WSRC-MS-96-0685

Statistical Methods for Accurately Determining Criticality Code Bias (U)

CONF-970926--11

by
E. F. Trumble
Westinghouse Savannah River Company
Savannah River Site
Aiken, South Carolina 29808
K. D. Kimball

A document prepared for 1997 CRITICALITY SAFETY CHALLENGES IN THE NEXT DECADE at Lake Chelan, WA, USA from 9/7/97 - 9/11/97.

DOE Contract No. **DE-AC09-89SR18035**

DISCLAIMER

WSRC-MS-96-0685

Statistical Methods for Accurately Determining Criticality Code Bias (U)

by

K. D. Kimball

NISYS Coropration
Norcross, Ga. 30071

and

E. F. Trumble

Westinghouse Savannah River Company
Savannah River Site
Aiken, SC 29808

A paper for presentation at the
The 1997 Criticality Topical Meeting
proposed for session "Close Enough: Criticality Safety Challenges in Accurate Determination of Subcritical Margins"
Lake Chelan, Wa.
September 7-11, 1997

STATISTICAL METHODS FOR ACCURATELY DETERMINING CRITICALITY CODE BIAS

K. D. Kimball, P.E.
NISYS Corporation
6055 Atlantic Blvd, Suite G-2
Norcross, GA 30071
(770) 263-0026

E. F. Trumble
Westinghouse Savannah River Company
Savannah River Site
Aiken, SC 29808
(803) 952-8348

I. ABSTRACT

A system of statistically treating validation calculations for the purpose of determining computer code bias is provided in this paper. The following statistical treatments are described: weighted regression analysis, lower tolerance limit, lower tolerance band, and lower confidence band. These methods meet the criticality code validation requirements of ANS 8.1[1].

II. INTRODUCTION

Westinghouse Savannah River Company recently established a methodology for statistically treating validation calculations to determine computer code bias. This methodology, which meets the criticality code validation requirements of ANS 8.1[1], allows for:

- the use of experimental as well as statistical uncertainty data;
- the "weighting" of validation calculations based on the degree of uncertainty;
- the use of a small number of data points when limited benchmark experiment data is available;
- the treatment of data that is not normally distributed;
- the ability to extrapolate results beyond the range of applicability; and
- the ability to "adjust" the subcritical margin based on the degree of knowledge of the neutronics and the cause for any bias trends.

III. OVERALL APPROACH

The approach developed for determining criticality code bias, illustrated in Figure 1, is a tiered approach that uses knowledge about the benchmark data and the neutronics resulting in observed bias trends.

The first step in determining the code bias is to see if there is a clear trend in the calculated results. Such a trend will indicate a tendency of the code to bias the results based on an independent parameter. The ability to predict this trend will lead to a more accurate determination of the true bias for the desired parameter (or range of applicability).

Two statistical treatments are available to evaluate a bias that has a clear trend: the single-sided lower confidence band and the single-sided lower tolerance band. The single-sided lower confidence band defines the region where the true bias is expected to be, within a prescribed degree of confidence. This method contains the smallest statistical determination of bias and bias uncertainty among the methods, but should be used only if the reason for the trend in bias is well known.

