

Determination of Bias and Bias Uncertainty for Criticality Safety Computational Methods

Office of Nuclear Material Safety and Safeguards

AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

NRC Reference Material

As of November 1999, you may electronically access NUREG-series publications and other NRC records at the NRC's Library at <u>www.nrc.gov/reading-rm.html.</u> Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and Title 10, "Energy," in the *Code of Federal Regulations* may also be purchased from one of these two sources:

1. The Superintendent of Documents

U.S. Government Publishing Office Washington, DC 20402-0001 Internet: <u>https://bookstore.gpo.gov/</u> Telephone: (202) 512-1800 Fax: (202) 512-2104

2. The National Technical Information Service 5301 Shawnee Road Alexandria, VA 22312-0002 Internet: https://www.ntis.gov/ 1-800-553-6847 or, locally, (703) 605-6000

A single copy of each NRC draft report for comment is available free, to the extent of supply, upon written request as follows:

Address: U.S. Nuclear Regulatory Commission

Office of Administration Program Management and Design Service Branch Washington, DC 20555-0001 E-mail: <u>Reproduction.Resource@nrc.gov</u> Facsimile: (301) 415-2289

Some publications in the NUREG series that are posted at the NRC's Web site address <u>www.nrc.gov/reading-rm/</u> <u>doc-collections/nuregs</u> are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

Non-NRC Reference Material

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, transactions, *Federal Register* notices, Federal and State legislation, and congressional reports. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—

The NRC Technical Library Two White Flint North 11545 Rockville Pike Rockville, MD 20852-2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

American National Standards Institute 11 West 42nd Street New York, NY 10036-8002 Internet: https://www.ansi.org/ (212) 642-4900

Legally binding regulatory requirements are stated only in laws; NRC regulations; licenses, including technical specifications; or orders, not in NUREG-series publications. The views expressed in contractor prepared publications in this series are not necessarily those of the NRC.

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG–XXXX) or agency contractors (NUREG/CR–XXXX), (2) proceedings of conferences (NUREG/CP–XXXX), (3) reports resulting from international agreements (NUREG/IA–XXXX),(4) brochures (NUREG/BR–XXXX), and (5) compilations of legal decisions and orders of the Commission and the Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of the NRC's regulations (NUREG-0750), (6) Knowledge Management prepared by NRC staff or agency contractors (NUREG/KM-XXXX).

DISCLAIMER: This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any employee, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this publication, or represents that its use by such third party would not infringe privately owned rights.





Determination of Bias and Bias Uncertainty for Criticality Safety Computational Methods

Manuscript Completed: March 2025 Date Published: April 2025

Prepared by: J. B. Clarity (Currently at Pacific Northwest National Laboratory) W. J. Marshall D. E. Mueller S. S. Powers B. T. Rearden S. M. Bowman

Oak Ridge National Laboratory Oak Ridge, TN 37831-6283

Andrew Barto, NRC Project Manager

Office of Nuclear Material Safety and Safeguards

ABSTRACT

Nuclear criticality safety evaluations must demonstrate that operations are subcritical under both normal and credible abnormal conditions, and such evaluations often rely upon computational techniques to determine the neutron multiplication factor for complex three-dimensional systems. Validation of the computer codes and data used to model these systems establishes their suitability for specific applications. The validation activity also determines the computational bias and the uncertainty in that bias that is relevant to the application. The bias is developed from calculations of known laboratory critical experiments that are similar to the intended application of interest. This report describes techniques that can be used by criticality safety analysts to perform the validation activity, including determination of calculational bias, bias uncertainty, and the application of those values to develop limits that can be applied in safety analyses. This report builds upon earlier works in the criticality safety validation area and incorporates modern analytical techniques developed over the last twenty years, as well as lessons learned from observations of previous validation efforts.

ABSTR	ACT	iii		
LIST OF	FIGURES	vii		
LIST OF TABLESix				
EXECU	TIVE SUMMARY	xi		
ABBRE	VIATIONS AND ACRONYMS	xiii		
1 INTR	ODUCTION AND BACKGROUND	1-1		
2 THE	PURPOSE OF VALIDATION	2-1		
3 DEFI	NITION OF COMPUTATIONAL METHOD	3-1		
4 CRIT DETE	ICAL EXPERIMENT SELECTION AND AREA OF APPLICABILITY ERMINATION	4-1		
4.1 C 4.2 S 4 4 4	election of Critical Experiments	4-1 4-3 4-3 4-5 4-6 4-7		
4.3 D 4	Defining the Area of Applicability	4-10 4-11		
5 STAT 5.1 H 5.2 A 5 5 5 5.3 G	ISTICAL BACKGROUND lypothesis Testing ssessment of Normality .2.1 Graphical Techniques .2.2 Normality Tests .2.3 Power Comparison of Normality Tests .30000ness-of-Fit Testing	5-1 5-1 5-2 5-4 5-8 5-10 5-14		
6 DETE 6.1 N 6.2 N 6 6 6 6 6 6 6 6 6 6 6 6 6	ERMINATION OF BIAS AND UNCERTAINTY Iontrending Methods Ionparametric Methods .2.1 Historical Nonparametric Method .2.2 Whisper Method .alysis of Trends .3.1 Calculation of the Bias for a Statistically Significant Trend .3.2 Single-Sided Lower Tolerance Band .3.3 Confidence Band with Administrative Margin (USL-1) .3.4 Single-Sided Uniform Width Closed Interval Approach (USL-2) Ising Positive Validation Analysis Biases	6-1 6-3 6-8 6-8 6-10 6-11 6-13 6-14 6-15 6-16 6-17		
6.5 ls	sue, Impact, and Potential Inclusion of Correlated Critical Experiment Results	6-17		

TABLE OF CONTENTS

	6.6 Using Va	lidation Analysis Results	6-18
7	IDENTIFYIN GAPS	G AND ADDRESSING VALIDATION WEAKNESSES AND	7-1
8	DOCUMENT	TATION	8-1
9	SUMMARY		9-1
10	REFERENC	ES	10-1
A	PPENDIX A	EXAMPLE CALCULATIONS OF BIAS AND BIAS UNCERTAINTY	A-1

LIST OF FIGURES

Figure 2-1	Flow Chart of the Validation Activity	2-3
Figure 4-1	Comparison of the Fission Cross Section Sensitivity Profiles of HMF-024- 001 and HST-014-001 Showing the Lack of Overlap of the Sensitivity Profiles Between Fast and Thermal Systems	4-13
Figure 4-2	Comparison of HST and HMF Systems with an HEU D_2O Application as a Function of EALF	4-14
Figure 5-1	Comparison of Deviated Distributions with Standard Normal Distribution	5-3
Figure 5-2	Histogram (Left) and Normal Q-Q Plot (Right) of a 200-Point Sample Drawn from a Normal Distribution	5-6
Figure 5-3	Histogram (Left) and Normal Q-Q Plot (Right) of a 200-Point Sample Drawn From a Negatively Skewed Distribution	5-6
Figure 5-4	Histogram (Left) and Normal Q-Q Plot (Right) of a 200-Point Sample Drawn from a Positively Skewed Distribution	5-7
Figure 5-5	Histogram (Left) and Normal Q-Q Plot (Right) of a 200-Point Sample Drawn from a Leptokurtic Distribution	5-7
Figure 5-6	Histogram (Left) and Normal Q-Q Plot (Right) of a 200-Point Sample Drawn from a Platykurtic Distribution	5-8
Figure 5-7	Comparison of Platykurtic (Black), Negatively Skewed (Red), and Combined Platykurtic and Negatively Skewed (Orange) Distributions Used in the Power Study to the Normal Distribution (Green) for 90% and 85% of Population Coverage	5-11
Figure 6-1	Statistical Method Selection Process	6-2
Figure 6-2	Summary of Statistical Methods for Determining Bias and Bias Uncertainty	6-3
Figure 7-1	Nuclear Data Uncertainty for ¹⁰ B Total Cross Section	7-2

LIST OF TABLES

Table 5-1	Summary of Possible Results of Hypothesis Test5	-2
Table 5-2	List of Generally Acceptable Normality Tests5	-9
Table 5-3	Probability of Rejection for the 90% Coverage of Population Cases5-	12
Table 5-4	Probability of Rejection for the 85% of Coverage of Population Cases5-	13
Table 5-5	Probability of Rejection for the Cases Drawn from a Normal Distribution5-	14
Table 6-1	Single-Sided Tolerance Factors for 95% Confidence that 95% of the True Population Lies Above the Tolerance Limit as a Function of the Number of Points in the Sample	-7
Table 6-2	Sample Size as a Function of Lower Rank of Data Points Necessary to Give 95% Confidence that 95% of the Population Lies Above that Point6	-9
Table 6-3	NPM as a Function of the Degree of Confidence Calculated with Eq. (18)6	-9
Table 6-4	Explanation of the Variables and Statistical Functions in Eq. (32)	15
Table 6-5	Summary of Bias and Bias Uncertainty Calculation Techniques	19
Table 8-1	Example Validation Report Layout with Description of Each Section8	-1

EXECUTIVE SUMMARY

The purpose of criticality safety is to prevent any inadvertent criticality from occurring during the handling or storage of fissile material. Calculations are frequently used to demonstrate that a sufficient subcritical margin exists. A key facet of the evaluation process is *validation*, which establishes the suitability of and determines the accuracy (i.e., bias) and the associated uncertainty of the computational method and data for the intended application.

Several documents have been generated to support validation of criticality safety computational methods over a range of systems. In some cases, guidance has been developed to accommodate specific regulatory requirements. Many of these reports are targeted for specific types of applications, such as transportation of light-water reactor fuel. No single guidance document has been identified that is intended to apply to all types of systems while also incorporating both traditional and sensitivity/uncertainty (S/U)-based validation techniques. S/U-based validation techniques have been added to traditional validation. The purpose of this report is to provide criticality safety computational method validation techniques for analyses involving all types of fissionable material operations. This is a work of synthesis, combining recommendations for acceptable validation approaches from their various sources. New recommendations are developed in areas where existing guidance is vague, incomplete, or lacking.

In addition, this document was prepared to address typical validation issues identified from observations of previous validation efforts. These observations include (1) inappropriate critical experiment selection, (2) insufficient trending analysis, (3) incorrect application of bias and/or bias uncertainty, (4) failure to meet bias method prerequisites, (5) failure to identify validation gaps and weaknesses, and (6) inadequate documentation. This document was developed to augment previous criticality validation guidance documents, providing recommendations to more effectively address some of the identified issues.

Section 1 of this document provides an introduction and some background information on criticality safety validation. Section 2 discusses the purpose of validation, and Section 3 discusses the definition of a computational method. Section 4 provides suggestions for characterizing the safety analysis model, critical experiment selection and area of applicability determination. Section 5 provides background in several relevant statistical topics, and Section 6 discusses a range of statistical methods for determining the bias and bias uncertainty. Section 7 provides methods for identifying and addressing gaps and weaknesses in the validation. Documentation of the validation is discussed in Section 8. Finally, Section 9 summarizes the content of this report.

ABBREVIATIONS AND ACRONYMS

AEF	average energy of neutrons causing fission
AEG	average neutron energy group causing fission
ANS	American Nuclear Society
ANSI	American National Standards Institute
BWR	boiling water reactor
C/E	calculation/experiment
CFR	US Code of Federal Regulations
CRC	commerical reactor critical [statepoint]
DICE	Database for the International Criticality Safety Benchmark Experiments
DOE	US Department of Energy
EALF	energy of the average lethargy of neutrons causing fission
ENDF	Evaluated Nuclear Data File
EVT	extreme value theory
HEU	highly enriched uranium
HMF	HEU-MET-FAST
HST	HEU-SOL-THERM
HTC	Haut Taux de Combustion
H/X	ratio of hydrogen atoms to fissile atoms
ICSBEP	International Criticality Safety Benchmark Evaluation Project
LEU	low-enriched uranium
LTL	lower tolerance limit
MCNP	Monte Carlo N-Particle
MOX	mixed-oxide (uranium and plutonium) fuel
NCSP	Nuclear Criticality Safety Program
NEA	Nuclear Energy Agency
NIST	National Institute of Standards and Technology
NPM	nonparametric margin
NRC	US Nuclear Regulatory Commission
ORNL	Oak Ridge National Laboratory
PWR	pressurized water reactor
QA	quality assurance
SNF	spent nuclear fuel
S/U	sensitivity and uncertainty
USL	upper subcritical limit (in other documents, USL is used as upper safety limit)
ZAID	nuclide identifier consisting of atomic number (Z) and atomic mass (A)

1 INTRODUCTION AND BACKGROUND

The purpose of criticality safety is to prevent any inadvertent criticality from occurring during the handling or storage of fissile material [1]. Calculations are frequently used to demonstrate that a sufficient subcritical margin exists. A key facet of the evaluation process is *validation*, which establishes the suitability of and determines the accuracy (i.e., bias) and the associated uncertainty of the computational method and data for the intended application. Validation is required by consensus standards [2], [3] which are endorsed by the US Nuclear Regulatory Commission (NRC) [4].

Several documents have been generated to support validation of criticality safety computational methods over a range of systems. In some cases, guidance has been developed to accommodate specific regulatory requirements. Many of these reports are targeted for specific types of applications, such as transportation of light-water reactor fuel. No single guidance document has been identified that is intended to apply to all types of systems while also incorporating both traditional and sensitivity/uncertainty (S/U)-based validation techniques. S/U-based validation techniques have been added to traditional validation techniques over the last 25 years and serve as additional tools to assist analysts in performing validation. The purpose of this report is to provide criticality safety computational method validation techniques for analyses involving all types of fissionable material operations. This is a work of synthesis, combining recommendations for acceptable validation approaches from their various sources. New recommendations are provided for areas in which existing guidance is vague, incomplete, or lacking.

This document was prepared to address typical validation issues identified from observations of previous validation efforts. These observations include (1) inappropriate critical experiment selection, (2) insufficient trending analysis, (3) incorrect application of bias and/or bias uncertainty, (4) failure to meet bias method prerequisites, (5) failure to identify validation gaps and weaknesses, and (6) inadequate documentation. This document was developed to augment previous criticality validation guidance documents and to provide recommendations to address some of the identified issues more effectively.

Section 2 discusses the purpose of validation, and Section 3 discusses the definition of a computational method. Section 4 provides suggestions for characterizing the safety analysis model, selecting critical experiments, and determining the area of applicability. Section 5 provides background in several relevant statistical topics, and Section 6 discusses a range of statistical methods for determining the bias and bias uncertainty. Section 7 provides methods for identifying and addressing gaps and weaknesses in the validation. Documentation of the validation is discussed in Section 8. Finally, Section 9 summarizes the content of this report.

2 THE PURPOSE OF VALIDATION

Validation establishes the applicability of a computational method to a particular safety analysis model (referred to as system or process model in some contexts), thus quantifying the suitability of the computer codes and nuclear data to a specific application. The validation process is performed by comparing the results of critical experiments with the calculated results from models of the experiments using the computational method to be validated [2], [3]. Laboratorycritical experiments (referred to as *critical experiments* for the remainder of the report) are controlled systems that achieve a k_{eff} of approximately 1. These experiments are used to investigate the parameters at which such a critical condition is achieved [1]. Thousands of critical experiments have been conducted, evaluated, and reported in the literature for validation. Currently, the most complete source of evaluated critical experiment descriptions is developed by the International Criticality Safety Benchmark Evaluation Project (ICSBEP) in the International Handbook of Evaluated Criticality Safety Benchmark Experiments, also known as the ICSBEP Handbook [5]. A key part of the validation process is the selection of experiments that are representative of the system or systems to be analyzed. The bias of the computational method is dependent on the materials in the model and the neutron energy spectrum in the system, so the selection of inappropriate experiments can lead to significant errors in the apparent bias for the system of interest. The experiment selection process is therefore carefully documented and reviewed during the validation process. All of the critical experiment models are also developed, documented, and reviewed to ensure that the bias is not impacted by modeling errors. This document focuses on the generation of the bias and bias uncertainty-not the more general topic of demonstrating the applicability of computational methods to criticality safety analysis. Well-established computational tools are generally used in analyses; the use of new or less commonly used tools would require a more complete demonstration of applicability.

For each experiment used in the validation, the difference between the calculated and evaluated k_{eff} results is determined. The experiments used in the validation are selected based on their similarity to the safety analysis model being analyzed. It is important to determine the bias for the system of interest, which is expected to be the same for similar systems. Various methods of assessing this similarity and selecting experiments are discussed in Section 4. Because the experiments are selected based on their expected similarity to the application, the application can be treated statistically as another member of the population of systems represented by the set of critical experiments. Thus, the estimate of the bias of the computational method used in modeling the experiments, also referred to as the computational bias, is determined as the mean of these differences. The bias should also be examined as a function of a system parameter using a trending technique, which is frequently linear regression. Several independent system parameters should be examined individually as part of this assessment. The system parameters used are typically a property of the fissionable material, such as enrichment, or they can be a property of the system, such as the neutron energy spectrum. Trending analysis may provide a more rigorous method to determine the bias and its uncertaintv.

Several factors contribute to the uncertainty in the bias estimate. First, there are uncertainties associated with each experiment. These can be measurement or dimensional uncertainties, the result of incompletely characterized materials, or other unknown or uncertain characteristics in the experimental materials or configuration. Secondly, there is also uncertainty in the bias estimate because it is the result of sampling a fixed set of experiments. There is also a computational uncertainty associated with calculating k_{eff} for the experiment models. Criticality safety calculations most frequently use methods involving Monte Carlo neutron transport, a

stochastic technique which therefore yields a k_{eff} value with some associated uncertainty. This calculation uncertainty is generally significantly lower than the experimental uncertainties. Deterministic methods have uncertainties associated with discretization of the problem geometry to conform to the required spatial mesh. Deterministic methods and some Monte Carlo implementations use multigroup representation of the energy variable, which may also contribute to the bias or its uncertainty. Multigroup cross section processing requires a flux solution for a representative simplified model. Differences between this model and the full system should be minimized but are typically unavoidable and are reflected in the bias of the multigroup computational method. The uncertainty in the bias is also increased to provide greater statistical confidence that the estimated bias and uncertainty bound the actual bias. The variance, which is a measure of the variability of the differences between measured and calculated results within the chosen set of experiments, is used in the overall determination of the bias uncertainty. The statistical margins often lead to the bias uncertainty being significantly larger than the bias itself, so proper quantification of uncertainty is essential.

Figure 2-1 provides a flow chart of the overall validation process. The first step, which is discussed in Section 4.1 below, is to identify the range of parameters in the safety analysis models. These characteristics identify the critical experiments needed for validation. This is followed by the selection of applicable benchmark experiments, which is described in Section 4.2. A variety of accepted statistical methods are used for performing validation based on the calculated results of the benchmark models, and several of these methods are described in Section 6. The final result of the statistical analysis is the bias and bias uncertainty of the computational method. The validation only applies to systems that are similar to the critical experiments used, so an area of applicability must be defined; recommendations for establishing the area of applicability for a validation suite are provided in Section 4.3. The bias and bias uncertainty values are combined with any reactivity allowances to account for gaps or weaknesses in the set of critical benchmark experiments and the margin of subcriticality, as discussed in Section 7. All of these data are used as discussed in Section 6.6 to establish the upper subcritical limit (USL). Finally, the validation activity must be documented and reviewed: recommendations for the documentation are provided in Section 8.

Verification is distinct from *validation. Verification* is the process of confirming that the algorithms used in the computational methods are coded correctly and functioning properly. This can be accomplished with a range of tests, including solving simplified problems with known solutions that are often analytical, processing known inputs to ensure that the expected outputs are generated, running inputs designed to fail, and other tests. Generally, the verification testing performed at installation is accomplished by executing a suite of test problems provided by the code developer and comparing the results to those provided. In addition to post-installation testing, the verification should be confirmed to remain valid after any changes to the computer operating system, and it should be rerun periodically to ensure that no unexpected changes have been made [4]. Both verification and validation must be performed and documented before the results from the computer code can be used in any safety assessment.



Figure 2-1 Flow Chart of the Validation Activity

3 DEFINITION OF COMPUTATIONAL METHOD

The validation process establishes the bias and bias uncertainty for a particular computational method. American National Standards Institute (ANSI)/American Nuclear Society (ANS)-8.24 [3], the consensus standard on validation, defines the *calculational method* as "The mathematical procedures, equations, approximations, assumptions and associated numerical parameters (e.g., cross sections) that yield calculated results." The standard uses a slightly different term—*calculational method* instead of *computational method*—because the standard applies to both hand calculation methods and computer code calculations. This document is primarily focused on validating computer codes.

It is necessary that important features of the computational method be the same for both benchmark and safety analysis model calculations to ensure that the bias and bias uncertainty generated during the validation is reflective of the bias and bias uncertainty expected when performing safety analysis calculations. This section discusses the important aspects of the computational method to be considered when making this determination. The components of the computational method include the code and cross section set, nuclear data, some input options, multigroup cross section processing techniques, variance reduction techniques, result selection, and the computer hardware and operating system. Each aspect of the computational method is discussed in further detail in this section.

The most obvious component of the computational method is the **computer code** being used. The validation applies only to the code and version installed; any updates, including patches, require that the applicability of the validation to the updated code be assessed. In most cases, a new validation is required following patch installation, and it is always required following a version update.

Other key aspects of the computational method are the cross sections and other nuclear data used by the computer code. These are often clearly delineated, but both SCALE [6] and the Monte Carlo N-Particle (MCNP®) code [7] allow user-specified cross sections to be loaded. For MCNP, this is accomplished with the extension on the ZAID of the relevant material card. SCALE users can load specific cross sections for continuous-energy calculations by specifying the file name (*fname*=) for an isotope in the composition block. Both codes have multiple thermal scattering data sets, also known as $S(\alpha,\beta)$ data. These data sets are available for many common materials. The thermal scattering data in SCALE are tied to the names specified in the composition block, whereas MCNP thermal scattering data are specified with the appropriate data card. Additional codes such as MONK [8] are used in some instances for licensing activities in the United States, and similar capabilities are available in these codes. An increasing number of cross section data treatments are provided as user options in recent releases of major computer codes. These options include Doppler broadening of cross sections to specific temperatures at time of execution, Doppler broadening resonance corrections especially important at elevated temperatures, and unionized or nuclide-specific energy points. In all instances, use should be reviewed carefully to ensure that the same nuclear data are applied as consistently as possible in the validation and application models.

Some code input options may need to be included in the computational method description and used consistently between the validation and application models. Many of these issues apply primarily to deterministic codes, and they include quadrature set and scattering order. The polynomial expression order for scattering cross sections is typically not a user option, but it will impact results and should thus be used consistently in both types of models. These items are

closely related to the cross section consistency requirements discussed in the previous paragraph.

Multigroup cross section processing is also an aspect of the computational method that should be as similar as possible between the validation and application models. The same type of processing used in the application models must be exercised in the validation models. For fuel rod arrays such as fuel assemblies, this is readily achieved because the validation cases will also involve arrays of rods. It may be a more subtle aspect of the computational method in process criticality models, in which different models may be appropriate for accurate modeling of different process cases. In some instances, it is impossible to achieve a complete match between cross section processing options, but this aspect should be included and addressed as a validation weakness. Validation gaps and weaknesses are discussed more fully in Section 7.

Variance reduction techniques are rarely used in Monte Carlo k_{eff} calculations. Two exceptions to this rule are (1) implicit capture and (2) rouletting and splitting. Implicit capture, or survival biasing, does not kill particles on absorption, but it does reduce their weight to account for the probability of capture. Rouletting and splitting are used to control the number of particles being tracked and to maintain a fairly even distribution of weights in the problem. The rouletting and splitting inputs can generally be changed by the user, whereas implicit capture is almost always applied to improve the code's calculational efficiency. Any variance reduction techniques that may be used must be included in the validation models and exercised in the same manner as in the application models.

Regarding **results selection**, most Monte Carlo codes report multiple estimates of k_{eff} as the final result of the calculation. Universally, the code developers have defined a preferred "final answer" that is the most correct result of the calculation. This result should be used, but various site procedures may dictate the use of a different value. The same method should be used to select the final k_{eff} estimate for the calculation for both validation and safety analysis cases.

The final aspect of the computational method discussed herein is the **computer hardware and operating system**. A validation only applies to the combination of computer hardware and software with which it is performed. This can be a challenging requirement given the frequency of operating system updates in contemporary computing, especially for purposes of cybersecurity vulnerabilities. It is incumbent on the code users to confirm that any system updates have not changed the operation of the computer code, thus rendering the validation inapplicable.

In summary, a wide range of hardware, software, inputs, and procedures can be viewed as aspects of the computational method, each of which must be identified and used consistently between the validation cases and the safety analysis cases to confirm that the bias and bias uncertainty determined in the validation are applicable to the safety evaluation being performed.

4 CRITICAL EXPERIMENT SELECTION AND AREA OF APPLICABILITY DETERMINATION

The primary function of criticality safety validation is to quantify the suitability of computational methods for specific fissionable material applications. The bias is the systematic difference between the calculated k_{eff} and benchmark k_{eff} . It may vary as a function of the materials present, the physical configuration and properties of the materials (i.e., density, chemical compound), and with environmental conditions such as temperature. Calculated k_{eff} values are biased estimates of real systems for a number of reasons, including:

- neutron transport method approximations such as meshing, scattering angle sampling, and calculation of $k_{\mbox{\scriptsize eff}},$
- nuclear data errors, including errors in cross section measurements, use of conflicting measurement data, and use of assumed data based on nuclear models, and
- approximations built into the data, such as reduction to multigroup data, resonance data descriptions, and use of $S(\alpha,\beta)$ thermal scattering data in continuous-energy calculations.

Variations in the materials used, their arrangement, and their physical properties all affect which nuclear data are important to the bias of the computational method. For example, different portions of the cross section set are used when the neutron flux energy and spatial distribution change. For modern, rigorously tested transport methods, the nuclear data errors have a greater impact on the bias than the transport method approximations. However, analysts should not assume that the contributions of the transport method to the bias are negligible.

In the area of critical experiment selection, the validation study documentation should explicitly describe:

- the process and criteria used to select critical experiments,
- the critical experiments selected,
- expected k_{eff} values and k_{eff} value uncertainties,
- references to critical experiment descriptions, and
- the critical experiments excluded and the justifications for their exclusion.

It is often efficient to approach a criticality code validation as a three-step process. The first step is to characterize the neutronically important features of the safety analysis applications to be analyzed. Secondly, critical experiments which are similar to the safety analysis models are selected, as well as the bias and bias uncertainty calculated using an appropriate method from Section 6. Thirdly, the characteristics of the suite of experiments selected should be documented as an area of applicability so its appropriateness can be judged for other applications. Further discussion of the documentation of the validation is provided in Section 8.

4.1 <u>Characterization of Safety Analysis Calculations</u>

Validation of the computational method is performed for specific safety analysis models. The characterization of the systems being modeled and the selection of critical experiments with similar features is essential for ensuring that the resulting bias and bias uncertainty are applicable. Both the normal and limiting abnormal conditions are relevant to making this

determination. For cases such as wet storage of fuel assemblies at reactor plants, the overall characteristics of the system and resulting models do not vary greatly between the normal and upset conditions. However, for cases such as process criticality assessments, the normal and credible upset conditions can vary greatly in terms of moderation, neutron energy spectrum, fissile mass, or other parameters. The importance of characterizing the safety analysis models is discussed in more detail in the remainder of this section.

