SAND2022-1784C

A Comparison of Sensitivity/Uncertainty-Based Upper Subcritical Limit Estimates

Christopher M. Perfetti, Robert D. Busch, Forrest B. Brown
University of New Mexico
1 University of New Mexico; Albuquerque, NM 87131
cperfetti@unm.edu; busch@unm.edu; fbrown@unm.edu

Shawn J. Henderson and John A. Miller
Sandia National Laboratories
P.O. Box 5800, Albuquerque, NM 87185
sjhende@sandia.gov; millerj@sandia.gov

[Digital Object Identifier (DOI) placeholder – to be added by ANS during production]

ABSTRACT

Nuclear Criticality Safety evaluations often rely on the availability of relevant benchmark experiments to determine justifiable margins of subcriticality. Recently developed sensitivity and uncertainty analysis-based upper subcritical limit (USL) methods allow NCS engineers to automatically identify benchmark experiments that share similar sources of uncertainty with a target validation application. This aspect supports and augments what was historically an element of expert judgment based on experience, while assisting less experienced staff in this determination. However, few method-to-method or code-to-code comparisons exist for these sensitivity-based methods.

This study provides a code-to-code comparison of USL estimates using the SCALE USLSTATS trending analysis method, the SCALE TSURFER data assimilation method, and the MCNP/Whisper non-parametric method. These three methods are applied to the example application from Appendix D of ANSI/ANS–8.24, which models buckets of UO_2F_2 waste material in both nominal dry (H/X = 80) conditions, and wet (H/X=600) upset conditions. This study reviews and contrasts the three sensitivity methods, compares their identified pertinent benchmarks for this example application, and compares computational bias and USL estimates for this example under both normal and upset conditions.

Key Words: Upper subcritical limits, computational bias, USLSTATS, TSURFER, Whisper

1 INTRODUCTION

Determining that a system will be subcritical under a number of process conditions is an integral part of a Nuclear Criticality Safety (NCS) evaluation. Hand calculation methods or handbook information may be appropriate for simple problems, but more sophisticated calculation methods are required when the system's geometry is complex or when intricate surfaces or arrays are present. Recent developments in high fidelity sensitivity analysis methods have opened the door for increasingly quantitative statistical methods for estimating computational bias and the upper subcritical limit (USL) [1][2][3]. Rather than basing USL estimates on quasi-quantitative parameters, such systems' EALF or H-X ratio, these methods generally estimate USLs using sensitivity analysis methods that identify common sources of computational bias.

However, despite the additional fidelity offered by these methods developments, few rigorous code-to-code comparisons of these methods exist. While there are some notable code-to-code comparisons [4][5], this study attempts to compare the USL estimates offered by the Whisper, USLSTATS, and TSURFER techniques [2][6][7]. This study applies all three methods to an example application and compares the resulting computational bias and USL estimates from the three methods.

2 SAMPLE PROBLEM DESCRIPTION

This study applied the three computational bias and USL estimation methods to the example application from Appendix D of ANSI/ANS–8.24, which models an array of 5-gallon buckets of 5 wt% UO_2F_2 waste material, as described in Table I. The material is normally dry (minimum *H*/*X* ratio of ~80), and upset conditions include moderation scenarios up to an *H*/*X* ratio of ~600.

Table I. Sample Problem Description and Process Parameters

Parameter	Process Data
Fissile Material	^{235}U in UO_2F_2
Fissile Form	UO_2F_2 compound form (normal condition) UO_2F_2 solution (upset condition)
Moderation (H/X)	= 80 (normal condition) = 600 (abnormal condition = estimated optimum moderation)
Enrichment (wt% ^{235}U)	1 wt% to 5 wt% (analyzed at 5 wt%)
Uranium Concentration	Minimal to 3000 g/L
Moderating Material	Water
Other Materials	Predominant Elements are Na and F
Reflecting Materials	Unreflected (light steel) (normal condition) Water reflected (abnormal condition)
Geometry	Cylinders and Arrays
Heterogeneity / Homogeneity	Homogeneous System
Neutron Energy	Unknown, estimated to be thermal up to intermediate energy spectrum

