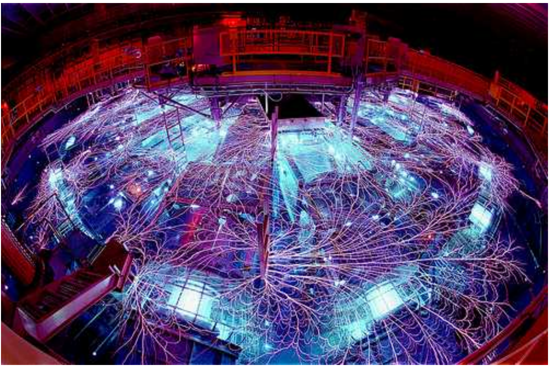
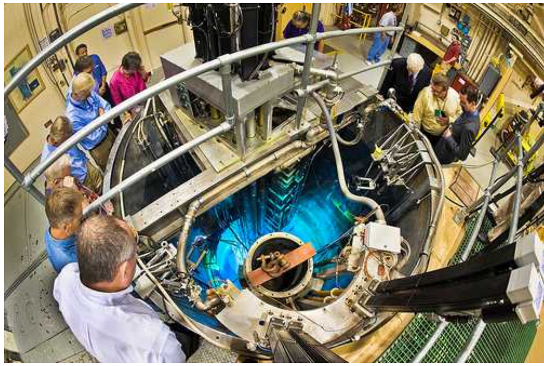
Exceptional service in the national interest

SAND2013-6438C

Sandia National Laboratories





A Tolerance Interval Approach for Physical Simulation Quantification of Margins and Uncertainties

Justin T. Newcomer

August 6, 2013

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

What is QMU and Why Do We Do It

- Quantification of Margins and Uncertainties (QMU) is a framework with applications that include the following:

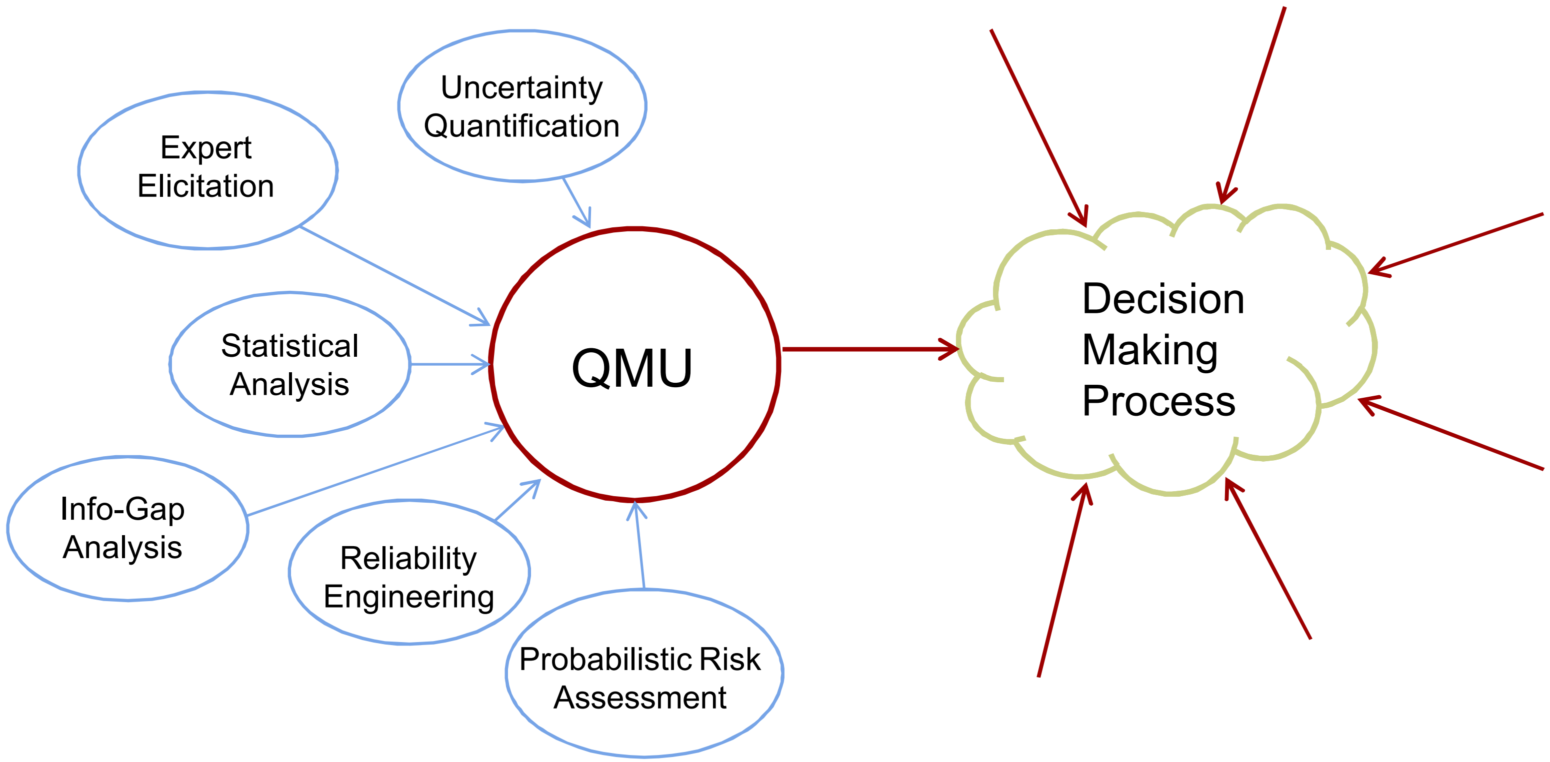
  - **As a measure of vulnerability to change**: By knowing the present margin (M) and uncertainty (U), one has a sense of how vulnerable a particular parameter may be to small changes. A parameter with a large margin relative to its uncertainty can tolerate more change before jeopardizing performance compared to a parameter with a small margin relative to its uncertainty.

  - **As a means of detecting a trend**: A trend analysis on the actual measurements is useful, but perhaps even more meaningful is identification of a trend or change in the amount of margin over time.

  - **As a means of determining a performance impact**: If a credible estimate can be made of a parameter's distribution through QMU and there is a credible understanding of the pass/fail limit relative to meeting performance requirements, then an estimate can be made of the proportion of units that would fail to achieve their required output.

