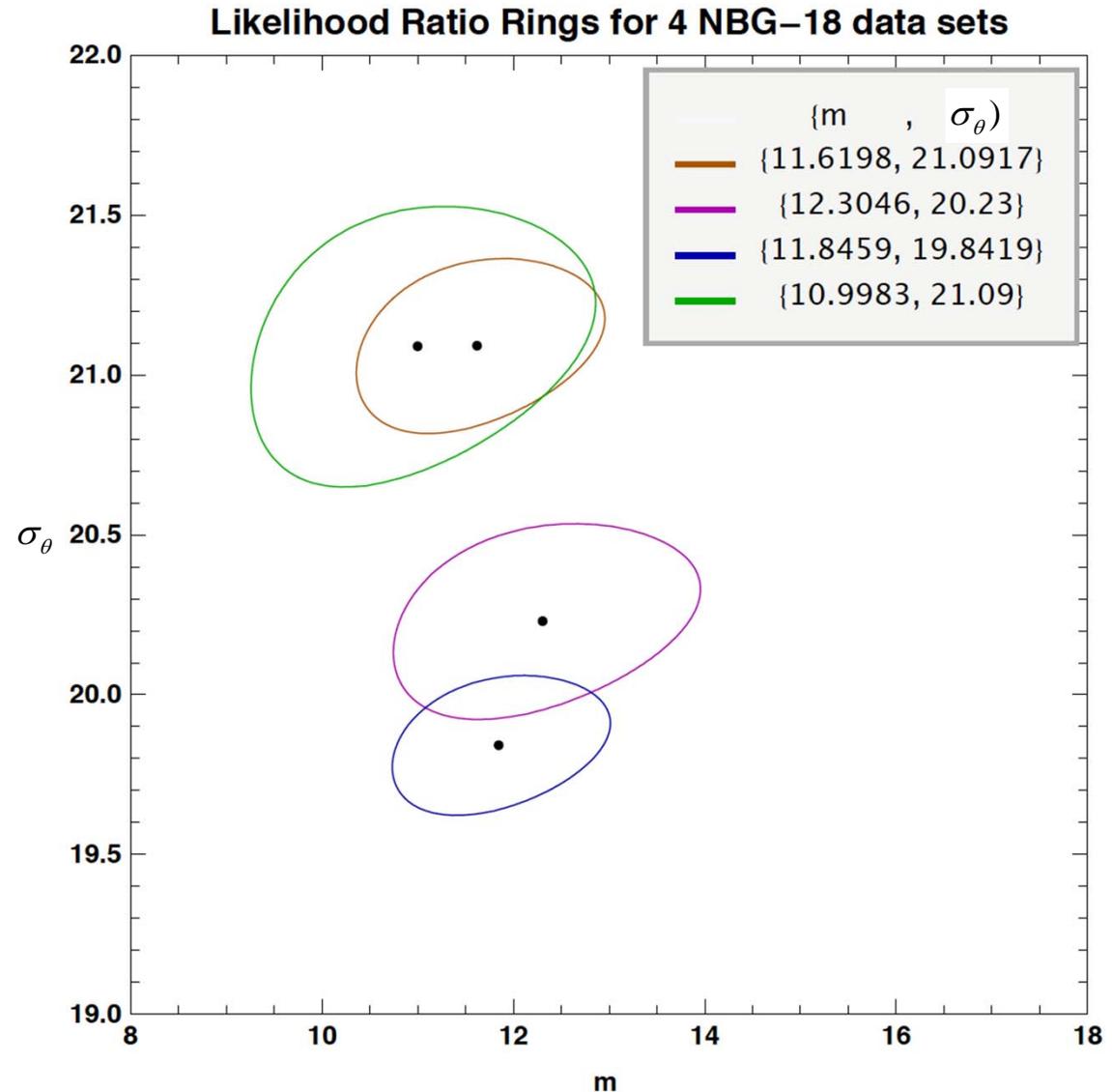


Figure courtesy of R.L. Bratton (INL TPOC)

Component Performance Curves

The acceptance criterion depends on establishing an acceptable probability of failure for the component under design. This can be quantified using a hazard rate format, i.e., expressed as a fraction with a numerator of one. The hazard rate is simply the reciprocal of the probability of failure a quantity usually expressed as a percentage.

