Uncertainty quantification of mass models using ensemble Bayesian model averaging

Yukiya Saito,[1,2,3,4,*] I. Dillmann,[1,5] R. Krücken,[1,2,6] M. R. Mumpower,[7,8] and R. Surman[3]

[1]*TRIUMF, 4004 Wesbrook Mall, Vancouver, BC V6T 2A3, Canada*
[2]*Department of Physics and Astronomy, The University of British Columbia, Vancouver, BC V6T 1Z1, Canada*
[3]*Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA*
[4]*Department of Physics and Astronomy, University of Tennessee, Knoxville 37996, Tennessee, USA*
[5]*Department of Physics and Astronomy, University of Victoria, Victoria, BC V8P 5C2, Canada*
[6]*Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[7]*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[8]*Center for Theoretical Astrophysics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

(Received 18 August 2023; revised 22 January 2024; accepted 1 April 2024; published 1 May 2024)

Developments in the description of the masses of atomic nuclei have led to various nuclear mass models that provide predictions for masses across the whole chart of nuclides. These mass models play an important role in understanding the synthesis of heavy elements in the rapid neutron capture (r) process. However, it is still a challenging task to estimate the size of uncertainty associated with the predictions of each mass model. In this work, a method called *ensemble Bayesian model averaging* (EBMA) is introduced to quantify the uncertainty of one-neutron separation energies (S_{1n}) which are directly relevant in the calculations of r-process observables. This Bayesian method provides a natural way to perform model averaging, selection, and uncertainty quantification, by combining the mass models as a mixture of normal distributions whose parameters are optimized against the experimental data, employing the Markov chain Monte Carlo method using the no-u-turn sampler. The EBMA model optimized with all the experimental S_{1n} from the AME2003 nuclides are shown to provide reliable uncertainty estimates when tested with the new data in the AME2020.

DOI: 10.1103/PhysRevC.109.054301

I. INTRODUCTION

Since the first introduction of the nuclear liquid drop model, the theoretical description of nuclear masses has seen great progress, which gave rise to many related but different approaches. It is now possible to describe the ground state properties of nuclei across the chart of nuclei with theories of different scales: Macroscopic-microscopic theories such as the finite-range droplet model (FRDM) [1,2], Weizsäcker-Skyrme (WS) models [3–6], microscopically inspired Duflo-Zucker models [7], and more microscopic theories such as nuclear density functional theory (DFT) with different interactions or energy density functionals (EDFs) [8–10].

Among the theoretical models that describe various aspects of nuclear structure, reactions, and decays, global mass models play an important role in understanding the origin of heavy elements in the Universe via the rapid neutron capture (r) process [11–13]. This is because the nuclear masses determine the Q value (energy release or absorption) of nuclear reactions and decays and are always required for the calculation of the relevant rates. The masses of the vast majority of neutron-rich nuclei relevant to the r process have yet to be experimentally studied; therefore, theoretical predictions of the masses must be used instead. This means that the mass models used in nucleosynthesis studies may have a significant impact on the prediction of abundance patterns and kilonova lightcurves [14,15].

One of the challenges in understanding the impact of mass models on nucleosynthesis is that, in general, uncertainty estimates associated with the theoretical masses are not available. Although there has been an effort to quantify the uncertainty in microscopic theories [16,17], the mass models that are commonly used in nucleosynthesis studies, especially macroscopic-microscopic and phenomenological models, do not come with a quantified prediction uncertainty. The root-mean-square (RMS) error of each mass model can be calculated with respect to the observations, but it most likely underestimates the uncertainty where there is no data (see Fig. 1). This poses a challenge in quantifying the uncertainty in the r-process nucleosynthesis that arises from uncertain nuclear masses.

As the next-generation radioactive isotope beam facilities allow us to gain access to more neutron-rich isotopes, it will become possible to test the performances of mass models in extremely neutron-rich regions of the chart of nuclides. It would be ideal to have a method that can continuously incorporate new experimental results, test agreement to theoretical predictions, and update their associated uncertainties in extrapolation to further neutron-rich regions.

*yukiya@alum.ubc.ca

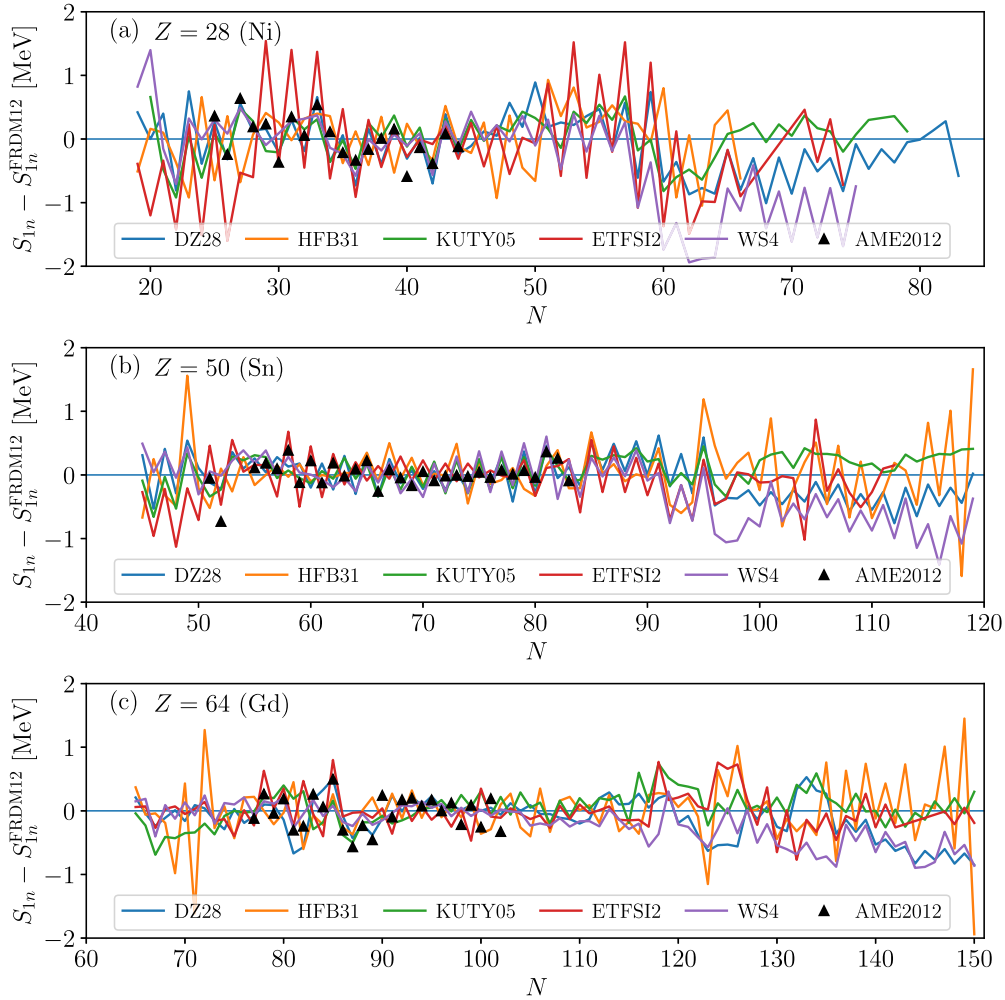

FIG. 1. Comparison of one-neutron separation energies (S_{1n}) predicted by each mass model used in this study and the experimental masses from the AME2020 [18], relative to the predictions of the FRDM2012 for (a) $Z = 28$ (Ni), (b) $Z = 50$ (Sn), and (c) $Z = 64$ (Gd) isotopes. The mass models and references are listed in Sec. II C.

In this work, we will apply a method called *ensemble Bayesian model averaging* (EBMA) introduced by Ref. [19] to combine available experimental data and multiple theoretical mass models, as well as to quantify the mass uncertainty. This method models an ensemble of theoretical mass models as a mixture of normal distributions, whose parameters are estimated based on the observations. EBMA combines model selection, averaging, and uncertainty quantification in a single framework. Although we focus on theoretical one-neutron separation energies (S_{1n}) in this work, this Bayesian method is quite general and can potentially be applied to other nuclear physics observables. We note, however, that the application of this method to quantities such as reaction and decay rates relevant to the *r* process will require further investigations, due to the smaller number of models available and the fact that the rates can vary by orders of magnitudes between nuclei.

Recently, data-driven modeling of nuclear masses using machine learning techniques has gained popularity [20–29]. Especially, Refs. [22,23] applied Bayesian model averaging (BMA, see Sec. II A) to nuclear mass models for the first time, opening up the possibility to perform probabilistic analyses on competing mass models (see Sec. II B 4 for further discussions). One of the advantages of probabilistic models is the ability to achieve high accuracy while providing uncertainty estimates. Although there have been attempts to construct physically interpretable models [27], it is generally challenging to gain insight into the underlying physics from machine learning models. Nevertheless, the advantages of machine learning models are that they can be created rapidly and often achieve similar performance to state-of-the-art theoretical models. Although they may not necessarily be able to predict new or unknown physics, the flexibility of the models may allow us to combine known physics in potentially novel ways that are difficult to produce through standard modeling.

The purpose of this study is not to create another mass model or to improve existing ones with machine learning. Rather, the aim is to investigate how well an ensemble of theoretical models can reproduce experimental data and quantify the performance of each model in the ensemble. This also quantifies the uncertainty in extrapolating the experimental data. Our approach should be considered as a method for

model averaging, selection, and uncertainty quantification, using only existing theoretical models.

This paper is divided into the following sections: in Sec. II, we discuss the details of the EBMA method and the numerical experiment where we construct EBMA models; in Sec. III, we discuss the results of the different approaches for constructing EBMA models and the details of the quantified uncertainties for one neutron separation energies (S_{1n}); finally, we summarize the work presented in the paper and describe possible future developments in Sec. IV.

