Beyond constant error: Heteroscedastic Bayesian model combination for modeling unmeasured nuclei

B. Knight,[1,2,*] S. Lalit,[1,†] P. Giuliani,[1,‡] K. Godbey,[1,§] W. Nazarewicz,[1,2,‖] A. Ravlić,[1,3,¶] and P.-G. Reinhard[4,**]

[1]*Facility for Rare Isotope Beams, Michigan State University, East Lansing, Michigan 48824, USA*

[2]*Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*

[3]*Department of Physics, Faculty of Science, University of Zagreb, Bijenička c. 32, 10000 Zagreb, Croatia*

[4]*Institut für Theoretische Physik, Universität Erlangen, D-91054 Erlangen, Germany*

(Received 15 July 2026; accepted 19 October 2026; published 4 November 2026)

Experimentally inaccessible regions of the nuclear chart remain a challenge for global models of atomic nuclei to predict. This includes exotic nuclei near particle drip lines, superheavy elements at the extremes of mass and charge, and the neutron-rich pathways of astrophysical processes in explosive stellar environments where heavy elements are created. Given that individual nuclear models are imperfect, deep extrapolations are best approached using model ensembles, which allow for the systematic combination of diverse theoretical predictions. In this study, we employ the recently introduced Bayesian model combination (BMC) method, based on statistical machine learning, that provides robust uncertainty quantification for forecasts using model ensembles. To account for the inherent degradation of predictive power as models extrapolate into the yet-unexplored domain, we introduce a heteroscedastic BMC framework in which the combined theoretical uncertainty is treated as a dynamic quantity. We apply this methodology to an ensemble of realistic energy density functionals with a specific focus on the $Z = 46$–52 isotopic chains. We rigorously validate the approach using both experimental data and synthetic data designed to assess performance in the deep extrapolation regime. Our results demonstrate that the proposed heteroscedastic approach yields superior calibration metrics and provides statistically principled assessments of the particle drip lines.

DOI: 10.1103/z7g9-7d74

I. INTRODUCTION

The limits of the nuclear landscape have not been firmly established [1–3]. The neutron-rich frontier, in particular, is challenging for nuclear theory because of the limited experimental data and the massive extrapolations involved. This part of the nuclear chart is of great astrophysical importance, as it hosts the rapid neutron-capture process (r-process) that is in part responsible for creating heavy nuclei [4]. For modeling these nuclei, nuclear density functional theory (DFT) [5,6] has been a recognized global microscopic framework. However, many energy density functionals (EDFs) exist, and their predictions for unmeasured nuclei often disagree.

Along with the inherent systematic uncertainties associated with extrapolating far beyond the domain of data used to calibrate the functionals, discrepancies between models introduce significant, often unquantified uncertainties into predictions of nuclear properties [7]. This deficiency underscores the critical need for robust uncertainty quantification methodologies to determine the predictive capabilities of nuclear DFT. In recent years, numerous machine learning techniques have been applied to nuclear mass predictions, often with the primary goal of minimizing the root mean square error (RMSE) of point predictions (see, e.g., Refs. [8,9]). While valuable, this focus can leave the reliability of the associated uncertainties underexplored.

In response to the challenge of reconciling different EDFs, recent efforts have increasingly turned to the robust framework of Bayesian statistics [10,11]. A landmark application of these techniques in nuclear DFT was presented in Ref. [12] using Bayesian model averaging (BMA), a multimodel approach. The BMA systematically combines predictions from a suite of different EDFs, weighting each model based on its ability to reproduce known experimental data by itself. A known drawback of BMA is its assumption that one of the models is the "true model" that perfectly describes reality, a deficiency that has motivated the exploration of other model-mixing approaches [11]. In particular, Ref. [13] introduced the Bayesian model combination (BMC) framework that contains model preselection and model orthogonalization. This approach has recently been used to predict the proton drip lines for the tin isotopes [14].

While powerful, these multimodel approaches rely on a simplifying assumption: The underlying intrinsic theoretical error can be modeled with constant (homoscedastic) variance.

*Contact author: knight@frib.msu.edu
†Contact author: lalit@frib.msu.edu
‡Contact author: giulianp@frib.msu.edu
§Contact author: me@kyle.ee
‖Contact author: witek@frib.msu.edu
¶Contact author: ravlic@frib.msu.edu
**Contact author: paul-gerhard.reinhard@fau.de

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

This assumption is not motivated by the properties of the models. The natural expectation is that a model's predictive power will degrade and its associated uncertainties will grow as it extrapolates further from the region of known data. In this work, we remove this restriction by introducing a more flexible and physically realistic statistical model. We propose that the theoretical uncertainty is not constant but dynamic (heteroscedastic), allowing it to vary across the nuclear chart. We demonstrate that such a framework provides a robust description of theoretical error and yields demonstrably superior uncertainty quantification for extrapolations of observables associated with nuclear binding energies.

The outline of this paper is as follows. In Sec. II, we describe the nuclear DFT models used in this work. The BMC methodology is discussed in Sec. III. Section IV describes the validation of BMC on the synthetic data. Predictions trained on experimental data are presented in Sec. V. Finally, in Sec. VI we present our conclusions and outlook.

II. NUCLEAR MODELS AND BINDING ENERGY INDICATORS

All the theoretical models we use in this work are based on nuclear DFT, with varying EDFs and calibration schemes. In particular, we use three covariant EDFs: DD-PC1 [15], DD-PCX [16] and DD-ME2 [17]; two Skyrme EDFs: UNEDF1 [18] and UNEDF2 [19]; and one Fayans EDF: Fy(IVP) [20]. The covariant EDFs employ the separable pairing functional as described in Ref. [21] and the Skyrme-based EDFs have mixed contact delta pairing [18], while the Fayans EDF has more sophisticated pairing terms including the gradient terms [22]. The underlying Hartree-Fock-Bogoliubov (or relativistic Hartree-Bogoliubov in case of covariant EDFs) equations are solved either by expanding the quasiparticle states in the basis of an axially deformed harmonic oscillator [23,24] or by discretizing them in the coordinate space [25]. All calculations assume axially deformed and reflection-invariant nuclear shapes. Calculations of odd-A and odd-odd nuclei are performed by self-consistent quasiparticle blocking assuming the equal-filling approximation [26].

To establish a rigorous framework for our evaluation, we partition the statistical domain of our models into distinct datasets for calibration and testing. Specifically, the dataset is divided into three distinct subsets:

(1) *Training set*. This subset comprises the core of the experimentally known domain from the atomic mass evaluation (AME2020), localized near the valley of stability. It serves as the calibration dataset to determine both the BMC model weights and the coefficients of the error model.

(2) *Validation set*. To evaluate performance under extrapolation, we define a validation set isolated at the outermost boundaries of the experimentally known domain. This boundary region consists of the lightest and heaviest n measured nuclei along each isotopic chain, where n is dynamically chosen to guarantee a 30/70 validation-to-training data split within the known nuclear chart. The explicit role of this validation data is to evaluate the performance of each error model against each other.

(3) *Prediction (extrapolation) set*. This subset is made up of every nucleus where theoretical model predictions exist but experimental data do not, serving as our deep extrapolation domain.

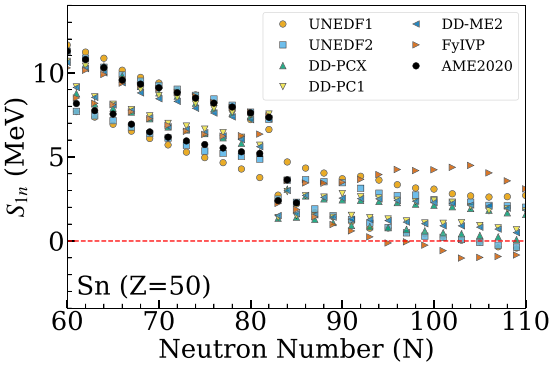


FIG. 1. Predictions of S_{1n} for the Sn ($Z = 50$) isotopic chain from the six DFT models, compared to experimental AME2020 data [27]. In the experimentally known region ($N \lesssim 82$), the models show reasonable agreement with data, but their spread grows substantially as they extrapolate toward the neutron drip line. The drip line threshold ($S_{1n} = 0$) is marked in red.