If the reason for the trend in bias is not known, then

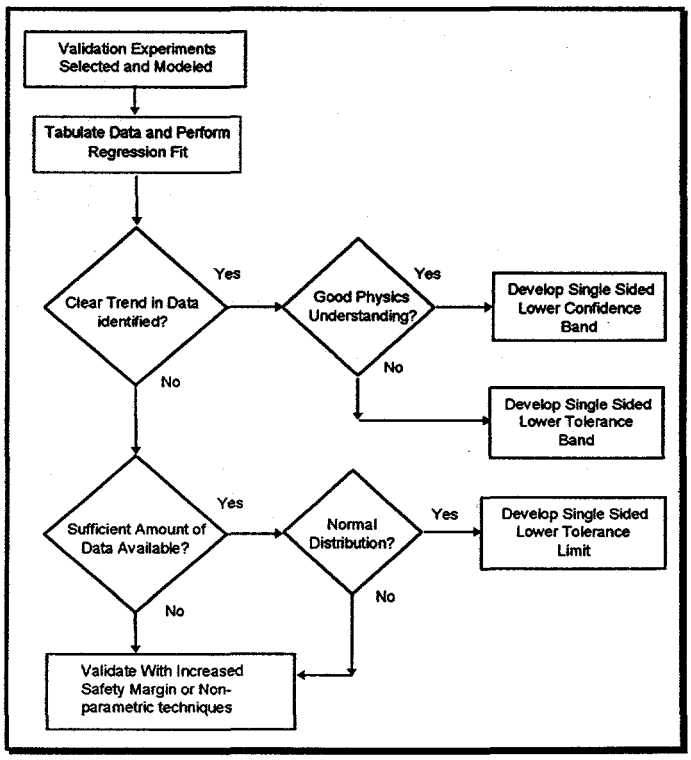

Figure 1 - Approach for Determining Criticality Code Bias

a more conservative single-sided lower tolerance band is used. The tolerance band defines the region above which a proportion of bias will lie for some prescribed degree of confidence.

The benchmark validation results may not have a clear trend. In this case, a lower tolerance limit should be used. This type of statistical treatment is only valid over the range of applicability of the benchmark experiments and assumes that the benchmark results have a normal distribution.

There will be situations where there are an insufficient number of benchmark experiments applicable to a specific problem or the data exhibits a non-normal distribution. For these cases, a validation may still be performed, with an increased minimum subcritical margin (MSM). This approach is also presented in this paper.

IV. DATA REGRESSION FITS

Looking for a trend in data is not always obvious. Therefore, regression techniques are used as an aid to identifying clear trends. Developing a fit to data requires the analyst to consider the following questions:

- Is all of the data of equal weight or is some data more reliable than other data?
- What independent variables should be used for developing the fit?
- What type fits are there and what will work?
- How does the analyst know there is a "good" fit?

A. Data Uncertainties

Knowing the uncertainties associated the with validation experiments and computer calculations will allow the analyst to develop a "truer" fit to the data. If some data are more reliable than other data, then the good data should be treated with more weight. This will prevent the fit from being skewed toward data that are more or less suspect.

"Weighted" data has known statistical uncertainties and/or experimental uncertainties associated with each data point. The three types of uncertainties are:

Experimental Uncertainty (σ_e) - Modeling of validation experiments frequently result in assumptions about experimental conditions. In addition, experimental uncertainties (such as measurement tolerances) influence the development of a computer model. Recent efforts by the Organization for Economic Cooperation and Development - Nuclear Energy Agency[2] (OECD-NEA) have resulted in the quantification of such uncertainties in validation experiments.

Statistical Uncertainty (σ_s) - Monte Carlo calculation techniques result in a statistical uncertainty associated with the actual calculation. This type of uncertainty is dependent upon many factors, including the number of neutron generations, the number of starting neutrons, and the problem geometry. For this paper, σ_s refers to the statistical Monte Carlo uncertainty associated with the computer modeled validation experiment, or the convergence criteria for deterministic (Sn) type methods.

Total Uncertainty - This is the total uncertainty associated with a calculated k_{eff} on a benchmark experiment. The total uncertainty for an individual benchmark calculation is:

$$\sigma_i = \sqrt{\left(\sigma_{e,i}^2 + \sigma_{s,i}^2\right)} \qquad (1)$$

where the subscript (i) refers to an individual benchmark calculation.

B. Normalizing k_{eff}

In many instances, benchmark experiments used for validation may not be exactly critical. Experimental results may show that the experiment is slightly above or below a $k_{eff} = 1.0$. For these cases, a possible approach to calculating a bias would be to use a difference in k_{eff}, (Δk_{eff}) (i.e., k_{eff}(calculated) - k_{eff}(experimental)). Note that this approach could also allow the limited use of subcritical experiments to determine a bias and bias uncertainty.

