frontiers

Bayesian averaging for ground state masses of atomic nuclei in a Machine Learning approach

M. R. Mumpower [1,*], **M. Li** [1,2] **and T. M. Sprouse** [1] **and B. S. Meyer** [2] **and A. E. Lovell** [1] **and A. T. Mohan** [3]

[1] *Los Alamos National Laboratory, Theoretical Division, Los Alamos, NM 87545, USA*
[2] *Clemson University, Department of Physics and Astronomy, Clemson, SC 29634-0978, USA*
[3] *Los Alamos National Laboratory, Computational Division, Los Alamos, NM 87545, USA*

Correspondence*:
Corresponding Author
mumpower@lanl.gov

ABSTRACT

We present global predictions of the ground state mass of atomic nuclei based on a novel Machine Learning (ML) algorithm. We combine precision nuclear experimental measurements together with theoretical predictions of unmeasured nuclei. This hybrid data set is used to train a probabilistic neural network. In addition to training on this data, a physics-based loss function is employed to help refine the solutions. The resultant Bayesian averaged predictions have excellent performance compared to the testing set and come with well-quantified uncertainties which are critical for contemporary scientific applications. We assess extrapolations of the model's predictions and estimate the growth of uncertainties in the region far from measurements.

Keywords: Atomic Nuclei, Bayesian Averaging, Binding Energies and Masses, Machine Learning - ML, Nuclear Physics, Computational Physiscs

1 INTRODUCTION

Mass is a defining quantity of an atomic nucleus and appears ubiquitously in research efforts ranging from technical applications to scientific studies such as the synthesis of the heavy elements in astrophysical environments [1, 2]. While accurate nuclear data of masses is available for nuclei that are relatively stable, the same is not true for nuclei farther away from beta stability because measurements on radioactive nuclei are exceedingly challenging [3]. As a consequence, theoretical models of atomic nuclei are required for extrapolations used in present-day scientific applications [4].

The goal of theoretical nuclear models is to describe all atomic nuclei (from light to heavy) using fundamental interactions. Attainment of this challenging goal remains elusive, however, due to the sheer complexity of modeling many-body systems with Quantum Chromodynamics [5]. To understand the range of nuclei that may exist in nature, mean-field approximations are often made which simplify complex many-body dynamics into a non-interacting system of quasi-particles where remaining residual interactions

can be added as perturbations [6]. A consequence of this approximation is that current nuclear modeling efforts are unable to describe the rich correlations that are found across the chart of nuclides.

In contrast, Machine Learning (ML) based approaches do not have to rely on the assumption of modeling nuclei from a mean-field. This provides freedom in finding solutions that contemporary modeling may not be capable of ascertaining. Furthermore, Bayesian approaches to ML afford the ability to associate predictions with uncertainties [7, 8]. Such tasks are more difficult to achieve in modern nuclear modeling due to relatively higher computational costs.

ML approaches in nuclear physics were pioneered by J.W. Clark and colleagues [9, 10]. These studies were the first to show that networks could approximate stable nuclei, learn to predict masses and analyze nuclear systematics of separation energies as well as spin-parity assignments [11, 12]. Powered by open-source frameworks, research into ML methods has seen a recent resurgence in nuclear physics [13]. ML approaches have shown promise in optimizing data and experiments [14], building surrogate models of density functional theory [15], and describing quantum many-body wave functions for light nuclei[16, 17].

Several research groups are actively pursuing the problem of describing nuclear systems with ML from a more data-centric approach. These efforts currently attempt to improve existing nuclear models by adding correction terms [18]. Gaussian Processes (GP) have also been used for model averaging [19], but this approach is inherently limited to where data is known as GP methods typically revert to the mean when extrapolating. A further limitation to training ML models on residuals (or the discrepancy of theoretical model predictions with experimental data) is that the methods are arbitrary. The changes learned by the network to improve one model will not be applicable to another. These approaches thus provide limited insight into the underlying missing physics in modern models of the atomic nucleus.

In Lovell et al. [20], a different approach was taken, where the masses of atomic nuclei were modeled directly with a neural network. It was shown that the masses of nuclei can be well described, and model predictions with increased fidelity correlate strongly with a careful selection of physically motivated input features. The selection of input features is especially important in ML applications [21, 22]. Following this work, Mumpower et al. [23] showed that the size of the training set can greatly be reduced, and the fidelity of model solutions increases drastically, when an additional physical constraint is introduced as a second loss function during model training.

The focus of this work is to present a Bayesian approach for combining precision data with theoretical predictions to model the mass of atomic nuclei. In Section 2, we present our ML algorithm and define the model hyperparameters. In Section 3, we show the results of our approach and assess the quality of model extrapolations. We end with a short summary.