II. METHOD

A. Bayesian model averaging

Bayesian model averaging (BMA) is applicable when more than one statistical model that describes the data reasonably well is available, and one wishes to account for the uncertainty in the analysis arising from conditioning on a single model. BMA computes a weighted average of the probability density functions (PDFs), weighted by the posterior probability of the "correctness" of each model given the training data. Following the description in Refs. [19,30], the posterior distribution of the observable of interest Δ (in our case, it correponds to one-neutron separation energies S_{1n}), defined by BMA, is

$$p(\Delta|D) = \sum_{k=1}^{K} p(\Delta|M_k, D)\; p(M_k|D), \tag{1}$$

where $p(\Delta|M_k, D)$ is the posterior PDF of the observable of interest based on a single statistical model M_k, and $p(M_k|D)$ is the corresponding posterior model probability, which represents how well the model M_k fits the data D. The posterior model probabilities can be considered as weights, since their sum is equal to 1.

B. Ensemble Bayesian model averaging

One of the limitations in the applicability of the BMA method is that the participating models themselves must be probabilistic. In nuclear physics, most models are not probabilistic. Therefore, we need to extend the BMA framework to handle such models. Raftery *et al.* [19] introduced the EBMA method, which computes the weighted average of an ensemble of bias-corrected models, as a finite mixture of normal distributions. In the EBMA framework, the predictive model is

$$p(\Delta|m_1, \ldots, m_K) = \sum_{k=1}^{K} w_k\; g_k(\Delta|m_k), \tag{2}$$

where Δ is again the quantity of interest (in our case S_{1n} of some nucleus), w_k is the weight of the model k, whose posterior represents the probability of the model k being the best one, based on the observed data D. Model prediction m_k can be a vector or scalar. In our case, it is a vector of theoretical S_{1n} values predicted by the mass model k. Since the size of the weight represents the posterior probability of the model being the best one, even if the ensemble includes an extremely inaccurate model, its effect on the predictive distributions of EBMA will be quite small, since an extremely small weight would be assigned to the model. $g_k(\Delta|m_k)$ is a normal PDF with its mean defined by the model prediction m_k and the standard deviation σ_k:

$$g_k(\Delta|m_k) = N\left(\Delta\middle|m_k, \sigma_k^2\right). \tag{3}$$

Although one needs to be cautious of overfitting, the mean of the normal PDF m_k may be replaced by the bias-corrected model predictions, which are discussed in more detail in the following section. In the original EBMA by Raftery *et al.* [19], a constant standard deviation was used across all the models in the ensemble; however, we take it as model dependent (denoted by the subscript k), which is a more natural way to construct a mixture model.

1. Bias correction

In constructing EBMA models, although not strictly necessary, Ref. [19] suggests linearly correcting the bias in the prediction of each model, prior to Bayesian inference of weights and standard deviations. This means replacing m_k in Eq. (3) with $a_k + b_k m_k$, where a_k is the intercept coefficient, b_k is the slope coefficient, and m_k is the prediction of the model k. In our case, this corresponds to linearly correcting a vector of theoretical S_{1n} values predicted by each mass model. The original prediction of the model k corresponds to $a_k = 0$ and $b_k = 1$.

Although the bias correction may provide one way to fine-tune the predictions of theoretical models (for more sophisticated approaches of correcting mass models, see, e.g., Ref. [22]) in case the model was constructed based on an older set of experimental data, or the predictions for the nuclei of interest are known to systematically deviate from the experimental results, we note that the use of bias correction can lead to overfitting. We investigate the effect of bias correction in Sec. III D. Throughout the manuscript, we do not apply bias correction unless explicitly noted otherwise.

2. Bayesian inference

The parameters of interest in our statistical inference are the weights w_k $(k = 1, \ldots, K)$ and the standard deviations σ_k $(k = 1, \ldots, K)$ of the normal distributions that correspond to each of the theoretical mass models in the ensemble. Therefore, prior distributions for these parameters must be specified. In general, we try to choose the prior distributions to be as weakly informative as possible. For the weights, since they have to sum up to one ($\sum_{k=1}^{K} w_k = 1$), we model the parameters with a Dirichlet distribution of order K, which naturally meets this requirement. Dirichlet distribution of order K has hyperparameters α_k $(k = 1, \ldots, K)$, thus, the prior distribution is expressed as

$$\begin{aligned} &p(w_1, w_2, \ldots, w_K) \\ &\quad = \text{Dirichlet}(w_1, w_2, \ldots, w_K|\alpha_1, \alpha_2, \ldots, \alpha_K). \end{aligned} \tag{4}$$

We set all the concentration parameters of the Dirichlet distribution to 1, which makes the prior distribution uniform for all weights $w_1, \ldots, w_K$, to represent the belief that we do not know which model would perform best. The prior distributions for the standard deviations are chosen to be exponential

distributions with the rate parameters equal to 1, which has been suggested to be one of the weaker priors [31].

The likelihood of the normal mixture model is defined as

$$L\left(w_1,\ \ldots,\ w_K,\ \sigma_1^2,\ \ldots,\ \sigma_K^2\right) = \prod_{(N,Z)} \left(\sum_{k=1}^{K} w_k g_k(\Delta_{(N,Z)} | m_{k,(N,Z)}) \right), \tag{5}$$

where Δ is again the quantity of interest (in our case S_{1n}), and m_k is the model predictions (a vector of S_{1n} predicted by the mass model k). The subscript (N, Z) represents pairs of neutron number N and proton number Z of the nuclei where experimental values exist. In practice, the logarithm of likelihood (log-likelihood) is often used for computation to avoid numerical problems.

With the prior distributions and the likelihood function, it is now possible to formulate the posterior distributions for the parameters of the EBMA model,

$$p(\boldsymbol{w}, \boldsymbol{\sigma^2}|D) \propto L(\boldsymbol{w}, \boldsymbol{\sigma^2})\ p(\boldsymbol{w})\ p(\boldsymbol{\sigma^2}), \tag{6}$$

where $\boldsymbol{w} = w_1, \ldots, w_K$, $\boldsymbol{\sigma^2} = \sigma_1^2, \ldots, \sigma_K^2$, and D denotes observational (experimental) data. The prior distributions are denoted as $p(\boldsymbol{w})$ and $p(\boldsymbol{\sigma}^2)$, respectively.

3. Predictive variance

In EBMA models, the uncertainty of the quantity of interest can be interpreted in the form of variance of the posterior predictive distribution. Based on Ref. [19] but reflecting the fact that our σ_k depends on model k, the predictive variance can be written as

$$\mathrm{Var}(\Delta|m_1, m_2, \ldots, m_K) = \sum_{k=1}^{K} w_k \left((m_k) - \sum_{i=1}^{K} w_i (m_i) \right)^2 + \sum_{k=1}^{K} w_k \sigma_k^2, \tag{7}$$

where the first term corresponds to the spread of predictions by the member mass models of the ensemble, and the second term corresponds to the expected deviation from the observations of each mass model, weighted by the posterior weights. If bias-corrected models predictions are used, then m_k is replaced by $a_k + b_k m_k$.

4. Differences to related works and discussion of models

It is worth discussing the key differences between our framework and related studies that use the BMA method, namely Refs. [22,23,32–34]. In their BMA framework, the uncertainty quantification of the considered mass models is performed by constructing Gaussian process (GP) emulators, which learn the corrections to the mass models from the residuals with respect to the observed values. Therefore, the quality of the prediction and the corresponding uncertainty mainly depend on the performance of the GP emulator. The BMA weights are calculated either based on some criteria such as nuclei being bound or the performances of each mass model on the test data. One of the drawbacks of this method is that the derived weights are point estimates, and the resulting BMA uncertainty is a deterministic weighted average of the GP uncertainties. Furthermore, one has to be cautious when performing extrapolations using GPs, since an unconstrained GP converges to its mean with fixed uncertainty away from the data [35,36].

However, the EBMA framework keeps the point predictions of the mass models in the ensemble. Instead, the weights and variances associated with each mass model are modeled probabilistically based on the experimental data. The probabilistic distributions are reflected onto the resulting predictive uncertainty through the Bayesian framework. This framework directly uses the predictions of each mass model that constitute the EBMA model; therefore, the local trend of the predictions remains unchanged.

One of the shortcomings of the current method is that inference of posterior weights is performed assuming that all observed data points are equally relevant. In the case of uncertainty quantification of mass models for neutron-rich nuclei, for example, one may wish to estimate the weights by focusing on the data for neutron-rich nuclei. However, this may pose a trade-off since the weights are better estimated using all available data, while only using data in a specific region may better capture the local performances of the mass models. The concept of such location- (input domain-) dependent weights is referred to as "Bayesian model mixing" or BMM, put forward by Refs. [37,38]. Recently, Ref. [29] applied BMM to nuclear mass models, using weights defined as a Dirichlet distribution. They modeled location-dependent (local) weights by designing the hyperparameter α for the Dirichlet distribution to be location-dependent. Another point is that they directly combined the predictions of the mass models, whereas in the current work the mass models are treated as components of a normal mixture model.

Further technical development would be required to incorporate location-dependent weights into the averaging of mass models, which will be investigated in the future. In this work, we provide a general methodology for averaging nuclear mass models. The dependence of uncertainty on location is assumed to be represented by the spread of the predictions of different mass models, as shown in Fig. 1.

C. Setup of a numerical experiment

In the numerical experiments discussed in the current work, all probabilistic models have been implemented using PyMC [39], which is a probabilistic programming language written in Python. PyMC offers an implementation of a highly efficient sampler called no-u-turn-sampler (NUTS), which adaptively tunes the parameters associated with the Hamiltonian (or Hybrid) Monte Carlo method [40,41]. Conventionally, parameter estimation in mixture models is performed with the expectation maximization (EM) algorithm to avoid the so-called "label switching problem" [42,43]. The label switching problem arises in mixture models such as EBMA models, since the likelihood [Eq. (5)] remains unchanged under permutation of the labels ($k = 1, \ldots K$) of the mixture components $g_k(\Delta|m_k)$. This makes the analysis of the posterior distributions challenging. Although, the EM algorithm does not guarantee convergence to the global

optimal weights and variances, especially in high-dimensional problems.

Furthermore, MCMC methods would be able to provide much more complete information on the posterior distributions. In our numerical experiments, we did not find evidence of a label switching problem due to employing the MCMC method. This is most likely because, in our normal mixture models, the means of the normal distributions are always specified by the predictions of bias-corrected mass models, which works as an identifiability constraint.

