SAND2011-2204C

Some Statistical Procedures to Refine Estimates of Uncertainty when Sparse Data are Available for Model Validation and Calibration[1]

Vicente Romero[2], Brian Rutherford, Justin Newcomer
Sandia National Laboratories,[3] Albuquerque, NM

Abstract

This paper presents some statistical concepts and techniques for refining the expression of uncertainty arising from: a) random variability (aleatory uncertainty) of a random quantity; and b) contributed epistemic uncertainty due to limited sampling of the random quantity. The treatment is tailored to handling experimental uncertainty in a context of model validation and calibration. Two particular problems are considered. One involves deconvolving random measurement error from measured random response. The other involves exploiting a relationship between two random variates of a system and an independently characterized probability density of one of the variates.

I. Introduction

Model validation involves a comparison between model predictions and experimental responses to assess whether or not the computational model is "valid" for the application (see e.g. the extensive survey and review [1]). Part of the challenge in these comparisons is the appropriate characterization of the various sources of experimental variability and uncertainty in the response data. Several complicating factors may be present: experiment-to-experiment boundary condition variability and/or systematic uncertainty, and measurement error variability and/or systematic uncertainty, may contribute uncertainty regarding output responses. References [2] and [3] discusses the implications and treatment of systematic uncertainty in a context of model validation and calibration. This paper concentrates on the random component of uncertainty encountered in multiple repeat tests.

Whether one or several units or systems ("devices") are tested in multiple repeat tests, portions of the response variance may be attributable to characterized sources of experimental randomness that are not affiliated with the tested device/s, while the remainder is attributed to variance of the tested devices/s. It is the variance attributed to the device/s being modeled that is of lasting importance in model validation or calibration and should be variously embodied in the model as it "travels" ([1], [2]) to downstream prediction use beyond the validation or calibration activity. Limited repeat tests in experimental campaigns introduce epistemic uncertainty to the identified device variability. This uncertainty shows up as epistemic uncertainty on the statistical parameters and/or form of the uncertainty distribution associated with the measured response variability.

Hence, 1) the epistemic uncertainty in the response distribution due to sparse sampling be accounted for; and 2) sources of measurement error variability and experiment-to-experiment test condition variability should be characterized in the tests or in separate tests so that their effects can be separated or deconvolved from the stochastic variability of the measured response of the tested device/s (thus isolating the device response variability for comparison against computational predictions).

---

[2] AIAA Senior Member, corresponding author: vjromer@sandia.gov
[3] Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Accordingly, this paper presents some statistical concepts and techniques for refining the expression of uncertainty for model validation and calibration when random variability in repeat experiments is involved. Two particular problems are considered. One problem involves the decomposition or deconvolution of response variability in a set of data where measurement error variability is "confounded" (indistinguishable based on the response measurements) with the device-to- device variability in the set of repeat experiments, but a good independent characterization of the measurement error variability exists. Then the contribution of measurement error can be deconvolved from the response data, yielding a sharpened estimate of variability of device response (which is the variability sought for comparison with predictions). When a large number of experimental repeats are involved, the well characterized variance of the measurement error can be subtracted from the variance of the measured data leaving a reasonable estimate of the device variance. However, when few experiments are involved, this simple “variance-subtraction” approach does not acknowledge the potentially large epistemic uncertainty due to the limited number of experiments. A simple approach is described in Section III of this paper to address this “deconvolution problem” in the common situation where only a few repeat experiments are available or can be conducted.

A second problem involves sharpening the estimate of the stochastic variability of a sparsely tested device by exploiting a causal relationship between a governing device attribute that varies randomly in the experiments and the related variability of device response. If the probability distribution of the governing attribute is well characterized through a large number of separate experiments, then this knowledge can be combined with the causal relationship to refine the estimate of device variability from the few experiments. The methodology, demonstrated in Section IV, uses the results of the few experiments and standard methods for variance component estimation and regression analysis to estimate the causal attribute/response relationship. Then a parametric bootstrap approach ([4]) is used to derive estimates of the true variability of device performance.

We start by providing essential starting background next in Section II. We close with Section V.

II. Some elementary background: Limiting cases of all measured variability caused by device variability or all caused by measurement error variability

Let the measured response from a set of repeat experiments be governed by a Normal distribution or PDF (probability density function). For simplicity, let the normal PDF have a mean μ = zero and a standard deviation $\sigma = 1/1.96$ such that the $\mu \pm 1.96\sigma$ extents of the PDFs, which mark its 0.025 to 0.975 percentiles, have values of -1 and 1. Figure 1 shows three sets of 8 random samples drawn from the PDF.

It is common in engineering to fit experimental data with a normal distribution. The plot at the top of Figure 2 shows 20 uncertainty intervals defined by $\mu_i \pm 1.96\sigma_i$ from normal PDFs fit to the three sample data sets shown in Figure 1 and 17 other 8-sample data sets. (The normal fits are obtained by forming a normal PDF_i with mean μ_i and standard deviation σ_i calculated from the *i*th set of 8 samples.) The intervals mark the 0.025 and 0.975 percentiles of the normal PDFs. In other words, the intervals show the central 95% probability ranges of the fitted normal PDFs. The interval ranges can be compared to the displayed (-1,1) central 95% probability range of the exact normal PDF from which the sample sets were drawn. Only four of the 20 intervals contain the true 95% range of the exact PDF. A desirable outcome when fitting random-variable data is that the fit be somewhat ***conservative*** so as to bound a desired large percentage of the actual PDF, say 95% included probability, with reasonable reliability. This does not occur in the figure.

A broader study in [5] finds that of 3000 such trials only 25.6% of computed 95% intervals contained the true 95% interval of the sampled Normal PDF. A mean shortfall error of 28% of the true 95% range occurred. When only 2 samples are drawn from the normal PDF (instead of 8), only 20.7% of 3000 computed intervals contained the true 95% interval of the sampled PDF. A mean shortfall error of 63% was found. Even for the relatively large number of 32 samples per trial, only 27.8% computed intervals contain the true 95% interval of the sampled PDF, but a relatively small mean shortfall error of 14% occurred.