With the tensile strength characterized by a two parameter Weibull distribution then a component probability of failure curve can be generated in an $m - \sigma_\theta$ graph. Using the expression

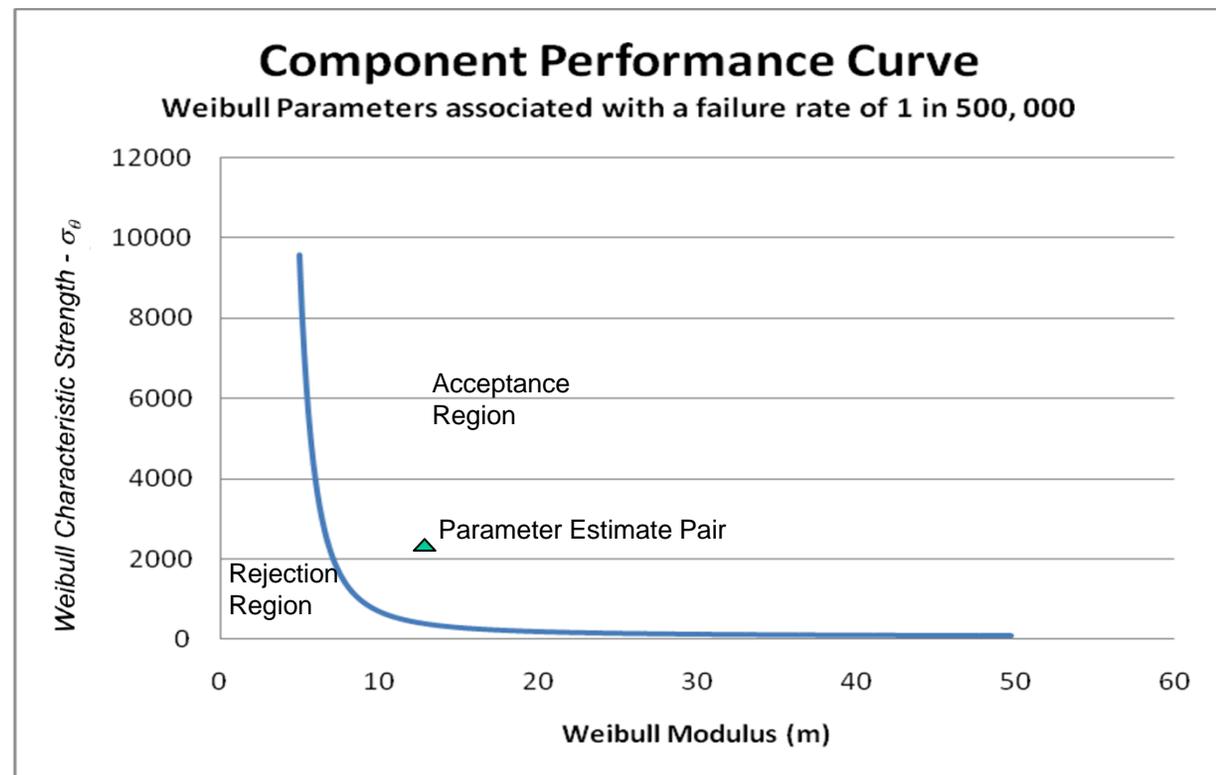
$$P_f = 1 - \exp\left[-\left(\frac{\sigma}{\sigma_\theta}\right)^m\right]$$

then points along the curve in the following figure represent parameter pairs equal to a specified stress and probability of failure given by this, or similar expressions. The curve generated by this expression will be referred to as a component performance curve.