Statistical treatments can certainly be used with a Δk_{eff}. However, the majority of linear and non-linear regression fits will not be possible if there are both negative and positive values in the data. Therefore, use of a Δk_{eff} approach will eliminate many statistical methods from consideration.

As an alternative, the calculated k_{eff} values could be normalized to the experimental value. This assumes that any inherent bias in the calculation is not affected by the normalization, which is valid for small differences in k_{eff}. To normalize k_{eff}, the following formula applies:

$$k_{eff}(\text{normalized}) = k_{eff}(\text{calculated}) \,/\, k_{eff}(\text{experimental})$$

C. Identifying Independent Variables

The independent variables selected for fitting the k_{eff} results need to be selected with care. These variables must relate to parameters used in establishing the range of applicability of the benchmark calculations.

Table 1 presents several types of independent variables that may be tried in establishing the fit. The list is not all inclusive and it is important to note that the calculational bias is rarely the function of a single parameter. Understanding the reason for a trend in bias with respect to a particular parameter is important in establishing the bias statistical treatment.

D. Types of Data Fits

Data fits are used to correlate data to an equation. The types of data fits include linear, non-linear, and polynomial fits.

A linear fit is determined by a linear regression analysis that results in an equation of the form:

$$y = a + bx \tag{2}$$

where $y = k_{eff}$
x = independent variable
a = y intercept
b = slope of the line

Each benchmark calculation has a total uncertainty that can be determined using equation (1). This total uncertainty can then be used to "weight" the data when performing a linear regression analysis. From Bevington[3], the coefficients of the linear equation can be found using a least squares fit from the following equations:

$$\begin{aligned} a &= \frac{1}{\Delta}\left(\sum W_i x_i^2 \sum W_i y_i - \sum W_i x_i \sum W_i x_i y_i\right) \\ b &= \frac{1}{\Delta}\left(\sum W_i \sum W_i x_i y_i - \sum W_i x_i \sum W_i y_i\right) \\ \Delta &= \sum W_i \sum W_i x_i^2 - \left(\sum W_i x_i\right)^2 \\ W &= \frac{1}{\sigma_i^2} \end{aligned} \tag{3}$$

Inserting the coefficients a and b from equation (3) into equation (2) will yield a straight line fitted to the weighted data.

Non-linear equations are handled by "transforming" the non-linear equation into a linear equation and then solving for the applicable coefficients. The transformed non-linear equation is then represented by:

$$Y_T = a_T + b_T X_T \tag{4}$$

Table 1 - Typical Independent Variables for k_{eff} Regression Fits

Example Bias Fit Independent Variables	
Variable	**Comment**
Enrichment	Used as a parameter to characterize bias from low to high enriched systems
Fissile Concentration	Primarily used for solution forms of fissile material.
Average Energy Group of Fissions	Used to characterize the energy spectrum of the fissioning neutron. Other measures include the average energy (ev) of the fissioning neutrons. This helps to characterize the system as a fast, intermediate, or thermal system.
Geometric Measurement	Variables include: Unit Cell Dimensions, Lattice Spacing, Interstitial Moderator Dimension, Array Size, etc.
Neutron Absorber	Primarily used for poisoned systems.
Moderation (H/X) Ratios	Used to characterize the moderating capability of a system. The common parameter is an H/X ratio. However, this parameter does not provide measurement of moderating capabilities of compound systems.
Other Dimensionless Parameters	Thermalness: p/[n(thermal)*f(thermal)] Thermal Absorption Ratio: (Thermal Absorptions/Total Absorptions)

Table 2 presents the linear transformations for several types of non-linear equations. Note that the total uncertainty must also be transformed to provide a weighted fit.