3 CALCULATIONAL METHODS

This study compares USL estimates for the sample problem under both the normal and upset conditions using the Whisper, USLSTATS, and TSURFER techniques. All calculations were performed using continuous-energy ENDF/B-VII.1 cross sections and fission source convergence was confirmed in each Monte Carlo simulation. These methods use sensitivity coefficients to either estimate the degree of similarity between benchmark experiments and a target application, and/or to estimate the computational bias present for that application. The sensitivity coefficient for a response, here the critical eigenvalue k, to an input parameter, Σ_x, is defined as

$$S_{k,\Sigma_x} = \frac{\partial k/k}{\partial \Sigma_x/\Sigma_x}. \quad (1)$$

These sensitivity coefficients describe how changes to or uncertainties in the input parameters affect the eigenvalue of a system. Sensitivities can account for uncertainty in the mass, density, or geometry of materials in an integral experiment, and in the NCS field these uncertain parameters are typically nuclear data parameters (i.e., nuclear cross sections, the energy distribution of fission neutrons, etc.).

The USLSTATS and Whisper methods use sensitivity coefficients for a given application to estimate the similarity between that application and various benchmark experiments. To achieve this, first the *sandwich equation* must estimate the amount of nuclear data–induced uncertainty that is shared between the benchmark experiment and the application eigenvalues (k_1 and k_2, respectively):

$$\sigma^2_{k_1,k_2} = S_{k_1\Sigma_x} \cdot Cov_{\Sigma_x,\Sigma_y} \cdot S^T_{k_2\Sigma_y}\,, \tag{2}$$

where Cov_{Σ_x,Σ_y} is the matrix or nuclear cross section covariance (i.e., uncertainty) data. Having obtained the uncertainty shared between the benchmark and application systems, the *similarity coefficient* c_k can be calculated for the two systems [8]:

$$c_k = \frac{\sigma^2_{k_1,k_2}}{\sigma_{k_1}\sigma_{k_2}}\,. \tag{3}$$

where σ_{k_1} and σ_{k_2} represent the data-induced uncertainty in the eigenvalues of systems 1 and 2, respectively. This c_k coefficient is analogous to a Pearson coefficient of correlation and describes the fraction of nuclear data–induced uncertainty that is shared between the benchmark experiment and the application. Because NCS analyses typically use high-fidelity Monte Carlo neutron transport simulations to estimate the critical eigenvalue of a system, it is reasonable to assume that any bias or disagreement encountered between the computed and experimental eigenvalues from high quality benchmark evaluations is typically attributed to errors or deficiencies in the underlying nuclear data. Therefore, a benchmark experiment that shares the same sources of nuclear data–induced uncertainty with an application is likely to experience a computational bias in its eigenvalue that is similar to the bias in the application.

Benchmark experiment sensitivity data files (SDFs) for the USLSTATS and TSURFER calculations for the benchmark experiments were obtained using the VALID library of 315 reference SDFs distributed with the Database for the International Handbook of Evaluated Criticality Safety Benchmark Experiments (DICE), and benchmark experiment eigenvalues and measurement uncertainties were extracted from DICE [9][10]. All Whisper calculations used SDFs for 1,100 ICSBEP benchmark experiments [10] that were generated using MCNP6.2 and distributed with Whisper in the MCNP 6.2 code release distribution. Automated benchmark rejection are performed for benchmarks that are distributed with Whisper. An administrative margin is not assumed for USL estimates calculated in this study.

3.1 USLSTATS Trending Analysis Method

USLSTATS is a generic parameter trending analysis code within the SCALE Code System that uses the results of benchmark experiments to estimate the bias present in modeling and simulation tools for application cases [6]. To perform USLSTATS simulations, a user enters values for a trending parameter for each benchmark experiment and USLSTATS then develops a linear regression fit and confidence intervals for ratios of the calculated-to-experimental (C/E) benchmark experiment eigenvalues values. USLSTATS either interpolates or extrapolates this trend to estimate the bias and USL for application cases.

This study used c_k similarity coefficients as the trending parameter and computed these similarity coefficients using SCALE 56-group covariance data. When trending on c_k values, USLSTATS estimates quantities by extrapolating the C/E regression fit to a value of $c_k = 1$; this extrapolation is chosen to estimate the computation bias because, assuming that errors in the nuclear data predominantly drive computational bias, any benchmark that shares all sources of neutronic uncertainty with an application would be expected to experience the same computational bias. Examining c_k considers to which nuclear cross section data an application case is most sensitive, which cross section data contain the highest degree of uncertainty, and how much of the nuclear data-induced uncertainty is shared between the target application and the library of benchmark experiments. However, these trending methods generally rely on the availability of a significant number of sufficiently similar benchmark experiments to accurately predict the computational bias for an application.

