



Bayesian approach to heterogeneous data fusion of imperfect fission yields for augmented evaluations

Z. A. Wang ¹, J. C. Pei,^{1,*} Y. J. Chen,² C. Y. Qiao,¹ F. R. Xu ¹, Z. G. Ge,² and N. C. Shu²

¹State Key Laboratory of Nuclear Physics and Technology, School of Physics, Peking University, Beijing 100871, China

²China Nuclear Data Center, China Institute of Atomic Energy, Beijing 102413, China



(Received 28 November 2021; revised 22 June 2022; accepted 9 August 2022; published 22 August 2022)

We demonstrate that Bayesian machine learning can be used to treat the vast amount of experimental fission data which are noisy, incomplete, discrepant, and correlated. To supply the application needs, the two-dimensional cumulative fission yields (CFY) of neutron-induced fission of ^{238}U are evaluated for energy dependencies and uncertainty qualifications by cross-experiment data fusion. For independent fission yields (IFY) with very few experimental data, the heterogeneous data fusion of CFY and IFY is employed to interpolate the energy dependence. This work shows that Bayesian data fusion can facilitate the maximum utilization of imperfect raw nuclear data.

DOI: [10.1103/PhysRevC.106.L021304](https://doi.org/10.1103/PhysRevC.106.L021304)

Machine learning now impacts many scientific fields including physics by offering new data-driven methodologies. The nuclear data needs [1,2] for applications are much concerned about the treatment of vast raw nuclear fission data which are generally noisy, incomplete, and discrepant [3], and they are also very costly. Conventionally, the evaluation of nuclear fission data is a combined inference based on experiments and semi-empirical models [4–6]. Differently, machine learning provides a new route to infer by including all possible data correlations, as well as by reducing influences of subjective factors. Hence the augmented evaluation by machine learning is very anticipated, so that maximum values of raw imperfect nuclear data can be exploited for novel utilizations.

To supply nuclear application needs, more accurate fission data are imperative in upgrading energy productions towards a sustainable and more compact, cleaner, and safer way [1,2]. It is also crucial for understanding the reactor antineutrino anomaly [7,8] and the astrophysical r process [9]. The fission product yields (FPYs) are the most important fission observables, being correlated with multiple observables. In major nuclear data libraries [4,10–12], evaluated fission yields are only available for neutron incident energy at thermal energy, 0.5, and 14 MeV. The energy dependence of FPYs is ascribed to energy-dependent shell effects, dissipation effects, and prompt neutron emissions [13–15]. However, it is still challenging for microscopic fission theories to provide satisfying FPYs at the application level. The uncertainty quantification of nuclear fission data is another major issue for the safety design in nuclear engineering [16]. Since uncertainties are correlated, the covariance matrix is widely used for uncertainty assessments [17,18], which describes actually the first-order sensitivity of uncertainty propagation. It is also not

realistic to obtain an ideal covariance matrix from experiments [16].

To overcome the issues in evaluations of imperfect fission yields regarding the energy dependence and the uncertainty quantification, we propose to use data fusion for augmented evaluations based on Bayesian machine learning. There have been an increasing number of applications of machine learning in nuclear physics (see a review [19]). For example, machine learning has been applied for inferences of nuclear structures [20–25] and nuclear reactions [26–29]. In particular, the machine-learning evaluation of fission data has promising application interests [30–35]. Besides, the Gaussian process can also solve regression problems of nuclear data evaluations but it assumes local correlations. The applications of machine learning in nuclear physics so far have not seriously considered the uncertainty propagation inherent in imperfect data.

Data fusion is a prevalent way to deal with imperfect data [36]. Data fusion refers to the process whereby information from individual datasets sharing at least a number of variables is merged [37]. The fused data are expected to produce more consistent, accurate, and useful information than separated data sources, due to underlying data correlations. It is possible that some nonlocal and high-dimensional correlations could be weak but non-negligible, in analogy to long-range and many-body interactions in quantum systems. The inference would be less precise when data in some energies is sparse, however, its correlations with other data in other energies can be beneficial to improve the inference. The case treated in this work reveals that the evaluation of extremely incomplete independent fission yields (IFYs) in terms of energy dependencies, namely, a large gap between incident-neutron energies of 2 and 14 MeV can profit from the data fusion of a more comprehensive coverage of cumulative fission yields (CFYs). It is known that Bayesian machine learning is ideal for uncertainty quantification [38] which is the key in fusion

*peij@pku.edu.cn

of imperfect data. Furthermore, the fusion of complex heterogeneous data can be naturally treated by Bayesian machine learning, which exhibits the powerful capabilities of machine learning by including correlations beyond the Gaussian process and the covariance matrix.