The observables we focus on in this work are a set of binding energy indicators, derived from the total binding energy BE(Z, N) of a nucleus with Z protons and N neutrons. In particular, these are the one- and two-neutron, and one- and two-proton separation energies, defined as the finite differences in binding energy between neighboring isotopes and isotones, respectively:

$$S_{1n}(Z,N) = \mathrm{BE}(Z,N) - \mathrm{BE}(Z,N-1), \tag{1a}$$

$$S_{2n}(Z,N) = \mathrm{BE}(Z,N) - \mathrm{BE}(Z,N-2), \tag{1b}$$

$$S_{1p}(Z,N) = \mathrm{BE}(Z,N) - \mathrm{BE}(Z-1,N), \tag{1c}$$

$$S_{2p}(Z,N) = \mathrm{BE}(Z,N) - \mathrm{BE}(Z-2,N). \tag{1d}$$

We also calculate the alpha-decay Q values:

$$Q_\alpha(Z,N) = \mathrm{BE}(Z-2,N-2) + \mathrm{BE}(2,2) - \mathrm{BE}(Z,N). \tag{2}$$

While it is possible to train the BMC model entirely on BE(Z, N) and derive the separation energies from the predicted total binding energies, this method fails to account for the local correlations between adjacent nuclei in the isotopic chain. To ensure the most accurate predictions possible, we perform calibrations on each separation energy indicator separately. This yields significantly tighter uncertainty bounds across the training and extrapolation datasets. The average predicted uncertainty for S_{1n} is 0.35 MeV, compared to 1.12 MeV when propagating errors from a BE(Z, N) calibration using the same error model.

Figure 1 displays the S_{1n} predictions of the six models for the tin chain. The models track the experimental separation energies well in the region calibrated by AME2020 data [27], but diverge beyond the last measured isotope ^{135}Sn. This growing intermodel spread signals increasing theoretical uncertainty across the collection of models and provides one of the key quantities we use to estimate the error scale we introduce in Sec. III.

III. BAYESIAN MODEL COMBINATION

Our application of BMC closely follows Ref. [13], where the models discussed in the previous section satisfy the "reasonable model" criterion defined therein. In the following, we present a brief overview of the method and highlight new developments, particularly those concerning the heteroscedastic likelihood model.

To construct the BMC, a set of m existing models $\mathcal{M}_k$, $k \in \{1, 2, \ldots, m\}$, with n predictions for observables $\boldsymbol{y} \equiv (y_i)$, $i \in \{1, 2, \ldots, n\}$, is carefully combined through an experimental dataset. In contrast to other model combination approaches [11,12], the BMC method combines the dominant principal components (PCs) of the model ensemble, aiming to enhance predictive capability by mitigating overfitting while reducing the influence of redundant or closely related models on the final mixture.

Following Ref. [13] we construct a matrix of model calculations $\boldsymbol{X^0} \equiv (X^0_{i,k})$, where each column k represents the predictions of model $\mathcal{M}_k$ across the observables of interest y_i (binding energies, separation energies, etc.). The centered matrix of model calculations $\boldsymbol{X^c} \equiv (X^c_{i,k})$ is then given by

$$X^c_{i,k} = X^0_{i,k} - \phi_0(x_i), \tag{3}$$

where

$$\phi_0(x_i) \equiv \frac{1}{m}\sum_{k=1}^{m} \mathcal{M}_k(x_i) \tag{4}$$

is the average model prediction vector and $\boldsymbol{x} \equiv (x_i)$ the vector of inputs defining observable y_i such as proton Z_i and neutron N_i numbers, as well as the observable class such as the separation energy indicators in Eq. (1). Each column k of the centered matrix therefore characterizes how a model $\mathcal{M}_k$ deviates from the average model prediction ϕ_0. We decompose this matrix using the singular value decomposition (SVD) algorithm [28],

$$\begin{aligned} \boldsymbol{X^c}_{n\times m} &= \boldsymbol{U}_{n\times n}\ \boldsymbol{S}_{n\times m}\ \boldsymbol{V}^T_{m\times m} \\ &\approx\ \hat{\boldsymbol{U}}_{n\times p}\ \hat{\boldsymbol{S}}_{p\times p}\ \hat{\boldsymbol{V}}^T_{p\times m}, \end{aligned} \tag{5}$$

and create an approximate representation of the centered matrix $\hat{\boldsymbol{X}}^c_{n\times m}$ by truncating its dimensions and keeping the first p singular vectors. This truncation is a dimensionality reduction approach that provides the principal components [29], which represent the axes of maximum variability from the mean ϕ_0 among the collection of models. These components, denoted by $\phi_j(\boldsymbol{x})$ with $j \in \{1, 2, \ldots, p \leqslant m\}$, are the jth retained principal components obtained from the SVD of the centered model matrix X^c. These form the basis vectors for a lower-dimensional PC space. The subset of components retained for the BMC is represented by this basis. The number of retained components p is a hyperparameter of the model and is usually selected by evaluating the performance on a set of validation data, similar to traditional machine learning approaches [30].

For a chosen value p, the BMC is written as

$$f^\dagger(x_i;\boldsymbol{b}) = \phi_0(x_i) + \sum_{j=1}^{p} b_j \phi_j(x_i), \tag{6}$$

where the weights $\boldsymbol{b} \equiv (b_j)$ represent the free parameters of the BMC model and are calibrated within a Bayesian framework using the selected experimental data. The likelihood model we use expresses the relationship between observations y_i and BMC predictions as

$$y_i = f^\dagger(x_i;\boldsymbol{b}) + \varepsilon_i\sigma_i, \quad \varepsilon_i \sim \mathcal{N}(0,1), \tag{7}$$

where the notation $\varepsilon_i \sim \mathcal{N}(0,1)$ indicates that the random variable ε_i is drawn from a normal (Gaussian) distribution with mean 0 and variance 1. We represent the error, or discrepancy, between model predictions $f^\dagger(x_i;\boldsymbol{b})$ and experimental data y_i through the term $\varepsilon_i\sigma_i$, with $\boldsymbol{\sigma} \equiv (\sigma_i)$ being the collection of error scales for all data.

In Ref. [13], the authors modeled the discrepancies in binding energies to be statistically independent of each other. In this work, we focus on the separation energies defined in Eq. (1). For the size of the errors, possibly the simplest assumption is that they share the same scale $\sigma_i \equiv \sigma_0$, and this overall value becomes a variable to be estimated through the Bayesian framework [31]. We refer to this approach as the *homoscedastic error model* (HoEM), which is the same procedure adopted in Refs. [13,14].

One particular weakness with the homoscedastic variance assumption when extrapolating is the highly discrepant behavior of different functionals outside of the region constrained by experimental observation. As illustrated in Fig. 1, while each model is well constrained in the region near stability where training data exist, the differing physical assumptions among the models, as well as variations in their calibration datasets, drive the models' predictions apart as they approach the extrapolation regime. In regions of significant model disagreement, a homoscedastic assumption artificially enforces a constant uncertainty bandwidth, underestimating the quality of fit in regions of the training domain that the ensemble reproduces well, while overestimating it in the predictive domain.

An alternative approach is to allow the error scale to vary across data points, a more flexible method that can improve the robustness of the predictions [32]. Since in this study we are interested in the ability to extrapolate beyond the regions of experimental data, we need to construct a function that estimates how the error scale varies across all isotopes [33]:

$$\sigma_i \equiv g(\boldsymbol{z}_i;\Lambda). \tag{8}$$

The error scale σ_i depends on each data point through $\boldsymbol{z}_i$, which are a set of features—or predictors—that, as we explain in more detail later, can themselves depend on the inputs x_i, and we expect to be correlated with the actual model error. The parameters Λ are learned within a Bayesian framework and help characterize the relationship between σ_i and the features $\boldsymbol{z}_i$. We refer to this approach as the *heteroscedastic error model* (HeEM).

Once we have specified the model combination $f^\dagger$ in Eq. (6), its relationship with the experimental observations in Eq. (7), and the expression for the error scale σ_i in Eq. (8), we construct our likelihood model [31] assuming conditional independence on the observations as

$$p(\boldsymbol{y}|\boldsymbol{b},\Lambda) \propto \prod_{i=1}^{n} \frac{1}{\sigma_i} \exp\left(-\frac{(f^\dagger(x_i;\boldsymbol{b}) - y_i)^2}{2\sigma_i^2}\right). \tag{9}$$

Equation (9) assumes that the discrepancies are conditionally independent across all data points entering the likelihood. This is a simplifying assumption, as neighboring separation energy indicators share a common binding energy term that is expected to induce nonzero correlations that the likelihood does not capture. In this work, we adopt this approximation, leaving the study of the impact of a correlated likelihood error model to a future work.

