

Methodology and issues of integral experiments selection for nuclear data validation

Ivanova Tatiana^{1,a}, Evgeny Ivanov², and Ian Hill¹

¹ OECD-NEA, Nuclear Science Division, Boulogne-Billancourt, France

² Institut de Radioprotection et de Sûreté Nucléaire, PSN-RES/SAG, 92262 Fontenay-aux-Roses, France

Abstract. Nuclear data validation involves a large suite of Integral Experiments (IEs) for criticality, reactor physics and dosimetry applications. [1] Often benchmarks are taken from international Handbooks. [2, 3] Depending on the application, IEs have different degrees of usefulness in validation, and usually the use of a single benchmark is not advised; indeed, it may lead to erroneous interpretation and results. [1] This work aims at quantifying the importance of benchmarks used in application dependent cross section validation. The approach is based on well-known General Linear Least Squared Method (GLLSM) extended to establish biases and uncertainties for given cross sections (within a given energy interval). The statistical treatment results in a vector of weighting factors for the integral benchmarks. These factors characterize the value added by a benchmark for nuclear data validation for the given application. The methodology is illustrated by one example, selecting benchmarks for ^{239}Pu cross section validation. The studies were performed in the framework of Subgroup 39 (Methods and approaches to provide feedback from nuclear and covariance data adjustment for improvement of nuclear data files) established at the Working Party on International Nuclear Data Evaluation Cooperation (WPEC) of the Nuclear Science Committee under the Nuclear Energy Agency (NEA/OECD).

1. Introduction

Nuclear data (ND) validation is part of the global validation of analytical tools and is intended to support nuclear industry applications, and so, the results should provide essential information for decision making in support of nuclear safety, design and operation. Results of validation could be, and commonly are, used to establish problem oriented research programs, guiding future experimental campaigns to support validation cases.

Validation is a process based on experimental observations; otherwise inconsistencies with reality will be present. Nuclear data validation, in particular, involves a large suite of Integral Experiments (IEs) for criticality, reactor physics and dosimetry applications. [1] Experimental benchmarks can be obtained from international the Handbooks [2,3], or other sources.

Nuclear data validation is a complex process owing to the need to take into account several correlated physical phenomena – nuclear fission, radiation capture, neutrons scattering and slowing-down, all elastic and inelastic neutron-nuclei interactions etc. Each of these phenomena are present, to various degrees, in the available IEs. For a robust validation case, it is required that a statistically significant number of IEs be used in validation [1].

Another consideration when performing validation is the importance of separating the experimental data to

be used in a validation process from the data used for ND and/or nuclear models calibration. There should be two sets of IEs data: 1) integral data applied during ND evaluation and 2) integral data to be used during validation.

The ND validation process, like others, deals with a given application domain in which the validator wishes to accurately quantify the biases and uncertainties. Worldwide practices distinguish four groups of phenomenological validation methodologies. Each has advantages and drawbacks, and none can be considered superior. However, we consider the Bayesian based methodologies to be the reference [1] since they combine knowledge from both differential and integral experiments. In this paper we illustrate the reference validation method using deterministic version of Bayesian approach – Generalized Linear Least Squared Method (GLLSM) [1].

Of primary importance is that conclusion of validation activity gives a clear understanding of the biases and uncertainties in the given domain of applicability. Otherwise the validation methodology is inconsistent.

In this article we present examples of one criticality safety case [4] and of an extended pseudo- application – the resonance integral of ^{239}Pu fission – in order to demonstrate application to practical cases as well as directly to nuclear data. In addition it is demonstrated that the Validation and Uncertainty Quantification (V&UQ) process gives enough information to establish a ranking table of IEs, characterizing their importance in a given application domain.

^a e-mail: tatiana.ivanova@oecd.org

2. Phenomenological validation strategies

Reference [5] defines phenomenological validation as the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model. In other words, validation characterizes the predictive capability of given tools/data within an application domain based on available observations using different methodologies.

