Uncertainty Quantification of Mass Models using Ensemble Bayesian Model Averaging

Yukiya Saito*

TRIUMF, 4004 Wesbrook Mall, Vancouver, BC V6T 2A3, Canada and
Department of Physics and Astronomy, The University of British Columbia, Vancouver, BC V6T 1Z1, Canada

I. Dillmann

TRIUMF, 4004 Wesbrook Mall, Vancouver, BC V6T 2A3, Canada and
Department of Physics and Astronomy, University of Victoria, Victoria, BC V8P 5C2, Canada

R. Krücken

TRIUMF, 4004 Wesbrook Mall, Vancouver, BC V6T 2A3, Canada and
Department of Physics and Astronomy, The University of British Columbia, Vancouver, BC V6T 1Z1, Canada†

M. R. Mumpower

Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA and
Center for Theoretical Astrophysics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

R. Surman

Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA
(Dated: May 4, 2023)

The development in the description of the masses of atomic nuclei has led to various nuclear mass models that can predict the 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 process (the 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 to quantify the mass uncertainty using ensemble Bayesian model averaging (EBMA) is introduced. This Bayesian method provides a natural way to perform model averaging, selection, calibration, 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 (MCMC) method using the No-U-Turn sampler (NUTS). The average size of our best uncertainty estimates of neutron separation energies based on the AME2003 data is 0.48 [MeV] and covers 95 % of new data in AME2020. The uncertainty estimates can also be used to detect outliers with respect to the trend of experimental data and theoretical predictions.

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].

These global mass models play an important role in understanding the origin of heavy elements in the Universe via the rapid neutron capture process (r-process) [11–13]. This is because the nuclear masses determine the Q-value (energy release) of nuclear reactions and decays, which also affect their rates (e.g., β-decay rates $\propto Q^{3-7}$ [14, 15]). However, the masses of the vast majority of neutron-rich nuclei relevant to the r-process have yet to be experimentally studied. Therefore, the mass models used in nucleosynthesis studies have a significant impact on the resulting abundance patterns and kilonova lightcurves [16, 17].

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 uncertainty in microscopic theories [18, 19], the mass models that are typically used in nucleosynthesis studies, especially macroscopic-microscopic and phenomenological models, do not come with quantified prediction uncertainty. One may calculate the root mean square error of each mass model with respect to the observations, but it most likely underestimates the uncertainty where there are no data (see Figure 1). Furthermore, the possibility of quantifying the uncertainty by combining multiple mass models and observations has been largely unexplored. This poses a challenge in quantifying the

---

* yukiya@alum.ubc.ca
† Present address: Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

uncertainty in the r-process nucleosynthesis that arises from uncertain nuclear masses.

As the next-generation radioactive isotope beam facilities allow us to have access to more neutron-rich isotopes, it becomes possible to test the performances of the mass models in the extremely neutron-rich regions of the chart of nuclides. In experimental studies of nuclear masses, usually only the new experimental results are compared with the theoretical predictions. It is often done by calculating the root mean square error (RMSE) (for examples of mass measurements relevant to the r-process, see Refs. [21–24]), but this does not fully combine all the available experimental data. Therefore, a statistical method to test the predictions of various theoretical models and evaluate the impact of new measurements on the uncertainty of extrapolated masses would be an improvement to the current situation.

In this work, we will apply a method called *ensemble Bayesian model averaging* (EBMA) introduced by Ref. [25] to combine available experimental data and multiple theoretical mass models as well as 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. The EBMA combines model calibration, selection, averaging, and uncertainty quantification in a single framework. The resulting probabilistic model is highly interpretable. This Bayesian method is quite general and it can be readily applied to other nuclear physics observables.

Recently, data-driven modeling of nuclear masses using machine learning techniques has been quickly gaining popularity [26–33]. Especially, probabilistic models have achieved high accuracy while providing an estimate of uncertainty. Although there have been attempts to construct physically interpretable models [33], 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. While they may not be able to predict new or unknown physics, they can combine physics that they learn 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 quantifies the uncertainty in extrapolating the experimental data. Our approach should be considered as a method for model averaging, selection, calibration, 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 in which 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; finally, we summarize the work presented in the paper and describe possible applications in Sec. IV.

II. METHOD

A. Bayesian model averaging

We start by describing the general framework of Bayesian model averaging (BMA). 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. [25, 34], the posterior distribution of the observable of interest Δ, defined by BMA, is

$$p(\Delta \mid D) = \sum_{k=1}^{K} p(\Delta \mid M_k, D) \, p(M_k \mid D), \qquad (1)$$

where $p(\Delta \mid M_k, D)$ is the posterior PDF of the observable of interest based on a single statistical model M_k, and $p(M_k \mid D)$ 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.* [25] introduced the *ensemble Bayesian model averaging* (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

$$p(\Delta \mid m_1, \ldots, m_k) = \sum_{k=1}^{K} w_k \, g_k(\Delta \mid m_k), \qquad (2)$$