We employ Bayesian neural networks (BNNs) [39] for machine learning of distributions of fission yields. The posterior distribution $p(\omega|x, t)$ of BNN is based on a prior distribution $p(\omega)$ of network parameters ω and a likelihood function $p(x, t|\omega)$,

$$p(\omega|x, t) = \frac{p(x, t|\omega)p(\omega)}{p(x, t)}. \quad (1)$$

Thus the resulting inference has a distribution and naturally provides the associated uncertainty. The data set is given as $D = \{x_i, t_i\}$, where x_i is the input and t_i is the output fission yields. The details of BNN have been described in previous works [31,38]. For the data fusion, it is crucial to take into account the experimental uncertainties by the likelihood function,

$$p(x, t | \omega) = \exp(-\chi^2/2). \quad (2)$$

The cost function $\chi^2(\omega)$ read

$$\chi^2(\omega) = \sum_{i=1}^N \frac{[t_i - f(x_i, \omega)]^2}{\delta_i^2 + \sigma_{i,\text{expt}}^2}, \quad (3)$$

where $f(x_i, \omega)$ denotes the network values. Note that the weights in the likelihood include a noise scale δ_i^2 and the experimental uncertainty $\sigma_{i,\text{expt}}^2$. δ_i^2 is uniform for all data points and changes in the learning process until numerical convergence. It is reasonable to see that data points with large experimental uncertainties would have small weights in the data fusion. It is also interesting to study how the experimental uncertainties propagate to the prediction uncertainties. The data fusion is in some sense similar to the conception of model averaging and model mixing [20] but does not suffer from the bias in model selections.

First, we demonstrate the uncertainty propagation by varying experimental uncertainties, or by deleting or adding data points deliberately to mimic incomplete and discrepant data. We use BNNs to evaluate the fission-mass distribution of ^{239}U as an illustrative example, with experimental data from Ref. [40]. In this experiment [40], the compound nucleus ^{239}U is produced by transfer reactions with an excitation energy of 8.3 MeV (corresponds to a neutron-incident energy of 3.5 MeV). The data set also includes the evaluated IFY of ^{238}U with neutron incident energies at 0.5 and 14 MeV from JENDL [10]. Note that the nominal energy of 0.5 MeV, given in JENDL as the mean energy of fission of ^{238}U induced by fast neutrons, is not realistic, while the true value is higher by a few MeV [41]. Here the BNN adopts a single hidden layer with 10 neurons because the tests employ a small data set of 309 data points. Figure 1(a) displays the evaluation of fission yields of ^{239}U by including original experimental uncertainties. The BNN evaluations generally agree with experimental data except that the evaluation at mass $A = 100$ is lower than experiments, since the JENDL evaluation is also lower. Figure 1(b) shows the inferred uncertainties by varying the