Observations, in ND validation, encompass all available set of IEs – including criticality, reactor physics and transition ones – or a subset of these. V&UQ deals with IE data, e.g., with observations that came from different types of measurements.

Some basic concepts are required before discussing the performance of different V&UQ techniques.

- 1) Observation means the ratios between measured and calculated values.
- 2) Topology of the experimental case or the application is a concept that gives unique (unambiguous) characterization of its neutron state, physics (spectra, importance functions etc.). For example [7]:
 - EALF – energy corresponding to the average lethargy of neutrons causing fission
 - ANECF – average energy of neutrons causing fission
 - $H/^{239}\text{Pu}$ or $H/^{235}\text{U}$ ratios, for solutions
 - Fraction of fissions caused by fast ($E < 100$ keV), intermediate ($1 \text{ eV} < E < 100$ keV), and thermal ($E < 1 \text{ eV}$) neutrons
 - $^{238}\text{U}(n,f)/^{235}\text{U}(n,f)$, $^{237}\text{Np}(n,f)/^{235}\text{U}(n,f)$, other ratios
 - etc.

The V&UQ techniques differ in terms of transposition, e.g., how they operate with information included in observations and how they take into account the different types of information. Following this we can divide the V&UQ techniques in four groups.

1. The first group – traditional weighted mean bias – is intended to characterize the tools and the libraries generally, e.g., uniformly for all plausible applications [6].
2. The second group – trending analysis [7] – performs extrapolation along given trending variable.
3. The third group – based on non-parametric statistics – includes Principle Component Analysis (PCA) [8] and different kinds of factor and cluster analysis.
4. The fourth group of V&UQ techniques consists of the Bayesian-based ones – deterministic [1] or stochastic [9].

The first group is the most popular because it is both the simplest method and has applied frequently. In this approach the user computes the following weighted bias:

$$\Delta_m = \left\{ \sum_i [(R_C - R_E)_i / \sigma_{Ei}^2] \right\} \cdot \left\{ \sum_i [1 / \sigma_{Ei}^2] \right\}^{-1}, \quad (1)$$

where Δ_m is the weighted bias for the m-th library, R_C , R_E and σ_{Ei} are; calculated and measured i-th benchmark's

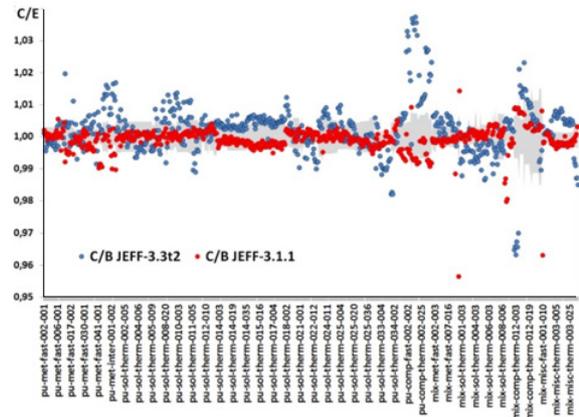


Figure 1. The observations for full list (635 cases) of the plutonium containing benchmarks.

Table 1. Weighted average biases in traditional analysis, pcm.

N° cases	N=635	N=238	N=139
JEFF-3.3t2	−88	−297	−220
JEFF-3.1.1	−54	−205	−176

values and experimental uncertainty, respectively [6]. The reduced weighted standard deviations (σ_m) are computed as follows:

$$\sigma_m = \sqrt{\left\{ \sum_i [(R_C - R_E)_i - \Delta_m]^2 / \sigma_{Ei}^2 \right\} \cdot \left\{ \sum_i 1 / \sigma_{Ei}^2 \right\}^{-1}}, \quad (2)$$

The application domain is often vaguely determined in this analysis – for example “all systems containing plutonium” or “all reactors with thermal spectra”. Furthermore, from a regulatory perspective it is probably the least attractive as the results depend on the number of benchmarks, i.e., traditional validation is strongly user-dependent.