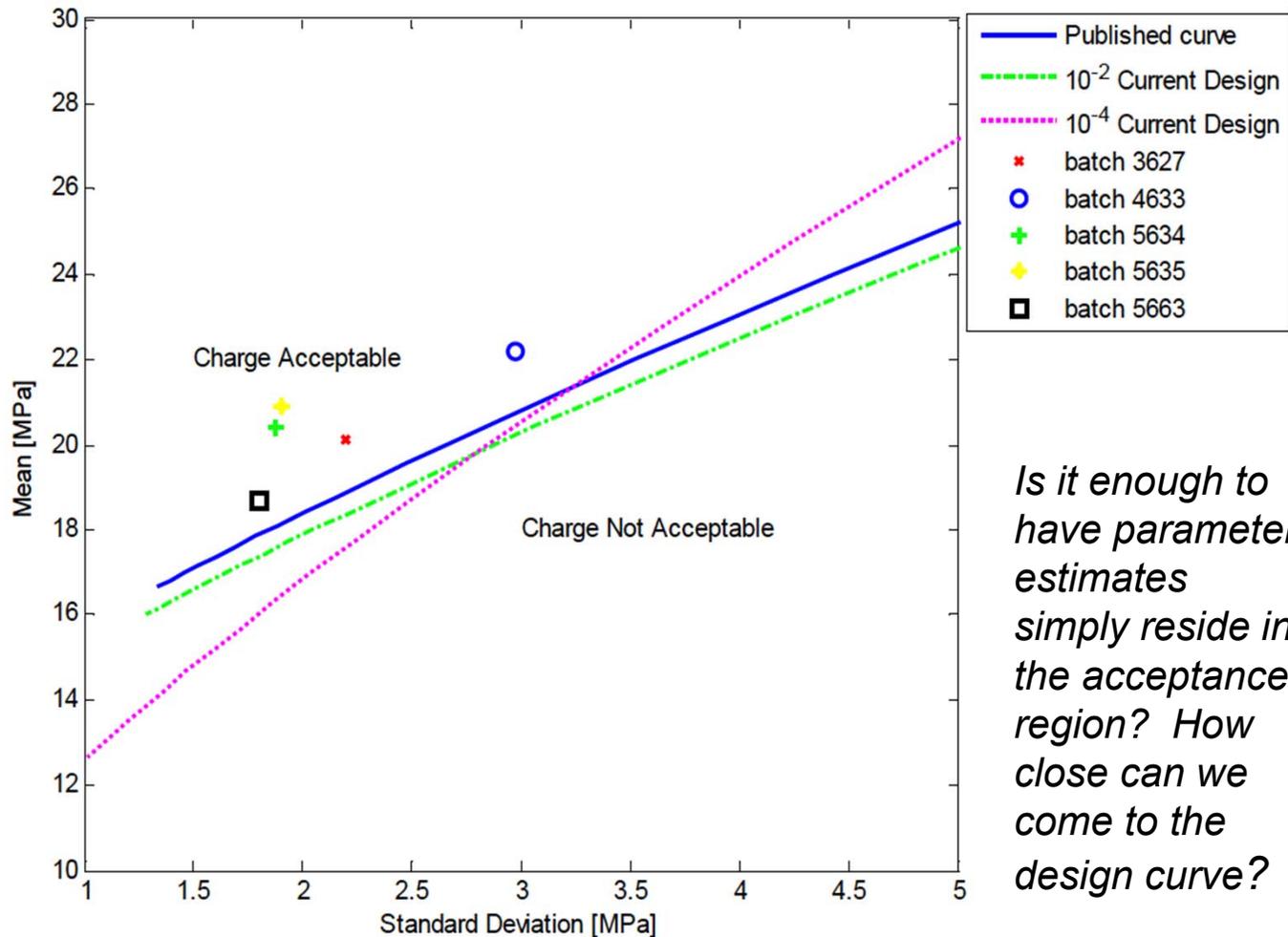
One component design curve defines two regions of the $m - \sigma_\theta$ space, an acceptable performance region and a rejection region relative to a specified component probability of failure.

We overlay this curve with point estimates of the distribution parameters obtained from tensile strength data.

Parameter pairs to the right of the component performance curve have lower probability of failures for a given design stress, and parameter pairs to the left are associated with higher probability of failures.



Alternative component performance curves exist. The graph to the right presents Weibull parameter estimates in terms of their equivalent mean and standard deviation.



Is it enough to have parameter estimates simply reside in the acceptance region? How close can we come to the design curve?