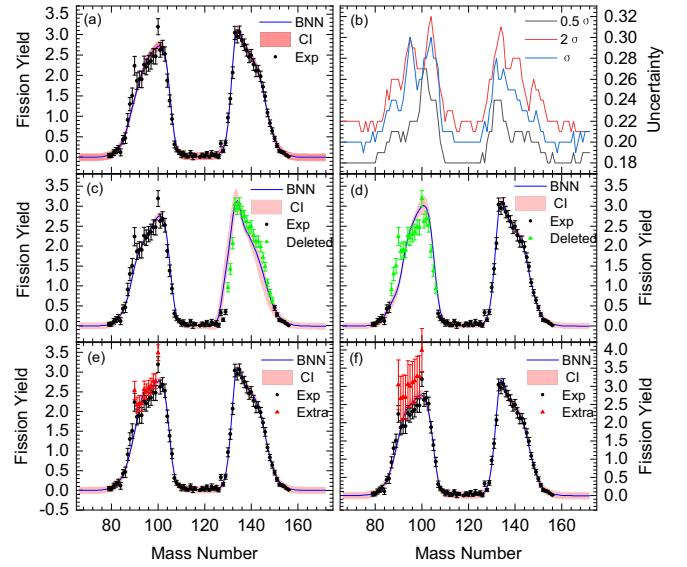


FIG. 1. The BNN evaluations of IFY from fission of ^{239}U , with experimental data from Ref. [40]. The shadows denote the BNN uncertainties as given by the confidence interval (CI) at 95%. Panel (a) shows the BNN evaluation with experimental uncertainties. Panel (b) shows the evaluated uncertainties by varying experimental uncertainties with a factor from 0.5 to 2.0. Panels (c) and (d) show the evaluations when data points of the right peak or the left peak are deleted, respectively. Panels (e) and (f) show the evaluations with some extra noisy data (red color).

experimental uncertainties in Fig. 1(a) with a factor from 0.5 to 2.0. We see that uncertainties indeed increase with increasing experimental uncertainties but do not increase linearly. It is reasonable to see that the uncertainties are generally larger around two peaks. Figures 1(c) and 1(d) show the predictions when the experimental data are deleted at the right or left peak, respectively. We clearly see that the uncertainties as a result of absent data are larger than that with experimental data. Figures 1(e) and 1(f) show the influences of discrepant data. In Fig. 1(e), we add some extra data at the left peak that are not far from experiments with small uncertainties. We see the evaluation moves toward the extra data with increased uncertainties. In Fig. 1(f), we add some extra data that are far from experiments but with large uncertainties. However, we see the evaluation does not move significantly towards the extra data. Therefore the uncertainty propagation is a comprehensive effect and would not significantly change due to a few specific data. This also indicates that Bayesian machine learning includes complex uncertainty correlations.

Next we did the practical Bayesian data fusion of cumulative fission yields (CFY) of neutron-induced fission of ^{238}U from different experiments with different incident energies. The raw imperfect experimental fission yields are taken from the EXFOR library [3,42]. The energy-dependent fission yields of $n + ^{238}\text{U}$ are key quantities in next-generation fast-neutron reactors. In BNN, input variables are given in terms of (Z, N, E) , i.e., the atomic number Z , neutron number N of fragments, and the neutron incident energy E . There are about 1221 scattered data points in a total of 33 experiments

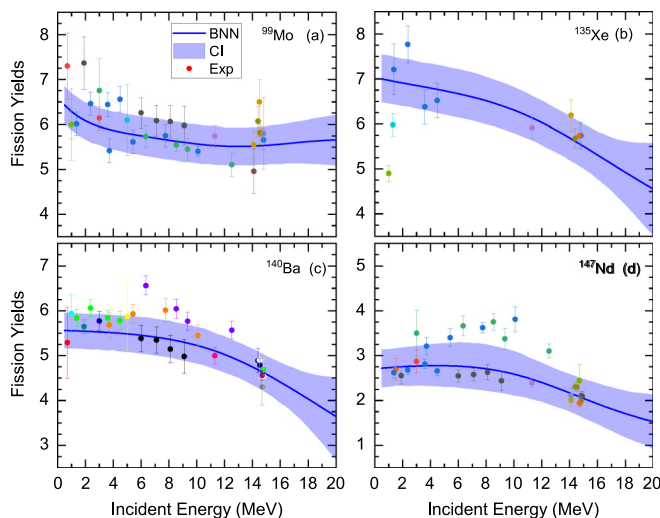


FIG. 2. The BNN evaluated CFY yield-energy relations of fragments (a) ^{99}Mo , (b) ^{135}Xe , (c) ^{140}Ba , and (d) ^{147}Nd from $n + ^{238}\text{U}$ fission are shown. The raw experimental data are taken from EXFOR [3]. Different experiments are denoted by different colors. The shadows denote the corresponding BNN uncertainties given by a CI of 95%.

with different energies of CFY of $n + ^{238}\text{U}$. In addition, 2064 data points of evaluated CFY from JENDL at energies of 0.5 and 14 MeV are included as a learning constraint. Here BNN adopts a double-layer network with 20-20 neurons.