For example, let consider calculated using two versions of JEFF library (3.1.1 and 3.3t2) [10] of ICSBEP [2] benchmarks.

Taking all available (635) cases we have a mean weighted biases of −88 pcm and −54 pcm for JEFF 3.3t2 and 3.1.1 respectively. Since there is no clear guidance how many benchmarks to be taken for validation we could reduce the number as we want, for instance, excluding all homogeneous thermal solutions. The resulting subset (238 cases) gives mean biases of −294 and −205 respectively. Next reduction (withdrawing all thermal cases) remaining 139 cases gives −220 and −179 pcm respectively (see Table 1).

The results are that there is a drift based on the chosen cases; no rules of the traditional method were broken in filtering cases. It is apparent that the major issue with the traditional approach is ambiguity both in application characterization and in choice of benchmarks.

The second group of V&UQ – trending analysis [7] – is performed by calculating a linear fit of the experiment results as a function of the trend parameters. The issue is the same – ambiguity in the use of trending parameters.

The third group of V&UQ, non-parametric statistics based Principal Component Analysis (PCA) [8] collapses

Table 2. The parameters of the applications – criticality safety case and fission integral for ²³⁹Pu.

Application object	Safety case				XS integral I _f (²³⁹ Pu)
	Case 1	Case 2	Case 3	Case 4	
humidity/energy	ε = 0.1%	ε = 1%	ε = 3%	ε = 5%	1 MeV–100 eV
EALF	4 keV	1 keV	300 eV	90 eV	–

given computational outputs or assigned parameters – topology – into few synthetic variables and builds extrapolation/scaling along them. Such methods are non-invasive they depend on completeness of the initially postulated variables. The major issue, though, is that if trial (prior) variables are incomplete and/or redundant, the validation process will lead to ambiguous results.

The fourth group, Bayesian based approaches can be deterministic and/or stochastic, but are always based on Bayes theorem to find an optimal solution using mathematical statistics to ill-posed inverse problem solution [1,9]. The deterministic alternative Generalized Linear Least Square Methods (GLLSM) [1] is intrusive while stochastic ones are non-intrusive [9].

The observations – validation domain – for Bayesian approaches do not differ from other techniques. One could notice that any Bayesian technique – both stochastic and GLLSM – provides users with ranking tables characterizing similarity between each IEs case and a given application object.

The application domain in Bayesian methodology is physically determined using the “phase space” of sensitivity functions to ND. This is why transposition is always unambiguous when validation is done with the Bayesian approach.

Since they are based on the same theory, we illustrate the reasoning using GLLSM formulae as it is more transparent than other realizations of the theory.

3. Illustration: Criticality safety case

Let consider wet plutonium safety cases [4] as the application domain (see Table 2).

The last case is a pseudo-application – ²³⁹Pu fission resonance integral computed as follows:

$$RI_{FISS} = \int_{INTR} \sigma_{239}^{FISS} \cdot \frac{dE}{E}, \quad (3)$$

where RI_{FISS} , σ_{239}^{FISS} and E are the resonance integral, fission cross section and energy (integration with Fermi spectra), respectively.

Figure 2 presents traditional validation using recent JEFF’s versions. Our example case corresponds to the last set of bars (139 cases) column and gives ~ 200 pcm of bias.

Using GLLSM we mapped the 139 observations onto biases and uncertainties for the previously noted applications (see Table 3 and Fig. 3).

The biases for different application cases have even different signs – the fastest spectrum case ~ +800 pcm and the slowest ~ -3500 pcm with uncertainties up to 1000 pcm.

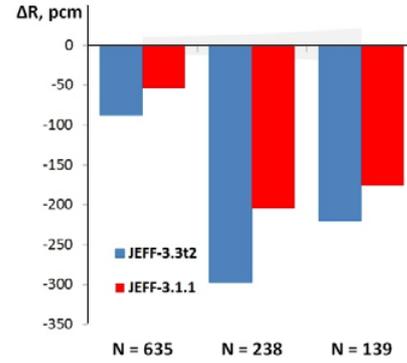


Figure 2. Traditional mean biases for ²³⁹Pu containing systems.