Table 2 - Linear Transformations for Nonlinear Equations

Form of Equation	Calculate The Transformed Values			Convert Straight Line Constants (a_T and b_T) To Original Constants	
	$Y_T =$	$X_T =$	$\sigma_{i,T} =$	$a_T =$	$b_T =$
$Y = a + b/X$	Y	1/X	σ_i	a	b
$Y = 1/(a + bX)$	1/Y	X	σ_i / Y^2	a	b
$Y = X/(a+bX)$	X/Y	X	$X \sigma_i / Y^2$	a	b
$Y = ab^X$	log Y	X	$(\log e) \sigma_i / Y$	log a	log b
$Y = ae^{bX}$	log Y	X	$(\log e) \sigma_i / Y$	log a	b log e
$Y = aX^b$	log Y	log X	$(\log e) \sigma_i / Y$	log a	b
$Y = a + bX^c$ (where c is known)	Y	X^n	σ_i	a	b

Figure 2 presents an example case using fictitious data. This figure shows the effect of weighting data with uncertainties for a linear fit. For this example, the two largest value keff data points have been assigned an uncertainty significantly greater than the remaining data points, representing benchmark experiments with large experimental uncertainties. As can be seen from the figure, the shift in linear slope when weighting is applied shows the impact of the "less" certain data. This shift in curve shape, can result in a significant shift in the bias represented by a tolerance or confidence band.

E. Goodness of Fit

There are two steps that should be employed when determining the goodness of fit. The first is to plot the data against the independent variable using different scales of axes. This allows for a visual evaluation on the effectiveness of the regression fit.

The second step is to numerically determine a goodness of fit after linear or non-linear relations are fitted to the data. This adds a useful measure because visual inspection of the data plot and the associated fit will not necessarily reveal how good the fit is to the data. The linear correlation coefficient is one standard method used to numerically measure goodness of fit.

The linear-correlation coefficient is a quantitative measure of the degree to which a linear relationship exists between two variables. For weighted data, the linear-correlation coefficient is

$$r \equiv \frac{\sum \frac{1}{\sigma_i^2}(x_i - \bar{x})(y_i - \bar{y})}{\left[\sum \frac{1}{\sigma_i^2}(x_i - \bar{x})^2\right]^{1/2}\left[\sum \frac{1}{\sigma_i^2}(y_i - \bar{y})^2\right]^{1/2}} \quad (5)$$

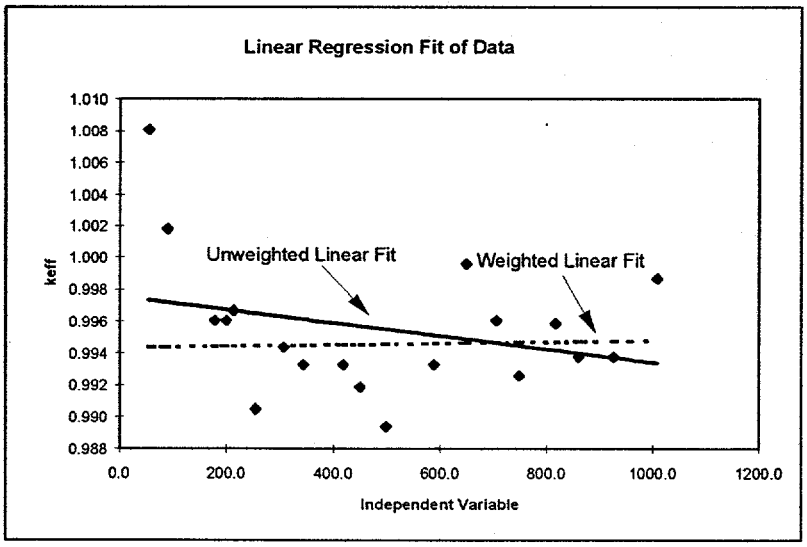

Figure 2 - Example showing the effect of weighted data and unweighted data on a linear regression fit.