Figure 4: Strength acceptance curves – Current Design and Published

The figure to the right is an example of a component performance curve and the previous examples of likelihood ratio rings for NBG-18. An increase in the failure rate would shift the curve to the left, a decrease would shift the curve to the right. Curves are graphed based on the following expression

$$P_f = 1 - \exp \left[- \left(\frac{S_g}{S_{c95\%}} \right)^{m_{95\%}} \right]$$

which is from ASME Boiler and Pressure Vessel Code, "Article HHA-II-3000," Section III, Division 5

Probability of Failure Curves – ASME Appendix II Equation 21

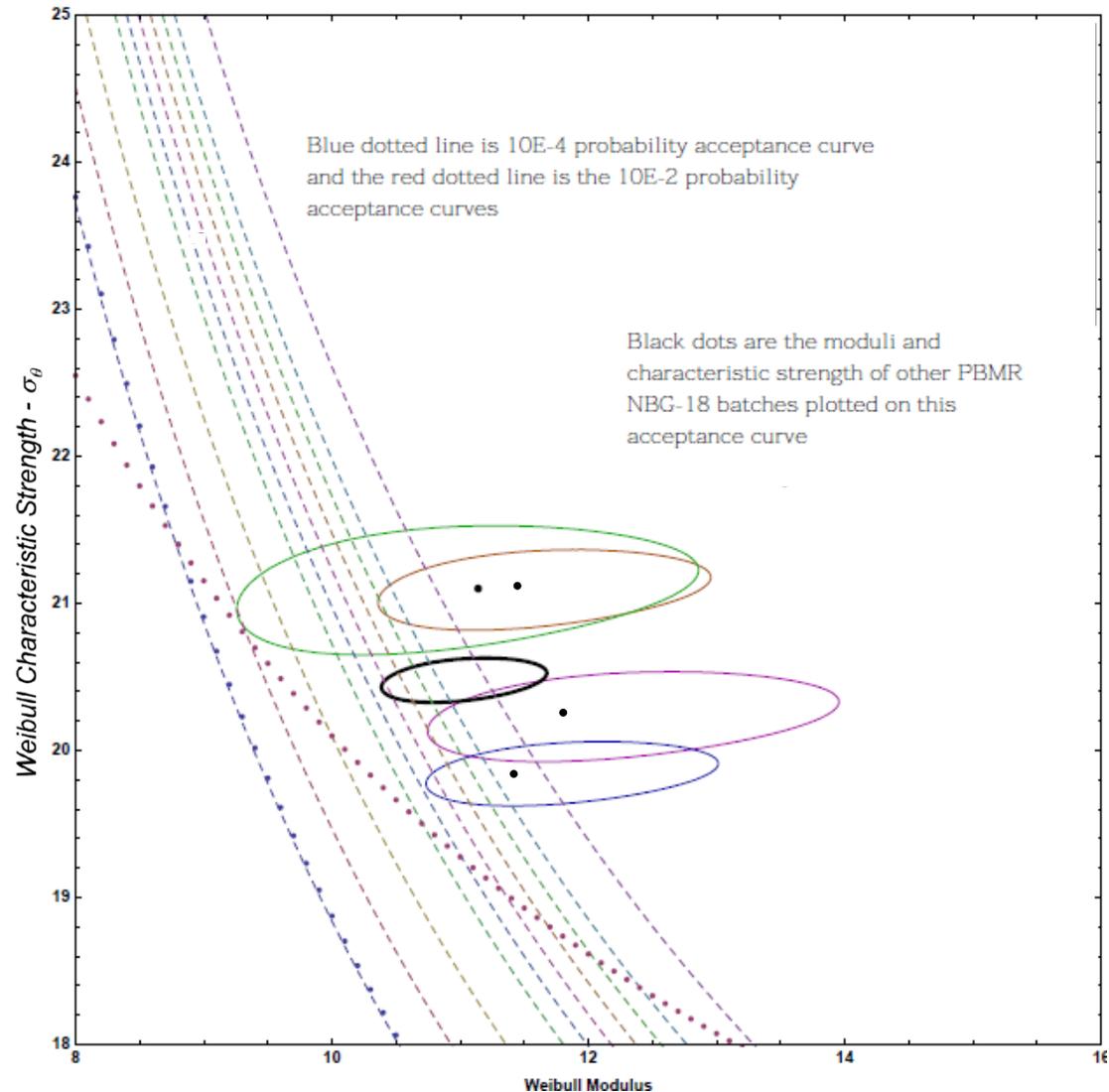
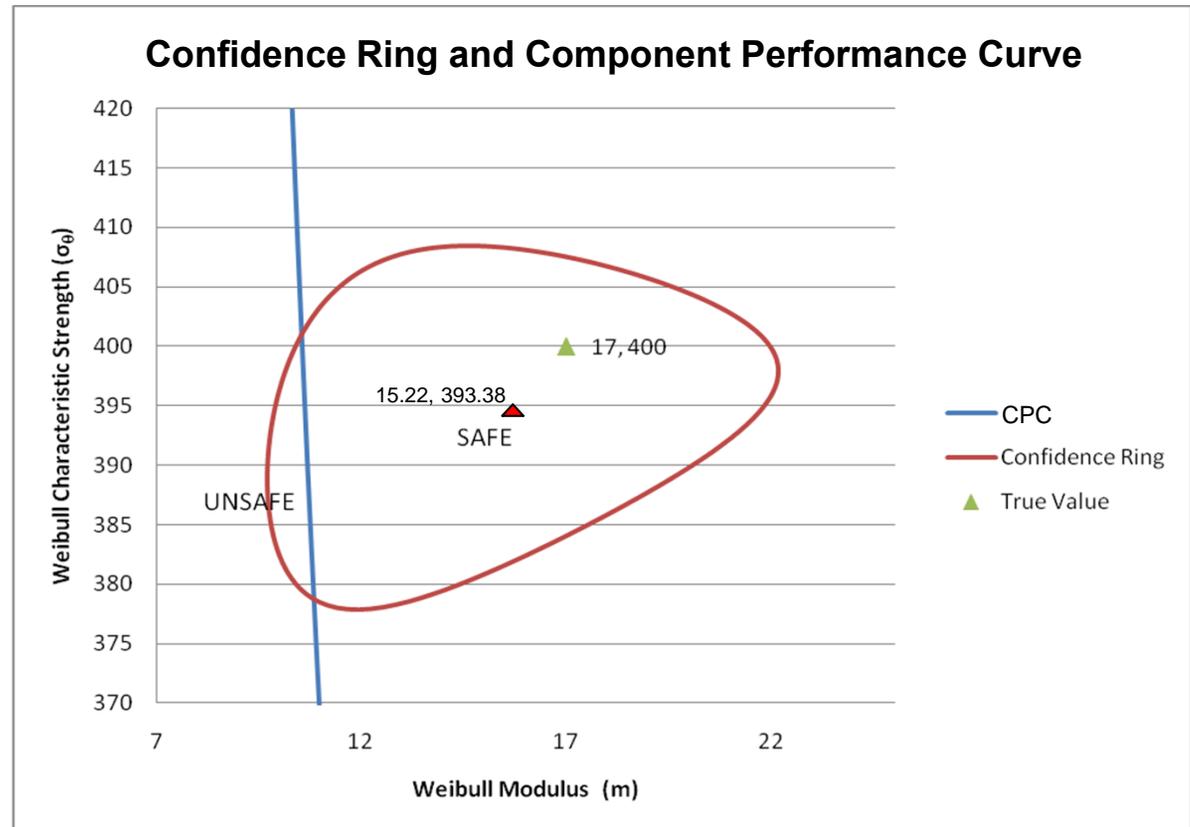