Hence, the common practice of fitting data with a normal PDF is not very reliable even if the underlying random process being sampled is Normal. In general, when relatively few samples are obtained from a source of random variability, a substantial amount of epistemic uncertainty exists in addition to the aleatory variability information in the samples from the random quantity. This epistemic uncertainty undermines accurate estimation of the source variability function or PDF. Indeed, in [5] the goal was not to pursue accurate PDF estimation from sparse data (an impossible goal to reach), but instead to identify uncertainty representation methodologies that are some conservative in that they reliably bound a desired range (95% included probability) of the sampled PDF with reasonable reliability. A second, opposing objective was that the representation not be overly conservative; that it minimally over-estimate the said exact range. The presence of the two opposing objectives makes the sparse-data uncertainty representation problem an interesting and difficult one.

It was provisionally found in [5], based on a very simple zeroth-order performance ranking scheme and an initial set of tests, that the classical approach of using statistical tolerance intervals [6] computed from sparse random data provides reliably conservative interval representation for the combined epistemic and aleatory uncertainty associated with the limited data. Furthermore, the Tolerance Interval method is simple to implement and use, as explained at the end of this section. However, a drawback is that the tolerance intervals often substantially overestimate the target percentile range of the sampled PDF at low numbers of samples (see [5]). For eight samples the overestimation is not egregious. The middle plot in Figure 1 shows calculated 90% confidence / 95% coverage (0.9/0.95) tolerance intervals for the same 20 sets of 8 samples that were used in the top plot for the Normal-Fitting method. Sixteen of the twenty tolerance intervals contain the true 95% interval of the sampled Normal PDF. Of the 3000 trials in [5], 84.4% of the computed tolerance intervals contained the true 95% interval—close to the advertised 90% reliability (confidence) of the 0.9/0.95 tolerance intervals.

The context of the above discussion is that the sampled Normal PDF ($\mu = 0$, $\sigma = 1/1.96$) represents stochastic behavior of the tested device/s in the repeat experiments. The stochastic variation of the response samples are assumed to be measured with zero error. Next we consider an opposite limiting case where the variability of sample values arises purely from random measurement errors in the experiments. In section III we consider the mixed case where the randomness in the samples comes from both sources: a) stochastic device response in the experiments; and b) random error in the measurements of response.

For the opposite limiting case just mentioned, let all the variability in measured device response be due to random measurement error. Let the measurement error be governed by the same Normal PDF ($\mu = 0$, $\sigma = 1/1.96$) as discussed above. Figure 1 shows three sets of 8 random samples drawn from this PDF. In the new context the sample values are instances of measurement errors. The goal is to determine, from the sample data, the actual value of the quantity being measured (i.e., the value of the ***measurand***). Let the measurement errors be approximately symmetrically distributed about the actual value of the measurand (i.e. the mean of the distribution from which the random errors arise approximately coincides with the actual value of the measurand). Then the problem becomes one of estimating the mean of the error distribution and taking this for the value of the measurand.

If the errors are approximately normally distributed, then the mean of the distribution can be obtained within an uncertainty determined from use of Student's t-distribution ([6]). The mean μ of the error distribution is said to lie within a distance or "confidence interval" of $t_{x\%,\nu}\sigma_i n^{-1/2}$ from the calculated mean μ_i of the sample data. The subscript ν signifies the "degrees of freedom" in the problem, equal to the number of samples minus one, $n - 1$, and the x% subscript designates the desired confidence level or odds of assurance that the true mean lies within the said distance of the sample mean μ_i. Desiring in this paper a confidence level of 95%, the value of $t_{95\%,\nu=7} = 2.365$ (see Table 1 and associated text at end of this section). The associated 95% confidence intervals defined by

$$\mu_i \pm t_{95\%,\nu}\sigma_i n^{-1/2} \qquad (1)$$

are shown in the lowest plot in Figure 2. It is seen that the 19th and 20th mean estimates with their uncertainty bars or intervals do not contain the true mean $\mu = 0$ of the sampled normal PDF. Hence, the advertised 95% confidence intervals only worked in 90% of the 20 trials shown. In 1000 trials run for the