The important characteristics of the safety analysis model are identified in Appendix A of the ANSI/ANS-8.24 standard [3] and are considered here, starting with the **materials present** in the system. The materials of interest must be present in the critical experiments conducted to ensure that a useful validation of the computational method is performed. The fissionable, moderating, and primary absorbing species are most important in this regard and must be represented in the validation suite. In many cases, minor constituents of the safety analysis system are not used in critical experiments, so they cannot be directly validated. Fission products in criticality analyses of burned fuel systems are a common example, but other examples exist in many other applications. Criticality analyses for burned fuel systems present an additional challenge in this area because the fissile and absorbing materials present are changing with burnup and cooling time. In an ideal validation, different experiments would be available to validate these different fissile compositions. Methods to address validation gaps—such as insufficient validation for the full range of materials present in the safety analysis case—are discussed in Section 7, and a review of burnup credit validation considerations is presented in Section 4.2.4.

It is also important to use experiments in validation with the same **neutron energy spectrum** as is present in the safety application models. A large number of different parameters are used as measures of the neutron energy spectrum, including (1) the energy of the average lethargy causing fission (EALF), (2) the average fission energy (AEF) or average fission energy group (AEG), (3) the moderation ratio (H/X), (4) or the pitch of fissile material units. The ICSBEP Handbook characterizes experiments using a scheme with three energy bins in which (1) the thermal range is defined as below 0.625 eV, (2) the fast range is above 100 keV, and (3) the intermediate range is between these two energies. Systems are categorized as fast, intermediate, or thermal if 50% or more of the fissions are caused by neutrons with energies in the associated range. Systems that do not have a majority of fissions in any single energy range are categorized as *mixed* [5]. This energy grid is generally sufficient for ensuring that the experiments' energy spectra used in validation are similar to the safety analysis models' energy spectra. Multiple validation analyses may be required if the normal and credible upset conditions of the application system differ significantly, especially with respect to moderation.

Another parameter identified as important for selecting benchmarks is the geometry of the system [3]. The important aspects of the **geometry** are homogeneity versus heterogeneity, size and spacing of reflectors, presence of interstitial absorbers or moderators, and the number and distribution of fissile units.

The characteristics of the safety analysis model are also important in trending analysis. The details of trending analysis are discussed in Section 6.3, but in general, these methods allow for the determination of the bias and bias uncertainty as a function of some independent parameter. The value of this independent parameter for the safety analysis model is used to determine the bias and uncertainty most relevant to the application. Therefore, different biases can be extracted for various application models from the same set of trended experiments.

The area of applicability of a validation suite, which is discussed more completely in Section 4.3, is also defined in terms of system characteristics. Each criticality safety evaluation must demonstrate that the validation suite being used is applicable to the safety analysis models used to demonstrate the safety basis for the operation. These characteristics include the material and spectrum metrics discussed earlier in this section, but they could also include reflection, physical form, geometry considerations, and multigroup cross section processing techniques. Many of these parameters should already be incorporated via the consideration of the materials present, but the form and placement of the material may also be important. For instance, thick steel reflectors may not provide adequate validation of thin interstitial steel panels typically included in fuel storage racks and baskets. A careful assessment of the characteristics of the safety analysis models will result in an effective, applicable validation.

Validation does not justify modeling approximations, poor characterization of safety analysis models, or failure to perform a valid transport calculation. Modeling approximations must be justified independently of the validation process, preferably with explicit calculations to demonstrate their acceptability. Similarly, documentation of the safety analysis model must include justification of parameters used, even if the exact values of one or more of the parameters are unknown. The validity of the transport calculation is separate portion of the safety analysis that must also be ensured. Additional margin in the validation cannot compensate for a poorly characterized system or inaccurate calculation.

4.2 <u>Selection of Critical Experiments</u>

4.2.1 Traditional Selection Criteria for Critical Experiments

For decades, criticality safety specialists have evaluated the relationship between calculated and actual k_{eff} values by performing criticality calculations for critical experiments with measured k_{eff} values. The relationship is typically quantified through development of a bias and its uncertainty or USL, either of which is used to ensure that k_{eff} values calculated for safety analysis models are subcritical. This analysis is frequently referred to as *validation* or as *a validation study*. Consistent with the requirements of ANSI/ANS-8.1 [2] and ANSI/ANS-8.24 [3], this validation is performed by comparing calculated results to measurements of real systems.

Critical experiments similar to the safety analysis models are selected for use in a validation study because code and data biases vary significantly as a function of target nuclide and incident neutron energy. Analysts must select critical experiments that they expect will have the same computational bias as the safety analysis model(s) or from which they can extract the correct bias through trend analysis. The bias and uncertainty developed from the selected set of such critical experiments are then considered applicable to the model of the application of interest; it is crucial to select critical experiments that are adequately similar to the application. Critical experiments would ideally not be used in the validation study if (1) they have extra materials that may significantly affect the bias, (2) they are missing materials that may significantly affect the bias, or (3) the critical experiment configuration causes the materials to be exposed to a neutron flux energy spectrum that is significantly different from the application(s).

In the absence of critical experiments designed to validate specific fissionable material systems, assessing the similarity of critical experiments to safety analysis models has generally been problematic. Historically, this has been accomplished using engineering judgement based qualitative comparisons, comparisons of key parameters, or comparisons of global figures of merit.

Criticality analysts use gualitative and guantitative information from the application of interest to select critical experiments from various reference documents. The ICSBEP Handbook contains detailed, independently reviewed descriptions of critical experiments. The Database for the International Criticality Safety Benchmark Experiment Project software, known as DICE [9], is also available. DICE is a searchable database of information characterizing the critical configurations described in the ICSBEP Handbook. Additional critical experiment description sources are available on the US Department of Energy (DOE) Nuclear Criticality Safety Program (NCSP) web page [10]. This site provides links to searchable databases of summary descriptions and references in which more complete experiment descriptions are available. Note that obtaining copies of the original reports listed in these databases may be difficult because of the age and limited distribution of the documents. Some reports may be obtained from one of the collections of reports held at various DOE sites or by searching the internet. For cases in which experiment descriptions are available in both the ICSBEP Handbook and in experimentspecific technical reports, the ICSBEP Handbook version should be used because it has been independently reviewed specifically for use in criticality safety validation studies. If critical experiment descriptions are taken from sources other than the ICSBEP Handbook, then the analyst must develop his or her own critical experiment models, generate expected k_{eff} values that include consideration of modeling approximations, and quantify experimental uncertainties. The ICSBEP Handbook critical experiment descriptions provide this information. Sample inputs from the ICSBEP Handbook should not be used directly without a thorough review to ensure that the models are correct. Any discrepancies identified must be corrected to match the benchmark model description prior to inclusion in the validation suite.

The analyst can evaluate similarity of a safety analysis model and an experiment in some scenarios by performing perturbation calculations. For example, if it is not clear whether an important neutron absorber is adequately similar in the critical experiments and safety analysis applications, then one might perform calculations with the density of the absorber varied by $\pm 5\%$ in the critical experiments and safety analysis application models and then compare the impacts on k_{eff} . The two systems would likely yield different biases associated with the perturbed material if the two models exhibit significantly different k_{eff} sensitivities to the variations.

In some cases, validating the range of variations in the safety analysis applications will require use of critical experiments that also have a range of parameters. In these cases, it is sometimes appropriate to use trending analysis, which may yield a range of bias and uncertainty values that can be applied to the range of safety analysis model results. This process is often referred to as *defining the area of applicability of a validation suite* and is discussed in more detail in Section 4.3.

In the traditional approach, the critical experiment set or sets used for the validation study are selected so that they have similar materials in similar geometries that result in similar energy-dependent neutron spectra. Ideally, critical experiments would be performed using materials in configurations intended to be very similar to the actual fissionable material operation, as was done for the Oak Ridge National Laboratory (ORNL) High Flux Isotope Reactor during its design and construction [11], [12], [13]. Unfortunately, few critical experiments simulating real fissionable material operations are performed today. Consequently, critical experiments must be selected that are as similar to the operation's safety analysis models as possible.

As discussed in Section 4.1, the typical process is for the analyst to gather characteristic data from the safety analysis application model and then to use the data to select critical experiments considered as similar as possible to the safety analysis application. Some frequently used comparison parameters are listed below:

- Uranium and/or plutonium isotopic distributions, and for mixed U+Pu systems, the U/Pu ratio
- Fuel pin pitch, fuel rod outer diameter, fuel bundle array size
- Solution fissile concentration and composition
- Dry compound composition and moisture content
- Neutron energy spectrum indices (e.g., EALF, AEG, AEF, etc.)
- Moderator ratios (e.g., H/U, H/²³⁵U, H/Pu—more generally referred to as H/X)
- Moderator density and temperature
- Neutron leakage fraction
- Geometry
- Similarity index (e.g., *c_k*)
- Reflector materials
- Neutron absorber material, composition, and geometry
- Structural materials

4.2.2 Use of Sensitivity/Uncertainty Analysis for Critical Experiment Selection

A technique using S/U analysis to select critical experiments was developed at ORNL [14] and implemented in the SCALE TSUNAMI code suite. In this technique, nuclide-, reaction- and energy-dependent k_{eff} sensitivity data are prepared for application models and critical experiment models. The sensitivity data are then combined with nuclear data uncertainty information to yield nuclide-, reaction- and energy-dependent k_{eff} uncertainty information for each critical experiment model and safety analysis model. A correlation coefficient, c_k , indicating the degree to which two systems have similar k_{eff} sensitivity to nuclear data uncertainties, is then calculated for each critical experiment–application pair. The definition of the integral index c_k can be found in Section 6.5.1.1.1 of the SCALE 6.2.3 manual [6]. A high correlation coefficient value (i.e., c_k approaching 1.0) indicates that the critical experiments and safety analysis models have sensitivities similar to those of nuclear data with significant uncertainties. This technique assumes that the most likely nuclear data bias sources in the criticality calculations are the nuclear data with the highest model-specific uncertainties. Both the SCALE TSUNAMI tools and the MCNP Whisper code use the c_k parameter to select applicable benchmarks for validation.

An introduction to the use of the SCALE TSUNAMI tools is available in a TSUNAMI Primer [15]. Use of this technique is facilitated by the availability of critical experiment sensitivity data for over 450 configurations that were generated by ORNL, as well as over 4,000 configurations generated for the Nuclear Energy Agency (NEA) [16]. These sensitivity data files can be obtained with the DICE database. It is acceptable to use these sensitivity data for screening critical experiments. Further use of the sensitivity data—such as use of c_k values as a trending or weighting parameter (see Section 6.3) or for generation of reactivity margins associated with validation gaps or weaknesses (see Section 7)—would require the sensitivity data files to be under a quality assurance (QA) plan. This can be accomplished by creating new sensitivity data within the requirements of a site's QA plan. It may also be possible to confirm the accuracy of the existing distributed sensitivity data with direct perturbation calculations that are performed and documented within requirements of a site's QA plan.

The MCNP Whisper code is discussed in Section 6.2.2. Whisper makes use of an MCNP data format for tabulating sensitivities and does not read SCALE-formatted sensitivity data, including that distributed by the NEA. Instead, a library of over 1,000 experiments is provided with MCNP via the distribution. These sensitivity data are, like the NEA data, not generated by the

organization performing the validation activity. Whisper uses the c_k values as a weighting function for individual experiments in its nonparametric statistical approach, as discussed in Section 6.2.2.

Using S/U methods for critical experiment selection provides a defensible, quantitative approach to critical experiment selection. Prior to the advent of S/U methods, analysts had no choice but to use key parameter comparison and engineering judgement. The primary weakness associated with the use of S/U methods is that it requires use of nuclear data uncertainty information distributed as cross section covariance data files. The covariance data are not as mature as the cross section data currently in use and are not generated using any consistent, standardized methodology. Consequently, covariance data changes have raised questions concerning the impact of covariance data on critical experiment selection. The apparent degree of similarity between experiments and an application can be significantly affected by these updates [17]. Additionally, analysts may be required to demonstrate validation of some key safety features such as credited neutron absorbers, even though the S/U method results may suggest otherwise. Fixed neutron absorbers in many systems have a small sensitivity because the strong absorbing properties make them essentially black to neutrons. Small changes in the cross section, such as those estimated with the sensitivity coefficient, have very small k_{eff} impacts, despite the importance of the reactivity hold-down provided by the fixed absorber. A range of different nuclear covariance data is available, and comparison of results with different libraries may prove useful in bolstering experiment selection choices. Other available similarity metrics, such as the E parameter in the SCALE code system, do not use the covariance data in comparing experiments and applications. The integral index E is defined in Section 6.5.1.1.3 of the SCALE 6.2.3 manual [6]. The independence of these metrics from the covariance data is both a strength and a weakness: the strength is that changes in the covariance data do not change the apparent similarity between an experiment and application, and the weakness is that the sensitivity data are not weighted by the uncertainties. Therefore, large sensitivities will dominate the similarity assessment, even for reactions which have small uncertainties and are therefore unlikely to contribute significantly to the bias. Also, published guidance on the use of E and other parameters is extremely limited.

Data adjustment methods attempt to develop a set of cross section adjustments to generate a consistent set of results from a large set of available benchmarks. The data adjustments are constrained by the nuclear covariance data, which are not as mature as the nuclear cross section data. Data adjustment also requires knowledge of the correlations among the benchmarks. As discussed in Section 6.5, there is no standard methodology for determining these correlations, and a wide range of correlation coefficients can be calculated from the same experimental evaluation by different analysts. Some questions also remain regarding the uniqueness of the solutions provided in these data adjustment methods, as well as the uncertainty that remains following the adjustment process. It is not clear how to process this residual uncertainty into a tolerance interval to assess bias uncertainty. Taken together, these difficulties preclude the use of data adjustment or assimilation methods as the primary method for determining bias and bias uncertainty in nuclear criticality safety validation at the time of writing. Developments addressing the shortcomings discussed above could allow for the adoption of such techniques in the future.

4.2.3 Number of Critical Experiments Selected

A validation analysis must consider the number of critical experiments to include and the bases for excluding some configurations. The number of critical experiments included may be informed by the requirements and desired results of the statistical analysis technique used to calculate bias and uncertainty and by the ranges of parameters characterizing the applications of interest. The maximum number of critical experiments to include in the validation set is most likely limited by the number of critical experiments that are similar to the safety analysis models and by the amount of time available for critical experiment modeling. Today, it is not unusual to use 10–100s of critical experiments in a validation study. It should be noted that larger validation suites are not necessarily better than smaller suites if the additional experiments included are less similar to the intended application or if they are highly correlated to each other. The impact of experimental correlations is discussed further in Section 6.5.

As noted in ANSI/ANS-8.24 [3] Section 5.6, "To minimize systematic error, benchmarks should be drawn from multiple, independent experimental series." Critical configurations performed using the same fissionable material, using the same apparatus, or being located at the same facility may have systematic errors. For example, if the same fissionable material is used in one or more series, then the critical experiment results (critical spacing, water level, concentration, etc.) will all include similar bias contributions related to errors in isotopic description of the material. This bias is then included in the computational method bias, which ideally would not include any bias related to experimental description. Using critical experiments that are independent minimizes the effect of systematic errors on the bias and uncertainty generated by the validation study. The clustering of results seen in different series of critical experiments indicates that there are correlations among the experiments and that different evaluations may indicate different biases. The impact of correlated critical experiments is discussed in more detail in Section 6.5.

The biases for some groups or individual critical experiments may vary significantly from the rest of the group(s). In some cases, this leads to a conclusion that the individual bias estimates cannot be characterized as being normally distributed. This is relevant because many statistical methods for defining confidence intervals assume that the distribution of the bias estimates is normal. Section 6.3.2 of ANSI/ANS-8.24 [3] provides the following guidance on rejection of critical experiments: "[R]ejection of outliers shall be based on the inconsistency of the data with known physical behavior in the experimental data." Analysts can reject some configurations within a series if there are valid technical reasons, such as lack of similarity to the application of interest, or results that are inconsistent with the other configurations in the series. Justification for removing some critical experiment results from a validation set must be carefully considered and documented. It is always inappropriate to remove results purely to obtain a less restrictive limit or to pass a normality test.

4.2.4 Critical Experiment Selection Considerations for Burnup Credit Validation

Validation of criticality calculations for spent nuclear fuel (SNF) systems has been and continues to be a technical challenge. Useful information on the validation of these systems can be found ANSI/ANS-8.27 [18], although the combined validation approach is not endorsed by Regulatory Guide 3.71 [4]. Ideally, criticality validation would be performed based on analysis of critical experiments with well-characterized fissionable material having compositions (i.e., actinides and fission products) similar to those of actual SNF. The fuel composition, which starts out as low-enriched uranium dioxide fuel, changes significantly as plutonium and fission products are created. Various critical experiment fissionable material compositions are needed to evaluate the change in bias and bias uncertainty with SNF burnup. The fundamental problem is that there are insufficient critical experiments involving either well-characterized SNF or fissionable material designed to be similar to SNF that also have actinide and fission product content similar to that of used fuel.

Whereas many technical challenges associated with validation of burnup credit are discussed throughout this document, a more thorough discussion is provided in NUREG/CR-7109 [19]. This subsection provides a discussion of the advantages and drawbacks of the Haut Taux de Combustion (HTC) critical experiments, the commercial reactor critical (CRC) statepoints, and uranium and plutonium mixed oxide (MOX) experiments from the ICSBEP Handbook. The HTC experiments are a series of critical experiments performed in France in the 1980s that incorporated an actinide mixture specifically designed to represent discharged SNF. A mixture of uranium and plutonium oxides was fabricated specifically for these experiments. The CRC statepoints are a series of power reactor hot zero-power or hot full-power critical conditions for which detailed information has been collected. These statepoints were developed and defined for use in burnup credit validation as part of the Yucca Mountain Project [20].

4.2.4.1 HTC Experiments

For validation of criticality calculations involving SNF, the HTC experiments will likely prove useful. These 156 critical configurations were assembled using fuel rods containing a mixture of enriched uranium and plutonium oxides designed to produce an actinide composition similar to that found in pressurized water reactor (PWR) fuel with an initial enrichment of 4.5 wt% ²³⁵U and a discharge burnup of 37,500 MWd/MTU. This is the only critical experiment set currently available with uranium and plutonium compositions very similar to spent fuel. An evaluation of these experiments is documented in NUREG/CR-6979 [21], and the experiments are described in four proprietary reports available from ORNL [22], [23], [24], [25]. These reports are available for limited approved purposes in the United States, contingent upon completion of a nondisclosure agreement with ORNL.

As stated in NUREG/CR-6979 [21], the "HTC experiments have many characteristics that make them very valuable for the validation of actinide-only burnup credit calculations." The fissile material is an extremely good match for the actinides present in SNF at typical discharge burnups. The distribution of the plutonium isotopes in particular is better than most of the mixedoxide critical experiments in the ICSBEP Handbook. The actinide content is a strength for validation of SNF at typical discharge burnups but is also a weakness for lower burnup assemblies. Damaged PWR assemblies or analyses involving boiling water reactor (BWR) assemblies at peak reactivity likely involve significantly lower burnups for which the HTC experiments are not very similar. Furthermore, the HTC rods do not contain fission products, which creates a validation gap for SNF analyses crediting these nuclides.

4.2.4.2 CRC Statepoints

It has been proposed that data from operating nuclear power plants could be used in the same way as critical experiments. Rather than using well-characterized fissile material in a critical experiment, the CRC statepoints would use well-characterized initial fuel and well-characterized in-reactor depletion conditions. To support this approach, CRC data have been accumulated for several nuclear power plants, and analysts have generated CRC models using these data. Because of challenges associated with this validation approach, no NRC-regulated license has yet been approved for which the use of CRC statepoints was the primary validation data source.

Two major challenges must be addressed before CRC statepoint data may be successfully used as a primary data source in the validation of criticality calculations for SNF systems. The first challenge is that the expected k_{eff} value and its uncertainty must be determined for the model of the CRC statepoint. The second challenge is that the analyst must address the differences between CRC statepoint models and the safety analysis models involving SNF

systems. The differences between CRC statepoints and SNF systems (e.g., temperatures and moderator densities) may indicate that these systems have different biases and uncertainties, so the analyst may need to determine and justify bias and uncertainty adjustments to cover extrapolation from CRC statepoint model-based validation to SNF system models.

The expected k_{eff} and uncertainty values are needed for comparison with the calculated k_{eff} value and uncertainty. Initially, one might think that the expected k_{eff} value should be exactly 1.0 because the reactor was critical. However, the assumptions, simplifications, and approximations made in generating the CRC statepoint models all contribute to the bias and uncertainty in the k_{eff} value expected from the CRC statepoint model. Numerous assumptions and approximations are typically made concerning the detailed, spatially, and temporally varying distributions of temperatures, power density, and moderator density. These quantities vary significantly during reactor operation, and the local distributions are not measured. The differences between the simplified model conditions and the actual reactor conditions could introduce bias and uncertainty. Uncertainties in the materials and dimensions of assemblies and other core components introduce additional uncertainty in the expected k_{eff} value.

If the challenges associated with (1) determining the expected k_{eff} value and its uncertainty for CRC statepoint models and (2) determining the adjustments that need to be made to the bias and uncertainty generated using CRC statepoint models can be overcome, then it may be possible to use CRC statepoints to validate criticality calculations involving SNF in storage and transportation configurations. This approach may have the added benefit of validating both SNF composition calculations and criticality k_{eff} calculations simultaneously. NRC staff have previously recommended against using the CRC statepoints for burnup credit validation in the Division of Spent Fuel Storage and Transportation Interim Staff Guidance 8, Revision 3 [26].

4.2.4.3 Mixed-Oxide Critical Experiments

Almost 300 MIX-COMP-THERM critical experiments are documented in the ICSBEP Handbook [5]; these or other experiments with mixed uranium and plutonium fissile material may prove useful for validation of SNF system criticality safety analyses. The applicability of some of these experiments for PWR SNF analyses was assessed and is presented in NUREG/CR-6979 [21], and it was also assessed for multiple PWR analyses and for one BWR SNF analysis as presented in in NUREG/CR-7109 [19]. The impact of updates to the SCALE 6.2 covariance data on the apparent applicability of these experiments is also examined in a paper by Marshall et al. [17]. The SCALE 6.2 covariance data are largely based on Evaluated Nuclear Data File (ENDF)/B-VII.1 evaluations. Results from previous assessments indicate that a limited number of experiments exceeded the *c*_k threshold of 0.8, but investigations using SCALE 6.2 revealed that none of the non-HTC MOX experiments exceeded the *c*_k threshold of 0.8 for the PWR SNF system analyzed. An analysis of the ENDF/B-VIII.0 covariance data [27] indicates that these evaluations indicate that some non-HTC MOX experiments are applicable for SNF system validation. Future covariance data updates will likely also impact the apparent applicability of MOX experiments for the validation of SNF criticality safety analyses.

4.2.4.4 Isotopic Validation

Although this document is dedicated to the validation of k_{eff} calculations, burnup credit calculations also require the validation of the SNF isotopic compositions used as inputs to the criticality calculations. Detailed guidance on acceptable methods for performing isotopic validation can be found in the literature [28] and [29]. One approach to an integrated validation of SNF isotopic compositions and k_{eff} calculations is described in [30].

4.3 Defining the Area of Applicability

Once the experiments have been selected to meet the validation needs of the application model or models, it is convenient to define the area of applicability, referred to as *validation applicability* in ANSI/ANS-8.24-2017 [3], so that future applications can be compared to the validation suite. The *area of applicability* is the breadth of the physically relevant parameters for a validation suite that can be used to judge how similar it would be to a given application. This section draws on Section 2.5 of NUREG/CR-6698 [31], Section 5 of NUREG/CR-6361 [32], and Appendix A of ANSI/ANS-8.24-2017 [3].

As in the initial selection of critical experiments, defining the area of applicability properly ensures that the bias generated by the validation suite reflects the bias that should be expected for the application or applications that fall within it. Having similar computational biases between the validation suite and the application is ensured by selecting critical experiments in which the same nuclear data are important to k_{eff} as the application of interest. Two important conditions necessary to ensure that similar biases are generated are (1) that the area of applicability sufficiently covers the important neutronic features of the application of interest, and (2) that there is no artificial reduction of the bias introduced by features present in experiments in the validation suite that are not present in the application. Further discussion of the potential challenges with regard to defining the area of applicability are discussed in Section 4.3.1.

The process of defining the area of applicability to ensure that the bias is appropriate for the application is inseparable from the selection of the benchmark experiments. A good approach to defining the area of applicability is to examine the neutronically relevant key parameters associated with analyzed normal and credible abnormal conditions and then compare those to the set of critical experiments. If the new application model fits within the set of critical experiments, then the bias and bias uncertainty are appropriate for its validation.

As with critical experiment selection, engineering judgement has historically been used to determine which parameters provide the most insight into the selection of appropriate experiments and the degree of deviation of those parameters from the application cases deemed acceptable in the critical experiments. The categories of physical parameters considered can be roughly subdivided into those related to the fissionable materials, those related to moderators and reflectors, those related to significant absorbers, and those related to the neutron energy spectrum within the system. Parameters considered for the fissionable materials should be the isotopic composition of the material (i.e., highly enriched uranium [HEU], low-enriched uranium [LEU], and ratios of Pu to U), the chemical form of the material (i.e., metal or oxide), the material concentration, and the material density. Parameters considered for the moderator and reflector should include chemical composition and density. Neutron cross sections of moderators and reflectors can vary substantially at thermal energy between chemical forms because of differences in the $S(\alpha,\beta)$ scattering data. The presence of any interstitial absorber materials should be considered, as well as whether that material is present in the fissile material, a liquid moderator, or as solid absorber plates or pins located within the system. Parameters considered to be representative of the neutron energy spectrum have included the EALF, AEG, and AEF, as well as ratios of moderating to fissile concentration (e.g., H/X). Ratios of the amount of moderating species to the amount of fissile species should not be used as trending parameters for validation of fuel lattices because the lumping of the fuel introduces spectral effects that may vary in unexpected ways. Useful recommendations on the allowed deviation of experiments from the parameters listed above can be found in Table 2.3 of NUREG/CR-6698 [31], which is adapted from Appendix E of LA-12683 [33]. The values of

deviations prescribed in Table 2.3 of NUREG/CR-6698 should be considered guidelines, and departure from these values may be acceptable with adequate justification.