The value of the linear-correlation coefficient is often expressed as a squared term, r^2. The closer r^2 approaches the value of 1, the better the fit of the data to the linear equation.

Note that neither the coefficient by itself, nor the comparison of coefficients can provide an absolute measure of how good the fit is.

V. STATISTICAL TREATMENTS

As stated earlier, the approach to determine bias and bias uncertainty relies on three types of statistical treatments: a tolerance limit, a tolerance band, and a confidence band.

A. Tolerance Limit

This method defines a limit above which the true population of k_{eff} benchmark calculations is expected to lie. It is used when data are shown to have a normal distribution and the data cannot be reasonably fit to an independent variable. This method cannot be used to extrapolate beyond the range of applicability.

There are several tests to determine if data are normally distributed. One common test is the Chi-square test, which has a limitation that the sample be at least 50 data points. Since there are typically fewer than 50 benchmark experiments for validation, the Shapiro-Wilk[4] test is another test that can be used.

The one-sided lower tolerance limit is defined by the equation:

$$K_L = \bar{k}eff - U S_P \quad (6)$$

K_L = one – sided lower tolerance limit

$\bar{k}eff$ = mean keff value

S_P = *sqrt*(pooled variance)

U = one – sided lower tolerance factor

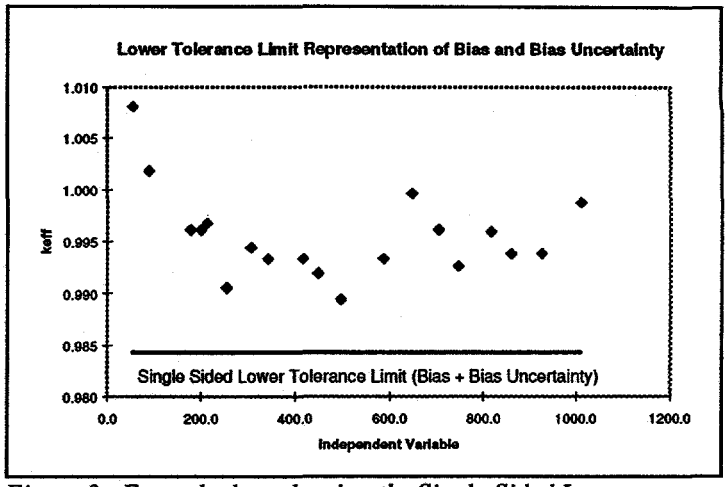

Figure 3 - Example data showing the Single-Sided Lower Tolerance Limit as a representation of bias and bias uncertainty.

The weighted equation variables for the single sided lower tolerance limit are presented in Table 3. The one-sided lower tolerance factor, U, is found in Natrella[5] using confidence (γ) and proportion factors (P) as variables. Note that Natrella gives a method to approximate U; however, this approximation is non-conservative.

The pooled variance is:

$$S_P = \sqrt{s^2 + \bar{\sigma}^2} \quad (7)$$

where: s^2 = variance about the mean

$\bar{\sigma}^2$ = average total uncertainty

Figure 3 presents an example of the single-sided lower

Table 3 - - Single-Sided Lower Tolerance Limit Equation Variables

Variable	Weighted Data
Variance About The Mean	$s^2 = \dfrac{\left(\frac{1}{n-1}\right)\sum\frac{1}{\sigma^2}(y-\bar{y})^2}{\frac{1}{n}\sum\frac{1}{\sigma^2}}$
Average Total Uncertainty	$\bar{\sigma}^2 = \dfrac{n}{\sum\frac{1}{\sigma_i^2}}$
Mean k_{eff} Value	$\bar{k}eff = \dfrac{\sum\frac{1}{\sigma_i^2}keff_i}{\sum\frac{1}{\sigma_i^2}}$

tolerance limit as it applies to bias and bias uncertainty. It is important to note that this method is only good for the range of applicability of the benchmark experiments.