where w_k is the weight of the model m_k, whose posterior represents the probability of the model k being the best one, based on the observed data D. Since the size of the weight depends on the relative performance of the model, even if the ensemble includes a pathological model, its effect on the predictive distributions of EBMA is quite
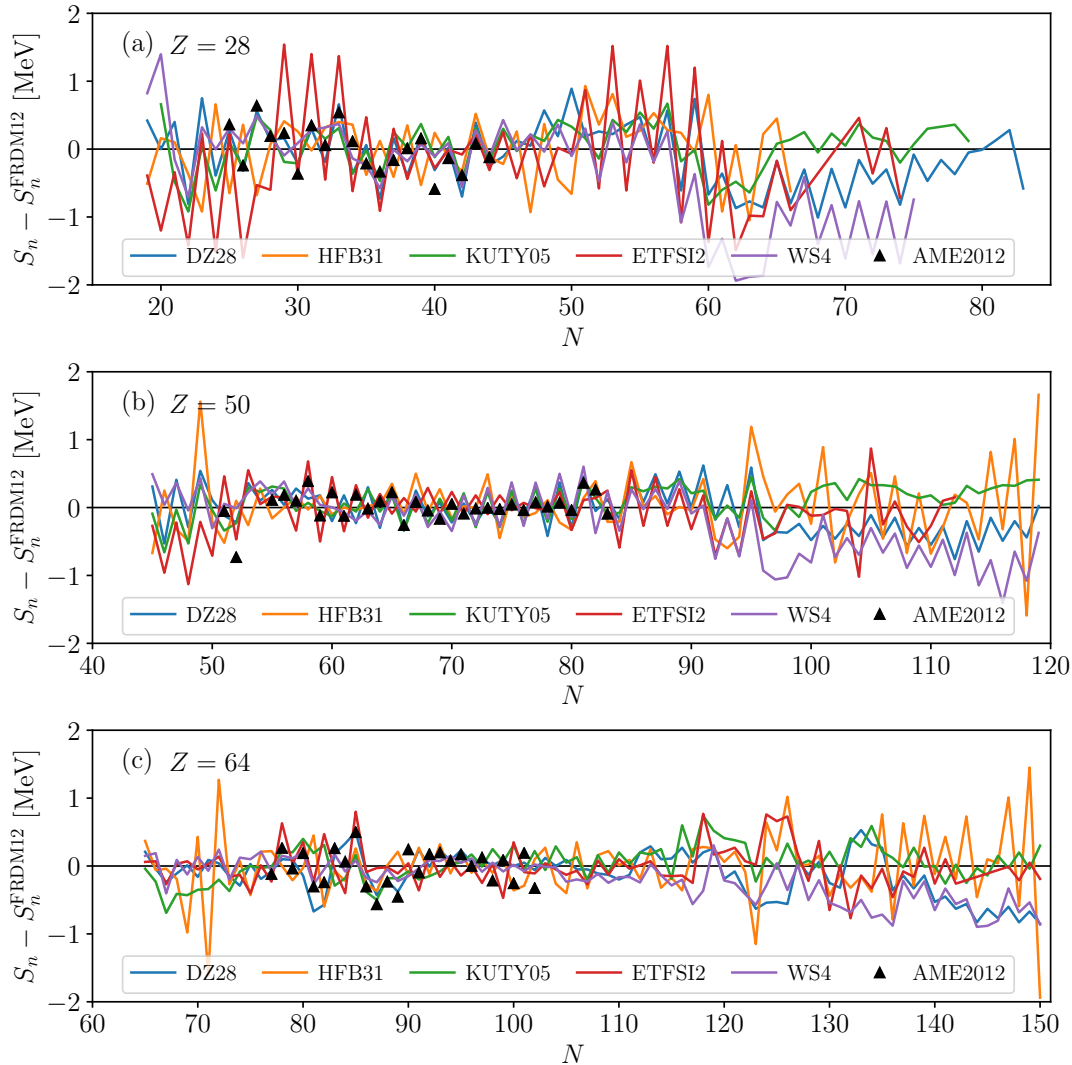
FIG. 1. Comparison of one-neutron separation energies (S_n) predicted by each mass model used in this study and the experimental masses from the AME2020 [20], relative to the predictions of the FRDM2012 for (a) $Z = 28$ (Ni), (b) $Z = 50$ (Sn), and (c) $Z = 64$ (Gd) isotopes.

small, since an extremely small weight would be assigned to the model. $g_k(\Delta \mid m_k)$ is a normal PDF with mean defined by the bias-corrected model predictions and standard deviation σ_k:

$$g_k(\Delta \mid m_k) = N(\Delta \mid a_k + b_k m_k, \sigma_k^2), \quad (3)$$

where a_k and b_k are the bias-correction coefficients, which are discussed in more detail in the following section. In the original EBMA by Raftery *et al.* [25], 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. [25] linearly corrects the bias of the prediction of each model, as shown in Equation 3. Since most mass models are already fitted to the experimental data, the values of the bias-correction coefficients are expected to be $a_k \sim 0$ and $b_k \sim 1$, where the original prediction of the model is obtained when $a_k = 0$ and $b_k = 1$.

Ref. [25] suggests that a_k and b_k for each $k = 1, \ldots, K$ are determined by linear regression. 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 is 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 AME2020 data are shown in Table I, for example.

2. Bayesian inference

The parameters of interest in our statistical inference are the weights $w_k (k = 1, \ldots, K)$ and the standard deviations of the normal distributions that correspond to each of the theoretical mass models in the ensemble. Therefore, prior distributions for the parameters must be specified. In general, we try to choose the prior distributions to be as weakly informative as possible. For the weights, since the weights 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 meets this requirement. Therefore, the prior for the weights is

$$
\begin{aligned}
& p(w_1, w_2, \ldots, w_k) \\
& = \text{Dirichlet}(w_1, w_2, \ldots, w_k \mid \alpha_1, \alpha_2, \ldots, \alpha_k) \\
& = \frac{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} w_k^{\alpha_k - 1},
\end{aligned} \tag{4}
$$

where $\alpha_1, \alpha_2, \ldots, \alpha_k$ are called the "concentration parameters", and $\Gamma(\cdot)$ is the gamma function [35]. The concentration parameters of the Dirichlet distribution are set to 1 to ensure that the prior distributions are only weakly informative. 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 [36].

The likelihood of the normal mixture model is defined as

$$
\begin{aligned}
& L(w_1, \ldots, w_K, \sigma_1^2, \ldots, \sigma_K^2) \\
& = \prod_{(n,p)} \left(\sum_{k=1}^{K} w_k g_k(\Delta_{(n,p)} \mid m_{k,(n,p)}) \right),
\end{aligned} \tag{5}
$$

TABLE I. Bias-correction coefficients determined from the MAP values of Bayesian linear regression (labeled as "BLR MAP") and usual linear regression (labeled as "LR"), respectively, using the AME2020 [20] data. a_k and b_k are the coefficients for intercept and slope, respectively, for mass model k. The values are rounded to four decimal places.