The yield-energy relations of some long-lived isotopes are of particular application interests for monitoring fission environments. For example, ^{135}Xe has large neutron absorption cross sections and is a “poison” for reactors [43]. Figure 2 shows the evaluated energy dependence of fission yields of ^{99}Mo , ^{135}Xe , ^{140}Ba , and ^{147}Nd . In Fig. 2, we see that the experimental data are indeed noisy, sparse, and even discrepant. We see that data fusion can give reasonably the yield-energy relations and uncertainty quantifications. Since JENDL evaluations are included, consequently BNN results are close to JENDL [10] and ENDF [4] evaluations at 0.5 and 14 MeV for the above four fragment yields. Our key motivation is to infer the energy dependence between 0.5 and 14 MeV. For ^{99}Mo , CFY decreases first but does not decrease after 10 MeV. For ^{140}Ba , the CFY decreases smoothly as the energy increases. There are more data for ^{99}Mo and ^{140}Ba and they are usually adopted as standards to determine other yields. For ^{135}Xe , there are very few experimental data and the resulting uncertainties are larger than in other cases. For ^{147}Nd , there are some significantly discrepant data between 5 and 10 MeV. Our BNN evaluations are flat before 10 MeV and are close to the latest measurement [44]. The corresponding uncertainties become larger in this energy regime. In Fig. 2, the inference uncertainties are larger than some experimental uncertainties but they are acceptable regarding the influences of discrepant data. The extrapolated yields at higher energies generally have increasing uncertainties. The BNN uncertainty includes two parts: the overall regression noise and the data-dependent posterior uncertainty [39], while they are not exclusive. In Fig. 2, there is an overall noise scale of ≈ 0.6 which is related

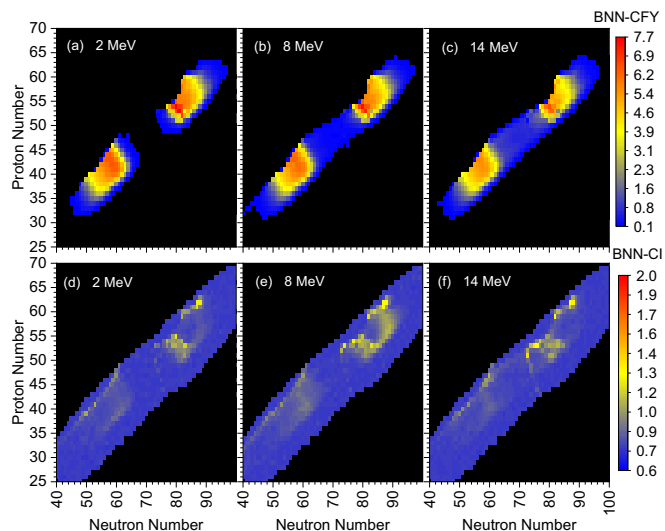


FIG. 3. The two-dimensional CFY distributions of $n + ^{238}\text{U}$ fission obtained by BNN data fusion. Panels (a)–(c) show CFY at incident neutron energies of 2, 8, and 14 MeV, respectively. Panels (d)–(f) show the corresponding uncertainties at different energies.

to the description capability of BNN and can be reduced by a more complicated neural network or a physics-informed neural network. Besides, Figs. 1 and 2 demonstrate that the experimental uncertainty, data sparsity, and data discrepancy are reflected by the data-dependent uncertainty.

Figure 3 displays the two-dimensional CFY distributions of neutron-induced fission of ^{238}U at energies of 2, 8, and 14 MeV from BNN data fusion. It is known that the distributions of fission yields are often incomplete. Recent experiments have made great progress in measurements of the isotopic mass yields using the inverse kinetic method [40,45]. The two-dimensional CFY distribution is very different from a sum of two-dimensional Gaussian distributions. In principle, our results include the yield-energy relations of all fragments. We can see that the highest yields are around $Z = 53\text{--}54$ and $N = 79\text{--}81$, which decrease gradually with increasing energies. It is known that symmetric fission would become more prominent as excitation energy increases [46,47]. The two peaks associated with asymmetric fission modes reduce with increasing energies, while the central distributions associated with the symmetric fission mode are increasing. Therefore, the energy dependence of fission modes can be reasonably described by BNN data fusion. Figures 3(d)–3(f) shows the corresponding uncertainties. There is an overall background noise scale ≈ 0.6 for all data. Obviously the uncertainty is energy dependent. The average uncertainties of Figs. 3(d)–3(f) are 0.789, 0.805, and 0.793 for 1032 points, respectively. The larger uncertainty at 8 MeV is related to the data sparsity around this energy regime. The uncertainties of some CFY are as large as 1.4 where CFY have large values. The relative uncertainties around peaks are actually smaller. The one-dimensional charge yields can also be extracted from the two-dimensional CFY distributions. The precise and complete CFY are crucial to estimate the abnormal antineutrino