B. Tolerance Band

When a relationship between a calculated k_{eff} and an independent variable can be determined, a one-sided lower tolerance band may be used. This is a conservative method that provides a fitted curve above which the true population of k_{eff} is expected to lie. The tolerance band equation is actually a calibration curve relation. This was selected because it was anticipated that a given tolerance band would be used multiple times to predict bias. Other typical predictors such as a single future value, can only be used for a single prediction to ensure the degree of confidence desired.

A similar method involving a fixed-width tolerance band is proposed by Dyer[6], however the technique presented by Dyer cannot be used to extrapolate beyond the range of applicability.

The equation for the one-sided lower tolerance band[7] is

$$K_L = K_{fit}(x) - S_p\left\{\sqrt{2F_\alpha^{(2,n-2)}\left[\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]}+z_{2P-1}\sqrt{\frac{(n-2)}{\chi^2_{1-\gamma,n-2}}}\right\}$$

where:

S_P = Square root of the pooled variance

$F_\alpha^{(fit,n-2)}$ = the F distribution percentile with (degree of fit, n - 2 degrees of freedom The degree of fit is 2 for a linear fit.

α = 1 minus the desired confidence, β, $= (1-\beta)$

n = the number of benchmark keff values

x = the independent fit variable

$\bar{x}$ = the mean of the independent variables

z_{2P-1} = the symetric percentile of the Gaussian or normal distribution that contains the P fraction

$\chi^{2,(n-2)}_{1-\gamma}$ = the upper Chi - square percentile

$$\gamma = 1-\frac{\alpha}{2}-\beta$$

This equation was developed with the least squares technique and is applicable to linear equations. For non-linear equations, the tolerance band equation is used with the transformed non-linear values and then the final result is transformed back.

Figure 4 presents an example of the one-sided lower tolerance band for a weighted linear fit of the example data.

C. Confidence Band

If a clear relationship exists between an independent variable and k_{eff}, and a good understanding of the physics causing this relationship exists, then a one-sided lower confidence band may be used. The confidence band is a fitted curve, above which the mean bias is expected to lie.

The equation for a confidence band is very similar to the tolerance band.

$$K_L = K_{fit}(x) - S_p\left\{\sqrt{2F_{2\alpha}^{(2,n-2)}\left[\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]}\right\}$$

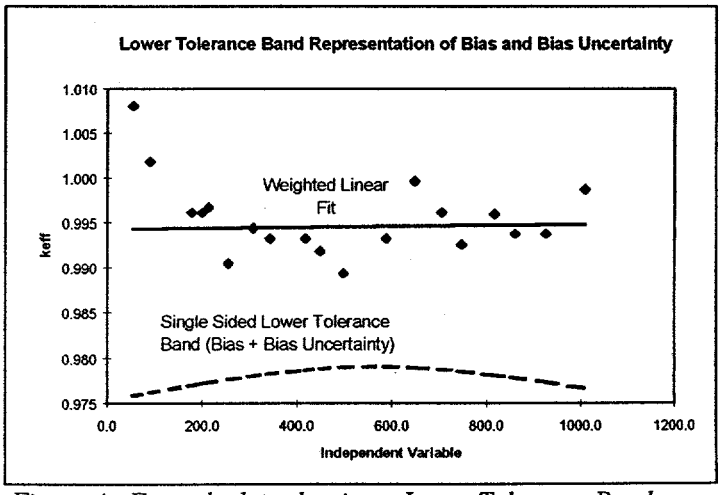

Figure 4 - Example data showing a Lower Tolerance Band as a representation of Bias and Bias Uncertainty.

where:

S_P = Square root of the pooled variance

$F_{2\alpha}^{(fit,n-2)}$ = the F distribution percentile with degree of fit, n - 2 degrees of freedom. The degree of fit is 2 for a linear fit.