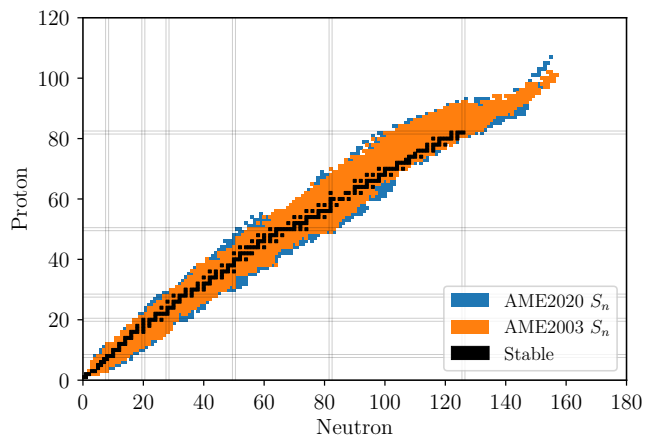| Mass model	a_k BLR MAP	a_k LR	b_k BLR MAP	b_k LR
WS4	-0.0230	-0.0230	1.0022	1.0022
DZ29	-0.1176	-0.1176	1.0141	1.0141
FRDM12	0.0528	0.0528	0.9930	0.9930
KUTY05	0.4373	0.4373	0.9441	0.9441
ETFSI2	0.7694	0.7694	0.9036	0.9036
HFB31	0.2093	0.2093	0.9749	0.9749

FIG. 2. Comparison of the AME2003 data with the latest AME2020 data for one-neutron separation energies S_n, illustrated on the chart of nuclides. The blue squares show the new S_n in the AME2020 values that did not exist in the AME2003. The S_n values listed in the AME2003 are shown in orange color.

where the subscript (n, p) represents pairs of neutron number n and proton number p of the nuclei where observations exist. In practice, the logarithm of the 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.

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

where $\boldsymbol{w} = w_1, \ldots, w_K$, $\boldsymbol{\sigma^2} = \sigma_1^2, \ldots, \sigma_K^2$, and D denotes observational 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 is provided in the form of variance of the posterior predictive distribution. Based on Ref.[25] but reflecting the fact that our σ_k depends on model k, the predictive variance can be written as

$$
\begin{aligned}
& \text{Var}(\Delta \mid m_1, m_2, \ldots, m_K) \\
& = \sum_{k=1}^{K} w_k \left((a_k + b_k m_k) - \sum_{i=1}^{K} w_i(a_i + b_i m_i) \right)^2 \\
& + \sum_{k=1}^{K} w_k \sigma_k^2,
\end{aligned} \tag{7}
$$

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.
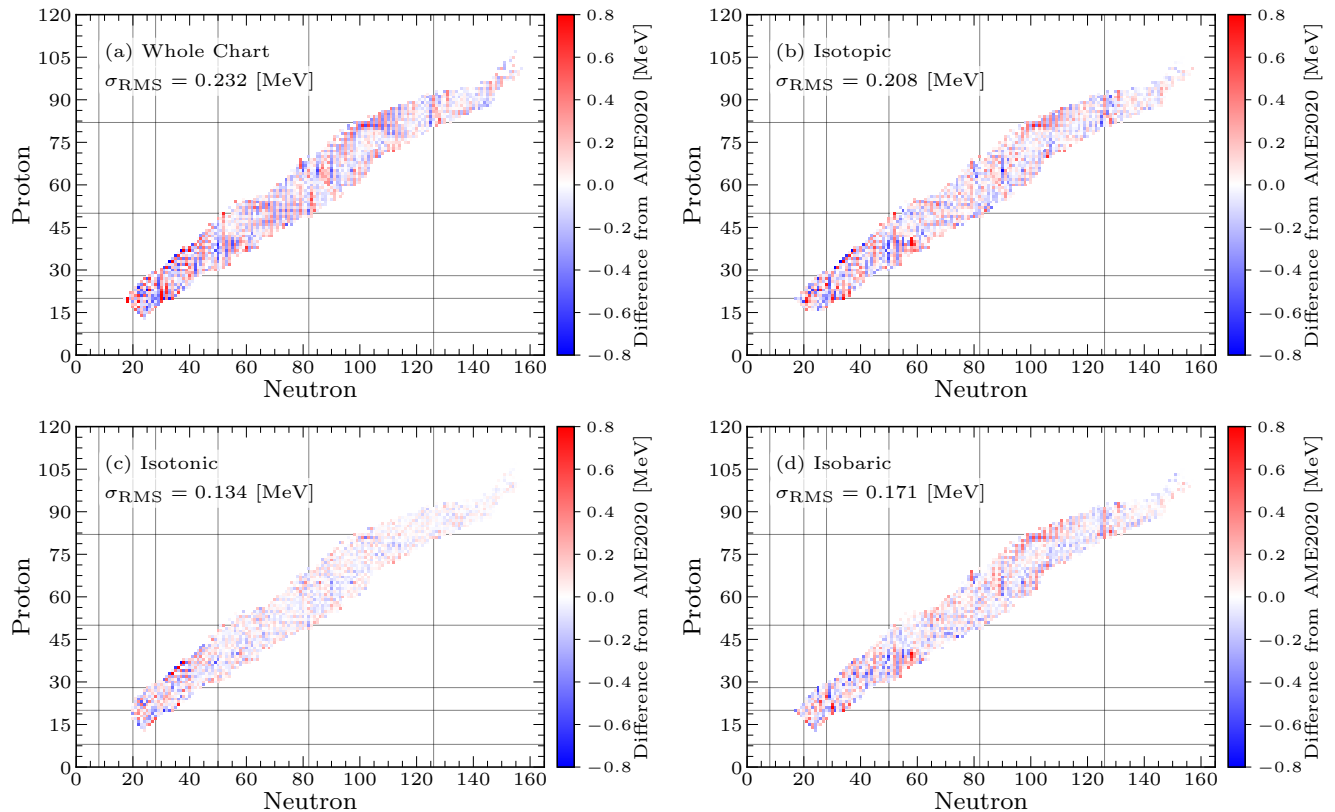
FIG. 3. Deviation and root mean square error (RMSE) σ_{RMS} [MeV] of the neutron separation energies S_n reconstructed by the EBMA models fitted with (a) the whole chart of nuclides, (b) each isotopic chain ($Z = \mathrm{const.}$), (c) each isotonic chain ($N = \mathrm{const.}$), and (d) each isobaric chain ($A = \mathrm{const.}$), compared to the AME2020 [20].

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.[28, 29, 37–39]. 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 [40, 41].

On the other hand, 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. In this framework, the predictions of each mass model that constitute the EBMA model are only linearly calibrated; therefore, the local trend of the predictions remains unchanged.