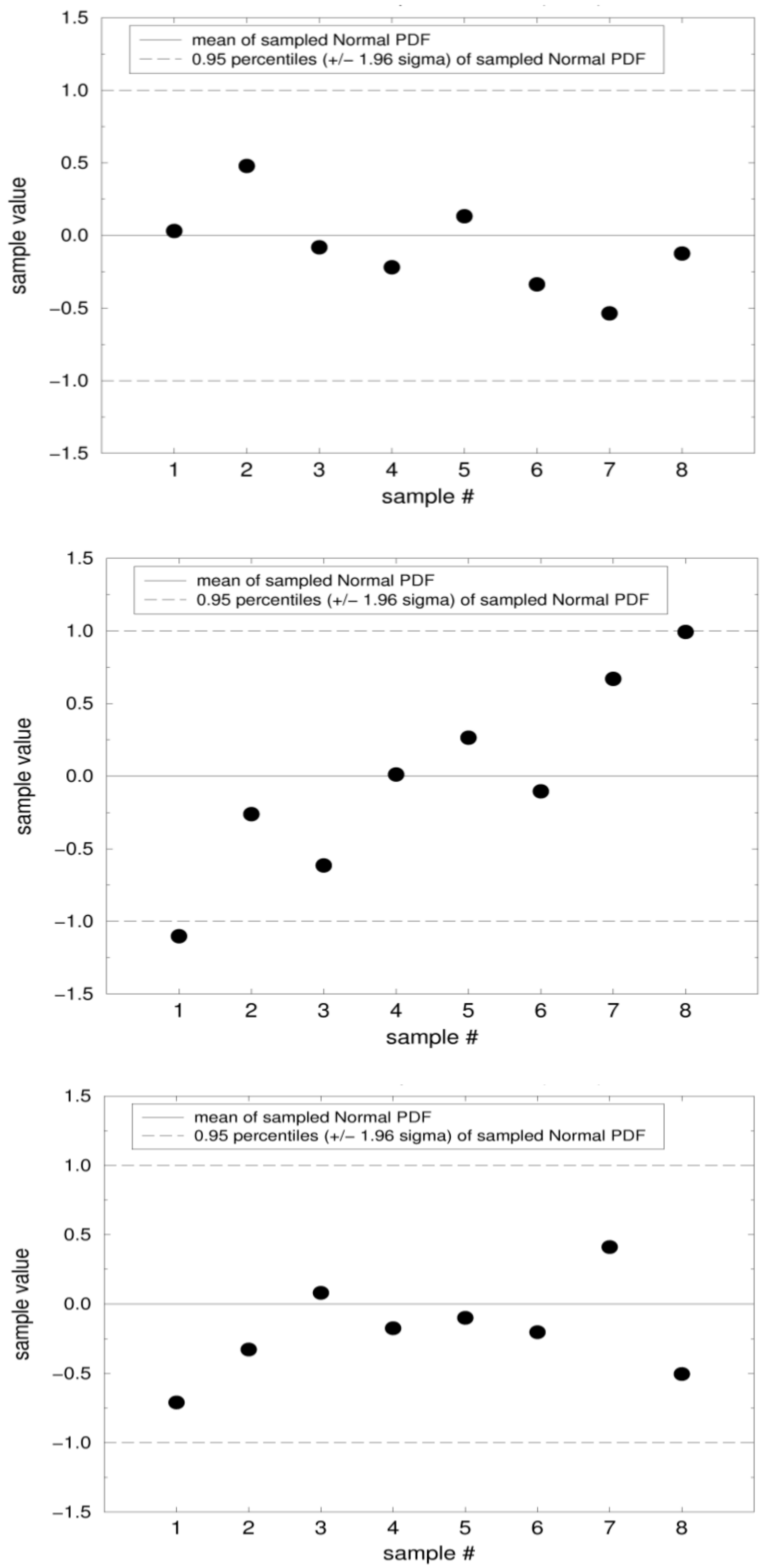

Figure 1. **Three sets of 8 random samples from Normal PDF described in body of paper.**

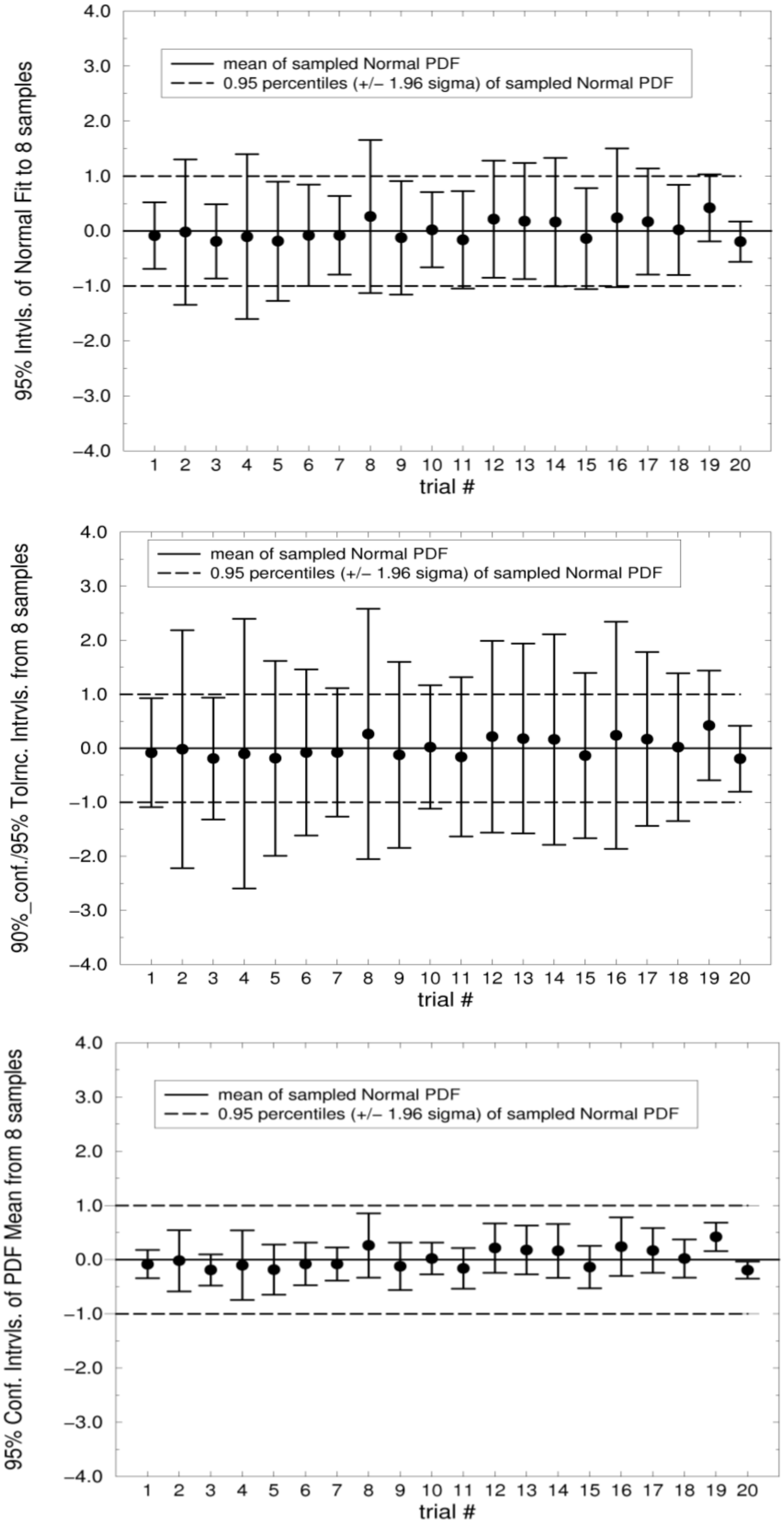

Figure 2. Various computed uncertainty intervals (labeled on plot ordinates) from 20 sets each consisting of 8 random samples of Normal PDF. Various interval are defined in body.

purposes of this paper the calculated 95% confidence intervals worked to their advertised level, in which 949 or 94.9% of the calculated intervals did bound the true mean.