The quantity of interest in our study is the one-neutron separation energy (S_{1n}), which is directly relevant to the r process. This is because in nucleosynthesis calculations (post-processing of hydrodynamical simulations), photodissociation rates (denoted as $\lambda_{(\gamma,n)}$ below), for example, are calculated from the neutron capture rate via detailed balance:

$$\lambda_{(\gamma,n)} = \langle \sigma v \rangle_{(n,\gamma)} \frac{G(N,Z)G(1,0)}{G(N+1,Z)} \left(\frac{A}{A+1}\right)^{3/2} \times \left(\frac{m_u kT}{2\pi\hbar^2}\right)^{3/2} \exp\left(-\frac{S_{1n}(N+1,Z)}{kT}\right), \quad (8)$$

where $\langle \sigma v \rangle_{(n,\gamma)}$ is the velocity-integrated neutron capture cross section for a nucleus with N neutrons and Z protons ($A \equiv N+Z$), $G(N,Z)$ is the partition function for the nucleus (N,Z) ($G(1,0)$ is the partition function for neutron), m_u is the mass of a nucleon, and T is the temperature of the environment. This shows that in addition to astrophysical conditions such as temperature T, the uncertainty in the reverse reaction rate, consequently the uncertainty of the final abundance pattern, arises from the nuclear physics inputs such as the integrated neutron capture cross section $\langle \sigma v \rangle_{(n,\gamma)}$, the partition functions $G(N,Z)$, and one-neutron separation energies S_{1n}. Note that the calculations of other reaction and decay rates are also dependent on the nuclear structure and masses.

The mass models included in our ensemble are the Duflo-Zucker mass model with 29 parameters (DZ29) [7], FRDM2012 [2], HFB31 [10], KUTY05 [44], ETFSI2 [45], and WS4 [6]. These models were chosen based on their popularity in r-process studies as well as the public availability of the data. When multiple versions of the same type of mass model are available, a newer version was selected. The included mass models are largely phenomenological in nature, and little has been investigated regarding uncertainty quantification.

In some of our numerical experiments, we take the S_{1n} values from the AME2020 [18] as experimental data. When evaluating the quality of uncertainty estimates for unseen data, we use the S_{1n} values from the AME2003 [46] for constructing our models (we refer to them as "training data") and then test them with the new data in the AME2020. In the AME2020, 318 new S_{1n} measurements with $Z = 16$–105 are available compared to the AME2003. The new data points in the AME2020 compared to the AME2003 are shown in Fig. 2.

We consider four different ways to categorize the S_{1n} data. The first category is the data for the whole chart of nuclides, which employs all the available experimental data in $Z = 16$–105 at once. The second and third are data for each isotopic and isotonic chain, respectively. This focuses on the evolution

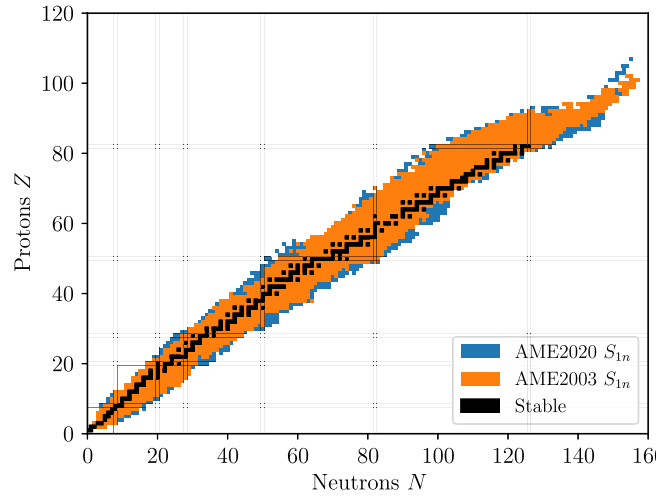

FIG. 2. Comparison of the AME2003 data with the latest AME2020 data for one-neutron separation energies S_{1n}, illustrated on the chart of nuclides. The blue squares show the new S_{1n} data in the AME2020 that did not exist in the AME2003. The S_{1n} values listed in the AME2003 are shown in orange color, aside from the stable nuclides, shown in black.

of the S_{1n} values as a function of proton and neutron number (isotopic and isotonic, respectively). The other category is isobaric (equal mass number A), which is relevant to the trend of β-decay Q values.

The main reason for considering EBMA models for each isotopic, isotonic, or isobaric chain is because it would also allow assessment of how well each mass model performs in different regions of the chart of nuclides. By reviewing the theoretical descriptions of the best performing models, i.e., models with largest weights, we may potentially gain insight into what is required for the more precise modeling of nuclear masses. However, we note that optimizing the parameters to capture the local trend may cause overfitting. Especially, when EBMA models are constructed for each chain across the chart of nuclides, the total number of parameters is significantly more than in the case where a single EBMA model is constructed using the S_{1n} values for the whole chart of nuclides. Therefore, a careful assessment of the quality of the uncertainty estimate is necessary. We will also investigate this aspect in Sec. III.

Except for Sec. III D where the effect of bias correction is investigated, EBMA models are constructed without bias correction (denoted as "raw") for the S_{1n} values predicted by each mass model.

III. RESULTS AND DISCUSSION

A. Comparison with experimental data

To investigate the performance of the EBMA model in reconstructing the experimental S_{1n} values, an EBMA model was constructed using the S_{1n} data of all the nuclei with $16 \leqslant Z \leqslant 105$ from the AME2020 (referred to as the "whole chart EBMA model"). The whole chart EBMA model was constructed without bias correction of the mass models (raw).

TABLE I. 95% posterior highest density intervals (HDI) of the EBMA weights and standard deviations (variances), fitted with all the AME2020 S_{1n} data in $16 \leqslant Z \leqslant 105$ (referred to as "whole chart" in the text). For the columns "Weight" and "Standard deviation", the values in the parenthesis denote an interval with the value on the left being the lower bound and on the right being the upper bound. $\sigma_{\rm RMS}$ [defined in Eq. (9)] shows the root-mean-square error of each mass model with respect to the same AME2020 data. Bias correction of mass models was not performed (raw).

Mass model (raw)	Weight	Standard deviation	$\sigma_{\rm RMS}$ [MeV]
WS4 [6]	(0.459, 0.596)	(0.183, 0.214)	0.239
FRDM12 [2]	(0.143, 0.251)	(0.129, 0.174)	0.312
DZ29 [7]	(0.113, 0.229)	(0.125, 0.245)	0.271
KUTY05 [44]	(0.034, 0.130)	(0.140, 0.322)	0.753
HFB31 [10]	(0.000, 0.027)	(0.183, 0.214)	0.428
ETFSI2 [45]	(0.000, 0.027)	(0.026, 0.761)	0.828

Table I lists the 95% posterior highest density intervals (HDIs), which are the narrowest intervals that include 95% of the posterior distributions, of the whole chart EBMA model. The posterior weight, which can be interpreted as the probability of the model being the best one, is the largest for the WS4 model, followed by FRDM2012 and DZ29. Further analysis of the weights is provided in Sec. III B.

The nominal predictions of the EBMA model are taken as the mode of the posterior predictive distributions of S_{1n}. The posterior distributions of weights and standard deviations σ_k [Eq. (3)] are determined from the AME2020 data through Bayesian inference. Since the AME2020 values are used both for fitting and evaluation of the performance, this analysis reveals how well the EBMA method can reproduce known experimental data using the constituent mass models.

Figure 3 shows the deviations of the modes of the EBMA poterior predictive for S_{1n} from the AME2020 values. The root-mean-square error ($\sigma_{\rm RMS}$) shown in the figure is defined as

$$\sigma_{\rm RMS} = \sqrt{\frac{\sum_{(N,Z)} \left(S_{1n}^{\rm AME2020}(N,Z) - S_{1n}^{\rm Model}(N,Z)\right)^2}{N^{\rm AME2020}}}, \quad (9)$$

where (N, Z) represents pairs of neutron number N and proton number Z of nuclei in the AME2020 whose S_{1n} values are used for the fit. $N^{\rm AME}$ is the total number of such nuclei ($N^{\rm AME2020} \equiv \sum_{(N,Z)}$). For the EBMA model shown in Fig. 3, which took into account the nuclei with $16 \leqslant Z \leqslant 105$, $N^{\rm AME2020} = 2021$. $S_{1n}^{\rm AME2020}(N,Z)$ and $S_{1n}^{\rm Model}(N,Z)$ are the S_{1n} values for a nucleus (N, Z) from the AME2020 and the mode of the posterior predictive distribution of S_{1n} given by a EBMA model (or the nominal prediction of a specific mass model in Table I), respectively.

The value of $\sigma_{\rm RMS}$ of the whole chart EBMA model [Fig. 3(a)] shows a slightly better $\sigma_{\rm RMS}$ (0.229 MeV) than the best performing model, which is the WS4 model [6] with $\sigma_{\rm RMS}^{\rm WS4} = 0.239$ MeV (see Table I). As shown in Fig. 3(b), while the constructed EBMA model is largely based on the WS4 model, the contribution of the other mass models in the ensemble is present throughout the chart of nuclides.

