

# MAPPER – A NOVEL CAPABILITY TO SUPPORT NUCLEAR MODEL VALIDATION AND MAPPING OF BIASES AND UNCERTAINTIES

U. Mertyurek<sup>1</sup>, H.S. Abdel-Khalik<sup>2</sup>, W.B.J. Marshall<sup>1</sup>

<sup>1</sup> Oak Ridge National Laboratory  
1 Bethel Valley Road, Oak Ridge, TN 37831

<sup>2</sup> School of Nuclear Engineering, Purdue University  
West Lafayette, IN 47906

[mertyureku@ornl.gov](mailto:mertyureku@ornl.gov); [abdelkhalik@purdue.edu](mailto:abdelkhalik@purdue.edu); [marshallwj@ornl.gov](mailto:marshallwj@ornl.gov)

## ABSTRACT

This paper overviews the initial results of a new project at the Oak Ridge National Laboratory, supported via an internal seed funding program, to develop a novel computational capability for model validation: MAPPER. MAPPER will eliminate the need for empirical criteria such as the similarity indices often employed to identify applicable experiments for given application conditions. To achieve this, MAPPER uses an information-theoretic approach based on the Kullback-Leibler (KL) divergence principle to combine responses of available or planned experiments with application responses of interest. This is accomplished with a training set of samples generated using randomized experiment execution and application of high-fidelity analysis models. These samples are condensed using reduced order modeling techniques in the form of a joint probability distribution function (PDF) connecting each application response of interest with a new effective experimental response. MAPPER's initial objective will be to support confirmation of criticality safety analysis of storage facilities which require known  $k_{eff}$  biases for safe operation. This paper reports some of the initial results obtained with MAPPER as applied to a set of critical experiments for which existing similarity-based methods have been shown to provide inaccurate estimates of the biases.

**KEYWORDS:** sensitivity analysis, uncertainty analysis, similarity indices, criticality safety, bias prediction

## 1. INTRODUCTION

Validation of computation method or model is a key requirement for all engineering applications, requiring the analyst to provide proof that models can accurately predict real behavior based on the available body of experiments and computer analysis results. In most practical situations, the experimental conditions are only partially similar to the application due to many factors, including infeasibility to reproduce application conditions, construction cost, etc. Therefore, the ultimate goal is to devise scientifically defensible methods that can predict the expected bias between model predictions and real behavior for conditions that are not exactly reproduced by the experiments. For validation of nuclear systems, experimental data are routinely collected from the critical experiments and reactor startup tests. These data can be used for uncertainty assessment and bias estimation, taking into account differences between experimental and application

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

conditions such as variations in fuel type, structural materials, geometry, etc., in order to support model validation for a wide range of application conditions. The reactor startup procedures for detector calibration and rod worth calculations are also used to support bias calculations as the reactor core is driven to criticality.

The first step in any model validation is to determine the set of experiments most relevant to the application conditions. Existing nuclear systems validation techniques [1,2,3] leave some ambiguity with regard to rigorously proving that the selected experimental data can guarantee confidence in the estimated biases and their uncertainties for applications that could be sufficiently different from the experiments. For example, critical experiments are typically conducted at zero power conditions using small mock-ups of reactor geometry/composition, which is not sufficient to analyze the wide range of steady-state and transient power conditions. Although reactor startup tests provide data for the actual reactor core configurations rather than a reactor core-like mockup, they continue to be challenged, as they are conducted at very low power. The credibility question persists: how can the mapped biases and uncertainties be credible at conditions sufficiently different from experimental conditions?

Existing methods have relied on the concept of similarity to construct experiments with conditions that experts consider to be representative to application conditions. Then biases and uncertainties which occur at the experimental conditions are assumed to be applicable at the application conditions as a function of experiment's similarity (usually linear).

The challenge lies in determining how to judge similarity between two different systems. Existing methods, which have been primarily used for criticality safety applications employ the first-order derivatives of key responses of interest (e.g., multiplication factor, and spectral indices) with respect to nuclear data, which are assumed to constitute the major source of uncertainty. These nuclear data are typically referred to as *nuclear cross sections*, and they characterize the probabilities of interaction between radiation and matter.