α = 1 minus the desired confidence, β, $= (1-\beta)$

n = the number of benchmark keff values

x = the independent fit variable

$\bar{x}$ = the mean of the independent variables

$$\gamma = 1 - \frac{\alpha}{2} - \beta$$

Note that there is a change in the probability value for the F-distribution lookup; the F-distribution percentile is determined at a value of 2α.

As with the tolerance band, values for the confidence limit are calculated at various values of the independent variable, X, and are then plotted to give the bias and the bias uncertainty. Figure 5 is a graphical example of the confidence band. Note that this band is much closer to the fitted equation than the tolerance band because it is a measure of the mean bias.

D. Non-Parametric Data

Two situations where the statistical techniques described above may not adequately determine the computational bias are: (1) data that do not have a correlation with an independent variable and the data is not normally distributed, and (2) an insufficient amount of data. Data that do not have a normal distribution may be handled with non-parametric techniques. For insufficient data, expansion of the range of applicability with an increased minimum subcritical margin (as needed) can be used. Each of these techniques is described below.

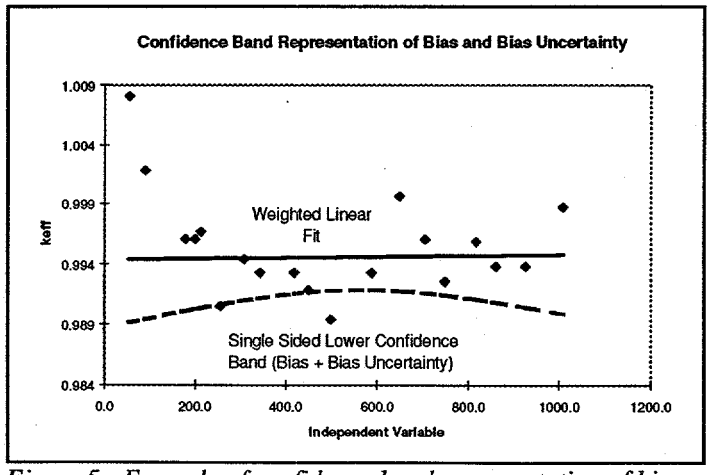

Figure 5 - Example of confidence band representation of bias and bias uncertainty.

1. Non parametric Techniques

Data that do not follow a normal distribution can be analyzed by non-parametric techniques. The analysis results in a determination of the degree of confidence that a fraction of the true population of data lies above the smallest observed value. The more data available in the sample, the higher the degree of confidence.

The following equation determines the percent confidence that a fraction of the population is above the lowest observed value:

$$\beta = 1 - \sum_{j=0}^{m-1} \frac{n!}{j!\cdot(n-j)!}(1-q)^j\, q^{n-j}$$

where: q is the desired population fraction (normally 0.95)
n is the number of data samples
m is the rank order of the smallest sample (m = 1 for the smallest sample)

For a desired population fraction of 95%, and a rank order of 1 (m = 1), the equation reduces to:

$$\beta = 1 - q^n$$

$$\beta = 1 - 0.95^n$$

For example, if the data sample size is n = 19, then β = 62.3%, or there is a 62.3% confidence that 95% of the population lies above the smallest observed value.

The equation can also be rewritten to provide information about the fraction of the population, given a confidence level.

$$\log(q) = \frac{\log(1-\beta)}{n}$$

If the sample size is n = 19, and the degree of confidence is β = 0.95, then q = 85.4%, or there is a 95% confidence that 85.4% of the population lies above the smallest value.

By using these equations to determine a confidence level or a population fraction, an additional subcritical margin is assigned when determining k_{eff}(safe). As stated earlier, the normal value of k_{eff}(safe) is:

k_{eff}(safe) = 1.0 + (Bias) - (Bias Uncertainty) - MSM

For non-parametric data analysis, k_{eff}(safe) is determined by:

k_{eff}(safe) = Smallest k_{eff} value - (Bias Uncertainty) - MSM - Non-parametric margin

where: Smallest k_{eff} value is the lowest calculated value in the data sample.

