

## Informing nuclear physics via machine learning methods with differential and integral experiments

Denise Neudecker<sup>1,\*</sup>, Oscar Cabellos<sup>2</sup>, Alexander R. Clark<sup>1</sup>, Michael J. Grosskopf<sup>1</sup>, Wim Haeck<sup>1</sup>,  
 Michal W. Herman<sup>1</sup>, Jesson Hutchinson<sup>1</sup>, Toshihiko Kawano<sup>1</sup>, Amy E. Lovell<sup>1</sup>, Ionel Stetcu<sup>1</sup>,  
 Patrick Talou<sup>1</sup> and Scott Vander Wiel<sup>1</sup>

<sup>1</sup>*Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

<sup>2</sup>*Department of Energy Engineering and Instituto de Fusión Nuclear, Universidad Politécnica de Madrid, 28006 Madrid, Spain*



(Received 15 April 2021; accepted 23 August 2021; published 10 September 2021)

**Background:** Information from differential nuclear-physics experiments and theory is often too uncertain to accurately define nuclear-physics observables such as cross sections or energy spectra. Integral experimental data, representing the applications of these observables, are often more precise but depend simultaneously on too many of them to unambiguously identify issues in the observable with human expert analysis alone.

**Purpose:** We explore how we can leverage physics knowledge gained from differential experimental data, nuclear theory, integral experiments, and neutron-transport calculations to better understand nuclear-physics observables in the context of the application area represented by integral experiments. We support this task with machine-learning methods to discern trends in a large amount of convoluted data.

**Methods:** Differential and integral information was used in an analysis augmented by the random forest and the Shapley additive explanations metric. We chose as an application area one that is represented by criticality measurements and pulsed-sphere neutron-leakage spectra.

**Results:** We show one representative example (<sup>241</sup>Pu fission observables) where the combination of differential and integral information allowed to resolve issues in data representing these observables. As a starting point, the machine learning (ML) algorithms highlighted several observables as leading potentially to bias in simulating integral experiments. Differential information, paired with sensitivity to integral quantities, allowed us then to pinpoint one specific observable (<sup>241</sup>Pu fission cross section) as the main driver of bias. The comparison to integral experiments, on the other hand, allowed us to indicate a likely reliable experiment among several discrepant ones for this observables. In other cases (e.g., <sup>239</sup>Pu observables), we were not able to resolve the confounding introduced by integral experiments but instead highlighted the need for targeted new experiments and theory developments to better constrain the nuclear-physics space for the application area represented by integral experiments.

**Conclusions:** We were able to combine information from differential experimental data, nuclear-physics theory, integral experiments, and neutron-transport simulations of the latter experiments with the help of the random forest algorithm and expert judgment. This combination of knowledge allows to improve our description of nuclear-physics observables as applied to a particular application area.

DOI: [10.1103/PhysRevC.104.034611](https://doi.org/10.1103/PhysRevC.104.034611)

### I. INTRODUCTION

Experimental data of nuclear-reaction processes are typically obtained for a particular target isotope at a given incident energy  $E$  of a projectile. Examples of such observables are reaction cross sections  $\sigma(E)$ , energy or angular distributions of secondary particles  $d\sigma/d\Omega$ ,  $d\sigma/dE$ , etc. Theoretical model calculations for these reaction processes provide the same type of information: nuclear-reaction probabilities at a fixed energy. More explicitly, nuclear models describe the same quantity over a more complete energy range. In this sense, these physical quantities, both experimental and theoretical, are commonly termed “differential” information. This definition is often contrasted with the term “integral.” In integral experiments, data are obtained as an average over a spectrum and often include the effects of several isotopes and materi-

als. These data are in general representative of a particular application area of nuclear physics covering fundamental science, nuclear astrophysics, and other applications in nuclear technology, such as reactor physics, criticality safety, neutron dosimetry, etc.

The simplest example illustrating these two concepts of differential and integral data is the spectrum-averaged cross section,  $\bar{\sigma}$ , measured in a well-defined neutron spectrum,  $w(E)$ ,

$$\bar{\sigma} = \int \sigma(E)w(E)dE. \quad (1)$$

The data resulting from differential information, termed nuclear data, appear in this equation as  $\sigma(E)$  as a representative example;  $\sigma(E)$  is a particular observable (for instance, the <sup>239</sup>Pu inelastic cross section) at a certain incident or outgoing energy or angle. They serve as input for various calculations of  $\bar{\sigma}$  in the application areas mentioned above. When experimental integral data of a specific application area are

\*dneudecker@lanl.gov

available, e.g., experimental data on  $\bar{\sigma}$ , these experimental data provide further information on validating the differential data by simulating the experiment with Eq. (1) and comparing to the associated experimental value [1]. However, one could not obtain  $\sigma(E)$  by only optimizing Eq. (1) to measured data of  $\bar{\sigma}$  as  $\sigma(E)$  has many degrees of freedom due to the dependence of  $\sigma$  on  $E$ . Hence, while the reliability of numerical data can be tested through simulating integral experiments and comparing to their experimental values, the numerical data are only validated for the application domain represented by integral experiments.

This differential and integral information is rarely brought together in a systematic manner to inform nuclear physics. However, combining them could alleviate issues encountered in both differential and integral data. Differential information is often too uncertain to accurately define nuclear-physics observables. Most angle-integrated cross sections cannot be measured to more than 1% given the corrections necessary to obtain them from actually measured data [2,3]. Measurements of the average prompt-fission neutron multiplicity, which are arguably among the most precise ones providing input for nuclear data, report a precision as low as 0.5%, but these low uncertainty values are extremely difficult to achieve [2,4,5]. Knowledge gained from nuclear theories complements differential data where experiments are inaccessible. However, this theory knowledge cannot define nuclear-physics observables with lower uncertainties than precise experimental data given the possible variations in parameter space and approximations made in the models. These physical variations in parameter space are often so large that the model is unable to clearly identify which differential experimental data set is closest to nature if several of them differ systematically by a few percent.

In contrast, some integral experiments can be measured to a much higher precision. For instance, some criticality experiments [6] that provide the effective neutron multiplication factor,  $k_{\text{eff}}$ , are reported with realistic uncertainties as low as 0.05%. However, these integral experiments only provide convolved information on many differential data. This confounding lies in the fact that the integral data cannot identify a set of differential data uniquely as shown with the example in Eq. (1). If one aims to identify issues in the differential data by taking advantage of integral information, one has to explore an extremely large parameter space that includes all materials and reactions taking place in the experiment to simulate it. In the simulation of  $k_{\text{eff}}$ , all neutron-scattering and fission nuclear data enter—adding easily up to several hundreds of nuclear-data values being used for the simulations. To give a particular example, in Ref. [7], 875 integral experiments were simulated as a whole set with approximately 20 000 nuclear-data values, i.e., a  $875 \times 20\,000 \approx 18$  million dimensional sensitivity space. It is challenging for a human mind on its own to find a trend in such a high-dimensional space. Therefore, our aim is to apply a machine-learning (ML) technique to excavate issues and inconsistencies that reside between the differential and integral information, which should be more efficient than conventional methods.

A variety of ML techniques have been applied recently to solve open issues in nuclear physics. Gaussian processes

have been used to estimate theory-truncation errors for nuclear forces [8], to systematically improve theory calculations compared to experimental data [9], and as efficient emulators for complex, computationally expensive physics models [10]. Support vector regression, elastic net, and random forest have been used to extract potential physics causes for experimental data being outliers based on measurement information [11]. Artificial neural networks have been successfully applied to parametrize nuclear wave functions for  $A \leq 4$  nuclei [12,13]. Bayesian neural networks were applied to mass models [9], to astrophysical applications [14,15], and to determining the limits of the neutron and proton driplines [16]. Bayesian neural networks and the probabilistic mixture density networks have been used to provide insight on fission yields [17,18]. In Ref. [7], we applied the random forest and Shapley additive explanations (SHAP) metric to identify groups of correlated nuclear data that are likely describing a nuclear-physics observable imperfectly. We also discussed that one cannot unambiguously disentangle which of the several nuclear-physics observables is biased just with integral input as explained at the beginning of the introduction. Information from differential experiments was already highlighted as potentially being able to aid in solving this confounding.

Here, we demonstrate how one can combine knowledge from differential and integral regimes to partially disentangle this convolved information and inform more clearly nuclear physics as applied to a specific application area. The resulting information can also provide valuable input on what future measurements and theory developments could unravel the confounding between different nuclear-physics observables as well as theoretical predictions. In addition, we emphasize that the techniques explored are so general that they not only reconcile differential data with integral information but may also be applicable to nuclear-physics problems where entangled model parameters concentrate into a few degrees of freedom. An example could be the prompt-fission neutron multiplicity,  $\bar{\nu}_p$ , or the delayed one,  $\bar{\nu}_d$ , which can be calculated by a variety of nuclear-reaction and structure models [19–22]. We could achieve a much higher accuracy on predicting  $\bar{\nu}$  with these models through the ML techniques presented here, and we envisage ML will become a common tool used in extracting physics information out of heterogeneous data.

Section II describes the differential and integral input data used for this analysis along with the analysis methods employed. Section III shows some examples on how this analysis can be applied to resolve issues in understanding differential data or highlight the need for future measurements and theory developments. Such recommendations are briefly summarized for Pu-isotope nuclear data in Sec. III as applied to the application areas represented by criticality and pulsed-sphere neutron-spectra experiments. While we focus on the application area of criticality safety and shielding in this particular paper, it is highlighted in the conclusions and outlook, Sec. IV, that this technique can be applied to other areas of nuclear physics.