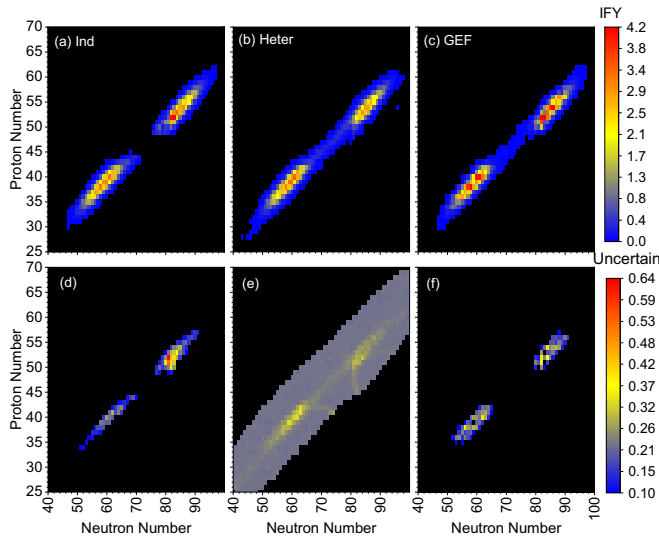


FIG. 4. The two-dimensional IFY distributions of $n + {}^{238}\text{U}$ fission with the neutron incident energy at 8 MeV. Panel (a) shows the inference by BNN learning pure IFY data, panel (b) shows the inference by the BNN heterogeneous data fusion of CFY and IFY, panel (c) shows the inference by the GEF model for comparison. Panels (d)–(f) show the uncertainties corresponding to panels (a)–(c), respectively.

spectrum at reactors [8] and our results will be useful for such studies.

Different from CFY, there are much fewer experimental data of IFY and its evaluation is more challenging. IFY distributions are particularly useful for posing constraints on fission theories. For the neutron-induced fission of ${}^{238}\text{U}$, there are only 211 IFY experimental data [3] and 2064 evaluated data points from JENDL. For better interpolations of the energy dependence of IFY, we employ the heterogeneous data fusion of CFY and IFY by utilizing the correlations between CFY and IFY. The learning dataset is $(A_i, Z_i^{C/I}, E_i, Y_i^{C/I})$, in which CFY and IFY share the mass number A of fragments and the incident energy E , but their proton numbers are deliberately separated into two groups as Z_i^I and Z_i^C . Such a scheme is based on the fact that CFY mainly results from β decays of IFY but they have very different distributions. Actually the conversion between CFY and IFY is nontrivial although they are related by β decays. The heterogeneous datasets are learned by the same network so that their multidimensional correlations are naturally included. The evaluated IFY from JENDL at 0.5 and 14 MeV are used in the learning and our main purpose is to interpolate the energy dependence at other energies.

Figure 4 displays the evaluation of IFY at 8 MeV with learning of pure IFY and heterogeneous CFY and IFY, respectively. The homogeneous fusion of IFY adopts a 16-16 network and the heterogeneous fusion of CFY and IFY adopts a 20-20 network. There are almost no experimental IFY data

between 2 and 14 MeV. It is known that the energy dependence of FPYs is nonlinear from Fig. 2. Thus the evaluation of IFYs at 8 MeV is not reliable by learning only IFY data. We see that there are almost no IFYs from the symmetric fission between two peaks in Fig. 4(a), but there are considerable symmetric fission IFY in Fig. 4(b). This demonstrated that the heterogeneous data fusion of CFYs and IFYs can build the energy-dependence information of CFYs into IFYs. The resulting IFYs in Fig. 4(b) are consistent with the CFY in Fig. 3(b). For comparison, the same evaluation by the GEF model [6] is shown in Fig. 4(c), which also has considerable IFY from the symmetric fission. Figures 4(d)–4(f) display the uncertainties, correspondingly. We see that the homogeneous fusion has a much smaller overall noise but some points have large uncertainties. The heterogeneous fusion has an overall background noise scale about 0.2 due to the influence of CFY. Indeed, it has been pointed out that cross-experiment correlations would result in increased final estimated uncertainties [18]. On the other hand, without the background noise, the data-dependent uncertainties of heterogeneous data fusion are generally smaller than that of the homogeneous fusion. The BNN evaluation will be further improved by considering the subtle features such as the odd-even effects in charge yields and the second-chance fission, through additional dedicated data or physics knowledge.