Adjoint sensitivity analysis is typically used to calculate the derivatives of  $k_{\text{eff}}$  with respect to all nuclear cross sections for both the experimental and application conditions, collected into two  $n$ -dimensional vectors, called *sensitivity profiles*, with  $n$  being the number of cross sections used in the models. Using linear algebra, a single number, the *similarity index*, which has a value between 0 and 1, is calculated. The similarity index represents a weighted inner product of the two sensitivity profiles, with the weights calculated using the cross sections' covariance matrix. If the number is above a preset threshold such as 0.85, for example, then the experiments are deemed to be similar to the application [4].

Experiment selection based on a single similarity index can be misleading due to dominant sensitivities or uncertainties that are not relevant at application conditions. For example, two experiments could be deemed similar even though their associated biases are significantly different. This is one of the limitations of similarity-based approaches for validation which have become well-known over the years. Other limitations include the inability to account for errors/uncertainties from sources not common to both experiment, and application conditions, uncertainties in modeling parameters other than nuclear data, possible nonlinearities resulting from feedback terms at the application conditions, etc.

After selecting applicable experiments, cross sections are calibrated to minimize discrepancies between the measured and predicted experimental responses. The calibrated data are used in the application model, and the differences in the application model responses with the calibrated data and with the original data are accepted as the application model bias. The premise of calibration techniques for cross sections is that with the experimental discrepancies reduced, the adjustments will likely be adequate at the application conditions, with the caveat that the experiments are similar to the application.

Also, correcting for uncertainties might be a reasonable approach, as most calibration techniques cannot guarantee that they have adjusted for the correct sources of discrepancies. The limitations of calibration techniques have also been recognized over the years. These limitations include ill-posedness and error compensation phenomena when dealing with many model parameters, as well as the inability to include application conditions in the adjustment procedure [5]. The implication is that the adjusted parameters are blind to the application conditions. As in the similarity methods, the success of calibration relies heavily on the cleverness of the experimentalists to establish experiments that are very similar to the application conditions, and it also relies on the analysts to exclude experiments that could degrade the calibration results. For example, it has been observed that the bias could change dramatically upon adding/removing some experiments with low relevance, and the bias could be overconfidently estimated with small uncertainties, causing the results to show sensitivities to the experiments used. Explanation for this complex behavior has not been recorded in the nuclear engineering literature, which has diminished the value of calibration-based techniques.

The need for the MAPPER capability is based on the fact that existing validation philosophy has not kept pace with the information theory advances introduced in the mid 20<sup>th</sup> century. These advances started with Shannon's information content principle, which was further developed in the 1950s into the Kullback-Leibler (KL) divergence principle [6]. These advances have proved that the most rigorous approach to judge similarity is via a full PDF, which, in this context, can encode all possible variations for the experimental and application responses. Interestingly, the similarity index, which is a single number, can be derived from this PDF after making several simplifying assumptions, such as linearity, Gaussian uncertainties, and PDF marginalization over the common sources of uncertainties only. This PDF provides a natural approach to mapping the biases between the experimental and application domains by marginalizing biases over the PDF describing the measured response. This approach is also not restricted to being Gaussian, allowing all uncommon sources of uncertainties to be considered.

If the experiments provide little information about the application, then this joint PDF will not be informative, as it will have very wide scatter, indicating that application conditions cannot be confidently predicted using available experiments. The degree of correlation between the experiment and the application can be quantitatively measured using the concept of mutual information, which provides an acceptable approach to optimize and select the experiments that are best correlated with the application of interest.