## II. METHODS AND DATA

Figure 1 illustrates how the analysis was undertaken to bring together integral and differential information for

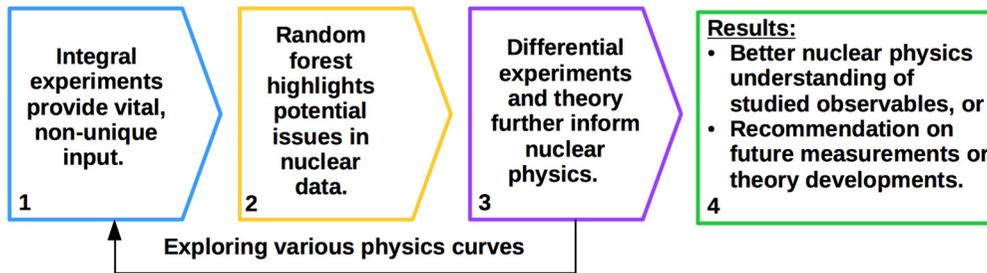


FIG. 1. Schematic figure of the analysis pathway that informs nuclear physics with differential and integral information.

improving our description of nuclear-physics observables represented by nuclear data. The following steps were undertaken in this analysis:

- (1) In a first step, integral experiments are selected that will be used to investigate the reliability of nuclear data in step 2. The input, sensitivities of integral quantities to nuclear data, that allows to link integral experiments to these data is described in Sec. II A.
- (2) In the second step, the validity of nuclear data is studied with respect to integral experiments selected in step 1 (experiments providing criticality and pulsed-sphere neutron-leakage spectra) employing the random forest algorithm and SHAP metric described in Sec. II B. The output of these machine-learning algorithms is shown in Fig. 2; namely, listings of the most important nuclear data related to bias between simulated values and experimental data of integral measurements. The higher a SHAP value is for a particular observable, the more it is likely to contribute to bias. We studied here those Pu-isotope nuclear data that appear with SHAP values among the highest 3 000 ones (out of approximately 20 000).
- (3) The nuclear data thus identified as potentially inaccurate are then compared to differential experimental data and assessed by means of theoretical considerations as showcased in Sec. III. If one cannot unambiguously define a more accurate description of a nuclear-physics observable with differential information on its own, the feedback loop shown in Fig. 1 is executed. This loop explores the potential space of solutions. Such a situation could arise, for instance, if there are several discrepant differential data sets, and one cannot decide which one is closer to truth. To this end, several nuclear-data curves are established that span the space of possible nuclear data given differential experimental data and theory considerations. These curves are then used to simulate integral experiments. In some cases, the agreement between simulated values and experimental data can help decide on the curve closest to nature.
- (4) The results of this analysis are in the best case a better understanding of the observable by differential and/or integral information. If no decisive information can be gleaned from step 3, recommendations can be formulated on which nuclear-physics observables should be further investigated, either experimentally or with better modeling. These recommended experiments and theory developments would not only improve our understanding of differential data but would also better constrain the nuclear-physics space for the application area represented by the integral experiments.

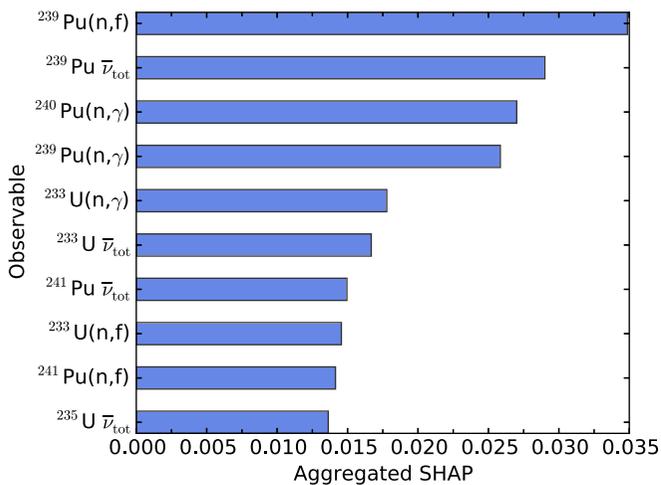


FIG. 2. The 10 most important nuclear-physics observables related to bias in predicting 875  $k_{\text{eff}}$  values and 15 LLNL pulsed-sphere neutron-leakage spectra are shown using ENDF/B-VIII.0 nuclear data. The SHAP values shown here are aggregated by observable [7]; i.e., no dependence on energy is shown.

## A. Input data

### 1. Integral experiments

Two types of integral experiments are considered: 875 criticality benchmarks documented in the International Handbook of Evaluated Criticality Safety Benchmark Experiments (ICS-BEP) [6] and 15 Lawrence Livermore National Laboratory (LLNL) pulsed-sphere measurements [23], the latter consisting of a total of 2 343 data points.

Criticality benchmarks provide the effective multiplication factor,  $k_{\text{eff}}$ , i.e., the ratio of the number of neutrons in a system in one generation<sup>1</sup> to the number of neutrons in the next one. One  $k_{\text{eff}}$  value is given per assembly configuration and can

<sup>1</sup>A neutron generation relates to the progeny of preceding neutrons in a fission chain.

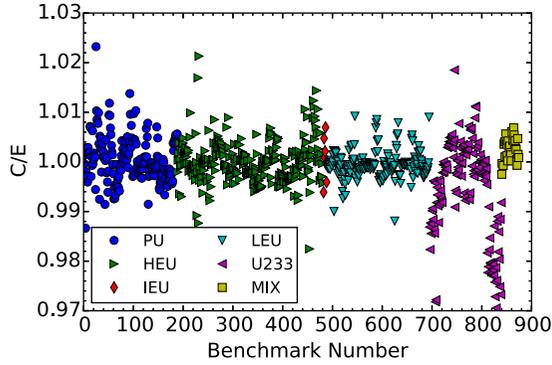


FIG. 3. Simulated  $k_{\text{eff}}$  values ( $C$ ) are shown in ratio to experimental data ( $E$ ) for all criticality benchmarks used as part of the current analysis. The different colors distinguish between different subclasses of experiments identified by their fuel type following an abbreviated ICSBEP nomenclature.

be determined to a precision as low as 0.05%; although  $k_{\text{eff}}$  uncertainties in the range of 0.1–0.4% are reported for many assemblies. Thousands of these experiments are assembled, after a stringent review, in the ICSBEP handbook [6]. In these experiments, there is always a fuel present, consisting of major and minor actinides, while light elements (e.g.,  $^1\text{H}$ ,  $^{14}\text{N}$ ) and structural materials (e.g.,  $^{56}\text{Fe}$ ,  $^{208}\text{Pb}$ ) may appear within the fuel, moderator, or reflector material. Here, we use  $k_{\text{eff}}$  values of 875 ICSBEP critical assemblies as detailed in Table I of Ref. [7] and shown in Fig. 3.

In LLNL pulsed-sphere experiments [23], neutron-leakage spectra are measured, i.e., the time-of-flight spectra of all neutrons emitted from the investigated material; these neutrons are mainly produced through scattering or fission reactions (if an actinide is present). Simple spheres of various materials containing only few distinct isotopes are studied in this measurement series; the spheres are pulsed by a 14-MeV neutron source produced by a deuteron beam hitting a tritiated target in the center of the sphere. The resulting neutron-leakage spectra are measured at a specific angle and in some cases for different thicknesses of the same material. An example of such a neutron-leakage spectra is shown in Fig. 4 for a  $^{238}\text{U}$  sphere. Contrary to criticality measurements, this class of experiments studies fuel materials independently from combinations of light and structural materials. Here, we use 15 pulsed spheres as detailed in Table 2 of Ref. [24].

## 2. Differential experiments

For the current study, we retrieved differential experimental data from the EXFOR database [25] for  $^{238-242}\text{Pu}$  neutron-induced total ( $n,\text{tot}$ ), elastic ( $n,\text{el}$ ), inelastic ( $n,\text{inl}$ ), capture ( $n,\gamma$ ), fission ( $n,\text{f}$ ), and ( $n,2n$ ) cross sections along with prompt-fission neutron spectra (PFNS) and average prompt-fission neutron multiplicities ( $\bar{\nu}_p$ ). The main difference between differential and integral experiments is that usually one studies in the former case one specific observable of one reaction per experiment using small amounts of materials consisting of few isotopes, e.g., the ( $n,\text{f}$ ) cross section of  $^{241}\text{Pu}$ .

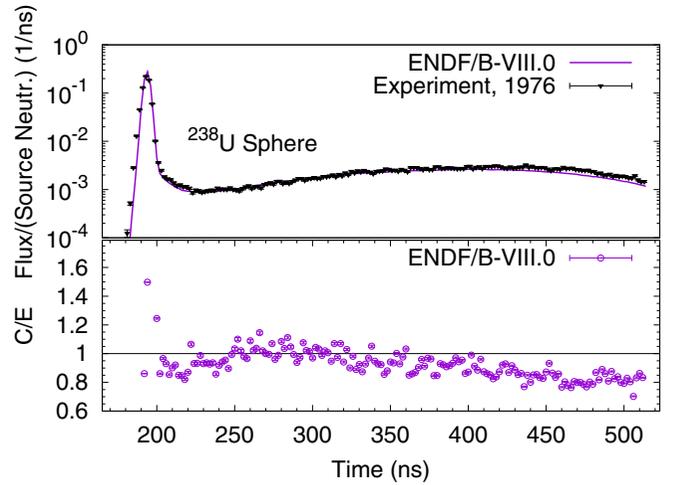


FIG. 4. A  $^{238}\text{U}$  LLNL pulsed-sphere neutron-leakage spectrum induced by neutrons of 14 MeV is shown as an example. Experimental data are given in black as measured with a sphere of 0.8 mean-free path, a NE-213A detector, and at an angle of  $117^\circ$  with respect to the deuteron beam, while simulated values are shown in violet for ENDF/B-VIII.0. Calculated versus experimental values,  $C/E$ , are shown in the lower panel.

## 3. Linking integral experiments to nuclear data

While it is clear that differential experimental and theoretical information enter nuclear data as direct input for their evaluation of the same observable and energy range, the connection of a specific nuclear-data value to simulations of integral experiments is more complicated.