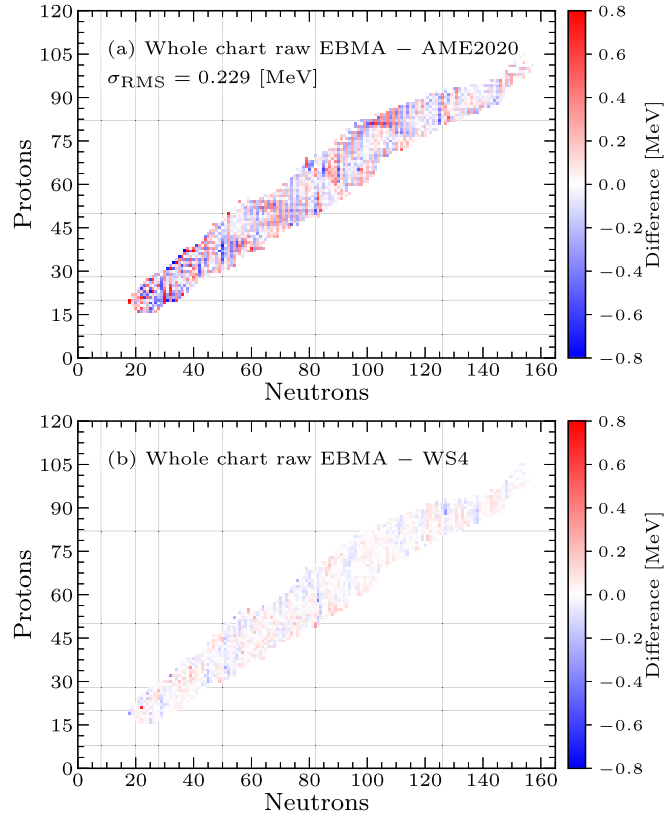

FIG. 3. Deviation and root-mean-square (RMS) error $\sigma_{\rm RMS}$ [MeV] of the neutron separation energies S_{1n} reconstructed by the EBMA model whose parameters are optimized using the whole AME2020 data (Whole chart raw EBMA), compared to the AME2020 data themselves (a), and the differeces of the predictions of the EBMA model from the WS4 model, which is the model in the ensemble with the largest weight (b). "raw" indicates that bias correction was not performed for the models in the ensemble.

B. Parameters in the EBMA models

As shown in Table I, the top three mass models in the size of weight corresponds to the models with the smallest $\sigma_{\rm RMS}$ with respect to the AME2020. Interestingly, while $\sigma_{\rm RMS}$ of the HFB31 model is significantly smaller than KUTY05 or ETFSI2, the assigned weight is one of the smallest. This implies that, although HFB31 can reproduce the experimental S_{1n} values relatively well, the addition of the mass model does not necessarily contribute to improving the overall fit to the experimental data.

Table I also shows that the standard deviations (variances) of the normal distributions in the mixture model do not correspond to the $\sigma_{\rm RMS}$ values. This is because EBMA is a normal mixture model, that is, a weighted sum of normal distributions whose parameters are inferred simultaneously based on the experimental data, while $\sigma_{\rm RMS}$ is calculated separately for each model.

To investigate the weights in EBMA, in addition to the whole chart EBMA model discussed above, EBMA models were further constructed for:

(1) each isotopic chain,

(2) each isotonic chain, and
(3) each isobaric chain,

respectively, using the AME2020 S_{1n} data with the raw mass models.

The colors in Fig. 4 show the mass model with the largest MAP value (maximum *a posteriori*; mode of posterior distribution) of weight within the ensemble for each isotopic [Fig. 4(a)], isotonic [Fig. 4(b)], and isobaric chain [Fig. 4(c)]. The color scale represents the value of the MAP value of the weight. Weight can be understood as the posterior probability of the mass model being the best [19], based on the training using the AME2020 data.

In Fig. 4, especially for the isotopic [Fig. 4(a)] and isotonic [Fig. 4(b)] EBMA models, clustering of the same models in specific regions is visible. For the isotopic EBMA models, it can be seen that the FRDM2012 model tends to have the largest weights just below the proton magic numbers $Z = 50$ and 82. The presence of the DZ29 model is also notable in $59 \leqslant Z \leqslant 71$ for odd proton numbers and in $91 \leqslant 95$. The WS4 model has largest weights at various locations on the chart of nuclides. For the isotonic EBMA models, the most notable trend is the presence of the DZ29 model just above the neutron magic number $N = 82$. It can also be seen that the KUTY05 model is much more present across the chart of nuclides compared to the isotopic case. However, such clustering of specific model is less visible in the isobaric case.

This analysis shows that, by inferring the weights, it is possible to quantify the performance of each mass model in the ensemble in different regions of the chart of nuclides. Some mass models have been shown to perform better than others in reproducing experimental S_{1n} values in specific regions. By further analyzing the well performing models, it may be possible to gain insight into what theoretical description of the nuclei can effectively reliably reproduce and predict nuclear masses in different parts of the chart of nuclides.

C. Uncertainty quantification with EBMA

One of the main goals of this study is to quantify the uncertainty of theoretical S_{1n} values when a variety of mass models are available. EBMA estimates it by creating a weighted average of a collection of mass models based on the performance of each model during the training. In the EBMA model, predictive uncertainty includes not only the spread of the forecasts among the members of the ensemble, but also takes into account the weighted variance of each member model according to the performance during the training [19].

1. Size of uncertainty estimates

Figure 5 shows the size of the 68% HDI, which is roughly analogous to the $\pm 1\sigma$ interval of the normal distribution. EBMA models were fitted for the whole chart of nuclides [Fig. 5(a)], each isotopic chain [Fig. 5(b)], each isotonic chain [Fig. 5(c)], and each isobaric chain [Fig. 5(d)], respectively. Bias correction was not performed for the mass models in the ensemble (raw). Figure 6 also shows the 68% HDIs, focusing on the $Z = 28$ (nickel), $Z = 50$ (tin), and $Z = 64$ (gadolinium) isotopes, compared to the mass models (gray dashed lines) shown in Fig. 1, relative to the nominal predictions of the FRDM2012 model. Figures 6(a)–6(c) show the EBMA models fitted for the whole chart of nuclides and each isotopic

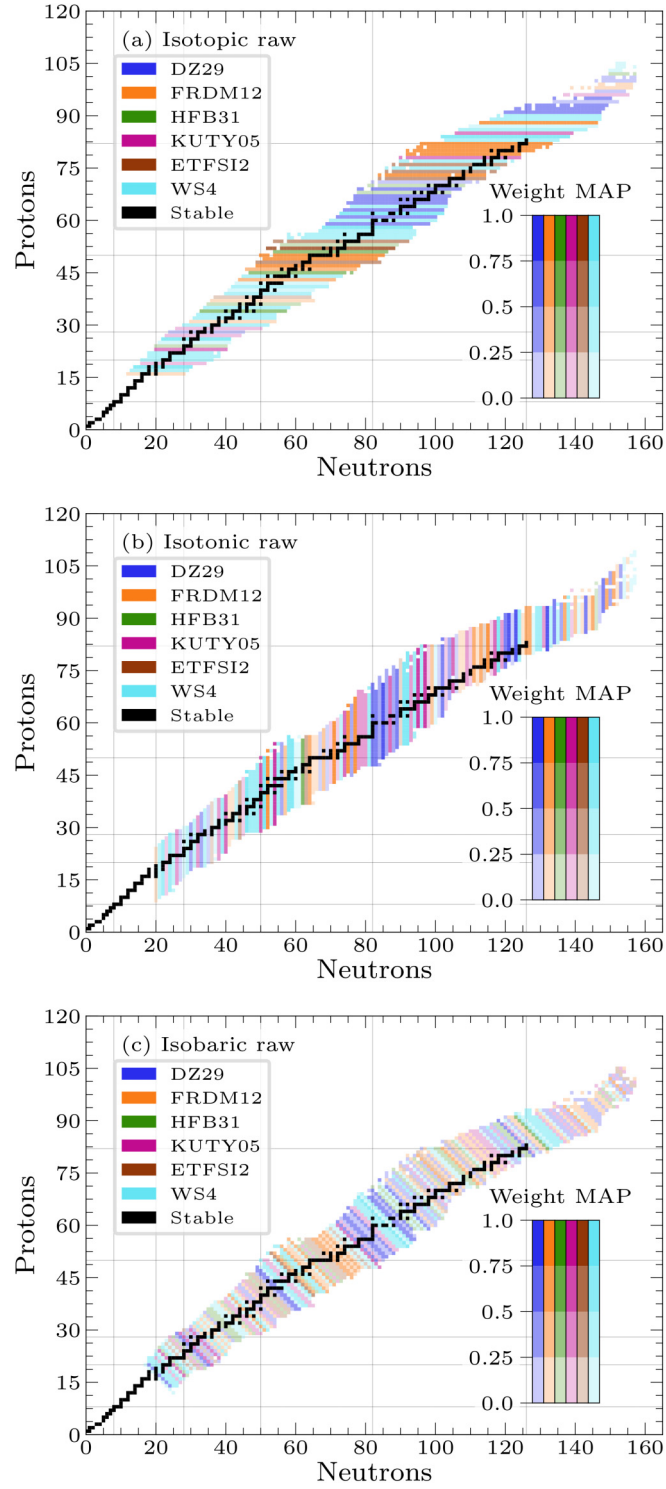

FIG. 4. Maximum *a posteriori* (MAP) values of the largest weight in the EBMA ensemble determined from the AME2020 data for each (a) isotopic chain, (b) isotonic chain, and (c) isobaric chain. "raw" indicates that bias correction was not performed for the models in the ensemble.