Within the past 25 years, S/U methods have been introduced as a means to determine if the area of applicability is defined appropriately for an application. S/U techniques explicitly account for the contributions of each nuclide-reaction pair to the sensitivity of k_{eff} as a function of energy. The sensitivity profiles can be generated for the application and for a large number of candidate critical experiments. The sensitivity profiles from the application and the experiments can then be folded together to generate integral indices (e.g., c_k). Integral indices give a measure of how much sensitivity the underlying nuclear data shares between the experiment and the application. The integral index values that most closely approach 1 indicate the highest similarities. The cross section covariance data are propagated with the sensitivity profiles to generate c_k values. The inclusion of the covariance data weights the overlap in the sensitivity profiles between the experiment and the application by the estimated uncertainty in the nuclear data as a function of nuclide, reaction, and energy. Historically, experiments with values of c_k greater than 0.9 have been considered similar to an application, whereas experiments with c_k values between 0.8 and 0.9 have been considered marginally similar by ORNL [34], [19]. The NRC has previously considered values of c_k of greater than 0.95 to indicate "a very high degree of similarity" and values of c_k greater than 0.90 to indicate "a high degree of similarity" with regard to providing justification of a minimum margin of subcriticality for fuel cycle facilities [35]. c_k is not the only integral index, and S/U methods continue to be an area of active research. The values of c_k that indicate the acceptability of an experiment in the validation of any particular application should be judged on an application-specific basis. Higher values always indicate a higher degree of similarity between cases.

4.3.1 Extrapolation and Wide Interpolation

Two of the most common errors introduced when defining the area of applicability are (1) inappropriately extrapolating the validation bias, and (2) defining the area of applicability so broadly that the bias generated by the validation suite does not reflect any individual application being validated. Existing guidance suggests limiting extrapolation to less than 10% and interpolation to be over ranges of less than 20% [35], although this should not be viewed purely as a mathematical exercise, and underlying physical phenomena should be considered.

The first error, extrapolating the bias and bias uncertainty beyond the bounds of the range of the physically relevant parameters explicitly tested by the validation suite, assumes that the trend of the bias is not changing outside the area tested. At the very least, this introduces additional uncertainty into the validation process. It is possible that the data used by the application not covered by the validation suite are less well known or are in some way more flawed than the data tested by the validation suite.

When dealing with unique applications, it is possible that some feature of the application will require extrapolation of the validation suite. The first and most obvious recommendation is to add experiments to cover the unvalidated feature, provided that those experiments do not add features which are highly dissimilar to the application. Assuming that the analyst has already used a large fraction of the applicable experiments, it then becomes necessary to examine ways to extend the area of applicability. A method to extend the area of applicability is to use the tolerance band method discussed in Section 6.3.2. When the statistical bands are used to extrapolate bias and bias uncertainty, they are accounting for the quadratic behavior of the statistical uncertainty, but they are not accounting for the potential issues associated with the nuclear data that are not tested appropriately. One approach to further investigate the nuclear

data component is to introduce a series of small material composition perturbations to the application model to show that the variation in k_{eff} is smooth over the range of the extrapolation. The trend from the perturbations to the application model should be included in the bias assessment. It is valid to include the statistical bands, even if the slope of the line is statistically indistinguishable from 0. The range of the extrapolation of the area of applicability with this method should be less than 10% [31]. Other means of addressing validation gaps and weaknesses are discussed in Section 7.

A second common error in the definition of the validation suite that can lead to a nonconservative application of the bias is defining the area of applicability so broadly that the bias is diluted by experiments not relevant to potential applications. This practice can result in what is known as *wide interpolation*. Broad expansion of the area of applicability can result in a condition in which the bias for the entire validation suite is less restrictive than a subset of the suite which is more appropriate to the validation of a particular application. The consensus validation standard [3] contains an explicit warning regarding this practice.

An overly broad definition of the area of applicability leads to the inclusion of experiments which use very different cross sections. A common example of this type of error is to attempt to define the area of applicability to include all potential applications that may be present at a facility. An example of this could occur at a facility that processes HEU. It is possible that HEU would be present in the form of metal ingots as initial feed to a process or as the final product. It is also possible that the process would include steps in which the HEU would be present in solution form. In such a situation, one might consider including both HEU metal and solution experiments so that one validation would produce a bias that could be applied to all of the criticality calculations performed at the site. However, this approach is problematic because the cross sections used by each calculation are vastly different and are known with different degrees of certainty. To illustrate this difference, fission cross section sensitivity profiles are shown from one HEU-SOL-THERM (HST) solution experiment and one HEU-MET-FAST (HMF) metal experiment in Figure 4-1. Figure 4-1 shows that there is virtually no overlap between the portions of the fission cross section used between the two calculations. A bias present in the fast energy portion of the cross section would impact the k_{eff} of a fast system, but it would not impact a thermal system, or vice versa.


Figure 4-1 Comparison of the Fission Cross Section Sensitivity Profiles of HMF-024-001 and HST-014-001 Showing the Lack of Overlap of the Sensitivity Profiles Between Fast and Thermal Systems

To further illustrate the issues caused by widely defining the area of applicability, a plot of many HST and HMF experiments is constructed in Figure 4-2. The unweighted bias would be -0.00133 Δk_{eff} for the entire validation suite, -0.00380 Δk_{eff} for the HST experiments only, and 0.00124 Δk_{eff} for the HMF experiments only. This example shows that if the analyst were to try to validate a solution application with a validation suite that included all of the experiments, then the bias would be underpredicted compared to a validation suite that included only the HST experiments. Another example of problems associated with wide interpolation would be the attempted validation of an HEU solution application with heavy water rather than light water as the primary moderating species. A heavy water HEU solution experiment with an EALF of 13.42 eV, which lies between the EALF values of the HST and HMF experiments, is also shown in Figure 4-2. The bias predicted by an unweighted trend line associated with this EALF is -0.00216 Δk_{eff} , but the true bias for the selected experiment is -0.01252 Δk_{eff} . The magnitude of this misprediction, over 1% Δk , illustrates that it is possible to make errors with large impacts on reactivity margins by improperly defining the area of applicability.



Figure 4-2 Comparison of HST and HMF Systems with an HEU D_2O Application as a Function of EALF

5 STATISTICAL BACKGROUND

A significant amount of statistical terminology is used in validation methods, and a good understanding of the relevant statistics is important to properly perform a validation. One common area of misunderstanding is hypothesis testing. Therefore, Section 5.1 provides the background on hypothesis testing necessary to ensure understanding of its effective use in the relevant validation sections that follow. The assessment of normality is important when confirming the applicability of several statistical techniques and is discussed in Section 5.2. Goodness-of-fit testing for trends is important to establish that a trend is statistically meaningful, as discussed in Section 5.3. These concepts are addressed in this section so that the validation techniques can be discussed with appropriate statistical rigor without complicating the descriptions with this background information.

5.1 <u>Hypothesis Testing</u>

This section discusses the basic nomenclature and framework that apply to all types of hypothesis testing so that it can be referenced during more specific discussions in later sections. The need for statistical hypothesis testing arises within the context of this document when testing the assumption of normality (Section 5.2) and assessing the statistical validity of a trend (Section 5.3). Additional information related to hypothesis testing is available in most statistics textbooks or in Chapter 10 of NUREG-1475 [36].

When a hypothesis is being tested, the null hypothesis (H_0) is the presupposed condition that may be rejected. The null hypothesis is either successfully or unsuccessfully rejected in favor of the alternative hypothesis (H_A) based on a statistical procedure and selected values of statistical parameters. For example, in normality tests (see Section 5.2.2), the null hypothesis is that the data are drawn from a normal distribution, and the alternative hypothesis is that the data are drawn from a distribution that is not normal. In hypothesis testing, it must be accepted that there is some probability of error, because the decisions are being based on a sample rather than the entire population, and the sample may not behave like the entire population.

In general, there are two kinds of errors than can be made:

- The error of inadvertently rejecting a true null hypothesis, known as a *false positive*, or *Type I error*. The symbol most commonly associated with the probability of committing a Type I error is α.
- The error of inadvertently accepting a false null hypothesis is known as a *false negative* or *Type II error*. The symbol most commonly associated with the probability of making a Type II error is β.

The probability of avoiding a Type I error is known as the *confidence level* and is calculated as $1 - \alpha$. The probability of avoiding a Type II error is known as the *power of a hypothesis test* and is calculated as $1 - \beta$. A summary of this discussion is provided in tabular format in Table 5-1.

Table 5-1 Summary of Possible Results of Hypothesis Test

	<i>H</i> ₀ is true	<i>H</i> ₀ is false
<i>H</i> ₀ rejected	Type I error (false positive) has a probability of α	Correct rejection has a probability of <i>1-</i> β (power)
H_0 accepted	Correct acceptance has a probability of <i>1- α</i> (confidence)	Type II error (false negative) has a probability of β

All hypothesis tests considered in this document are posed in terms of α or the probability of inadvertently discarding a correct null hypothesis. As α increases, β decreases, although the exact numerical relationship cannot be calculated. In other words, as the probability of incorrectly rejecting a true null hypothesis increases, the probability of failing to reject a false null hypothesis decreases (or the ability to detect a faulty one increases). This indirect relationship is explored in Section 5.2.3 for common normality tests. An alternate means of implementing hypothesis testing that is often used in modern statistical software is to calculate the probability that the value of the test statistic for the sample could occur by chance if the null hypothesis were true, known as the *p*-value. The p-value can be compared to a predefined value of α . A p-value smaller than α indicates that it is unlikely that the sample could have occurred by chance if the null hypothesis were true, so the null hypothesis should be rejected.

A common statistical misconception is that passing (failing to reject the null hypothesis) a hypothesis test at a given confidence level (1- α) implies that there may only be an α probability that null hypothesis is true. This is different from the true statement that the test has an α probability that the test inadvertently rejected a true null hypothesis.

5.2 Assessment of Normality

The assessment of normality is important during validation because the statistical tolerance limits used to calculate the bias uncertainty are sensitive to departures from normality. When used properly, the tolerance limits ensure that an appropriate fraction of the true population of applicable critical experiments lies above the calculated lower tolerance limit (LTL) with the required statistical confidence level. One of the conditions necessary to ensure that the appropriate proportion of the population of $k_{\rm eff}$ values in the validation suite lies above the LTL is that the assumption of the normality of the underlying population of critical experiments is valid or conservative. This is because the LTL is determined so that the specified proportion of the population is in the lower tail of a normal distribution. The population distribution for the normal distribution is known, so the number of standard deviations (K from Section 6.1) necessary to achieve the desired population fraction is also known. For other distributions, these factors may not capture the desired fraction of the population. Therefore, if it is not possible to show that the distribution from which the validation suite is drawn is normal or conservatively bounded by the normal distribution, then other statistical techniques (typically the nonparametric techniques discussed in Section 6.2) must be used to develop the LTL. In this context, it is conservative if the normal distribution has a larger proportion of the population in the lower tail than the distribution of critical experiments. This section discusses techniques used to assess the validity of the assumption that the validation suite is drawn from a normal distribution.

Before discussing normality assessment, the ways in which a distribution can depart from normality must be addressed. Distributions can depart from normality in two meaningful ways: (1) in the symmetry about the mean, or the skewness, and (2) the peakedness of the distribution, or kurtosis. As shown in Eq. (1), the value of the estimated skewness [37] of a sample describes the direction in which an unusually long tail is located relative to the mean. The normal distribution has a skewness of 0. Negative values indicate that the distribution has a heavy left tail, and positive values indicate that the distribution has a heavy right tail. The value of the calculated kurtosis of a distribution describes how broad the distribution is compared to a normal distribution. As shown in Eq. (2), kurtosis can either be expressed as absolute kurtosis [37], which has a value of 3 for a normal distribution, or excess kurtosis, which is calculated by subtracting 3 from the absolute kurtosis. A kurtosis of less than 3 indicates that the distribution is broader than normal and is termed *platykurtic*, and a kurtosis of greater than 3 indicates that the distribution is narrower than normal and is termed *leptokurtic*. Rigorously, kurtosis relates to the weight of the distribution tails, but either platykurtic or leptokurtic distributions may have heavier or lighter tails than the normal distribution. Figure 5-1 shows examples of distributions with positive and negative skewness (left) and leptokurtic and platykurtic distributions (right) compared to a normal distribution. Departures from normality that result from a combination of skewness and kurtosis also occur, and although they are ignored here for simplicity, they are considered later where appropriate.

$$S_{k} = \frac{\frac{\sum_{i=1}^{N} (k_{i} - \bar{k})^{3}}{N}}{\sigma_{\bar{k}}^{3}}$$
(1)

$$K_{u} = \frac{\frac{\sum_{i=1}^{N} (k_{i} - \bar{k})^{4}}{N}}{\sigma_{\bar{k}}^{4}}$$
(2)



Positively and negatively skewed distributions compared to standard normal distributions

Leptokurtic and platykurtic distributions compared to the standard normal distribution



Within the context of a criticality safety validation, not all departures from normality result in a nonconservative estimate of the bias uncertainty. As discussed in Section 6.1, the bias uncertainty for a nontrending analysis is calculated by multiplying the standard deviation of k_{eff} by the K factor. The factor provides coverage of a specified proportion of the population (typically 95%) with a specified statistical confidence level (usually 95% confidence). Distributions which result in a smaller fraction of the population below that value will yield conservative validation parameters. For example, if the kurtosis of the distribution exceeds that of a normal distribution, then less of the population could be in the tails than would be for the normal distribution. The bias uncertainty would therefore be overestimated when normality-based tolerance limit methods are applied. Applications of normality-based statistics to positively skewed distributions are similarly conservative because there are fewer k_{eff} values that deviate on the low reactivity side of the distribution. Conversely, distributions that result in more data in the negative tail than would be predicted by a normal distribution would lead to a nonconservative application of the LTL methods.

The remainder of this subsection addresses the methods used to assess whether statistical techniques based on a normal distribution are appropriate or conservative for the development of the LTL. The techniques discussed in the remainder of this section can be broadly categorized as *graphical techniques* or *formalized normality tests*. Graphical techniques are discussed in Section 5.2.1, and formalized normality tests such as omnibus normality tests and single parameters tests for skewness and kurtosis are discussed in Section 5.2.2.

5.2.1 Graphical Techniques

Two main methods are used to graphically assess whether a sample was drawn from a normal distribution. The first method uses a histogram on which a normal curve has been superimposed. A histogram allows for visual inspection of the general features of the data sample and provides a good first look at the data. Histograms are also easy to construct in modern statistical software packages such as Dataplot [38] and R [39]. Standardizing the data by converting it to Z-scores (number of standard deviations a data point is from the mean) can help in the evaluation. The Z-score places the data on the same scale as a normal distribution, with a mean of 0 and a standard deviation of 1. The Z-score can be calculated for each data point in the sample using Eq. (3),

$$Z_i = \frac{k_{norm_i} - \overline{k}_{norm}}{\sigma_{\overline{k}_{norm}}},\tag{3}$$

where: k_{norm} are the normalized k_{eff} values,

 \overline{k}_{norm} is the average of the normalized k_{eff} values, and $\sigma_{\overline{k}_{norm}}$ is the standard deviation about \overline{k}_{norm} .

Equations to calculate the values for these three variables are provided in Section 6.1.

The second graphical normality assessment technique is the *quantile-quantile plot*, or the *Q*-*Q plot*, which provides a mechanism to evaluate whether the data might be drawn from a normal distribution. Many software packages implement the Q-Q plot: the one presented here is taken from R and requires no alteration to the default settings. The process to implement the Q-Q plot is as follows.

- 1. Sample data are ordered from smallest to largest. The result is referred to as *ordered data*.
- 2. Ordered data are plotted against the appropriate quantiles from a normal distribution, creating a scatterplot of the sample data, or *sample quantiles*, where the *x* values are the theoretical quantiles, and the *y* values are the sample quantiles.
- 3. If the points fall along a straight line (y=x), then this indicates that the data are likely drawn from the normal distribution. Departure from linearity indicates a deviation of the sample from the normal distribution.

To demonstrate Q-Q and histogram plots for the normal distribution vs. other non-normal distributions, 200 data points were sampled from various distributions using the R software package. The data from these samples are presented in Figure 5-2 through Figure 5-6. Figure 5-2 shows data sampled from a normal distribution. As expected, the results from a normal distribution fall reasonably close to the superimposed line, with little deviation of any of the points. Figure 5-3 and Figure 5-4 show how skewness would manifest itself in a Q-Q plot. Figure 5-3 shows negatively skewed data, which produces values that are more extreme in the left tail and less extreme in the right tail than expected from data sampled from a normal distribution. This results in a downward-shaped C curve that departs from the superimposed line. Conversely, Figure 5-4 shows a Q-Q plot based on the data sampled from a positively skewed distribution. The Q-Q plot in Figure 5-4 shows that there are less extreme data than expected in the left tail of the sample and more extreme data than expected in the right tail of the data. These plotted data result in a characteristic upward-shaped C curve compared to the superimposed line. Having less data in the negative or left tail of the distribution than would be present in a normal distribution would result in a conservative application of the single-sided lower tolerance band, whereas having more data in left tail would result in a nonconservative application of normal statistics.

Figure 5-5 and Figure 5-6 show the effects of kurtosis on the Q-Q plot. Figure 5-5 shows a histogram of a leptokurtic distribution, along with the corresponding Q-Q plot. The Q-Q plot shows that the values of the Z-score in each tail are less than expected from a normal distribution. This results in values being lower than the superimposed line in the left tail and above the line for values in the right tail. Figure 5-6 shows a histogram of a platykurtic distribution, along with the corresponding Q-Q plot. The Q-Q plot in Figure 5-6 shows that the values of the Z-score in each tail are greater than expected, resulting in values greater than expected in the left tail.

In practice, it is often difficult to make definitive judgements solely using graphical techniques, but they can be used to identify potentially nonconservative departures from normality. Distributions often depart from normality, resulting from a combination of skewness and excess kurtosis. However, showing that the left tail of the distribution contains a smaller fraction of the data than would be present in a sample from a normal distribution provides some confidence that the application of normality-based statistics would be conservative. These graphical methods should always be combined with either omnibus or single-parameter tests for the normal distribution.



Figure 5-2 Histogram (Left) and Normal Q-Q Plot (Right) of a 200-Point Sample Drawn from a Normal Distribution



Figure 5-3 Histogram (Left) and Normal Q-Q Plot (Right) of a 200-Point Sample Drawn From a Negatively Skewed Distribution



Figure 5-4 Histogram (Left) and Normal Q-Q Plot (Right) of a 200-Point Sample Drawn from a Positively Skewed Distribution



Figure 5-5 Histogram (Left) and Normal Q-Q Plot (Right) of a 200-Point Sample Drawn from a Leptokurtic Distribution



Figure 5-6 Histogram (Left) and Normal Q-Q Plot (Right) of a 200-Point Sample Drawn from a Platykurtic Distribution

5.2.2 Normality Tests

This section discusses statistical tests used to determine if the data sample being investigated might come from a normal distribution. There are two types of normality tests discussed here: omnibus tests and single-sided, single-parameter tests for skewness and kurtosis. Omnibus tests provide a single statistic that measures the goodness-of-fit of a normal (or potentially other) distribution to the sample. The test statistic is calculated for the sample and compared with tabulated values of the test statistic distribution for the specified confidence level and the number of points in the sample (degrees of freedom). Based on the relative values of the calculated and tabulated statistics, the analyst either rejects or fails to reject the assumption of normality. The single-sided tests for skewness and kurtosis similarly develop test statistics for comparison to a distribution based on the number of points in the sample and the desired level of confidence; however, these tests individually check for the skewness and kurtosis of a normal distribution. The single-sided tests can provide more information than the omnibus tests regarding what potential departures from normality may exist, and they may be more powerful at detecting those departures. It is noted that modern statistical software will generally calculate a p-value for these, which is the level of confidence that would be required to reject the assumption of normality. The p-value is typically compared to the prescribed level of confidence to decide if the assumption of normality is appropriate.

All normality tests are posed with the null hypothesis (as previously discussed in Section 5.1) that the sample came from a normal distribution, with the alternative being that the data were not drawn from a normal distribution. This is formally stated below.

 H_0 : The data are normally distributed.

H_A : The data do not follow a normal distribution.

Because the null hypothesis is posed in the affirmative, the test requires that there be a preponderance of evidence to reject normality. Historically the 95% confidence level is used for most statistical applications, including criticality safety. Because the use of normal statistics generally results in a smaller bias uncertainty, σ_{β} when compared to the nonparametric methods discussed in Section 6.2, this may allow for a nonconservative validation if not closely examined. As discussed in Section 5.1, the p-value for the normality test is *not* the probability that H_{A} is true.

Omnibus normality tests are discussed in Section 5.2.2.1, single-sided single parameter normality tests are discussed in Section 5.2.2.2, and a power comparison of normality tests is presented in Section 5.2.3.

5.2.2.1 Omnibus Tests

Several generally accepted omnibus normality tests are used in practice. Table 5-2 lists some of the generally acceptable tests, along with the references providing information on how to calculate test statistics and p-values. The methods are not presented here because statistical software can often generate these values with minimal user input. It is also noted that the list of tests presented here is not an exhaustive list of normality tests, and there are other tests that may be acceptable. All tests specified in Table 5-2 are available in the "nortest" package in R [40].

Test	Reference
Chi-square	Pg. 106 [41]
Anderson-Darling	Pg. 104 [41]
Cramer-Von Mises	Pg. 103 [41]
Lilliefors	Pg. 102 [41]
Shapiro-Wilk	Pg. 27 [41]
Shapiro-Francia	Pg. 29 [41]
Jarque-Bera	[42]

Table 5-2 List of Generally Acceptable Normality Tests

Two potential issues are associated with using omnibus normality tests, as the normal assumption relates to criticality safety analyses. The first is that they may fail to reject normality at the traditional 95% confidence level for small sample sizes [43], [44]. The second issue associated with omnibus normality tests is that they may reject normality for samples which violate normality in ways that will not result in an underestimate of the bias uncertainty. The Kolmogorov-Smirnov test, a popular normality test, is cited as having particularly poor results for applications sensitive to the tails of a distribution [36], [44]. One such application is the calculation of single-sided LTLs discussed in Section 6.1. For this reason, the Kolmogorov-Smirnov test should be avoided for criticality safety validation applications. The Chi-square test for normality is unique relative to the other tests described here because it requires a means for

grouping the data. It is important to note the sensitivity of the test depends on the choice of the number and width of the groups [36], as well as the confidence level and sample size.

5.2.2.2 Single-Sided, Single-Parameter Tests for Normal Skewness and Kurtosis

As previously discussed, it is possible to fail the omnibus normality tests discussed in Section 5.2.2.1 for samples that would overestimate the bias uncertainty. This section presents methods that provide single-sided tests to determine if the skewness estimated by a sample is less than or statistically equal to 0 and if the kurtosis is greater than or statistically equal to 3 (or an excess kurtosis of 0). The test used to determine acceptable skewness is given by D'Agostino [45], and the test for acceptable kurtosis is given by Anscombe and Glynn [46]. Both the D'Agostino skewness test and Anscombe-Glynn kurtosis test are available in the R moments package [47]. As discussed in Section 5.2, distributions with excess kurtosis and/or positive skewness are conservatively represented with a normality-based LTL.

5.2.3 Power Comparison of Normality Tests

As previously noted, the tests for normality used here are posed so that the null hypothesis is that the data are consistent with a normal distribution. Because of the way the test is posed, the confidence level specified is the allowed error rate of rejecting the assumption of normality when the data are in fact drawn from a normal distribution. This does not directly indicate the probability that samples drawn from a non-normal distribution will pass the test. Rather, the statistical power of the test indicates how likely it is to detect non-normal behavior. Power studies documented by Razali et al. [44] test a few normality tests with some extreme distributions.

Power studies performed for this work use unimodal distributions based on the perturbable normal distribution put forward by Jones and Pewsey [48]. The Jones and Pewsey distribution allows the user to enter factors that will transform the normal distribution into one that has different skewness and kurtosis. The normality tests discussed above are exercised here on distributions with negative skewness, negative excess kurtosis, and a combination of both. The distributions were produced so that they would result in 90% and 85% coverages of the population, respectively. This means that 90% of the data points in the first case and 85% of the data points in the second case lie above the point that is 1.645 standard deviations below the mean. This value was selected because it corresponds to 95% of the population for a normal distribution. Plots of these distributions are shown in Figure 5-7. For the 90% case, these deviations from normality would result in a 26.1% underprediction of the single-sided tolerance factor for the purely skewed case, a 32.3% underprediction for the purely kurtotic case, and a 28.7% underprediction for the mixed case. For the 85% case, the skewed case would result in a 50.5% underprediction in the mixed case.

Using the distributions described above, 1,000 samples of 50, 125, and 250 points were randomly generated for each of the conditions for the 90% and 85% population coverage distributions, as well as for a normal distribution. The samples were then subjected to all the normality tests discussed in Section 5.2.2.1 with varying values of α (0.05, 0.10, 0.15, and 0.20). For each combination of sampled distribution, value of α , and sample size, the probability that the distribution would be rejected was calculated by dividing the number of cases for which the p-value was less than α by 1,000. The probabilities of rejection for the samples drawn from a distribution that would cover 90% of the population with normal statistics are given in Table 5-3,

the probabilities of rejection for the 85% of the population case are given in Table 5-4, and the probabilities of rejection for the normal distribution are given in Table 5-5.



Perturbed distributions offering 90% coverage of the population under the application of normal statistical one-sided tolerance factors.

Perturbed distributions offering 85% coverage of the population under the application of normal statistical one-sided tolerance factors.

Figure 5-7 Comparison of Platykurtic (Black), Negatively Skewed (Red), and Combined Platykurtic and Negatively Skewed (Orange) Distributions Used in the Power Study to the Normal Distribution (Green) for 90% and 85% of Population Coverage

In general, the results in Table 5-3 and Table 5-4 show that the omnibus tests perform similarly when compared with one another. The results also show that all of the tests do a better job of rejecting distributions that depart from normality through skewness rather than kurtosis. The results in Table 5-5 show that the omnibus tests reproduce the values of α effectively for samples drawn from a normal distribution.

This information shows that graphical techniques should be used to supplement quantitative normality testing because of the potentially low power of common statistical tests at small sample sizes and low values of α .