One of the shortcomings of the current method is that the 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 poses 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-dependent weights is referred to as "Bayesian Model Mixing", put forward by Refs. [42, 43]. However, 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 methodology for the averaging of general nuclear mass models, using data across the chart of nuclides as described in Sec. II C. It is

assumed that the dependence of uncertainty on location is represented by the spread of the predictions of different mass models, as shown in Fig. 1.

C. Setup of numerical experiment

In the numerical experiments discussed in the current work, all probabilistic models have been implemented using PyMC [44], 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 [45, 46]. Conventionally, parameter estimation in mixture models is performed with the Expectation Maximization (EM) algorithm to avoid the so-called "label switching problem" [47, 48]. 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 $g_k(\Delta \mid m_k)$. This makes the analysis of the posterior distributions challenging. However, 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_n), 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) are often calculated from the neutron capture rate via detailed balance:

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

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), m_u is the mass of a nucleon, and T is the temperature of the environment. Note that the reverse rate is exponentially dependent on S_n.

The mass models included in our ensembles are the Duflo-Zucker mass model with 29 parameters (DZ29) [7], FRDM2012 [2], HFB31 [10], KUTY05 [49], ETFSI2 [50], and WS4 [6]. In most of our numerical experiments, we take the S_n values from the AME2020 [20] as observations. In evaluating the quality of uncertainty estimates for unseen data, we use the S_n values from AME2003 [51] for constructing our models and then test them with the new data points in AME2020. In the AME2020, 319 new S_n data points with proton number $Z = 16\text{-}105$ are available compared to the AME2003. The new data points in the AME2020 compared to the AME2003 are shown in Figure 2.

We consider four different ways to categorize the S_n data. The first category is the data for the whole chart of nuclides, which employs all the available experimental data at once. The second and third are data for each isotopic and isotonic chain, respectively. This focuses on the evolution of the S_n values as a function of proton and neutron number (isotopic and isotonic, respectively). The last is isobaric (equal mass number A), to demonstrate that it is possible to create an EBMA model for each isobaric chain, which is relevant to the trend of β^--decay Q-values.

III. RESULTS AND DISCUSSION

A. Comparison with observations

To compare the predictions of the EBMA model with the observations (data from AME2020), the nominal predictions of EBMA are taken as the MAP values of the predictive distributions of S_n, with the bias-correction parameters, weights, and standard deviations σ_k (Equation 3) also determined from the AME2020 values. 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 EBMA predictions of the one-neutron separation energies (S_n) from the AME2020 values. The root mean square error (σ_{RMS}) shown in the figure is defined as

$$\sigma_{\text{RMS}} = \sqrt{\frac{\sum_{(n,p)} \left(S_{n,(n,p)}^{\text{AME}} - S_{n,(n,p)}^{\text{EBMA}} \right)^2}{N^{\text{AME}}}}, \quad (9)$$

where (n,p) represents pairs of neutron number n and proton number p of nuclei in the AME2020 whose S_n values are used for the fit. N^{AME} is the total number of such nuclei ($N^{\text{AME}} \equiv \sum_{(n,p)}$). For the fit of the entire chart of nuclides, $N^{\text{AME}} = 2030$. This number is slightly different in different fits, since some of the fits do not converge at the edge of the chart. $S_{n,(n,p)}^{\text{AME}}$ and $S_{n,(n,p)}^{\text{EBMA}}$ are the S_n values for a nucleus (n,p) from the AME2020 and EBMA models, respectively.

The value of σ_{RMS} of the fit of the whole chart of nuclides (panel (a)) shows that the averaged mass model created by the EBMA method can reproduce the experimental values slightly better ($\sigma_{\text{RMSE}}^{(a)} = 0.232$ MeV) than
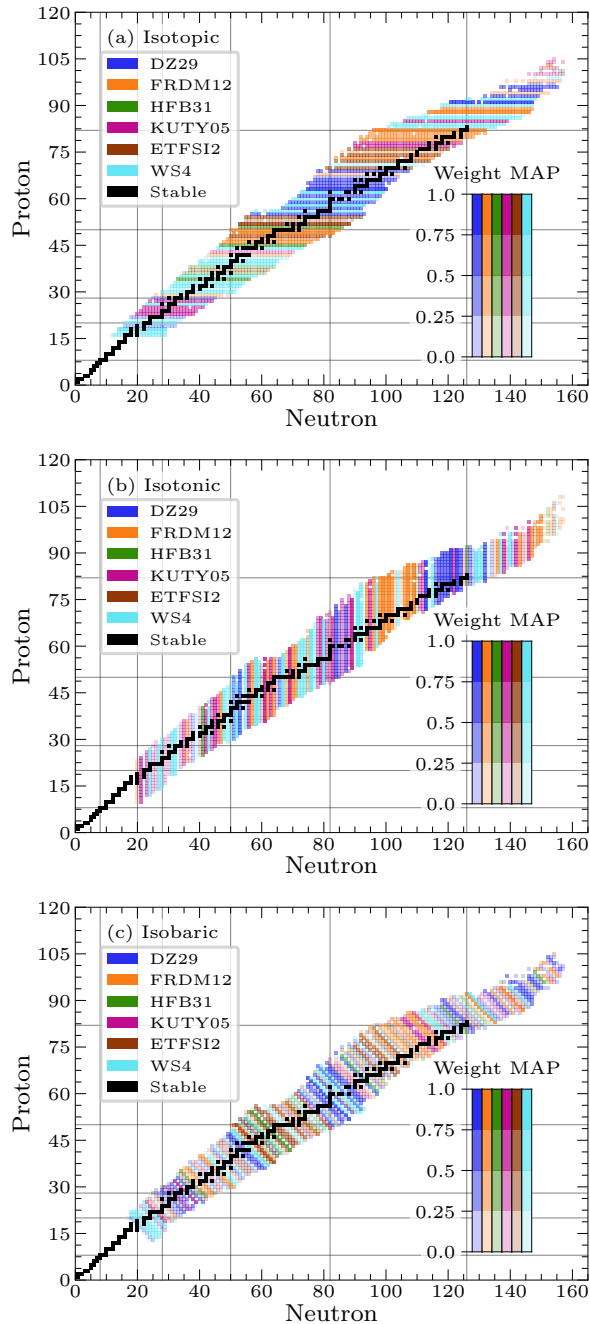
FIG. 4. Maximum a posteriori (MAP) values of the largest weight in the EBMA ensemble determined for each (a) isotopic chain, (b) isotonic chain, and (c) isobaric chain.