Figure courtesy of R.L. Bratton (INL TPOC)

In Parikh's thesis (2011) generic data sets were generated via Monte Carlo simulation. The true distribution parameters were known. The red curve is a likelihood ratio ring that corresponds to a data set with 10 test specimens. The blue curve is the component performance curve.

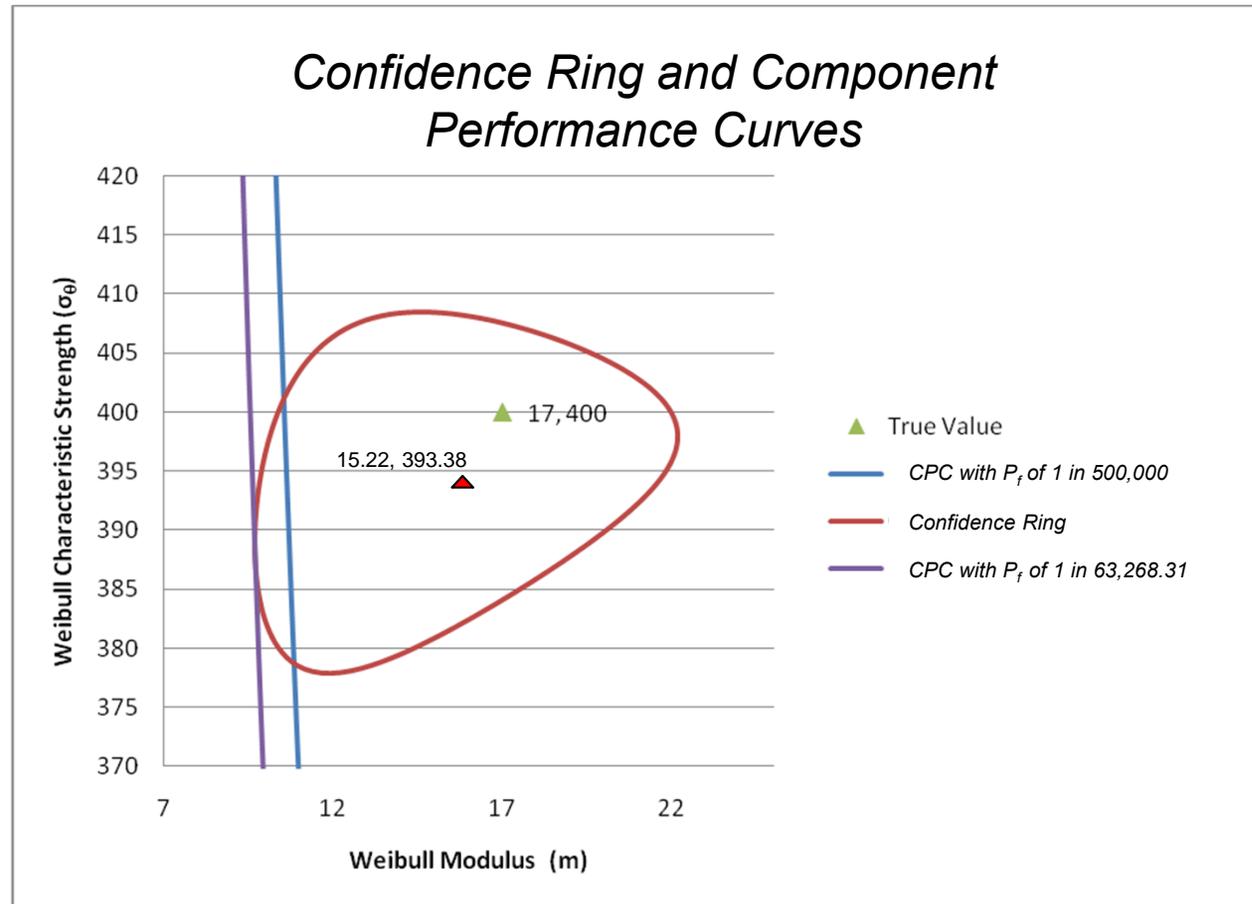
Point estimates were made using maximum likelihood estimators.