Studies by Broadhead and others suggest that an application should have at least 20 experiments with c_k values greater than 0.80 [8]; guidance from the US Nuclear Regulatory Commission (NRC) recommends that NCS analysts have access to multiple experiments with c_k values in excess of 0.90 or 0.95 [11]. If a sufficient number of similar experiments does not exist, the licensee must justify that the most important

sources of uncertainty have been accounted for in the application and/or provide additional margin of subcriticality.

As shown in Fig. 1, USLSTATS was able to develop a regression trend for the sample application using a sufficient number of similar benchmark experiments ($c_k > 0.90$) under both the normal and upset conditions, although no highly similar benchmarks ($c_k > 0.95$) were identified under normal conditions.

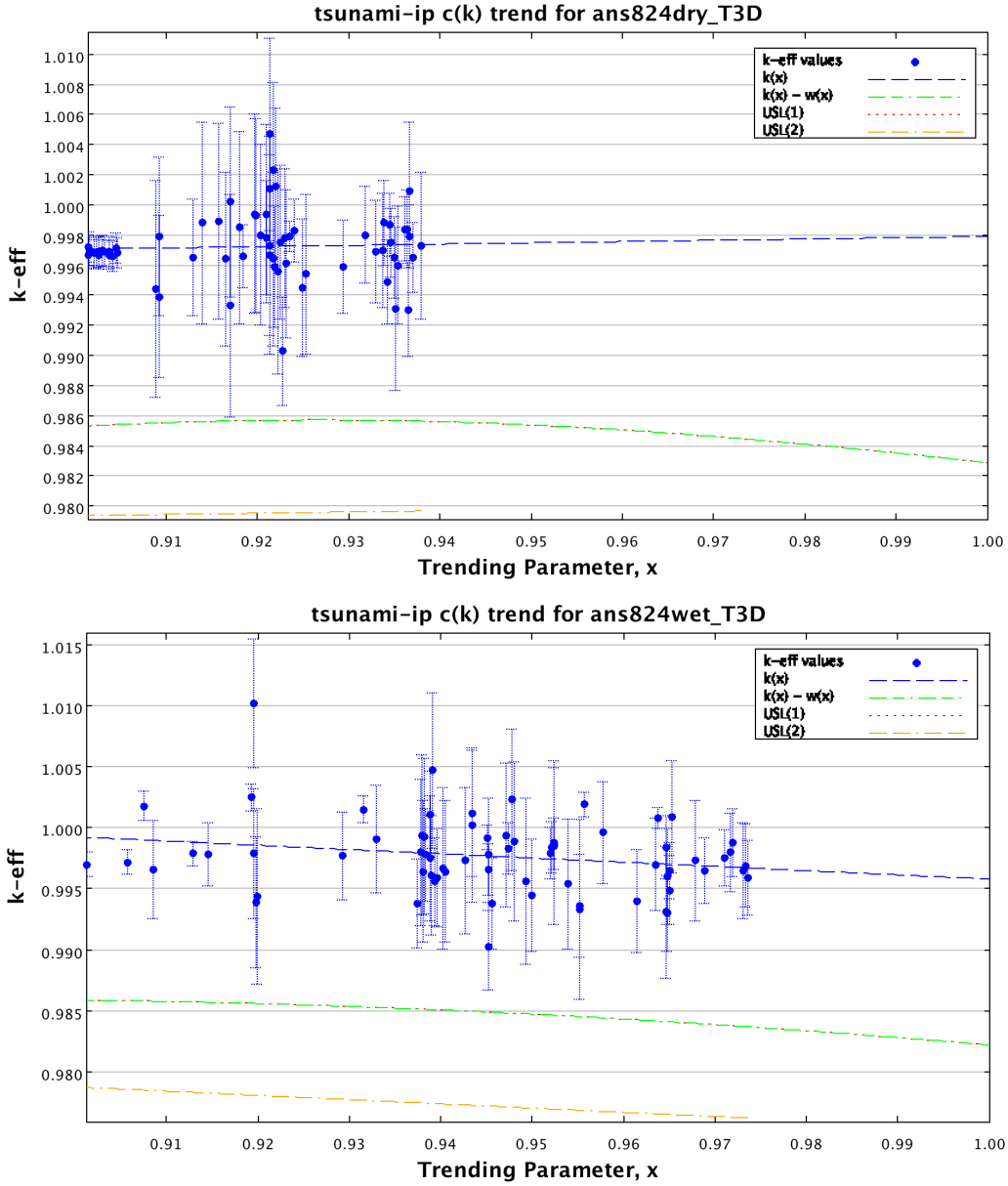

Figure 1. USLSTATS regression trends for the ANSI/ANS–8.24 Sample Problem under normal (top) and upset (bottom) conditions.

3.2 TSURFER Data Assimilation / Data Adjustment Method

The TSURFER tool relies on the Bayesian-based Generalized Linear Least-Squares (GLLS) methodology to estimate USLs [7]. This methodology compares benchmark experiment eigenvalues with those predicted by a modeling and simulation code (here, KENO), and adjusts nuclear data in a way that

maximizes agreement between the code and the experiments. This adjustment is performed such that no data is adjusted by an unreasonable number of its standard deviations and so that benchmark experiment results that are identified as outliers are automatically omitted from the adjustment process. Again, it is reasonable to assume that nuclear data is primarily responsible for computational biases in NCS simulations that use high-fidelity methods (i.e. KENO and MCNP) because of the minimal number of approximations inherent to these methods.