the best performing model, which is the WS4 model [6] with $\sigma_{\mathrm{RMSE}}^{\mathrm{WS4}} = 0.257$ MeV (see Table II). Further reduction in σ_{RMSE} is achieved when S_n of each isotopic chain is fit separately (panel (b) of Figure 3). This suggests that some models in the ensemble perform better than the overall best-performing mass model (WS4) for some isotopic chains. This can be verified by inspecting the weights of the EBMA model and will be discussed in

more detail in Section III B. The best performance in reproducing the experimental S_n is obtained when separate EBMA models are optimized for each isotonic chain (panel (c) of Figure 3). This means that the mass models can capture the isotonic trends of S_n (S_n as functions of neutron number N) much better than the isotopic trends (S_n as functions of proton number Z), at least for the available experimental data. This is most likely because the values of S_n evolve rather smoothly as a function of the proton number, compared to as a function of the neutron number where odd-even staggering is present. The fit of each isobaric chain resulted in the RMSE value of 0.171 MeV, which sits between the isotopic and isotonic models (panel (d) of Figure 3).

B. Weights in the EBMA models

EBMA models are constructed using the S_n data from AME2020 in three different ways: optimizing the EBMA models with 1) the data for the whole chart of nuclides, 2) each isotopic chain, and 3) each isotonic chain, respectively. Table II lists the 95 % posterior highest density intervals (HDIs) of the EBMA weights and variances, which are the narrowest intervals that include 95 % of the posterior distributions, when the EBMA model is optimized with the observed S_n data for the whole chart of nuclides. 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 DZ29 and FRDM12. The order of the top three mass models is the same as the order of the smallest root mean square errors (RMSEs) with respect to the AME2020 [20]. On the other hand, the standard deviations or variances of the normal distributions in the mixture model do not agree with the RMSE values. This is because the mixture model is a weighted sum of the normal distributions and is fitted to the data at once, not individually. It is not apparent why the posterior weight of the HFB-31 model is much smaller than the others, although the RMSE value (0.472 MeV) is smaller than the KUTY05 (0.746 MeV) and ETFSI2 (0.828 MeV) models. The HFB-31 mass model performs relatively well on average in reproducing the observed data, but it is possible that the local trend of the S_n values does not agree with the observation, therefore resulting in a small weight. Such local disagreements could also explain the large interval of the posterior standard deviation of the model (Table II).

The colors in Figure 4 show the mass model with the largest weight within the ensemble for each isotopic (panel (a)), isotonic (panel (b)), and isobaric chain (panel (c)). The color scale represents the value of the weight. The weight is the posterior probability of the mass model being the best one, based on the training using the AME2020 data. Trends of the best models in different regions are visible, especially when the ensembles are created for each isotopic chain. Note that the current method cannot provide explanations for why
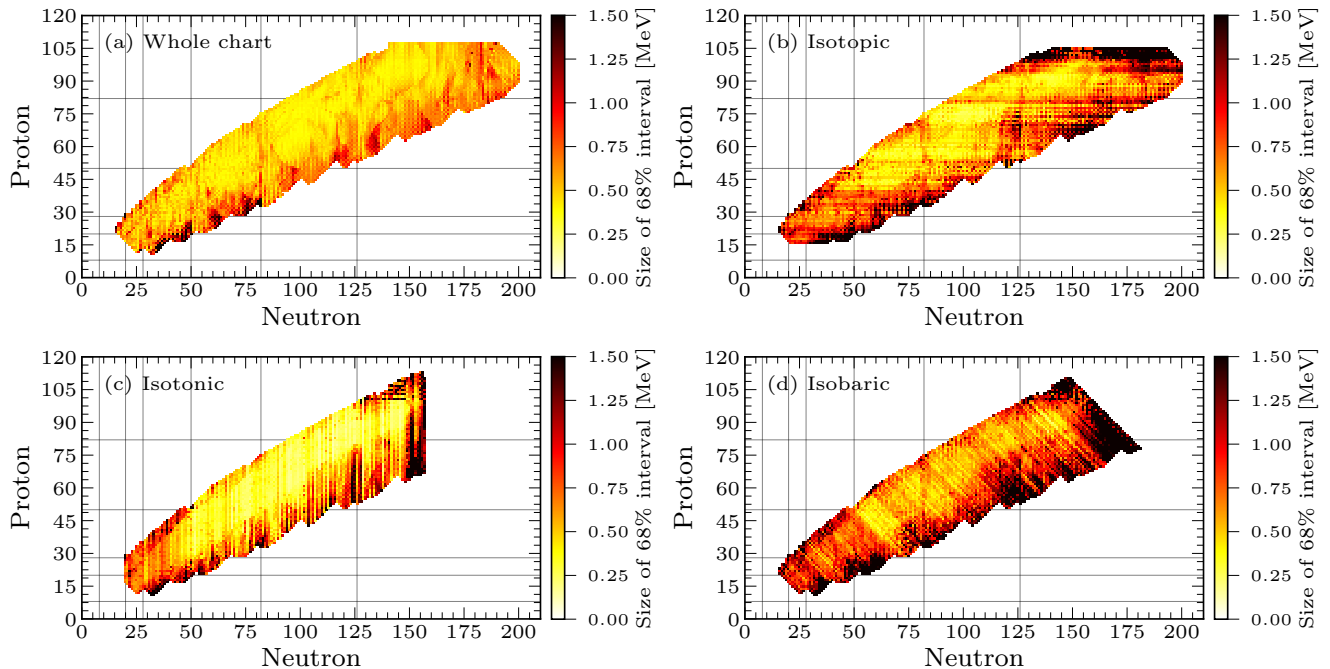
FIG. 5. Trends of the sizes of 68% highest density intervals of the EBMA models, fitted with the AME2020 S_n data for (a) the whole chart of nuclides, (b) each isotopic chain, (c) each isotonic chain, and (d) each isobaric chain. The charts with isotonic and isobaric fits are truncated at a large neutron number and mass number, respectively, because there are not enough data points within the chains to determine the EBMA parameters.

certain weights are large or small. Understanding the weights would require a careful analysis of the relative performance of each model.