We note that the HoEM approach is contained within the more general HeEM approach by setting $g(z_i;\Lambda)=\sigma_0$ in Eq. (8), with $\Lambda=\sigma_0$ the only parameter to be estimated. By combining the constructed likelihood with a suitable prior $p(\boldsymbol{b},\Lambda)$, we construct the posterior distribution that will allow us to constrain the parameters of the BMC given the available experimental data:

$$p(\boldsymbol{b},\Lambda|\boldsymbol{y})\propto p(\boldsymbol{y}|\boldsymbol{b},\Lambda)p(\boldsymbol{b},\Lambda). \tag{10}$$

For the prior distribution of the weights $\boldsymbol{b}$, we follow the same approach as in Ref. [13], creating an independent normal distribution centered at zero. For the HoEM, we also follow the approach in Ref. [13] with a weakly informed prior through a Gamma distribution for the precision $1/\sigma_0^2$. For the heteroscedastic model, the same shape of Gamma distribution is used for the variance model parameters. This resulted in healthy convergence, and several choices of the Gamma rate parameter resulted in nearly identical posteriors, demonstrating the robustness of the fit to different priors. In all cases, samples from the posterior, Eq. (10), are obtained by using a Gibbs sampler. For the homoscedastic model, this is a standard two-block Gibbs sampler, while the heteroscedastic model uses a Metropolis-Hastings step embedded within each Gibbs sweep (a Metropolis-within-Gibbs scheme) [34]. Within each Gibbs step, candidate parameters Λ are updated via a random-walk Metropolis step with a Gaussian proposal. During a burn-in period of 5000 iterations, the step size is adaptively tuned to target an acceptance rate of 25%. Because successive Markov chain Monte Carlo (MCMC) draws are correlated, one can assess the sampling process by measuring the effective sample size (ESS). The ESS is defined as the number of independent draws that would give the same precision as the correlated chain [31]. Across observables and error models, the ESS of the error metrics ranges from the hundreds to the thousands, with higher parameter models having smaller ESS. The point predictions reported in our validation tables and figures represent the posterior expectation values obtained by averaging over the full MCMC chain generated by the Gibbs sampler. Because the BMC framework is linear in its weights, this posterior predictive mean corresponds to evaluating the model at the center of mass of the posterior distribution, preserving full alignment with our sampled uncertainty calibration metrics.

The uncertainties we report here are constructed from the full posterior predictive distribution. We propagate every retained sample of $(\boldsymbol{b},\Lambda)$ from the Gibbs chain through the model $f^{\dagger}$, generating a set of posterior predictions for each nucleus and observable. We then draw and add a simulated noise at the error scale implied by that sample's Λ rather than evaluating only the mean prediction $f^{\dagger}(x_i;\boldsymbol{b})$. This procedure therefore incorporates both parameter uncertainty from $\boldsymbol{b}$ and predictive uncertainty from Λ. The reported uncertainty is then taken to be the 68% credible interval of the set of sampled predictions, rather than an analytic propagation of a single fitted σ_i. The numerical hyperparameters of the MCMC chain are given in Table I.

TABLE I. MCMC sampler configuration (identical across all observables and error models).

Total iterations	200 000
Burn-in iterations	5000
Thinning	None (195 000 retained draws)
Sampler ($\boldsymbol{b}$)	Gibbs
Sampler (error model)	Metropolis-within-Gibbs
Prior on $\boldsymbol{b}$	$\mathcal{N}(0,\ 10^6\,I)$
Prior on $\alpha,\ \beta_j,\ \gamma_j,\ \delta$	Gamma(2, 10)

A. Heteroscedastic error model design

In order to develop an error model that describes the confidence of the BMC, we must identify statistical indicators that are correlated with the model's performance. One primary driver of model performance degradation is calibration drift: A statistical framework is typically well calibrated only where training data are available. Consequently, the distance of a given nucleus from the training domain of the Bayesian calibration emerges as a natural indicator for error scaling.

An intuitive space for evaluating this proximity is the (N,Z) grid of the nuclear domain, where the distance to the nearest training isotope is

$$\Delta(N,Z)=\min_{(N_i,Z_i)\in\mathcal{T}}\sqrt{(N-N_i)^2+(Z-Z_i)^2}, \tag{11}$$

where $\mathcal{T}$ is the training set. However, evaluating distance over a discrete integer lattice severely restricts the granularity and fidelity of the metric. Empirically, this coordinate distance exhibits a weak correlation with the observed residuals, rendering it an ineffective predictor of model confidence (see Fig. 2).

To obtain a more physically informative metric, we may instead assess nuclei in our predictive domain within the continuous space spanned by the dominant PCs. Projecting the

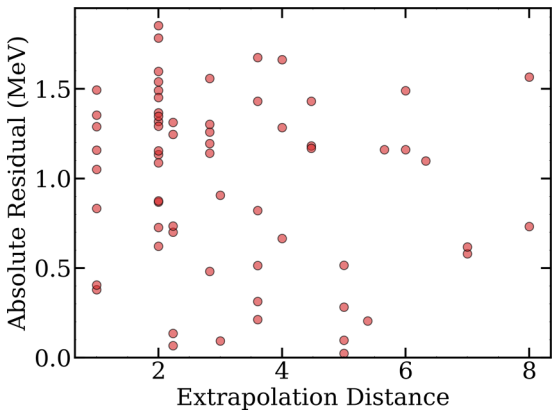


FIG. 2. For each nucleus in the validation set plotted is the distance (11) and the residual between the mean homoscedastic prediction and the experimental value.

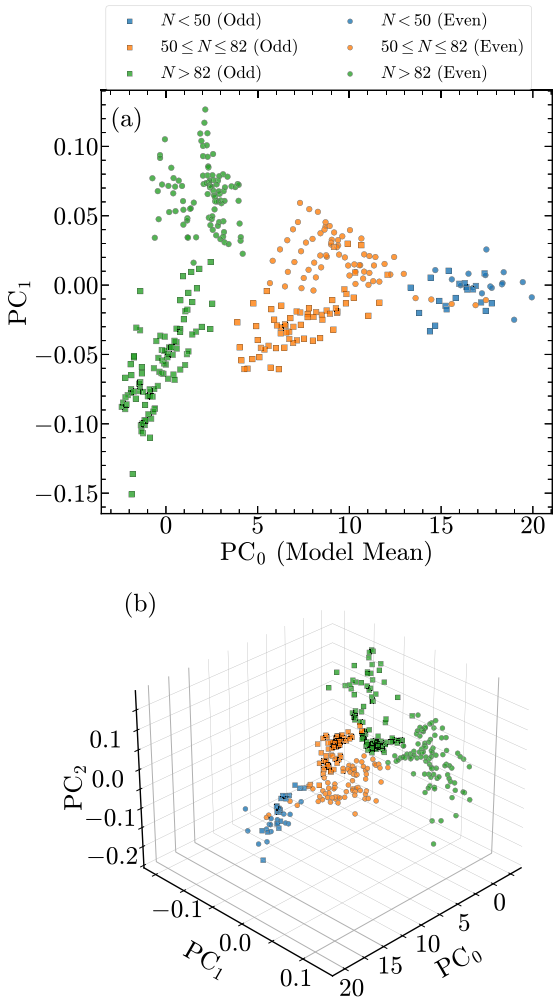


FIG. 3. Two-dimensional and three-dimensional visualizations of the model data in PC space. The grouping of nuclei of the same neutron shell and parity shows how nuclear structure is preserved in the principal component space.