TSURFER produces a set of adjustments to the nuclear data (see Fig. 2) that improves the accuracy of the computational simulation results compared to benchmark experiment measurements (see Fig. 3). After obtaining these recommended data adjustments, TSURFER applies the adjustments to the application case to examine how they would change the eigenvalue of the application simulations; this change in the application eigenvalue is analogous to the computational bias for these simulations. Sensitivity coefficient data (i.e. SDFs) for the benchmark experiments are required for TSURFER to solve for an optimal set of nuclear data adjustments, and SDFs for the application cases allow TSURFER to predict how these nuclear data adjustments will change the application eigenvalues.

The TSURFER data adjustments and bias estimates for the sample application were performed using the same SDF and benchmark experiment data that was used for the previously mentioned USLSTATS simulation (i.e., 315 of the ORNL-validated SDFs that are distributed with DICE) to estimate the computational bias present in the application simulations. The resulting TSURFER data adjustments and pre- and post-adjustment C/E values are given in Figs. 2 and 3. While the TSURFER method has great potential for estimating the computational bias of simulations, it has not yet been adapted to provide USL estimates with a known statistical confidence interval (e.g., a 99/99 confidence interval). Studies by Perfetti have proposed methods for generating confidence intervals using the TSURFER data assimilation method, but this study will limit its scope to TSURFER computational bias estimates [5].