Kurtosis												
Sample size		5	0		125				250			
Test/α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Anderson-Darling	0.07	0.16	0.22	0.28	0.12	0.20	0.28	0.34	0.21	0.30	0.38	0.46
Cramer–Von Mises	0.06	0.14	0.20	0.27	0.12	0.19	0.25	0.30	0.19	0.28	0.36	0.43
Lilliefors	0.06	0.13	0.18	0.25	0.08	0.15	0.21	0.26	0.13	0.24	0.33	0.40
Shapiro-Wilk	0.09	0.16	0.22	0.28	0.13	0.22	0.28	0.34	0.20	0.30	0.36	0.43
Shapiro-Francia	0.12	0.20	0.27	0.32	0.17	0.26	0.34	0.42	0.24	0.36	0.43	0.52
Jarque-Bera	0.10	0.14	0.17	0.20	0.17	0.22	0.26	0.32	0.26	0.33	0.39	0.44
	Combination of skewness and kurtosis											
Sample size	50					12	25			25	0	
Test/α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Anderson-Darling	0.09	0.16	0.23	0.28	0.16	0.24	0.30	0.37	0.25	0.35	0.44	0.50
Cramer–Von Mises	0.09	0.16	0.21	0.27	0.15	0.24	0.30	0.36	0.23	0.33	0.41	0.48
Lilliefors	0.08	0.15	0.19	0.25	0.13	0.21	0.28	0.34	0.18	0.31	0.39	0.44
Shapiro-Wilk	0.11	0.17	0.23	0.27	0.16	0.24	0.32	0.38	0.28	0.39	0.46	0.51
Shapiro-Francia	0.11	0.18	0.24	0.29	0.18	0.27	0.34	0.40	0.30	0.41	0.49	0.54
Jarque-Bera	0.09	0.13	0.14	0.17	0.17	0.23	0.29	0.34	0.28	0.38	0.43	0.50
				Sk	kewnes	S						
Sample size		5	0		125					25	0	
Test/α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Anderson-Darling	0.12	0.20	0.27	0.33	0.23	0.35	0.42	0.49	0.48	0.59	0.68	0.72
Cramer–Von Mises	0.11	0.20	0.26	0.32	0.22	0.33	0.40	0.48	0.45	0.57	0.64	0.70
Lilliefors	0.11	0.18	0.25	0.31	0.18	0.28	0.35	0.43	0.36	0.50	0.59	0.64
Shapiro-Wilk	0.13	0.22	0.29	0.36	0.26	0.38	0.46	0.51	0.53	0.64	0.70	0.74
Shapiro-Francia	0.14	0.22	0.28	0.35	0.24	0.37	0.44	0.50	0.50	0.61	0.69	0.73
Jarque-Bera	0.09	0.14	0.17	0.20	0.19	0.28	0.36	0.43	0.46	0.57	0.65	0.70

Table 5-3Probability of Rejection for the 90% Coverage of Population Cases

Kurtosis												
Sample size		5	0		125				250			
Test/α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Anderson-Darling	0.15	0.24	0.32	0.38	0.32	0.44	0.53	0.60	0.61	0.72	0.77	0.81
Cramer–Von Mises	0.15	0.23	0.32	0.39	0.30	0.43	0.52	0.59	0.59	0.70	0.76	0.80
Lilliefors	0.12	0.21	0.29	0.35	0.24	0.36	0.45	0.52	0.42	0.59	0.68	0.74
Shapiro-Wilk	0.13	0.20	0.27	0.34	0.27	0.38	0.45	0.51	0.51	0.62	0.70	0.75
Shapiro-Francia	0.17	0.26	0.36	0.42	0.32	0.45	0.53	0.59	0.56	0.68	0.75	0.80
Jarque-Bera	0.12	0.16	0.20	0.25	0.26	0.33	0.39	0.44	0.44	0.56	0.64	0.70
Combination of skewness and kurtosis												
Sample size	50			125				250				
Test/α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Anderson-Darling	0.19	0.30	0.37	0.43	0.41	0.53	0.61	0.66	0.69	0.77	0.83	0.90
Cramer–Von Mises	0.19	0.29	0.36	0.42	0.38	0.50	0.57	0.64	0.65	0.75	0.80	0.88
Lilliefors	0.17	0.26	0.33	0.39	0.31	0.44	0.51	0.58	0.54	0.66	0.72	0.87
Shapiro-Wilk	0.21	0.31	0.37	0.44	0.44	0.55	0.61	0.66	0.70	0.79	0.85	0.87
Shapiro-Francia	0.23	0.33	0.41	0.46	0.45	0.58	0.63	0.69	0.71	0.82	0.86	0.90
Jarque-Bera	0.18	0.23	0.28	0.32	0.40	0.49	0.57	0.62	0.68	0.78	0.83	0.85
				Skew	ness					-		
Sample size		5	0			12	25			25	50	
Test/α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Anderson-Darling	0.29	0.42	0.49	0.86	0.64	0.76	0.82	0.86	0.94	0.97	0.98	0.99
Cramer–Von Mises	0.28	0.39	0.47	0.85	0.59	0.71	0.79	0.84	0.91	0.96	0.97	0.98
Lilliefors	0.21	0.31	0.40	0.81	0.48	0.61	0.69	0.77	0.82	0.90	0.94	0.96
Shapiro-Wilk	0.32	0.44	0.53	0.87	0.68	0.78	0.83	0.87	0.96	0.98	0.99	0.99
Shapiro-Francia	0.30	0.43	0.51	0.88	0.66	0.76	0.82	0.86	0.94	0.97	0.99	0.99
Jarque-Bera	0.19	0.27	0.34	0.85	0.53	0.68	0.76	0.81	0.92	0.96	0.98	0.99

Table 5-4Probability of Rejection for the 85% of Coverage of Population Cases

Sample Size	50			125				250				
Test/α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Anderson-Darling	0.04	0.09	0.15	0.19	0.05	0.11	0.16	0.21	0.04	0.10	0.15	0.21
Cramer–Von Mises	0.04	0.09	0.14	0.20	0.05	0.11	0.15	0.20	0.05	0.11	0.15	0.20
Lilliefors	0.04	0.10	0.15	0.20	0.06	0.12	0.16	0.22	0.04	0.11	0.15	0.20
Shapiro-Wilk	0.05	0.10	0.15	0.20	0.05	0.10	0.16	0.21	0.04	0.08	0.13	0.19
Shapiro-Francia	0.05	0.10	0.15	0.20	0.05	0.10	0.16	0.22	0.04	0.09	0.14	0.19
Jarque-Bera	0.04	0.06	0.08	0.09	0.04	0.07	0.11	0.15	0.04	0.07	0.11	0.16

 Table 5-5
 Probability of Rejection for the Cases Drawn from a Normal Distribution

5.3 Goodness-of-Fit Testing

When evaluating the bias and bias uncertainty as a function of the independent variable, there are statistical and physical limitations. The primary statistical limitation of a linear regression-based bias uncertainty is that, similar to the single-sided LTL calculation, the approach assumes that the residuals about the trendline are normally distributed. The physical limitations of trend-based statistical procedures are discussed in Section 4.3.1.

The information gathered thus far is sufficient to test the following hypothesis:

 $H_0: \beta_1 = 0$

 $H_A: \beta_1 \neq 0$,

where β_1 and β_0 are the slope and intercept of the trendline.

The test statistic t is calculated using Eq. (4) and the information derived from the validation suite:

$$t_{fit} = \frac{|\beta_1|}{\left(\frac{\sigma_{fit}}{\sqrt{S_{xx}}}\right)}.$$
(4)

 S_{xx} is the sum of the squared errors for the independent variable and is determined using Eq. (21) for unweighted data and Eq. (22) for weighted data.

The test statistic is then compared to the *t*-statistic, with n-2 degrees of freedom and a confidence level of 1- α . Therefore, conclusions reached based on the comparison with the *t*-distribution are as follows:

If $t_{fit} > t_{n-2,\frac{\alpha}{2}}$, then the null hypothesis is rejected, and the trended approach should be used for processing the validation results. The slope is statistically significant—nonzero—

if $t_{fit} < t_{n-2,\frac{\alpha}{2}}$, then the null hypothesis cannot be rejected, and the trend should be ignored in favor of techniques that view the sample as uncorrelated.

Historically, the value of α has been taken to be 0.05, which corresponds to having 95% confidence that the analyst has not inadvertently concluded there is a trend when one is not present. The *t* test provides a rigorous determination of whether or not the slope of the proposed trend is statistically significant. Trends that are statistically significant have a slope that is nonzero. This approach does not differ based on the sign of the slope.

or

6 DETERMINATION OF BIAS AND UNCERTAINTY

This section discusses the methods used to convert the k_{eff} values calculated for selected critical experiments into parameters which can be applied to the results of safety analysis calculations. This will provide an appropriate level of confidence that the final value has the appropriate probability that the true system k_{eff} is lower than the regulatory limit.

The validation parameters traditionally applied to safety analysis calculations are the *bias*, β , and the *bias uncertainty*, σ_{β} . The *bias* is the deviation of the average k_{eff} of the validation suite from unity. The *bias uncertainty* accounts for the statistical uncertainty in the bias based on the standard deviation, sample size, and distribution of k_{eff} values of the validation suite. The values of β and σ_{β} ensure that the systems that are predicted to be subcritical by the computational method will indeed be subcritical.

Rigorously evaluated critical experiments are recognized as the best source of the integral measurement of k_{eff} values and are the current standard for criticality code and cross section library validation [31]. Part of the critical experiment evaluation process is the development of estimates of the expected value of experimental k_{eff} (k_{exp}) and the uncertainty in k_{exp} resulting from experimental uncertainties (σ_{exp}). The value of k_{exp} takes any physical deviations from 1.0 of the experiment into account, as well as the changes in k_{eff} from the measured configuration introduced by modeling simplifications included in the evaluation model (e.g., removal of support hardware from model). To account for deviation of the expected value of the evaluated experiment model from 1.0, the normalized k_{eff} (k_{norm}) should be calculated for each critical experiment model. The total uncertainty for each experiment (σ_i) is determined using Eq. (6), where σ_{MC} is the Monte Carlo uncertainty in k_{calc} . The σ_{MC} term would be 0 for deterministic methods that lack a stochastic uncertainty. The methods discussed in this section are taken from those used in work by Dean and Tayloe [31] and can also be traced to work by Trumble and Kimball [49].

$$k_{norm} = \frac{k_{calc}}{k_{exp}} \tag{5}$$

$$\sigma_i = \sqrt{\sigma_{exp}^2 + \sigma_{MC}^2} \tag{6}$$

With regard to selecting a technique to calculate the bias and bias uncertainty, it is recommended that trends on physically relevant parameters be investigated first, because these techniques allow the analyst to generate a bias and bias uncertainty that are more reflective of the system being analyzed than nontrending methods. If no statistically significant trend is present in the data, then the univariate normality of the validation data set is assessed to determine whether the normality-based bias uncertainty methods are acceptable. If it is not possible to demonstrate that the normality-based bias uncertainty technique is appropriate, then nonparametric methods are used. A flow chart of the selection process is provided in Figure 6-1.



Figure 6-1 Statistical Method Selection Process

The methods presented above are discussed in detail in the following sections. Section 6.1 discusses the methods used to calculate β and σ_{β} using nontrending methods under the assumption that the validation suite is a sample drawn from a normal distribution. These techniques are the simplest and are most applicable when no statistically significant trends are identified in the data with respect to independent system variables. Section 6.2 discusses the nonparametric methods that can be used to develop β and σ_{β} from an uncorrelated sample that does not satisfy the necessary criteria to be treated as having been drawn from a normal distribution. Nonparametric statistical methods are used for data sets that fail to meet the assumption of normality because no assumptions are made regarding the underlying population distribution. Section 6.3 discusses the methods used to determine β and σ_{β} as a function of a physically relevant independent system variable, commonly referred to as *trending analysis*. These techniques can provide insight into the underlying physical cause of the bias of a computational method and can generate application-specific bias and bias uncertainty estimates for various similar safety analysis models. Figure 6-2 provides an overview of the methods presented in this section grouped into the different statistical approaches used in each technique. The methods are presented in the same order in Figure 6-2 as the subsections are presented in the text.



Figure 6-2 Summary of Statistical Methods for Determining Bias and Bias Uncertainty

Section 6.4 examines issues related to treatment of positive biases in validation analysis, and Section 6.5 addresses the potential impact of correlated critical experiments on bias and bias uncertainty. Section 6.6 discusses the methods that have traditionally been used to incorporate the validation bias and bias uncertainty into a criticality analysis to demonstrate compliance with regulatory limits.

6.1 Nontrending Methods

For cases in which no statistically significant k_{norm} trend exists (see Section 6.3 for trending analysis) and the data are appropriately or conservatively represented by a normal distribution, it is possible to develop a bias and bias uncertainty using the LTL approach. The first step is to determine the mean value of k_{norm} (\bar{k}_{norm}), the standard deviation of k_{norm} , the average total uncertainty, and the square root of the pooled uncertainty. \bar{k}_{norm} is given by Eq. (7) for unweighted calculations and by Eq. (8) for uncertainty weighted calculations. The uncertainty

weighting uses the inverse of the variance to reduce the weight applied to experiments with higher uncertainties.

Uncertainty weighting is a generally recommended statistical approach for handling a series of similar measurements, and the purpose of the weighting is to reduce the importance of measurements with higher uncertainties [50]. Because the uncertainty in k_{norm} shown in Eq. (6) is typically dominated by the experimental uncertainty derived by the individual evaluating the critical experiment, it is possible that the uncertainties in k_{norm} can vary significantly from evaluation to evaluation due simply to the variability in the judgement of the evaluators between cases. Both weighted and unweighted approaches to the determination of the bias and bias uncertainty should be considered. When deciding whether a weighted or unweighted analysis is more appropriate, the variation in the experimental uncertainties from the critical experiment evaluations should be assessed. The uncertainty in k_{norm} should still be used in the calculation of the average total uncertainty as given in Eq. (11) and Eq. (12) for cases in which it is decided that a weighted analysis is not appropriate and should only be excluded from the calculation of the average k_{norm} and its variance.

$$\overline{k}_{norm} = \frac{\sum_{i=1}^{N} k_{norm_i}}{N} \tag{7}$$

$$\overline{k}_{norm} = \frac{\sum_{i=1}^{N} \frac{k_{norm_i}}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$
(8)

The variance about \bar{k}_{norm} is given by Eq. (9) for unweighted calculations and in Eq. (10) for weighted calculations.

$$\sigma_{\overline{k}_{norm}}^2 = \frac{\sum_{i=1}^N (k_{norm_i} - \overline{k}_{norm})^2}{N-1}$$
(9)

$$\sigma_{\overline{k}_{norm}}^{2} = \frac{\left(\frac{1}{N-1}\right)\sum_{i=1}^{N}\frac{1}{\sigma_{i}^{2}}(k_{norm_{i}}-\overline{k}_{norm})^{2}}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_{i}^{2}}}$$
(10)

The average total uncertainty is calculated using Eq. (11) for unweighted calculations and in Eq. (12) for weighted calculations.

$$\bar{\sigma}^2 = \frac{\sum_{i=1}^N \sigma_i^2}{N} \tag{11}$$

$$\bar{\sigma}^{2} = \frac{N}{\sum_{i=1}^{N} \frac{1}{\sigma_{i}^{2}}}$$
(12)

The square root of the pooled variance is computed using Eq. (13). The square root of the pooled variance accounts for the uncertainty associated with the individual k_{eff} values and the uncertainty from the statistical scatter of the k_{eff} values about the mean.

$$S_p = \sqrt{\sigma_{\bar{k}_{norm}}^2 + \bar{\sigma}^2} \tag{13}$$

The methods presented here calculate the variance *about* the mean (term from NUREG/CR-6698 [31]), which will subsequently be referred to simply as *the variance*, and its square root will be referred to as the *standard deviation*. The variance *of* the mean is not an appropriate substitute, as was documented in NRC Information Notice 2011-03 [51]. The variance about the mean ensures that a suitable proportion of the population of validation k_{effs} are above the LTL with the specified confidence level.

The single-sided lower tolerance factor necessary for calculating the bias uncertainty must be determined once the average normalized k_{eff} and its standard deviation have been calculated. The assumption of a normal distribution must also have been assessed to be applicable or conservative (See Section 5.2) for the validation suite. The method presented for calculating the single-sided lower tolerance factor is originally taken from the National Institute of Standards and Technology (NIST) Engineering Statistics Handbook [52] and is applied by calculating the non-centrality parameter in Eq. (14) and subsequently using that value to calculate the single-sided lower tolerance factor in Eq. (15). This method of calculating the single-sided lower tolerance factor is calculated using the inverse cumulative noncentral t-distribution, which may be found in statistical tables, or it may be calculated with most statistical packages. It may be difficult to calculate these values with a spreadsheet. If it is desired to calculate-single sided tolerance factors without the noncentral t-distribution, then alternate methods are provided in Natrella [53], but they may produce inaccurate results for small values of N.

$$\delta = z_n \sqrt{N} \tag{14}$$

$$K = \frac{t_{\gamma, N-1, \delta}}{\sqrt{N}},\tag{15}$$

where z_p is the z-score corresponding to the desired proportion of the population, p, *N* is the number of experiments in the validation suite and K is the single-sided lower tolerance factor,

 δ is the non-centrality parameter,

 $t_{\gamma,N-1,\delta}$ is the inverse of the cumulative noncentral t-distribution corresponding to the desired confidence level, γ , and

K is the single-sided lower tolerance factor.

For convenience, values of *K* are presented in Table 6-1 for a 95% confidence level that 95% of the true population of k_{eff} values lies above the LTL for various values of *N*. The values in Table 6-1 were calculated using Eq. (14), Eq. (15), and the R statistical package. It is conservative to use a value of *K* corresponding to a smaller sample size than is present in the validation suite if the confidence level and proportion of the population are both 95%. If it is necessary to change either the confidence level or the fraction of the population lying above the LTL, then Eq. (14) and Eq. (15) must be used in lieu of Table 6-1.

Once the statistical parameters have been calculated, it is possible to determine the bias and bias uncertainty. Determination of the bias is independent of the underlying distribution. The bias uncertainty, however, requires that a specific proportion of the population of true k_{eff} values

lies above the calculated LTL, which is underpinned by the normal distribution. The bias is calculated using Eq. (16):

$$\beta = \overline{k_{norm}} - 1. \tag{16}$$

Equation (17) may be used to calculate the bias uncertainty if the assumption that the data used for the validation of the criticality code was drawn from a normal distribution can be justified. If the assumption of normality cannot be defended, then the methods presented in Section 6.2 should be used.

$$\sigma_{\beta} = KS_P \tag{17}$$

Table 6-1Single-Sided Tolerance Factors for 95% Confidence that 95% of the True
Population Lies Above the Tolerance Limit as a Function of the Number of
Points in the Sample

Number of points in sample (<i>N</i>)	Single-sided lower tolerance multiplier (K)	Number of points in sample (<i>N</i>)	Single-sided lower tolerance multiplier (K)
10	2.911	115	1.905
11	2.815	120	1.899
12	2.736	125	1.894
13	2.671	130	1.888
14	2.614	135	1.883
15	2.566	140	1.879
16	2.524	145	1.874
17	2.486	150	1.870
18	2.453	155	1.866
19	2.423	160	1.862
20	2.396	165	1.858
21	2.371	170	1.855
22	2.349	175	1.852
23	2.328	180	1.849
24	2.309	185	1.846
25	2.292	190	1.843
30	2.220	195	1.840
35	2.167	200	1.837
40	2.125	205	1.835
45	2.092	210	1.832
50	2.065	215	1.830
55	2.042	220	1.828
60	2.022	225	1.825
65	2.005	230	1.823
70	1.990	235	1.821
75	1.976	240	1.819
80	1.964	245	1.817
85	1.954	250	1.815
90	1.944	255	1.814
95	1.935	260	1.812
100	1.927	265	1.810
105	1.919	270	1.809
110	1.912	275	1.807

6.2 Nonparametric Methods

Nonparametric methods, also known as *distribution-free methods*, differ from the nontrending methods presented in the previous subsection in that there is no assumption of normality underlying the analysis of the bias uncertainty. This approach tends to result in a larger bias uncertainty because more conservative analysis is required without an assumed underlying distribution of the benchmark results. Historically, the use of nonparametric methods has been based on rank-order approaches such as those outlined in Dean and Tayloe [31] and in Trumble and Kimball [49]. This approach will be presented in this subsection. Recently, a new USL determination method has been developed at Los Alamos National Laboratory for use with MCNP. This new method, Whisper, incorporates automatic experiment selection based on the c_k parameter and determines the bias and bias uncertainty using a method derived from the extreme value theory (EVT) [54]. Some portions of the administrative margin (see Section 6.6) are also estimated in Whisper. A summary of this new method is also provided in this subsection.

6.2.1 Historical Nonparametric Method

If normality cannot be demonstrated for a validation suite, then methods must be used that do not rely upon the assumption of a normal distribution to calculate the bias uncertainty. The methods used to calculate the bias uncertainty that do not rely on the normal distribution are referred to as *distribution free* or *nonparametric methods*. The nonparametric methods presented here were previously presented in Dean and Tayloe [31] and Trumble and Kimball [49] and are originally traceable to Conover [55]. Nonparametric methods compute the bias uncertainty by determining the confidence level (C) that the lower ranks of each of the data points in the sample bound a prescribed proportion of the population. Once the data point has been selected, the bias uncertainty is calculated as the difference between the mean k_{eff} and that data point. The confidence level can be calculated using Eq. (18),

$$C = 1 - \sum_{j=0}^{m-1} \frac{n!}{j! (n-j)!} (1-q)^j q^{n-j},$$
(18)

where

q is the desired proportion of the population (usually 0.95), n is the number of points in the sample, and m is the rank of the data point under consideration order indexing from lowest to

highest k_{norm_i} values, (m=1 for lowest point; m=2 for second lowest point).

The nonparametric method presented here is used to calculate the confidence level that the specified proportion of the population lies above the lowest ranked points of the sample. The analyst works from the lowest data point to the point that yields a value of C less than the desired confidence level (0.95 most commonly.) For example, if a data set has 105 points with 95% of the population covered with a confidence level of 95%, then the value of C for the lowest data point (m=1, n=105, and q=0.95) would be 0.9954. Because this is higher than the 95% confidence level, the next point is evaluated. The confidence level of the second lowest point (m=2, n=105, and q=0.95) is 0.9701, also satisfying the required confidence level. Finally, the third lowest data point would be examined. It fails to meet the desired confidence level (C of 0.9008). From this process, it is determined that the second lowest point would be used. Using a lower order rank is always more conservative, so this process does not need to be used if the first rank is chosen for the sake of conservatism as long as the lowest rank provides the needed level of confidence.

For convenience, Table 6-2 provides the minimum number of points corresponding to 95% confidence that 95% of the population lies above each of the five lowest ranks.

Lower rank (<i>m</i>)	Number of points in sample
1	59
2	93
3	124
4	153
5	181

Table 6-2Sample Size as a Function of Lower Rank of Data Points Necessary to Give
95% Confidence that 95% of the Population Lies Above that Point

As shown in Table 6-2, at least 59 points are required to have 95% confidence that at least 95% of the population is above the lowest point when making no assumption regarding the distribution of the population. For smaller samples, it is necessary to compensate for the lack of confidence that can be demonstrated with nonparametric methods. Based on Eq. (18), a sample with 30 points, for example, is only sufficient to establish 78.5% confidence that 95% of the population is above the lowest point. The reactivity margin for this lack of statistical confidence, termed the *nonparametric margin* (NPM), was put forward in Dean and Tayloe [31] and Trumble and Kimball [49]. The original basis for the values used in Dean and Tayloe [31] and Trumble and Kimball [49] was not documented in either reference. The values are believed to be conservative because they are large in comparison to typical bias and bias uncertainty values seen with modern codes and data sets. Whereas lower NPM values may be appropriate, this is difficult to demonstrate without knowing the basis for the currently accepted values. The values of NPM from Trumble and Kimball [49] are repeated in NUREG/CR-6698 [31] and are presented in Table 6-3 below.

Degree of confidence for 95% of the population	NPM
>90%	0.00
>80%	0.01
>70%	0.02
>60%	0.03
>50%	0.04
>40%	0.05
≤40%	Need additional data

Table 6-3 NPM as a Function of the Degree of Confidence Calculated with Eq. (18)

The bias uncertainty can be extracted from the other terms used in the historical nonparametric assessment. The underlying distribution has no impact on the average normalized k_{eff} , so Eq. (16) is still appropriate for the calculation of the bias. Because predictions of the extrema of a distribution are highly sensitive to the underlying distribution, the bias uncertainty must be calculated using Eq. (19) rather than Eq. (17):

$$\sigma_{\beta} = (\bar{k}_{norm} - k_{norm_r}) + \sigma_{k_{norm_r}} + NPM \text{ if } \bar{k}_{norm} \le 1, \text{ or}$$

$$\sigma_{\beta} = (1 - k_{norm_r}) + \sigma_{k_{norm_r}} + NPM \text{ if } \bar{k}_{norm} > 1,$$
(19)

where

 \bar{k}_{norm} is the mean normalized $k_{\rm eff}$,

 k_{norm_r} is the normalized k_{eff} of the point corresponding to the lower rank calculated by iterating to the desired level of confidence,

 $\sigma_{k_{norm_r}}$ is the combined uncertainty in k_{norm_r} calculated with Eq. (6), and

NPM is the nonparametric margin for cases with less than the needed number of points necessary to justify a lower rank of one.

6.2.2 Whisper Method

The Whisper methodology brings several aspects of validation analysis into a single automated process. The methodology has been developed into a computational tool which is also called Whisper. The initial public release of the capability was in MCNP version 6.2, which was made publicly available in April 2018. A brief synopsis of the Whisper methods is presented here. More information is available in Kiedrowski [54].

Whisper determines a calculation margin, which is the sum of the bias and bias uncertainty. The bias is estimated but is used solely for extracting the value of the bias uncertainty from the calculational margin. Both the bias and its uncertainty can then be reported, but only the sum (total calculational margin) is used in the USL determination.

Experiment selection within the Whisper method is performed using S/U methods, specifically with the c_k integral index (see Section 4.2.2 for more information about c_k). Sensitivity data for a large number of critical experiments (on the order of 1,000) is distributed with MCNP to facilitate these comparisons. Prior to c_k comparisons, a data adjustment process is completed to identify potential outlier experiments. These cases may then be excluded from the validation suite. As discussed below, the residual uncertainty from this data adjustment is used to determine the recommended minimum administrative margin for the application. The c_k value is then determined for each remaining experiment and the application being validated. The c_k values are then normalized so that the maximum value in the validation suite is set to 1. The normalized c_k values will be used as weights in the determination of calculational margin.

The Whisper software also determines the number of experiments to include in the bias calculation. The default minimum desired total weight is 25, assuming a perfectly correlated benchmark experiment (i.e., $c_k=1$) exists within the validation suite. A default penalty factor of 100 is applied to the difference between $c_k=1$ and the maximum c_k value in the selected set of benchmarks to raise the minimum desired total weight and thus increase the number of benchmark experiments required. For example, if the highest c_k value is 0.95 the difference is

0.05; this increases the target weight to 30. Both the penalty factor (100) and minimum desired total weight (25) may be arbitrarily changed by the user. The normalized c_k values are summed, starting with the largest c_k value, until the desired weight is achieved. There are two notable effects of this method: larger maximum c_k values reduce the total weight requirement; and smaller maximum c_k values allow less correlated benchmark experiments to be used in meeting the total weight requirement. There may also be scenarios in which smaller maximum c_k values allow the target weight to be achieved with fewer critical experiments than for larger maximum c_k values. An additional margin is included in the recommended administrative margin if the required weight cannot be achieved with the available benchmarks. The calculational margin is then determined based on the EVT methods as described in Kiedrowski [54].

The final step of the Whisper USL determination method is the calculation of the recommended minimum administrative margin. Whisper partitions the administrative margin into three terms: the first is assigned to the neutron transport software, the second to the nuclear data, and the third to the application. The determination of the application-specific portion of the administrative margin is left to the analyst for each application. The residual uncertainty remaining after the initial data adjustment process is used here as the nuclear data contribution to the administrative margin. Finally, a 0.5% Δk contribution to the administrative margin is added for potentially undetected errors in the neutron transport code. This term is unique to Whisper and goes beyond the requirements of ANSI/ANS-8.24 [3]. This term is additional to those required by traditional validation approaches, as the purpose of validation is to quantify the performance of the computation method for the application; it is conservative to add such a penalty. A larger margin for software errors is recommended for neutron transport codes that are not widely used [54]. Ultimately, the analyst controls the final USL by adjusting the application-specific portion of the administrative margin.