For instance, one simulates  $k_{\text{eff}}$  of a criticality experiment by solving the time-independent Boltzmann forward neutron-transport equation,

$$\begin{aligned} & \hat{\Omega} \cdot \nabla \psi(\vec{r}, E, \hat{\Omega}) + \Sigma_t(\vec{r}, E) \psi(\vec{r}, E, \hat{\Omega}) \\ &= \int_0^\infty dE' \int_{4\pi} d\hat{\Omega}' \Sigma_s(\vec{r}, E' \rightarrow E, \hat{\Omega}' \cdot \hat{\Omega}) \psi(\vec{r}, E', \hat{\Omega}') \\ &+ \frac{1}{k} \frac{1}{4\pi} \int_0^\infty dE' \int_{4\pi} d\hat{\Omega}' \bar{\nu}_{\text{tot}} \Sigma_f(\vec{r}, E') \\ &\times \chi_{\text{tot}}(\vec{r}, E', E) \psi(\vec{r}, E', \hat{\Omega}'), \end{aligned} \quad (2)$$

for the largest eigenvalue of  $k$ . One then compares the simulated  $k_{\text{eff}}$  value to the experimental one. The variable  $\psi(\vec{r}, E, \hat{\Omega})$  is the angular flux in the assembly as a function of location in the assembly,  $\vec{r}$ , outgoing energy,  $E$ , and outgoing direction,  $\hat{\Omega}$ . The equation balances neutrons lost through the streaming and collision terms on the left-hand side versus the neutrons produced via in-scatter and fission on the right-hand side. Probabilities of nuclear processes (e.g., fission and scattering) are represented by nuclear data in this equation. For instance, neutrons produced by fission are described by the average total-fission neutron multiplicity,  $\bar{\nu}_{\text{tot}}$ , the total-fission neutron spectrum,  $\chi_{\text{tot}}$ , and the macroscopic total fission cross section,  $\Sigma_f$ .  $\Sigma_s$  and  $\Sigma_t$  are the macroscopic scattering and total cross sections. All of these observables consider the isotopes present in and included in the simulation of the assembly; e.g., they are the cross sections,  $\chi_{\text{tot}}$ , etc., for a

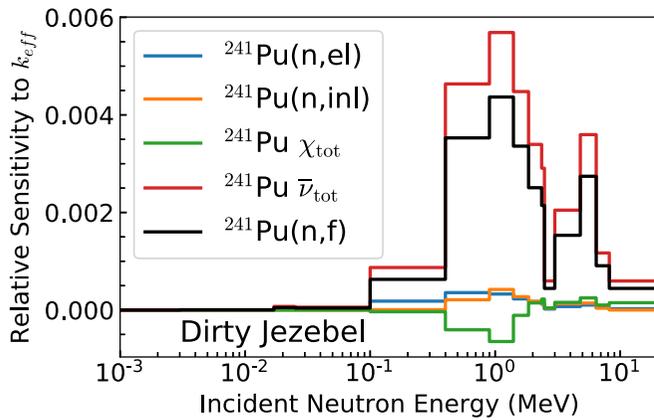


FIG. 5. The sensitivities of Dirty Jezebel’s  $k_{\text{eff}}$  to various nuclear data are shown (sensitivities to  $\chi_{\text{tot}}$  are constrained).

material that may contain  $^{239}\text{Pu}$  but also at the same time  $^{240}\text{Pu}$  and  $^{241}\text{Pu}$ . The Monte Carlo technique is often employed to numerically solve the Boltzmann equation and such simulations are usually undertaken with neutron-transport codes; we employed here MCNP-6.2 [26].

One way to understand how various nuclear data influence the simulation of, for instance,  $k_{\text{eff}}$  of a particular integral experiment is by calculating sensitivity profiles. To be more specific, first-order sensitivity coefficients  $S_{ij}$  are numerically approximated by [24]

$$S_{ij} = \frac{\Delta C_i / C_i}{\Delta \sigma_j / \sigma_j}, \quad (3)$$

with the relative change of a simulated value,  $\Delta C_i / C_i$ , induced by a relative change,  $\Delta \sigma_j / \sigma_j$ , of the nuclear data for energy group  $j$  of 50 groups spanning from thermal to 20 MeV. The calculated values,  $C_i$ , could correspond to a single  $k_{\text{eff}}$  value for one ICSBEP assembly, leading to  $S_{ij}$  turning into the vector  $S_j$  for one criticality experiment, or could be a function of time-of-flight bins  $i$  and nuclear-data energy bins  $j$ .

When one studies the sensitivity of  $k_{\text{eff}}$  of the “Dirty Jezebel” critical assembly (listed with the identifier PU-MET-FAST-002 in Ref. [6]) to various  $^{241}\text{Pu}$  nuclear data in Fig. 5, one can clearly understand that the one simulated  $k_{\text{eff}}$  value is sensitive to several hundreds of nuclear-data values—especially considering that many more isotopes than  $^{241}\text{Pu}$  are present and the sensitivity to  $^{241}\text{Pu}$  nuclear data is by a factor of 50 smaller than to  $^{239}\text{Pu}$  data. The same is true for simulating pulsed-sphere neutron-leakage spectra. In addition to that, the relative sensitivity of  $k_{\text{eff}}$  to  $\bar{v}$  versus the  $(n,f)$  cross section as a function of incident-neutron energy is very similar across all critical assemblies, tied to Eq. (2), leading to the fact that one cannot disentangle bias coming, e.g., from the  $(n,f)$  cross section or  $\bar{v}$  only via critical assemblies. To this end, one would need an integral observable with different sensitivities to those two nuclear-data observables in the same energy range. While sensitivities of pulsed-sphere neutron-leakage to  $(n,f)$  cross sections and, e.g.,  $(n,\text{el})$  cross sections are different compared to those for  $k_{\text{eff}}$ , the pulsed spheres conclusively query nuclear data from 15 down to 5 MeV, while  $k_{\text{eff}}$  is mostly sensitive to nuclear data up to 5 MeV.

Hence, one cannot uniquely trace back a shortcoming in simulating  $k_{\text{eff}}$  or pulsed-sphere neutron-leakage spectra to one specific nuclear-data value. On the contrary, there is a one-to-one correspondence between differential information and nuclear data, allowing for a more unique identification of biased nuclear data.

## B. Analysis methods

As mentioned above, several hundreds to thousands of nuclear data may enter the simulation of one integral experiment, dependent on the makeup of the assembly. Here, we simulate 890 experiments with approximately 20 000 nuclear data from ENDF/B-VIII.0. About a fourth of these experiments have plutonium in their core, which we study in detail.

The following question is asked as part of the second step in the pathway in Fig. 1: What specific nuclear-data values,  $\sigma_j$ , are related to a difference between simulated integral values,  $C$ , and the associated experimental data,  $E$ ? To cast this into a machine-learning algorithm, one builds a prediction model for the bias  $\Delta = (E - C) / \sqrt{\delta E^2 + \delta C^2}$  as a nonlinear function of all potentially informative features ( $S_{i,j}$  and 1 000 measurement features described below):

$$\Delta = \frac{E - C}{\sqrt{\delta E^2 + \delta C^2}} = f(X_1, \dots, X_{21000}). \quad (4)$$

Contrary to Ref. [7],  $C$  and  $E$  do not only encompass  $k_{\text{eff}}$  values but also pulsed-sphere neutron-leakage-spectra values. The bias is weighted with the experimental variances,  $\delta E^2$ , and the calculated ones,  $\delta C^2$ ; the latter account for only Monte Carlo statistics uncertainties from the neutron-transport code. This weighting was undertaken as pulsed-sphere neutron-leakage spectra are distinctly (by a factor of 5–200) more uncertain than experimental criticality values. All of them are assumed to be independent given the lack of information on possible correlations. The features that should be explored here are those nuclear data that lead to bias between  $C$  and  $E$  of integral experiments. Hence, one has to use for feature,  $X$ , values that encompass the information on how  $C$  depends on nuclear data,  $\sigma$ ; that information is represented by a total of 20 000 sensitivities,  $S_{ij}$ , of Eq. (3). Another class of features that is added, in addition to  $S_{ij}$ , are about 1 000 measurement features describing the experiments (see Ref. [7]) to account for the fact that some biases might have a trend related to appearing with a specific measurement feature. Measurement features could be, for instance, the core geometry (sphere, cylinder, etc.), number of fission units, or percent weight of actinides.

The function  $f(X_1, \dots, X_{21000})$  is modeled with the random forest [27]. As integral input (step 1 in Fig. 1), one uses  $E$  of the integral experiments described in Sec. II A,  $C$  calculated with the neutron-transport codes MCNP-6.2 [26] and ENDF/B-VIII.0 [28], and  $S_{ij}$  calculated with the codes MCNP-6.2, FRENDRY, and SANDY [26,29,30] for both criticality and pulsed spheres responses [24].

The random forest algorithm then randomizes decision trees, as its name suggests, by entering in a set of trees each a subset of biases  $\Delta$  and a subset of features  $X$ . Each tree partitions  $\Delta$  then in several subsequent splits such that one

part of the data has the feature  $X_i$  present while the other part does not depend on  $X_i$ . One could imagine that one part of the experiment depends on the  $^{241}\text{Pu}(n,f)$  cross section from 0.4–0.9 MeV while the other part does not (i.e., experiments with and without this isotope present). The first split is selected by the algorithm such that it maximally reduces the spread in  $\Delta$  by splitting with respect to a particular  $X$ . Subsequent splits apply the same principle with diminishing returns on reducing the spread in  $\Delta$ . Randomizing over several decision trees guarantees that not one feature  $X$  dominates as the main reason to split but that one can explore several important factors.

An importance metric, here the SHAP metric [31,32], is applied to the resulting random forest to extract which features were most important in predicting  $\Delta$ . The SHAP metric measures importance of a feature by comparing the incremental predictive improvement obtained by adding the feature to a subset of all other predictive features. SHAP averages these incremental improvements over all subsets of other features and over all observed data points, resulting in an aggregate measure of importance for a given feature. Figure 2 shows SHAP values of only the top 10 most important reactions from approximately 400 in the study (all 20 000 nuclear-data features were aggregated by reaction), but also SHAP values for individual energy groups of specific reactions are used for the analysis below. The SHAP metric as applied to this problem was described in more detail in Refs. [7,33]. It is important to know that the higher a SHAP value is for a particular feature  $X$ , the more  $X$  is indicated to be related to bias. Such results are for instance shown in Fig. 2. The results serve then as input for step 2 of Fig. 1, namely to identify which nuclear data are possibly incorrect. These particular nuclear data are then plotted in step 3 of Fig. 1 with respect to differential data and analyzed including theory understanding to discern if there is truly a shortcoming in the nuclear data.