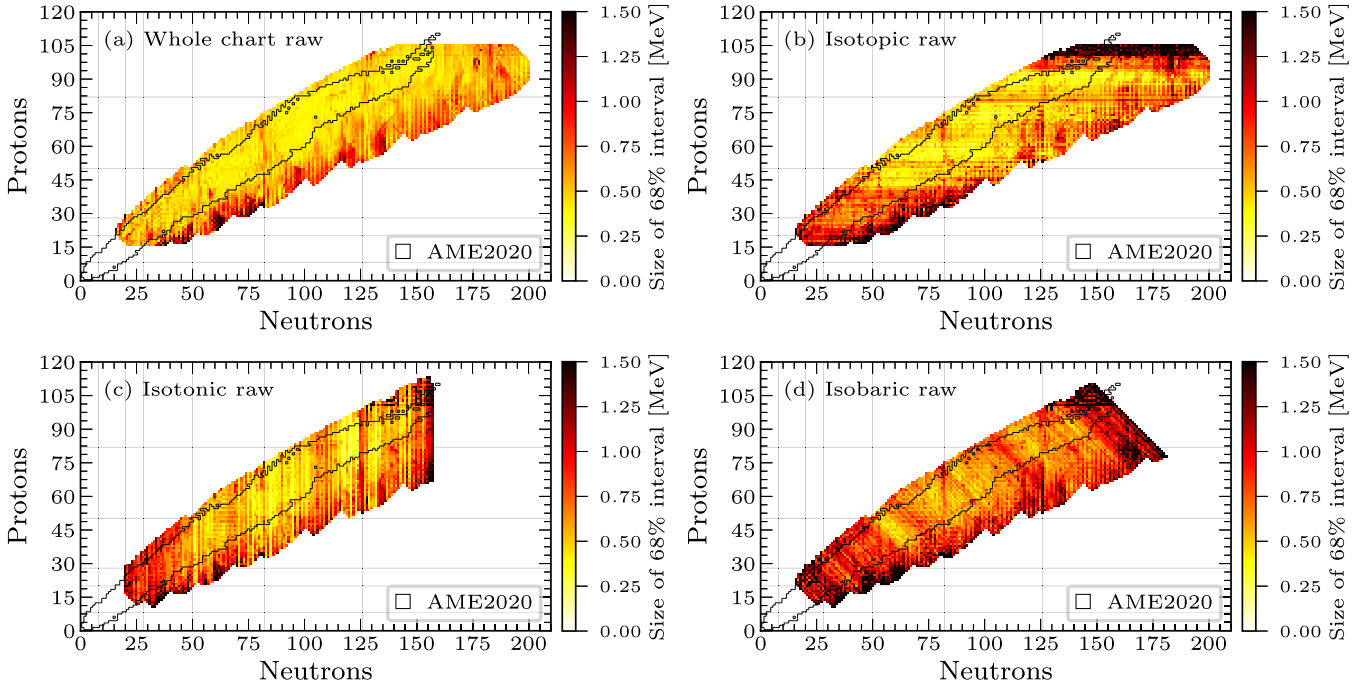

FIG. 5. The sizes of 68% highest density intervals (HDIs) of the EBMA models across the chart of nuclides, fitted with the AME2020 S_{1n} data for (a) the whole chart of nuclides, (b) each isotopic chain, (c) each isotonic chain, and (d) each isobaric chain. The area enclosed by the black contour shows where the AME2020 data exist. The charts with isotonic and isobaric fits are truncated at large neutron number and mass number, respectively, because there are not enough data points within the chains to determine the EBMA parameters. "raw" indicates that bias correction was not performed for the models in the ensemble.

chain and Figs. 6(d)–6(f) show the models for each isotopic chain and each isobaric chain for the same set of nuclei. The fits are performed using the AME2020 data, and predictions are made for all the nuclei available in all the member mass models within the ensemble. In all cases, it can be seen that the size of the uncertainty is more constrained where the data exist (black contour in Fig. 5 and black triangles in Fig. 6), but increases towards the edge of the chart of nuclides. This is especially visible in the neutron-rich direction, where many of the nuclear masses are yet to be experimentally measured. This is a reflection of the fact that the predictions of the mass models that make up the ensemble start to diverge as we move further away from the last data point, as shown in Fig. 1.

Comparing the four plots in Fig. 5, the increase in the size of uncertainty in the neutron-rich region is the smallest for the fit using the whole chart of nuclides [Fig. 5(a)]. This is because the weights for the whole chart EBMA model are determined using all available data, whereas for each isotopic, isotonic, and isobaric chain, the weights are determined only from the data in each chain. The smaller number of training data points (experimental S_{1n} values) in each chain also results in larger uncertainties for reproducing the experimental values, compared to the fit with the whole chart data.

As shown in Fig. 6, for some isotopic chains, the uncertainty estimates are different between different types of fits. Especially, for $Z = 50$ isotopes, it can be seen in Fig. 6(b) that the uncertainty estimated by the whole chart EBMA model constantly falls below 0 (the predictions by FRDM2012) beyond $N = 90$, while the isotopic EBMA model is dominated by FRDM2012. It is not yet possible to experimentally test the predictions for such neutron-rich nuclei. However, differences in the predictions in neutron-rich regions may have a significant impact on the prediction of observables from the *r* process such as the abundance pattern.

2. *Quality of uncertainty estimates*

We now investigate the quality of the uncertainty estimate. For this purpose, we construct EBMA models and quantify the prediction uncertainties using the S_{1n} data for nuclei with $16 \leqslant Z \leqslant 105$ from the AME2003 [46] (we will refer to the data as "training data"), then evaluate the quality of the uncertainties based on the new S_{1n} data in the AME2020 [18]. In the AME2020, there are 318 S_{1n} values newly registered since the AME2003. Bias correction was not performed for the mass models in the ensemble (raw).

Figure 7 shows the distribution of the new S_{1n} data relative to the sizes of uncertainties given by the EBMA models fitted with the data for the whole chart of nuclides [Fig. 7(a)], each isotopic chain [Fig. 7(b)], isotonic chain [Fig. 7(c)], and isobaric chain [Fig. 7(d)]. Note that the number of new data points included in the fit (n in Fig. 7) is not necessarily 318 for some cases. This is because at the edge of the chart of nuclides, there are fewer experimental data points and the posterior weights do not converge for some chains. Such chains were excluded from the fits. The summary of the analysis is shown in the upper half of Table II. In the following, we describe our analysis and findings.

The size of the estimated uncertainty is represented by the 68% highest density interval ($\mathrm{HDI}_{0.68}$), which is the narrowest

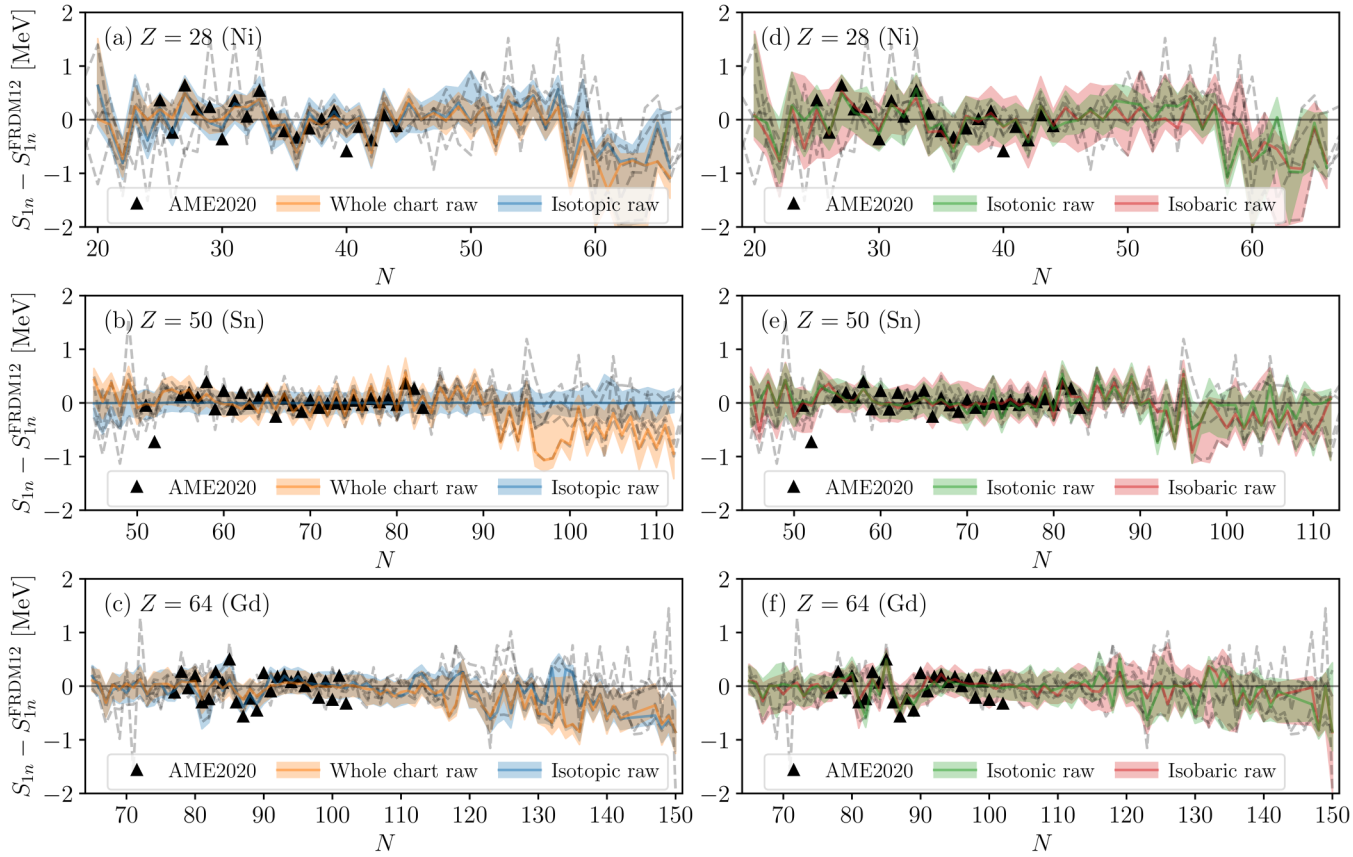

FIG. 6. Trends of the sizes of 68% highest density intervals of the EBMA models compared to the predictions by the FRDM2012 model for the S_{1n} values in the $Z = 28$ (Ni) (a), (d), $Z = 50$ (Sn) (b), (e), and $Z = 64$ (Gd) (c), (f) isotopic chains, similar to Fig. 1. Panels (a)–(c) show the EBMA models fitted with the AME2020 S_{1n} data for the whole chart of nuclides (orange) and each isotopic chain (blue), and panels (d)–(f) show the models for each isotonic chain (green) and each isobaric chain (red). The solid lines associated with the bands show the modes of the predictive distributions. The gray dashed lines are the mass models shown in Fig. 1. "raw" indicates that bias correction was not performed for the models in the ensemble.

interval that contains 68% of the distribution, of the posterior predictive distribution of the S_{1n} values. To investigate how the new experimental S_{1n} values are distributed relative to the predicted $\mathrm{HDI}_{0.68}$, we define δ, which represents an experimental S_{1n} value normalized by the size of $\mathrm{HDI}_{0.68}$. Let h^{low} and h^{up} represent the lower and upper boundaries of the $\mathrm{HDI}_{0.68}$, respectively, then

$$\delta = \frac{S_{1n} - h^{\mathrm{low}}}{h^{\mathrm{up}} - h^{\mathrm{low}}} - 0.5, \tag{10}$$

where 0.5 is subtracted to symmetrize the distribution around 0. This means that $\delta = -0.5$ and 0.5 correspond to the S_{1n} values at the lower and upper boundaries of the $\mathrm{HDI}_{0.68}$, respectively.