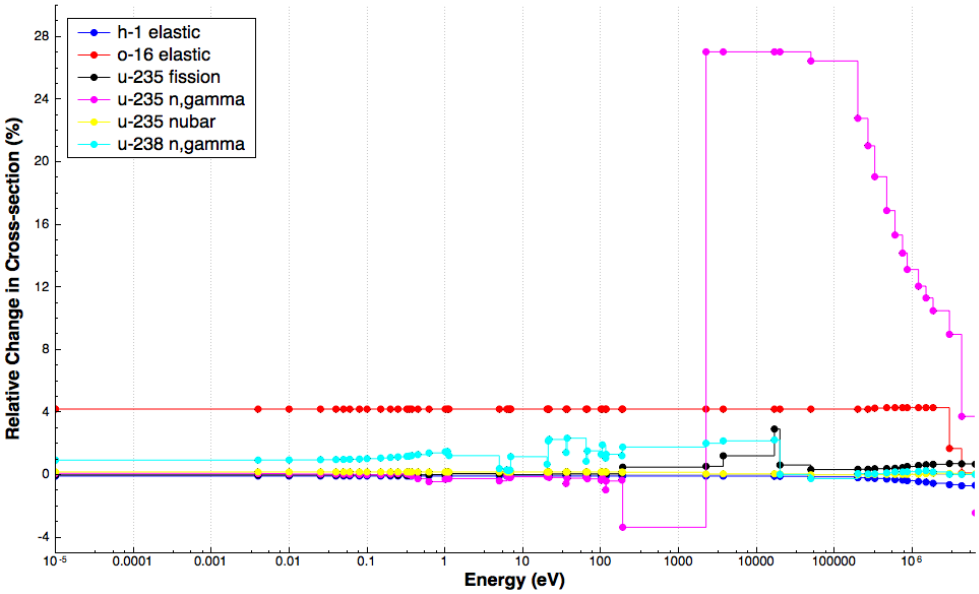

Figure 2. TSURFER-recommended adjustments to nuclear cross section data. ^{235}U (n,γ) reactions experienced substantial adjustments (up to 26%) between 1 keV and 100 keV because these cross sections contain a large degree of uncertainty (up to 34%) in this energy range.

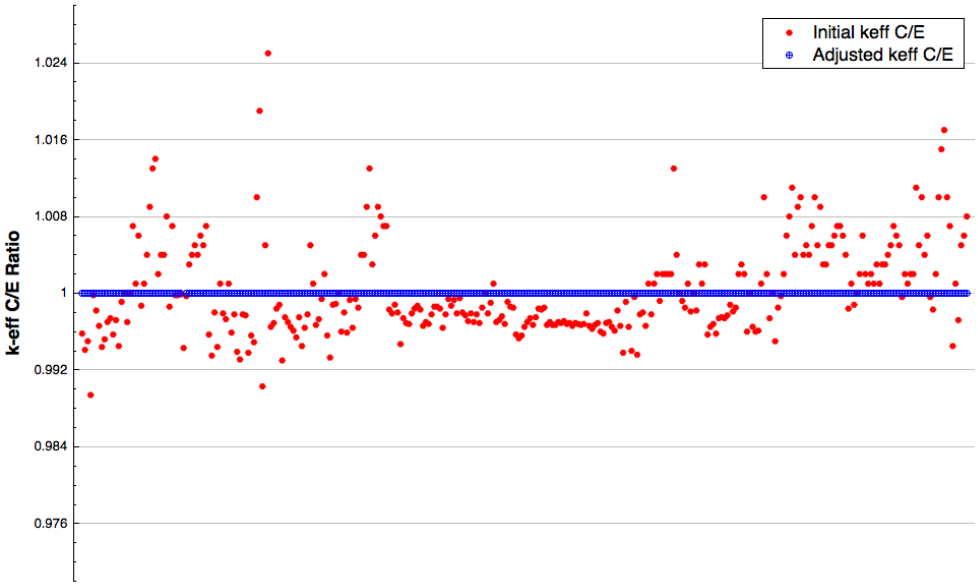

Figure 3. TSURFER-recommended nuclear data adjustments significantly improved the accuracy of benchmark experiment eigenvalue computational estimates.

3.3 Whisper Extreme Value Theory Method

The Whisper method uses extreme value theory to estimate a conservative margin of subcriticality for an application based on the observed computational biases in relevant benchmark experiments. The Whisper method assigns each benchmark experiment a weight for computing the Margin of Subcriticality (MOS) that is related to the experiment's degree of similarity to a target application; these weights are computed as

$$w_{exp.} = max\left\{0, \frac{(c_{k,exp.} - c_{k,acc.})}{(c_{k,max.} - c_{k,acc.})}\right\}, \quad (4)$$

where $c_{k,exp.}$ is the c_k value for a benchmark experiment, $c_{k,max.}$ is the maximum c_k from all available experiments, and $c_{k,acc.}$ is the smallest c_k that has been deemed acceptable for including a benchmark in the validation analysis. Benchmarks can have a maximum weight of one. As described in detail by Kiedrowski, these weights can then be used to estimate the computational bias by estimating the mean of a weighted, extreme value probability density function (PDF) [3]. Making use of this extreme value PDF provides an intentional degree of conservatism when estimating the computational bias because this bias estimate is dictated predominantly by the benchmark experiments with the most conservative bias estimates. Thus, the Whisper method is not intended to estimate the true bias in the computational methods in the same way as the USLSTATS and TSURFER methods, but rather the maximum likely bias that an application may experience.