The Role of QMU

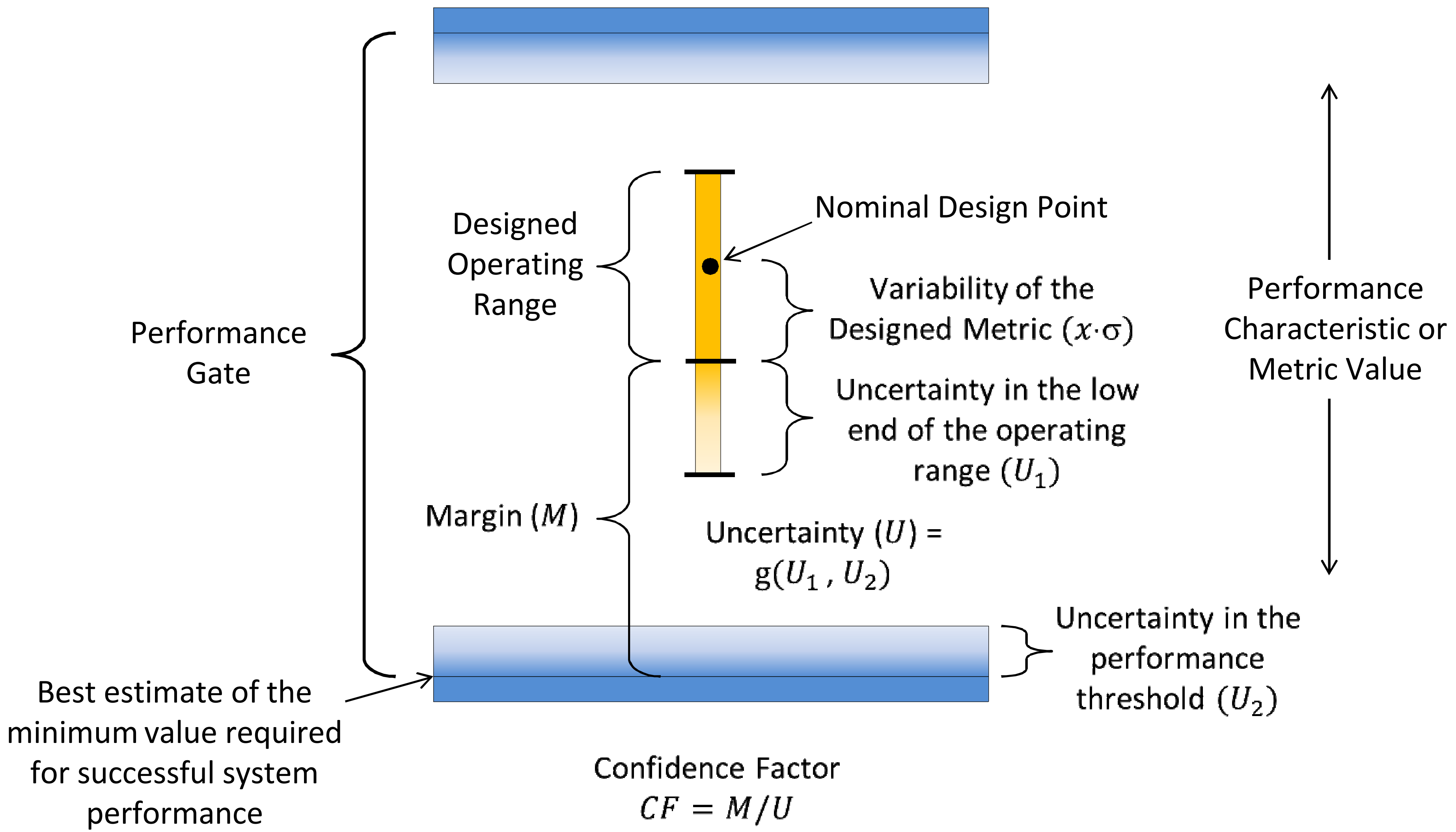
- QMU is a conceptual framework that has evolved over time that outlines a process for ***communicating*** the ***confidence*** in our components and systems.



Key Elements of QMU

1. Specification of **Performance Thresholds**
   - *Performance* is the ability of a system or a component to provide the proper function (e.g., timing, output, response) when exposed to the sequence of design environments and inputs.
2. Identification of associated **Performance Margins**
   - A *performance margin* is the difference between the required performance of a system and the demonstrated performance of a system, with a positive margin indicating that the expected performance exceeds the required performance.
3. Quantification of **Uncertainty** in the performance thresholds and the performance margins as well as in the larger framework of the decisions being contemplated
   - There are two general types of uncertainty that must be separately accounted for, quantified, and aggregated within QMU:
     1. *Aleatory uncertainty* – also called irreducible uncertainty or stochastic variability. Aleatory uncertainty (or variability) is naturally characterized, quantified, and communicated in terms of probability. Common examples are variability in manufacturing processes, material composition, test conditions, and environmental factors, which lead to variability in component or system performance.
     2. *Epistemic uncertainty* – also called reducible uncertainty. This type of uncertainty is due to lack of knowledge or incomplete knowledge. Common examples of epistemic uncertainty are the so-called model form uncertainty (that is, uncertainty in how well the equations in the model capture the physical phenomena of interest), both known and unknown unknowns in scenarios, and poor-quality physical test data.