2 METHODS

In this section, we outline our methodology: describe the neural network, define our physics-based feature space, list model hyperparameters, and discuss training.

2.1 Mixture Density Network

In a feed-forward neural network, inputs, $\mathbf{x}$, are mapped to outputs, $\mathbf{y}$, in a deterministic manner. We employ the Mixture Density Network (MDN) of Bishop [24] which differs from the standard approach. This ML network takes as inputs stochastic realizations of probability distributions and maps this to a mixture of Gaussians. Thus, the network fundamentally respects the probabilistic nature of both known

data and model predictions by both sampling the prior distribution of inputs and predicting the posterior distribution of the outputs.

Formally, the conditional probability can be written as

$$p(\mathbf{y}|\mathbf{x}) = \sum_{i=1}^{K} \pi_i(\mathbf{x}) \mathcal{N}\left(\mathbf{y}|\mu_i(\mathbf{x}), I\sigma_i(\mathbf{x})\right) \ , \tag{1}$$

where $\mathcal{N}$ is the normal distribution with means, $\mu_i(\mathbf{x})$, and standard deviations, $\sigma_i(\mathbf{x})$. The $\pi_i(\mathbf{x})$ represent the weighting of each Gaussian respectively. The covariance matrix is assumed to be diagonal, as indicated by the use of the notation $I\sigma_i$.

The neural network outputs are $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, and $\boldsymbol{\sigma}$ which depend only on the input training set information $\mathbf{x}$ and the network weights. For ease of reading the equation we have kept the dependence of the network weights implicit.

The hyperbolic tangent function $a(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ is used as the activation function for the neurons in the linear layers of the network . At the final layer a softmax function is used for the π_i so that the previous layer's output can be mapped to a vector that sums to unity. This choice ensures that the mixture of Gaussians can be safely interpreted as a probability. Our MDN uses the `PyTorch` Paszke et al. [25] framework and can be run on either CPU or GPU architectures.

2.2 Physics-based feature space

We now discuss the components of the input vector, $\mathbf{x}$. The ground state of an atomic nucleus comprises Z protons and N neutrons. While it is reasonable to start from these two independent features as inputs, [21] and Lovell et al. [20] reported that a modestly larger physics-based feature space drastically improves the prediction of masses. For this reason, we employ a combination of macroscopic and microscopic features that are of relevance to low-energy nuclear physics properties.

In addition to the proton number Z and neutron number N, we also use the mass number $A = Z + N$, and a measure of isospin asymmetry, $P_{\text{asym}} = \frac{N-Z}{A}$, as relevant macroscopic features. For the microscopic features that encode the quantized nature of atomic nuclei, we employ notions of pairing by considering the even-odd behavior of the proton, neutron, and mass numbers. This can calculated by observing the binary values of these quantities modulo 2; $Z_{\text{eo}} = Z \div 2$, $N_{\text{eo}} = N \div 2$, $A_{\text{eo}} = A \div 2$. A notion of shell structure is also important. To encode this information we include the number of valence nucleons or holes (beyond the mid-shell) from the nearest major closed shell for protons, V_p, and neutrons, V_n, respectively. The value of V_p or V_n is zero at a closed shell and reaches a maximum at the mid-shell. The number of valence nucleons is correlated with more complex excitations in nuclei, including collective behavior that may appear [26, 27]. The closed proton shells are set to 8, 20, 28, 50, 82, and 114. The closed neutron shells are set to 8, 20, 28, 50, 82, 126 and 184. These choices are free parameters in our modeling and can be modified to explore different physics.

The input feature space is then a nine component vector:

$$\mathbf{x} = (Z, N, A, P_{\text{asym}}, Z_{\text{eo}}, N_{\text{eo}}, A_{\text{eo}}, V_p, V_n) \ , \tag{2}$$

where the first four components can be considered macroscopic features and the last five are microscopic features. All remaining features beyond the second are functions of Z and N exclusively.

2.3 Hybrid data for training

Our training is hybrid data consisting of two distinct input sets. The first is the mass data provided by the 2020 Atomic Mass Evaluation (AME2020) [28]. The information in this set is very precise with an average reported mass uncertainty of roughly 25 keV. Modern experimental advances, such as Penning trap mass spectrometers enhanced with the Phase-Imaging Ion-Cyclotron-Resonance technique, enable such high precision measurements [29, 30].

The second mass data are provided by modern theoretical models. The information in this set is less precise, owing to the approximations made in the modeling of atomic nuclei. This set can be calculated for nuclei that have not yet been measured, providing a valuable new source of information. Nuclear models in this second set include macroscopic-microscopic approaches like the Duflo-Zuker model [31], the 2012 version of the Finite Range Droplet Model (FRDM) [32], the WS4 model [33] and microscopic approaches like UNEDF [34], and HFB32 [35].