The Whisper benchmark library distributed with MCNP 6.2 contains over 1,100 MCNP input files spanning a large set of uranium and plutonium systems containing a variety of fissionable isotopes, forms (metal, oxide, solution), geometries, spectral characteristics, etc. Whisper provides two sets of 44-group covariance data: the first set is the same data that is distributed with SCALE 6.1 in a format that Whisper can parse. The second set is an adjusted nuclear data library based upon a GLLS fitting of the benchmarks following rejection (using the same methodology as TSURFER). Whisper uses the latter to quantify the effect of nuclear data uncertainties within the MOS.

Whisper method separates the MOS into several components when computing USL estimates:

$$USL = 1 - MOS_{code} - MOS_{data} - m - \Delta k_{admin}\,, \quad (5)$$

where m represents the calculational margin from the extreme value theory calculation. The MOS_{code} term represents additional margin that is added to account for any undetected code errors or bugs in the sensitivity analysis and radiation transport tools. Kiedrowski suggests assuming a value of 0.005 for MOS_{code}, but this study assumed a value of 0.000 to allow for a fairer method-to-method comparison. The MOS_{data} term equals the nuclear data-induced uncertainty in the application following a GLLS data assimilation calculation and is intended to account for any additional uncertainty in the subcritical margin that is introduced due to uncertainty and errors in the nuclear data.

4 RESULTS

Table II shows the eigenvalue estimates produced by MCNP and SCALE/KENO for the sample problem under normal and upset conditions. The two codes produced eigenvalue estimates for both conditions that agreed well: a 107 per cent mille (pcm) difference and a 146 pcm difference for the normal and upset condition cases, respectively.

Table II. Sample Problem Eigenvalue Estimates

Case	SCALE/KENO	MCNP
Dry (Normal)	0.37746 ± 0.00020	0.37639 ± 0.00028
Wet (Upset)	0.87079 ± 0.00027	0.87225 ± 0.00032

Table III gives the computational bias estimates from the USLSTATS and TSURFER methods. The Whisper bias estimates were not included in this table because Whisper does not produce bias estimates that are analogous to those produced by other methods; rather than estimating the most likely computational bias, Whisper estimates the most likely, most conservative bias present for the application.

The TSURFER method does not currently provide traditional USL estimates, but a method for using TSURFER to generate USL estimates has been proposed by Perfetti [5]:

$$TSURFER\ USL = 1 + \min(0, \beta) - \kappa\, \sigma_\beta - \Delta k_{admin}, \quad (6)$$

where κ is a single-sided tolerance factor for the 99/99 tolerance interval where a percentage p of the benchmarks is bounded with confidence level q, and σ_β is the nuclear data-induced uncertainty present in the application following the TSURFER data adjustment (i.e. the bias uncertainty).

Table III. Sample Problem Bias and USL Estimates

Case	TSURFER Bias Estimates	USLSTATS Bias Estimates	USLSTATS USL Estimates	TSURFER USL Estimates	Whisper USL Estimates
Dry (Normal)	-0.00282	-0.00214	0.9822	0.9939	0.9759
Wet (Upset)	-0.00165	-0.00421	0.9829	0.9953	0.9800

The USLSTATS and TSURFER methods were found to produce very similar computational bias estimates for the sample problem under normal conditions but produced bias estimates that disagreed more significantly under upset conditions. The TSURFER method used the same underlying data adjustments to compute both bias estimates and the USLSTATS method had more highly similar benchmark experiments available for the upset condition case, so it is difficult to say which method produced a more accurate bias estimate under upset conditions.

The USLSTATS and Whisper methods produced similar USL estimates for their normal and upset condition cases. Overall, the Whisper method produced more conservative USL estimates than the USLSTATS method, which given that the Whisper method was designed to be intentionally produce more conservative USL estimates than USLSTATS. These methods assumed a 99% confidence interval for their USL estimates, or a 99/99 confidence interval as applicable.

The TSURFER data adjustment process has the effect of reducing the size of the nuclear data-induced uncertainty in the application – this occurs because comparing the simulated benchmark responses against the experimental responses mitigates some of the initial uncertainty in the nuclear data [7]. The TSURFER-Perfetti method produced the highest USL estimates because the post-adjustment nuclear data uncertainty (σ_β) was roughly 10-20% as large as the initial nuclear data-induced uncertainty, suggesting that the TSURFER data assimilation process was able to successfully identify which nuclear data components contributed most significantly to the code computational bias. It should be noted that if the TSURFER USL calculation used the pre-adjustment nuclear data-induced uncertainty for σ_β, that it would have produced USL estimates that were significantly closer to, yet significantly more conservative than, the USLSTATS and Whisper estimates (0.9777 and 0.9657 for the normal and upset cases, respectively, compared to USLSTATS estimates of 0.9822 and 0.9829 for the normal and upset cases, respectively).