C. Uncertainty quantification with EBMA

One of the main goals of this study is to quantify the uncertainty of the theoretical values S_n 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

TABLE II. 95 % posterior highest density intervals (HDI) of the EBMA weights and standard deviations (variances), fitted with the AME2020 S_n values. The notation (a, b) denotes an interval with a being the lower bound and b being the upper bound, respectively. RMSE shows the root mean square error of each mass model with respect to the AME2020 values.

Mass model	Weight	Standard deviation	RMSE [MeV]
WS4	(0.392, 0.539)	(0.186, 0.221)	0.257
DZ29	(0.154, 0.277)	(0.134, 0.196)	0.292
FRDM12	(0.145, 0.264)	(0.143, 0.196)	0.350
KUTY05	(0.021, 0.127)	(0.134, 0.292)	0.746
ETFSI2	(0.001, 0.048)	(0.063, 0.450)	0.828
HFB31	(0.000, 0.029)	(0.138, 1.073)	0.472

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 [25]. The interpretability of the uncertainty is another advantage of the EBMA method.

1. Size of uncertainty estimates

Figure 5 shows the size of the 68% highest density interval (HDI), which is roughly comparable to the $\pm 1\sigma$ interval of the normal distribution, with EBMA models fitted for the whole chart of nuclides (panel (a)), each isotopic chain (panel (b)), each isotonic chain (panel (c)), and each isobaric chain (panel (d)), respectively. The fit is performed using the AME2020 data, and the predictions are made for all the nuclei available in all the member mass models within the ensemble. In all the cases, it can be seen that the size of the uncertainty is constrained where the data exist, but increases towards the edge of the chart of nuclides, especially in the neutron-rich direction. This reflects the fact that the predictions of the mass models constituting the ensemble start to diverge as we move further from the last data point, as shown in Figure 1.
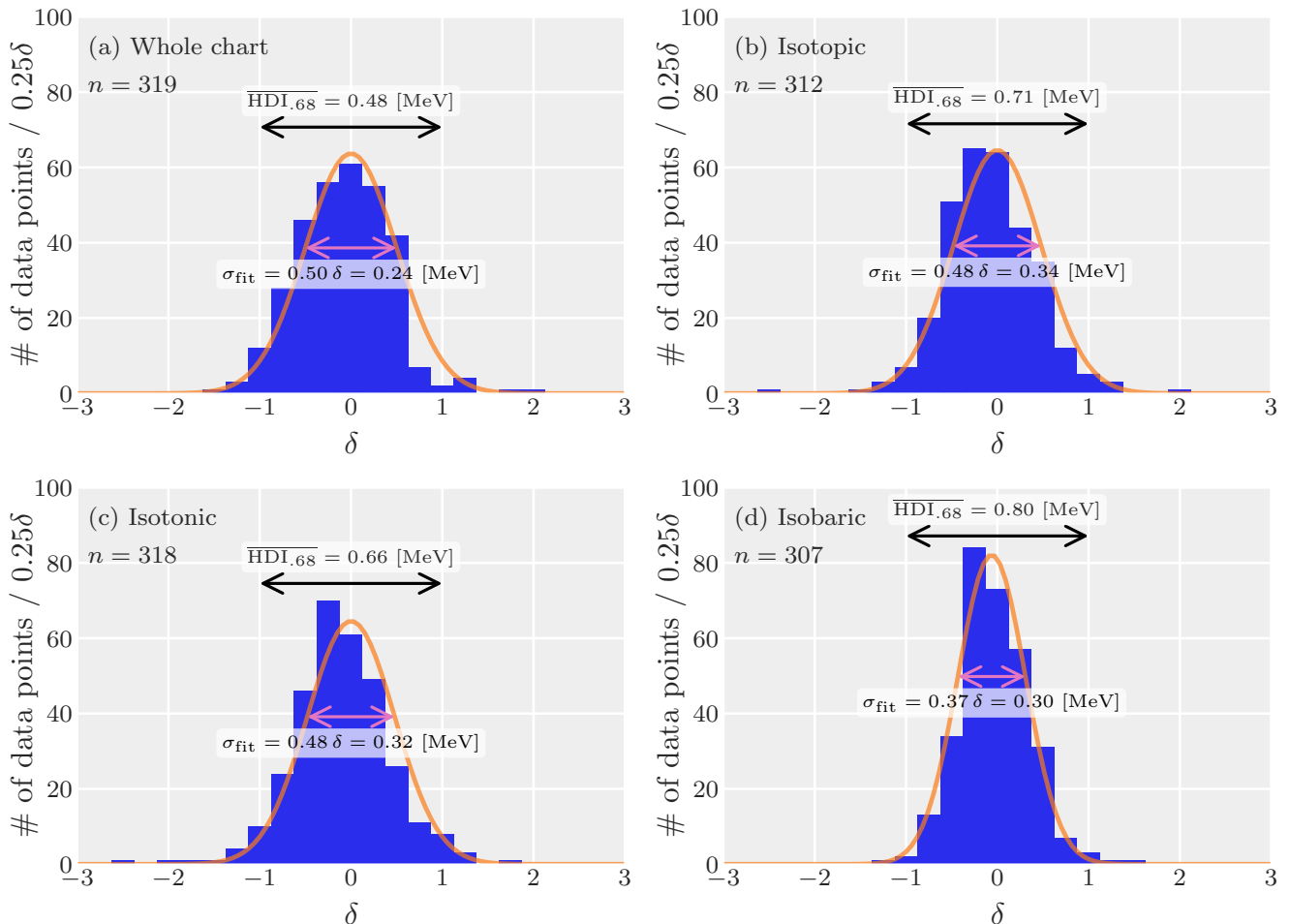
Comparing the four plots of Figure 5, the increase

FIG. 6. Distributions of the new data points in AME2020 compared to 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.

in the size of uncertainty in the neutron-rich region is the smallest for the fit using the whole chart of nuclides (panel (a)). This is likely because the weights for the whole chart of nuclides can be determined using all available data, whereas for each isotopic, isotonic, and isobaric chain, the weights are determined only from the data in each chain. Although the isotonic fit was best performing in terms of reproducing the experimental data (Figure 3), it can be seen from the panel (c) of Figure 5, the size of uncertainty rapidly grows where experimental data do not exist.