The lengths of the confidence and tolerance intervals in Figure 2 are given by multiplication factors that scale the calculated standard deviation σ_i of a set of random data. The multiplication factors, and thus the relative lengths of the confidence and tolerance intervals, are listed in Table 1 and plotted in Figure 3 according to the number of data samples available. For the purposes here we use 95% confidence intervals and tolerance intervals with a 95% percentile range coverage at 90% confidence or assurance odds. The confidence intervals are calculated from equation 1 where the multiplication factor $t_{95\%,\nu}n^{-1/2}$ is given in Table 1 and plotted in Figure 3. Values of $t_{x\%,\nu}$ are found in standard confidence interval tables, e.g. [6] and [7]. The tolerance intervals are defined by

$$\mu_i \pm f_{0.9/0.95}\sigma_i \tag{2}$$

where the multiplication factor $f_{0.9/0.95}$ comes from standard tolerance interval tables, e.g. [6]. This factor is listed in Table 1 and plotted in Figure 2.

The figure and table reveal that the confidence and tolerance intervals are very large at n=2 samples but quickly decrease in size when more samples are available. A knee in the rate of uncertainty decrease (per added sample) occurs somewhere between 4 to 6 samples, with the rate of decrease being fairly small after n=8 samples. At n=2 samples the confidence intervals are about half the extent of the tolerance intervals. The relative extent decreases to about 1/4 at n=8 samples and to little more than 1/10 at 40 samples, as Table 2 shows and Figure 4 plots. As might be expected, the uncertainty range (confidence interval) for the estimated mean of the sampled normal PDF is considerably less and decreases at a faster rate than the tolerance interval for the uncertainty about the estimated 95 percentile range of the sampled PDF. The tolerance interval has an asymptotic standard-deviation multiplier value of 1.96 for an infinite number of samples. This corresponds to the exact 95 percentile range of the sampled normal PDF. The confidence interval has an asymptotic multiplier value of zero, consistent with zero estimation error (an exact estimate) for the PDF mean when infinite number of samples are used.

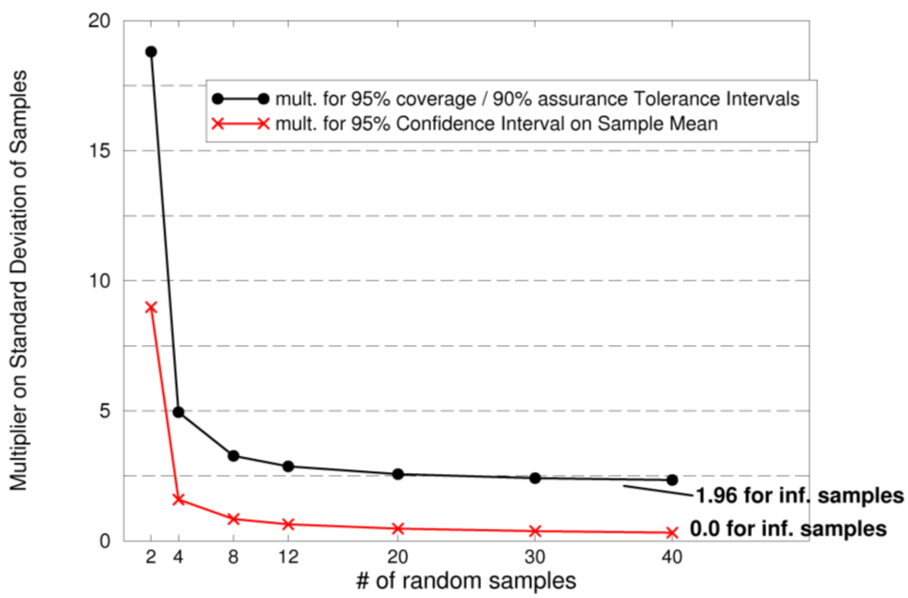

Figure 3. Multiplier on calculated standard deviation used to form uncertainty interval ranges for confidence intervals and tolerance intervals.

Table 1. Standard deviation multipliers for 95% Confidence Intervals and Tolerance Intervals of 95% coverage at 90% assurance

# samples	$t_{95\%,\nu}$	$t_{95\%,\nu}/\sqrt{n}$	$f_{0.9/0.95}$
2	12.71	8.98	18.80
4	3.18	1.59	4.94
8	2.37	0.84	3.26
12	2.20	0.64	2.86
20	2.09	0.47	2.56
30	2.05	0.37	2.41
40	2.02	0.32	2.33
∞	1.96	0	1.96

III. Variability Reduction by Deconvolution for Mixed cases having both Device Variability and Measurement Error Variability

The previous section defined the limiting cases of uncertainty associated with a set of measured random data. At one extreme where all the variability is caused by random measurement error, confidence intervals are used to provide an estimated uncertainty range within which the true value of the measured quantity is expected to lie with expectation odds or confidence that the user sets. At the other extreme, no random measurement error exists and all the measured variability is caused by randomness in the device behavior in the repeat experiments. Then tolerance intervals are used to provide an estimated uncertainty range within which a specified percentage of the entire population of random values lies (only a few samples of which were measured).

In this section we examine cases that lie between the two extremes, where the variability in measured results is due to random measurement error and also to variability of the quantity being measured. In particular, we consider the situation where the PDF of measurement error is well characterized by the measurement instrumentation manufacturer or by calibration activities from the experimentalists. In this case, variability can be subtracted off the data measurements using the following simple technique.