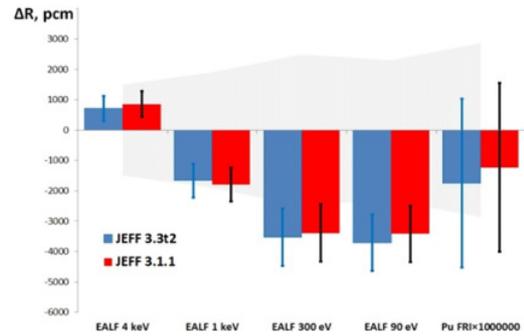


Figure 3. Biases and uncertainties in application domain.

Table 3. The biases and uncertainties for given applications.

		Case 1	Case 2	Case 3	Case 4	I _f (²³⁹ Pu)
σ	prior	1492	1904	2489	2295	2863
	posterior	417	564	943	927	2780
bias	JEFF 3.3t2	710	-1669	-3536	-3713	-1743
	JEFF-3.1.1	856	-1796	-3391	-3419	-1234

Thus the comfortable errors of 200 ÷ 300 pcm being extrapolated onto application domain give ~ 4000 pcm of the error demonstrating the limited progress in knowledge improvement.

4. The extension of Bayesian techniques to IEs selection

It should be addressed how to select the benchmarks for validation. Bayesian tools operate with two kinds of the output – 1) observation free and 2) observation dependent. These points are expanded upon in this section.

4.1. Observation free outputs

Posterior uncertainties are observation independent. They are represented as follows:

$$\sigma_{AO}^2 = \bar{S}_{AO}^T \cdot \hat{W}' \cdot \bar{S}_{AO}. \quad (4)$$

where notations are the same as before.

They depend on the relevant sensitivity coefficients, e.g., on the physics. Also they depend on nuclear data and IEs covariance matrices. But there is no room for observations $\Delta \bar{R}$.

Therefore we could arrange them before and independently based on an estimation of the calculation-measurement discrepancies.

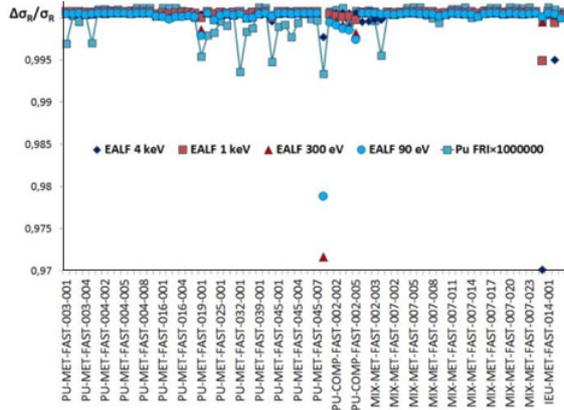


Figure 4. The URF computations for different benchmarks cases and for different applications.

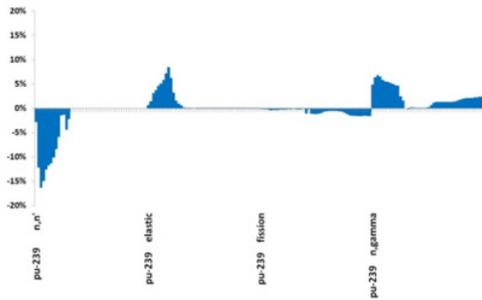


Figure 5. Correction factors computed using GLLS approach for ^{239}Pu cross sections.

In the past such analysis included representatively factors (\mathbf{r}) computed as follows:

$$\mathbf{r} = \sqrt{\frac{(\bar{\mathbf{S}}_{AO}^T \cdot \hat{\mathbf{W}} \cdot \hat{\mathbf{S}}_B) \cdot (\hat{\mathbf{S}}_B^T \cdot \hat{\mathbf{W}} \cdot \bar{\mathbf{S}}_{AO})}{\hat{\mathbf{S}}_B^T \cdot \hat{\mathbf{W}} \cdot \hat{\mathbf{S}}_B}} \quad (5)$$

The factors were derived assuming IEs independence.