Bias uncertainty = the total statistical and experimental uncertainty

MSM = Minimum Subcritical Margin

Non-parametric Margin = additional safety margin from Table 4.

Table 4 - Non-Parametric Subcritical Margins

Degree of Confidence for 95% of the Population	Non-parametric Margin
> 90%	0.0
> 80%	0.01
> 70%	0.02
> 60%	0.03
> 50%	0.04
> 40%	0.05
≤ 40%	Additional data needed (This corresponds to less than 10 data points.)

2. *Insufficient Data*

The range of applicability of an analysis may result in fewer than 10 benchmark experiments. This is particularly true for cases with unusual moderating materials.

When this occurs, the analyst should examine the problem to see if the range of applicability can be expanded to include more benchmark experiments. This is particularly acceptable if the behavior of the fissile material is well known.

In the event that less than 10 benchmark experiments are available, then the analyst will need to apply a minimum subcritical margin to account for the lack of data. The amount of margin must be determined by engineering judgment because no deterministic guidance is available.

E. Extrapolation of Results

Often, available experimental data do not completely cover the range of applicability of the problem to be analyzed. One of the advantages of performing a linear or non-linear fit to benchmark validation calculations is the ability to extrapolate beyond the range of applicability.

For obvious reasons, extrapolation cannot be performed if using a single sided lower tolerance limit as the statistical treatment. However, if a good fit of the data to an independent variable is possible, then a single sided lower tolerance band or confidence band can be used for extrapolation. Some key points on extrapolating these bands follow.

- The equations of fit do not necessarily retain their "shape" beyond the range of applicability. For example, some polynomial equations will oscillate in shape. Therefore, the shape of the fitted equations need to be checked to verify that the equation still reflects the expected trend in the data.
- Published guidance[8] recommends that extrapolation be limited to +/- 5% of the independent variable. This percentage may not reflect the degree of extrapolation fairly and must be evaluated in each case. In any event, justification of the degree of extrapolation must be documented in the analysis.
- Extrapolation of a tolerance band or a confidence band should not result in a positive bias value (i.e., a k_{eff} greater than or equal to 1.0).
- Extrapolation should only be used in cases where there is a good fit to the data and experimental data suitable for benchmark calculations is not available in the desired range of applicability.

VI. ACKNOWLEDGMENTS

The authors would like to acknowledge Mr. Jim Mincey for his conceptualization of these statistical treatments in the application to criticality code bias determination.

---

[1] *American National Standard for Nuclear Criticality Safety in Operations with Fissionable Materials Outside Reactors,* ANSI/ANS 8.1-1983, American Nuclear Society, La Grange Park, Illinois (1983).

[2] *International Handbook of Evaluated Criticality Safety Benchmark Experiments*, NEA/NSC/DOC(95)03, Organization for Economic Cooperation and Development - Nuclear Energy Agency.

[3] *Data Reduction And Error Analysis For The Physical Sciences*, Philip R. Bevington, McGraw-Hill Book Company, 1969.

[4] *Goodness of Fit Techniques*, D'Agostino and Stephens, marcel Decker, Inc. New York.

[5] *Experimental Statistics, National Bureau of Standards Handbook 91*, National Institute of Standards and Technology, M.G. Natrella, (August 1963).

[6] *A Technique for Code Validation for Criticality Safety Calculations*, H.R. Dyer, W.C. Jordan, V.R. Cain, Transactions of the American Nuclear Society 63, page 238 (June 1991).

[7] *Simultaneous Statistical Inference*, Miller, 2nd Edition, Rupert G., Jr., Springer-Verlag New York Inc.

[8] *Forecast of Criticality Experiments and Experimental Programs Needed to Support Nuclear Operations in the United States of America: 1994 - 1999*, LA-12683, Appendix E, Initial Draft of Criteria for Establishing Area of Applicability, Los Alamos National Laboratory.