For demonstration, consider a case where the variability in the measured data comes from equal sources of Normal variability in measurement error and in the quantity being measured. Let the model problem have a combined variance σ^2 of 1.0 for convenience. Then the normal PDF to be sampled has a standard deviation of $\sigma = 1.0$, and let its mean μ be equal to zero. By the rule that the variances of the contributing random sources sum to the variance of the resultant PDF, and noting that the resulting PDF has a variance of 1.0 and the two contributing sources have equal variances, the following is obtained: $\sigma_{meas.\_error} = \sigma_{device.\_variability} = 0.707$.

The deconvolution problem is posed such that the variability of the combined PDF is not known a priori; information on this PDF comes from experimental measurements (samples). However, the PDF of random measurement error contributed to the measurement *is* known a priori. This latter information can be used to refine the estimate of true variability of the quantity being measured, which is the variability that is important to characterize in the experiments (and for model validation and calibration purposes if applicable).

A simple way to approach the deconvolution problem is to form tolerance intervals for the measured data and then subtract from these a variance contribution due to the measurement error. The tolerance intervals as we have seen are fairly conservative, especially for sparse data. Therefore, any fair scheme to reduce the extent of these intervals based on knowledge independently gained (not from the current experiments) should be taken advantage of. In the present case, a classical result from [8] is used to subtract, from the 95% percentile range represented by the 0.9/0.95 tolerance intervals, the variance contribution from measurement error (represented by the 95% percentile range of the measurement error PDF, here = $2*1.96\sigma_{meas.\_error}$). The subtraction operation is as follows:

$$\textit{reduced 95\% uncertainy interval} = \mu_i \pm \Delta_i\ , \quad \Delta_i = [(f_{0.9/0.95}\sigma_i)^2 - (1.96\sigma_{meas.\_error})^2]^{1/2}\ . \qquad (3)$$

This relationship is applied to the model problem above, having equal variability sources of the quantity being measured and of the measurement error, where $\sigma_{meas.\_error}$ = 0.707. For illustration in the following, also let the sample standard deviation σ_i be the value of the actual standard deviation σ = 1.0 of the sampled PDF. Then Table 2 and Figure 4 give the variance reductions versus number of experimental samples. The reduced variance is presented in terms of fraction of the full tolerance-interval range that would exist if subtraction of the measurement error variability is not done.

The plot shows that even if the measurement error variability constitutes half the contributed variability in the measured data, the effect of subtracting this out is swamped by the epistemic uncertainty coming from low numbers of samples. As the number of samples increases, the epistemic uncertainty and hence the tolerance interval size quickly decreases, causing the subtracted measurement variability to have a greater proportionate effect on total uncertainty. The uncertainty reduction reaches a value of about 10% reduction at 8 samples, nearly 20% reduction at 30 samples, with an asymptotic reduction of about 30% for infinite sampling. If the problem is changed so that the variability (standard deviation) of measurement errors is only 1/10 the actual variability of quantity being measured, then the achievable reduction is less than 1% even at the large number of 40 samples.

The results here are somewhat sobering. While deconvolving measurement errors can help to reduce uncertainty, the gains to be made appear relatively minimal unless a relatively large number of tests can be run. It may be that more sophisticated deconvolution methods can shave off more uncertainty, but this has not been explored by the authors yet.

Table 2. Relative sizes of listed uncertainty intervals as a fraction of size of 0.9/0.95 Tolerance Intervals (95% coverage at 90% assurance)

# samples	reduced tol. intvl. by deconvolved 50% source of rand. meas. error	$(t_{95\%,\nu}/\sqrt{n})/f_{0.9/0.95}$
2	0.997	0.478
4	0.960	0.322
8	0.905	0.256
12	0.875	0.220
20	0.841	0.183
30	0.819	0.155
40	0.805	0.137
∞	0.707	0.0

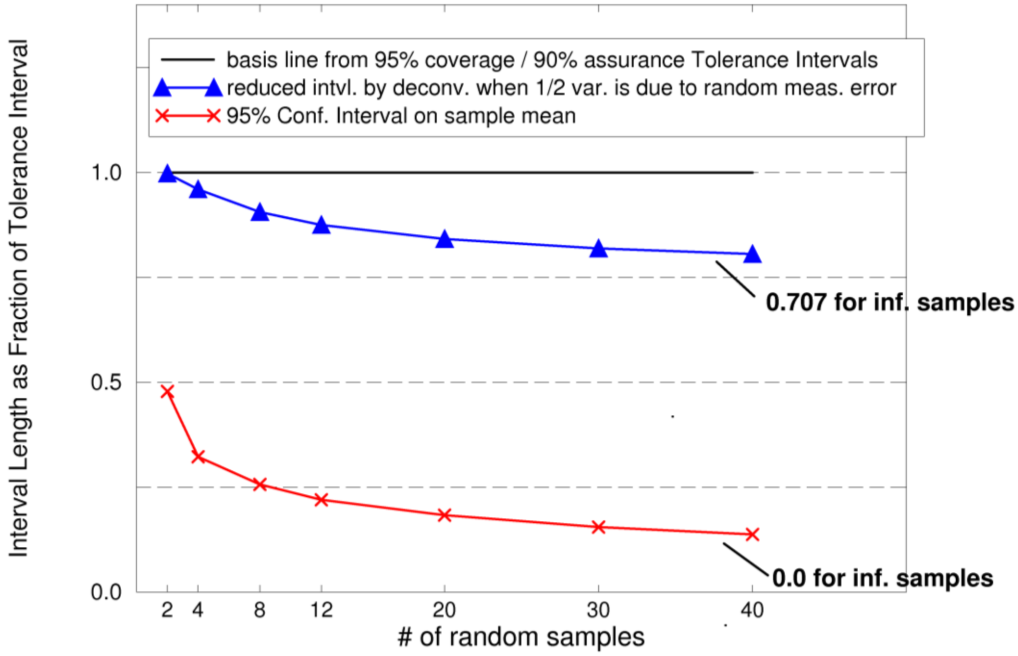

Figure 4. Plot of attainable variance reduction by subtracting out well-characterized random measurement error from a set of measurements.