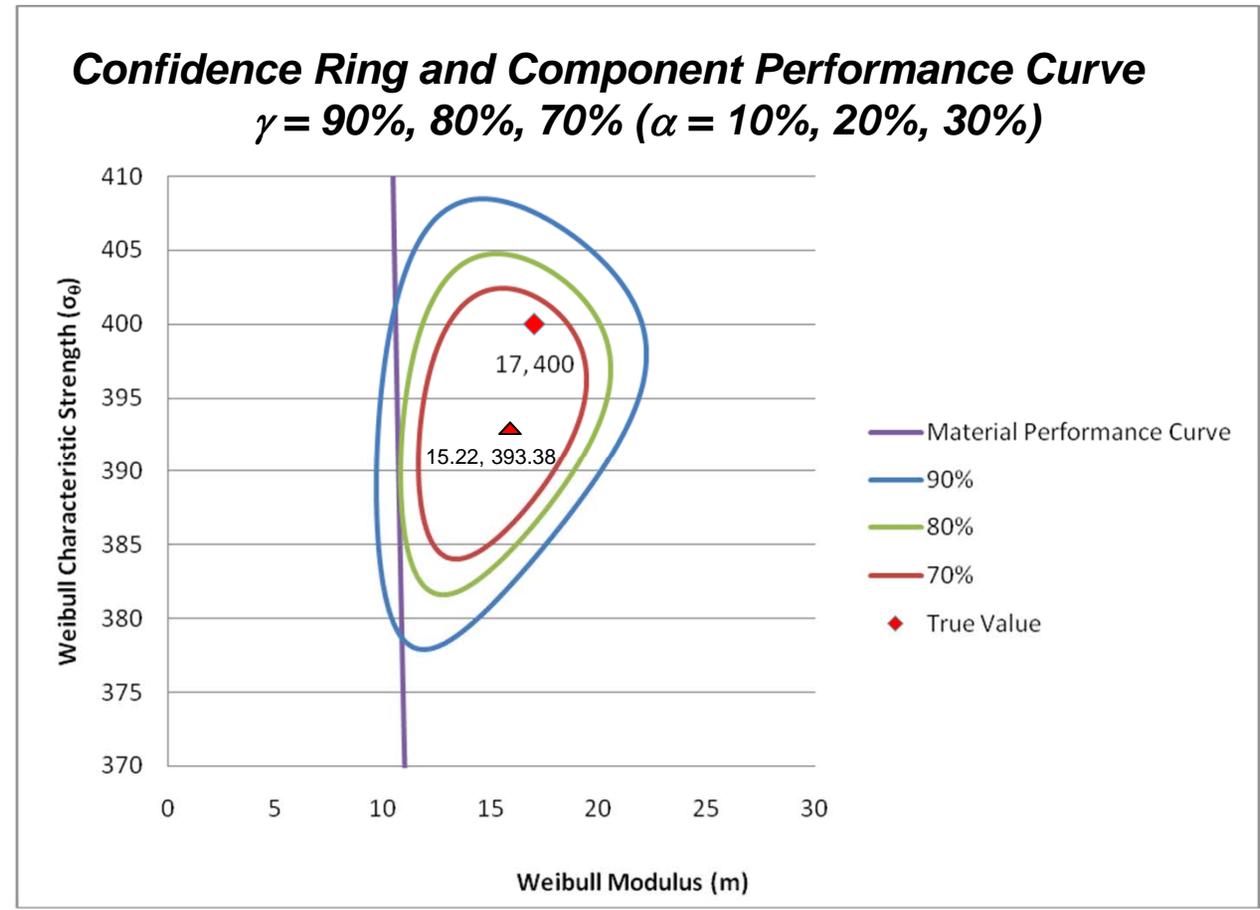
The ring was generated at the 90th confidence level ($\alpha = 10\%$) and it is partitioned by the component performance curve.

To insure that an acceptance region enclosed by the likelihood ratio ring is completely to the right of the component reliability curve we can either adjust the probability of failure by increasing it (not a good approach)....