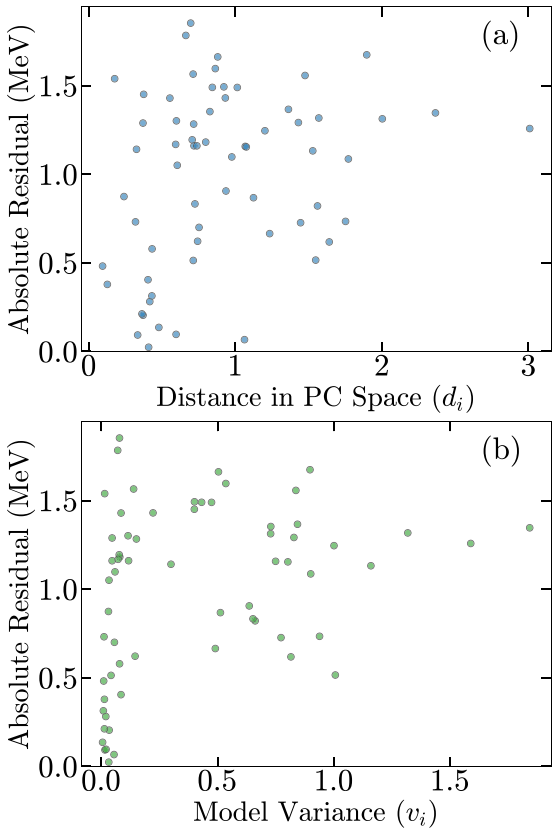


FIG. 4. Two uncertainty predictors plotted against the residual between the mean homoscedastic prediction and the true experimental values. Training data are omitted to avoid cluttering the plot.

model predictions onto the primary axes of variability maps each nucleus into a continuous, lower-dimensional PC space. For a given nucleus i, we construct a centered deviation vector $\mathbf{x}_i^c = \mathbf{x}_i^0 - \phi_0(x_i)$, where $\mathbf{x}_i^0$ represents the raw ensemble predictions and $\phi_0(x_i)$ is the average model prediction vector. The coordinates ζ_i of the nucleus in the reduced p-dimensional PC space are then obtained by projecting this deviation vector onto the truncated right singular vectors $V_{m\times p}$ determined from the SVD of the training matrix:

$$\zeta_i = \mathbf{x}_i^c V_{m\times p}, \tag{12}$$

where $V_{m\times p}$ contains the first p columns of V, defining the orthogonal axes of maximum variability across the model ensemble. As shown in Fig. 3, the nuclei in the PC space are clustered by familiar nuclear properties such as nuclear shell closure and odd-even parity. This emergent structure demonstrates that the principal components capture systematic trends in the model predictions. Principal component analysis naturally isolates the physical variations of the nuclei into orthogonal principal components, resulting in a continuous PC space that acts as a physical spectrum, organizing nuclei automatically by their structural properties. The continuous nature of this space ensures sufficient resolvable fidelity in the uncertainty predictor to accurately quantify a model's distance from the training data.

A second driver of performance degradation we consider is a larger intermodel spread, where the individual models exhibit significant mutual disagreement. In particular, a high model spread for a specific nucleus could indicate disagreement in the underlying physics model assumptions. This spread can be quantified as the variance σ^2 of the individual model predictions for each nucleus.

These two numerical predictors are taken as quantifiers of the primary mechanisms of degradation in prediction quality:

(1) distance d_i from the center of the training data in PC space;

(2) variance v_i of model predictions.

Figure 4 shows that the cross-variation between the two defined predictors and the absolute residual of the constant-variance BMC indeed share some level of correlation. Using these two predictors, we can write the error models we explore in this work:

$$\text{(HoEM) } \sigma_i^2 = \alpha, \tag{13a}$$

$$\text{(HeEM linear) } \sigma_i^2 = \alpha + \beta_1 d_i + \gamma_1 v_i, \tag{13b}$$

$$\text{(HeEM quadratic) } \sigma_i^2 = \alpha + \beta_1 d_i + \beta_2 d_i^2 + \gamma_1 v_i + \gamma_2 v_i^2 + \delta d_i v_i, \tag{13c}$$

where α, β_j, γ_j, and δ are the error model parameters Λ sampled from the constructed posterior distribution given by Eq. (10). Positivity of σ_i^2 is enforced by restricting $\alpha, \beta_j, \gamma_j, \delta > 0$ through their priors and by rejecting any Metropolis-Hastings proposal that violates this constraint. The first expression, Eq. (13a), corresponds to the constant-variance homoscedastic error model.

B. Evaluation metrics

To evaluate these error models rigorously, we introduce several statistical metrics to quantify the quality of the uncertainty calibration. RMSE is a metric that evaluates the average magnitude of the residual between a model's predictions and the target values. While RMSE is capable of evaluating the performance of the mean predictions of a model, it is not sensitive to the uncertainty of a model, and therefore not useful in describing the performance of an error model.

A better-suited metric that takes into account the model uncertainty is the reduced χ^2, calculated as

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{(y_{\mathrm{pred},i} - y_{\mathrm{exp},i})^2}{\sigma_i^2}, \tag{14}$$

where N is the total number of data points. While the χ^2 score is an accurate estimator of whether a model's variance is globally well scaled, it is not a complete measure of the distribution's quality. For example, if a model is generally underconfident but exhibits extreme overconfidence in specific cases, these opposing deviations may aggregate to a misleading total error. To obtain a more robust metric that accounts for these conflicting deviations, we analyze the empirical coverage across various credible intervals (CI). A posterior predictive distribution for a dataset is compared with the percentage of true data points that fall within a given nominal credible interval p.

To aggregate the deviation of the empirical coverage (C_{emp}) [35,36] from the nominal credible interval (p) across the full probability distribution, we compute the *mean absolute calibration error* (MACE):

$$\mathrm{MACE} = \frac{1}{M}\sum_{j=1}^{\mathrm{M}}|C_{\mathrm{emp}}(p_j) - p_j|, \tag{15}$$

where $M(=21)$ is the number of credible-interval levels used to compute the MACE, $p_j \in \{0, 0.05, 0.1, \ldots, 1\}$. A lower MACE value indicates a model that is better calibrated, meaning its predicted probabilities more accurately reflect the frequency of observed outcomes. Global metrics across credible intervals generalize the concept of predictive interval calibration to continuous regression tasks [37].

IV. VALIDATION OF THE BMC ON SYNTHETIC DATA

Having defined the various BMC error models and our evaluation metrics, in this section we proceed to test their extrapolation performance in a synthetic data experiment. In this procedure, each constituent DFT model is treated as a surrogate for the experimental data. We construct the BMC framework using the remaining set of models and calibrate it through synthetic data constrained exclusively to the experimentally known domain (AME2020). These calibration data consist of the numerical predictions generated by the DFT model that were excluded from the ensemble. By analyzing the performance of the BMC on the holdout test dataset, we can compare how adopting the various error models impacts the calculated observables well outside the training domain, for example, near the driplines.

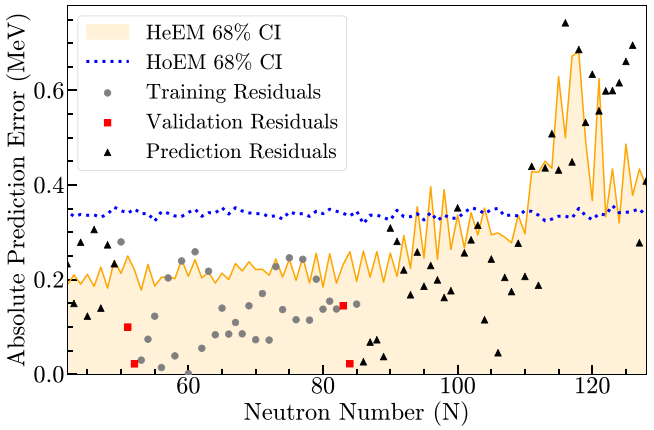


FIG. 5. Visualization of the synthetic data test. In this case, S_{1n} is predicted for the tin chain using DD-PC1 as the synthetic data model. The linear HeEM is used.

Figure 5 visualizes the synthetic validation for the case where the synthetic data are produced by DD-PC1. The homoscedastic model retains a static error scale determined by the training data, leading to underconfident predictions in the training regime and overconfident predictions in the extrapolation regime. In contrast, the linear heteroscedastic error model allows the CI to expand and successfully encompass the synthetic data even as they deviate significantly from the predicted mean.

Table II shows the RMSE and reduced χ^2 results of the homoscedastic and heteroscedastic error models compared against each other. For four of the six synthetic data models tested, the HoEM deviates significantly from unity. However, in almost all of these cases, both the linear and quadratic heteroscedastic models are closer to unity. This indicates that allowing the error magnitude to vary with identified indicators

TABLE II. RMSE and reduced χ^2 values for both error models predicting the S_{1n} values for the holdout validation set. Lower RMSE is better and χ^2 closer to unity is better.