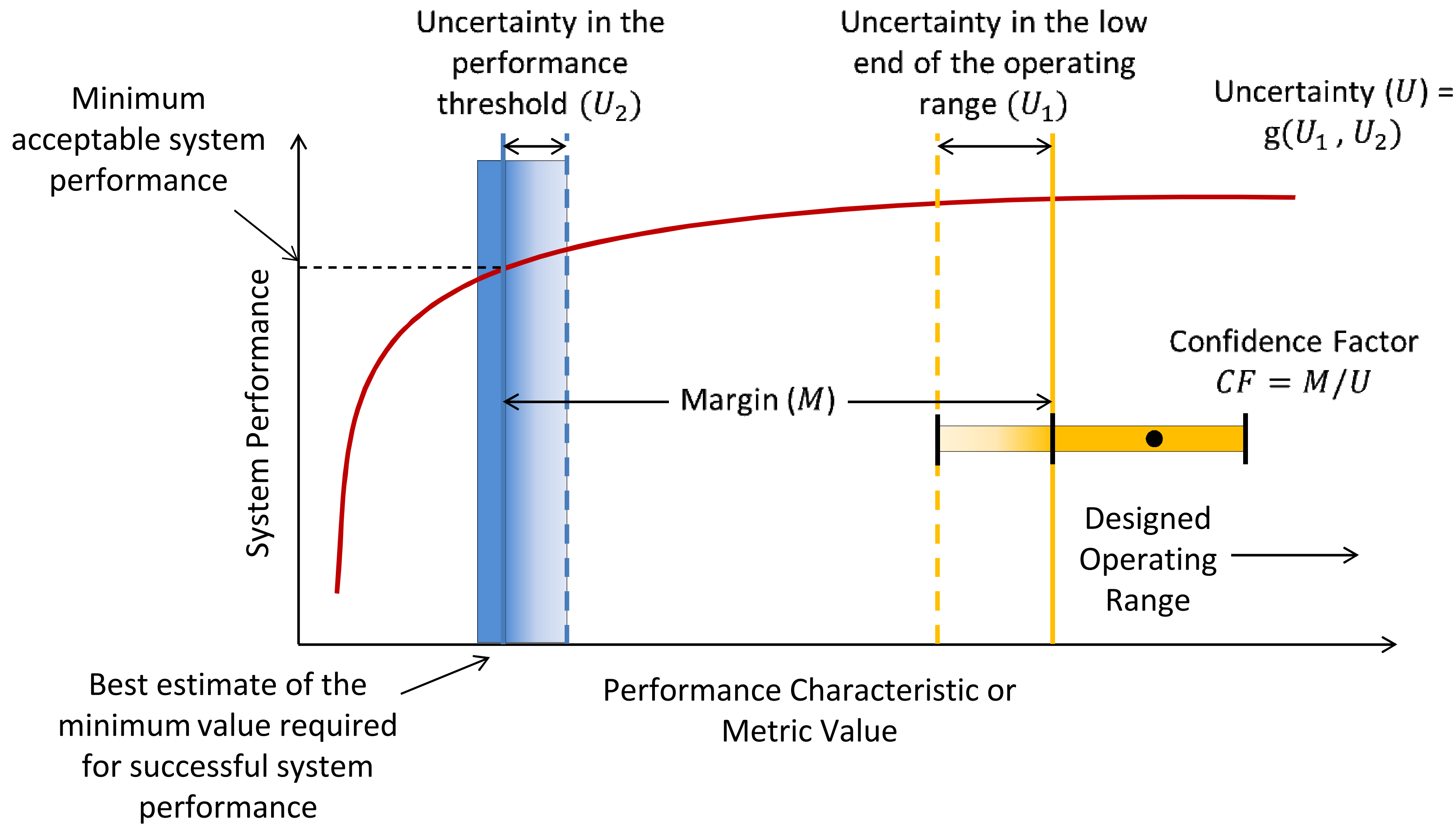
Conceptual Framework



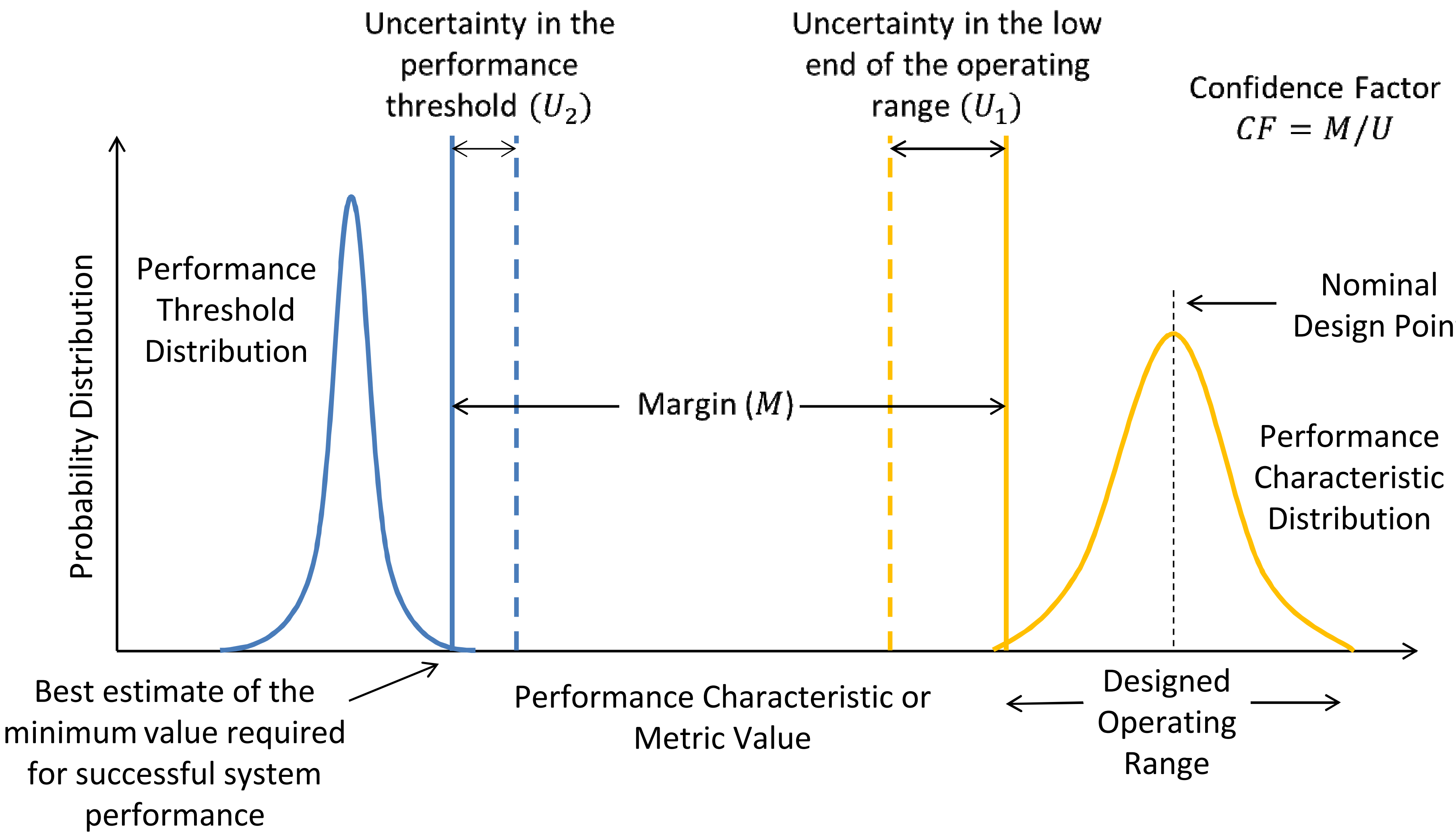
Confidence Factor

$CF = M/U$

Conceptual Framework



Conceptual Framework



Statistical Framework for QMU Questions

- QMU attempts to answer the following types of questions:
  - Are we $YY\%$ certain that **at-least** $XX\%$ of the unit population will yield a response **greater than** the threshold T?
  - Are we $YY\%$ certain that **at-least** $XX\%$ of the unit population will yield a response **greater than** the threshold T after Z years of life?
  - At which age will we no longer be $YY\%$ certain that **at-least** $XX\%$ of the unit population will yield a response **greater** than the threshold T?
- The values of XX, YY, Z, and T and the comparisons '**at-least**' versus '**at-most**' and '**greater than**' versus '**less than**' are all parameters of the requirement.
- Examples:
  - Are we 95% certain that at least 99.5% of the unit population will yield a response greater than the lower performance requirement of 10?
  - Are we 95% certain that at least 99.5% of the unit population will yield a response less than the upper performance requirement of 20?
  - At which age will we no longer be 99% certain that 99.99% of the unit population will yield a response greater than the lower performance requirement of 10?
  - Are we 90% certain that 99% of the unit population will yield a response less than the upper performance requirement of 20 after 35 years of life?