## 2. DETAILS OF IMPLEMENTATION

MAPPER is a computational sequence designed to automate the construction of the noted joint PDF. For this initial proof-of-principle implementation, MAPPER is designed to leverage the SAMPLER code under the Oak Ridge National Laboratory (ORNL) SCALE suite of codes [8], which facilitates automated execution of uncertainty analysis. Given two sets of models—one representing the experiments, and one the application—MAPPER executes the SAMPLER code for all models and records all generated samples for the experimental and application responses. For this initial implementation, MAPPER uses the samples calculated directly by SAMPLER which are generated in a manner that preserves the statistical consistency of the cross section based on the prior covariance matrix.

The recorded samples for the response of interest represent a discrete representation of the sought joint PDF. Since the experiments contain numerous responses, a single effective response will be generated by maximizing the mutual information between the response and the application response of interest. This maximization process can be achieved using any number of parametric or nonparametric regression techniques, such as alternating conditional estimation or projection pursuit or reduced order modeling techniques. In this initial work, the alternating conditional expectations (ACE) algorithm is used to determine optimum relationships between the experiments and the application. Given the high similarity

between the experiments, the training samples for ACE must be preconditioned to ensure their independence. Details of input preparation for ACE will be discussed in a follow-up full-length journal article. Next, given available measurements and their uncertainties either from an existing or hypothesized (i.e., future, experiments), the constructed joint PDF will be used to calculate the full PDF for each application response. These PDFs can be reduced to biases and uncertainties representing mean values and standard deviations of the PDFs, per the user’s selection. The full implementation of MAPPER comprises four steps [7] (Figure 1):

- a. A comprehensive uncertainty quantification exercise using SAMPLER is executed for both experimental and application conditions by sampling all uncertain parameters within their prior uncertainties, including parameters that are both common and uncommon to the experimental and application conditions. Step a generates  $N$  (on the order of  $\sim 300\text{--}500$ ) samples for all responses of interest, which do not necessarily have to be of the same type. In the initial implementation, only cross section uncertainties will be propagated. For future work, both modeling and cross section uncertainties will be propagated.
- b. An automated procedure using nonparametric regression combines all samples from the experimental domain to construct a new effective experimental variable,  $k\text{-eff}\text{-exp}$ , which is selected to have the highest mutual information possible with an application response of  $k\text{-eff}\text{-app}$ . This effective response serves as a general nonparametric function of all available experimental data.
- c. A joint PDF is constructed between  $k\text{-eff}\text{-exp}$  and  $k\text{-eff}\text{-app}$  using kernel density estimation techniques.
- d. This joint PDF is used to propagate the biases and uncertainties of measured responses to the given application response. Steps b–d are repeated for each application response.

The computational cost of this algorithm is primarily associated with the cost of executing the simulation in step a. However, this must be done only once for available or new experiments. If a new experiment or a new application must be added, then the uncertainty quantification is executed for new models only.