The Whisper method is new and has seen limited applications at DOE-regulated facilities in the last few years. Whisper was developed to be a conservative approach to validation, and the overall conservatism of the methodology framework is likely sufficient to address concerns on many questions regarding implementation details. Largely administrative concerns remain regarding the use of the sensitivity data and calculated k_{eff} values distributed with MCNP that are generated on LANL computers and not by the organization performing the validation. There are also reported errors in some of the LANL benchmark models [56], and it is not clear how these errors may impact Whisper-based validation of MCNP. At this writing, the methodology has not been thoroughly reviewed for generic use at NRC licensed facilities.

6.3 Analysis of Trends

This section discusses trending techniques used to correlate k_{eff} bias with an independent variable. The techniques and descriptions used in this section are largely paraphrased from Chapter 10 of Tamhane and Dunlop [57] for unweighted analysis and from Bevington and Robinson [50] for weighted analysis. As discussed in Section 6.1, uncertainty weighting is generally recommended for analysis of a series of similar measurements. The uncertainties in critical experiment evaluations vary for many reasons and may not provide equivalent estimates of the total experimental uncertainties. Consideration should therefore be given to the use of weighted or unweighted statistics.

Nuclear data sets have historically included errors which result in increasing biases with increasing dependence on specific portions of the data. Trending analysis of the critical benchmarks can reveal these dependencies. Appropriately characterized trends may capture

the bias of the application of interest better than a nontrending statistical model. The linear model typically used in this approach is given in Eq. (20),

$$k_{fit}(x) = \beta_0 + \beta_1 x, \tag{20}$$

where $k_{fit}(x)$ is the mean of the normalized k_{eff} as a function of x, x is the physically relevant independent variable upon which a trend is

being investigated, and β_0 and β_1 are the intercept and slope of the regression line correlating *x* and the normalized k_{eff} values.

The generation of a linear trend requires the sum of squared differences about the mean of the independent variable (x) and the sum of the product of the differences for the k_{norm} values and the independent variable. The first term, the sum of squared differences about the mean for the independent variable (denoted S_{xx}), can be calculated with Eq. (21) for an unweighted analysis and with Eq. (22) for a weighted analysis. The sum of the cross products of the independent variable with k_{norm} (denoted S_{kx}) can be calculated with Eq. (23) for an unweighted analysis and with Eq. (24) for a weighted analysis. It is noted that the value of \bar{k}_{norm} should be calculated with Eq. (7) for unweighted analyses and Eq. (8) for weighted analyses. It is also appropriate to calculate \bar{x} with Eqs. (7) and (8) by substituting the trend parameter for k_{norm} .

$$S_{xx} = \sum_{i=1}^{N} (x_i - \bar{x})^2$$
(21)

$$S_{xx} = \frac{\sum_{i=1}^{N} (\frac{1}{\sigma_i^2} (x_i - \overline{x})^2)}{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$
(22)

$$S_{kx} = \sum_{i=1}^{N} (k_{norm_i} - \bar{k}_{norm})(x_i - \bar{x})$$
(23)

$$S_{kx} = \frac{\sum_{i=1}^{N} (\frac{1}{\sigma_i^2} (x_i - \overline{x}) (k_{norm_i} - \overline{k}_{norm}))}{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$
(24)

The slope of the regression line, β_1 , can be calculated using Eq. (25), and the intercept of the regression line, β_0 , can be calculated using Eq. (26).

$$\beta_1 = \frac{S_{kx}}{S_{xx}} \tag{25}$$

$$\beta_0 = \bar{k}_{norm} - \beta_1 \bar{x} \tag{26}$$

It is also necessary to calculate the residual error associated with the regression line for each data point (e_i) , as shown in Eq. (27):

$$e_i = k_{norm_i} - k_{fit}(x_i). \tag{27}$$

The residual error term can be used to calculate variance about the trend line. The residual standard error about the trend line is given by Eq. (28) for an unweighted analysis and in Eq. (29) for a weighted analysis:

$$\sigma_{fit} = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N-2}}$$
(28)

$$\sigma_{fit} = \sqrt{\frac{\frac{1}{N-2}\sum_{i=1}^{N}\frac{1}{\sigma_{i}^{2}}e_{i}^{2}}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_{i}^{2}}}}$$
(29)

where n-2 corresponds to the number of degrees of freedom associated with a linear fit solving for the two unknowns— β_0 and β_1 .

6.3.1 Calculation of the Bias for a Statistically Significant Trend

For cases in which a statistically significant trend has been found to exist (see Section 5.3 for assessment techniques), the bias as a function of the independent trend variable can be calculated using Eq. (30), which is equivalent to Eq. (16) for the nontrending technique.

$$\beta(x) = k_{fit}(x) - 1 \tag{30}$$

A single-sided uncertainty about the trend line must also be developed. This uncertainty band establishes a limit above which a specified proportion of the population of critical experiments lies with sufficient statistical confidence. The statistical techniques used for this purpose are analogous to the single-sided LTL discussed in Section 6.1, but they differ in that they represent the statistical lower bound of the data as a function of an independent variable rather than as a constant value. Historically, there have been at least three methods used to calculate a lower tolerance band. These methods are referred to as the *single-sided lower tolerance band*, the *confidence band with administrative margin*, and the *single-sided uniform width closed interval*. Each of the methods uses the square root of the pooled variance as determined using Eq. (31). Note that these methods are sensitive to the departure of the distribution of the residuals from normality, as is the single-sided tolerance factor used in the LTL approach. The normality of the residuals can be assessed using the same methods discussed in Section 5.2.

$$S_p = \sqrt{\sigma_{fit}^2 + \bar{\sigma}^2} \tag{31}$$

where

 σ_{fit} is the uncertainty in the fit given by Eq. (28), and

 $\overline{\sigma}$ is the average total uncertainty given by Eq. (11) or Eq. (12).

6.3.2 Single-Sided Lower Tolerance Band

The single-sided lower tolerance band is the method of bias uncertainty calculation presented in the "Tolerance Band" section of NUREG/CR-6698 [31] (p.12). The source of these methods for use in the criticality safety analysis is Trumble and Kimball [49]. The original statistical model is from Chapter 3 of Miller [58]. The equation for the single-sided lower tolerance band is given by Eq.(32):

$$\sigma(x) = S_p \left\{ \sqrt{2F_{\alpha}^{(2,n-2)} \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}} \right]} + z_p \sqrt{\frac{(n-2)}{\chi_{1-\frac{\alpha}{2},n-2}^2}} \right\}.$$
 (32)

The variables and statistical functions in (32) are explained in Table 6-4, which also includes the equation numbers for variables defined elsewhere in this document and the Excel and R input for the statistical functions, where α is the significance level (typically 0.05), and *P* is the proportion of the population covered (typically 0.95).

Another method discussed in Trumble and Kimball [49], *the confidence band approach*, is not an appropriate trend validation technique because it does not account for the spread of the data around the trend line. This shortcoming is analogous to the difference between the variance *of* the mean vs. the variance *about* the mean identified in NRC Information Notice 2011-03 [51].

Table 6-4 Explanation of the Variables and Statistical Functions in Eq. (32)
---	-----

Variable	Variable description	Method of calculation
S_p	The square root of the pooled variance for trend calculations	Eq. (31)
$F_{lpha}^{(2,n-2)}$	The inverse of the F probability distribution corresponding with a degree of fit of 2 and n-2 degrees of freedom	R: qf(1-α,2,n-2) Excel: FINV(α,2,n-2) or F.INV.RT(α,2,n-2)
п	The number of critical experiments in the validation suite	-
$S_{_{XX}}$	The sum of squared errors for the independent variable	Eq. (21) for weighted data Eq. (22) for unweighted data
Zp	The inverse of the standard normal cumulative distribution that contains the fraction p of the distribution	R: qnorm(P) Excel: NORMSINV(P) or NORM.S.INV(P)
$\chi^2_{1-\frac{\alpha}{2},n-2}$	The inverse of the right-tailed probability of the Chi-squared distribution.	R: qchisq(p=1-α/2, df=n-2, lower.tail=FALSE) Excel: CHIINV(1-α/2, n-2) or CHISQ.INV.RT(1-α/2,n-2)

6.3.3 Confidence Band with Administrative Margin (USL-1)

The second implementation of the trended validation approach for developing the bias uncertainty is from NUREG/CR-6361 [32]. Statistically, the method is a lower prediction band (the name in the title of the section is retained for continuity with previous documents) and is the basis of USL-1 from the USLSTATS code. The lower prediction band should be interpreted as the line above which there is 95% confidence that a future individual k_{eff} value would lie. The original statistical methods are discussed in Miller, Chapter 3, Section 2 [58]. For this method, the bias uncertainty as a function of the independent variable is calculated with Eq. (33). S_{kk} is calculated in the same way as S_{xx} , except the normalized k_{eff} values are used instead of the independent variable. The value of $t_{1-\alpha,n-2}$ is the inverse t distribution corresponding to the desired confidence (usually 95%) that the next point will be above the prediction band, with n-2 degrees of freedom. The inverse t distribution can be calculated in Excel using *TINV((2\alpha), n-2)* or *T.INV((1- \alpha), n-2)*:

$$\sigma_{\beta}(x) = t_{1-\alpha,n-2} S_p \left\{ \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}} \right\}.$$
(33)

Historically, USL-1 is calculated by evaluating $\sigma_{\beta}(x)$ at the upper and lower ends of the range of the independent trending variable in the validation suite. The larger of those two values is then

used within the range as a constant bias uncertainty. This is expressed mathematically in Eq. (34):

$$\sigma_{\beta} = \max\{\sigma(x_{\min}) | \sigma(x_{\max})\}$$
(34)

Taking the maximum of the bias uncertainty across the range of applicability is a conservative treatment, and Lichtenwalter et al. [32] indicate that it is done for convenience. It is also acceptable to evaluate the bias uncertainty at the value of the trending variable corresponding to the application being validated. If used for extrapolation outside the range of critical experiments, $\sigma_{\beta}(x)$ should be used.

6.3.4 Single-Sided Uniform Width Closed Interval Approach (USL-2)

This method of calculating the bias uncertainty is referred to as the *USL-2 calculation* in Lichtenwalter et al. [32] and is also included in the USLSTATS program. This method is a lower tolerance band approach similar to the method discussed in Section 6.3.2. The original statistical methods for this approach are taken from Bowden and Graybill [59], as well as Johnson [60]. Calculation of the lower tolerance band is accomplished by first calculating the statistical parameters g, h, ρ , and A using Eq. (35), Eq. (36), Eq. (37), and Eq. (38). These parameters are used to determine the appropriate tolerance band.

$$g = \sqrt{\frac{1}{n} + \frac{(x_{\min} - \bar{x})^2}{S_{xx}}}$$
(35)

$$h = \sqrt{\frac{1}{n} + \frac{(x_{\max} - \bar{x})^2}{S_{xx}}}$$
(36)

$$\rho = \frac{1}{gh} \left[\frac{1}{n} + \frac{(x_{\min} - \overline{x})(x_{\max} - \overline{x})}{S_{xx}} \right]$$
(37)

$$A = \frac{g}{h} \tag{38}$$

The value of D can then be calculated iteratively using these parameters in the integral in Eq. (39) to arrive at the desired level of confidence. The value of D is then used with the proper form of Eq. (40) based on the value of A to calculate the value of confidence constant, C^* . $C_{\alpha/P}$ is calculated using Eq. (41), which is in turn used to calculate the bias uncertainty in Eq. (42).

$$\int_{-D}^{D} \int_{-AD}^{AD} \frac{1}{2\pi\sqrt{1-\rho^2}} \left[1 + \frac{u^2 - 2\rho uv + v^2}{(n-2)(1-\rho^2)} \right]^{\frac{-n}{2}} du dv \le \alpha$$
(39)

$$C^* = D \cdot g$$
 for $0.5 \le A \le 1.5$

$$C^* = D \cdot h$$
 for other values of A (11)

$$C_{\alpha/p} = C^* + t_{P,n-2} \sqrt{\frac{n-2}{\chi^2_{\alpha,n-2}}}$$
(41)

$$\sigma_{\beta} = C_{\alpha/p} \cdot S_p \tag{42}$$

(40)

6.4 Using Positive Validation Analysis Biases

For the purposes of this discussion and consistency with the definition provided in footnote 3 of ANSI/ANS-8.24 [3], a *positive bias* is produced by a validation study when the average of the calculated k_{eff} values exceeds the average of the expected k_{eff} values. Existence of a positive bias indicates that the computational method tends to produce k_{eff} values higher than the true value. Taking credit for this tendency is referred to as *crediting a positive bias*. Crediting the positive bias lowers calculated safety analysis k_{eff} values after adjustment for bias and uncertainties, or it generates less restrictive USLs.

Section 6.1.3 of ANSI/ANS-8.24 [3] states, "If a positive bias is used in the determination of the calculational margin, its use shall be justified by the close applicability of the benchmarks." The standard requires a justification based on the confidence generated by having closely applicable, highly similar benchmarks. This requirement is an amplification of the primary requirement of a validation to be based on critical benchmark experiments that are as similar as possible to the safety analysis model. Justification for use of a positive bias must be documented.

Generally, the NRC has not accepted credit for positive bias in licensing analyses. Regulatory Guide 3.71, Revision 3 [4] specifically takes exception to the positive bias allowance in ANSI/ANS-8.24 and states that staff, "may choose to evaluate its use on a case-by-case basis with suitable demonstration that the causes of the bias are known and in accordance with Section 6.1.3."

The motivation behind treating positive biases differently is to eliminate the potential nonconservatism that would result if a positive bias were overestimated. This is a valid concern but is equally applicable to any bias determination: an estimate of the bias that is less negative than the actual bias for the application system is also nonconservative. The best way to ensure an accurate bias estimate is to perform a thorough validation covering the entire parameter space in which the normal and credible abnormal conditions exist and to perform trending analyses to expose any changes of the bias as a function of a variable parameter in the analysis models. This is the reason that close applicability of the benchmarks is required to justify the use of a positive bias, and such close applicability of the benchmarks is a condition that should be satisfied for all validations.

6.5 <u>Issue, Impact, and Potential Inclusion of Correlated Critical Experiment</u> <u>Results</u>

A series of critical experiments is often performed with a limited number of parameters that are varied systematically to cover a range. This approach serves multiple purposes. Primarily, performing experiment series allows for the determination of system sensitivity to specific

parameters such as lattice pitch or reflector thickness. Unlike some types of experiments, critical experiments cannot vary in only a single independent parameter. Any change to a critical system makes it either subcritical or supercritical, so an offsetting additional change must be made to restore criticality. Some experiments are controlled with the mass of fuel present, and others are controlled with material separation, the amount of moderating material present, or the concentration of neutron absorber in the system. Generally, the system response to one parameter change, such as material separation, is well understood and is therefore used to offset changes in a different parameter. This allows for an estimation of the sensitivity of the system to changes in the second parameter, although the sensitivity is not necessarily known with the same accuracy as is possible in experiments with single variable controls. An additional benefit to performing experiments in series is that several related experiments can be performed at lower cost per experiment and in less time than if each experiment had been performed in isolation.

The use of experiment series in traditional, non-data adjustment validation techniques creates additional complexities because the correlations among the individual experiments within the series are generally not treated in the statistics used in the analysis. All of the techniques presented in Sections 6.1, 6.2, and 6.3 assume that uncorrelated data are used. The correlation between a pair of experiments is a result of shared experimental components that include but are not limited to fissile, reflector, or absorbing materials, detector systems, and procedures. Many of these shared characteristics should have very little effect on the results of the experiments or the independence of the data measured or derived from the experiments. The use of common materials and fixtures, however, can create correlations among the experiments that demonstrably reduce the independence of each experiment in a series. This can impact the determination of the computational bias, but it is far more likely to affect the uncertainty in the bias. The uncertainty is increased because several measurements of the same system do not provide as much unique information as the same number of measurements of different systems. Thus, the correlation among experiments in a series acts to reduce the effective number of experiments in a validation set. The smaller number of effective experiments would lead to a larger uncertainty, so neglecting the correlations is nonconservative because it results in a lower bias uncertainty. The impacts of these correlations on the bias uncertainty are not considered in any of the methods presented in the previous sections.

In some critical experiments, a high degree of correlation is a desired characteristic. The maximum amount of information can be extracted from substitution experiments only when other parameters are constant or nearly so. For these experiments, a lack of correlation would cause the impact of the substitution to be difficult to determine. The value of these experiments, especially when incorporated into data adjustment, is greatly increased by a high degree of correlation.

A current challenge facing criticality safety practitioners and regulators is to establish a reliable method of determining the correlations among the critical experiments and ultimately to determine methods to incorporate them into usable validation techniques. More information on the determination of correlation coefficients is available in Hoefer [61] and in Marshall and Rearden [62]. One proposed validation technique incorporating correlations into trending analysis is also available [63].

6.6 Using Validation Analysis Results

Two methods have traditionally been used to incorporate the validation bias and bias uncertainty into a criticality analysis to demonstrate compliance with the regulatory limit. The
first method calculates the USL against which the statistically bounded eigenvalue ($k_{eff} + 2\sigma_{calc}$) of the safety analysis model is compared. This method is used when all physical uncertainties associated with the composition and arrangement of materials in the application of interest are included directly in the safety analysis model. The equations for the USL are given in Eq. (43) for the cases of both positive and negative values of β :

$$USL = 1 + \beta - \sigma_{\beta} - M_A - M_D \text{ if } \beta \le 0 \text{, or}$$

$$USL = 1 - \sigma_{\beta} - M_A - M_D \text{ if } \beta > 0 \text{,}$$
(43)

where

 β is the bias, σ_{β} is the bias uncertainty, M_A is the administrative margin, and M_D is margin added to account for any deficiencies associated the validation suite.

The methods for calculating the bias (β) and bias uncertainty (σ_{β}) are described in Sections 6.1, 6.2, and 6.3 and are summarized in Table 6-5. The administrative margin used in calculation of the USL depends on regulatory guidance for the system being evaluated and the fidelity with which the system being analyzed is known. In general, the administrative margin is viewed to be reserved for unanticipated issues and is smaller for systems that are well known than for systems that are more difficult to characterize. M_D is the margin used to account for identified deficiencies in the validation report such as unvalidated nuclides that are present in the safety analysis model. These deficiencies are discussed in more detail in Section 7. For more information on using a positive bias, see Section 6.4.

Table 6-5 Summary of Bias and Bias Uncertainty Calculation Techniques

Method	Equation to calculate β	Equation to calculate σ_{eta}
Nontrending	Eq. (16)	Eq. (17)
Historical nonparametric	Eq. (16)	Eq. (19)
Single-sided lower tolerance band	Eq. (30)	Eq. (32)
Confidence band with administrative margin (USL-1)	Eq. (30)	Eq. (33)
Confidence band with administrative margin (USL-1), single band width	Eq. (30)	Eq. (34)
Single-sided uniform width closed interval (USL-2)	Eq. (30)	Eq. (42)

The second method of developing a USL includes reactivity allowances for the treatment of uncertainties in the safety analysis model. Typically, these allowances are calculated using the square root of the sum of the squares of the k_{eff} uncertainties caused by independent uncertainties in material composition, geometry, or conditions. This representation of the USL is provided in Eq. (44). Examples of this approach are the new fuel vault and the spent fuel pool analysis regulated under Title 10 of the US Code of Federal Regulations (CFR) 50.68 [64]. It is

possible that a deficiency in the validation suite may be considered as part of σ_{SAM} rather than as part of M_D .

$$USL = 1 + \beta - \sum \beta_{SAM} - \sqrt{\sigma_{\beta}^{2} + \sum \sigma_{SAM}^{2}} - M_{A} - M_{D} \text{ if } \beta \leq 0 \text{ , and}$$

$$USL = 1 - \sum \beta_{SAM} - \sqrt{\sigma_{\beta}^{2} + \sum \sigma_{SAM}^{2}} - M_{A} - M_{D} \text{ if } \beta > 0 \text{ ,}$$
(44)

where

- β_{SAM} are the biases associated with known deviations of the model from reality or account for parameters which have no best estimate value (e.g., spent fuel pool temperature), and
- σ_{SAM} are uncertainties in the safety analysis model which have best estimate values and known uncertainties.

The use of Eq. (43) represents a worst-case bounding treatment of parameters. On the other hand, Eq. (44) is used when the system is modeled with best-estimate values, and reactivity allowances are made for independent uncertainty assessments.

An equivalent way to implement the validation parameters is to add them to the calculated k_{eff} (k_{calc} , adjusted for the Monte Carlo uncertainty of the calculation) to develop a maximum k_{eff} (k_{max}) that would then be compared to the limit. Using this approach, Eq. (43) would take the form of Eq. (45), and Eq. (44) would take the form of Eq. (46):

$$k_{\max} = k_{calc} - \beta + \sigma_{\beta} + M_A + M_D \text{ if } \beta \le 0 \text{ , and}$$

$$k_{\max} = k_{calc} + \sigma_{\beta} + M_A + M_D \text{ if } \beta > 0 \text{ ;}$$
(45)

$$k_{\max} = k_{calc} - \beta + \sum \beta_{SAM} + \sqrt{\sigma_{\beta}^2 + \sum \sigma_{SAM}^2} + M_A + M_D \text{ if } \beta \le 0 \text{ , and}$$

$$k_{\max} = k_{calc} + \sum \beta_{SAM} + \sqrt{\sigma_{\beta}^2 + \sum \sigma_{SAM}^2} + M_A + M_D \text{ if } \beta > 0 \text{ .}$$
(46)

7 IDENTIFYING AND ADDRESSING VALIDATION WEAKNESSES AND GAPS

Frequently, criticality analysts will find that there is some feature (i.e., physical component, material, element, or nuclide) in their application models that cannot be validated by comparison with critical experiments. This might be the result of materials present in the safety analysis application that are not present in critical experiments or that are not present in critical experiments in a neutronic environment similar to the application. In some cases, one or more materials may be present in safety analysis models in materials with unique compositions.

In this report, deficiencies in the validation study results or their applicability to an analysis are referred to as *validation deficiencies*. There may be a bias associated with the deficiency that is not captured by the validation study, and ignoring the missing bias component may be nonconservative. It is the responsibility of the analyst using the validation study results in the criticality safety analysis to identify and address validation deficiencies. There are several ways that some validation deficiencies may be addressed. Addressing major validation deficiencies with the potential for significant nonconservatisms will likely require discussions among all stake holders (i.e., analysts, operations, management, license holders and regulators).

The most obvious way to address validation deficiencies is to identify additional critical experiments that can be used to extend the area of applicability of the validation or to demonstrate that, if there is a bias, it is conservative to neglect it. An analyst may be able to demonstrate that ignoring a deficiency is conservative by using critical experiments that are not similar to the application of interest, provided that those critical experiments can account for the effect of the material of interest. An example of this would be the ability to examine a set of experiments that were not like the application of interest but had some cases with an absorber present and some cases that did not include the absorber. From that series of experiments, the analyst could determine whether the absorber caused a change in the bias or not.

In some situations, analysts have attempted to add a few experiments to cover a feature, material, or nuclide that is not present in the bulk of the validation suite to extend the area of applicability to include this missing feature(s). In general, this practice is not acceptable if the difficult-to-validate feature is only present in a small subset of the experiments and therefore cannot impact the overall bias. Hypothetically, one could attempt to quantify the bias using trending analysis. This would require more than a few experiments to support a statistically significant trend analysis. Without trending analysis, the extra bias associated with the few added experiments becomes averaged with the other critical experiments, reducing the associated bias. Therefore, one cannot generally address a validation gap by adding a few additional critical experiments. It is also important to carefully consider trending analysis for a parameter when many cases in the validation suite do not vary with respect to the parameter. One example of this problem is soluble boron in LEU fuel lattice experiments. As a general rule. only cases that contain soluble boron should be included in the trend analysis as a function of boron concentration because including a large number of cases with 0 ppm will statistically dilute the trend. An exception to this would be to include a case with no boron as part of a series of experiments that varied the soluble boron concentration.

It may be possible to address validation deficiencies by modifying the safety analysis model. If removing an unvalidated nuclide, material, or feature causes k_{eff} to rise and the maximum k_{eff} is still acceptable, then the validation deficiency may then be avoided by removing the material or feature. The conservative modeling simplification can be justified by the increased reactivity of

the model, and validation of the absorber is no longer needed, as it is not credited. In all cases, the analyst is responsible for demonstrating that the modeling changes made to address the validation deficiency are conservative under all conditions to be analyzed. In general, these are changes that must be documented in the generation of the safety analysis model, not in the validation.

One may be able to address a validation deficiency using sensitivity studies and knowledge of the nuclear data uncertainties. An example is validation of a neutron absorber panel heavily loaded with ¹⁰B. As a result of saturation of neutron absorbing effects and physical self-shielding, changing the ¹⁰B content in the panel by 10% may have little or no effect on k_{eff} . As can be seen in Figure 7-1, the one-standard-deviation uncertainty in the ¹⁰B cross sections are no more than 2% over the entire relevant neutron energy range. Consequently, the ¹⁰B cross sections cannot be in error by 10%, and inclusion of a penalty equal to the 10% perturbation study effect on k_{eff} should adequately address the validation deficiency. The selection of the factor of 5 from the reported uncertainty to the uncertainty evaluated with perturbation calculations was purely arbitrary and was chosen because it is easy to defend. Other factors could be used, although the magnitude of the factor may also vary with confidence in the nuclear covariance data.



Figure 7-1 Nuclear Data Uncertainty for ¹⁰B Total Cross Section

A more detailed, less conservative analysis of this type may be performed using sensitivity and uncertainty analysis tools. In this analysis, nuclide-, reaction-, and energy-dependent k_{eff} sensitivity data are calculated for the safety analysis model. These sensitivity data are then combined with the nuclear data uncertainty information to yield problem-, nuclide- and reaction-specific k_{eff} uncertainty values. This uncertainty analysis technique provides a quantitatively defensible estimate for the potential biases that may be associated with unvalidated nuclides

and features. Such an approach was used to address the lack of fission product validation in SNF calculations in NUREG/CR-7109 [19].

One last approach to dealing with unvalidated materials and features is to adopt a suitable Δk_{eff} margin that is large enough to cover the potential biases associated with them. This approach should be based on the reactivity worth of the feature or material in the safety analysis model. Adopting this margin will require acceptance by technical reviewers, criticality safety program managers, and the relevant regulators. The difficulty of building this consensus will vary, depending on the available margin to the k_{eff} limits.

The analyst is responsible for comparing the safety analysis models and the supporting validation study to identify validation deficiencies as part of each criticality safety assessment. The potential impact of validation deficiencies must be addressed. If a preexisting validation study is used to support criticality analysis, then the analyst must identify and address any existing impact of the validation deficiencies in the criticality analysis.