$\overline{\mathrm{HDI}_{0.68}^{2020^*}}$ in Fig. 7 and Table II shows the average size of the 68% intervals for the new data in the AME2020 (denoted as 2020*). On average, the whole chart EBMA model provides the most constrained size of the uncertainty of 0.480 MeV. Table II also shows the root-mean-square errors of the modes of the posterior predictive distributions of S_{1n} given by the EBMA models from the experimental S_{1n} values from the AME2003 ($\sigma_{\mathrm{RMS}}^{2003}$), similarly to Eq. (9). While the models fitted for each isotopic/isotonic/isobaric chain exhibit smaller $\sigma_{\mathrm{RMS}}^{2003}$ values, the sizes of $\overline{\mathrm{HDI}_{0.68}^{2020^*}}$ are larger than the whole chart EBMA model. This is because the EBMA models are constructed and fine-tuned for each chain, which also means that the total number of parameters across the chart is significantly larger than the fitting of the whole chart with a single EBMA model. At the same time, the parameters for each chain are inferred from a smaller number of data points compared to the whole chart EBMA model; therefore, the parameters are less sharply determined. This suggests that the large total number of parameters of the EBMA models constructed for individual chains across the chart of nuclides may lead to overfitting. However, also considering the associated uncertainty, the overfitting may partially be mitigated by the increased uncertainty through the Bayesian framework. We will further investigate this issue below.

Insets of Fig. 7 show that the distribution of the training S_{1n} values from the AME2003 around the center of $\mathrm{HDI}_{0.68}$ approximately follows a normal distribution. Performing a similar fit to the new data in the AME2020, we obtain the approximate 68% interval of the distribution of the new experimental S_{1n} values in the AME2020 around the center of

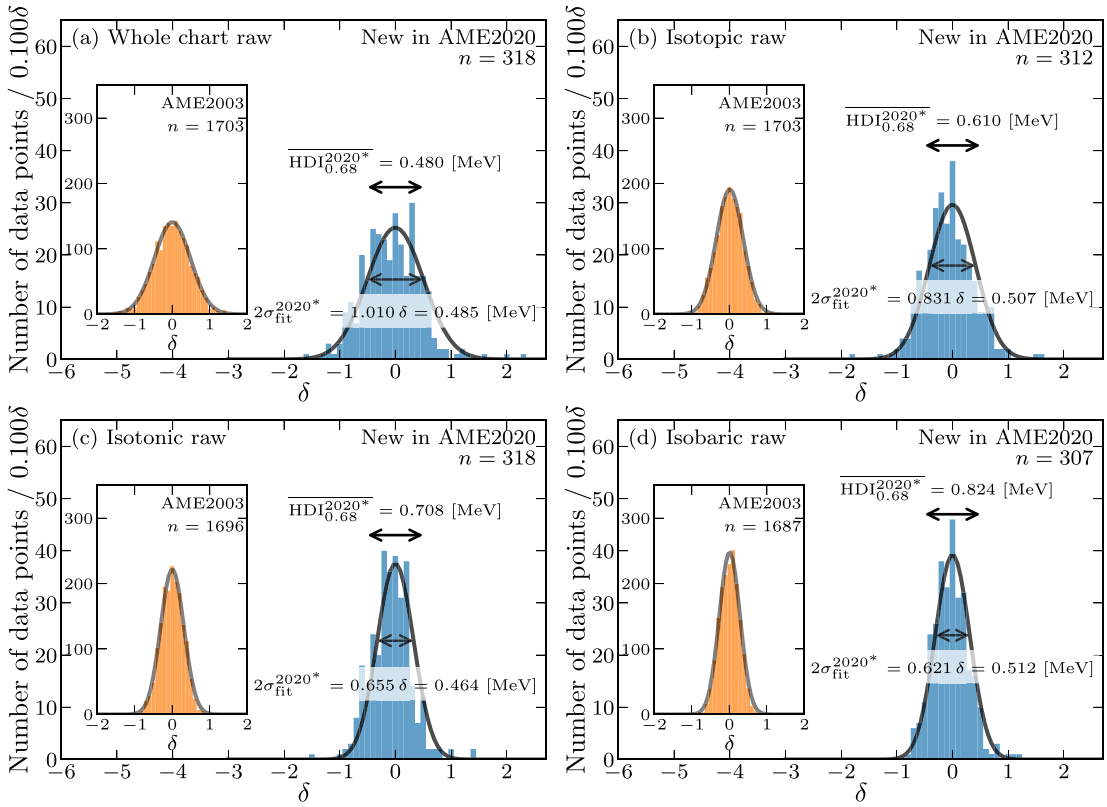

FIG. 7. Distributions of the new data points in the AME2020 compared to the AME2003, with respect to the 68% HDIs predicted by the EBMA models fitted with the AME2003 data for (a) the whole chart of nuclides, (b) each isotopic chain, (c) each isotonic chain, and (d) each isobaric chain. The definition of δ is given in Eq. (10). "raw" indicates that bias correction was not performed for the models in the ensemble.

$\mathrm{HDI}_{0.68}$, which is denoted as $2\sigma^{2020*}_{\mathrm{fit}}$. Comparing the sizes of $\overline{\mathrm{HDI}^{2020*}_{0.68}}$ and $2\sigma^{2020*}_{\mathrm{fit}}$ for the whole chart EBMA model, it can be seen that the sizes are comparable, which shows that this EBMA model can appropriately estimate the uncertainty for the new data. However, the sizes of $\overline{\mathrm{HDI}^{2020*}_{0.68}}$ for the other EBMA models are larger than $2\sigma^{2020*}_{\mathrm{fit}}$, indicating that the uncertainty estimates are inflated (underconfident). The tendency of inflated uncertainty for the EBMA with individual chains can also be seen from the comparison of the corresponding quantities for the AME2003 training data ($\overline{\mathrm{HDI}^{2003}_{0.68}}$ and $2\sigma^{2003}_{\mathrm{fit}}$), shown in Table II.

This observation is further supported by looking at the ratios of the S_{1n} values that fall within $\mathrm{HDI}_{0.68}$ for the AME2003 and the new data in the AME2020, also shown in Table II

TABLE II. Quantities that characterize the uncertainty estimates and their quality for different EBMA models, which are fitted to the S_{1n} data for nuclei with $16 \leqslant Z \leqslant 105$ from the AME2003. σ_{RMS} is the root-mean-square error calculated similarly to Eq. (9), $\overline{\mathrm{HDI}_{0.68}}$ is the average size of the 68% highest density intervals, $2\sigma_{\mathrm{fit}}$ is the spread of the experimental S_{1n} values around the center of $\mathrm{HDI}_{0.68}$ approximated by a normal distribution, and the $\mathrm{HDI}_{0.68}$ coverage is the ratio of experimental S_{1n} values that fall within $\mathrm{HDI}_{0.68}$. The superscripts "2003" and "2020*" denote the AME2003 data and the new data in the AME2020 (absent in the AME2003), respectively. The upper half shows the results without bias correction (raw) and the lower half with bias correction (corrected).

EBMA model	$\sigma^{2003}_{\mathrm{RMS}}$ [MeV]	$\overline{\mathrm{HDI}^{2003}_{0.68}}$ [MeV]	$2\sigma^{2003}_{\mathrm{fit}}$ [MeV]	$\mathrm{HDI}^{2003}_{0.68}$ coverage [%]	$\overline{\mathrm{HDI}^{2020*}_{0.68}}$ [MeV]	$2\sigma^{2020*}_{\mathrm{fit}}$ [MeV]	$\mathrm{HDI}^{2020*}_{0.68}$ coverage [%]
Whole chart (raw)	0.239	0.466	0.448	71.5	0.480	0.485	69.5
Isotopic (raw)	0.218	0.562	0.401	85.1	0.610	0.507	75.3
Isotonic (raw)	0.201	0.598	0.368	92.5	0.708	0.464	82.7
Isobaric (raw)	0.219	0.746	0.411	95.6	0.824	0.512	89.6
Whole chart (corrected)	0.238	0.466	0.449	71.3	0.480	0.490	70.1
Isotopic (corrected)	0.206	0.517	0.373	84.4	0.706	0.664	72.4
Isotonic (corrected)	0.134	0.370	0.219	89.8	0.658	0.627	68.3
Isobaric (corected)	0.170	0.618	0.310	95.4	0.804	0.614	82.1

as the $\mathrm{HDI}_{0.68}^{2003}$ coverage and the $\mathrm{HDI}_{0.68}^{2020^*}$ coverage, respectively. Except for the whole chart EBMA model, the coverages significantly exceed 68% for both data sets. Comparing the coverages of $\mathrm{HDI}_{0.68}^{2003}$ and $\mathrm{HDI}_{0.68}^{2020^*}$, except for the whole chart EBMA, the coverages of $\mathrm{HDI}_{0.68}^{2020^*}$ are smaller than $\mathrm{HDI}_{0.68}^{2003}$. This points to the fact that the quality of uncertainty estimates does not extrapolate well for the EBMA models constructed for individual chains. In other words, this can be seen as a symptom of overfitting to the training data.