IV. Refinement of variability estimates using a behavioral relationship between two variates and independently characterized PDF of one of the variates

The methodology presented in this section was motivated by an application where a small number of hardware devices were stress tested to destructive levels (only a small number of devices could be sacrificed for this testing). The devices came from a very large population of devices (nearly 200) whose performance in normal environments was well characterized. By comparing the levels of performance of the stressed devices with their normal-environment performance it was found that a relatively strong correlation existed between their performance in the two environments. This relationship was used, along with the well-characterized PDF of device performance in normal environments, to provide estimates of the stress-performance variability of the full population of devices. These estimates were substantially more precise than could have been obtained by calculating tolerance intervals of device performance from the responses of the few stress tested devices. The methodology and the resulting precision improvements are illustrated here using data from two related variates in Example 12.5 from [7]. Table 4 presents the data from the example. The data is plotted in Figure 5.

Table 3. Table of related variates from example 12.5 in Ref. [7].

X value	Y value
78	850
75	775
78	750
81	975
84	915
86	1015
87	1037

This example was established to illustrate two points. First, it can be extremely advantageous to make use of independent variables were appropriate in the construction of tolerance bounds. Second, if we have an independent assessment of the measurement uncertainty we can further reduce the width of the tolerance bounds to accommodate this. In order to demonstrate this second point we compare the two extremes: a) all variability not explained through the relationship between the dependent and independent variable is measurement error -- we refer to this extreme as the **measurement variation case** and b) all variability not explained through the relationship is device to device variability -- we refer to this extreme as the **device variation case**.

We supplement the problem above with an assumption that we know (through separate analysis) the distribution of the independent variable *X*. We assume that this has a normal distribution, *X* ~ Normal(84.1,4.1), where ~ implies "is distributed as" and 84.1 and 4.1 are the mean and standard deviation respectively. With these parameters, one can compute distribution percentiles. We will use the 2.5th and 97.5th percentiles, 73.1 and 89.5 respectively. We seek to construct tolerance bounds that cover this central 95% of the population with 90% certainty. Figure 5, below, provides an illustration of the problem.

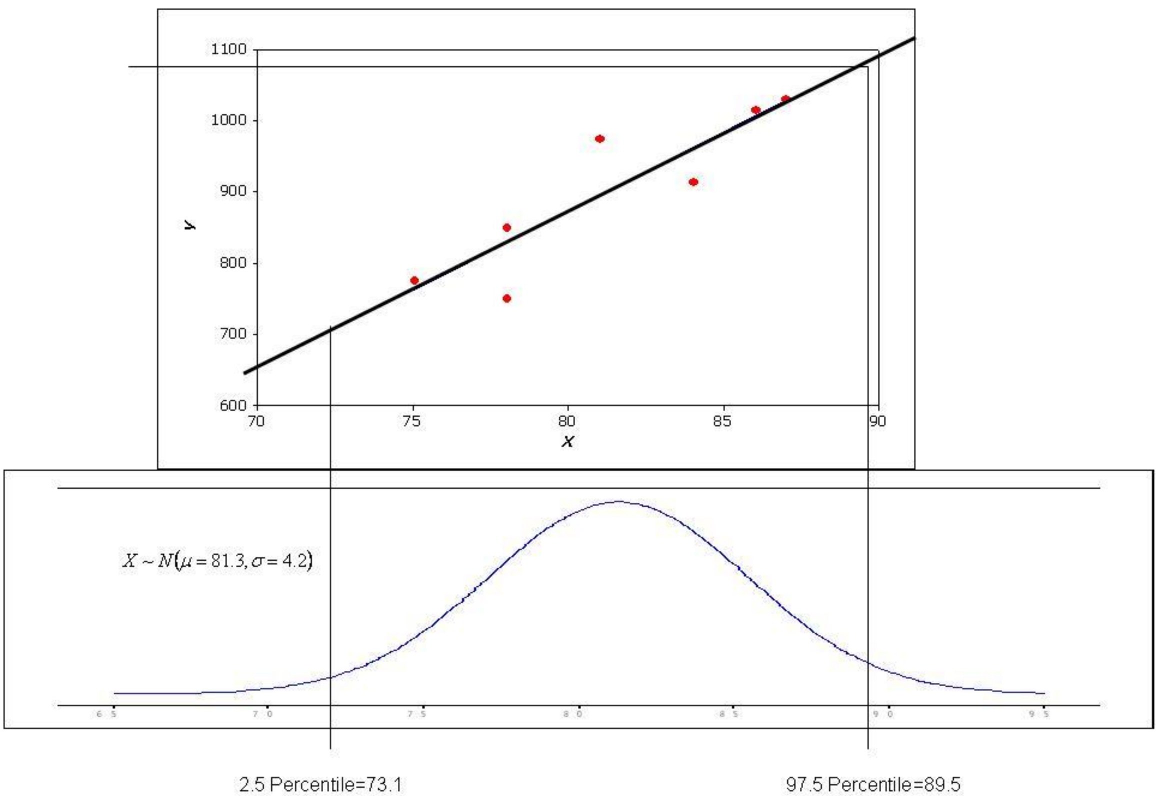

Figure 5. Data from Example 12.5 in [7], with additional specification of the distribution of *X*.

In order to address this problem, we use a simple linear regression model $y = \beta_0 + \beta_1 x$. Using the data provided, we estimate the regression parameters β_0, β_1 and σ^2 using the notation $\hat{\beta}_0, \hat{\beta}_1$ and s^2 respectively. We also make use of the "typical" assumptions for simple linear regression (simple linear model with independent homoscadastic (of equal variance) normally distributed residuals is appropriate).