8 DOCUMENTATION

The validation of the computational method must be documented in a formal report that is controlled under the appropriate QA program for the organization performing the criticality analyses. In general, there are many acceptable forms for a validation report, as long as the layout and content of the report allow a competent reviewer to reproduce the critical experiment selections, understand the applicability of the methods used to determine the bias and bias uncertainty, understand the calculation of the bias and its uncertainty, and implement the bias and bias uncertainty to determine the subcritical limits for a process. The layout of the report presented in Table 8-1 is a recommended organization of sections to be included in a validation report, with descriptions of information that should be included in each section. Other formats are acceptable as long as the required elements are documented.

Table 8-1 Example Validation Report Layout with Description of Each Section

Section title	Section description
Front matter/cover	Includes names and documentation of authors' and independent
	reviewers'/verifiers' qualifications, titles and tracking numbers, and any QA
page	information needed to uniquely identify the document.
Introduction	Provides a brief description of the purpose of the particular valdiation in terms of
minoduction	the code and general system type being validated.
	Describes the computer hardware and operating system, code, cross section
	library, and any cross section processing codes used to prepare the cross
Computational	sections for use in the Monte Carlo calculations. Descriptions of the options used
method	within the cross section processing and Monte Carlo code should be included in
	this section so that judgements may be made with respect to the applicability to
	the options used in the safety analysis models.
	Defines the types of application systems to be validated by the experiments. The
	definition of the area of applicability is discussed further in Section 4.3 but should
Definition of the	include the dominant fissile and absorbing nuclides in the appropriate energy
area(s) of	spectrum, with applicable moderators and reflectors. It is acceptable for the
applicability	validation report to document more than one area of applicability. This section
	should also identify weaknesses to be addressed, including extrapolations and
	unvalidated nuclides.
	Describes each critical experiment used and defends their selection. The
	experiments should be identified by ICSBEP experiment number (e.g., LEU-
	COMP-THERM-003) if selected from the ICSBEP Handbook [5]. If selected from
Selection of critical	a different source, information to locate the source documentation should be
experiments	provided. Information used to establish the area(s) of applicability should also be
	listed here (e.g., EALF, H/X). The rejection of any experiments, especially from
	within a series of similar experiments, must be documented and defended.
	Discusses the methods chosen to process the validation suite for each area of
	applicability. This section should include a justification for the applicability to the
Validation methods	data set of the statistical method used to generate the bias and bias uncertainty.
	The final bias and bias uncertainty values and equations should also be
	documented here. The application of the bias and bias uncertainty to the safey
	analysis models can be discussed here or on an application-specific basis.
•	A summary of the validation parameters and areas of applicability may be desired
Summary	to facilitate easier use of the document by other analysts.

9 SUMMARY

This document provides recommendations for determining the bias and bias uncertainty of computational methods associated with nuclear criticality safety. Proper implementation of the recommendations presented here should result in a validation in compliance with past guidance [31], [49], [32] and the consensus standard on validation [3].

Section 2 discusses the purpose of validation, which is to establish an appropriate margin of subcriticality for an application by performing criticality calculations for a set of critical experiments with similar materials, geometries, and neutronic characteristics. Validation involves calculating the estimate of the computational method bias and its uncertainty based on the differences between the calculated and experimental k_{eff} values.

Section 3 discusses the definition of a computational method. The most obvious components of the computational method are the computer code and nuclear data being used. Other components include multigroup cross section processing, variance reduction techniques, selection of results, and the computer hardware and operating system.

The selection of critical experiments and determination of the area of applicability are discussed in Section 4, incorporating traditional and/or S/U-based approaches. Appropriate experiment selection is essential to a correct validation. The underlying fundamental assumption is that the safety application case is a member of the same population as the critical experiments in the validation case; this underscores the need for neutronic physics similarity of the application to the critical experiments, and it ensures that the bias and bias uncertainty calculated for the validation suite are appropriate to the application. Section 4.1 describes the importance of characterizing safety analysis models and the impact of the characteristics of the application system on the validation effort. The most important characteristic of the safety analysis model is the set of materials that are present. It is also important to use experiments in validation with the same neutron energy spectrum as the safety application models. Section 4.2 describes critical experiment selection using both traditional and S/U techniques. The discussion also addresses the number of experiments required for validation and specific considerations for burnup credit applications. Section 4.3 discusses establishing the area of applicability. The analyst must confirm that each safety analysis model falls within the applicability of a pre-existing validation or that an appropriate set of experiments has been selected for each specific application. In this regard, having a large library of acceptable experiments available under QA at a site would allow an analyst to select the most appropriate experiments and statistical treatments for each application. This more technically rigorous approach to validation may be problematic at some sites; however, using huge validation suites covering many different types of systems in a validation is not advisable and is discouraged in ANSI/ANS-8.24 [3].

Statistical background information is provided in Section 5. This section introduces hypothesis testing because correct understanding and interpretation of the statistical tools and tests are important to ensure that analysts draw appropriate conclusions from the results. Methods for assessing normality are provided because several of the statistical methods used in validation rely on the assumption that samples have been drawn from a normal distribution. Finally, tools for assessing the goodness of a fit to the sampled data are described.

A range of statistical methods drawn from previous guidance documents is provided in Section 6. The methods presented include nontrending methods (Section 6.1), two nonparametric methods (Section 6.2), and three trending methods (Section 6.3). The equations for weighted and unweighted analysis of both trending and nontrending methods are provided. A wide range of viable statistical methods is presented in Section 6. The application of the results and a summary of the methods (Table 6-5) are discussed in Section 6.6. Two methods are described that have traditionally been used to incorporate the validation bias and bias uncertainty into a criticality analysis to demonstrate compliance with regulatory limits.

No validation suite is perfect. Critical experiments are no longer performed in great number to establish safe operating parameters for individual operations, so contemporary validation suites are constructed from experiments performed for different purposes and/or applications. This can lead to a lack of validation or a weak validation for some components of a safety analysis system. A rigorous, comprehensive assessment of these gaps and weaknesses is important for ensuring conservative safety analyses. Methods for conservatively addressing these gaps and weaknesses are described in Section 7. As in Section 4, both traditional and S/U-based techniques are discussed.

Documentation of the validation is discussed in Section 8. Documentation is important to support independent and regulatory reviews and to facilitate the correct implementation of the results into safety analyses. A variety of implementation options are available based on different regulatory requirements and site practices.

Validation of computational methods is an essential step in the criticality safety analysis process. Careful and correct validation is crucial for the use of simulation results to provide uncompromising safety to workers, the general public, and the environment.

10 REFERENCES

- [1] R. A. Knief, *Nuclear Criticality Safety: Theory and Practice*, American Nuclear Society, La Grange Park, IL (1985).
- [2] ANSI/ANS-8.1-2014 (R2018), *Nuclear Criticality Safety in Operations with Fissionable Materials Outside Reactors*, American Nuclear Society, La Grange Park, IL (2014).
- [3] ANSI/ANS-8.24-2017, Validation of Neutron Transport Methods for Nuclear Criticality Safety Calculations, American Nuclear Society, La Grange Park, IL (2017).
- [4] *Nuclear Criticality Safety Standards for Nuclear Materials Outside Reactor Cores*, Regulatory Guide 3.71, Revision 3, US Nuclear Regulatory Commission, October 2018.
- [5] International Handbook of Evaluated Criticality Safety Benchmark Experiments, NEA/NSC/DOC(95)03, Organisation for Economic Co-operation and Development, Nuclear Energy Agency, Paris, France (2018).
- [6] B. T. Rearden and M. A. Jessee, Eds., SCALE Code System, ORNL/TM-2005/39, Version 6.2.2, Oak Ridge National Laboratory, Oak Ridge, Tennessee (2017). Available from Radiation Safety Information Computational Center as CCC-834.
- [7] C. J. Werner, Ed., *MCNP Users Manual Code Version 6.2, Rev. 0*, LA-UR-17-29981, Los Alamos National Laboratory, 2017.
- [8] D. Long, S. D. Richards, P. N. Smith, C. M. J. Baker, A. J. Bird, N. Davies, G. Dobson, T. C. Fry, D. Hanlon, R. Perry and M. Shepherd, "MONK10: A Monte Carlo Code for Criticality Analysis," ICNC 2015, Charlotte, NC, USA, September 13–17, 2015
- [9] DICE: User's Manual, Rev. 7, NEA/NSC/DOC(95)03/II, Organisation for Economic Cooperation and Development, Nuclear Energy Agency, Nuclear Science Committee, September 2016.
- [10] NCSP Main, <u>http://ncsp.llnl.gov</u>, retrieved May 28, 2019.
- [11] R. D. Cheverton and T. M. Sims, *HFIR Core Nuclear Design*, ORNL-4621, Oak Ridge National Laboratory, July 1971.
- [12] S. J. Raffety and J. T. Thomas, Experimental Determination of Safe Handling Procedures for High Flux Isotope Reactor Fuel Elements Outside the Reactor, ORNL-TM-1488, Oak Ridge National Laboratory, July 1966.
- [13] E. B. Johnson, *Critical Lattices of High Flux Isotope Reactor Fuel Elements*, ORNL-TM-1808, Oak Ridge National Laboratory, March 1967.
- [14] B. T. Rearden, M. L. Williams, M. A. Jessee, D. E. Mueller, and D. A. Wiarda, "Sensitivity and Uncertainty Analysis Capabilities and Data in SCALE," *Nucl. Technol.* **174**, pp. 236– 288 (2011).
- [15] B. T. Rearden et al., *TSUNAMI Primer: A Primer for Sensitivity/Uncertainty Calculations* with SCALE, ORNL/TM-2009/027, Oak Ridge National Laboratory, January 2009.

- [16] I. Hill, "Generation of Low Fidelity Experimental Covariance Matrices for ICSBEP Cases," *Trans. Am. Nucl. Soc.* **115**, pp. 692–695 (2016).
- [17] W. J. Marshall et al., "Development and Testing of Neutron Cross Section Covariance Data for SCALE 6.2," *Proceedings of International Conference on Nuclear Criticality Safety*, Charlotte, NC (2015).
- [18] ANSI/ANS- 8.27-2015, "Burnup Credit for LWR Fuel," American Nuclear Society, La Grange Park, IL, 2015.
- [19] J. M. Scaglione et al., An Approach for Validating Actinide and Fission Product Burnup Credit Criticality Safety Analyses-Criticality (k_{eff}) Predictions, NUREG/CR-7109 (ORNL/TM-2011/514), prepared by Oak Ridge National Laboratory for the US Nuclear Regulatory Commission, April 2012.
- [20] J. J. Sapyta, C. W. Mays, and J. W. Pegram, Jr., "Use of Reactor-Follow Data to Determine Biases and Uncertainties for PWR Spent Nuclear Fuel," *Trans. Am. Nucl. Soc.* 83, 137 (2000).
- [21] D. E. Mueller, K. R. Elam, and P. B. Fox, Evaluation of the French Haut Taux de Combustion (HTC) Critical Experiment Data, NUREG/CR-6979 (ORNL/TM-2007/083), prepared for the US Nuclear Regulatory Commission by Oak Ridge National Laboratory, September 2008.
- [22] F. Fernex, Programme HTC Phase 1: Réseaux de Crayons dans l'Eau Pure (Water-Moderated and Reflected Simple Arrays) Reevaluation des Expériences, DSU/SEC/T/2005-33/D.R, Valduc, France, IRSN (2008). PROPRIETARY document.
- [23] F. Fernex, Programme HTC Phase 2: Réseaux Simples en Eau Empoisonnée (Bore et Gadolinium) (Reflected Simple Arrays Moderated by Poisoned Water with Gadolinium or Boron) Réévaluation des Experiences, DSU/SEC/T/2005-38/D.R, Valduc, France, IRSN (2008). PROPRIETARY document.
- [24] F. Fernex, Programme HTC Phase 3: Configurations "Stockage en Piscine" (Pool Storage) Réévaluation des Experiences, DSU/SEC/T/2005-37/D.R, Valduc, France, IRSN (2008). PROPRIETARY document.
- [25] F. Fernex, *Programme HTC Phase 4: Configurations "Châteaux de Transport"* (*Shipping Cask*) *Réévaluation des Experiences*, DSU/SEC/T/2005-36/D.R., Valduc, France, IRSN (2008). PROPRIETARY document.
- [26] Standard Review Plan for Spent Fuel Dry Storage Systems and Facilities, Chapter 7, Appendix A "Technical Recommendations for the Criticality Safety Review of Pressurized-Water-Reactor Transportation Packages and Storage Casks that Use Burnup Credit," NUREG-2215, April 2020, ADAMS Accession Number ML20121A190.
- [27] V. Sobes, W. J. Marshall, D. Wiarda, F. Bostelmann, A. M. Holcomb, and B. T. Rearden, ENDF/B-VIII.0 Covariance Data Development and Testing for Advanced Reactors, ORNL/TM-2018/1037, Oak Ridge National Laboratory, March 2019.

- [28] I. C. Gauld, Strategies for Application of Isotopic Uncertainties in Burnup Credit, NUREG/CR-6811 (ORNL/TM-2001/257), prepared for the US Nuclear Regulatory Commission by Oak Ridge National Laboratory, June 2003.
- [29] G. Radulescu, I. C. Gauld, G. Ilas, and J. C. Wagner, An Approach for Validating Actinide and Fission Product Burnup Credit Criticality Safety Analyses – Isotopic Composition Prediction, NUREG/CR-7108 (ORNL/TM-2011/509), prepared for the US Nuclear Regulatory Commission by Oak Ridge National Laboratory, December 2011.
- [30] H. Akkurt and K. Cummings, *Utilization of the EPRI Depletion Benchmarks for Burnup Credit Validation–Revision 2*, EPRI Report 3002016888, Electric Power Research Institute, August 2019.
- [31] J. C. Dean and R. W. Tayloe, *Guide for Validation of Nuclear Criticality Safety Calculational Methodology*, NUREG/CR-6698, January 2001.
- [32] J. J. Lichtenwalter, S. M. Bowman, M. D. DeHart, and C. M. Hopper, *Criticality Benchmark Guide for Light-Water-Reactor Fuel in Transportation and Storage Packages*, NUREG/CR-6361, March 1997.
- [33] D. Rutherford, Forecast of Criticality Experiments and Experimental Programs Needed to Support Nuclear Operations in the United States of America: 1994–1999, LA-12683, Los Alamos National Laboratory, March 1994.
- [34] B. L. Broadhead et al., "Sensitivity- and Uncertainty-Based Criticality Safety Validation Techniques," *Nucl. Sci. Eng.* **146**, 340–366, 2004.
- [35] Standard Review Plan for Fuel Cycle Facilities, Chapter 5, Appendix B "Justification for Minimum Margion of Subcriticality for Safety," NUREG-1520, June 2006, ADAMS Accession Number ML061650370.
- [36] D. Lurie, L. Abramson, and J. Vail, *Applying Statistics*, US Nuclear Regulatory Commission, NUREG-1475, Rev. 1, March 2011.
- [37] "Measures of Skewness and Kurtosis," *NIST Engineering Statistics Handbook* <u>http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm</u>, retrieved May 28, 2019.
- [38] N. A Heckert and J. J. Filliben, . *NIST Handbook 148: DATAPLOT Reference Manual, Volume I: Commands*, National Institute of Standards and Technology Handbook Series, June 2003.
- [39] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016 <u>https://www.R-project.org/</u>
- [40] J. Gross and U. Ligges, "Package 'nortest." July 30th 2015, Retrieved from <u>https://cran.r-project.org/web/packages/nortest/nortest.pdf</u>
- [41] H. C. Thode, *Testing for Normality*, New York, New York, 2002, Marcel Dekker.

- [42] C. M. Jarque and A. K. Bera, "Efficient Test for Normality, Homoscedasticity and Serial Independence of Residuals," *Economic Letters*, Vol. 6 Issue 3, 255–259, 1980.
- [43] A. Ghasemi and S. Zahediasl, "Normality Tests for Statistical Analysis: A Guide for Non-Statisticians," *Int. J. Endocrinol. Metab.*, **10**(2) 486–489, 2012.
- [44] N. M. Razali and Y. B. Wah, "Power Comparisions of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests," *JOSMA*, **2**(1), 21–33, 2011.
- [45] R. B. D'Agostino, "Transformation to Normality of the Null Distribution of G1." *Biometrika* 57(3), 679–681, 1970.
- [46] F. J. Anscombe and W. J. Glynn, "Distribution of Kurtosis Statistic for Normal Statistics." *Biometrika* 70(1), 227–234, 1983.
- [47] L. Komsta and F. Novomestky, "Package 'Moments," Retrieved from <u>https://cran.r-project.org/web/packages/moments/moments.pdf</u>. Feb. 20, 2015.
- [48] M. C. Jones and A. Pewsey, "Sinh-Arcsinh Distributions," *Biometrika* **96**(4): 761–780, 2009.
- [49] E. F. Trumble and K. D. Kimball, "Statistical Methods for Accurately Determining Criticality Code Bias," American Nuclear Society Topical Meeting Meeting on Criticality Safety, Chelan, WA, September 7–10, 1997.
- [50] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis*, New York, NY, McGraw-Hill, 2003.
- [51] NRC Information Notice 2011-03, "Nonconservative Criticality Safety Analyses for Fuel Storage," February 2011, ADAMS Accession Number 103090055.
- [52] "Tolerance Intervals for a Normal Distribution," *NIST Engineering Statistics Handbook* <u>http://www.itl.nist.gov/div898/handbook/prc/section2/prc263.htm</u>, retrieved May 28, 2019.
- [53] M. G. Natrella, *Experimental Statistics*, National Bureau of Standards, August 1963. <u>http://www.keysolutionsinc.com/references/NBS%20Handbook%2091.pdf</u>, retrieved May 28, 2019.
- [54] B. C. Kiedrowski, *Methodology for Sensitivity and Uncertainty-Based Criticality Safety Validation*, LA-UR-14-23202, Los Alamos National Laboratory, Los Alamos, NM, 2014.
- [55] W. J. Conover, *Practical Nonparametric Statistics*, 2nd Edition, John Wiley & Sons, New York, 1980.
- [56] I. Duhamael, J. L. Alwin, F. B. Brown, et al., "International Criticality Benchmark Comparison for Nuclear Data Validation," *Trans. Am. Nucl. Soc.* **121** pp. 873–876 (2019).
- [57] A. C. Tamhane and D. D. Dunlop, *Statistics and Data Analysis from Elementary to Intermediate*, Prentice Hall, Upper Saddler River, NJ, 2000.

- [58] R. G. Miller, *Simultaneous Statistical Inference*, 2nd Edition, Springer-Verlag, New York, 1981.
- [59] D. C. Bowden and F. A. Graybill, "Confidence Bands of Uniform and Proportional Width for Linear Models," *Am. Stat. Associ.* **61**, pp. 182–198, March 1966.
- [60] N. L. Johnson, Editor, "Query," *Technometrics* 10, 207–209, Feburuary 1968.
- [61] A. Hoefer et al., "Proposal for Benchmark Phase IV Role of Integral Experiment Covariance Data for Criticality Safety Validation," Benchmark Proposal, OECD/NEA (2015).
- [62] W. J. Marshall and B. T. Rearden, "Determination of Critical Experiment Correlations Using the Sampler Sequence within SCALE 6.2," *Proceedings of International Conference on Nuclear Criticality Safety*, Charlotte, NC, 2015.
- [63] V. Sobes, B. T. Rearden, D. E. Mueller, W. J. Marshall, J. M. Scaglione, and M. E. Dunn, "Upper Subcritical Limit Calculations Based on Correlated Experimental Data," International Conference on Nuclear Criticality Safety (ICNC 2015), Charlotte, NC, 2015.
- [64] Criticality Accident Requirements, 10 CFR 50.68, November 2006.

APPENDIX A EXAMPLE CALCULATIONS OF BIAS AND BIAS UNCERTAINTY

APPENDIX A EXAMPLE CALCULATIONS OF BIAS AND BIAS UNCERTAINTY

This section provides three examples of how the methods described in the body of the report can be implemented to calculate the biases and bias uncertainties needed to perform validation. Each of the data sets is drawn from the SCALE 6.2.2 validation report [A-1].

The bias of the computational method should be examined as a function of potentially meaningful independent variables first to provide an estimate of the bias that may be more reflective of the application system. The examples used here investigate the bias as a function of EALF to demonstrate the method, but several parameters should generally be examined. If no statistically significant trend in the data is found, then other nontrending analyses are used. The thought process associated with bias and bias uncertainty calculation techniques is discussed further in Section 6. The following subsections discuss the implementation of the techniques that have been discussed throughout the document. It is noted that there are other potential interpretations that may also be valid and what follows is only one implementation.

A.1 Example 1

Experiment	C/E	$\sigma_{c/E}$	EALF
U233-COMP-THERM-001-002	1.00082	0.00250	1.54E+00
U233-COMP-THERM-001-003	1.00227	0.00241	7.82E-01
U233-COMP-THERM-001-004	1.00058	0.00250	4.64E-01
U233-SOL-INTER-001-001	0.98522	0.00818	6.86E+00
U233-SOL-INTER-001-002	0.98053	0.00834	8.01E+00
U233-SOL-INTER-001-003	0.98126	0.00648	8.62E+00
U233-SOL-INTER-001-004	0.99217	0.00605	3.72E+00
U233-SOL-INTER-001-005	0.98462	0.00807	9.22E+00
U233-SOL-INTER-001-006	0.98553	0.00601	4.29E+00
U233-SOL-INTER-001-007	0.98174	0.00579	9.65E+00
U233-SOL-INTER-001-008	0.98015	0.00549	4.55E+00
U233-SOL-INTER-001-009	0.97978	0.00666	7.35E+00
U233-SOL-INTER-001-010	0.97862	0.00519	1.01E+01
U233-SOL-INTER-001-011	0.98018	0.00559	7.76E+00
U233-SOL-INTER-001-012	0.98123	0.00893	4.46E+00
U233-SOL-INTER-001-013	0.98207	0.00697	5.09E+00
U233-SOL-INTER-001-015	0.98000	0.00735	5.47E+00
U233-SOL-INTER-001-017	0.98848	0.00544	2.55E+00
U233-SOL-INTER-001-018	0.97843	0.00558	5.85E+00
U233-SOL-INTER-001-019	0.97523	0.00810	6.10E+00

Table A-1 U-233 Critical Experiment Data Used in Example 1

Experiment	C/E	$\sigma_{c/E}$	EALF
U233-SOL-INTER-001-020	0.97965	0.00549	3.01E+00
U233-SOL-INTER-001-021	0.97295	0.00487	6.36E+00
U233-SOL-INTER-001-022	0.97805	0.00479	6.51E+00
U233-SOL-INTER-001-023	0.99000	0.00465	4.72E+00
U233-SOL-INTER-001-024	0.99236	0.00804	1.99E+00
U233-SOL-INTER-001-025	0.98533	0.00798	2.28E+00
U233-SOL-INTER-001-026	0.98907	0.00643	2.40E+00
U233-SOL-INTER-001-028	0.98357	0.00600	2.56E+00
U233-SOL-INTER-001-029	0.97727	0.00958	2.66E+00
U233-SOL-INTER-001-031	0.99085	0.00704	2.70E+00
U233-SOL-INTER-001-032	0.97571	0.00517	2.81E+00
U233-SOL-INTER-001-033	0.99382	0.00457	2.08E+00
U233-SOL-MIXED-001-014	0.98998	0.00515	2.30E+00
U233-SOL-MIXED-001-016	0.97444	0.00273	1.81E+00
U233-SOL-MIXED-001-030	0.97763	0.00518	1.46E+00
U233-SOL-MIXED-002-003	0.98645	0.00671	1.30E+00
U233-SOL-MIXED-002-005	0.98620	0.00543	1.37E+00
U233-SOL-MIXED-002-006	0.97657	0.00967	1.42E+00
U233-SOL-MIXED-002-008	0.97311	0.00652	1.47E+00
U233-SOL-MIXED-002-009	0.96863	0.00484	1.50E+00
U233-SOL-THERM-001-001	1.00141	0.00311	3.94E-02
U233-SOL-THERM-001-002	1.00067	0.00330	4.00E-02
U233-SOL-THERM-001-003	1.00012	0.00330	4.06E-02
U233-SOL-THERM-001-004	1.00103	0.00331	4.12E-02
U233-SOL-THERM-001-005	1.00025	0.00330	4.18E-02
U233-SOL-THERM-002-001	1.00187	0.00868	1.74E-01
U233-SOL-THERM-002-002	0.98990	0.00858	1.34E-01
U233-SOL-THERM-002-003	1.00592	0.00872	1.05E-01
U233-SOL-THERM-002-004	1.00252	0.00869	8.50E-02
U233-SOL-THERM-002-005	1.00738	0.00873	7.41E-02
U233-SOL-THERM-002-006	0.99400	0.00861	6.60E-02
U233-SOL-THERM-002-007	0.98375	0.00853	6.25E-02
U233-SOL-THERM-002-008	0.99783	0.00865	5.75E-02
U233-SOL-THERM-002-009	0.98647	0.00855	5.19E-02
U233-SOL-THERM-002-010	0.99937	0.00866	5.01E-02
U233-SOL-THERM-002-011	1.00803	0.00874	4.67E-02

Experiment	C/E	$\sigma_{c/e}$	EALF
U233-SOL-THERM-002-012	0.98872	0.00857	2.73E-01
U233-SOL-THERM-002-013	0.98852	0.00857	4.88E-01
U233-SOL-THERM-002-014	0.99754	0.00864	1.39E-01
U233-SOL-THERM-002-015	1.00412	0.00870	9.56E-02
U233-SOL-THERM-002-017	1.00640	0.00872	5.44E-02
U233-SOL-THERM-003-001	1.00214	0.00873	3.15E-01
U233-SOL-THERM-003-002	1.01754	0.01541	3.50E-01
U233-SOL-THERM-003-003	0.99854	0.00869	3.36E-01
U233-SOL-THERM-003-004	1.00206	0.01259	7.88E-01
U233-SOL-THERM-003-005	1.00933	0.01233	1.06E+00
U233-SOL-THERM-003-006	1.02084	0.00888	1.30E-01
U233-SOL-THERM-003-007	1.01405	0.00882	8.37E-02
U233-SOL-THERM-003-008	1.01003	0.00878	6.91E-02
U233-SOL-THERM-003-009	1.00984	0.00878	6.20E-02
U233-SOL-THERM-003-010	1.00751	0.00877	4.62E-02
U233-SOL-THERM-004-001	0.99887	0.00876	1.72E-01
U233-SOL-THERM-004-002	1.00241	0.00859	1.33E-01
U233-SOL-THERM-004-003	0.99177	0.00879	2.73E-01
U233-SOL-THERM-004-004	0.98154	0.00869	4.89E-01
U233-SOL-THERM-004-005	0.98867	0.00887	3.86E-01
U233-SOL-THERM-004-006	1.00145	0.01049	4.86E-01
U233-SOL-THERM-004-007	0.99817	0.01036	3.83E-01
U233-SOL-THERM-004-008	1.00466	0.01023	1.38E-01
U233-SOL-THERM-005-001	1.00160	0.00401	6.19E-02
U233-SOL-THERM-005-002	1.00456	0.00492	5.45E-02
U233-SOL-THERM-008-001	1.00089	0.00290	3.72E-02
U233-SOL-THERM-009-001	0.99933	0.00441	3.79E-02
U233-SOL-THERM-009-002	1.00118	0.00401	3.75E-02
U233-SOL-THERM-009-003	1.00163	0.00381	3.71E-02
U233-SOL-THERM-009-004	0.99981	0.00380	3.67E-02
U233-SOL-THERM-011-027	0.99022	0.00505	1.25E+00
U233-SOL-THERM-012-001	1.00010	0.00280	1.73E-01
U233-SOL-THERM-012-002	1.00004	0.00250	1.66E-01
U233-SOL-THERM-012-003	1.00998	0.00233	1.48E-01
U233-SOL-THERM-012-004	1.00247	0.00151	1.08E-01
U233-SOL-THERM-012-005	1.00412	0.00713	9.14E-02