	RMSE (MeV)			Reduced χ^2		
Synthetic data	HoEM	Linear HeEM	Quadratic HeEM	HoEM	Linear HeEM	Quadratic HeEM
DD-ME2	0.38	0.37	0.37	1.70	1.29	1.31
DD-PC1	0.22	0.22	0.22	0.46	0.71	0.74
DD-PCX	0.42	0.42	0.42	2.45	1.43	1.52
Fy(IVP)	1.62	1.64	1.65	1.30	1.16	1.26
UNEDF1	0.40	0.40	0.39	1.05	0.92	0.88
UNEDF2	0.29	0.30	0.30	0.95	0.94	1.00
Average	0.55	0.56	0.56	1.32	1.08	1.12

TABLE III. MACE score averaged across error models predicting synthetic S_{1n}. The performance is given for two prediction sets: Prediction (Nearby) includes the eight closest nuclei on either side of the known data along each isotopic chain, and Prediction (All) encompasses the full predictive domain. Lower MACE is better.

			Prediction	
Error model	Train	Validation	(Nearby)	(All)
HoEM	5.88	10.90	11.66	11.10
HeEM (linear)	4.15	7.31	7.61	8.26
HeEM (quadratic)	4.56	7.21	8.02	8.44

can address and rectify both systematically overconfident and systematically underconfident error models.

However, the improvement in χ^2 does not definitively demonstrate a more accurate error model. To do so, the MACE scores are calculated sequentially, treating each DFT model as the source of synthetic data. The resulting MACE scores were averaged to establish comprehensive benchmarks for the synthetic data test. The results for S_{1n} are listed in Table III. We find that both the linear and quadratic HeEMs lower the calibration error compared to the HoEM baseline within the validation boundaries. Crucially, even as model performance degrades within the extrapolation domain, metrics on the validation set remain highly predictive of performance across both prediction sets considered.

The theoretical models share a microscopic framework that introduces systematic biases, and consequently, the entire ensemble may collectively deviate from reality. Therefore, the true experimental values may be more challenging to predict than the synthetic calibration data. Despite this limitation, the synthetic experiment shows that the HeEM's calibration is not just an artifact of a single benchmark. Averaged over all six synthetic tests, the linear HeEM's validation MACE (7.31) is close to its prediction-set MACE (8.26). Overall, validation-set metrics are a reasonable proxy for extrapolation accuracy, at least within the scope of the DFT models used for this calibration.

Taken together, these synthetic benchmarks inform our choice of error model for the analysis of the real experimental data. While the quadratic HeEM achieves marginally tighter calibration on the validation set in some cases (see Table III), this advantage does not consistently carry over to the prediction sets, where the linear HeEM matches or outperforms its quadratic counterpart. Given that the linear HeEM achieves this performance with three error-model parameters compared to six for the quadratic form, we consider the linear form to be the more robust choice for extrapolation. We therefore adopt the linear HeEM as our primary heteroscedastic model when applying the BMC framework to the experimental AME2020 dataset in Sec. V.

V. MODEL APPLICATION TO AME2020 DATASET

Having validated the framework using synthetic data, we now apply BMC to extrapolate from the AME2020 dataset in the region of $Z = 46$ to $Z = 52$. For the training data, we use the binding energies, Q_α values, and neutron separation energies from $N = 40$ to $N = 130$ and $Z = 46, 48, 50, 52$. We use proton separation energies for $Z = 48$ to $Z = 52$ and from $N = 40$ to $N = 80$.

TABLE IV. RMSE results on the holdout validation set for separation energies (in MeV) for the BMC models and the individual models used to train BMC.

	Neutron (MeV)		Proton (MeV)	
Model	S_{1n}	S_{2n}	S_{1p}	S_{2p}
BMC (homoscedastic)	0.22	0.27	0.26	0.36
BMC (linear HeEM)	0.22	0.27	0.27	0.37
BMC (quadratic HeEM)	0.22	0.28	0.31	0.40
DD-ME2	0.80	0.70	0.81	0.93
DD-PC1	0.80	0.81	0.76	0.80
DD-PCX	0.66	0.75	0.79	1.12
Fy(IVP)	0.58	0.70	1.26	1.12
UNEDF1	0.43	0.53	1.05	0.84
UNEDF2	0.35	0.44	0.90	0.88

Table IV shows the RMSE of the separation energies for the validation set using homoscedastic and heteroscedastic error models. For both HoEM and HeEM, the BMC process yields RMSE performance that far exceeds the performance of any individual microscopic model. This, again, demonstrates BMC's core capability to predict nuclear observables with greater accuracy than its constituent models.

As shown in Table V, the HoEM consistently produces χ^2 scores significantly below unity. This indicates that the global variance assumption forces the model to adopt overly conservative uncertainties on the validation set to accommodate outliers. In contrast, the HeEM formulations achieve values closer to unity, demonstrating superior calibration of the error scale. Additionally, the linear form of the error model outperforms the quadratic form, especially for proton separation energies.

In Fig. 6, we compare the empirical coverage probability [38,39] for different error models of S_{1n}. For each model, a posterior predictive distribution for the validation dataset is compared with the percentage of true validation data points that fall within a given credible interval. While each model performs well for this case, both heteroscedastic models outperform the homoscedastic model. To quantify the overall quality of uncertainty estimates across all credible intervals, we show the MACE (15) in Table VI. The HoEM generally exhibits the largest average spread about the mean; however, the optimal functional form of the error model depends on the observable under consideration. For example, for the

TABLE V. Reduced χ^2 values for BMC models on the holdout validation set. χ^2 closer to unity is better.

	Neutron		Proton	
Model	S_{1n}	S_{2n}	S_{1p}	S_{2p}
HoEM	0.57	0.66	0.49	0.69
Linear HeEM	0.86	0.86	0.71	0.74
Quadratic HeEM	0.81	0.75	0.55	0.53

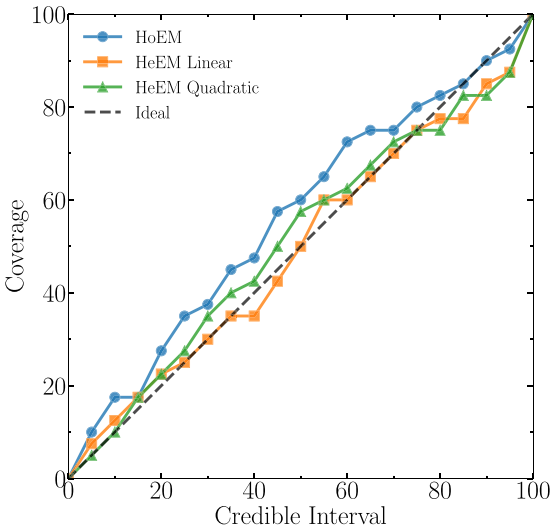


FIG. 6. Empirical coverage probability for different models of S_{1n}.

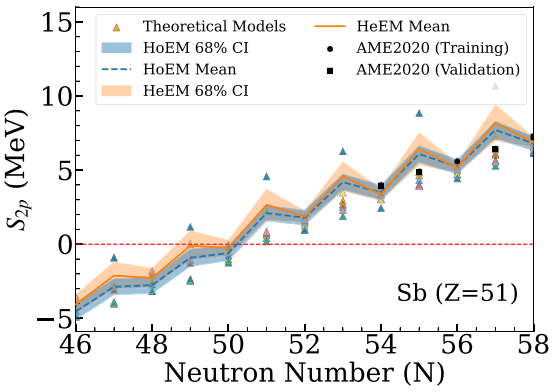


FIG. 7. HoEM and linear HeEM predictions for S_{2p} of the Sb ($Z = 51$) chain alongside the DFT model predictions. The HeEM band adapts to model divergence when N is odd, while the HoEM uncertainty band remains constant across the entire chain.

single-proton separation energy, the quadratic HeEM outperforms both the homoscedastic error model and the linear error model. However, for S_{2p}, it exhibits the worst performance of any error model across all observables. This demonstrates the importance of exploring different error models when applying the BMC across multiple observables.

While the primary motivation for introducing a varying uncertainty scale is to account for the degradation in mean performance as the model extrapolates progressively farther from the training data, the choice of a varying error scale also enables the error model to reproduce other qualities of the theoretical models. Figure 7 demonstrates the qualitative difference between different approaches for a representative case of Sb ($Z = 51$). The two error models produce similar central values and similar uncertainty bounds for even N. However, for odd-N nuclei, model predictions strongly deviate from experiment, suggesting missing physics in theoretical descriptions of odd-odd nuclei. Since the variance of model calculations v_i is one of the predictors in Eq. (8), the HeEM BMC is naturally able to capture this behavior for odd-N nuclei.