We use the fact that under the assumptions of our analyses:

(4) $\left.(\mathrm{n}-2)\mathrm{s}^2\middle/\sigma^2\right. \sim \chi^2(n-2)$ where $\chi^2(n-2)$ is the Chi Square distribution with (n - 2) degrees of freedom;

(5) let $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \\ 1 & x_6 \\ 1 & x_7 \end{bmatrix}$, then $C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = s^2\left(X'X\right)^{-1}$ provides an estimate of the variance covariance matrix for the coefficients $\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$; and

(6) $\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \sim MVN\left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, C^*\right)$, where *MVN* is the multivariate normal distribution and $C^* = \sigma^2\left(X'X\right)^{-1}$.

In order to construct the tolerance intervals we simulate values for the line $y = \beta_0 + \beta_1 x$ that reflect our uncertainty in the parameters $\hat{\beta}_0, \hat{\beta}_1$ and s^2 based on a sample of size 7. This is illustrated in figure 6 below. The four distributions on the left-hand side are based on the intersection of the simulated lines with the percentiles of the soon normal distribution for *X*. the two solid PDFs are constructed for the **measurement variation case** in the two dashed PDFs are for the **device variation case**.

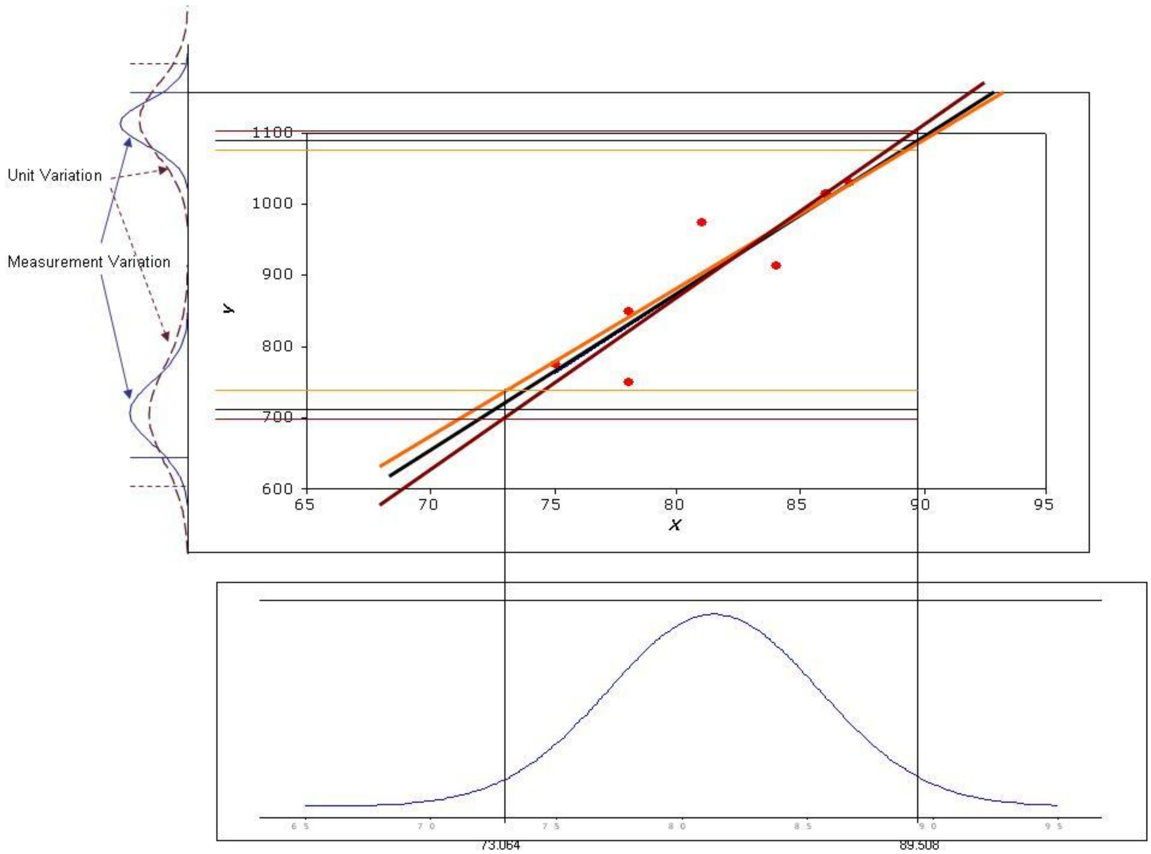

Figure 6. Examples of simulated regression lines used to construct percentiles of the distribution of the dependent variable *Y*.

The following algorithm was used to create tolerance bound estimates for the **measurement variation** and **device variation cases**.

For 100000 iterations, $i = 1,\ldots,100000$

using relationship (1), generate $\tilde{s}_i$

established the covariance matrix $\tilde{C}_i = \begin{bmatrix} \tilde{c}_{11,i} & \tilde{c}_{12,i} \\ \tilde{c}_{21,i} & \tilde{c}_{22,i} \end{bmatrix} = \tilde{s}_i^2 \left(X'X\right)^{-1}$

using relationships (2) and (3) with $\sigma^2 = \tilde{s}_i^2$ generate $\begin{bmatrix} \tilde{\beta}_{0,i} \\ \tilde{\beta}_{1,i} \end{bmatrix}$

Measurement variation case

calculate the upper response value $\tilde{y}_{i,meas} = \tilde{\beta}_{0,i} + \tilde{\beta}_{1,i} 89.5$

calculate the lower response value $\tilde{y}_{i,meas} = \tilde{\beta}_{0,i} + \tilde{\beta}_{1,i} 73.1$

Device variation case

calculate the upper response value $\tilde{y}_{i,unit} = \tilde{\beta}_{0,i} + \tilde{\beta}_{1,i} 89.5 + \varepsilon_{ii}$ where

ε_{ii} ~ Normal(0, $\tilde{s}_i$)

calculate the lower response value $\tilde{y}_{i,unit} = \tilde{\beta}_{0,i} + \tilde{\beta}_{1,i} 73.1 - \varepsilon_{ii}$ where

ε_{ii} ~ Normal(0, $\tilde{s}_i$)

end

end.