Experiment	C/E	$\sigma_{\text{C/E}}$	EALF
U233-SOL-THERM-012-006	1.00523	0.00111	8.06E-02
U233-SOL-THERM-012-007	1.00126	0.00381	5.43E-02
U233-SOL-THERM-012-008	0.99854	0.00479	5.41E-02
U233-SOL-THERM-013-001	1.00548	0.00735	1.57E-01
U233-SOL-THERM-013-002	1.00557	0.00705	1.57E-01
U233-SOL-THERM-013-003	1.00562	0.00695	1.58E-01
U233-SOL-THERM-013-004	1.00595	0.00735	1.59E-01
U233-SOL-THERM-013-005	1.00653	0.00675	1.60E-01
U233-SOL-THERM-013-006	1.00647	0.00504	1.51E-01
U233-SOL-THERM-013-007	1.00629	0.00544	1.52E-01
U233-SOL-THERM-013-008	1.00681	0.00504	1.52E-01
U233-SOL-THERM-013-009	1.00707	0.00454	1.53E-01
U233-SOL-THERM-013-010	1.00769	0.00464	1.53E-01
U233-SOL-THERM-013-011	1.00529	0.00543	1.50E-01
U233-SOL-THERM-013-012	1.00621	0.00504	1.52E-01
U233-SOL-THERM-013-013	1.00372	0.00623	1.51E-01
U233-SOL-THERM-013-014	1.00642	0.00514	1.53E-01
U233-SOL-THERM-013-015	1.02100	0.00787	1.05E-01
U233-SOL-THERM-013-016	0.99338	0.00686	8.88E-02
U233-SOL-THERM-013-017	0.99596	0.00518	8.63E-02
U233-SOL-THERM-013-018	1.00020	0.00200	8.05E-02
U233-SOL-THERM-013-019	0.99634	0.00887	8.07E-02
U233-SOL-THERM-013-020	0.99820	0.00559	6.31E-02
U233-SOL-THERM-013-021	1.00238	0.00341	5.67E-02
U233-SOL-THERM-015-001	0.98999	0.00743	1.10E+00
U233-SOL-THERM-015-002	0.98558	0.00690	1.24E+00
U233-SOL-THERM-015-004	0.98933	0.00406	7.21E-01
U233-SOL-THERM-015-007	0.98653	0.00691	7.95E-01
U233-SOL-THERM-015-010	0.98994	0.00505	1.13E+00
U233-SOL-THERM-015-011	0.99315	0.00745	6.93E-01
U233-SOL-THERM-015-012	0.99395	0.00686	7.67E-01
U233-SOL-THERM-015-013	0.99177	0.00684	8.07E-01
U233-SOL-THERM-015-014	0.99644	0.00359	4.66E-01
U233-SOL-THERM-015-015	0.98968	0.00594	8.47E-01
U233-SOL-THERM-015-016	0.98839	0.00425	8.68E-01
U233-SOL-THERM-015-017	0.99653	0.00289	5.05E-01

Experiment	C/E	$\sigma_{\text{C/E}}$	EALF
U233-SOL-THERM-015-018	0.97430	0.00546	8.98E-01
U233-SOL-THERM-015-019	0.97484	0.00507	9.12E-01
U233-SOL-THERM-015-020	0.99483	0.00786	2.89E-01
U233-SOL-THERM-015-021	0.99794	0.00699	3.15E-01
U233-SOL-THERM-015-022	0.99612	0.00618	3.30E-01
U233-SOL-THERM-015-023	0.99407	0.00547	3.43E-01
U233-SOL-THERM-015-024	0.99084	0.00505	3.51E-01
U233-SOL-THERM-015-025	0.99585	0.00229	2.25E-01
U233-SOL-THERM-015-026	0.99396	0.00656	1.27E-01
U233-SOL-THERM-015-027	0.99871	0.00629	1.31E-01
U233-SOL-THERM-015-028	0.99682	0.00578	1.33E-01
U233-SOL-THERM-015-029	0.99532	0.00508	1.35E-01
U233-SOL-THERM-015-030	0.99465	0.00478	1.37E-01
U233-SOL-THERM-015-031	0.99381	0.00547	1.38E-01
U233-SOL-THERM-016-001	1.00439	0.00372	2.91E-01
U233-SOL-THERM-016-002	1.00577	0.00443	2.92E-01
U233-SOL-THERM-016-003	1.00458	0.00362	2.92E-01
U233-SOL-THERM-016-004	1.00600	0.00353	2.91E-01
U233-SOL-THERM-016-006	0.99620	0.00339	2.91E-01
U233-SOL-THERM-016-007	0.99530	0.00338	2.91E-01
U233-SOL-THERM-016-008	0.99461	0.00278	2.91E-01
U233-SOL-THERM-016-009	0.99569	0.00269	2.91E-01
U233-SOL-THERM-016-010	1.00389	0.00301	2.88E-01
U233-SOL-THERM-016-011	1.00456	0.00412	2.88E-01
U233-SOL-THERM-016-012	1.00485	0.00473	2.88E-01
U233-SOL-THERM-016-013	1.00505	0.00362	1.45E-01
U233-SOL-THERM-016-014	1.00482	0.00262	1.45E-01
U233-SOL-THERM-016-015	1.00582	0.00272	1.45E-01
U233-SOL-THERM-016-016	1.00965	0.00313	1.46E-01
U233-SOL-THERM-016-017	0.99429	0.00279	1.47E-01
U233-SOL-THERM-016-018	0.99588	0.00359	1.46E-01
U233-SOL-THERM-016-019	0.99486	0.00348	1.46E-01
U233-SOL-THERM-016-021	1.00902	0.00283	1.44E-01
U233-SOL-THERM-016-022	1.00882	0.00343	1.44E-01
U233-SOL-THERM-016-023	1.00895	0.00313	1.44E-01
U233-SOL-THERM-016-024	1.00839	0.00242	1.44E-01

Experiment	C/E	$\sigma_{\text{C/E}}$	EALF
U233-SOL-THERM-016-025	1.00106	0.00401	8.16E-02
U233-SOL-THERM-016-026	1.00679	0.00343	8.16E-02
U233-SOL-THERM-016-027	1.00384	0.00372	8.15E-02
U233-SOL-THERM-016-028	0.99960	0.00371	8.14E-02
U233-SOL-THERM-016-029	1.00033	0.00311	8.14E-02
U233-SOL-THERM-016-030	0.99923	0.00320	8.14E-02
U233-SOL-THERM-016-030	1.01071	0.00344	5.74E-02
U233-SOL-THERM-016-031	1.01307	0.00325	5.74E-02
U233-SOL-THERM-016-032	1.01283	0.00396	5.74E-02
U233-SOL-THERM-016-033	1.00454	0.00322	1.15E-01
U233-SOL-THERM-017-001	1.00051	0.00250	1.12E-01
U233-SOL-THERM-017-002	1.00519	0.00352	1.09E-01
U233-SOL-THERM-017-003	1.00581	0.00403	8.32E-02
U233-SOL-THERM-017-004	1.00240	0.00291	8.15E-02
U233-SOL-THERM-017-005	1.00021	0.00290	5.44E-02
U233-SOL-THERM-017-006	0.99990	0.00370	5.54E-02

A.1.1 Assessment of Trends

To determine whether the bias in the computational method is best expressed as function of EALF for the ²³³U solution systems in the Example 1 data set provided in Table A-1, weighted and unweighted trendlines are constructed, and their statistical significance is tested using the methods discussed Sections 6.3 (trend calculation) and 5.3 (t-test). As the data span a few orders of magnitude, trends using the raw EALF values and the natural logarithm of the EALF values were investigated. The best fit on EALF was produced when first taking the natural logarithm of the EALF values so that set of calculations is discussed here. A similar approach to trending analysis is performed in Appendix D to ANSI/ANS-8.24-2017 [3]. The data set contains 180 experiments, with an unweighted average C/E = 0.996612 and an unweighted average $\ln(EALF) = -1.19943$. Using the uncertainty of each experiment to weight the data, the weighted average C/E = 0.999391, and the weighted $\ln(EALF) = -1.63147$.

To develop the trendlines, the S_{xx} and S_{kx} values were calculated using Eq. (21) and Eq. (23) for an unweighted analysis or Eq. (22) and Eq. (24) for a weighted analysis. The slope and intercept of the trendlines are then calculated using Eq. (25) and Eq. (26), respectively. The residuals about the trendline are then calculated using Eq. (27) and the uncertainty in the fit is calculated using Eq. (28) for a unweighted analysis and (29) for a weighted analysis. The t statistics for the fits are then calculated using Eq. (4).

Once the t-statistic for the fit is calculated, there are two equivalent methods by which the statistical significance of the fit can be judged. The first method is by comparison of the t statistic

for the fit to the critical value from the t-distribution corresponding to *n*-2 degrees of freedom and 95% confidence. For this example, the comparison is made to the t-statistic corresponding to $\alpha/2=0.025$ (two-sided test because the slope can depart from zero in either the positive or negative direction, so the value of α is divided by 2) and 178 degrees of freedom (*n*-2). The t-statistics for the weighted and unweighted fits were found to be 13.34 and 14.41, respectively. Because these values are larger than the appropriate critical values from the t-distribution of 1.97, the null hypothesis that there is no slope is rejected in favor of the calculated trendline. The second method for assessing the statistical significance of a trendline is to calculate the p-value of the t-statistic with statistical software and compare it to the prescribed value of α . In this case, the p-values are 4.34e-26 and 3.34e-31 for the weighted and unweighted calculations, respectively. Because these values are less than 0.025, the null hypothesis of zero slope is again rejected in favor of the calculated trendline. The parameters calculated in the determination of the trendline and the assessment of the statistical significance of the trendline are presented in Table A-2.

Parameter	Weighted value	Unweighted value
S _{xx}	300.987172	425.815251
S _{kx}	-1.41940943	-2.12011093
β_0	0.991697698	0.99063965
βı	-0.00471585	-0.0049789
$\sigma_{\it fit}$	0.00612913	0.0071257
<i>t_{fit}</i>	13.348571	14.418539
tcritical	1.9734	1.9734
p-value	4.3450e-26	3.3417e-31

Table A-2 Trend Statistical Significance Parameters for the Example 1 Data

Once the calculated trend is determined to be statistically significant, it can be used in conjunction with the values of the trend parameter from the application cases to calculate the bias and bias uncertainty. For this example, the EALFs of three hypothetical application cases considered are 0.05, 1, and 50 eV, which have natural logarithms of -2.99573, 0, and 3.91202, respectively.

To evaluate the bias and bias uncertainty as a function of EALF, a number of additional statistical parameters must be evaluated. Two bands are used for the bias uncertainty evaluation in this appendix: the *single-sided lower tolerance band method* from NUREG/CR-6698, and the *single-sided lower confidence band* from NUREG/CR-6361, which is often referred to as the *USL-1 method*. For both methods, the average total uncertainty ($\bar{\sigma}$) is needed and is calculated using Eq. (11) for unweighted analyses and Eq. (12) for weighted analyses. The pooled standard deviation (S_p) can then be calculated using Eq.(31) for a trending analysis. Additionally, the statistical parameters for the NUREG/CR-6698 lower tolerance band from Table 6-4 and the inverse t-distribution calculation are needed for the NUREG/CR-6361 lower confidence band calculation. The values of these parameters are presented in Table A-3.

Parameter	Weighted value	Unweighted value
$\bar{\sigma}$	0.004147	0.006218
S _p	0.007400	0.009457
$F_{\alpha}^{(2,n-2)}$	3.04672	3.04672
n	180	180
Z _P	1.64485	1.64485
$\chi_{1-\frac{\alpha}{2},n-2}^2$	142.94862	142.94862
$t_{1-\alpha,n-2}$	1.65345	1.65345

Table A-3 Parameters Necessary for Trended Bias Uncertainty Evaluation for the Example 1 Data

Using the parameters calculated above, the trended value of calculation/experiment (C/E) $(k_{fit}(x))$, the bias and bias uncertainty are estimated for each of the three application cases and are reported in Table A-4. The values of $k_{fit}(x)$ and the trended biases in Table A-4 show that the bias is set to zero in cases that would result in a positive bias. In this example, the 0.05 eV case has the bias conservatively set to zero, whereas the 1 eV and 50 eV cases use the trendline to estimate the biases. The 50 eV case requires extrapolation of the trend because the maximum EALF in the data set is 10.1 eV. The bias uncertainties are estimated by the NUREG/CR-6698 and USL-1 bands. It is noted that for the USL-1 calculation, the bias uncertainty is calculated as the maximum of the values found by evaluation at the end points of the energy range and applied to all values in between (0.05 and 1 eV). For the 50 eV case, the bias uncertainty is evaluated at that point.

fc	or the Example	es Cases				
	0.0	05 eV	1	eV	Ę	50 eV
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
$k_{fit}(x)$	1.00583	1.00556	0.99170	0.99064	0.97325	0.97116
Tranded bias	0	0	-0.00830	-0.00036	-0.02675	-0 02884

Table A-4 Sample Calculation of Bias, Bias Uncertainties, and Calculational Margins

	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
$k_{fit}(x)$	1.00583	1.00556	0.99170	0.99064	0.97325	0.97116
Trended bias	0	0	-0.00830	-0.00936	-0.02675	-0.02884
NUREG/CR-6698 bias uncertainty	0.01556	0.02003	0.01577	0.01956	0.01958	0.02339
NUREG/CR-6698 calculational margin	0.01556	0.02003	0.02407	0.02892	0.04633	0.05223
USL-1 bias uncertainty	0.01258	0.01590	0.01258	0.01590	0.01288	0.01615
USL-1 calculational margin	0.01258	0.01590	0.02088	0.02526	0.03963	0.04499

A.1.2 Nontrending Assessment

Even though there is a statistically significant trend associated with the Example 1 data set, the process necessary to determine the bias and bias uncertainty without the trend is shown here. When assessing the nontrending bias, there are two methods for calculating the bias uncertainty: the normality-based LTL and the nonparametric LTL.

To apply the normality-based LTL, the assumption that the data come from a normal distribution must be shown to be appropriate or conservative. To determine if a parametric assessment of the untrended bias and bias uncertainty is appropriate, the assumption of normality is assessed for the Example 1 data set. The data were standardized using Eq. (3) and the standardized data were used to create the histogram and Q-Q plot in Figure A-1, and they were also used to conduct a number of omnibus normality tests, the results of which are shown in Table A-5. In the plot of the Z-scores in the left half of Figure A-1, the data appear to be slightly negatively skewed. Similarly, the Q-Q plot has a slight downward facing "C," which is similar to the right portion of Figure 5-3, also indicating a negative skewness. Additionally, except for the Kolmogorov-Smirnov test, the p-values for each of the omnibus normality tests produced here are less than 0.05. The p-value of less than 0.05 indicates that the null hypothesis that the data are a sample drawn from a normal distribution is rejected in favor of the alternative hypothesis that the data have some other underlying distribution at the 95% confidence level. It is noted that the Kolmogorov-Smirnov results are included to show the insensitivity to the data in the tails and to highlight why it may not be effective for these assessments.



Figure A-1 Histogram (Left) and Normal Q-Q Plot (Right) of the Example 1 Data Set

Test Test	p-value
Chi-square (16 bins)	5.633e-05
Anderson-Darling	2.605e-06
Cramer-Von Mises	1.475e-05
Lilliefors	1.987e-04
Shapiro-Wilk	1.159e-04
Shapiro-Francia	4.274e-04
Jarque-Bera	0.01948
Kolmogorov-Smirnov	0.05950

Table A-5 Summary of Example 1 Normality Test P-Values

Having concluded that the data are not drawn from a normal distribution, an assessment of the conservatism of applying a normal distribution-based calculation of the bias uncertainty is performed. In order to perform this evaluation, the skewness and kurtosis of the data are calculated, and the D'Agostino skewness and the Anscombe-Glynn kurtosis single-sided tests are performed, the results of which are shown in Table A-6. The results in Table A-6 show that the calculated skewness is -0.485, which is less than zero corresponding to a normal distribution, and that the skewness is statistically significantly less than zero at the 95% confidence level because the p-value for the single-sided D'Agostino skewness test is less than 0.05. Based on the skewness results, it is appropriate to conclude that the data are negatively skewed in a statistically significant way. The kurtosis of the data is assessed in a similar manner, and the results are also presented in Table A-6. The kurtosis of the data is calculated to be 2.664, which is less than a normal value of 3, indicating that the distribution data are slightly platykurtic (heavy tails). The Anscombe-Glynn kurtosis test results indicate that, although the data have a calculated kurtosis of less than 3, the data are not statistically significantly less than 3 at the 95% confidence level because the p-value is greater than 0.05.

Because the data are negatively skewed, it is not appropriate to use the normality-based LTL, and the nonparametric statistical treatment should be used. If the skewness had been positive or statistically indistinguishable from zero and kurtosis had been less than three or if it were statistically indistinguishable from three (as it is), then it would have been acceptable to the use the normal distribution-based LTL calculation.

Table A-6 Summary of Example 1 Skewness and Kurtosis Assessment

Parameter	Value
Skewness	-0.485
Skewness test p-value	4.009e-3
Kurtosis	2.664
Kurtosis test p-value	0.2213

The assessment of the bias is the same for the nontrending analysis, regardless of the whether a parametric or nonparametric analysis is used. The bias assessment is simply the difference

between the mean C/E value (\overline{k}_{norm}) and unity. The bias uncertainty is then evaluated using the traditional historical nonparametric method described in Section 6.2.1. Using the data in Table 6-2 , it is determined that, because the validation suite includes more than 153 points, the fourth lowest point can be used to attain 95% confidence that 95% of the population of C/E lie above it. Therefore, the nonparametric LTL is produced by subtracting the uncertainty in the C/E from the C/E value that is the fourth lowest in the suite. Hence, in this case, the fourth lowest C/E is 0.97430 with an uncertainty of 0.00546, which results in a nonparametric LTL of 0.96884. For cases in which the bias uncertainty can be convoluted with other uncertainties in the safety analysis case, the LTL can be split into the bias by subtracting the mean C/E value from the LTL and calculating the bias as one minus the mean C/E value. The \overline{k}_{norm} , the bias, and the bias uncertainty are calculated for the Example 1 data set, and the results are shown in Table A-7 as weighted and unweighted calculation of the mean C/E value.

Parameter	Weighted value	Unweighted value
\overline{k}_{norm}	0.99939	0.99661
β	-0.00061	-0.00339
σ_{eta}	0.03055	0.02777

Table A-7 Nonparametric Validation Parameters for the Example 1 Data Set

A graphical summary of the validation parameters is shown in Figure A-2 for the Example 1 data set. The calculations shown in Figure A-2 do not have any administrative margin and are meant to provide a comparison of the methods. Unweighted calculations are shown as dashed lines in the same color as the associated weighted calculations.





A.2 Example 2

Experiment	C/E	$\sigma_{c/E}$	EALF
HEU-SOL-THERM-001-001	0.99656	0.00598	8.16E-02
HEU-SOL-THERM-001-002	0.99280	0.00713	2.72E-01
HEU-SOL-THERM-001-003	1.00009	0.00350	8.00E-02
HEU-SOL-THERM-001-004	0.99648	0.00528	2.91E-01
HEU-SOL-THERM-001-005	0.99706	0.00489	4.34E-02
HEU-SOL-THERM-001-006	1.00044	0.00460	4.49E-02
HEU-SOL-THERM-001-007	0.99591	0.00398	7.74E-02
HEU-SOL-THERM-001-008	0.99729	0.00379	8.17E-02
HEU-SOL-THERM-001-009	0.99258	0.00536	2.91E-01
HEU-SOL-THERM-001-010	0.99158	0.00536	4.65E-02
HEU-SOL-THERM-013-001	0.99700	0.00259	3.25E-02
HEU-SOL-THERM-013-002	0.99673	0.00359	3.39E-02
HEU-SOL-THERM-013-003	0.99311	0.00357	3.53E-02
HEU-SOL-THERM-013-004	0.99526	0.00358	3.60E-02
HEU-SOL-THERM-014-001	0.99359	0.00278	4.57E-02
HEU-SOL-THERM-014-002	1.01012	0.00526	4.72E-02
HEU-SOL-THERM-014-003	1.01898	0.00887	4.93E-02
HEU-SOL-THERM-016-001	0.98969	0.00357	7.76E-02
HEU-SOL-THERM-016-002	1.00560	0.00694	8.15E-02
HEU-SOL-THERM-016-003	1.02444	0.00810	9.05E-02
HEU-SOL-THERM-028-001	0.99564	0.00229	4.75E-02
HEU-SOL-THERM-028-002	0.99650	0.00339	4.80E-02
HEU-SOL-THERM-028-003	0.99757	0.00260	4.76E-02
HEU-SOL-THERM-028-004	0.99814	0.00280	4.80E-02
HEU-SOL-THERM-028-005	0.99294	0.00308	4.76E-02
HEU-SOL-THERM-028-006	0.99650	0.00229	4.80E-02
HEU-SOL-THERM-028-007	0.99700	0.00379	4.79E-02
HEU-SOL-THERM-028-008	0.99695	0.00269	4.81E-02
HEU-SOL-THERM-028-009	0.99566	0.00488	1.43E-01
HEU-SOL-THERM-028-010	0.99430	0.00527	1.45E-01
HEU-SOL-THERM-028-011	0.99703	0.00509	1.44E-01
HEU-SOL-THERM-028-012	0.99460	0.00458	1.47E-01
HEU-SOL-THERM-028-013	0.99615	0.00578	1.47E-01
HEU-SOL-THERM-028-014	0.99617	0.00458	1.49E-01

Table A-8HEU Solution Critical Experiment Data Used in Example 2

Experiment	C/E	σ _{c/e}	EALF
HEU-SOL-THERM-028-015	1.00427	0.00643	1.48E-01
HEU-SOL-THERM-028-016	1.00034	0.00520	1.51E-01
HEU-SOL-THERM-028-017	0.99565	0.00657	1.50E-01
HEU-SOL-THERM-028-018	0.99645	0.00598	1.52E-01
HEU-SOL-THERM-029-001	0.99770	0.00659	1.55E-01
HEU-SOL-THERM-029-002	1.00158	0.00581	1.54E-01
HEU-SOL-THERM-029-003	0.99372	0.00676	1.56E-01
HEU-SOL-THERM-029-004	0.99267	0.00735	1.63E-01
HEU-SOL-THERM-029-005	0.99757	0.00668	1.66E-01
HEU-SOL-THERM-029-006	0.99771	0.00649	1.66E-01
HEU-SOL-THERM-029-007	0.99822	0.00629	1.65E-01
HEU-SOL-THERM-030-001	0.99569	0.00388	4.81E-02
HEU-SOL-THERM-030-002	0.99652	0.00319	4.88E-02
HEU-SOL-THERM-030-003	0.99558	0.00309	4.84E-02
HEU-SOL-THERM-030-004	0.99980	0.00640	1.56E-01
HEU-SOL-THERM-030-005	0.99595	0.00578	1.57E-01
HEU-SOL-THERM-030-006	0.99785	0.00589	1.59E-01
HEU-SOL-THERM-030-007	0.99725	0.00638	1.63E-01

 Table A-8
 HEU Solution Critical Experiment Data Used in Example 2 (continued)

A.2.1 Assessment of Trends

To determine if the bias in the computational method is best expressed as function of EALF for the HEU solution systems presented in the Example 2 data set presented in Table A-8, weighted and unweighted trendlines are constructed, and their statistical significance is tested using the methods discussed Sections 6.3 (trend calculation) and 5.3 (t-test). The better fit on EALF was obtained when first taking the natural logarithm of the EALF values and then fitting. The data set contains 52 experiments, with an unweighted average C/E = 0.997788 and an unweighted average In(EALF) = -2.42246. Using the uncertainty of each experiment to weight the data, the weighted average C/E = 0.996659, and the weighted In(EALF) = -2.73842.

First the S_{xx} and S_{kx} values are calculated using Eq. (21) and Eq. (23) for an unweighted analysis or Eq. (22) and Eq. (24) for a weighted analysis. The slope and intercept of the trendlines are then calculated using Eq. (25) and Eq. (26), respectively. The residuals about the trendline are then calculated using Eq. (27), and the uncertainty in the fit is calculated using Eq. (28) for a unweighted analysis and Eq. (29) for a weighted analysis. The t-statistics for the fits are then calculated using Eq. (4).

Once the t-statistic for the fit is calculated, there are two equivalent methods that can be used to determine the statistical significance of the fit. The first method is by comparing the t-statistic for the fit to the critical value from the t-distribution corresponding to n-2 degrees of freedom and 95% confidence. For this example, the critical value is compared to the t-statistic corresponding to $\alpha/2=0.025$ (a two-sided test because the slope can depart from zero in either the positive or

negative direction, so the value of α is divided by 2) and 50 degrees of freedom (n-2). The tstatistics for the weighted fit was 0.2947 and 0.5618 for the unweighted fit. Because these values are smaller than the appropriate critical values from the t-distribution of 2.0086, the null hypothesis that there is no slope is not rejected in favor of the calculated trendline. The second method of assessing the statistical significance of a trendline is to calculate the p-value of the tstatistic with statistical software and compare the result to the prescribed value of α . In this case, the p-value for the weighted calculation is 0.3798, and it is 0.3381 for the unweighted calculation. Because these values are greater than 0.025, the null hypothesis of zero slope is again not rejected in favor of the calculated trendline. The parameters calculated in the determination of the trendline and the assessment of the statistical significance of the trendline are presented in Table A-9.

Parameter	Weighted value	Unweighted value
S _{xx}	16.750758	20.8918558
S _{kx}	0.0045487	-0.0153275
β ₀	0.9974029	0.9960108
β ₁	0.0002716	-0.0007337
σ _{fit}	0.0037717	0.0059669
t _{fit}	0.2947	0.5618
t _{critical}	2.0086	2.0086
p-value	0.3798	0.3381

Table A-9 Trend Statistical Significance Parameters for the Example 2 Data

A.2.2 Nontrending Assessment

Because it has been determined that no statistically significant trend is associated with the Example 2 data set, the process necessary to determine the bias and bias uncertainty without the trend is shown here. When assessing the nontrending bias, there are two methods of calculating the bias uncertainty, the normality-based LTL, and the nonparametric LTL.