5 CONCLUSIONS

This study has provided a brief review of three statistical methods for estimating computational bias and USLs that are based on cutting-edge, high-fidelity sensitivity analysis methods, and has compared their bias and USL estimates for the sample application from Appendix D of ANSI/ANS–8.24 under both normal and upset conditions. The Whisper and USLSTATS methods produced similar USL estimates for both normal and upset conditions, while the TSURFER method's proposed USL estimation method produced significantly less conservative results. While it is not clear which method(s) produce the most accurate computational bias and USL estimates, it is our hope that future studies will investigate and benchmark these methods in greater detail.

NCS analysts must determine and understand the effects of approximations and errors in computer codes and nuclear data to quantify the accuracy of calculated eigenvalue estimates. The use of computer codes has become prevalent in the NCS industry, a trend that will be continued with the next generation of NCS staff. Coupled with a much larger set of NCS benchmarks available for use in validation, there is an industry need to develop tools and techniques to replace historical methods based solely on expert judgment, were often limited to benchmarks known to that expert or site, and that were often limited in their application to simple checks on quasi-quantitative parameters (EALF, H-X ratio, etc...)

It is almost impossible for any individual to be familiar with all of the benchmarks and their accepted application areas, and it is accepted that modern computer techniques can identify applicable benchmarks. Moving forward the industry needs to build trust in these sensitivity-based benchmark selection methods so that they may ultimately be used to determine computational biases and/or USLs for given applications. With these advancements and improved understanding of computational biases, the industry will have the opportunity to optimize operational activities based on reducing unnecessary conservatism. To reach this point there must be an increase in the scrutiny and importance placed on validation methods, as has been seen in recent years.

6 REFERENCES

1. B. T. Rearden, "Perturbation Theory Eigenvalue Sensitivity Analysis with Monte Carlo Techniques," *Nucl. Sci. Eng.*, **146**, 367–382 (2004).
2. C. M. Perfetti, B. T. Rearden, and W. R. Martin, "SCALE Continuous-Energy Eigenvalue Sensitivity Coefficient Calculations," *Nucl. Sci. Eng.*, **182**(3), 332–353 (2016).
3. B. C. Kiedrowski, F. Brown, and et al. "Whisper: Sensitivity/Uncertainty-Based Computational Methods and Software for Determining Baseline Upper Subcritical Limits." *Nucl. Sci. Eng.*, **181**(1), 17–47 (2017).
4. A. Santamarina, "Benchmark Phase-V Blind Benchmark on MOx wet powders," *OECD NEA Working Party on Nuclear Criticality Safety* (2015).
5. C. M. Perfetti, B. T. Rearden, "Estimating Code Biases for Criticality Safety Applications with Few Relevant Benchmarks," *Nucl. Sci. Eng.*, **193**(10), 1090–1128 (2019). DOI: 10.1080/00295639.2019.1604048
6. 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 (ORNL/TM-13211), Oak Ridge National Laboratory, Oak Ridge, Tennessee (1997).
7. M. L. Williams, et al., *TSURFER: An Adjustment Code to Determine Biases and Uncertainties in Nuclear System Responses by Consolidating Differential Data and Benchmark Integral Experiments*, ORNL/TM-2005/39 Version 6, Oak Ridge National Laboratory (2009).
8. B. L. Broadhead, et al., "Sensitivity- and Uncertainty-Based Criticality Safety Validation Techniques," *Nucl. Sci. Eng.*, **146**, 340–366 (2004).
9. W. J. Marshall and B. T. Rearden, *Criticality Safety Validation of SCALE 6.1*, ORNL/TM-2011/450 (Revised), Oak Ridge National Laboratory, Oak Ridge, Tennessee (2011).
10. *International Handbook of Evaluated Criticality Safety Benchmark Experiments*, Nuclear Energy Agency Nuclear Science Committee of the Organization for Economic Co-operation and Development, NEA/NSC/DOC(95)03, Paris (2019).