Unfortunately all IEs are correlated in certain sense – the simple formulae are not strictly correct for real cases. However, we can compute similar criteria iteratively as demonstrated in Fig. 4.

Hereinafter we will denote them as factors of posterior uncertainty reduction, or uncertainty reduction factors (URF). As far as the URFs clearly illustrate the gaps in IEs suites they can be used both for the selection of IEs cases and for the planning new experiments.

4.2. Observation dependent outputs

Values that depend on the observation are the bias and correction factors assigned nuclear reaction cross sections¹; examples are presented in Fig. 5.

Notably, the contribution of each benchmark can be computed and based on this we can estimate the role of each individual benchmark in the adjustment of nuclear data for each individual nuclide-reaction (see Fig. 6).

We can see that the contributions of different benchmarks have different sizes and even different signs.

¹ These correction factors relay to the elements of the group-wise transport operator but not to the nuclear data. They can be used indirectly in the further data correction/evaluation.

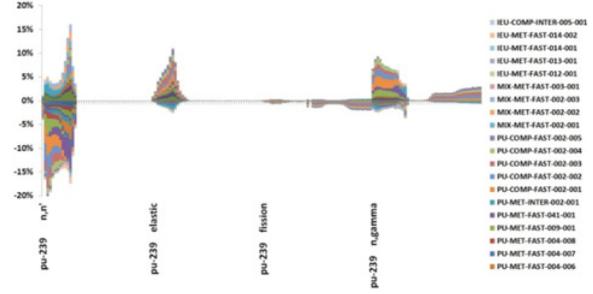


Figure 6. Contributions of the different benchmarks on ^{239}Pu cross sections adjustment.

Many authors have associate controversial contributions with a compensation effect [1] while it could be driven as well by inevitable roughness of the sensitivity formalism.

The results don't compromise the methodology but rather enlighten the fact that convergence in mathematical statistics should be considered a statistical process. Furthermore, this means that using the entire suite of the benchmarks may make sense, without the result of each component being correct.

At the same time it seems inconvenient to analyze correction factors needed to qualify each individual cross section. Instead of analysis of individual cross sections we can associate the target nuclear data with an application object. We did this previously by including resonance integral of the ^{239}Pu fission as one of the applications previously presented. Generally speaking any user is free to formulate any application object, with the only one requirement being the existence of linear sensitivity coefficients. Therefore using any sophisticated sensitivity vectors we can represent any data with their inter-correlations.

All considerations before were based on the sensitivity calculations. However not all codes are able to compute the sensitivity coefficients but can provide high-fidelity computations of the design or the safety parameters.

Taking into account the additivity of the bias assessment and the conservatism of the sensitivity computations we can construct the vector of coefficients associated with each given application in order to characterize the resulting bias, when applied to the observation cases.

The bias for an application is given by:

$$\mathbf{bias} = (\bar{\mathbf{S}}_{AO}^T \cdot \hat{\mathbf{W}} \cdot \hat{\mathbf{S}}_B) \cdot (\hat{\mathbf{V}}_{\text{EXP}} + \hat{\mathbf{V}}_{\text{CLC}} + \hat{\mathbf{S}}_B^T \cdot \hat{\mathbf{W}} \cdot \hat{\mathbf{S}}_B)^{-1} \cdot \Delta \bar{\mathbf{R}}, \quad (6)$$

where $\Delta \bar{\mathbf{R}}$ means the vector of observations. The equations can be re-written in the following form:

$$\mathbf{bias} = RF_1 \cdot \Delta R_1 + RF_2 \cdot \Delta R_2 + \dots + RF_N \cdot \Delta R_N, \quad (7)$$

where ΔR_i means the i -th observation and RF_i is relevant coefficient computed using matrices determined above. We can establish these bias ranking factors (BRF) – for each observation, for a set of given applications (see Fig. 7).