Background Definitions

- **Percentile**: A percentile of a distribution is the value of a variable (here the performance characteristic) below which a certain percent of the values for that variable will fall.
  - We denote a percentile by Q_r, where r represents the probability that a performance characteristic value will fall below Q_r.
  - $Prob(PC < Q_r) = r$.

- **Maximum Allowable Probability of Failure, P_{req}**: Acceptable probability of failure to meet performance requirements (margin failures).

- **Content, p**: The content of a distribution is the proportion of units that are expected to be within the performance requirements (proportion greater than a lower requirement or proportion less than an upper requirement)
  - $p = 1 - P_{req}$

- ***Required Performance***: The $(1-p) \cdot 100^{\text{th}}$ percentile of the performance characteristic distribution, Q_{1-p}, for a lower requirement or the $p \cdot 100^{\text{th}}$ percentile, Q_p, for an upper requirement.
  - For a lower requirement, $Prob(PC < Q_{1-p}) = 1 - p = P_{req}$.
  - For an upper requirement, $Prob(PC > Q_p) = 1 - Prob(PC < Q_p) = 1 - p = P_{req}$.

Background Definitions

- **Observed Performance**: The *estimate* of the chosen percentile and denote this by either $\hat{Q}_{1-p}$ or $\hat{Q}_p$ for a lower or upper percentile respectively
- **Statistical Tolerance Bound**: A one-sided $\gamma \cdot 100\%$ statistical confidence bound on an estimated percentile.
  - For a lower percentile a lower confidence bound is computed, denoted by $\hat{Q}_{1-p,\gamma}$
  - For an upper percentile an upper confidence bound is computed, denoted by $\hat{Q}_{p,\gamma}$
  - Accounts for the sampling uncertainty in the estimate. Formally,
    - $Prob\left(Prob\left(PC < \hat{Q}_{1-p,\gamma}\right) \leq 1 - p\right) \geq \gamma$
    - $Prob\left(Prob\left(PC < \hat{Q}_{p,\gamma}\right) \geq p\right) \geq \gamma$

Inner Probability Characterizes Aleatory Uncertainty

- If $\hat{Q}_{1-p,\gamma} > LPR$ then we are able to claim that $p \cdot 100\%$ of the performance characteristic values will be greater than the lower performance requirement with $\gamma \cdot 100\%$ confidence.
  - A similar statement could be made for an upper requirement however we would require $\hat{Q}_{p,\gamma} < UPR$.
- Therefore, the tolerance bound incorporates information about margin and uncertainty and can be compared directly to the performance requirement to draw conclusions.
  - This is appealing because all decisions remain on the engineering unit scale, which provides an easily interpreted result.

Physical Simulation QMU Metrics

- **Margin** $(\widehat{M})$: The difference between the observed performance (estimated percentile) and the performance requirement.
  - $\widehat{M} = \widehat{Q}_{1-p} - LPR$, for a lower requirement
  - $\widehat{M} = UPR - \widehat{Q}_p$, for an upper requirement
  - Note that in some applications the appropriate metric could be defined as the median Q_{50} or mean. In such cases, the definition of margin would still hold as defined here.

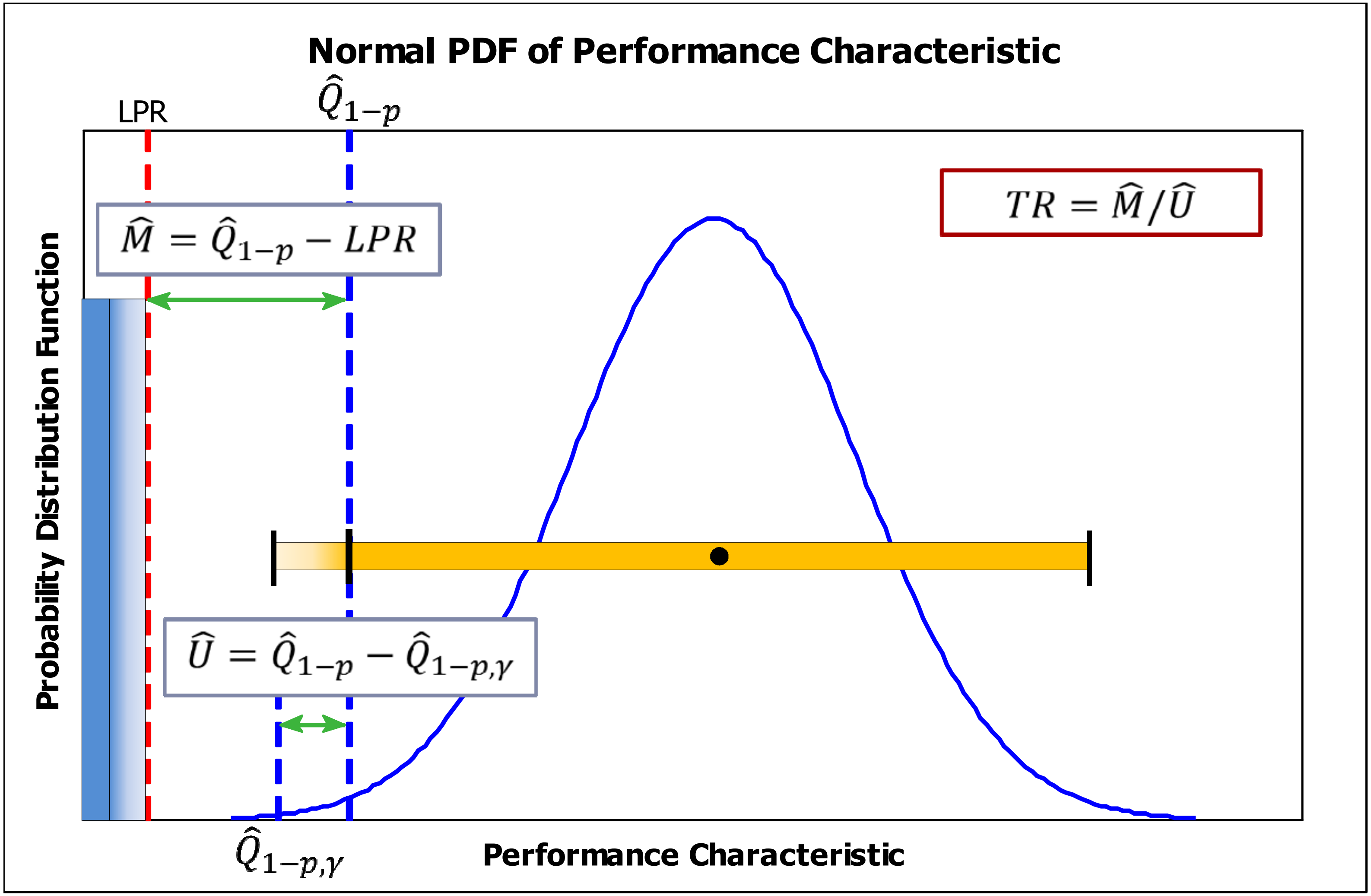
- **Uncertainty** $(\widehat{U})$: The width of the confidence bound on the percentile (absolute difference between the estimated percentile and its confidence bound).
  - $\widehat{U} = \widehat{Q}_{1-p} - \widehat{Q}_{1-p,\gamma}$ for a lower requirement
  - $\widehat{U} = \widehat{Q}_{p,\gamma} - \widehat{Q}_p$ for an upper requirement