The farthest extrapolation from experimental data is for the drip lines; hence, the limits of nuclear existence are uncertain. The most dramatic difference between our homoscedastic and heteroscedastic implementations manifests precisely at these limits. While the HoEM maintains an essentially constant uncertainty magnitude by construction, the HeEM reveals a progressive increase in predicted error, with uncertainties growing 5 keV on average per neutron number away from known data across all observables until $N = 107$. As the theoretical models diverge for $N \geqslant 108$, the predicted uncertainty increases dramatically (see Fig. 5), more than doubling on average in magnitude relative to the known data region. This behavior highlights a key strength of the heteroscedastic approach: By linking the error scale directly to the intermodel spread v_i rather than to a fixed function of distance alone, the HeEM is able to distinguish between regions where extrapolation is genuinely unconstrained and regions where, despite being far from data, the underlying physics still imposes a strong, shared constraint on the models.

The recent mass measurements of neutron-deficient Cd isotopes reported in Ref. [40] provide an opportunity to test the performance of our BMC frameworks. As detailed in Table VII, the measurements for S_{1n} at the closed shell demonstrate reasonable agreement with HeEM producing predictions within 1σ agreement for ^{96}Cd and ^{98}Cd. Moving just below the shell closure exposes the limitations of the constituent EDFs in capturing the extreme odd-even staggering in the empirical mass surface. At ^{97}Cd, the experimental S_{1n} drops precipitously, which is not captured by our DFT models. This common bias is not captured by the intermodel spread v_i driving the HeEM error scale, and so the predicted uncertainty can underestimate the true deviation.

Prior to recent direct observations, the empirical properties of ^{104}Te near the doubly magic ^{100}Sn boundary were constrained by Ref. [41], which reported a decay energy of 4.9(2)

TABLE VI. Mean absolute calibration error on the holdout validation set for proton and neutron separation energies. Lower MACE is better.

	Neutron		Proton		
Model	S_{1n}	S_{2n}	S_{1p}	S_{2p}	Average
HoEM	6.07	3.21	7.14	7.38	5.95
Linear HeEM	2.14	2.97	5.52	6.42	4.26
Quadratic HeEM	3.21	2.38	4.03	10.47	5.02

TABLE VII. Predictions and newly established experimental values (in MeV) for S_{1n} in $^{96-98}$Cd. The experimental values are derived from the mass excesses reported in Ref. [40], utilizing the extrapolated mass of ^{95}Cd for the ^{96}Cd calculation.

Obs.	Model	^{96}Cd	^{97}Cd	^{98}Cd
S_{1n}	HoEM	16.61(41)	14.15(41)	15.42(42)
	Linear HeEM	16.65(43)	14.13(35)	15.46(42)
	Quadratic HeEM	16.68(46)	14.11(34)	15.51(50)
	Expt.	16.99(14)	12.91(4)	15.21(3)

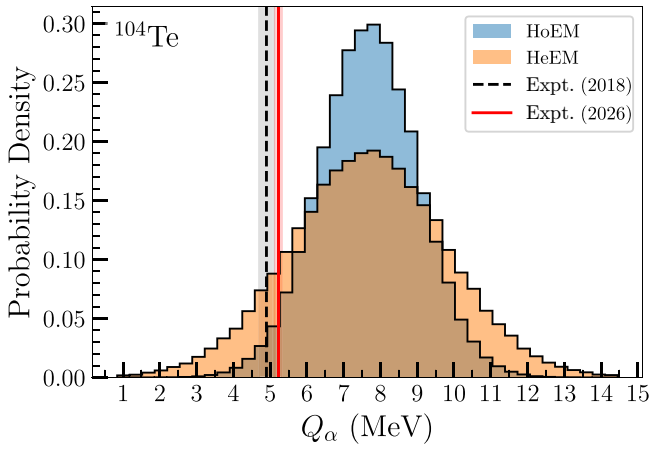


FIG. 8. A histogram of all generated HoEM and linear HeEM predictions for Q_α for ^{104}Te in the posterior alongside the old [41] and recent [42] experimental values.

MeV. The recent study [42] isolated the ground-state α-decay chain to reveal a significantly higher decay energy of $Q_\alpha = 5.23(9)$ MeV. As demonstrated in Fig. 8, the new BMC framework provides better agreement with both measurements than the previous HoEM framework. However, similar to Ref. [40], each of the individual DFT models predicts a higher Q_α value. This applies not just to ^{104}Te, but also down the rest of the Te chain all the way to ^{112}Te. This demonstrates how these Q values can be used as calibration data to further improve EDFs in this region.

To illustrate the practical impact of heteroscedastic uncertainties, we present detailed predictions for four neutron-deficient tin isotopes in Table VIII. These nuclei ($^{96-99}$Sn) lie beyond current experimental reach on the proton-rich side and demonstrate how the HeEM and HoEM approaches yield similar mean predictions but different uncertainty estimates. Even for the same nuclei, the outcome varies for different observables, suggesting that agreement between theoretical models is not necessarily uniform across observables.

The proton-rich frontier presents a challenging test case for local extrapolation. Figure 9 displays the predicted probabilities of existence for nuclei in the tin region approaching the proton drip line. The probability of existence [2] is defined as the probability that the selected nucleus is bound to one- and two-particle emission:

$$P(\mathrm{ex})(Z,N) = P(S_{1\mathfrak{n}}(Z,N) > 0,\ S_{2\mathfrak{n}}(Z,N) > 0), \quad (16)$$

where $\mathfrak{n} \in \{n, p\}$ denotes the neutron- or proton-rich region, and the probability is computed over the posterior predictive distribution of the BMC model.

While both error models largely agree on the location of the drip line, they differ substantially in their predictions for individual nuclei in this region. For example, the transition gradients from confidently bound to unbound exhibit distinct structural differences between the two frameworks. The HoEM, constrained by a static error scale, predicts a relatively sharp transition across the isotonic lines, with probabilistic assignments for transitional nuclei that remain rigid. In contrast, the HeEM produces a probability gradient that correlates with local model divergence. While the HoEM may confidently classify a transitional nucleus as unbound, the expanded uncertainty bands of the HeEM at the extrapolative frontier allow for a nontrivial probability of existence. A negative central separation-energy prediction does not, by itself, imply that a nucleus is theoretically excluded from existing, and only a fully probabilistic treatment of the uncertainty can reveal whether such marginal cases remain experimentally viable targets. Both the HoEM and the HeEM confidently report that ^{104}Te is proton unbound, a claim that agrees with the AME2020 record of the two-proton separation energy as -0.73 MeV [27].

The same process was used to predict neutron drip lines in Fig. 10. Similar to the proton drip line, the two error models identify similar locations for the dripline. However, for all three chains, the linear HeEM produces a gentler dropoff in the likelihood of existence than the HoEM, reflecting the genuine theoretical disagreement as to where exactly each chain becomes unbound, rather than the artificially sharp cutoff imposed by a fixed error scale. Across all models and chains, there is locally boosted model consensus at the $N = 126$ shell gap, where both the HoEM and HeEM predict a rapid decrease in existence likelihood.

VI. SUMMARY AND CONCLUSIONS

In this work, we presented an approach for nuclear uncertainty quantification by introducing physically motivated

TABLE VIII. Predictions in MeV for $^{96-99}$Sn nuclear observables. The linear HeEM was used for these data.