In this work, we combine predictions from three theoretical models: FRDM2012, WS4, and HFB32. Because these models do not report individual uncertainties on their predictions, we instead estimate theoretical model uncertainty using the commonly quoted root-mean-square error or

$$\sigma_{\rm RMS} = \sqrt{\frac{1}{N}\sum_{j}^{N}(d_j - t_j)^2}\,, \tag{3}$$

where d_j is the atomic mass from the AME and t_j is the predicted atomic mass from the theoretical model. The sum runs over each j which defines a nucleus, (Z,N). For FRDM2012, WS4, and HFB32, $\sigma_{\rm RMS}$ = 0.606, 0.295 and 0.608 MeV respectively, using AME2020 masses. While the $\sigma_{\rm RMS}$ is a good measure of overall model accuracy for measured nuclei, uncertainties are certainly larger for shorter-lived systems. In this work, we do not seek to preferentiate one model over another. For this reason, we increase the assumed uncertainty of WS4 to a more reasonable 0.500 MeV when probing its predicted masses further from stability.

Training for the hybrid input data is taken at random, rather than selected based on any given criteria. The number of unique nuclei from experimental data is a free parameter in our training. The best performance is found for models provided with approximately 20% of the AME, or 400 to 500 nuclei [23]. The number of unique nuclei from theory is also a free parameter. We find that as few as 50 additional unmeasured nuclei can influence training, and therefore use this minimal number. In the case of theory data, we sample the masses of 50 randomly chosen nuclei from each of the three mass models independently.

The benefit of using hybrid data is that the neural network is not limited to solutions of model averaging which can regress to the mean when extrapolating. Instead, the combination of hybrid data with ML-based methods affords the opportunity to create new models that are capable of reproducing data, capturing trends, and predicting yet to be measured masses with sound uncertainties.

2.4 Model training and hyperparameters

The network is set up with 6 hidden layers and 10 hidden nodes per layer. The final layer turns the network into a Gaussian ad-mixture. For masses we choose a single Gaussian, although other physical quantities, such as fission yields, may require additional components [36]. The Adam optimizer is used with learning rate 0.0002 [37]. We also implement a weight-decay regularization with value 0.01. These hyperparameters were determined from a select set of runs where the values were varied.

We perform model training with two loss functions. The first loss function, $\mathcal{L}_1$, captures the match to input data. The log-likelihood loss for data is written as,

$$\mathcal{L}_1 = -\ln\left[\sum_{i=1}^{K}\frac{\pi_i(\mathbf{x})}{(2\pi)^{K/2}\sigma_i(\mathbf{x})}\exp\left\{-\frac{||\mathbf{y}-\mu_i(\mathbf{x})||^2}{2\sigma_i(\mathbf{x})^2}\right\}\right], \tag{4}$$

where $\mathbf{y}$ is the vector of training outputs and K is the total number of Gaussian mixtures. The $\pi_i(\mathbf{x})$, $\mu_i(\mathbf{x})$, and $\sigma_i(\mathbf{x})$, variables define the Gaussians, as in Eqn. 1. The minimization of this loss function furnishes the posterior distributions of predicted masses.

The hybrid mixture of experimental data and theoretical data enter into training as the variable $\mathbf{y}$. Each nucleus defined uniquely by a proton number Z and neutron number N. A Gaussian distribution is assumed to represent the probability distribution for sampling both experimental and theoretical data,

$$f(y,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) . \tag{5}$$

For the high-precision experimental data taken from the AME, the mean is set to the evaluated mass, $\mu = M^{\rm AME}_{Z,N}$, and the variance is set to the reported uncertainty of a nucleus' mass, $\sigma = \delta M^{\rm AME}_{Z,N}$. For the theoretical data from the three mass models, the mean value is taken as the prediction of the given mass model respectively. The uncertainties in these models is not reported on a per nucleus basis. Therefore an approximation to the model's $\sigma_{\rm RMS}$, which is computed with respect to the AME, is used as the variance in the probability distribution.

In this work, we do not include masses of isomeric states in the training set. However, we note that since our previous works [20, 23], the AME2020 is now utilized, rather than the earlier AME2016. This data better refines the separation of ground state and isomeric states in evaluated masses, which continues to be a known source of systematic uncertainty in the evaluation of atomic masses.

For the AME data, we take roughly 500 nuclei for training, leaving the remaining 80% of the AME as testing data. The number of stochastic realizations per nucleus is 50. For theory data, we take only 50 additional nuclei explicitly outside the AME. These nuclei are also taken at random. A given nucleus is set to have 20 stochastic realizations per theory model, for a total of 60 samples overall. From a set of testing runs, the above choices produce suitable models. We summarize the model hyperparameters in Table 1. A more complete study of all model hyperparameters is the subject of future investigations.