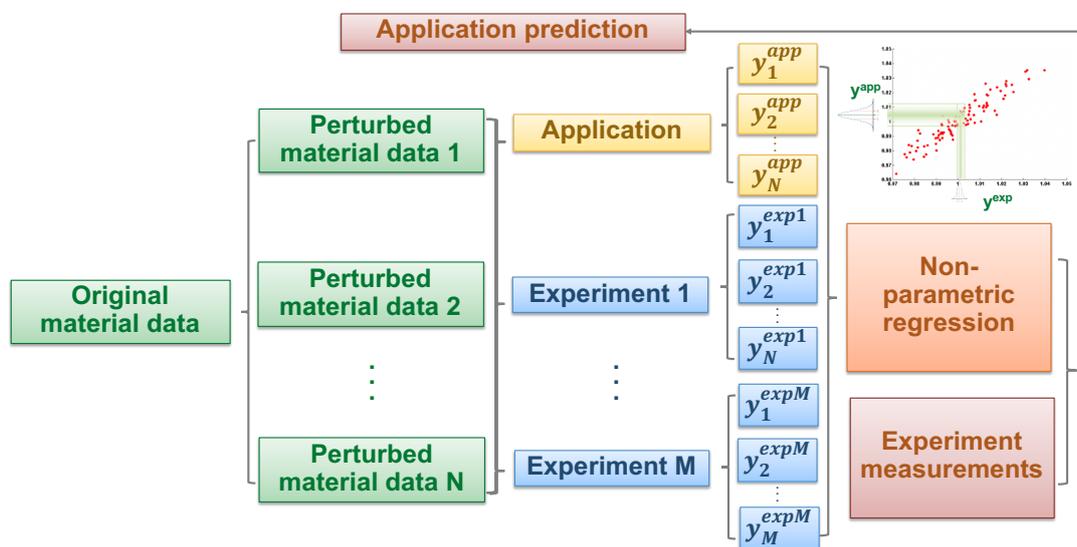


Figure 1. MAPPER Algorithm.

### 3. APPLICATION

A series of critical experiments were used to train the MAPPER algorithm and to predict biases for two experiments that are not included in the trained set. The two selected experiments (applications) are highly correlated (high similarity), but their biases are more than 600 pcm different, which represents a known challenge for similarity-based methods.

### 3.1. Critical Experiments

A set of critical experiments using highly enriched (> 60% U<sup>235</sup>) metal fuel and operating in fast spectrum (HEU-MET-FAST) were selected from the VALID [9] database for testing. The VALID database includes input models for more than 600 critical experiments from the ICSBEP Handbook [10]. These models are maintained by ORNL. The list of experiments from the selected set and the experimentally measured values [11] are provided in Table 1. The provided multigroup KENO (KENO V and KENO VI) input models were run for each experiment using SCALE 6.2.3 with the ENDF/B VII.1 56g cross section library. The difference between the measured and calculated  $k_{eff}$  values of Eq. (1) are also included in Table I.

$$Bias = k_{eff}^{calculated} - k_{eff}^{measured}, \quad (1)$$

Note that 56g library is optimized for thermal systems, and the biases are considerably reduced when the 252g library is used. Since the focus of this paper is bias estimation for a given model and nuclear data, the 56g library is adequate.

Some of the experiments include a series of cases with variations in the configuration; however, the modeling biases stay relatively consistent in general. Therefore, except for HEU-MET-FAST-025 experiments, only the first experiment from the series was included for testing. The HEU-MET-FAST-025 experiment includes 5 cases, and the modeling biases increase with increasing vanadium reflector thickness. Compared to other experiments with multiple cases, the first case (HEU-MET-FAST-025-1) and the last case (HEU-MET-FAST-025-5) show large differences in modeling biases (>600 pcm). Since these two systems are similar in configuration but show very different biases, predicting their biases is challenging. Therefore, the HEU-MET-FAST-025-1 and HEU-MET-FAST-025-5 experiments are used to simulate two applications with unknown biases, and they are also used to test the accuracy of MAPPER predictions. Axial configuration of HEU-MET-FAST-025-1 and HEU-MET-FAST-025-5 experiments are shown in Figure 2.