Observable	Model	^{96}Sn	^{97}Sn	^{98}Sn	^{99}Sn
BE	HoEM	762.09(84)	777.96(84)	795.10(86)	810.76(84)
	HeEM	762.17(101)	778.03(94)	795.15(101)	810.79(88)
S_{1n}	HoEM	17.87(28)	15.32(29)	17.14(29)	14.63(30)
	HeEM	17.88(24)	15.30(24)	17.13(23)	14.59(28)
S_{2n}	HoEM	33.63(32)	33.09(34)	32.38(34)	31.72(34)
	HeEM	33.66(32)	33.11(31)	32.35(34)	31.66(35)
S_{1p}	HoEM	0.84(37)	1.44(39)	1.96(38)	2.46(38)
	HeEM	0.77(32)	1.37(32)	1.89(31)	2.39(33)
S_{2p}	HoEM	−0.44(44)	0.66(43)	1.90(44)	2.93(43)
	HeEM	−0.61(53)	0.47(50)	1.75(47)	2.85(37)

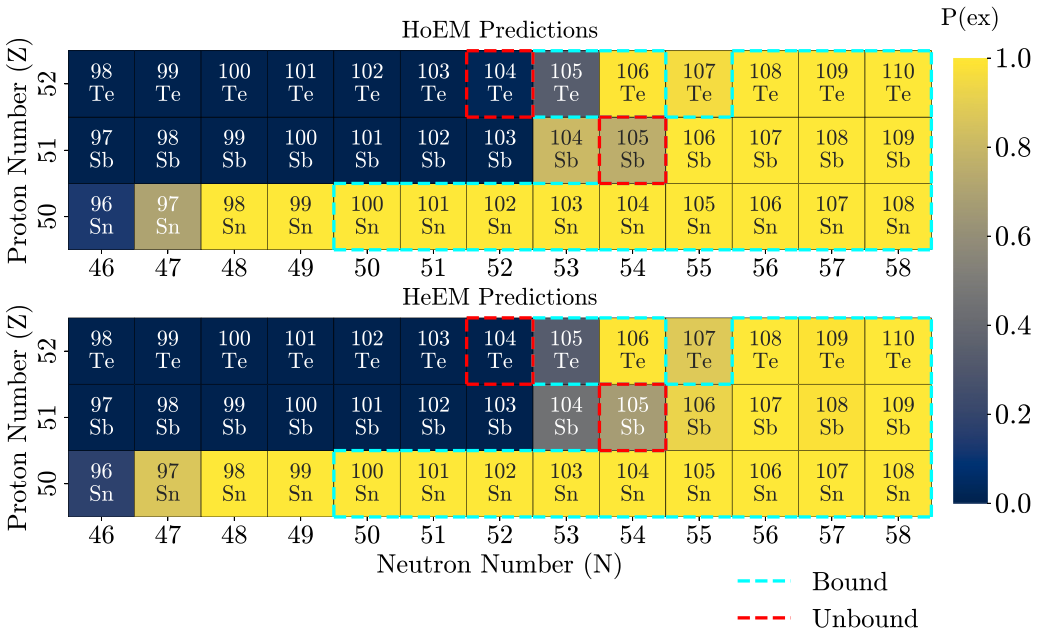


FIG. 9. A heatmap of the probability of existence [Eq. (16)] for bound and unbound nuclei predicted using homoscedastic and heteroscedastic approaches. The linear HeEM was used for the heteroscedastic predictions.

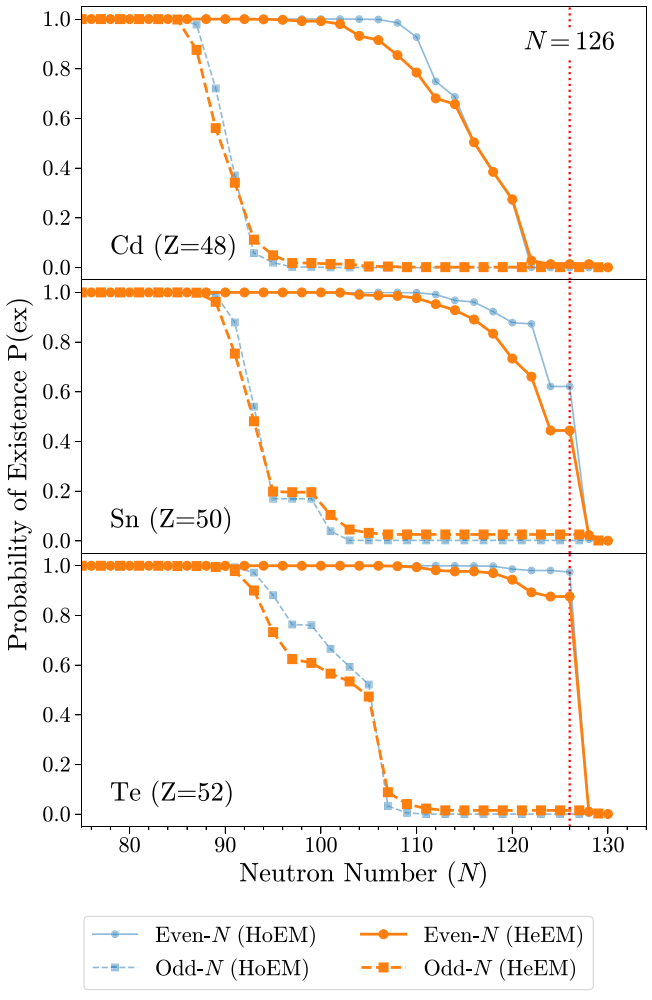


FIG. 10. Probability of existence for even-Z chains in the tin region. Predictions are made using the linear HeEM.

heteroscedasticity into the BMC framework of nuclear DFT models. While previous benchmarking of nuclear mass models has largely relied on an assumed constancy of the variance of predictions across the nuclear landscape, our results demonstrate that this simplification fails to account for the inherent degradation of model predictive power as one extrapolates toward the limits of nuclear stability. By allowing the error scale to vary across isotopes through a dependency on the disagreement across models and the principal component distance, the HeEM provides a more robust and statistically rigorous description of the systematic uncertainties associated with the nuclear mass surface.

Before applying this framework to experimental data, we validated it using a synthetic data test, in which each constituent DFT model was treated as a surrogate for the experimental data. This procedure allowed the simulation of genuine extrapolation to confirm that the HeEM consistently produces better-calibrated uncertainties than the HoEM and to motivate our choice of the linear HeEM as the primary model for the AME2020 application.

Using HeEM, we were able to achieve strong performance metrics (RMSE, χ^2, and MACE) on the validation set for AME2020 for every observable we investigated, whereas the HoEM was either conservative or overconfident. We identified the mean absolute calibration error as a strong predictor of extrapolative performance. Specifically, models that perform well on the validation set at the boundaries of known data, such as the linear HeEM, maintain favorable performance well into the predictive domain. In the extreme extrapolation regime for the neutron-rich region ($N \geqslant 100$), we observed a sharp surge in uncertainty, marking the transition from well-constrained theoretical predictions to domains where model agreement deteriorates.

The practical impact of this framework is most evident in the prediction of the nuclear drip lines. While homoscedastic

and heteroscedastic approaches may yield similar mean predictions, their probabilistic assessments of nuclear existence differ substantially in critical regions.

We would like to note the limitations of the present framework. First, the likelihood in Eq. (9) assumes conditional independence of the discrepancies across data points. One must remember that since separation energies are derived quantities of the binding energies, they have inherent nonzero correlations. Second, shared model biases are reproduced by all BMC mean estimates. Therefore, although the statistical framework we presented in this work definitely improves the predictive power of the *collection* of models, it cannot be used as a substitute for systematically improving the underlying nuclear theory.

Ultimately, this study underscores the necessity of moving beyond global error scales in nuclear theory. By adopting heteroscedastic uncertainties that correlate with local model divergence and distance from known data, we provide a more faithful representation of the limits of our theoretical understanding.

ACKNOWLEDGMENTS

Discussions with Franziska Maier are gratefully acknowledged. This material is based upon work supported by the U.S. Department of Energy under Awards No. DE-SC0013365 and No. DE-SC0023688 (Office of Science, Office of Nuclear Physics), No. DOE-DE-NA0004245 (NNSA, the Stewardship Science Academic Alliances program), No. DE-SC0023175 (Office of Science, NUCLEI SciDAC-5 collaboration), and by the National Science Foundation under Award No. 2004601 (CSSI program, BAND collaboration).

DATA AVAILABILITY

The data that support the findings of this article are openly available [43], embargo periods may apply.

[1] J. Erler, N. Birge, M. Kortelainen, W. Nazarewicz, E. Olsen, A. M. Perhac, and M. Stoitsov, The limits of the nuclear landscape, Nature (London) **486**, 509 (2012).

[2] L. Neufcourt, Y. Cao, S. A. Giuliani, W. Nazarewicz, E. Olsen, and O. B. Tarasov, Quantified limits of the nuclear landscape, Phys. Rev. C **101**, 044307 (2020).

[3] S. R. Stroberg, J. D. Holt, A. Schwenk, and J. Simonis, *Ab initio* limits of atomic nuclei, Phys. Rev. Lett. **126**, 022501 (2021).

[4] J. J. Cowan, C. Sneden, J. E. Lawler, A. Aprahamian, M. Wiescher, K. Langanke, G. Martínez-Pinedo, and F.-K. Thielemann, Origin of the heaviest elements: The rapid neutron-capture process, Rev. Mod. Phys. **93**, 015002 (2021).