It is worth noting that the ML algorithms are not being used to solve an inverse problem here, in the sense that they do not change input nuclear data to make the resulting simulations of integral experiments consistent with their experimental counterparts. Instead, the bias between simulated and experimental values is observed given ENDF/B-VIII.0 nuclear data; the ML algorithms then identify systematic relationships between the observed bias and nuclear data sensitivities, which serves as input for experts to identify nuclear data for potential adjustment. We manually explore possible changes in nuclear data rather than blindly following ML results because these data-driven algorithms cannot pinpoint individual predictors of bias due to correlation between nuclear-data sensitivities resulting from physics.

### III. RESULTS AND DISCUSSION

#### A. $^{241}\text{Pu}(n,f)$ cross section from 0.1–2 MeV

Several  $^{241}\text{Pu}$  nuclear data appear as strongly related to bias from 0.1–2.354 MeV. That is, several  $^{241}\text{Pu}$  nuclear data in that particular energy range were assigned SHAP values that place them among the first 0.1% of most important values related to  $\Delta$  across all isotopes, reactions, and energy ranges.

Among them were the  $(n,f)$  cross section, the PFNS, and  $\bar{v}_{\text{tot}}$ . The  $(n,f)$  cross section and  $\bar{v}_{\text{tot}}$  appear even among the seven most important observables in Fig. 2 despite the fact that  $^{241}\text{Pu}$  appears to only low percentages in critical assemblies. All three fission observables contribute to the fission-source term of  $k_{\text{eff}}$  in Eq. (2). Given that, one cannot disentangle the effect those three observables have on simulating  $k_{\text{eff}}$  by only studying these integral experiments. In addition to that, there seem to be issues in ENDF/B-VIII.0 nuclear data representing the PFNS,  $\bar{v}_{\text{tot}}$ , and the fission cross section of  $^{241}\text{Pu}$ . These issues are discussed below in the context of theory considerations and by comparing existing differential experimental data to nuclear data in the energy range highlighted by the random forest and SHAP as related to  $\Delta$ .

The PFNS,  $\chi(E_{\text{out}})$ , is often described by a Maxwellian-like shape

$$\chi(E_{\text{out}}) \sim \sqrt{E_{\text{out}}} \exp\left(-\frac{E_{\text{out}}}{T}\right), \quad (5)$$

at low outgoing-neutron energies,  $E_{\text{out}}$ , in the laboratory frame, using the parameter  $T$  conventionally termed the Maxwellian temperature. This was inferred by transforming the spectrum of neutrons evaporating from moving fission fragments into the laboratory system [34–36]. The ENDF/B-VIII.0  $^{241}\text{Pu}$  file contains only three values of  $T$  to represent PFNS for three different incident-neutron energy ranges,  $E$ . However, as noted by Watt [37], the PFNS behaves like an evaporation spectrum,  $E_{\text{out}} \exp(-E_{\text{out}}/T)$ , at higher  $E_{\text{out}}$ ; hence, the Maxwellian representation chosen for ENDF/B-VIII.0  $^{241}\text{Pu}$  PFNS poorly approximates the true shape of a PFNS [38].

The ENDF/B-VIII.0 fission cross section,  $\sigma_f$ , goes through the cloud of differential experimental data and approximates it as best as it can. However, it is obvious from Fig. 6 that these differential data are highly discrepant. Note that the  $^{241}\text{Pu}(n,f)$   $\sigma_f$  is determined from absolute measurements of the cross section (Butler, Simpson, and Szabo data [39–41]), and also by ratio data to other reactions [42] such as  $^{235}\text{U}(n,f)$  (Carlson, Fursov, Kaeppler, White, and Smith data [43–47] in Fig. 6). The ratio data of Tovesson *et al.* data [48] were not included in the analysis underlying ENDF/B-VIII.0 because the reasons for its difference to other data were not understood.

Theoretical prediction of the  $^{241}\text{Pu}(n,f)$  cross section is one powerful tool to estimate variations of  $\sigma_f$  as a function of incident-neutron energy. However, it is known that the predicted absolute magnitude as well as the exact shape may be too uncertain, or too sensitive to model parameters such as the fission barriers, to eliminate any of the inconsistent data sets. In fact, current modeling of the fission channel in the statistical Hauser-Feshbach theory often employs a very crude Wentzel-Kramers-Brillouin (WKB) approximation to calculate the penetration probability through the fission barriers, together with phenomenological models for fission level-density enhancement. It can be seen from Fig. 7 that a variation of only 150 keV in the  $^{241}\text{Pu}$  fission barrier, well within the uncertainties of this parameter, encloses most differential data. Therefore, one cannot exclude most of the discrepant  $^{241}\text{Pu}(n,f)$   $\sigma_f$  differential data solely by theory considerations; that also means that the  $^{241}\text{Pu}$   $\sigma_f$

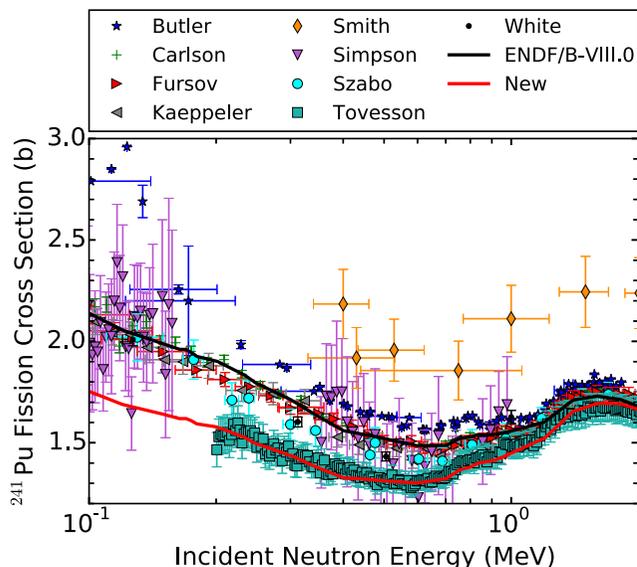


FIG. 6. Experimental  $^{241}\text{Pu}(n,f)$  cross sections [39–41,43–47] are shown in comparison to ENDF/B-VIII.0 (black line) and the new curve tested here (red line). This plot shows the considerable spread between the differently colored data set—that is distinctly more than their reported total uncertainties.

remains as a possible source of bias given differential information.

However, one can exclude the PFNS and  $\bar{\nu}_{\text{tot}}$  as the main driver of bias,  $\Delta$ , in simulations of integral experiments by exploring different shapes of the observable that are supported by both differential experiments and theory. While the  $^{241}\text{Pu}$  PFNS clearly needs to be improved by using a more physical shape, using such a PFNS (taken from the JENDL-4.0

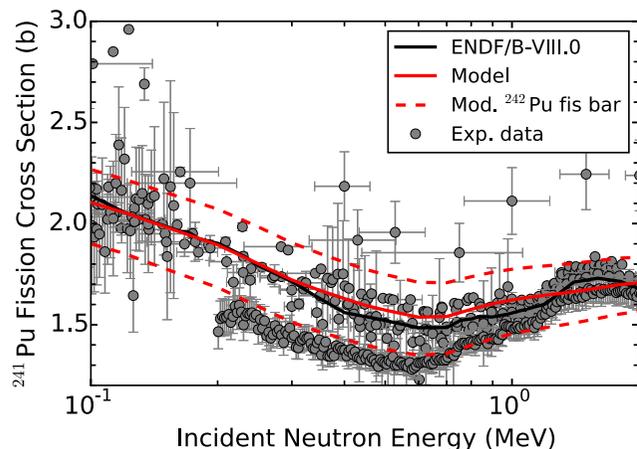


FIG. 7. The same experimental data as shown in Fig. 6 for the  $^{241}\text{Pu}(n,f)$  cross section are compared to ENDF/B-VIII.0 (black line) and model curves. The mean model curve (red, solid line) was obtained with model parameters optimized to reproduce ENDF/B-VIII.0 data. It is also shown that only a small change in the fission barrier of 150 keV suffices to enclose most discrepant experimental data.

[49] nuclear-data library, which is based on the Madland-Nix model [36]) changes  $C$  very little. The reason for that can be seen from Fig. 5. The Dirty Jezebel assembly is among the most sensitive to  $^{241}\text{Pu}$  data but is little sensitive to the  $^{241}\text{Pu}$  total-fission neutron spectrum,  $\chi_{\text{tot}}$ , which mainly consists of the PFNS. The mean energy of the PFNS changes by 1% from ENDF/B-VIII.0 to the physical shape of JENDL-4.0—a change much too small to strongly improve  $\Delta$  of criticality measurements. For instance,  $k_{\text{eff}}$  of Dirty Jezebel changes only by negligible 18 pcm (per cent mille) if one simulates  $k_{\text{eff}}$  with JENDL-4.0  $^{241}\text{Pu}$  PFNS versus ENDF/B-VIII.0. To give a qualitative measure of pcm, the  $k_{\text{eff}}$  of Dirty Jezebel is reported with 200 pcm uncertainty, with approximately 200 pcm being also the difference between a controlled assembly and a criticality accident for this particular configuration.

The same argument applies to  $\bar{\nu}_{\text{tot}}$ . While Dirty Jezebel's  $k_{\text{eff}}$  is distinctly more sensitive to  $\bar{\nu}_{\text{tot}}$ , one cannot change it significantly without disagreeing systematically from existing differential experiments given that ENDF/B-VIII.0  $\bar{\nu}_{\text{tot}}$  already agrees well with them except for the fact that the energy grid was chosen too coarse to capture experiments fully. Hence, only fission cross sections remain as possibly related to  $\Delta$ .