These results show that, in general, the current theoretical understanding and the accuracy of modeling of nuclear masses become increasingly uncertain towards the neutron-rich region. Therefore, in investigating the impact of uncertain masses on the prediction of the r-process observables, such an increase in the size of uncertainty should be taken into account. Using the EBMA uncertainty would allow for more realistic uncertainty quantification and sensitivity studies of the r-process ob-

servables.

2. Quality of uncertainty estimates

As discussed in Section II B, the predictive uncertainty given by the EBMA model has a straightforward interpretation. Now, we investigate the quality of the estimation of the size of uncertainty. We first construct EBMA models and quantify the prediction uncertainties using the data from the AME2003 [51], then evaluate the quality of the uncertainties based on the new data from the AME2020 [20].

Figure 6 shows the distribution of the new S_n data relative to the sizes of uncertainties given by the EBMA models fitted with the data for the whole chart of nuclides (panel (a)), fitted for each isotopic chain (panel (b)), isotonic chain (panel (c)), and isobaric chain (panel (d)). The number of new data points included in the fit (n in Figure 6) is not necessarily 319 for the iso-

topic/isotonic/isobaric fits because at the edge of the chart of nuclides, often there are not enough data points in the isotopic/isotonic/isobaric chains and the posterior weights do not converge. Such data points are excluded from the fits. The size of the EBMA uncertainty is taken as the 68% highest (posterior) density interval (HDI$_{.68}$), which is the narrowest 68% credible interval on the posterior distribution. If the distribution is a perfect normal distribution, 68% corresponds to the $\pm 1\sigma$ interval symmetric about the mean. To study how the observation distributes relatively to the HDI$_{.68}$, we define δ, which represents an observed S_n value normalized by the size of HDI$_{.68}$. Let h^{low} and h^{up} represent the lower and upper boundaries of the HDI$_{.68}$, respectively,

$$\delta = \frac{S_n - h^{\text{low}}}{h^{\text{up}} - h^{\text{low}}} - 0.5, \tag{10}$$

where 0.5 is subtracted to symmetrize the distribution around 0.

Comparing the average sizes of the 68% intervals ($\overline{\text{HDI}_{.68}}$), on average, the model fitted with the whole chart of nuclides (panel (a) of Figure 6) provides the most constrained size of the uncertainty of 0.48 MeV. On the other hand, the models fitted for each isotopic (panel (b)), isotonic (panel (c)), and isobaric (panel (d)) chain have larger credible intervals. This is most likely due to the fewer observation data points in each isotopic, isotonic, or isobaric chain compared to the data for the whole chart of nuclides. For the isotopic and isotonic fits (panels (c) and (d)), some data points around $\delta \sim 2$ can be seen, suggesting that the models optimized for isotopic/isotonic chains may be more sensitive to observations that do not follow the trend within the isotopic/isotonic chain. This means that the isotopic/isotonic models may be used to detect anomalous masses or separation energies with respect to the trend within the isotopic/isotonic chain. Since most of the new data points fall within the 68 % intervals, while the sizes of the uncertainties are somewhat overestimated, it can be concluded that the EBMA method is capable of extrapolating the S_n values with conservative uncertainty estimates, when the new data points are adjacent to the existing data. The quantified uncertainties may also be used to estimate the probabilities of certain nuclei to be bound as in Refs. [29, 38, 52], which will be addressed in future work.

IV. CONCLUSIONS

We have explored the possibility of quantifying the uncertainty of theoretical one-neutron separation energies (S_n) when a variety of mass models are available, using the Ensemble Baysian Model Averaging (EBMA) method. The EBMA method models an ensemble of bias-corrected mass models as a mixture of normal distributions, whose parameters are estimated by MCMC using the No-U-Turn-Sampler (NUTS). EBMA models have been constructed in four different ways of fitting, namely, the whole chart of nuclides, each isotopic chain, each isotonic chain, and each isobaric chain.

In reproducing the observed one-neutron separation energies (S_n), in all cases, the maximum a posteriori (MAP) estimates of the EBMA models result in smaller root mean square deviations from the AME2020 data than the best model in the ensemble, namely the WS4 model. While the EBMA model fitted for the whole chart of nuclides results in a larger σ_{RMS} value than the isotopic/isotonic/isobaric fit, it provides a more constrained and accurate uncertainty.

For all cases, the 68% HDIs estimated from AME2003 data contain roughly 95% of the new observations in AME2020. This suggests that the extrapolations of the S_n values provided by the current EBMA models based on the AME2003 data and the six theoretical mass models work well for the new data in AME2020. Furthermore, based on the distributions of the new AME2020 data with respect to the size of the 68% HDIs, we conclude that the EBMA method provides meaningful but conservative uncertainty estimates.

The advantage of the EBMA method is its simplicity in obtaining the uncertainty estimates of an ensemble of theoretical models. This would make it possible to apply the current method to various nuclear physics observables. Using an ensemble of theoretical models can also naturally model the fact that the predictions of nuclear physics models are not well constrained far from stability, especially in the neutron-rich region. Such realistic uncertainty estimates are necessary to accurately assess the impact of uncertain nuclear physics inputs in the studies of heavy element nucleosynthesis, especially the r-process.

[1] P. Moller, J. Nix, W. Myers, and W. Swiatecki, Atomic Data and Nuclear Data Tables **59**, 185 (1995).

[2] P. Möller, A. Sierk, T. Ichikawa, and H. Sagawa, Atomic

Data and Nuclear 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, Physics Letters B **734**, 215 (2014).

[7] J. Duflo and A. 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 **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, Progress in Particle and Nuclear Physics **86**, 86 (2016).

[14] B. W. Sargent, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **139**, 659 (1933).

[15] Y. Zhou, Z.-H. Li, Y.-B. Wang, Y.-S. Chen, B. Guo, J. Su, Y.-J. Li, S.-Q. Yan, X.-Y. Li, Z.-Y. Han, Y.-P. Shen, L. Gan, S. Zeng, G. Lian, and W.-P. Liu, Science China Physics, Mechanics & Astronomy **60**, 082012 (2017).

[16] Y. L. Zhu, K. A. Lund, J. Barnes, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower, and R. Surman, The Astrophysical Journal **906**, 94 (2021).

[17] J. Barnes, Y. L. Zhu, K. A. Lund, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower, and R. Surman, The Astrophysical Journal **918**, 44 (2021).