From this analysis, it can be concluded that the uncertainty estimates of the whole chart EBMA have desirable properties and extrapolate well to the new data in the AME2020. However, the EBMA models constructed for each isotopic/isotonic/isobaric chain tend to overfit the AME2003 data and at the same time provide underconfident uncertainty estimates. However, it is intriguing that the isotonic EBMA model resulted in the smallest values of $\sigma_{\mathrm{RMS}}^{2003}$, $\sigma_{\mathrm{fit}}^{2003}$, and $\sigma_{\mathrm{fit}}^{2020^*}$. This suggests the possibility that focusing on the isotonic chain may potentially provide a reliable way to extrapolate the S_{1n} values if overfitting can be avoided and the size of the uncertainty estimate can be appropriately controlled. We note that the new data in the AME2020 lie just outside the AME2003 data, as shown in Fig. 2. Therefore, these results may not hold in a more neutron-rich region. Further development of the EBMA method will be carried out to put more emphasis on the neutron-rich nuclei. Without abundant data in the neutron-rich region, alternative constraints may also be necessary, e.g., the *r*-process observables, which are known to involve neutron-rich nuclei.

D. Bias corrected EBMA models

As discussed in Sec. II B, Ref. [19] suggests a linear bias correction of each model, prior to inference of the EBMA weights and standard deviations. This means determining the bias correction coefficients a_k and b_k for each model, where a_k is the intercept (offset) and b_k is the slope. $a_k + b_k m_k$ is then the prediction of the bias-corrected model based on the mass model k. In our case, m_k is a vector of theoretical S_{1n} values for the whole chart of nuclides or a specific isotopic/isotonic/isobaric chain, predicted by the model k.

The simplest approach to determine a_k and b_k is to perform linear regression, as shown in Ref. [19]. Another way to determine these parameters a_k and b_k is by Bayesian linear regression and taking the *maximum a posteriori* (MAP) values, which would be a slightly more probabilistic treatment. In our case, the two approaches yield virtually identical values. The coefficients determined for each mass model using the entire AME2003 data are shown in Table III, for example. It can be seen that for some mass models, the coefficients a_k deviate from 0.

The summary of the models that quantify the quality of the uncertainty estimates is shown in the lower half of Table II. As expected with further calibration (bias correction) of the mass models in the ensemble, the values of $\sigma_{\mathrm{RMS}}^{2003}$ decreased compared to the raw EBMA models. For the bias-corrected whole chart EBMA model, since the corrections made by the coefficients for the dominating mass models (WS4, FRDM2012, and DZ29) are small, as shown in Table III, the difference from the raw model is smaller compared to other EBMA models. For the bias-corrected EBMA models for individual chains, indications of inflated uncertainty estimates are still present, following the same argument given for the raw EBMA models in the previous section.

TABLE III. Bias-correction coefficients determined for different mass models from the MAP values of Bayesian linear regression (labeled as "BLR MAP") and usual linear regression (labeled as "LR"), respectively, using all of the available (whole chart) one-neutron separation energy (S_{1n}) data from the AME2003 [46]. a_k and b_k are the coefficients for offset (intercept) and slope, respectively, for mass model k. The values are rounded to two digits.

	a_k		b_k	
Mass model	BLR MAP	LR	BLR MAP	LR
WS4	−0.01	−0.01	1.00	1.00
FRDM12	0.12	0.12	0.99	0.99
DZ29	−0.10	−0.10	1.01	1.01
KUTY05	−0.06	−0.06	1.01	1.01
ETFSI2	0.38	0.38	0.95	0.95
HFB31	0.30	0.30	0.96	0.96

A notable difference from the raw EBMA models is the more significant decrease in $\mathrm{HDI}_{0.68}^{2020^*}$ coverage from the $\mathrm{HDI}_{0.68}^{2003}$ coverage for the EBMA models for individual chains, indicating further overfitting problems.

In conclusion, since the linear bias correction further facilitates overfitting in the EBMA models for individual chains, which was already not as reliable as the whole chart EBMA model as discussed in Sec. III C, and does not have a significant effect on the whole chart EBMA model, the use of linear bias correction would not be recommended.

E. Effect of the choice of mass models in the ensemble

So far we have only considered the ensemble consisting of the mass models listed in Table I, which are some of the most popular models in the *r*-process studies. Nevertheless, it

TABLE IV. 95% posterior highest density intervals (HDI) of the EBMA weights and standard deviations (variances) without the WS4 model in the ensemble, fitted with all the AME2020 S_{1n} data in $16 \leqslant Z \leqslant 105$ (whole chart), similarly to Table I. For the columns "Weight" and "Standard deviation", the values in the parenthesis denote an interval with the value on the left being the lower bound and on the right being the upper bound. σ_{RMS} [defined in Eq. (9)] shows the root-mean-square error of each mass model with respect to the same AME2020 data, identical to Table I. Bias correction of mass models was not performed (raw).

Mass model (raw)	Weight	Standard deviation	σ_{RMS} [MeV]
DZ29	(0.360, 0.484)	(0.207, 0.248)	0.271
FRDM12	(0.273, 0.383)	(0.144, 0.175)	0.312
KUTY05	(0.119, 0.230)	(0.158, 0.243)	0.753
HFB31	(0.030, 0.089)	(0.128 0.264)	0.428
ETFSI2	(0.002, 0.030)	(0.029, 0.663)	0.828

TABLE V. Similar table to Table II, but for the whole chart EBMA model based on the ensemble without WS4. It shows the quantities that characterize the uncertainty estimates and their quality for different EBMA models, which are fitted to the S_{1n} data for nuclei with $16 \leqslant Z \leqslant 105$ from the AME2003. $\sigma_{\rm RMS}$ is the root-mean-square error calculated similarly to Eq. (9), $\overline{\rm HDI_{0.68}}$ is the average size of the 68% highest density intervals, $2\sigma_{\rm fit}$ is the spread of the experimental S_{1n} values around the center of $\rm HDI_{0.68}$ approximated by a normal distribution, and the $\rm HDI_{0.68}$ coverage is the ratio of experimental S_{1n} values that fall within $\rm HDI_{0.68}$. The superscripts "2003" and "2020*" denote the AME2003 data and the new data in the AME2020 (absent in the AME2003), respectively.

EBMA model (no WS4)	$\sigma_{\rm RMS}^{2003}$ [MeV]	$\overline{\rm HDI_{0.68}^{2003}}$ [MeV]	$2\sigma_{\rm fit}^{2003}$ [MeV]	$\rm HDI_{0.68}^{2003}$ coverage [%]	$\overline{\rm HDI_{0.68}^{2020*}}$ [MeV]	$2\sigma_{\rm fit}^{2020*}$ [MeV]	$\rm HDI_{0.68}^{2020*}$ coverage [%]
Whole chart (raw)	0.257	0.495	0.479	71.5	0.506	0.542	66.7

is possible to consider ensembles of any combination of mass models, and it is important to understand how the choice of models in the ensemble affects the uncertainty estimates.

In the case where an ensemble of mass models includes models with poor performance in reproducing the experimental values, in general, one can expect that such models will be assigned small weights, as seen for the weight of ETFSI2 shown in Table II, for example. This is because the likelihood [Eq. (5)] evaluates how well each mass model reproduces the experimental data. Models with small weights have a small effect on quantified uncertainty, as models in the ensemble have contributions to the uncertainty proportional to the weights, as shown in the expression of predictive variance [Eq. (7)].

As for the removal of models with large weights, the effect is not trivial. Since the WS4 model has the largest contribution to the whole chart EBMA model we have considered so far, we will investigate how the EBMA model is affected by removing the WS4 model from the ensemble. From this we aim to gain insight into the effect of the choice of mass models. Based on the quality of uncertainty estimates investigated in Secs. III C 2 and III D, we limit ourselves to the whole chart EBMA model without bias correction.

In order to compare with the inferred EBMA parameters shown in Table I, a new whole chart EBMA model was constructed using the same AME2020 data in $16 \leqslant Z \leqslant 105$, using an ensemble without the WS4 model. The 95% HDIs for the posterior weights and standard deviations are shown in Table IV, in order of the size of the posterior weights. It can be seen that the assignment of the weights is somewhat more democratic, without the dominant WS4 model. It is also notable that the DZ29 model now has the largest weight, while it previously had the third largest weights in the ensemble with WS4. This implies that DZ29 and WS4 may have provided somewhat similar predictions of S_{1n}, while the performance of WS4 is consistently better with respect to the experimental data; therefore, WS4 had absorbed the possible contribution of DZ29.

Similarly to the Sec. III C 2, we now construct the whole chart EBMA model with the AME2003 S_{1n} data and test the quality of uncertainty estimates using the new data in the AME2020. The summary of performance is shown in Table V, again focusing on 68% highest density intervals ($\rm HDI_{0.68}$) of the predictive distributions of S_{1n}. In terms of reproducing the experimental values in the AME2003 (the training data), $\sigma_{\rm RMS}^{2003}$ is 0.257 MeV, increasing from the 0.239 MeV with the ensemble with WS4. It is still an improvement over individual mass models, since the root-mean-square error of DZ29 is 0.274 MeV, which is the smallest in the current ensemble. Although the sizes of both $\overline{\rm HDI_{0.68}^{2003}}$ and $2\sigma_{\rm fit}^{2003}$ are also about 0.3 MeV larger than those with the ensemble with WS4, the coverage of $\rm HDI_{0.68}^{2003}$ is the same (71.5%), showing that the quality of the uncertainty estimates for the training data is still adequate.

We now look at the quality of uncertainty estimates for the new data in the AME2020. A notable difference from the case with the ensemble with WS4 is that $2\sigma_{\rm fit}^{2020^*}$, the average size of the deviation from the center of $\rm HDI_{0.68}^{2020^*}$, is significantly larger than the size of $\overline{\rm HDI_{0.68}^{2020^*}}$, resulting in the coverage of $\rm HDI_{0.68}^{2020^*}$ of 66.7%. The decrease in coverage from the AME2003 data is also greater than in the case with the ensemble with WS4, suggesting that the performance of extrapolation in uncertainty estimates is slightly degraded. However, considering that the ideal coverage is 68%, overall quality of the uncertainty estimates is still decent.

We again note that the new data in the AME2020 lie just outside of the available data in the AME2003, therefore, the current results may not apply to the extrapolation to the more neutron-rich region. With this in mind, the current results show that, while the removal of the best-performing mass model WS4 slightly degrades the performance in extrapolating the uncertainty estimates of S_{1n}, the whole chart EBMA model still provides decent uncertainty estimates. This in turn suggests that, although a small root-mean-square error does not necessarily mean a large contribution to the ensemble, as shown from the weights of HFB31, including performant mass models would likely contribute to constraining the size of uncertainty and improving the quality in extrapolation of uncertainty estimates. It would also be recommended to include a wide range of models in the ensemble, even when some of the models are similar in nature, since it is difficult to guess the possible contributions of each model. If there are models with similar but less performant predictions, then their contributions should automatically be absorbed by the more performant models.