It seems reasonable to reduce the number of benchmarks for practical needs (for efficient analysis). We can keep the only benchmarks with sufficient

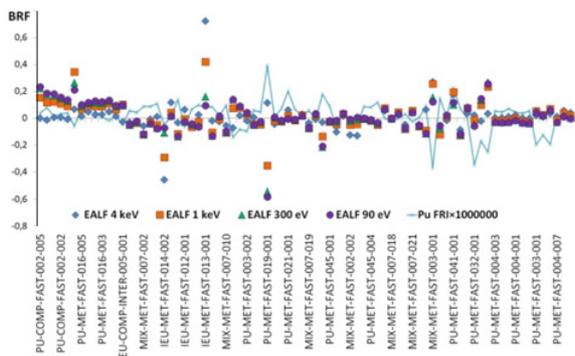


Figure 7. Computed bias ranking factors (BRF) for plutonium integral benchmarks for different applications.

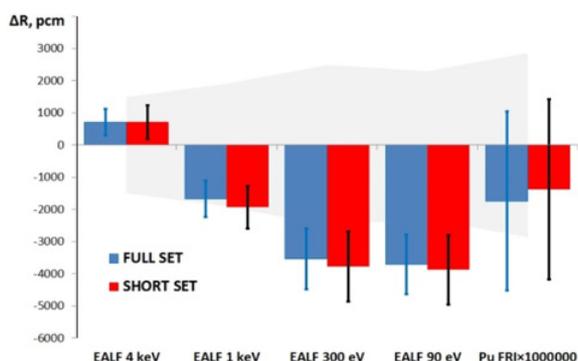


Figure 8. Biases and posterior uncertainties for different applications computed using short list (23 cases) of the benchmarks.

contributions in the resulting uncertainties². The figure below demonstrates the application of such reduced list of the benchmarks.

Comparing Figs. 3 and 8 we see that there are no notable differences between the full and the short suits of the benchmarks.

Thus the combination of the URF and BRF allow one to perform a comprehensive validation, and efficiently analyze the biases and uncertainties for different applications and tools. For example, a combination of the two factors can establish a ranking table with the most impactful benchmarks for any given purpose.

5. Conclusions

Due to progress in numerical tools and computational power we have entering into a new reality even when compared to ten years ago. The conjuncture is true for validation researchers as well. Nowadays we can practically neglect any methodological errors using enough powerful hardware and Monte Carlo methods.

In the Monte Carlo methods experimental validation of the evaluated nuclear data libraries becomes the first priority as the nuclear data uncertainties dominate over all others. Ideally validation should be physically based and robust, and user-independent, of course.

Such requirements for validation can be met using Bayesian-based approaches and available high-fidelity Handbooks of the IEs.

² The criteria of the sufficiency should be established taking into account the required accuracy etc..

It appears attractive to feedback the results for V&UQ on the planning of new IE programs. Bayesian approaches could help also here. For example as we discussed before we can identify gaps in the existing validation base.

The phenomenological validation can never be cheap as it certainly requires representative mockups for each application in the validation matrix – for a variety of applications such as critical as shielding or reaction rates measurements. It may also require a significant expense if there are no enough available benchmarks or the benchmarks were previously used in the nuclear data evaluation process as Godiva or Jezebel. In this case new integral experiments will be needed.

The new landscape can make validation more easily available and affordable: precise calculations of the particle transport and the criticality, fine-mesh ND treatment and high-fidelity IEs data (the Handbooks), and high-fidelity or even precise sensitivity analysis. These can circumvent the needs for mockups.

Crucial components for comprehensive validation are: availability of high-fidelity IE data with covariances, consistent ND covariances, and precise analytical and sensitivity analysis tools.

Users need to be informed about the IEs cases that have been yet applied for differential experiments calibration and for ND evaluation in order to avoid the double use of the IEs data.