Confidence Ring and Component Performance Curves



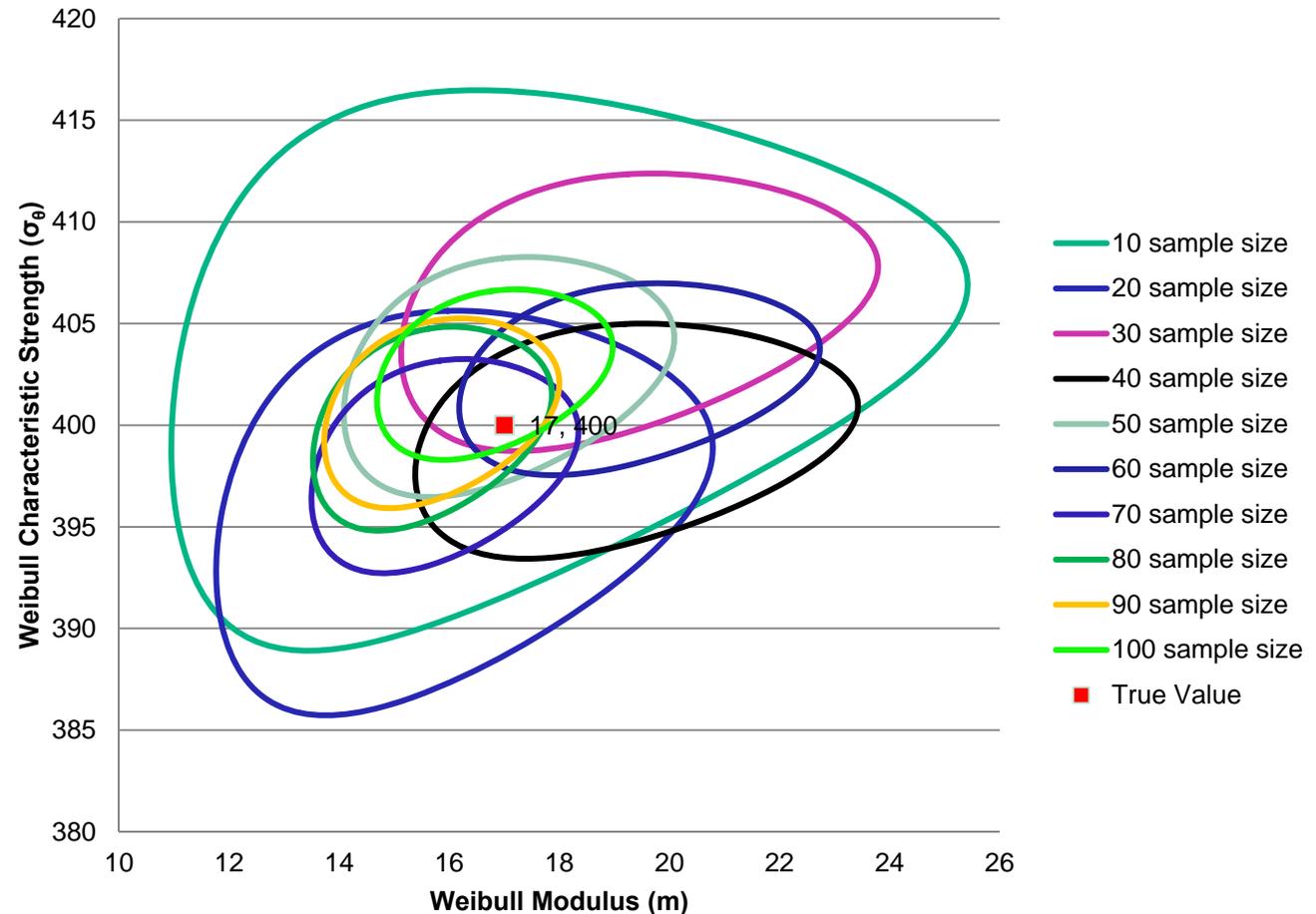
... or we can decrease the confidence level until the likelihood ratio ring is tangent to the material performance curve (see the figure to the right).



Likelihood Confidence Rings with increasing sample size (N) and fix $\gamma(=90\%)$

... or we increase the size of the data set and hold the confidence level steady to shrink the size of the confidence ring.

Should be the preferred approach by regulators.



Summary

- *A topic was presented to bridge one of several gaps between experiment and component modeling, i.e., how many tests should be conducted given a component design.*
- *The question was answered by combining two common statistical inference procedures: confidence interval estimation and hypothesis testing. Confidence interval estimates provide a range of possible values within which the true, unknown population parameters reside. Hypothesis testing speaks to the quality of the data. The two concepts can be combined to produce a material acceptance criterion. The concepts presented here are being considered for incorporation into the ASME nuclear design code.*
- *Other issues exist that. Size dependence in graphite appears to be a function of density. A more robust size scaling methodology is needed that incorporates density in a coherent fashion. A coherent solution requires a joint effort among experimentalists and designers.*