- **Tolerance Ratio** (TR): QMU Figure of Merit for Physical Simulation Data. Defined as the ratio of margin divided by uncertainty based on the tolerance bound methodology.
  - $TR = \frac{\widehat{M}}{\widehat{U}} = \frac{\widehat{Q}_{1-p} - LPR}{\widehat{Q}_{1-p} - \widehat{Q}_{1-p,\gamma}}$ or $TR = \frac{UPR - \widehat{Q}_p}{\widehat{Q}_{p,\gamma} - \widehat{Q}_p}$.

- To make decisions based on this margin divided by uncertainty figure-of-merit, we have, if $\widehat{Q}_{1-p,\gamma} > LPR$ then $\widehat{Q}_{1-p} - \widehat{Q}_{1-p,\gamma} < \widehat{Q}_{1-p} - LPR$ indicates we are meeting the requirement and hence

$\boldsymbol{TR = \frac{\widehat{Q}_{1-p} - LPR}{\widehat{Q}_{1-p} - \widehat{Q}_{1-p,\gamma}} > 1}$ is a single consistent decision rule that can be applied!

Tolerance Bound Methodology – Normal Data



Graphical Depiction of the New QMU Metrics Relative to a Normal Distribution

Non-Normal Data

- When the assumption of Normality is violated, there are two common approaches to analyzing the data:

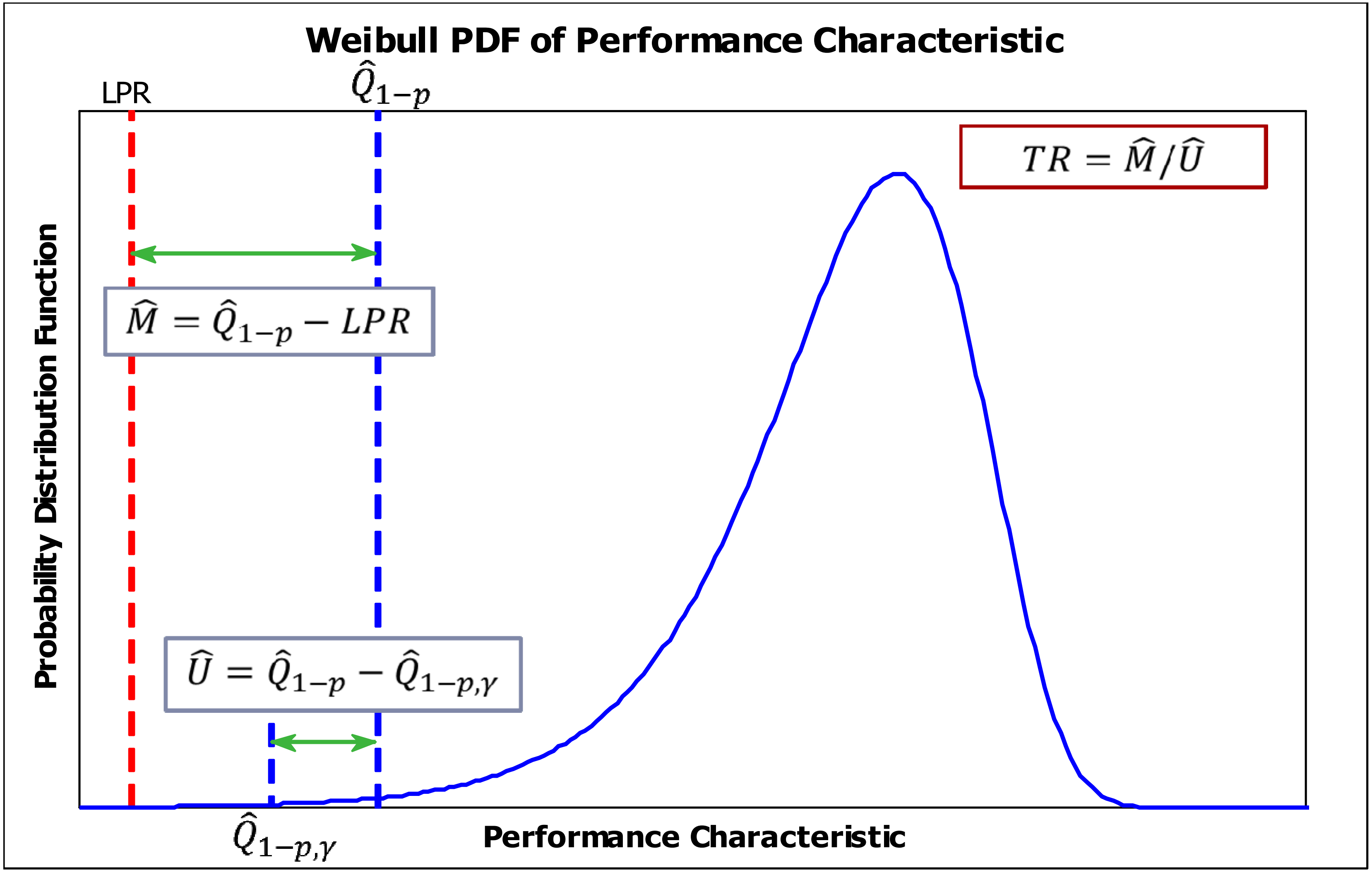
  Transformation Approach

  - Recommended when there is a one-to-one transformation available to transform the data to Normality (at least approximately).
    - Log transformation (Lognormal), Square root transformation, etc.
    - The Gamma distribution has an approximate transformation to Normality
      - $X^{1/3} \sim Normal$

  Direct Parametric Approach

  - Recommended when no one-to-one transformation is available but there is a standard statistical distribution available that fits the data reasonably well.
    - Weibull, Exponential, Extreme-value, etc.