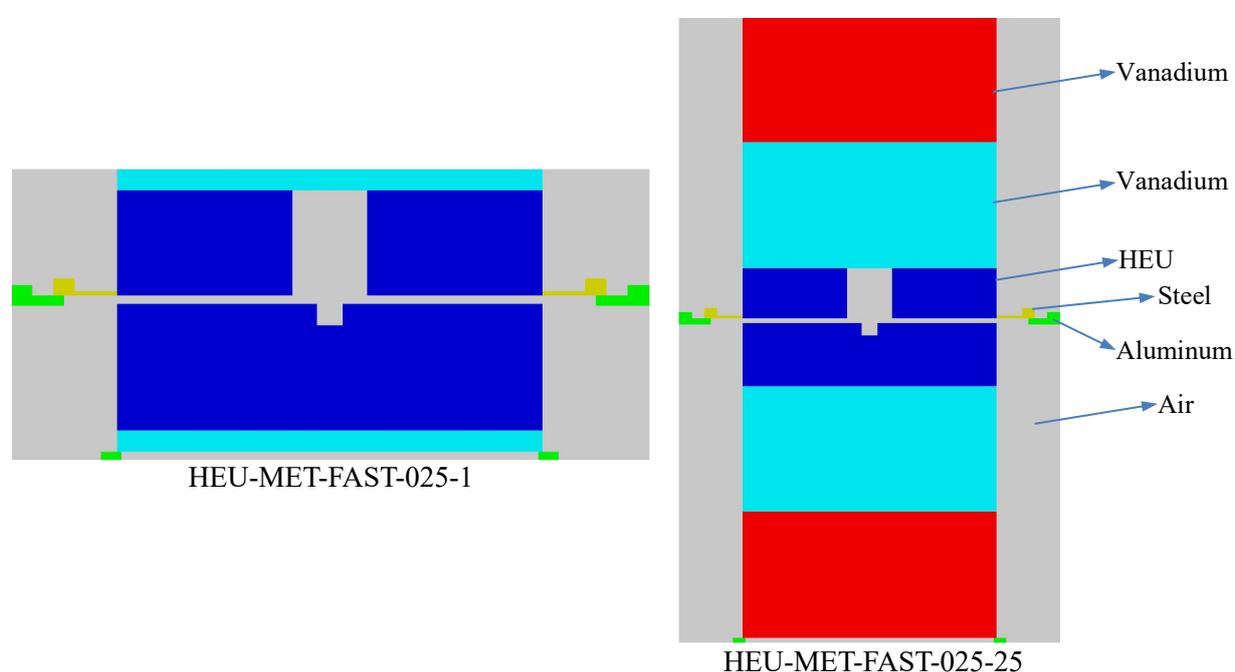
**Table I. Critical benchmark experiments [11].**

Evaluation	Case	Fissile material	Moderator	Measured*	Uncertainty (pcm)	Bias (pcm)	Similarity (c <sub>k</sub> )**
HEU-MET-FAST-005	1	90% enriched U metal	None	1.0000	360	-367.2	0.989
HEU-MET-FAST-008	1	90% enriched U metal	None	0.9989	160	-330.9	0.972
HEU-MET-FAST-009	1	90% enriched U metal	None	0.9992	150	-221.3	0.937
HEU-MET-FAST-010	1	90% enriched U metal	None	0.9992	150	-121.1	0.999
HEU-MET-FAST-011	1	90% enriched U metal	None	0.9989	150	-371.5	0.975
HEU-MET-FAST-013	1	90% enriched U metal	None	0.9990	150	-173.2	0.998
HEU-MET-FAST-015	1	96% enriched U metal	None	0.9996	170	-590.6	0.992
HEU-MET-FAST-016	1	96% enriched U metal	None	0.9996	180	164.5	0.994
HEU-MET-FAST-018	1	90% enriched U metal	None	1.0000	140	41.1	0.971
HEU-MET-FAST-019	1	90% enriched U metal	None	1.0000	280	715.6	1.000
HEU-MET-FAST-020	1	90% enriched U metal	None	1.0000	280	99.6	1.000
HEU-MET-FAST-021	1	90% enriched U metal	None	1.0000	240	694.9	0.999
HEU-MET-FAST-024	1	90% enriched U metal	None	0.9990	150	-181.9	0.998
HEU-MET-FAST-025	1	96% enriched U metal	None	0.9987	140	-25.9	0.998

HEU-MET-FAST-025	2	96% enriched U metal	None	0.9990	160	160.3	0.996
HEU-MET-FAST-025	3	96% enriched U metal	None	0.9991	160	426.6	0.999
HEU-MET-FAST-025	4	96% enriched U metal	None	0.9995	160	517.8	0.989
HEU-MET-FAST-025	5	96% enriched U metal	None	0.9991	160	607.4	0.997
HEU-MET-FAST-040	1	96% enriched U metal	None	0.9991	110	903.8	0.989
HEU-MET-FAST-065	1	96% enriched U metal	None	0.9995	130	-138.6	0.937
HEU-MET-FAST-080	1	93% enriched U metal	None	1.0000	120	692.3	0.999
HEU-MET-FAST-086	1	93% enriched U metal	None	0.9986	260	-1196.9	0.975
HEU-MET-FAST-092	1	96% enriched U metal	None	0.9986	110	237.9	0.998
HEU-MET-FAST-093	1	96% enriched U metal	None	0.9978	120	534.5	0.992

\*Evaluated value

\*\*similarity index  $c_k$  [7] calculated with respect to HEU-MET-FAST-025-1 experiment using TSUNAMI and TSUNAMI-IP sequences.