IV. CONCLUSIONS

We have explored the possibility of quantifying the uncertainty of theoretical one-neutron separation energies (S_{1n}) when a variety of mass models are available, using the EBMA method. The EBMA method creates a weighted average of theoretical mass models as a mixture of normal distributions, whose weights and standard deviations are estimated

by MCMC using the NUTS from experimental S_{1n} data. The resulting weights and standard deviations are expressed as distributions. The distributions of the EBMA parameters are then used to estimate the uncertainty in S_{1n}.

EBMA models have been constructed in four different ways of fitting the experimental data, namely, the whole chart of nuclides, each isotopic chain, each isotonic chain, and each isobaric chain. The mass models included in the ensemble were WS4, DZ29, KUTY05, HFB31, and ETFSI2, which are some of the most popular choices of mass models in r-process studies. For the whole chart EBMA model, WS4 was shown to have a dominant contribution.

Quality of uncertainty estimates has been examined by constructing EBMA models with the data from the AME2003 and testing the extrapolation performance on the new data from the AME2020. Our analysis shows that the whole chart EBMA model provides reliable uncertainty estimates.

The EBMA models constructed with the data for each isotopic, isotonic, and isobaric chain were shown to exhibit signs of underconfidence and overfitting. However, the weights in the EBMA models for individual chains, especially the isotopic and isotonic EBMA models, show which mass models perform well in different parts of the chart of nuclides, potentially giving insight into what theoretical descriptions of nuclei are effective in different regions.

Effect of linear bias correction of each mass model was examined, which was found to facilitate the overfitting of the EBMA models constructed for individual chains, and has little effect on the whole chart EBMA model. Therefore, linear bias correction is not recommended in general.

Effect of choices of mass models included in the ensemble was investigated by removing WS4, which was the dominant model in the original ensemble. It was found that the quality of uncertainty estimates provided by the whole chart EBMA model slightly degrades, but still decent. On the basis of the results and the general structure of the EBMA, it is recommended that a wide range of mass models be included for more reliable and constrained extrapolation of uncertainty estimates.

We note the performance of EBMA on extrapolation of uncertainty estimates has only been tested with the new data in the AME2020, which lie just outside the AME2003 data. In order for the EBMA method to gain more validity in the neutron-rich region, improvements of the method that put more emphasis on the neutron-rich region will be necessary. Alternatively, inferring the weights from the r-process observables may be considered, since the r process is known to involve extremely neutron rich nuclei.

ACKNOWLEDGMENTS

We thank S. Yoshida and D. R. Phillips for helpful feedback on the manuscript. We also thank the anonymous referee for their thorough and constructive feedback, which contributed to the improvement of the manusctipt. Y.S. and I.D. acknowledge funding from the Canadian Natural Sciences and Engineering Research Council (NSERC), NSERC Discovery Grants No. SAPIN-2019-00030, No. SAPPJ-2017-00026, and the NSERC CREATE Program IsoSiM (Isotopes for Science and Medicine). R.K. is supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Contract No. DE-AC02-05CH11231. M.R.M. is supported by the U.S. Department of Energy through the Los Alamos National Laboratory (LANL). LANL is operated by Triad National Security LLC for the National Nuclear Security Administration of the U.S. Department of Energy (Contract No. 89233218CNA000001). Y.S. and R.S. acknowledges support from the U.S. National Science Foundation under Grant No. 21-16686 (NP3M). R.S. acknowledges support from the U.S. Department of Energy Contract No. DE-FG02-95-ER40934.

[1] P. Moller, J. Nix, W. Myers, and W. Swiatecki, At. Data Nucl. Data Tables **59**, 185 (1995).
[2] P. Möller, A. Sierk, T. Ichikawa, and H. Sagawa, At. Data Nucl. Data Tables **109-110**, 1 (2016).
[3] N. Wang, M. Liu, and X. Wu, Phys. Rev. C **81**, 044322 (2010).
[4] N. Wang, Z. Liang, M. Liu, and X. Wu, Phys. Rev. C **82**, 044304 (2010).
[5] M. Liu, N. Wang, Y. Deng, and X. Wu, Phys. Rev. C **84**, 014333 (2011).
[6] N. Wang, M. Liu, X. Wu, and J. Meng, Phys. Lett. B **734**, 215 (2014).
[7] J. Duflo and A. P. Zuker, Phys. Rev. C **52**, R23 (1995).
[8] J. Erler, N. Birge, M. Kortelainen, W. Nazarewicz, E. Olsen, A. M. Perhac, and M. Stoitsov, Nature (London) **486**, 509 (2012).
[9] S. Goriely, S. Hilaire, M. Girod, and S. Péru, Phys. Rev. Lett. **102**, 242501 (2009).
[10] S. Goriely, N. Chamel, and J. M. Pearson, Phys. Rev. C **93**, 034337 (2016).
[11] J. d. J. Mendoza-Temis, M.-R. Wu, K. Langanke, G. Martínez-Pinedo, A. Bauswein, and H.-T. Janka, Phys. Rev. C **92**, 055805 (2015).
[12] D. Martin, A. Arcones, W. Nazarewicz, and E. Olsen, Phys. Rev. Lett. **116**, 121101 (2016).
[13] M. Mumpower, R. Surman, G. McLaughlin, and A. Aprahamian, Prog. Part. Nucl. Phys. **86**, 86 (2016).
[14] Y. L. Zhu, K. A. Lund, J. Barnes, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower, and R. Surman, Astrophys. J. **906**, 94 (2021).
[15] J. Barnes, Y. L. Zhu, K. A. Lund, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower, and R. Surman, Astrophys. J. **918**, 44 (2021).
[16] J. Dobaczewski, W. Nazarewicz, and P.-G. Reinhard, J. Phys. G: Nucl. Part. Phys. **41**, 074001 (2014).
[17] J. D. McDonnell, N. Schunck, D. Higdon, J. Sarich, S. M. Wild, and W. Nazarewicz, Phys. Rev. Lett. **114**, 122501 (2015).
[18] M. Wang, W. Huang, F. Kondev, G. Audi, and S. Naimi, Chin. Phys. C **45**, 030003 (2021).
[19] A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski, Monthly Weather Rev. **133**, 1155 (2005).
[20] Z. M. Niu and H. Z. Liang, Phys. Rev. C **106**, L021303 (2022).
[21] Z. Niu and H. Liang, Phys. Lett. B **778**, 48 (2018).
[22] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Phys. Rev. C **98**, 034318 (2018).

[23] L. Neufcourt, Y. Cao, W. Nazarewicz, E. Olsen, and F. Viens, Phys. Rev. Lett. **122**, 062502 (2019).
[24] R.-D. Lasseri, D. Regnier, J.-P. Ebran, and A. Penon, Phys. Rev. Lett. **124**, 162502 (2020).
[25] R. Utama, J. Piekarewicz, and H. B. Prosper, Phys. Rev. C **93**, 014311 (2016).
[26] A. E. Lovell, A. T. Mohan, T. M. Sprouse, and M. R. Mumpower, Phys. Rev. C **106**, 014305 (2022).
[27] M. R. Mumpower, T. M. Sprouse, A. E. Lovell, and A. T. Mohan, Phys. Rev. C **106**, L021301 (2022).
[28] M. Mumpower, M. Li, T. M. Sprouse, B. S. Meyer, A. E. Lovell, and A. T. Mohan, Front. Phys. **11**, 1198572 (2023).
[29] V. Kejzlar, L. Neufcourt, and W. Nazarewicz, Sci. Rep. **13**, 19600 (2023).
[30] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, Stat. Sci. **14**, 382 (1999).
[31] A. Gelman, D. Simpson, and M. Betancourt, Entropy **19**, 555 (2017).
[32] A. Hamaker, E. Leistenschneider, R. Jain, G. Bollen, S. A. Giuliani, K. Lund, W. Nazarewicz, L. Neufcourt, C. R. Nicoloff, D. Puentes, R. Ringle, C. S. Sumithrarachchi, and I. T. Yandow, Nat. Phys. **17**, 1408 (2021).
[33] L. Neufcourt, Y. Cao, S. A. Giuliani, W. Nazarewicz, E. Olsen, and O. B. Tarasov, Phys. Rev. C **101**, 044307 (2020).
[34] L. Neufcourt, Y. Cao, S. Giuliani, W. Nazarewicz, E. Olsen, and O. B. Tarasov, Phys. Rev. C **101**, 014319 (2020).
[35] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, Cambridge, MA, 2005)
[36] S. Yoshida, Phys. Rev. C **102**, 024305 (2020).
[37] 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: Nucl. Part. Phys. **48**, 072001 (2021).
[38] A. C. Semposki, R. J. Furnstahl, and D. R. Phillips, Phys. Rev. C **106**, 044002 (2022).
[39] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, PeerJ Comput. Sci. **2**, e55 (2016).
[40] R. M. Neal, *Bayesian Learning for Neural Networks*, Vol. 118 (Springer Science & Business Media, Berlin, 2012).
[41] R. M. Neal, *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. Jones, and X.-L. Meng (Chapman and Hall/CRC, New York, 2011), pp. 113–162.
[42] M. Stephens, J. R. Stat. Soc. B **62**, 795 (2000).
[43] A. Jasra, C. C. Holmes, and D. A. Stephens, Stat. Sci. **20**, 50 (2005).
[44] H. Koura, T. Tachibana, M. Uno, and M. Yamada, Prog. Theor. Phys. **113**, 305 (2005).
[45] Y. Aboussir, J. Pearson, A. Dutta, and F. Tondeur, At. Data Nucl. Data Tables **61**, 127 (1995).
[46] G. Audi, A. Wapstra, and C. Thibault, Nucl. Phys. A **729**, 337 (2003).