In summary, we have applied Bayesian data fusion for the evaluation of imperfect fission yields, in which the raw experimental data are generally noisy, incomplete, discrepant, and correlated. The Bayesian data fusion can be used to augment the inference of energy dependence and uncertainty quantification of imperfect fission data by including underlying correlations. The full two-dimensional distributions of CFYs in terms of energy dependence are obtained, which is valuable for designing advanced reactors. The evolution of fission modes is revealed by the smooth transition of the two-dimensional fission yields. The yield-energy relations of some key fragments are now given with uncertainty quantifications, in which the energy dependencies of the uncertainties are data dependent. Furthermore, we applied the heterogeneous data fusion to interpolate the energy dependence of IFYs, which have very few experimental data, by including additional information from CFYs. In the future, it is vital to develop physics-informed machine learning for more reliable evaluations. We expect that the heterogeneous data fusion of multiple fission observables by physics-informed machine learning is promising for a comprehensive and accurate modeling of nuclear fission.

We are grateful for valuable comments by W. Nazarewicz. This work was supported by the National Key R&D Program of China (Contract No. 2018YFA0404403) and by the National Natural Science Foundation of China under Grants No. 11975032, No. 11835001, No. 11790325, and No. 11961141003.

[1] L. A. Bernstein, D. A. Brown, A. J. Koning, B. T. Rearden, C. E. Romano, A. A. Sonzogni, A. S. Voyles,

and W. Younes, *Annu. Rev. Nucl. Part. Sci.* **69**, 109 (2019).