[5] M. Bender, P.-H. Heenen, and P.-G. Reinhard, Self-consistent mean-field models for nuclear structure, Rev. Mod. Phys. **75**, 121 (2003).

[6] *Energy Density Functional Methods for Atomic Nuclei*, edited by N. Schunck (IOP Publishing, Bristol, 2019).

[7] N. Schunck, J. D. McDonnell, J. Sarich, S. M. Wild, and D. Higdon, Error analysis in nuclear density functional theory, J. Phys. G: Nucl. Part. Phys. **42**, 034024 (2015).

[8] Q. Yuan, P. Qi, X. Xiao, X. Wang, J. He, G. Long, Z. Duan, Y. Dai, R. Yan, G. Yu, and H. Yang, Machine learning-driven high-precision model for α-decay energy and half-life prediction of superheavy nuclei, Phys. Scr. **101**, 176004 (2026).

[9] R. Utama, J. Piekarewicz, and H. B. Prosper, Nuclear mass predictions for the crustal composition of neutron stars: A Bayesian neural network approach, Phys. Rev. C **93**, 014311 (2016).

[10] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Bayesian approach to model-based extrapolation of nuclear observables, Phys. Rev. C **98**, 034318 (2018).

[11] V. Kejzlar, L. Neufcourt, and W. Nazarewicz, Local Bayesian Dirichlet mixing of imperfect models, Sci. Rep. **13**, 19600 (2023).

[12] L. Neufcourt, Y. Cao, W. Nazarewicz, E. Olsen, and F. Viens, Neutron drip line in the Ca region from Bayesian model averaging, Phys. Rev. Lett. **122**, 062502 (2019).

[13] P. Giuliani, K. Godbey, V. Kejzlar, and W. Nazarewicz, Model orthogonalization and Bayesian forecast mixing via principal component analysis, Phys. Rev. Res. **6**, 033266 (2024).

[14] C. M. Ireland, G. Bollen, S. E. Campbell, X. Chen, H. Erington, N. D. Gamage, K. Godbey, A. M. Houff, C. Izzo, B. Knight, *et al.*, Mass of ^{101}Sn and Bayesian extrapolations to the proton drip line, Phys. Rev. C **113**, L021302 (2026).

[15] T. Nikšić, D. Vretenar, and P. Ring, Relativistic nuclear energy density functionals: Adjusting parameters to binding energies, Phys. Rev. C **78**, 034318 (2008).

[16] E. Yüksel, T. Marketin, and N. Paar, Optimizing the relativistic energy density functional with nuclear ground state and collective excitation properties, Phys. Rev. C **99**, 034318 (2019).

[17] G. A. Lalazissis, T. Nikšić, D. Vretenar, and P. Ring, New relativistic mean-field interaction with density-dependent meson-nucleon couplings, Phys. Rev. C **71**, 024312 (2005).

[18] M. Kortelainen, J. McDonnell, W. Nazarewicz, P.-G. Reinhard, J. Sarich, N. Schunck, M. V. Stoitsov, and S. M. Wild, Nuclear energy density optimization: Large deformations, Phys. Rev. C **85**, 024304 (2012).

[19] M. Kortelainen, J. McDonnell, W. Nazarewicz, E. Olsen, P.-G. Reinhard, J. Sarich, N. Schunck, S. M. Wild, D. Davesne, J. Erler, and A. Pastore, Nuclear energy density optimization: Shell structure, Phys. Rev. C **89**, 054314 (2014).

[20] J. Karthein, C. M. Ricketts, R. F. G. Ruiz, J. Billowes, C. L. Binnersley, T. E. Cocolios, J. Dobaczewski, G. J. Farooq-Smith, K. T. Flanagan, G. Georgiev, *et al.*, Electromagnetic properties of indium isotopes elucidate the doubly magic character of ^{100}Sn, Nat. Phys. **20**, 1719 (2024).

[21] Y. Tian, Z.-y. Ma, and P. Ring, Axially deformed relativistic Hartree Bogoliubov theory with a separable pairing force, Phys. Rev. C **80**, 024313 (2009).

[22] P.-G. Reinhard and W. Nazarewicz, Toward a global description of nuclear charge radii: Exploring the Fayans energy density functional, Phys. Rev. C **95**, 064328 (2017).

[23] T. Nikšić, N. Paar, D. Vretenar, and P. Ring, DIRHB—A relativistic self-consistent mean-field framework for atomic nuclei, Comput. Phys. Commun. **185**, 1808 (2014).

[24] P. Marević, N. Schunck, E. M. Ney, R. N. Pérez, M. Verriere, and J. O'Neal, Axially-deformed solution of the Skyrme-Hartree-Fock-Bogoliubov equations using the transformed harmonic oscillator basis (IV) hfbtho (v4.0): A new version of the program, Comput. Phys. Commun. **276**, 108367 (2022).

[25] P.-G. Reinhard, B. Schuetrumpf, and J. Maruhn, The axial Hartree–Fock + BCS code SkyAx, Comput. Phys. Commun. **258**, 107603 (2021).

[26] S. Perez-Martin and L. M. Robledo, Microscopic justification of the equal filling approximation, Phys. Rev. C **78**, 014304 (2008).

[27] M. Wang, W. Huang, F. Kondev, G. Audi, and S. Naimi, The AME 2020 atomic mass evaluation (II). Tables, graphs and references, Chin. Phys. C **45**, 030003 (2021).

[28] A. Blum, J. Hopcroft, and R. Kannan, *Foundations of Data Science* (Cambridge University Press, Cambridge, 2020).

[29] I. T. Jolliffe, *Principal Component Analysis* (Springer, New York, 2002).

[30] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, Cambridge, 2019).

[31] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis, Texts in Statistical Science Series*, 3rd ed. (Chapman & Hall/CRC, Boca Raton, FL, 2013), pp. xxvi + 668.

[32] R. J. Carroll and D. Ruppert, *Transformation and Weighting in Regression* (Chapman and Hall/CRC, Boca Raton, FL, 2017).

[33] M. Davidian and R. J. Carroll, Variance function estimation, J. Am. Stat. Assoc. **82**, 1079 (1987).

[34] P. D. Hoff, *A First Course in Bayesian Statistical Methods* (Springer, New York, 2009), Vol. 580.

[35] R. Schall, The empirical coverage of confidence intervals: Point estimates and confidence intervals for confidence levels, Biom. J. **54**, 537 (2012).

[36] T. P. Morris, I. R. White, and M. J. Crowther, Using simulation studies to evaluate statistical methods, Stat. Med. **38**, 2074 (2019).

[37] V. Kuleshov, N. Fenner, and S. Ermon, Accurate uncertainties for deep learning using calibrated regression, in *Proceedings of the 35th International Conference on Machine Learning*, edited by J. Dy and A. Krause (PMLR, Stockholm, Sweden, 2018), Vol. 80, pp. 2796–2804.

[38] T. Gneiting, F. Balabdaoui, and A. E. Raftery, Probabilistic forecasts, calibration and sharpness, J. R. Stat. Soc. Ser. B Stat. Methodol. **69**, 243 (2007).

[39] T. Gneiting and A. E. Raftery, Strictly proper scoring rules, prediction, and estimation, J. Am. Stat. Assoc. **102**, 359 (2007).

[40] D. Lange, D. Atanasov, M. Au, A. Belley, M. Benhatchi, K. Blaum, R. B. Cakirli, P. F. Giesel, A. Herlert, J. D. Holt, *et al.*, Constraining the trend of the $N = 50$ shell gap towards ^{100}Sn with the masses of $^{96–98}$Cd, arXiv:2604.18287.

[41] K. Auranen, D. Seweryniak, M. Albers, A. D. Ayangeakaa, S. Bottoni, M. P. Carpenter, C. J. Chiara, P. Copp, H. M. David, D. T. Doherty, *et al.*, Superallowed α decay to doubly magic ^{100}Sn, Phys. Rev. Lett. **121**, 182501 (2018).

[42] I. Cox, R. Grzywacz, T. T. King, K. P. Rykaczewski, S. Nishimura, R. Yokoyama, N. Fukuda, N. Kitamura, S. Go, C. Mazzocchi, *et al.*, Direct observation of the superallowed α-decay of ^{104}Te, Nature (London) **654**, 52 (2026).

[43] B. Knight, S. Lalit, P. Giuliani, K. Godbey, W. Nazarewicz, A. Ravlić, and P.-G. Reinhard, Heteroscedastic error model data repository, https://github.com/syiman/ProtonDripPublic.