If one changes the fission cross section such that it corresponds to the experimental data of Tovesson *et al.* [48] (red line in Fig. 6) and simulates  $k_{\text{eff}}$  with it, the simulated  $C$  changes significantly, by 143 pcm. Moreover,  $E - C$  changes from a bias of 147 to 4 pcm with the  $^{241}\text{Pu}(n,f)$   $\sigma_f$  following Tovesson *et al.*

Of course, one critical assembly is not a good measure to test whether new evaluated data improve  $\Delta$  as  $k_{\text{eff}}$  is a severely underdetermined system and by introducing a mistake in one observable, one might mask it for another observable. Hence, we tested the effect of changing  $^{241}\text{Pu}$  nuclear data to be closer to Tovesson *et al.* data [48] on  $\Delta$  for many assemblies containing  $^{241}\text{Pu}$ . To this end, the change in  $\sigma_f$  was linearly propagated to  $k_{\text{eff}}$  of 50 critical assemblies where a change in the  $^{241}\text{Pu}$   $\sigma_f$  leads to a change in  $k_{\text{eff}}$  by more than 10 pcm. Linearity holds as the total change in simulated criticality is expected to be small despite the large change in  $^{241}\text{Pu}$   $\sigma_f$  given the overall low sensitivity. Lowering  $^{241}\text{Pu}$   $\sigma_f$  decreases the calculated  $k_{\text{eff}}$  for all these benchmarks; i.e., our proposed change will improve the  $C/E$  for overcalculated benchmarks and make it worse for undercalculated ones. The former are in majority so there is a slight overall improvement: The bias averaged over all the benchmarks studied,  $\langle \Delta \rangle = \langle E - C \rangle$ , is 181 pcm with ENDF/B-VIII.0  $^{241}\text{Pu}(n,f)$   $\sigma$  while it is 150 pcm for the new nuclear data. This improvement is statistically insignificant given measurement uncertainties typically in the range of 50–500 pcm. Out of all these benchmarks, Dirty Jezebel is the one most sensitive to  $^{241}\text{Pu}$  nuclear data. Hence, while we significantly improve  $k_{\text{eff}}$  simulations of Dirty Jezebel by using a  $\sigma_f$  closer to Tovesson *et al.* data, we preserve (actually slightly improve) good performance of simulating most other critical assemblies, some of which are among the primary test cases for any new  $^{239}\text{Pu}$  nuclear data.

It should be emphasized that random forest and SHAP informed by integral experiments and simulations was only able to identify a group of nuclear data as related to bias, namely,

the  $^{241}\text{Pu}$  fission cross section, PFNS and  $\bar{\nu}_{\text{tot}}$ . The reason for this correlation lies in the fact that the fission-source term in Eq. (2) combines these three observables always in the same way; that renders it impossible to disentangle their individual contribution to bias. One needs additional information, such as provided by differential experiments and theory, to find out which of them is likely incorrect. This information, combined with the sensitivities  $S_{i,j}$ , allow us to estimate whether a change of an observable within differential uncertainties is likely to yield a substantial improvement in  $C/E$ . In the case of  $^{241}\text{Pu}$ , examination of the differential data for the fission cross section, PFNS, and  $\bar{\nu}_{\text{tot}}$  led us to the conclusion that the  $^{241}\text{Pu}(n,f)$  cross section is most likely the major source of bias among the three. Integral information, on the other hand, allowed us to select from several discrepant differential measurements the one yielding improved  $C/E$  values, which would not be possible given only differential experimental and theoretical information. Hence, the combination of both differential and integral information, supported by ML and constrained by theory was the key to improving our understanding of nuclear-physics observables.

### B. $^{239}\text{Pu}(n,f)$ cross section from 10–15 MeV

The  $^{239}\text{Pu}$   $\sigma_f$ ,  $\chi$ ,  $\bar{\nu}_{\text{tot}}$ , and  $(n,\text{el})$  cross section,  $\sigma_{\text{el}}$ , from 10–15 MeV are all identified by the random forest and SHAP metric as among the 5–10% most important observables related to bias.  $^{239}\text{Pu}$   $\sigma_f$  and  $\bar{\nu}$  appear even as the top two most important observables related to bias in Fig. 2. Simulated  $k_{\text{eff}}$  values are sensitive to nuclear data up to maximally 6 MeV. Hence, nuclear data above 6 MeV are validated here only by LLNL pulsed-sphere neutron-leakage spectra.

Similar to the example above, one cannot disentangle shortcomings in scattering ( $\sigma_{\text{el}}$ , discrete and continuum inelastic scattering cross sections,  $\sigma_{\text{inl}}$ , and angular distributions) and fission observables ( $\sigma_f$ ,  $\chi$ ,  $\bar{\nu}_{\text{tot}}$ ) from each other solely by comparing  $E$  and  $C$  of LLNL pulsed-sphere experiments. Differential information, either from theory or experiment, is needed to identify which nuclear data are likely incorrect.

Unfortunately, resolving this confounding is more difficult than for the  $^{241}\text{Pu}$  case. For instance, no  $^{239}\text{Pu}$   $\sigma_{\text{el}}$  and  $\sigma_{\text{inl}}$  were found in EXFOR for  $E = 10\text{--}15$  MeV. The total cross section, a sum of all open reaction channels, cannot provide conclusive information as too many of its components are experimentally undefined ( $\sigma_{\text{el}}$  and discrete and continuum  $\sigma_{\text{inl}}$ ) or discrepancies are observed for others. Hence,  $^{239}\text{Pu}$   $\sigma_{\text{el}}$  and discrete and continuum  $\sigma_{\text{inl}}$  are purely defined by model calculations that themselves are uncertain. While optical or coupled-channels models are expected to grasp adequately essential physics of nucleon scattering, they depend on phenomenological optical model potentials that are derived from fitting to experimental data. If those are missing, global parametrization or extrapolation of potentials for other nuclei must be used. In such a case, the reliability of model calculations is considerably reduced. Even if the often existing experimental total cross sections allow us to constrain optical model parameters, the lack of other reaction cross sections and angular distributions prevents proper partition of the total cross section between

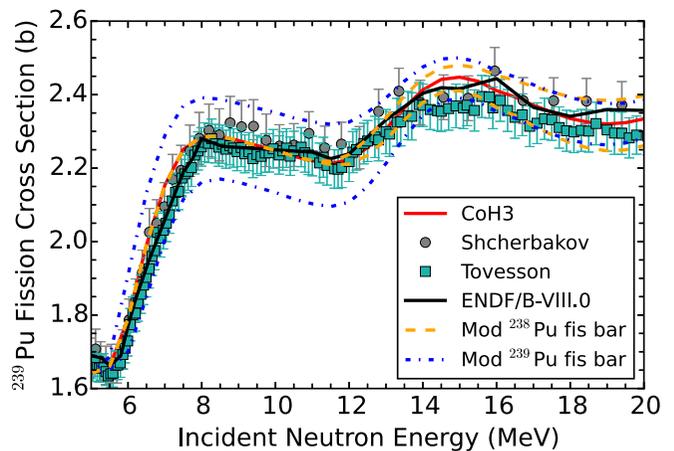


FIG. 8. The discrepancies between Shcherbakov [50] and Tovesson [48]  $^{239}\text{Pu}$   $\sigma_f$  are highlighted. The data are also compared to ENDF/B-VIII.0 (black line) and model curves. The mean model curve (red, solid line) was obtained with model parameters optimized to reproduce ENDF/B-VIII.0 data. It is also shown that only a small change in the  $^{239}\text{Pu}$  (blue, dash-dotted line) and  $^{238}\text{Pu}$  (orange, dashed line) fission barriers of  $\pm 150$  keV suffices to enclose most discrepant experimental data from 6 to 20 MeV.

shape elastic and absorption. This limits the model's predictive power when calculating individual reactions.

Also, sizable discrepancies are again observed between differential  $^{239}\text{Pu}$   $\sigma_f$ . The data of Tovesson *et al.* and Shcherbakov *et al.* [48,50] differ by up to 5% and span between them the space of possible values for the  $^{239}\text{Pu}$   $\sigma_f$ . Several other differential data sets exist [51–53] that go through the middle between those two data sets but are not shown in Fig. 8 for the sake of better visibility. Uncertainties in theoretically predicted  $\sigma_f$  are often larger than the differences observed between Tovesson *et al.* and Shcherbakov *et al.* data. To demonstrate this, we employ statistical Hauser-Feshbach model calculations to model the  $^{239}\text{Pu}(n,f)$  reaction and investigate how much the calculated  $\sigma_f$  vary based on empirical uncertainties on the fission-barrier parameters.

The fission cross-section calculation was carried out with the statistical Hauser-Feshbach model code CoH<sub>3</sub> [54]. First we fitted the calculation to the evaluated  $^{239}\text{Pu}$   $\sigma_f$  in ENDF/B-VIII.0 [28]. ENDF/B-VIII.0 values were obtained from a statistical analysis of many measured fission cross sections, either measured absolutely or as a ratio to other isotopes, e.g.,  $^{235}\text{U}$  [4]. The obtained fission-barrier heights are comparable with other fission cross section studies, such as by Iwamoto [49,55], where the fission barriers for each isotopes reside in the 4.5- to 6-MeV range. It should be noted that fission-barrier heights extracted in this way strongly depend on other nuclear structure properties, such as level densities, in both a naturally deformed shape and an elongated shape toward nuclear fission.

The CoH<sub>3</sub>-calculated  $\sigma_f$  is shown in Fig. 8 by the solid curve, which roughly goes through the data between Shcherbakov and Tovesson. Although the calculated  $\sigma_f$  in this energy range is sensitive to not only the fission barriers but also to the level densities on top of the fission barriers, here

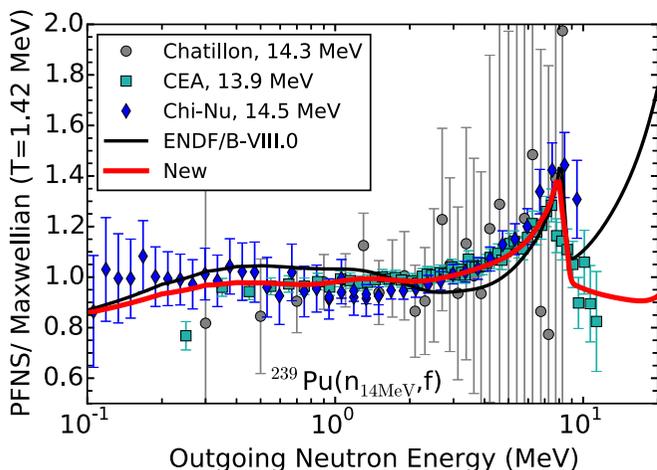


FIG. 9. It is shown that ENDF/B-VIII.0  $^{239}\text{Pu}$   $\chi$  (black line) at 14-MeV incident-neutron energy differs systematically from recently measured Chi-Nu (blue diamonds) and CEA (light blue squares) high-precision data. A new evaluation (in red) based on the Los Alamos and exciton models describes the new data better.

we just show how the calculations are modified by the fission barriers. Hence, the curves show only a lower bound of  $\sigma_f$  uncertainties due to model parameters. The most sensitive fission barriers in the energy range of interest are those of  $^{239}\text{Pu}$  (second chance fission) and  $^{238}\text{Pu}$  (third chance fission). The dot-dashed and dashed curves in Fig. 8 are roughly estimated to reproduce both Shcherbakov's and Tovesson's data, and we found that changes in the fission-barrier heights for both  $^{238}\text{Pu}$  and  $^{239}\text{Pu}$  of  $\pm 150$  keV reasonably cover the data points.