The functionals computed using Bayesian methodology – residual uncertainties (σ_{RES}), bias ranking factors (BRF), uncertainty reduction factors (URF) – can comprehensively characterize the available IEs data set and can provide sufficient basis to design new experiments.

In our practical example of criticality safety case we have seen much higher predicted biases of reactivity than traditional χ^2 analysis. So the real ND knowledge remains rather limited despite on recent effort on their re-evaluation.

The performed works revealed certain problems which need to be addressed and are encapsulated by the following suggestions.

Suggestion 1: Advanced validation process should be extended to the assess the knowledge in “spatial sense”, i.e., it should be intended to qualify the consistency of both ND and the covariance matrices.

Suggestion 2: Further efforts on new ND evaluation and new generations of analytical tools development should be harmonized with the establishment of ND covariance matrices, IEs covariances and with access to high-fidelity benchmarks (including proprietary).

Suggestion 3: it would be valuable for validation if the next generation of evaluated ND libraries contains information about how IEs cases have been used for differential experiments calibration and ND evaluation.

Suggestion 4: The validation process should be a systematic approach aimed at, amongst other goals, identification of the gaps in data and models and, more importantly, on comprehensive support of guiding the selection of further experiments to be performed or analyzed.

This research was carried out within the framework of Subgroup 39 (Methods and approaches to provide feedback from nuclear

and covariance data adjustment for improvement of nuclear data files) established at the Working Party on International Nuclear Data Evaluation Cooperation (WPEC) of the Nuclear Science Committee under the Nuclear Energy Agency (NEA/OECD). We thank our colleagues from different institutes and countries who participating in the Sub-group provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

We thank coordinators of WPEC/Sg39 M.Salvatores and G.Palmiotti for numerous fruitful discussions that encouraged us to perform this research work.

References

- [1] M. Salvatores, et al., “Methods and Issues for the Combined Use of Integral Experiments and Covariance Data: Results of a NEA International Collaborative Study”, *Nuclear Data Sheets* **118**, 38–71 (2014)
- [2] International Handbook of Evaluated Criticality Safety Benchmark Experiments, NEA/NSC/DOC(95)03, OECD Nuclear Energy Agency (2014)
- [3] International Handbook of Evaluated Reactor Physics Benchmark Experiments, Organization for Economic Cooperation and Development – Nuclear Energy Agency, NEA/NSC/DOC(2006)1 (2011 Edition)
- [4] The need for Integral Critical Experiments with Low-moderated MOX Fuels, OECD-NEA/NSC Workshop proceedings, Paris, France, 2004
- [5] William L. Oberkampf, Timothy G. Trucano, Verification and validation benchmarks, *Nuclear Engineering and Design* **238**(3), 716–743 (2008)
- [6] Rowlands, J.L. Accuracy of criticality calculations made using current nuclear data libraries for MOX fuelled fast critical assemblies/ *Annals of Nuclear Energy* **37**(12) (2010)
- [7] J.C. Dean, R.W. Tayloe, Jr., and D. Morey, Guide for Validation of Nuclear Criticality Safety Computational Methodology , NUREG/CR-6698, prepared for the US Nuclear Regulatory Commission by Science Applications International Corporation, Oak Ridge, Tenn., January 2001
- [8] F. Fernex, Y. Richet, and E. Létang, “MACSENS: A New MORET Tool to Assist Code Bias Estimation,” Integrating Criticality Safety into the Resurgence of Nuclear Power, Knoxville, Tennessee, September 19–22, 2005
- [9] A.J. Koning, Bayesian Monte Carlo Method for Nuclear Data Evaluation, *Nuclear Data Sheets* **123**, 207–213 (2015)
- [10] S. van der Marck, “Benchmarking ENDF/B-VII.1, JENDL-4.0 and JEFF-3.1.1 with MCNP6”, *Nuclear Data Sheets* **113**, 2935–3005 (2012)