- [2] K. Kolos, V. Sobes, R. Vogt *et al.*, *Phys. Rev. Research* **4**, 021001 (2022).
- [3] <https://www-nds.iaea.org/exfor/>.
- [4] M. B. Chadwick *et al.*, *Nucl. Data Sheets* **112**, 2887 (2011).
- [5] U. Brosa, S. Grossmann, and A. Müller, *Phys. Rep.* **197**, 167 (1990).
- [6] K.-H. Schmidt, B. Jurado, C. Amouroux, and C. Schmitt, *Nucl. Data Sheets* **131**, 107 (2016).
- [7] G. Mention, M. Fechner, Th. Lasserre, Th. A. Mueller, D. Lhuillier, M. Cribier, and A. Letourneau, *Phys. Rev. D* **83**, 073006 (2011).
- [8] A. A. Sonzogni, E. A. McCutchan, T. D. Johnson, and P. Dimitriou, *Phys. Rev. Lett.* **116**, 132502 (2016).
- [9] M. Eichler *et al.*, *Astrophys. J. Lett.* **808**, 30 (2015).
- [10] K. Shibata *et al.*, *J. Nucl. Sci. Technol. (Abingdon, UK)* **48**, 1 (2011).
- [11] A. J. M. Plompen *et al.*, *Eur. Phys. J. A* **56**, 181 (2020).
- [12] Z. G. Ge, Z. X. Zhao, H. H. Xia, Y. X. Zhuang, T. J. Liu, J. S. Zhang, and H. C. Wu, *J. Korean Phys. Soc.* **59**, 1052 (2011).
- [13] M. Bender *et al.*, *J. Phys. G* **47**, 113002 (2020).
- [14] Y. Qiang and J. C. Pei, *Phys. Rev. C* **104**, 054604 (2021).
- [15] T. Ethvignot, M. Devlin, H. Duarte, T. Granier, R. C. Haight, B. Morillon, R. O. Nelson, J. M. O'Donnell, and D. Rochman, *Phys. Rev. Lett.* **94**, 052701 (2005).
- [16] P. Talou, *Ann. Nucl. Energy* **164**, 108568 (2021).
- [17] K. Tsubakihara, S. Okumura, C. Ishizuka, T. Yoshida, F. Minato, and S. Chiba, *J. Nucl. Sci. Technol. (Abingdon, UK)* **58**, 151 (2021).
- [18] P. Talou, *EPJ Nuclear Sci. Technol.* **4**, 29 (2018).
- [19] P. Bedaque, A. Boehnlein, M. Cromaz *et al.*, *Eur. Phys. J. A* **57**, 100 (2021).
- [20] L. Neufcourt, Y. Cao, S. A. Giuliani, W. Nazarewicz, E. Olsen, and O. B. Tarasov, *Phys. Rev. C* **101**, 044307 (2020).
- [21] R. Utama, J. Piekarewicz, and H. B. Prosper, *Phys. Rev. C* **93**, 014311 (2016).
- [22] Z. M. Niu and H. Z. Liang, *Phys. Lett. B* **778**, 48 (2018).
- [23] D. Wu, C. L. Bai, H. Sagawa, and H. Q. Zhang, *Phys. Rev. C* **102**, 054323 (2020).
- [24] R. Utama, W.-C. Chen, and J. Piekarewicz, *J. Phys. G* **43**, 114002 (2016).
- [25] Y. Ma, C. Su, J. Liu, Z. Ren, C. Xu, and Y. Gao, *Phys. Rev. C* **101**, 014304 (2020).
- [26] C. W. Ma, D. Peng, H. L. Wei, Z. M. Niu, Y. T. Wang, and R. Wada, *Chin. Phys. C* **44**, 014104 (2020).
- [27] A. E. Lovell, F. M. Nunes, M. Catacora-Rios, and G. B. King, *J. Phys. G* **48**, 014001 (2021).
- [28] D. Neudecker, O. Cabellos, A. R. Clark, M. J. Grosskopf, W. Haeck, M. W. Herman, J. Hutchinson, T. Kawano, A. E. Lovell, I. Stetcu, P. Talou, and S. V. Wiel, *Phys. Rev. C* **104**, 034611 (2021).
- [29] Y. Wang, F. Li, Q. Li, H. Lü, and K. Zhou, *Phys. Lett. B* **822**, 136669 (2021).
- [30] P. Vicente-Valdez, L. Bernstein, and M. Fratoni, *Ann. Nucl. Energy* **163**, 108596 (2021).
- [31] Z. A. Wang, J. C. Pei, Y. Liu, and Y. Qiang, *Phys. Rev. Lett.* **123**, 122501 (2019).
- [32] C. Y. Qiao, J. C. Pei, Z. A. Wang, Y. Qiang, Y. J. Chen, N. C. Shu, and Z. G. Ge, *Phys. Rev. C* **103**, 034621 (2021).
- [33] Z. A. Wang and J. C. Pei, *Phys. Rev. C* **104**, 064608 (2021).
- [34] A. E. Lovell, A. T. Mohan, and P. Talou, *J. Phys. G* **47**, 114001 (2020).
- [35] L. Tong, R. He, and S. Yan, *Phys. Rev. C* **104**, 064617 (2021).
- [36] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, *Inf. Fusion* **57**, 115 (2020).
- [37] https://en.wikipedia.org/wiki/Data_fusion.
- [38] D. R. Phillips, R. J. Furnstahl, U. Heinz, T. Maiti, W. Nazarewicz, F. M. Nunes, M. Plumlee, M. T. Pratola, S. Pratt, F. G. Viens, and S. M. Wild, *J. Phys. G* **48**, 072001 (2021).
- [39] R. Neal, *Bayesian Learning for Neural Networks* (Springer, New York, 1996).
- [40] D. Ramos *et al.*, *Phys. Rev. C* **101**, 034609 (2020).
- [41] K. Kern, M. Becker, and C. Broeders, Assessment of fission product yields data needs in nuclear reactor applications, *PHYSOR 2012: International Conference on Advances in Reactor Physics, Knoxville, Tennessee, USA* (Curran Associates, Inc., Red Hook, NY, 2012), p. 4383.
- [42] N. Otuka *et al.*, *Nucl. Data Sheets* **120**, 272 (2014).
- [43] O. Moreira, *Ann. Nucl. Energy* **39**, 62 (2012).
- [44] M. E. Gooden *et al.*, *Nucl. Data Sheets* **131**, 319 (2016).
- [45] J.-F. Martin *et al.*, *Phys. Rev. C* **104**, 044602 (2021).
- [46] J. A. Sheikh, W. Nazarewicz, and J. C. Pei, *Phys. Rev. C* **80**, 011302(R) (2009).
- [47] J. Zhao, T. Nikšić, D. Vretenar, and S. G. Zhou, *Phys. Rev. C* **99**, 014618 (2019).