[18] J. Dobaczewski, W. Nazarewicz, and P.-G. Reinhard, Journal of Physics G: Nuclear and Particle Physics **41**, 074001 (2014).

[19] J. D. McDonnell, N. Schunck, D. Higdon, J. Sarich, S. M. Wild, and W. Nazarewicz, Phys. Rev. Lett. **114**, 122501 (2015).

[20] M. Wang, W. Huang, F. Kondev, G. Audi, and S. Naimi, Chinese Physics C **45**, 030003 (2021).

[21] J. Van Schelt, D. Lascar, G. Savard, J. A. Clark, S. Caldwell, A. Chaudhuri, J. Fallis, J. P. Greene, A. F. Levand, G. Li, K. S. Sharma, M. G. Sternberg, T. Sun, and B. J. Zabransky, Phys. Rev. C **85**, 045805 (2012).

[22] M. Vilen, J. M. Kelly, A. Kankainen, M. Brodeur, A. Aprahamian, L. Canete, R. P. de Groote, A. de Roubin, T. Eronen, A. Jokinen, I. D. Moore, M. R. Mumpower, D. A. Nesterenko, J. O'Brien, A. P. Perdomo, H. Penttilä, M. Reponen, S. Rinta-Antila, and R. Surman, Phys. Rev. C **101**, 034312 (2020).

[23] C. Izzo, J. Bergmann, K. A. Dietrich, E. Dunling, D. Fusco, A. Jacobs, B. Kootte, G. Kripkó-Koncz, Y. Lan, E. Leistenschneider, E. M. Lykiardopoulou, I. Mukul, S. F. Paul, M. P. Reiter, J. L. Tracy, C. Andreoiu, T. Brunner, T. Dickel, J. Dilling, I. Dillmann, G. Gwinner, D. Lascar, K. G. Leach, W. R. Plaß, C. Scheidenberger, M. E. Wieser, and A. A. Kwiatkowski,

[24] H. F. Li, S. Naimi, T. M. Sprouse, M. R. Mumpower, Y. Abe, Y. Yamaguchi, D. Nagae, F. Suzaki, M. Wakasugi, H. Arakawa, W. B. Dou, D. Hamakawa, S. Hosoi, Y. Inada, D. Kajiki, T. Kobayashi, M. Sakaue, Y. Yokoda, T. Yamaguchi, R. Kagesawa, D. Kamioka, T. Moriguchi, M. Mukai, A. Ozawa, S. Ota, N. Kitamura, S. Masuoka, S. Michimasa, H. Baba, N. Fukuda, Y. Shimizu, H. Suzuki, H. Takeda, D. S. Ahn, M. Wang, C. Y. Fu, Q. Wang, S. Suzuki, Z. Ge, Y. A. Litvinov, G. Lorusso, P. M. Walker, Z. Podolyak, and T. Uesaka, Phys. Rev. Lett. **128**, 152701 (2022).

[25] A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski, Monthly Weather Review **133**, 1155 (2005).

[26] Z. M. Niu and H. Z. Liang, Phys. Rev. C **106**, L021303 (2022).

[27] Z. Niu and H. Liang, Physics Letters B **778**, 48 (2018).

[28] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Phys. Rev. C **98**, 034318 (2018).

[29] L. Neufcourt, Y. Cao, W. Nazarewicz, E. Olsen, and F. Viens, Phys. Rev. Lett. **122**, 062502 (2019).

[30] R.-D. Lasseri, D. Regnier, J.-P. Ebran, and A. Penon, Phys. Rev. Lett. **124**, 162502 (2020).

[31] R. Utama, J. Piekarewicz, and H. B. Prosper, Phys. Rev. C **93**, 014311 (2016).

[32] A. E. Lovell, A. T. Mohan, T. M. Sprouse, and M. R. Mumpower, Phys. Rev. C **106**, 014305 (2022).

[33] M. R. Mumpower, T. M. Sprouse, A. E. Lovell, and A. T. Mohan, Phys. Rev. C **106**, L021301 (2022).

[34] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, Statistical Science **14**, 382 (1999).

[35] $\Gamma(z) \equiv \int_0^\infty t^{z-1} e^{-t} \mathrm{d}t$.

[36] A. Gelman, D. Simpson, and M. Betancourt, Entropy **19** (2017).

[37] 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, Nature Physics **17**, 1408 (2021).

[38] L. Neufcourt, Y. Cao, S. A. Giuliani, W. Nazarewicz, E. Olsen, and O. B. Tarasov, Phys. Rev. C **101**, 044307 (2020).

[39] L. Neufcourt, Y. Cao, S. Giuliani, W. Nazarewicz, E. Olsen, and O. B. Tarasov, Phys. Rev. C **101**, 014319 (2020).

[40] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, 2005).

[41] S. Yoshida, Phys. Rev. C **102**, 024305 (2020).

[42] 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, Journal of Physics G: Nuclear and Particle Physics **48**, 072001 (2021).

[43] A. C. Semposki, R. J. Furnstahl, and D. R. Phillips, Phys. Rev. C **106**, 044002 (2022).

[44] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, PeerJ Computer Science **2**, e55 (2016).

[45] R. M. Neal, *Bayesian learning for neural networks*, Vol. 118 (Springer Science & Business Media, 2012).

[46] R. M. Neal et al., Handbook of markov chain monte carlo **2**, 2 (2011).

[47] M. Stephens, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **62**, 795 (2000), https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9868.00265.

[48] A. Jasra, C. C. Holmes, and D. A. Stephens, Statistical Science **20**, 50 (2005).

[49] H. Koura, T. Tachibana, M. Uno, and M. Yamada, Progress of Theoretical Physics **113**, 305 (2005), https://academic.oup.com/ptp/article-pdf/113/2/305/5192381/113-2-305.pdf.

[50] Y. Aboussir, J. Pearson, A. Dutta, and F. Tondeur, Atomic Data and Nuclear Data Tables **61**, 127 (1995).

[51] G. Audi, A. Wapstra, and C. Thibault, Nuclear Physics A **729**, 337 (2003), the 2003 NUBASE and Atomic Mass Evaluations.

[52] S. R. Stroberg, J. D. Holt, A. Schwenk, and J. Simonis, Phys. Rev. Lett. **126**, 022501 (2021).