Tolerance Bound Methodology – Non-Normal Data

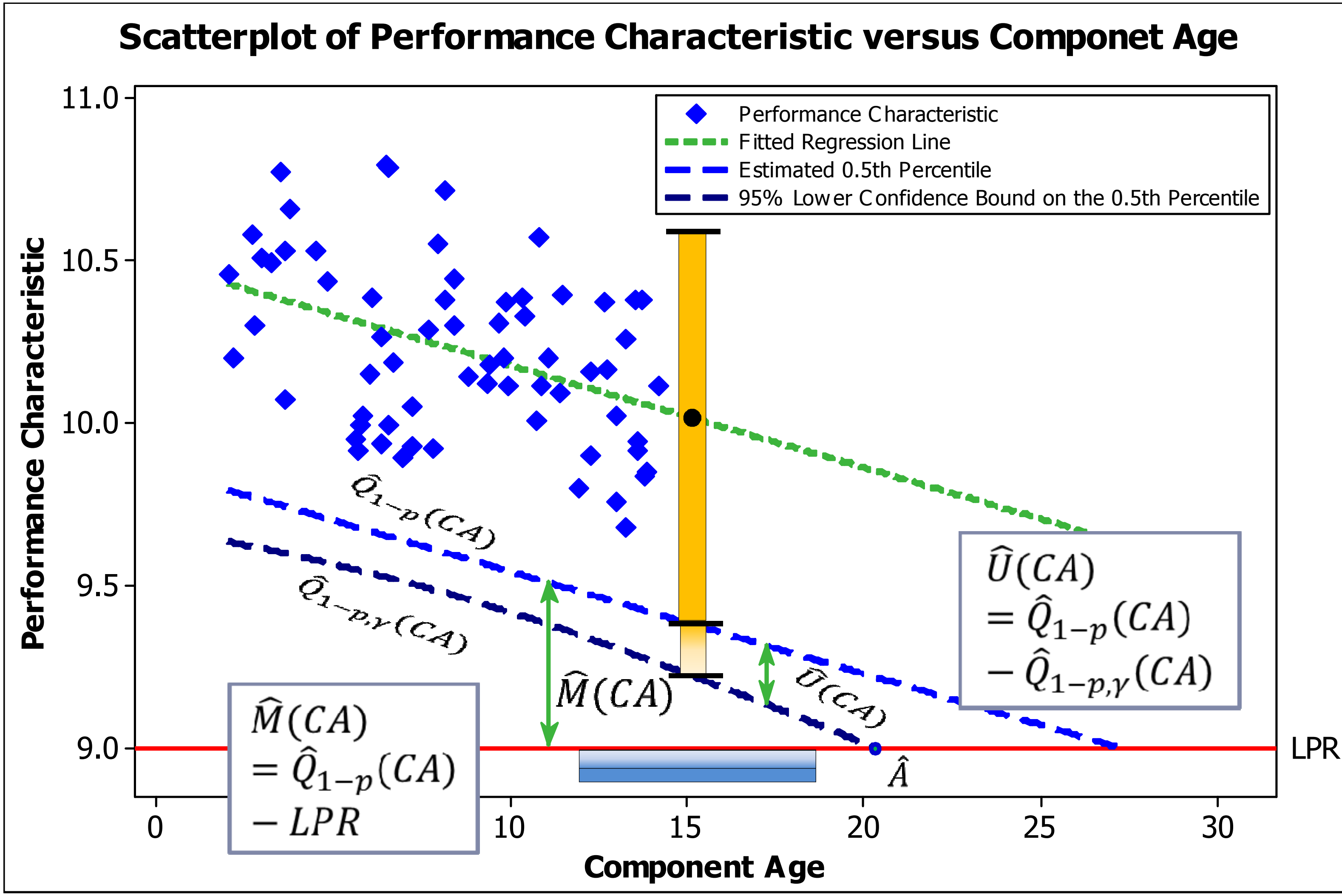


Graphical Depiction of the New Metrics Relative to a Weibull Distributed PC.

QMU with a Trending Performance Characteristic

- When an age trend is present in the data, the goal of the analysis tends to attempt to answer one of the following questions,
  - Are we $\boldsymbol{YY}\%$ certain that **at-least** $\boldsymbol{XX}\%$ of the unit population will yield a response **greater than** the threshold $\boldsymbol{T}$ after $\boldsymbol{Z}$ years of life?
  - At which age will we no longer be $\boldsymbol{YY}\%$ certain that **at-least** $\boldsymbol{XX}\%$ of the unit population will yield a response **greater than** the threshold $\boldsymbol{T}$?
- ***Alarm Age***: The component age at which we estimate certain percentage of the population is no longer contained by the performance requirement, with a given level of confidence.
  - The ***Alarm Age*** is estimated as the component age, $\hat{A}$, that satisfies, $\widehat{M}(\hat{A}) = \widehat{U}(\hat{A})$ and $TR = 1$
  - For ages less than $\hat{A}$ we can claim that we are $\gamma \cdot 100\%$ certain that at least $p \cdot 100\%$ of the units will meet the performance requirement.
  - For ages greater than $\hat{A}$, we can no longer make this statement.

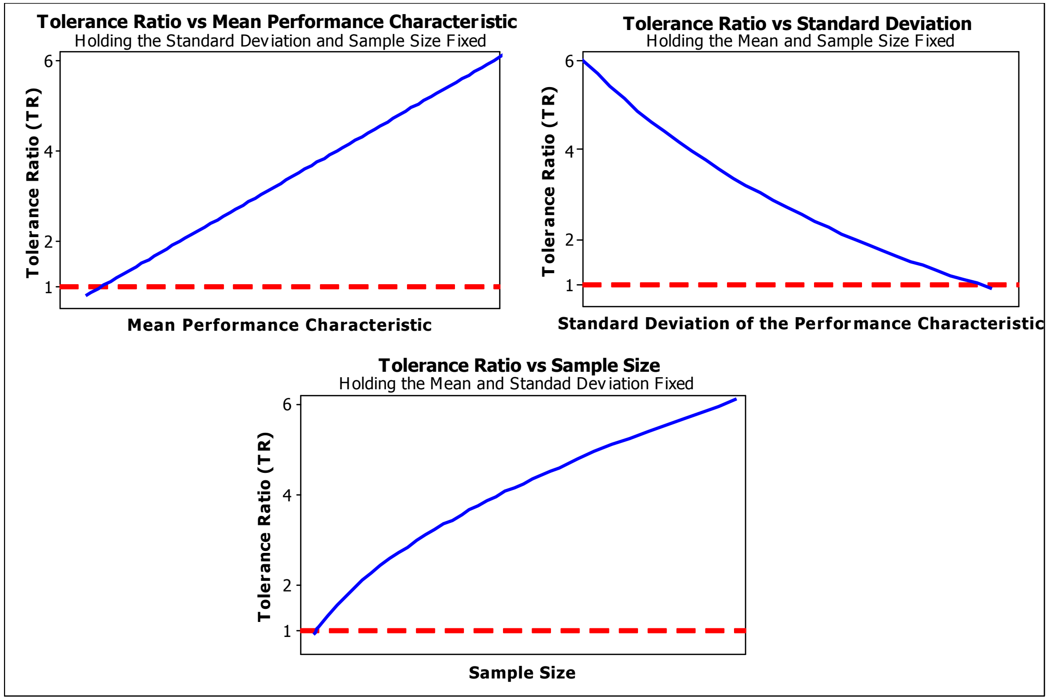
Linear Regression with a Tolerance Bound



Graphical Depiction of the QMU Metrics with a Linear Regression Model

Exploration of the Tolerance Ratio

- The tolerance ratio is a function of the content p and the confidence level γ, however these parameters are specific to the performance characteristic and the requirements, and for any given analysis these will be fixed.
  - With these considered constant, the tolerance ratio, in general, becomes a function of the mean, standard deviation, and the sample size.
- For a fixed content p and confidence level γ, we have
  - $TR(\bar{x}) \propto \bar{x}$,
  - $TR(s) \propto 1/s$,
  - $TR(n) \propto \sqrt{n}$.