**Figure 2. KENO V Application Models.**

#### 4. RESULTS

All experiments listed in Table I were sampled for 300 perturbations to the ENDFB/VII.1 56g library using the 56g covariance library provided by SCALE. The calculated  $k_{eff}$  values were used in training the MAPPER algorithm. Evaluations for the critical experiments—except the applications (HEU-MET-FAST-025-1 and HEU-MET-FAST-025-25) and their associated uncertainties—were also provided to MAPPER. Table I shows the high similarity between the HEU-MET-FAST-025-1 and HEU-MET-FAST-025-2 experiments (0.996), as well as the high similarity between HEU-MET-FAST-025-1 and HEU-MET-FAST-025-5 experiments (0.997), Therefore, the relative bias shown in Eq. (2) calculated for the HEU-MET-FAST-025-2 experiment should be applicable to the other two experiments based on their similarities.

The results are shown in Table II. MAPPER accurately predicts measured  $k_{eff}$  values and corollary modeling biases. On the other hand, the bias calculated based on a similarity-based method (SCALE 6.2.3/TSURFER) overpredicts the computational model bias for HEU-MET-FAST-025-1 by 81 pcm and underpredicts the computational model bias for HEU-MET-FAST-025-5 by 305 pcm. The reason for this over and underprediction is that while the included experiments have strong similarity to the application, their biases are inconsistent, which forces similarity-based methods to average the biases to calculate the application bias.

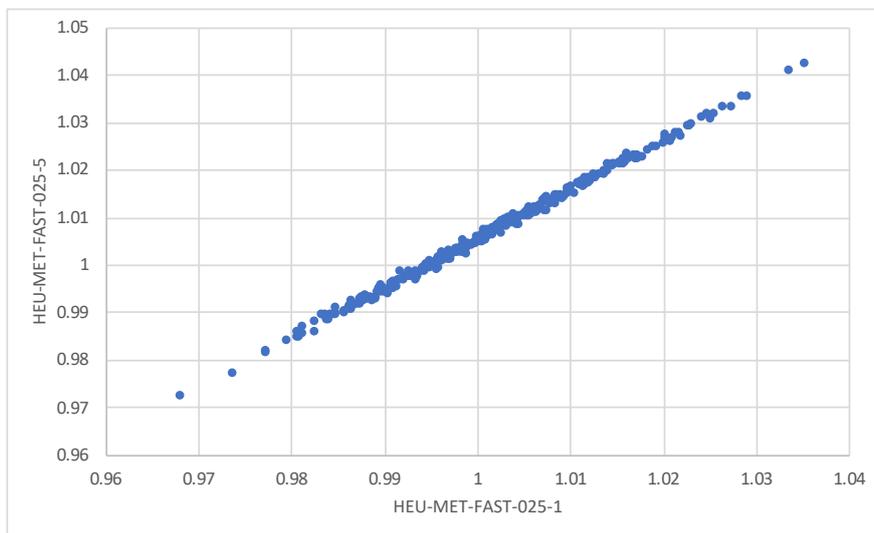
$$Relative\ Bias = \frac{k_{eff}^{calculated} - k_{eff}^{measured}}{k_{eff}^{measured}} \quad (2)$$

To illustrate this, Figure 3 plots all the recorded samples for two of the noted experiments: HEU-MET-FAST-025-1 and HEU-MET-FAST-025-5, which were selected in this study to represent applications showing an extremely high degree of correlation with a similarity index ( $c_k$ ) exceeding 0.997. While these two experiments behave similarly with respect to cross section perturbations, and the numerical model (KENO V) accurately predicts  $k_{eff}$  for HEU-MET-FAST-025-1, the numerical model for HEU-MET-FAST-025-5 shows a bias greater than 600 pcm as compared to the measured value. This forces similarity-based methods to calculate the bias as a weighted average of the two biases.