To apply the normality-based LTL, the assumption that the data come from a normal distribution must be demonstrated to be appropriate or conservative. To determine if a parametric assessment of the untrended bias and bias uncertainty can be conducted, the assumption of normality is assessed for the Example 2 data set. The data were standardized using Eq. (3), and the standardized data were used to create the histogram and Q-Q plot in Figure A-3 , and they were also used to conduct a number of omnibus normality tests, the results of which are shown in Table A-10. In the plot of the Z-scores in the left half of Figure A-3, the data appear very center peaked, with a few points that are positively skewed compared to the imposed normal curve. It is also noted that there is no data with a Z-score of less than -1.5 standard deviations. Similarly, The Q-Q plot shows that the most negative points are not as negative as would be produced by a normal distribution, and the positive points are more positive than would be produced by a normal distribution. The p-values of less than 0.05 indicates that the null hypothesis that the data are a sample drawn from a normal distribution is rejected in favor of the alternative hypothesis that the data have some other underlying distribution.



Figure A-3	Histogram (Left) and Norma	al Q-Q Plot (Right) of the Example 2 Data Se

Test	p-value
Chi-square (10 bins)	2.117e-08
Anderson-Darling	1.052e-12
Cramer-Von Mises	2.256e-09
Lilliefors	5.607e-09
Shapiro-Wilk	3.075e-09
Shapiro-Francia	2.706e-08
Jarque-Bera	2.200e-16
Kolmogorov-Smirnov	6.222e-04

 Table A-10
 Summary of Example 2 Normality Test P-Values

Having concluded that the data are not drawn from a normal distribution, an assessment of the conservatism of applying a normal distribution-based calculation of the bias uncertainty is performed. To perform this evaluation, the skewness and kurtosis of the data are calculated, and the D'Agostino skewness test and the Anscombe-Glynn kurtosis single sided tests are performed. The skewness results in Table A-11 show that the calculated skewness is 2.720, which is greater than zero corresponding to a normal distribution. The kurtosis of the data is similarly assessed (Table A-11). The kurtosis of the data is calculated to be 11.597, which is greater than a normal value of 3, indicating that the distribution is leptokurtic (center peaked). Because skewness is greater than zero and the kurtosis is greater than three, it is not necessary to perform the single-sided tests, which are useful for showing that the data do not statistically significantly have more data in the lower tail than a normal distribution. Regardless, the tests were performed here and calculate p-values of near 1.0, which are clearly higher than

the 0.05 value necessary to conclude that the data are taken from a distribution with skewness less than 0 or a kurtosis less than 3. The result of this assessment is that the parametric LTL may be applied conservatively to this data set even though it appears to come from a distribution that is not normal.

Parameter	Value
Skewness	2.720
Skewness test p-value	~1.0
Kurtosis	11.597
Kurtosis test p-value	~1.0

Table A-11 Summary of Example 2 Skewness and Kurtosis Assessment

The parametric LTL bias is calculated by determining the mean C/E using Eq. (7) for an unweighted analysis and Eq. (8) for a weighted analysis. The bias is then calculated using Eq. (16) and the values of the mean C/E value. If this value is greater than zero, it is typically set to zero for conservatism. The uncertainty in \overline{k}_{norm} is then calculated using Eq. (9) for an unweighted analysis or Eq. (10) for a weighted analysis. The average total uncertainty is calculated using Eq. (11) for an unweighted analysis or Eq. (12) for a weighted analysis. The average total uncertainty and the uncertainty in \overline{k}_{norm} are then combined to form pooled uncertainty using Eq. (13). The one-sided tolerance factor is then determined either by using Eq. (14) and Eq. (15). Alternatively, the one-sided tolerance factor can be taken from Table 6-1 for the next lowest number of experiments (50) if the regulatory statistical standard is 95% confidence that 95% of the population is covered. The calculated values are carried forward here. The bias uncertainty can then be obtained by multiplying the pooled uncertainty by the one-sided tolerance factor according to Eq. (17). The bias, bias uncertainty, and all of the statistical parameters needed to perform the parametric LTL calculation are provided in Table A-12.

Table A-12 Parametric Validation Parameters for the Example 2 Data Set

Parameter	Weighted value	Unweighted value
\overline{k}_{norm}	0.99666	0.99779
β	-0.00334	-0.00221
$\sigma_{\overline{k}_{norm}}$	0.00374	0.00593
ō	0.00411	0.00519
S _p	0.00555	0.00788
Κ	2.055	2.055
σ_{eta}	0.01141	0.01619

For comparison, the nonparametric LTL is also calculated. The assessment of the bias is the same for the nontrending analysis, regardless of whether a parametric or nonparametric

analysis is used, and it is simply the difference between the mean C/E value (\overline{k}_{norm}) and unity. The bias uncertainty is then evaluated using the historical nonparametric method described in Section 6.2.1. Using the data in Table 6-2, it is determined that even the lowest rank point is not sufficient to attain 95% confidence that 95% of the population of C/E lies above it. However, the confidence calculated with Eq. (18) is 93% for a 95% population coverage. Table 6-3 indicates that no additional NPM is necessary for validation suites with calculated confidences greater than 90%. Therefore, the nonparametric LTL is produced by subtracting the uncertainty in C/E from the lowest C/E value in the suite. In this case, the lowest C/E is 0.98969 with an uncertainty of 0.00357, which results in a nonparametric LTL of 0.98612. For cases in which the bias uncertainty can be convoluted with other uncertainties in the safety analysis case, the LTL can be split into the bias by subtracting the mean C/E value from the LTL and calculating the bias as one minus the mean C/E value. The \overline{k}_{norm} , the bias, and the bias uncertainty are calculated for the Example 2 data set, and the results are shown in Table A-13: weighted and unweighted calculation of the mean C/E value.

Parameter	Weighted value	Unweighted value
\overline{k}_{norm}	0.99666	0.99779
β	-0.00334	-0.00221
σ_{eta}	0.01054	0.01167

Table A-13 Nonparametric Validation Parameters for the Example 2 Data Set

A graphical summary of the validation parameters is shown in Figure A-4 for the Example 2 data set. It is noted that the calculations shown in Figure A-4 do not have any administrative margin and are meant to provide a comparison of the methods. Unweighted calculations are shown as dashed lines in the same color as the corresponding weighted calculations.




A.3 Example 3

Experiment	C/E	$\sigma_{c/E}$	EALF
PU-SOL-THERM-001-001	1.00459	0.00502	8.72E-02
PU-SOL-THERM-001-002	1.00687	0.00504	1.10E-01
PU-SOL-THERM-001-003	1.00913	0.00505	1.33E-01
PU-SOL-THERM-001-004	1.00366	0.00502	1.49E-01
PU-SOL-THERM-001-005	1.00770	0.00504	1.57E-01
PU-SOL-THERM-001-006	1.00897	0.00505	3.41E-01
PU-SOL-THERM-002-001	1.00366	0.00472	7.10E-02
PU-SOL-THERM-002-002	1.00437	0.00472	7.26E-02
PU-SOL-THERM-002-003	1.00335	0.00472	7.75E-02
PU-SOL-THERM-002-004	1.00634	0.00473	8.09E-02
PU-SOL-THERM-002-005	1.00895	0.00474	8.45E-02
PU-SOL-THERM-002-006	1.00496	0.00472	9.22E-02
PU-SOL-THERM-002-007	1.00744	0.00474	9.96E-02
PU-SOL-THERM-003-001	1.00226	0.00471	5.84E-02
PU-SOL-THERM-003-002	1.00188	0.00471	5.95E-02
PU-SOL-THERM-003-003	1.00438	0.00472	6.19E-02
PU-SOL-THERM-003-004	1.00402	0.00472	6.27E-02
PU-SOL-THERM-003-005	1.00516	0.00473	6.54E-02
PU-SOL-THERM-003-006	1.00556	0.00473	6.92E-02
PU-SOL-THERM-003-007	1.00637	0.00473	5.92E-02
PU-SOL-THERM-003-008	1.00511	0.00473	6.01E-02
PU-SOL-THERM-004-001	1.00351	0.00472	5.35E-02
PU-SOL-THERM-004-002	0.99840	0.00469	5.38E-02
PU-SOL-THERM-004-003	1.00045	0.00470	5.48E-02
PU-SOL-THERM-004-004	0.99843	0.00469	5.61E-02
PU-SOL-THERM-004-005	0.99932	0.00470	5.47E-02
PU-SOL-THERM-004-006	1.00134	0.00471	5.50E-02
PU-SOL-THERM-004-007	1.00521	0.00473	5.60E-02
PU-SOL-THERM-004-008	1.00084	0.00470	5.66E-02
PU-SOL-THERM-004-009	1.00029	0.00470	5.87E-02
PU-SOL-THERM-004-010	1.00185	0.00471	6.32E-02
PU-SOL-THERM-004-011	1.00020	0.00470	6.83E-02
PU-SOL-THERM-004-012	1.00253	0.00471	5.59E-02
PU-SOL-THERM-004-013	0.99972	0.00470	5.57E-02

Table A-14 Plutonium Solution Critical Experiment Data Used in Example 3

Experiment	C/E	σ _{C/E}	EALF
PU-SOL-THERM-005-001	1.00165	0.00471	5.57E-02
PU-SOL-THERM-005-002	1.00243	0.00471	5.67E-02
PU-SOL-THERM-005-003	1.00306	0.00472	5.77E-02
PU-SOL-THERM-005-004	1.00474	0.00472	6.01E-02
PU-SOL-THERM-005-005	1.00586	0.00473	6.29E-02
PU-SOL-THERM-005-006	1.00545	0.00473	6.60E-02
PU-SOL-THERM-005-007	1.00388	0.00472	6.91E-02
PU-SOL-THERM-005-008	0.99898	0.00470	5.68E-02
PU-SOL-THERM-005-009	1.00169	0.00471	5.80E-02
PU-SOL-THERM-006-001	1.00038	0.00350	5.26E-02
PU-SOL-THERM-006-002	1.00152	0.00351	5.35E-02
PU-SOL-THERM-006-003	1.00118	0.00351	5.54E-02
PU-SOL-THERM-007-001	1.00944	0.00475	2.70E-01
PU-SOL-THERM-007-002	1.00376	0.00472	2.57E-01
PU-SOL-THERM-007-003	1.00901	0.00474	1.11E-01
PU-SOL-THERM-007-004	1.00301	0.00472	1.13E-01
PU-SOL-THERM-007-005	1.00513	0.00473	1.11E-01
PU-SOL-THERM-007-006	0.99891	0.00470	1.14E-01
PU-SOL-THERM-007-007	0.99702	0.00469	1.13E-01
PU-SOL-THERM-007-008	1.00074	0.00470	1.05E-01
PU-SOL-THERM-011-001	1.00945	0.00525	6.24E-02
PU-SOL-THERM-011-002	1.01406	0.00527	6.36E-02
PU-SOL-THERM-011-003	1.01629	0.00529	6.60E-02
PU-SOL-THERM-011-004	1.00865	0.00525	6.66E-02
PU-SOL-THERM-011-005	1.00583	0.00523	7.40E-02
PU-SOL-THERM-011-006	0.99378	0.00517	5.12E-02
PU-SOL-THERM-011-007	0.99981	0.00520	5.22E-02
PU-SOL-THERM-011-008	0.99637	0.00518	5.21E-02
PU-SOL-THERM-011-009	0.99297	0.00516	5.34E-02
PU-SOL-THERM-011-010	1.00309	0.00522	5.47E-02
PU-SOL-THERM-011-011	0.99972	0.00520	5.84E-02
PU-SOL-THERM-011-012	0.99918	0.00520	5.32E-02
PU-SOL-THERM-020-001	1.00316	0.00592	6.55E-02
PU-SOL-THERM-020-002	1.00570	0.00593	6.46E-02
PU-SOL-THERM-020-003	1.00032	0.00590	5.90E-02

Table A-14Plutonium Solution Critical Experiment Data Used in Example 3
(continued)

Experiment	C/E	σ _{c/E}	EALF
PU-SOL-THERM-020-004	1.00381	0.00592	7.59E-02
PU-SOL-THERM-020-005	1.00419	0.00593	7.88E-02
PU-SOL-THERM-020-006	0.99815	0.00589	6.04E-02
PU-SOL-THERM-020-007	1.00314	0.00592	1.05E-01
PU-SOL-THERM-020-008	0.99443	0.00587	7.57E-02
PU-SOL-THERM-020-009	1.00401	0.00592	6.48E-02
PU-SOL-THERM-020-010	1.00092	0.00591	5.87E-02
PU-SOL-THERM-020-011	1.00221	0.00591	7.57E-02
PU-SOL-THERM-020-012	1.00341	0.00592	7.88E-02
PU-SOL-THERM-020-013	0.99258	0.00586	7.57E-02
PU-SOL-THERM-020-014	0.99609	0.00588	1.06E-01
PU-SOL-THERM-020-015	1.00349	0.00592	5.85E-02

Table A-14Plutonium Solution Critical Experiment Data Used in Example 3
(continued)

A.3.1 Assessment of Trends

To determine if the bias in the computational method is best expressed as function of EALF for the Pu solution systems contained in the Example 3 data set provided in Table A-14 weighted and unweighted trendlines are constructed, and their statistical significance is tested using the methods discussed Sections 6.3 (trend calculation) and 5.3 (t-test). The best fit on EALF was produced by first taking the natural logarithm of the EALF and then fitting. The data set contains 81 experiments, with an unweighted average C/E = 1.00295642 and an unweighted average ln(EALF) = -2.61567. Using the uncertainty of each experiment to weight the data, the weighted average C/E = 1.00296, and the weighted ln(EALF) = -2.62585.

First the S_{xx} and S_{kx} values are calculated using Eq. (21) and Eq. (23) for an unweighted analysis or Eq. (22) and Eq. (24) for a weighted analysis. The slope and intercept of the trendlines are then calculated using Eq. (25) and Eq. (26), respectively. The residuals about the trendline are then calculated using Eq.(27), and the uncertainty in the fit is calculated using Eq. (28) for a unweighted analysis and Eq. (29) for a weighted analysis. The t-statistics for the fits are then calculated using Eq. (4).

Once the t-statistic for the fit is calculated, there are two equivalent methods by which the statistical significance of the fit can be judged. The first method is by comparison of the t-statistic for the fit to the critical value from the t-distribution corresponding to n-2 degrees of freedom and 95% confidence. For this example, the critical value is compared to the t-statistic corresponding to $\alpha/2=0.025$ (a two-sided test is used because the slope can depart from zero in either the positive or negative direction, so the value of α is divided by 2) and 79 degrees of freedom (n-2). The t-statistics for the weighted fit was 3.40 and 3.12 for the unweighted fit. These values are larger than the appropriate critical values from the t-distribution of 1.99; the null hypothesis that there is no slope is rejected in favor of the calculated trendline. The second

method of assessing the statistical significance of a trendline is to calculate the p-value of the tstatistic with statistical software and then compare it to the prescribed value of α . In this case, the p-value was 0.0017 for the weighted calculation and 0.0038 for the unweighted calculation. Because these values are less than 0.025, the null hypothesis of zero slope is again rejected in favor of the calculated trendline. The parameters calculated in the determination of the trendline and the assessment of the statistical significance of the trendline are presented in Table A-15.

Parameter	Weighted value	Unweighted value
S _{xx}	11.83255	11.58647
S _{kx}	0.044352	0.042674
β ₀	1.0127992	1.0125901
β1	0.0037483	0.0036831
σ_{fit}	0.0037951	0.0040172
t _{fit}	3.3974	3.1208
t _{critical}	1.9905	1.9905
p-value	0.0017	0.0038

Table A-15 Thenu Statistical Significance Parameters for the Example 5 Do	able A-15	nce Parameters for the Example 3 Data
---	-----------	---------------------------------------

Once the calculated trend is determined to be statistically significant, it can be used in conjunction with the values of the trend parameter from the application cases to calculate the bias and bias uncertainty. For this example, the EALFs of two the hypothetical application cases considered are 0.05 and 1 eV, which have natural logarithms of -2.99573 and 0.0, respectively.

To evaluate the bias and bias uncertainty as a function of the natural logarithm of EALF, a number of additional statistical parameters need to be evaluated. Two bands are used for the bias uncertainty evaluation in this appendix: the Single-Sided Lower Tolerance Band method from NUREG/CR-6698, and the Single-Sided Lower Confidence Band method from NUREG/CR-6361, which is often referred to as the *USL-1 method*. For both methods, the average total uncertainty ($\bar{\sigma}$) is needed and is calculated using Eq. (11) for unweighted analyses and Eq. (12) for weighted analyses. The pooled standard deviation (S_p) for a trended bias assessment can then be calculated with Eq.(31). Additionally, the statistical parameters for the NUREG/CR-6698 lower tolerance band from Table 6-4 and the inverse t-distribution calculation are needed to perform the NUREG/CR-6361 lower confidence band calculation are required. The values of these parameters are presented in Table A-16.

Parameter	Weighted value	Unweighted value
$\bar{\sigma}$	0.004899	0.005019
S_p	0.006197	0.0.006428
$F_{\alpha}^{(2,n-2)}$	3.1122	3.1122
n	81	81
Z _P	1.64485	1.64485
$\chi^2_{1-\frac{\alpha}{2},n-2}$	56.3089	56.3089
$t_{1-\alpha,n-2}$	1.66437	1.66437

Table A-16Parameters Necessary for Trended Bias Uncertainty Evaluation for the
Example 3 Data

Using the parameters calculated above the trended value of C/E ($k_{fit}(x)$), the bias and bias uncertainty are estimated for each of the application cases and are reported in Table A-17. The values of $k_{fit}(x)$ and the trended bias in Table A-17 show that the bias is set to zero in both cases because otherwise the trendline would result in a positive bias. The 1 eV case requires extrapolation of the trend beyond the available data. The bias uncertainties are estimated by the NUREG/CR-6698 and USL-1 bands. It is noted that for the USL-1 calculation the bias uncertainty is calculated as the maximum of the values found by evaluating at the end points of the energy range. The larger of the two end point uncertainty values is then applied to the entire energy range. For the 1 eV case the bias uncertainty is evaluated at that point.

Table A-17Sample Calculation of Bias, Bias Uncertainties, and Calculational Margins
for the Example 3 Cases

Deterretor	0.05 eV		1 eV	
Parameter	Weighted	Unweighted	Weighted	Unweighted
$k_{fit}(x)$	1.00157	1.00156	1.01280	1.01259
Trended bias	0	0	0	0
NUREG-6698 bias uncertainty	0.01447	0.01505	0.02400	0.02498
NUREG-6698 calculational margin	0.01447	0.01505	0.02400	0.02498
USL-1 bias uncertainty	0.01137	0.01180	0.01302	0.01354
USL-1 calculational margin	0.01137	0.01180	0.01302	0.01354

A.3.2 Nontrending Assessment

Even though it has been determined that a statistically significant trend is associated with the Example 3 data set, the process necessary to determine the bias and bias uncertainty without the trend is shown here. When assessing the nontrending bias, there are two methods for calculating the bias uncertainty: the normality-based LTL and the nonparametric LTL.

To apply the normality-based LTL, the assumption that the data come from a normal distribution must be shown to be appropriate or conservative. To determine if a parametric assessment of the untrended bias and bias uncertainty is needed, the assumption of normality is assessed for the Example 3 data set. The data were standardized using Eq. (3), and the standardized data were used to create the histogram and Q-Q plot in Figure A-5. The data were also used to conduct a number of omnibus normality tests, the results of which are shown in Table A-18. The Z-scores in the left half of Figure A-5 appear slightly center peaked, with a few points that are positively skewed compared to the imposed normal curve. Similarly, The Q-Q plot looks reasonably like the Q-Q plot generated from a normal distribution in Figure 5-2, so there are no remarkable non-normal features identified with graphical methods. All of the omnibus normality tests used here calculate p-values that are greater than 0.05, such that the null hypothesis that the data are a sample drawn from a normal distribution is not rejected in favor of the alternative hypothesis that the data have some other underlying distribution. Therefore, it is concluded that it is acceptable to use a normality-based LTL approach.





 Table A-18
 Summary of Example 3 Normality Test P-Values

Test	p-value
Chi-square (12 bins)	0.1875
Anderson-Darling	0.1857
Cramer-Von Mises	0.2327
Lilliefors	0.3468
Shapiro-Wilk	0.1791
Shapiro-Francia	0.0933
Jarque-Bera	0.1755
Kolmogorov-Smirnov	0.775

The parametric LTL bias is calculated by determining the mean C/E using Eq. (7) for an unweighted analysis and Eq. (8) for a weighted analysis. The bias is then calculated using Eq. (16) and the values of the mean C/E values. If this value is greater than zero, it is typically set to zero for conservatism. The uncertainty in \overline{k}_{norm} is then calculated using Eq. (9) for an unweighted analysis or Eq. (10) for a weighted analysis. The average total uncertainty is calculated using Eq. (11) for an unweighted analysis or Eq. (12) for a weighted analysis. The average total uncertainty and the uncertainty in \overline{k}_{norm} are then combined to form pooled uncertainty (S_n) using Eq. (13) for the non-trending analysis. The one-sided tolerance factor is then determined either by calculation using Eq. (14) and Eq. (15) according to the number of experiments in the validation suite (81). Alternatively, the one-sided tolerance factor can be taken from Table 6-1 for the next lowest number of experiments (80), provided that the regulatory statistical standard is 95% confidence that 95% of the population is covered. The calculated tolerance factor is used here. The bias uncertainty can then be obtained by multiplying the pooled uncertainty by the one-sided tolerance factor according to Eq. (17). The bias, bias uncertainty, and all of the statistical parameters needed to perform the parametric LTL calculation are provided in Table A-19.

Parameter	Weighted value	Unweighted value
\overline{k}_{norm}	1.00296	1.00296
β	0	0
$\sigma_{\overline{k}_{norm}}$	0.00403	0.00423
ō	0.00489	0.0502
S _p	0.00635	0.00656
K	1.962	1.962
σ_{eta}	0.01245	0.01288

Table A-19 Parametric Validation Parameters for the Example 3 Data Set

For the sake of comparison, the nonparametric LTL is also calculated. The assessment of the bias is the same for the nontrending analysis, regardless of the whether a parametric or nonparametric analysis is used, and it is simply the difference between the mean C/E value (\overline{k}_{norm}) and unity. The bias uncertainty is then evaluated using the traditional historical nonparametric method described in Section 6.2.1. Using the data in Table 6-2, it is determined that the lowest rank point is sufficient to attain 95% confidence that 95% of the population of C/E values lie above it. However, there are not enough points to use the second lowest point. Therefore, the nonparametric LTL is produced by subtracting the uncertainty in C/E from the C/E value that is the lowest in the suite. The lowest C/E value is 0.99258 and its uncertainty is 0.00586, so the resulting LTL is 0.98672. The bias uncertainty calculation using Eq. (19) indicates that the bias uncertainty should be calculated by using a \overline{k}_{norm} value of 1 if $\overline{k}_{norm} > 1$, as is the case here. The resulting bias uncertainty (σ_{β}) is 0.01328. The \overline{k}_{norm} , the adjusted bias, and the bias uncertainty are calculated for the Example 3 data set, and the results are shown in Table A-20 as weighted and unweighted calculations of the mean C/E value.

Parameter	Weighted value	Unweighted value
\overline{k}_{norm}	1.00296	1.00296
β	0	0
σ_{eta}	0.01328	0.01328

Table A-20 Nonparametric Validation Parameters for the Example 3 Data Set

A graphical summary of the validation parameters is shown in Table A-6 for the Example 3 data set. The calculations shown in Table A-6 do not have any administrative margin and are meant to provide a comparison of the methods. Unweighted calculations are shown as a dashed line in the same color as the corresponding weighted calculations.





A.4 <u>References</u>

[A-1] E. M. Saylor, W. J. Marshall, J. B. Clarity, Z. J. Clifton, and B. T. Rearden, *Criticality Safety Validation of SCALE 6.2.2*, ORNL/TM-2018/884, Oak Ridge, TN, 2018.

NRC FORM 335 U.S. NUCLEAR REGULATORY COMMISSION (12-2010) NRCMD 3.7	1. REPORT NUMBER (Assigned by NRC, A and Addendum Num	ک dd Vol., Supp., Rev., bers, if any.)	
BIBLIOGRAPHIC DATA SHEET (See instructions on the reverse)	NUREG/(ORNL/TN	NUREG/CR-7311 ORNL/TM-2024/3	
2. TITLE AND SUBTITLE	3. DATE REPO	ORT PUBLISHED	
Determination of Bias and Bias Uncertainty for Criticality Safety Computational Methods	молтн April	YEAR 2025	
	4. FIN OR GRANT NU	JMBER	
5. AUTHOR(S)	6. TYPE OF REPORT		
J.B. Clarity, W.J. Marshall, D.E. Mueller, S.S. Powers, B.T. Rearden, and S.M. Bowman	Tech	nnical	
	7. PERIOD COVEREI	D (Inclusive Dates)	
8. PERFORMING ORGANIZATION - NAME AND ADDRESS (If NRC, provide Division, Office or Region, U. S. Nuclear Regula contractor, provide name and mailing address.) Oak Ridge National Laboratory Oak Ridge, TN 37831-6283	tory Commission, and r	mailing address; if	
 SPONSORING ORGANIZATION - NAME AND ADDRESS (If NRC, type "Same as above", if contractor, provide NRC Divisio Commission, and mailing address.) Division of Fuel Management Office of Nuclear Material Safety and Safeguards U.S. Nuclear Regulatory Commission Washington, D.C. 20555-0001 	n, Office or Region, U. S	S. Nuclear Regulatory	
10. SUPPLEMENTARY NOTES			
11. ABSTRACT (200 words or less) Nuclear criticality safety evaluations must demonstrate that operations are subcl and credible abnormal conditions, and such evaluations often rely upon compute determine the neutron multiplication factor for complex three-dimensional syster computer codes and data used to model these systems establishes their suitabi applications. The validation activity also determines the computational bias and that is relevant to the application. The bias is developed from calculations of kno experiments that are similar to the intended application of interest. This report de can be used by criticality safety analysts to perform the validation activity, includ calculational bias, bias uncertainty, and the application of those values to develo applied in safety analyses. This report builds upon earlier works in the criticality incorporates modern analytical techniques developed over the last twenty years learned from observations of previous validation efforts.	ritical under bot ational techniqu ns. Validation of lity for specific the uncertainty wm laboratory of escribes technic ing determination p limits that ca safety validatio , as well as less	th normal lies to of the in that bias critical ques that on of n be n area and sons	
12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.)	13. AVAILAB	ILITY STATEMENT	
nuclear criticality safety	14. SECURIT	Y CLASSIFICATION	
indical onliquity ourory	(This Page)	nclassified	
	(This Report	nclassified	
15. NUMBER OF PAGES			
	16. PRICE		



Federal Recycling Program



UNITED STATES NUCLEAR REGULATORY COMMISSION WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS



NUREG/CR-7311 Determination of Bias and Bias Uncertainty for Criticality Safety Computational Methods April 2025