The Tolerance Ratio as a Function of the QMU Metrics.

Mean (top left), Standard Deviation (top right), and Sample Size (bottom)

Exploration of the Tolerance Ratio

- Using these relationships, we can explore how to increase the tolerance ratio. The list shown here discusses possible practical ways to increase the tolerance ratio by improving either the component's performance or the amount of information obtained on the component.

  - Increase margin – Improvements in the component performance (possibly through design improvement) that shift the mean of the distribution of the performance characteristic away from the performance requirement will result in a larger tolerance ratio, **provided the unit-to-unit variability does not change significantly**.

  - Decrease uncertainty – Reducing the unit-to-unit variability of the component performance distribution (possibly through improvement or implementation of manufacturing process controls) will result in a larger tolerance ratio, **provided these changes do not shift the mean of the distribution significantly**.

  - Increase precision – Increasing the data acquired on a performance requirement (possibly through increased testing) will result in a larger tolerance ratio, **provided the additional data does not significantly change the estimated mean or standard deviation**.

Tolerance Bound Advantages

- Conclusions are made on the same scale as the original measurements = Interpretability.
- Assumes a specific distribution, but is not limited to the Normal distribution.
  - Any distribution that maps 1:1 with a normal can be obtained easily:
    - Lognormal: $X = \ln(Y)$ is Normal if Y is lognormal.
    - Gamma: $X = Y^{1/3}$ is approximately Normal if Y is gamma.
  - The Exponential and Weibull distributions are derived in Krishnamoorthy and Mathew (2009).
    - Pareto, Power, and Extreme Value are 1:1 transformations with Exponential and Weibull.
- Can be communicated in a QMU framework with a single critical value of 1 which provides a consistent and interpretable decision metric.

The Future of QMU at Sandia (In My Opinion)

- More rigorous engineering analyses and data reviews to better understand what is being tested, what it means, and what the limitations of the available data are.
  - Are there gaps in the current test programs?
  - Are we measuring what's important?
- QMU analyses at the subsystem level that look at interactions between components and performance characteristics.
  - How do we "roll up" analyses on several characteristics?
  - Does the current testing exercise potential failure mechanisms at the interfaces between components?
- Margin testing and analyses to better understand the thresholds for critical performance characteristics.
  - If an analysis indicates low margin relative to a requirement, what does that mean?
  - Do we feel the system will fail if a unit performs outside of its requirements?
  - Do we understand the point at which it will begin to fail?
- More consistent use of methodologies for both physical simulation and computational simulation QMU analyses.
  - Our ability to communicate the results of QMU analyses should be independent of the types of methodologies used!

Useful References for QMU

1. K.V. Diegert, S.E. Klenke, G. Novotny, R.A. Paulsen, M. Pilch, and T.G. Trucano, *Toward a More Rigorous Application of Margins and Uncertainties within the Nuclear Weapons Lifecycle—A Sandia Perspective*, SAND2007-6219, Sandia National Laboratories, Albuquerque, NM, December 2007.

2. J.T. Newcomer, *A New Approach to QMU for Physical Simulation Data*, SAND2012-7912, Sandia National Laboratories, Albuquerque, NM, September 2012

3. M. Pilch, T.G. Trucano, and J.C. Helton, *Ideas Underlying Quantitative Margins and Uncertainty (QMU): A White Paper*, SAND2006-5001, Sandia National Laboratories, Albuquerque, NM, September 2006.

4. J.T. Newcomer, B.M. Rutherford, E.V. Thomas, R.L. Bierbaum, L.M. Hickman, J.W. Lane, S Fitchett, A Urbina, A.A. Robertson, and L.P. Swiler, *Handbook of Statistical Methodologies for QMU*, SAND2012-9603, Sandia National Laboratories, Albuquerque, NM, November 2012

5. Goodwin BT, Juzaitis RJ. *National certification methodology for the nuclear weapons stockpile*, UCRL-TR-223486. Lawrence Livermore National Laboratory, Livermore, CA, 2006.

6. *Special Issue on Quantification of Margins and Uncertainties*, Reliability Engineering and System Safety, Volume 96, Issue 9, Pages 959-1256, Edited by J.C. Helton and M. Pilch, September 2011.

7. K. Krishnamoorthy and T. Mathew, *Statistical Tolerance Regions, Theory, Applications, and Computation*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 2009.

8. W.Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 1998.

9. D.B. Owen, *Factors for One-Sided Tolerance Limits and for Variables Sampling* Plan, Sandia Corporation Monograph SCR-607, Sandia National Laboratories, Albuquerque, NM, March 1963.

10. G.E. Box, W.G. Hunter, J.S. Hunter, *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 2005.

11. R.H. Myers, *Classical and Modern Regression with Applications*, Second Edition, Duxbury Press, 1990.

12. J.K. Patel, C.H, Kapadia, and D.B. Owen, *Handbook of Statistical Distributions*, Marcel Dekker, Inc., 1976.

BACKUPS

Normal Data Analysis

- For a univariate Normal distribution with mean μ and standard deviation σ, the $r \cdot 100^{\text{th}}$ percentile is,
  - $Q_r = \mu + \sigma \cdot \Phi^{-1}(r)$
  - $\Phi^{-1}(r)$ is the $r \cdot 100^{\text{th}}$ percentile of a standard Normal distribution
- The best estimate of this percentile is obtained by replacing the mean and standard deviation with their respective best estimates, $\bar{x}$ and s.
  - $\hat{Q}_r = \bar{x} + s \cdot \Phi^{-1}(r)$.
- For data that follows a Normal distribution the estimated tolerance bound from a sample of size n is of the form,
  - $\hat{Q}_{1-p,\gamma} = \bar{x} - s \cdot k_1$ for a lower tolerance bound
  - $\hat{Q}_{p,\gamma} = \bar{x} + s \cdot k_1$ for an upper tolerance bound
    - $k_1 = t_{n-1,\gamma}(\sqrt{n} \cdot \Phi^{-1}(p))/\sqrt{n}$, where $t_{df,\gamma}(\Delta)$ denotes the $\gamma \cdot 100^{\text{th}}$ percentile of a non-central t-distribution with df degrees of freedom and noncentrality parameter Δ.