Since typical fission barrier heights of the Pu isotopes are around 5.5 MeV, the deviation of 150 keV is only 2.7%. The present predictive capability of fission barriers, regardless of the modeling employed nowadays, based either on a phenomenological or microscopic point of view, is not that accurate. Hence, a purely theoretical investigation is not able to resolve which of these two discrepant experimental data sets is more realistic, as the uncertainties in the model calculation are about the same size.

$^{239}\text{Pu}$   $\bar{\nu}_{\text{tot}}$  and  $\chi$  were also highlighted as possible sources of bias in simulating  $^{239}\text{Pu}$  LLNL pulse-sphere neutron-leakage spectra. Significant bias in ENDF/B-VIII.0  $^{239}\text{Pu}$   $\bar{\nu}_{\text{tot}}$  can be excluded by a recent measurement [56] which confirms the trend of past experimental data [57] that ENDF/B-VIII.0 data are based upon. Nuclear theory does not yield tighter bounds on  $^{239}\text{Pu}$   $\bar{\nu}_{\text{tot}}$  as codes that predict it, such as CGMF or BeoH [19,20,58–60], depend on tuning their parameters to such high-precision experimental data and are thus less accurate than experimental  $\bar{\nu}_{\text{tot}}$ . However, these codes might be able to constrain PFNS and  $\bar{\nu}$  while simultaneously bringing in diverse physical data, e.g., neutron,  $\gamma$ , and fission-product yields.

Recent  $^{239}\text{Pu}$  PFNS measurements [61,62] clearly highlighted shortcomings in ENDF/B-VIII.0 nuclear data as can be seen from Fig. 9. These nuclear data were evaluated based on an extended Los Alamos model [36,63,64] that was described in Ref. [7]. One part of the extension was

for the physics-expected pre-equilibrium neutron component that was described with the exciton model [63,64] encoded in CoH<sub>3</sub>. The major issue for the ENDF/B-VIII.0 evaluation was that both the space of physically justifiable model parameters (fission barriers and heights, average total-kinetic energy of fission fragments, etc.) and experimental data available at the time [65] (i.e., the data by Chatillon *et al.*) were both too uncertain to significantly constrain  $\chi$ . New experimental data [61,62] verified the general trend predicted by nuclear models for ENDF/B-VIII.0, but defined a distinctly different evaluated curve in Fig. 9. However, if this updated PFNS curve is used for simulating LLNL pulsed-sphere experiments, the bias reduces only very slightly by an average of 0.3%, showing that  $^{239}\text{Pu}$  PFNS issues are only a small part of the bias in pulsed spheres.

In addition to that, we tested whether changing the  $^{239}\text{Pu}$   $\sigma_f$  within credible limits, i.e., to span the space between Tovesson and Shcherbakov data, would clearly improve the simulation of pulsed-sphere neutron-leakage spectra. However, again, only small differences in pulsed-sphere simulations are observed. Using a  $\sigma_f$  in agreement with Tovesson *et al.* data for the neutron-leakage spectra's simulation increases the bias by about 0.5%, while using a  $\sigma_f$  closer to Shcherbakov data leads to an improvement of about 0.5%, which is small given that the  $C/E$  for the Pu sphere considered here ranges from 0.8 to 1.2. While LLNL pulsed-sphere bias seems to favor a  $\sigma_f$  trend that is supported by Shcherbakov data, it is unfortunately not conclusive enough. That indicates that the remaining nuclear data that could be leading to significant bias in simulating pulsed-sphere neutron-leakage spectra are  $(n,\text{el})$ ,  $(n,\text{inl})$  discrete and continuum cross sections along with their angular distributions that counterbalance each other with  $\sigma_f$  and  $(n,2n)$  within the unitarity constraint of the total cross section. Also, angular distributions related to fission [66] should be explored as a possible source of bias. As mentioned above, given the lack of differential data and uncertain models, the resulting nuclear data can vary widely and their physics space is poorly constrained. While one can change the cross sections such that the bias in LLNL pulsed-sphere neutron-leakage spectra is reduced, there are too many moving pieces (angular distributions, cross sections, etc.) to clearly define nuclear data by this information. Hence, an  $(n,\text{inl})$  measurement with a precision of a few percent would be of high importance to better constrain the currently unconstrained physics space spanned between  $(n,\text{tot})$ ,  $(n,\text{el})$ ,  $(n,\text{inl})$  discrete and continuum cross sections and their angular distributions. Given that only scarce experimental data are found for the  $(n,\text{inl})$  cross section, this measurement should cover the energy range from 10 keV to 20 MeV. Such an experiment is recommended in Table I along with a high-precision measurement of less than 1% of the total cross section to derive  $\sigma_{\text{el}}$ . A high-precision absolute measurement of  $\sigma_f$  at the 0.8% level from 10 to 15 MeV is also needed that at the same time explains the biases in Shcherbakov and Tovesson data.

While differential experimental data allowed us to clearly rule out some issues ( $\bar{\nu}_{\text{tot}}$ ,  $\chi$ ) in nuclear data highlighted originally by ML, pulsed-sphere LLNL neutron-leakage spectra did not allow us to conclusively provide input on which differential data were more reliable. However, ML paired with

TABLE I. Unconstrained or poorly constrained physics spaces for  $^{238-242}\text{Pu}$  nuclear data are highlighted for particular energy ranges and observables. These were identified by a combination of (a) random forest and SHAP hinting at potential shortcoming in the data when simulating criticality and pulsed-sphere neutron-leakage spectra and (b) limitations in differential experimental data or theory information to unambiguously defining individual observables. Either new differential experiments, theory developments, or a combination of both are needed to better describe nuclear data. Italic and underlined energy ranges are those where no conclusive experimental information is available to guide theory, while in normal font are those where no clear conclusions can be drawn from existing theory and experiments. Energy ranges are specified with “th” for thermal ( $10^{-11} - 1.77 \times 10^{-6}$  MeV), “r” for resonance ( $1.77 \times 10^{-6} - 0.4$  MeV), “f” for fast (0.4–20 MeV) and “all” from thermal to 20 MeV.

Observable	$^{238}\text{Pu}$	$^{239}\text{Pu}$	$^{240}\text{Pu}$	$^{241}\text{Pu}$	$^{242}\text{Pu}$
PFNS	<u>all</u>	th	<u>all</u>	<u>all</u>	<u>all</u>
( $n,f$ ) cs		> 6 MeV	t; f	f	
( $n,f$ ) $\bar{\nu}$	<u>f</u>	<u>0.3–100 keV</u>	f	all	r, f
( $n, \gamma$ ) cs			<1 MeV	<1 MeV	t
( $n,\text{inl}$ ) cs		<u>f</u>	<u>f</u>	<u>f</u>	<u>f</u>
( $n, 2n$ ) cs			<u>f</u>		
( $n,\text{tot}$ ) cs	all	r, f	f	<u>f</u>	f

differential information gave us hints toward where the largest shortcomings in nuclear data could be hidden, namely in elastic and inelastic  $^{239}\text{Pu}$  nuclear data, which in turn motivates new experimental campaigns and theory developments. It also highlights the need for more sensitive (and maybe less uncertain) integral experiments if we want to conclusively validate PFNS and  $\sigma_f$  in the energy range of 10–15 MeV.

### C. Recommendations for future $^{238-242}\text{Pu}$ developments

We applied the procedure shown in Fig. 1 to identify potential issues in  $^{238-242}\text{Pu}$  ENDF/B-VIII.0 nuclear data using differential and integral experimental data as described in Sec. II. Similar to the example of  $^{239}\text{Pu}$  from 10 to 15 MeV, we found issues in nuclear data for particular observables and energy ranges that could not be resolved given current differential experimental data or nuclear-theory understanding. More explicitly, we identified unconstrained or poorly constrained nuclear-physics spaces given differential and integral information. These observables and energy ranges are listed in Table I; we distinguish between cases where predicting nuclear-physics observables by models and guided by integral information is unconstrained either due to lack of conclusive differential experiments (italic and underlined), or by a significant enough spread across existing measured data (normal font). Performing new differential experiments or improving theory for the observables listed could potentially improve simulations of critical assemblies, pulsed-sphere neutron-leakage spectra, or for application areas represented by these integral experiments. If a discrepant differential experimental database exists for a particular observables in this table, a new experiment should aim to achieve lower uncertainties than

the spread in the data and explore potential issues leading to systematic discrepancies across data sets. It should be noted that one should not translate each single entry in Table I into a request for a new experiment; a combination of theory developments and new experiments might be more effective. For instance, one could measure PFNS for  $^{240}\text{Pu}$  and  $^{242}\text{Pu}$  and then pair this information with existing  $^{239}\text{Pu}$  PFNS experiments and a PFNS model across isotopes to derive reliable  $^{238-242}\text{Pu}$  PFNS nuclear data. For a similar reason, no recommendations were made to investigate ( $n,\text{el}$ ) cross sections as these can be derived by nuclear models if conclusive experimental cross sections for all other open channels, e.g., ( $n, \gamma$ ) and ( $n,f$ ) at thermal and in the resonance range, and the total cross section are available.

## IV. CONCLUSIONS AND OUTLOOK

We explored how we can improve our description of nuclear-physics observables by bringing together three types of information:

- (1) differential experimental data of the observable itself,
- (2) nuclear-theory information, and
- (3) integral experiments that are simulated by integration over many thousands of nuclear-physics observables as a function of energy and angle.