The tolerance bound estimates can now be taken from the simulated responses - we use the 5th percentile of the lower distributions of each response (the 5000th ordered response) and the 95th percentile of the upper distributions (the 95000th ordered response). These intervals provide two-sided tolerance bound estimates for 95 percent of the response population with 90% confidence. The interval are: (642, 1160) for the **measurement variation case** and (600, 1201) for **the device variation case**. The latter range compares with a .9/.95 tolerance interval of (540.35, 1262.5) constructed from the raw *y* data in Table 4. The tolerance interval is about 20% larger than the reduced interval computed by the demonstrated methodology. Figure 7 below shows the response distributions in more detail. The density functions were estimated using kernel density estimation.

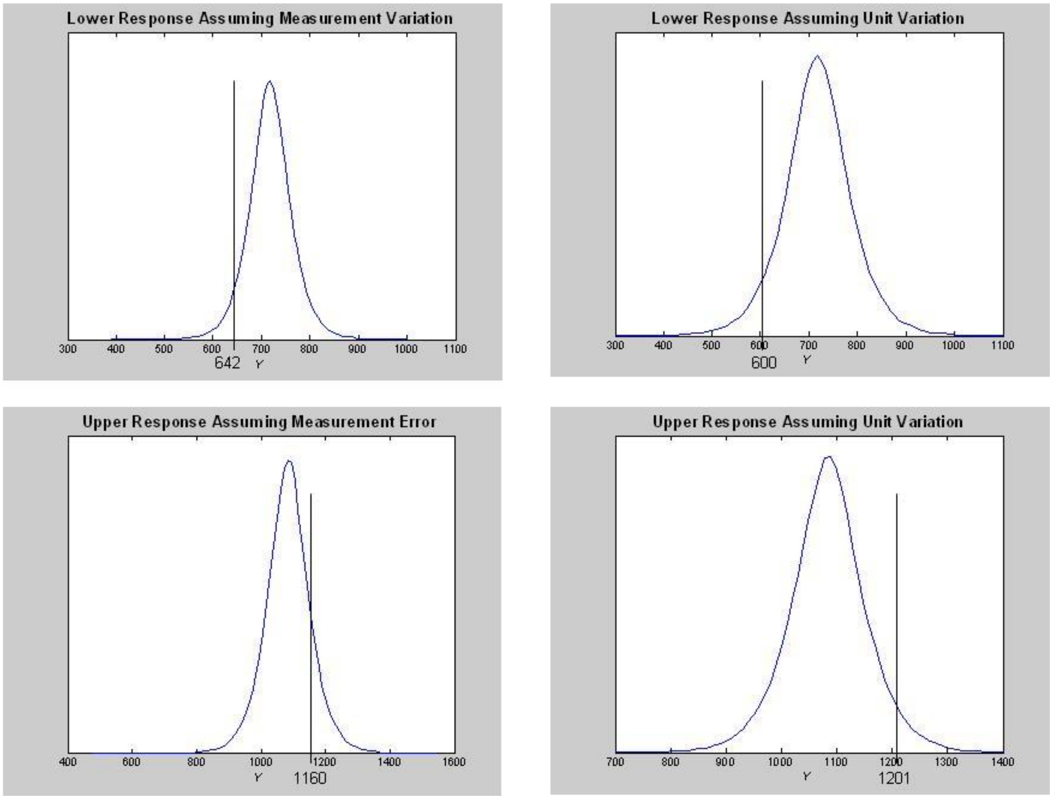

Figure 7. Kernel density estimates of the response distributions for the fifth and 95th percentile.

V. Closing

Some useful statistical concepts and techniques have been presented for expressing and refining uncertainty arising from: a) random variability (aleatory uncertainty) of a random quantity; and b) contributed epistemic uncertainty due to limited sampling of the random quantity. A number of experimental scenarios have been considered. The material is provided to assist modelers working in model validation and calibration who are non-statisticians and are not accustomed to working with random experimental data.

Acknowledgments

The authors thank the Advanced Simulation and Computing (ASC) Program of the National Nuclear Security Administration (NNSA) at Sandia National Laboratories for funding this research.

References

[1] Romero, V.J., "Comparison of Several Model Validation Conceptions against a "Real Space" End-to-End Approach," SAE paper 2011-01-0238 for Society of Automotive Engineers 2011 World Congress, April 12-14, 2011, Detroit, MI.

[2] Romero, V.J., "Type X and Y Errors and Data & Model Conditioning for Systematic Uncertainty in Model Calibration, Validation, and Extrapolation," SAE paper 2008-01-1368 for Society of Automotive Engineers 2008 World Congress, April 14-17, 2008, Detroit, MI.

[3] Romero, V.J., "Data & Model Conditioning for Multivariate Systematic Uncertainty in Model Calibration, Validation, and Extrapolation," paper AIAA-2010-2511, 12th AIAA Non-Deterministic Approaches Conference, 12-15 April 2010, Orlando, Fl.

[4] Efron, B., and R. Tibshirani (1986). "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy", *Statistical Science*, Vol. 1, No. 1, pp. 54-75

[5] Romero, V., L. Swiler, A. Urbina, "An Initial Comparison of Methods for Representing and Aggregating Experimental Uncertainties involving Sparse Data," 13th AIAA Non-Deterministic Approaches Conference, April 4-7, 2011, Denver, CO.

[6] Hahn, G.J., and W.Q. Meeker, *Statistical Intervals—A Guide for Practitioners*, Wiley & Sons, 1991.

[7] Devore, J.L., *Probability and Statistics for Engineering and the Sciences*, Brooks/Cole Publ. Co., division of Wadworth Inc., 1982.

[8] Kline, S.J., and McClintock, F.A., "Describing Uncertainties in Single-Sample Experiments," *Mechanical Engineering*, Jan. 1953, pp. 3 – 8.