Non-Normal Data: Transformation Approach

- Suppose, $y_1, y_2, \dots, y_n$ are a sample from a Lognormal performance characteristic distribution with location and scale parameters μ and σ respectively.
  - $x_1 = \ln(y_1), x_2 = \ln(y_2), \dots, x_n = \ln(y_n)$ is a sample from a Normal distribution with mean and standard deviation μ and σ respectively.
  - Normal based approaches discussed previously can be applied to construct tolerance bounds based on the transformed sample $x_1, x_2, \dots, x_n$.
- Recall, $\hat{Q}_{1-p,\gamma} = \bar{x} - s \cdot k_1$ and $\hat{Q}_{p,\gamma} = \bar{x} + s \cdot k_1$ are lower and upper tolerance bounds respectively for a sample from a Normal distribution as defined before
- Therefore, $\hat{Q}^*_{1-p,\gamma} = e^{\hat{Q}_{1-p,\gamma}} = e^{\bar{x}-s\cdot k_1}$ and $\hat{Q}^*_{p,\gamma} = e^{\hat{Q}_{p,\gamma}} = e^{\bar{x}+s\cdot k_1}$ are lower and upper tolerance bounds respectively from the original Lognormal distribution.
- The calculations and interpretations of M, U, and TR remain on the original engineering unit scale

Non-Normal Data: Direct Parametric Approach

Sandia National Laboratories

- Consider the two-parameter Weibull distribution $f(x) = \frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1} e^{-\left(\frac{x}{\eta}\right)^{\beta}}$
  - No one-to-one transformation with a Normal distribution exists
- The $r \cdot 100^{\text{th}}$ percentile is, $Q_r = \eta \cdot [-\ln(1-r)]^{1/\beta}$ and the best estimate of this percentile, $\hat{Q}_r = \hat{\eta} \cdot [-\ln(1-r)]^{1/\hat{\beta}}$, is given by replacing the shape and scale parameters with their respective maximum likelihood estimates, $\hat{\beta}$ and $\hat{\eta}$.
- The estimated lower tolerance bound from a sample of size n is of the form $\hat{Q}_{1-p,\gamma} = \hat{\eta} \cdot \exp\left(w_{1-p,1-\gamma}/\hat{\beta}\right)$ where $w_{1-p,1-\gamma}$ is the $(1-\gamma) \cdot 100^{\text{th}}$ percentile of $w_{1-p} = \beta^* \cdot [-\ln(\eta^*) + \ln\{-\ln(1-(1-p))\}]$.
- For an upper tolerance bound, $\hat{Q}_{p,\gamma} = \hat{\eta} \cdot \exp\left(w_{p,\gamma}/\hat{\beta}\right)$ where $w_{p,\gamma}$ is the $\gamma \cdot 100^{\text{th}}$ percentile of $w_p = \beta^* \cdot [-\ln(\eta^*) + \ln(-\ln(1-p))]$.
  - β^* and η^* are the maximum likelihood estimates calculated from a sample of size n from a standard Weibull $(\beta = 1, \eta = 1)$ distribution.
  - The distribution of w does not depend on any unknown parameters, and so its percentiles can be estimated using Monte Carlo simulation.
  - See Krishnamoorthy and Mathew (2009) for complete details and algorithm.
- Again, the calculations and interpretations of M, U, and TR remain on the original engineering unit scale

Regression with a Tolerance Bound

- A simple linear regression model is given by,
  - $PC_i = \beta_0 + \beta_1 \cdot CA_i + \varepsilon_i,\ i = 1,2,\ldots n.$
  - PC_i is the performance characteristic and CA_i is the age of the i^{th} unit
  - β_0 and β_1 are the regression model parameters to be estimated
  - ε_i is a random error assumed to follow a Normal distribution with mean zero and standard deviation σ_R.

- For a regression analysis assuming the random error follows a Normal distribution with mean zero and standard deviation σ_R, the $r \cdot 100^{\text{th}}$ percentile is a function of the component age and is defined to be,
  - $\hat{Q}_r(CA) = \hat{\beta}_0 + \hat{\beta}_1 \cdot CA + \hat{\sigma}_R \cdot \Phi^{-1}(r)$
    - $\hat{\beta}_0$, $\hat{\beta}_0$, and $\hat{\sigma}_R$ are obtained via maximum likelihood
    - $\Phi^{-1}(r)$ is the $r \cdot 100^{\text{th}}$ percentile of a standard Normal distribution

- For a regression with a random error that follows a Normal distribution, the estimated tolerance bound from a sample of size n is of the form,
  - $\hat{Q}_{1-p,\gamma}(CA) = \hat{\beta}_0 + \hat{\beta}_1 \cdot CA - \hat{\sigma}_R \cdot k_1(CA)$, for a lower tolerance bound
  - $\hat{Q}_{p,\gamma}(CA) = \hat{\beta}_0 + \hat{\beta}_1 \cdot CA + \hat{\sigma}_R \cdot k_1(CA)$, for an upper tolerance bound
    - $k_1 = d(CA) \cdot t_{n-2,\gamma}\left(\frac{\Phi^{-1}(p)}{d(CA)}\right)$
    - $d^2(CA) = \frac{1}{n} + \frac{(CA-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$, x_i is the ith component age, $\bar{x}$ is the average component age
    - $t_{df,\gamma}(\Delta)$ denotes the $\gamma \cdot 100^{\text{th}}$ percentile of a non-central t-distribution with df degrees of freedom and noncentrality parameter Δ.

Regression with a Tolerance Bound

- The estimates of Margin, Uncertainty, and the Tolerance Ratio remain the same as in the previous examples, however here they are a function of the component age (CA)
  - $\widehat{M}(CA) = \widehat{Q}_{1-p}(CA) - LPR$, $\widehat{U}(CA) = \widehat{Q}_{1-p}(CA) - \widehat{Q}_{1-p,\gamma}(CA)$, and $TR = \frac{\widehat{M}(CA)}{\widehat{U}(CA)} = \frac{\widehat{Q}_{1-p}(CA) - LPR}{\widehat{Q}_{1-p}(CA) - \widehat{Q}_{1-p,\gamma}(CA)}$, for a lower requirement
  - $\widehat{M}(CA) = UPR - \widehat{Q}_p(CA)$, $\widehat{U}(CA) = \widehat{Q}_{p,\gamma}(CA) - \widehat{Q}_p(CA)$, and $TR = \frac{UPR - \widehat{Q}_p(CA)}{\widehat{Q}_{p,\gamma}(CA) - \widehat{Q}_p(CA)}$, for an upper requirement

- The ***Alarm Age*** is estimated as the component age, $\hat{A}$, that satisfies, $\widehat{M}(\hat{A}) = \widehat{U}(\hat{A})$ and $TR = 1$
  - For ages less than $\hat{A}$ we can claim that we are $\gamma \cdot 100\%$ certain that at least $p \cdot 100\%$ of the units will meet the performance requirement.
  - For ages greater than $\hat{A}$, we can no longer make this statement.