To this end, we employed the random forest algorithm and SHAP metric to identify groups of nuclear data representing nuclear-physics observables that could potentially cause biases in the simulation of criticality experiments and LLNL pulsed-sphere neutron-leakage spectra. Only combinations of physics observables can be identified by the machine-learning algorithms. The reason for that lies in these data being correlated with one another through being jointly used for simulating integral experiments. We show as one example— $^{241}\text{Pu}$  fission observables from 0.1 to 2.354 MeV—how the combination of differential experimental information, theory, and simulated criticality experiments allows us to isolate a likely root cause of the bias in criticality calculations from a group of potentially biased nuclear-physics observables. To do this, we first investigated how far one can change the description of nuclear-physics observables given the spread in differential experimental data and constraints from theory. Then we tested whether one can significantly improve agreement with integral experiments given physically justifiable changes in the numerical data. This process allowed us to identify the most likely cause of bias, the  $^{241}\text{Pu}(n,f)$  cross section. Moreover, we were able to select the experimental-data set that leads to improved simulations of integral experiments out of several, discrepant data sets. Another example, for  $^{239}\text{Pu}$  nuclear data from 10 to 15 MeV, was shown where the combination of differential and integral information was unable to yield conclusive information on what specific  $^{239}\text{Pu}$  nuclear data led to bias in simulating integral experiments. In this case, we were still able to down-select to a smaller group of data that could be biased by comparing to differential data and exploring the impact of different theory-based curves on simulations of integral experiments. We then highlighted which future differential experiments or theory developments

could potentially resolve the remaining confounding and better constrain the nuclear-physics space.

While this work focused on exploring how to resolve issues in  $^{238-242}\text{Pu}$  nuclear data that impact the simulation of criticality and 14-MeV pulsed-sphere neutron-leakage spectra, there are several ways to extend the present approach or use it for similar applications. Obviously, one can extend this work by including broader classes of integral-experiment responses. The benefit of this extension is twofold: On the one hand, one explores the descriptive power of nuclear data for physics observables with respect to a broad range of application areas represented by these integral responses, and thus gives a more balanced assessment of nuclear data. Also, by carefully selecting the integral responses to have different relative sensitivity to various nuclear-physics observables, ML algorithms have more distinct trends at their disposal to aid in resolving the confounding between them. One challenge for this extension is that tools to calculate sensitivity profiles—that tie the nuclear-data observables to the integral responses—are missing for many integral quantities beyond criticality. These missing tools represent a barrier to include various integral responses into ML.

Another way to extend the current approach is by explicitly including differential experimental data in Eq. (4). While conceptually straightforward, there are some major obstacles: First of all, one has to preselect differential experimental data to exclude sets that are known to be biased and quantify realistic uncertainties in order not to skew ML results. While this task is routinely done by nuclear-data evaluators for each observable they provide, this information is not openly available. Hence, if one plans to use all differential information for observables used in criticality experiments, this task would amount to judging and estimating uncertainties for thousands of data sets again. The second challenge is that one has to extend the importance analysis with random forests to account for correlations in uncertainties of differential data, something the approach cannot currently handle.

Lastly, one can apply the methods used here for problems similar in character, i.e., where one is faced with calculating a well-known observables with many parameters that are only loosely defined by less accurate data. For instance, the CGMF and BeoH models calculate the average prompt-neutron multiplicity, which is known to less than 1%, by several hundreds

of parameters that are defined by fission-yield data, level densities, total kinetic energy, etc., that have often uncertainties on the level of several percent. One can apply the same procedure as described here and shown schematically in Fig. 1 to understand which groups of model parameters are associated with bias in calculating  $\bar{\nu}_p$  and then filter them down with experimental information on the parameters of CGMF and BeoH. One could then constrain the less well-known parameters with respect to  $\bar{\nu}_p$  [67]. Another potential area where this methodology could be applied is, for instance, nuclear network simulations of the  $r$ -process of nucleosynthesis. In this particular case, our “integral” observable would correspond to the abundances of the elements, and the “differential” ones would correspond to nuclear properties such as masses,  $\beta$ -decay half-lives, neutron-capture rates, fission probabilities, etc., for those nuclei on the neutron-rich side of the nuclear chart. Sensitivity studies have been performed, for instance, in Refs. [68,69], on which nuclear properties are the most important ones (specific masses,  $\beta$ -delayed neutron-emission probabilities, and  $\beta$ -decay and neutron-capture rates) to describe the final abundances of elements observed in nature; these studies aid in gauging which observables should be used to correspond to what we term “differential input” here. However, one might need to extend the random forest algorithm to account for correlations in the input since many nuclear data used are calculated by models and hence strongly correlated.

## ACKNOWLEDGMENTS

We thank K. J. Kelly, P. Koehler, and R. C. Little (all Los Alamos National Laboratory, LANL) for insightful comments and discussions. Work at LANL was carried out under the auspices of the National Nuclear Security Administration (NNSA) of the U.S. Department of Energy (DOE) under Contract No. 89233218CNA000001. Research reported in this publication was partially supported by the U.S. DOE Laboratory Directed Research & Development (LDRD) program at LANL. We gratefully acknowledge partial support of the Advanced Simulation and Computing program at LANL and the DOE Nuclear Criticality Safety Program (NCSP), funded and managed by NNSA for the DOE.

- 
- [1] A. Kahler, R. MacFarlane, R. Mosteller, B. Kiedrowski, S. Frankle, M. Chadwick, R. McKnight, R. Lell, G. Palmiotti, H. Hiruta, M. Herman, R. Arcilla, S. Mughabghab, J. Sublet, A. Trkov, T. Trumbull, and M. Dunn, *Nucl. Data Sheets* **112**, 2997 (2011).
  - [2] D. Neudecker, A. M. Lewis, E. F. Matthews, J. Vanhoy, R. C. Haight, A. D. Carlson, D. L. Smith, S. Croft, B. Pierson, A. Wallner, A. Al-Adili, D. P. Barry, L. Bernstein, R. C. Block, D. Brown, R. Capote, Y. Danon, M. Devlin, M. Drosig, D. L. Duke *et al.*, Tech. Report LANL LA-UR-19-31156 (LANL, US, 2019).
  - [3] P. Schillebeeckx, B. Becker, Y. Danon, K. Guber, H. Harada, J. Heyse, A. Junghans, S. Kopecky, C. Massimi, M. Moxon, N. Otuka, I. Sirakov, and K. Volev, *Nucl. Data Sheets* **113**, 3054 (2012).
  - [4] A. Carlson, V. Pronyaev, R. Capote, G. Hale, Z.-P. Chen, I. Duran, F.-J. Hamsch, S. Kunieda, W. Mannhart, B. Marcinkievicius, R. Nelson, D. Neudecker, G. Noguere, M. Paris, S. Simakov, P. Schillebeeckx, D. Smith, X. Tao, A. Trkov, A. Wallner *et al.*, *Nucl. Data Sheets* **148**, 143 (2018).
  - [5] A. De Volpi and K. G. Porges, *Phys. Rev. C* **1**, 683 (1970).
  - [6] International Handbook of Evaluated Criticality Safety Benchmark Experiments/Nuclear Energy Agency, Tech. Rep. NEA 7328 (OECD Nuclear Energy Agency, Paris, 2019).
  - [7] D. Neudecker, M. Grosskopf, M. Herman, W. Haeck, P. Grechanuk, S. Vander Wiel, M. Rising, A. Kahler, N. Sly, and P. Talou, *Nucl. Data Sheets* **167**, 36 (2020).
  - [8] J. A. Melendez, S. Wesolowski, and R. J. Furnstahl, *Phys. Rev. C* **96**, 024003 (2017).