**Table II. Comparison of measured and predicted  $k_{eff}$  and biases**

	HEU-MET-FAST-025-1		HEU-MET-FAST-025-5	
Application	Bias (pcm)	Uncertainty	Bias (pcm)	Uncertainty
Computational Model*	-25.9	140	607.4	160
Similarity-based	56.2	56	302.5	50
MAPPER-based	-35.9	67	577.4	160

\*reference value [11]



**Figure 3. Comparison of  $k_{eff}$  for Cross Section Perturbations.**

## 5. CONCLUSIONS

This paper discusses the initial results for a prototypic implementation of the MAPPER sequence, a new tool within SCALE which was developed to support model validation by addressing the challenges of existing similarity-based and calibration-based techniques. MAPPER relies on the use of information theory rather than sensitivity analysis to establish the relevance of experiments to the given applications. As supported by information-theoretic principles, some of the existing challenges facing similarity and calibration-based methods can be eliminated—such as the sensitivity of the solutions to the prior uncertainties—thus precluding the need for prescreening experiments based on an empirical approach, as in the use of similarity indices. This eliminates the need for evaluation of sensitivity coefficients, which could be computationally infeasible, etc. Future work will focus on expanding the implementation of MAPPER to analyze the impact of adding and removing experiments on the calculated biases and their uncertainties, with the goal of fully automating the experiments selection as opposed to the trial and error approach currently used in similarity-based methods.

## REFERENCES

1. B. Broadhead et al., “Sensitivity and Uncertainty-based Criticality Safety Validation Techniques,” *Nucl. Sci. Eng.*, **146**, pp. 340–366 (2004).
2. G. Palmiotti et al., “Developments in Sensitivity Methodologies and the Validation of Reactor Physics Calculations,” *Sci Tech. Nucl. Installations*, **2012** (2012).
3. B. Kiedrowski et al., “Whisper: Sensitivity/Uncertainty-Based Computational Methods and Software for Determining Baseline Upper Subcritical Limits,” *Nucl. Sci. Eng.*, **181** (1), pp. 17–47 (2017).
4. B. T. Rearden, et al., “Sensitivity and Uncertainty Analysis Capabilities and Data in SCALE”, *Nucl. Technol.* **174** (2), pp. 236–288 (2011).
5. A. Tarantola, “Inverse Problem Theory,” SIAM (2005).
6. D. Commenges, “Information Theory and Statistics,” <https://arxiv.org/pdf/1511.00860.pdf>
7. H. Abdel-Khalik et al., “Physics-Guided Covered Mapping: A New Approach for Quantifying Experimental Coverage and Bias Scaling,” *Trans ANS*, **112** (2015).
8. B. T. Rearden and M. A. Jessee, *SCALE Code System*, ORNL/TM-2005/39, Version 6.2, Oak Ridge National Laboratory, Oak Ridge, TN (2017).
9. W. J. Marshall and D. A. Reed, *SCALE Procedure for Verified, Archived Library of Input and Data (VALID)*, SCALE-CMP-012, Revision 2, Oak Ridge National Laboratory, Oak Ridge, TN (2013).
10. *International Handbook of Evaluated Criticality Safety Benchmark Experiments*, NEA/NSC/DOC(95)03, Nuclear Energy Agency/Organisation for Economic Co-operation and Development, Paris, France (2016).
11. E. M. Saylor et al., *Criticality Safety Validation of SCALE 6.2.2*, ORNL/TM-2018/884, Oak Ridge National Laboratory, Oak Ridge, TN (2018).
12. D. Huang and H. Abdel-Khalik, “Construction of Optimized Experimental Responses in Support of Model Validation via Physics Coverage Mapping Methodology,” PHYSOR 2016, Sun Valley, ID (2016).