- [9] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, *Phys. Rev. C* **98**, 034318 (2018).
- [10] E. Sangaline and S. Pratt, *Phys. Rev. C* **93**, 024908 (2016).
- [11] B. Whewell, M. Grosskopf, and D. Neudecker, *Nucl. Instrum. Methods Phys. Res., Sec. A* **978**, 164305 (2020).
- [12] J. Keeble and A. Rios, *Phys. Lett. B* **809**, 135743 (2020).
- [13] C. Adams, G. Carleo, A. Lovato, and N. Rocco, *Phys. Rev. Lett.* **127**, 022502 (2021).
- [14] R. Utama, J. Piekarewicz, and H. B. Prosper, *Phys. Rev. C* **93**, 014311 (2016).
- [15] R. Utama, W.-C. Chen, and J. Piekarewicz, *J. Phys. G: Nucl. Part. Phys.* **43**, 114002 (2016).
- [16] L. Neufcourt, Y. Cao, W. Nazarewicz, E. Olsen, and F. Viens, *Phys. Rev. Lett.* **122**, 062502 (2019).
- [17] Z.-A. Wang, J. Pei, Y. Liu, and Y. Qiang, *Phys. Rev. Lett.* **123**, 122501 (2019).
- [18] A. E. Lovell, A. T. Mohan, and P. Talou, *J. Phys. G: Nucl. Part. Phys.* **47**, 114001 (2020).
- [19] A. E. Lovell, T. Kawano, S. Okumura, I. Stetcu, M. R. Mumpower, and P. Talou, *Phys. Rev. C* **103**, 014615 (2021).
- [20] B. Becker, P. Talou, T. Kawano, Y. Danon, and I. Stetcu, *Phys. Rev. C* **87**, 014617 (2013).
- [21] J. Verbeke, J. Randrup, and R. Vogt, *Comput. Phys. Commun.* **222**, 263 (2018).
- [22] O. Litaize, O. Serot, and L. Berge, *Eur. Phys. J. A* **51**, 177 (2015).
- [23] C. Wong, J. Anderson, P. Brown, L. F. Hansen, J. L. Kammerdiener, C. Logan, and B. Pohl, Livermore pulsed sphere program: Program summary through July 1971, LLNL UCRL-51144 Rev. 1 (LLNL national laboratory, 1972) (unpublished).
- [24] D. Neudecker, O. Cabellios, A. Clark, W. Haeck, R. Capote, A. Trkov, M. C. White, and M. E. Rising, *Ann. Nucl. Energy* **159**, 108345 (2021).
- [25] N. Otuka, E. Dupont, V. Semkova, B. Pritychenko, A. Blokhin, M. Aikawa, S. Babykina, M. Bossant, G. Chen, S. Dunaeva, R. Forrest, T. Fukahori, N. Furutachi, S. Ganesan, Z. Ge, O. Gritzay, M. Herman, S. Hlavač, K. Katō, B. Lalremruata *et al.*, *Nucl. Data Sheets* **120**, 272 (2014).
- [26] C. Werner, J. Armstrong, F. Brown, J. Bull, L. Casswell, L. Cox, D. Dixon, R. Forster, J. Goorley, H. Hughes, J. Favorite, R. Martz, S. Mashnik, M. Rising, C. Solomon, A. Sood, J. Sweezy, A. Zukaitis, C. Anderson, J. Elson *et al.*, MCNP Users Manual—Code Version 6.2, Tech. Rep. LA-UR-17-29981 (Los Alamos National Laboratory, Los Alamos, NM, 2017).
- [27] L. Breiman, *Mach. Learn.* **45**, 5 (2001).
- [28] D. A. Brown, M. B. Chadwick, R. Capote, A. C. Kahler, A. Trkov, M. W. Herman, A. A. Sonzogni, Y. Danon, A. D. Carlson, M. Dunn, D. L. Smith, G. M. Hale, G. Arbanas, R. Arcilla, C. R. Bates, B. Beck, B. Becker, F. Brown, R. J. Casperson, J. Conlin *et al.*, *Nucl. Data Sheets* **148**, 1 (2018).
- [29] K. Tada, Y. Nagaya, S. Kunieda, K. Suyama, and T. Fukahori, *J. Nucl. Sci. Technol.* **54**, 806 (2017).
- [30] L. Fiorito, G. Žerovnik, A. Stankovskiy, G. Van den Eynde, and P. Labeau, *Ann. Nucl. Energy* **101**, 359 (2017).
- [31] S. M. Lundberg, G. G. Erion, and S.-I. Lee, Consistent individualized feature attribution for tree ensembles, [arXiv:1802.03888](https://arxiv.org/abs/1802.03888) [cs.LG].
- [32] S. Lundber and S. Lee, in *Proceedings of the 31st Conference on Neural Information Processing Systems* (Curran Associate, Inc., Morehouse Ln, Red Hook, 2017).
- [33] P. Grechanuk, M. Rising, and T. Palmer, *J. Comp. Theor. Transp.* **47**, 552 (2018).
- [34] N. Feather, Emission of Neutrons from Moving Fission Fragments, Tech. Rep. BM-148 (British Mission, Oak Ridge, Tennessee, 1942).
- [35] J. Terrell, *Phys. Rev.* **113**, 527 (1959).
- [36] D. G. Madland and J. R. Nix, *Nucl. Sci. Eng.* **81**, 213 (1982).
- [37] B. E. Watt, *Phys. Rev.* **87**, 1037 (1952).
- [38] R. Capote, Y.-J. Chen, F.-J. Hamsch, N. Kornilov, J. Lestone, O. Litaize, B. Morillon, D. Neudecker, S. Oberstedt, T. Ohsawa, N. Otuka, V. Pronyaev, A. Saxena, O. Serot, O. Shcherbakov, N.-C. Shu, D. Smith, P. Talou, A. Trkov, A. Tudora *et al.*, *Nucl. Data Sheets* **131**, 1 (2016).
- [39] D. K. Butler and R. K. Sjoblom, *Phys. Rev.* **124**, 1129 (1961).
- [40] O. Simpson, R. Fluharty, M. Moore, N. Marshall, B. Diven, and A. Hemmendinger, The fission cross section of Pu241 from 20-200eV as determined from a nuclear explosion, in *Proceedings of the Neutron Cross-Section Techn. Conf., Washington, 1966* (US Government Printing Office, Washington, DC, 1966), pp. 910 (F-6).
- [41] I. Szabo, J. Leroy, and J. Marquette, in *Proceedings of Second Conf. on Neutron Physics, Kiev, 1973* (F.E.I., Obninsk, 1973), Vol. 3, p. 27.
- [42] T. Kawano, H. Matsunobu, T. Murata, Y. Zukeran, A. Nakajima, M. Kawai, O. Iwamoto, K. Shibata, T. Nakagawa, T. Ohsawa, M. Baba, and T. Yoshida, *J. Nucl. Sci. Technol.* **37**, 327 (2000).
- [43] F. Käppeler and E. Pfletschinger, *Nucl. Sci. Eng.* **51**, 124 (1973).
- [44] B. Fursov, V. M. Kupriyanov, and G. N. Smirenkii, *Sov. At. Energy* **44**, 262 (1978).
- [45] H. L. Smith, R. K. Smith, and R. L. Henkel, *Phys. Rev.* **125**, 1329 (1962).
- [46] G. Carlson, J. Behrens, and J. Czirr, *Nucl. Sci. Eng.* **63**, 149 (1977).
- [47] P. White, J. Hodgkinson, and G. Wall, in *Proceedings of the Conference on Physics and Chemistry of Fission, Salzburg, 1965* (IAEA, Vienna, Austria, 1965), Vol. 1, p. 219.
- [48] F. Tovesson and T. S. Hill, *Nucl. Sci. Eng.* **165**, 224 (2010).
- [49] K. Shibata, O. Iwamoto, T. Nakagawa, N. Iwamoto, A. Ichihara, S. Kunieda, S. Chiba, K. Furutaka, N. Otuka, T. Ohsawa, T. Murata, H. Matsunobu, A. Zukeran, S. Kamada, and J. Katakura, *J. Nucl. Sci. Technol.* **48**, 1 (2011).
- [50] O. Shcherbakov, A. Donets, A. Evdokimov, A. Fomichev, T. Fukahori, A. Hasegawa, A. Laptev, V. Maslov, G. Petrov, S. Soloviev, Y. Tuboltsev, and A. Vorobyev, *J. Nucl. Sci. Technol. Suppl.* **2 39**, 230 (2002).
- [51] P. Staples and K. Morley, *Nucl. Sci. Eng.* **129**, 149 (1998).
- [52] P. Lisowski, J. Ullman, S. Balestrini, A. Carlson, O. Wasson, and N. Hill, in *Proceedings of the Conference on Nuclear Data for Science and Technol., Mito 1988* (Japan Atomic Energy Research Institute, Tokai-mura, 1988), pp. 97-99.
- [53] L. t. Snyder *et al.*, [arXiv:2107.02881](https://arxiv.org/abs/2107.02881).
- [54] Edited by J. Escher, Y. Alhassid, L. A. Bernstein, D. Brown, C. Fröhlich, P. Talou, and W. Younes, *Compound-Nuclear Reactions, Proceedings of the 6th International Workshop on Compound-Nuclear Reactions and Related Topics CNR\*18*, Springer Proceedings in Physics (Springer, Cham, 2021), Vol. 245, p. 28.
- [55] O. Iwamoto, *J. Nucl. Sci. Technol.* **44**, 687 (2007).

- [56] P. Marini, J. Taieb, G. Bélier, A. Chatillon, D. Etasse, B. Laurent, P. Morfouace, B. Morillon, M. Devlin, J. Gomez, R. Haight, K. Kelly, D. Neudecker, and J. O'Donnell, Energy dependence of prompt Fission Neutron Multiplicities in the  $^{239}\text{Pu}(n, f)$  Reaction, Techn. Report LANL LA-UR-20-28839 (LANL, 2020).
- [57] M. Soleilhac, J. Frehaut, and J. Gauriau, *J. Nucl. Energy* **23**, 257 (1969).
- [58] P. Talou, I. Stetcu, P. Jaffke, M. E. Rising, A. E. Lovell, and T. Kawano, *Comput. Phys. Commun.* **269**, 108087 (2021).
- [59] S. Okumura, T. Kawano, P. Talou, P. Jaffke, and S. Chiba, *J. Nucl. Sci. Tech.* **55**, 1009 (2018).
- [60] S. Okumura, T. Kawano, A. E. Lovell, and T. Yoshida, Energy dependent calculations of fission product, prompt, and delayed neutron yields for neutron induced fission on  $^{235}\text{U}$ ,  $^{238}\text{U}$ , and  $^{239}\text{Pu}$ , [arXiv:2102.01015](https://arxiv.org/abs/2102.01015) [nucl-th].
- [61] K. J. Kelly, M. Devlin, J. M. O'Donnell, J. A. Gomez, D. Neudecker, R. C. Haight, T. N. Taddeucci, S. M. Mosby, H. Y. Lee, C. Y. Wu, R. Henderson, P. Talou, T. Kawano, A. E. Lovell, M. C. White, J. L. Ullmann, N. Fotiades, J. Henderson, and M. Q. Buckner, *Phys. Rev. C* **102**, 034615 (2020).
- [62] P. Marini, J. Taieb, B. Laurent, G. Bélier, A. Chatillon, D. Etasse, P. Morfouace, M. Devlin, J. A. Gomez, R. C. Haight, K. J. Kelly, J. M. O'Donnell, and K. T. Schmitt, *Phys. Rev. C* **101**, 044614 (2020).
- [63] H. Gruppelaar, P. Nagel, and P. Hodgson, *Riv. Nuovo Cimento* **9**, 1 (1986).
- [64] E. Gadioli and P. Hodgson, *Pre-equilibrium Nuclear Reactions* (Clarendon Press, Oxford, UK, 1992).
- [65] A. Chatillon, G. Bélier, T. Granier, B. Laurent, B. Morillon, J. Taieb, R. C. Haight, M. Devlin, R. O. Nelson, S. Noda, and J. M. O'Donnell, *Phys. Rev. C* **89**, 014611 (2014).
- [66] K. J. Kelly, T. Kawano, J. M. O'Donnell, J. A. Gomez, M. Devlin, D. Neudecker, P. Talou, A. E. Lovell, M. C. White, R. C. Haight, T. N. Taddeucci, S. M. Mosby, H. Y. Lee, C. Y. Wu, R. Henderson, J. Henderson, and M. Q. Buckner, *Phys. Rev. Lett.* **122**, 072503 (2019).
- [67] J. Randrup, P. Talou, and R. Vogt, *Phys. Rev. C* **99**, 054619 (2019).
- [68] M. Mumpower, R. Surman, G. McLaughlin, and A. Aprahamian, *Prog. Part. Nucl. Phys.* **86**, 86 (2016).
- [69] M. R. Mumpower, R. Surman, D.-L. Fang, M. Beard, P. Möller, T. Kawano, and A. Aprahamian, *Phys. Rev. C* **92**, 035807 (2015).