One essential observation of ground-state masses is that they obey the eponymous Garvey-Kelson (GK) relations [38]. This result suggests a judicious choice of mass differences of neighboring nuclei that minimizes the interactions between nucleons to first order, resulting in particular linear combinations that strategically sum to zero.

If $N \geq Z$, the GK relations state that the mass difference is

$$\begin{aligned} &M_{Z-2,N+2} - M_{Z,N} + M_{Z-1,N} \\ &- M_{Z-2,N+1} + M_{Z,N+1} - M_{Z-1,N+2} \approx 0 \, , \end{aligned} \tag{6}$$

Table 1. The neural network hyperparameters used in this work.

Parameter	Value(s)	Comment
λ_{layers}	6	Defines the number of hidden layers.
λ_{nodes}	10	Defines the number of hidden nodes per hidden layer.
λ_{gauss}	1	Defines the number of Gaussians used in the MDN.
λ_{lr}	0.0002	Defines the learning rate of the Adam optimizer.
λ_{wd}	0.01	Defines the weight-decay regularization.
λ_{exp}	506	Defines the number of AME2020 data used in training.
$\lambda_{\text{exp-pulls}}$	50	Defines the number of samples per AME mass.
λ_{theory}	50	Defines the unique nuclei probed using the three models.
$\lambda_{\text{theory-pulls}}$	20	Defines the number of samples per theory mass.
λ_{physics}	1.0	Defines the strength of the physics loss enforcement.
$\lambda_{\text{Z-low}}$	5	Defines the minimal proton number for the network.
$\lambda_{\text{N-shells}}$	8, 20, 28, 50, 82, 126, 184	Defines the major closed neutron shells.
$\lambda_{\text{Z-shells}}$	8, 20, 28, 50, 82, 114	Defines the major closed proton shells.
$\lambda_{\text{FRDM2012}}$	0.6 MeV	Defines the uncertainty used in probing FRDM2012 masses.
λ_{WS4}	0.5 MeV	Defines the uncertainty used in probing WS4 masses.
λ_{HFB32}	0.6 MeV	Defines the uncertainty used in probing HFB32 masses.

and for $N < Z$,

$$\begin{aligned} &M_{Z+2,N-2} - M_{Z,N} + M_{Z,N-1} \\ &- M_{Z+1,N-2} + M_{Z+1,N} - M_{Z+2,N-1} \approx 0 \, . \end{aligned} \tag{7}$$

Higher order GK mass relations may also be considered, as in Ref. Barea et al. [39]. However, the use of these constraints alone does not yield viable predictions far from stability due to the accumulation of uncertainty as the relationship is recursively applied beyond known data [40].

As an alternative, we perform no such iteration in our application of the GK relations. Equations 6 and 7 are used directly, and it is important to recognize that these equations depend exclusively on the masses. Thus the second (physics-based) loss function can be defined purely as a function of the ML model's mass predictions.

To enforce this physics-based observation, the second loss function can be defined as

$$\mathcal{L}_2 = -\ln\left(\left|\sum_{\{C\}} GK(\boldsymbol{\mu})\right|\right) \, , \tag{8}$$

where GK is function that defines the left-hand side of Equations 6 and 7 and we only use the model's predicted mean value of the masses, $\boldsymbol{\mu}$. The sum is performed over any choice of subset, $\{C\}$, of masses and does not have to overlap with the hybrid training data. The absolute value is necessary to ensure that the log-loss remains a real number. As with the data loss, $\mathcal{L}_1$, we also seek to simultaneously minimize the physics-loss, $\mathcal{L}_2$, which amounts to reducing the error among the difference in masses defined in the above equations.

An alternative to Equation 8 that is potentially more restrictive, is to take the absolute value inside the summation

$$\mathcal{L}_2^{\text{alt}} = -\ln\left(\sum_{\{C\}} |GK(\boldsymbol{\mu})|\right) \, . \tag{9}$$

Because Equation 9 sums many non-zero items, it is a larger loss than using Equation 8. In this case, the strength of the physics hyperparameter (discussed below) should generally be lower than in the case of using Equation 8. A strong preference for selecting one functional form for the physics-based loss over the other has not been found.

The total loss function used in training is taken as a sum of the data and physics losses

$$\mathcal{L}_{\text{total}} = \mathcal{L}_1 + \lambda_{\text{physics}}\mathcal{L}_2 \tag{10}$$

where λ_{physics} is a model hyperparameter which defines the strength of enforcement of the physics loss. We have found that values between 0.1 and 2 generally enforce the physics constraint in model predictions.

2.5 Assembling a model

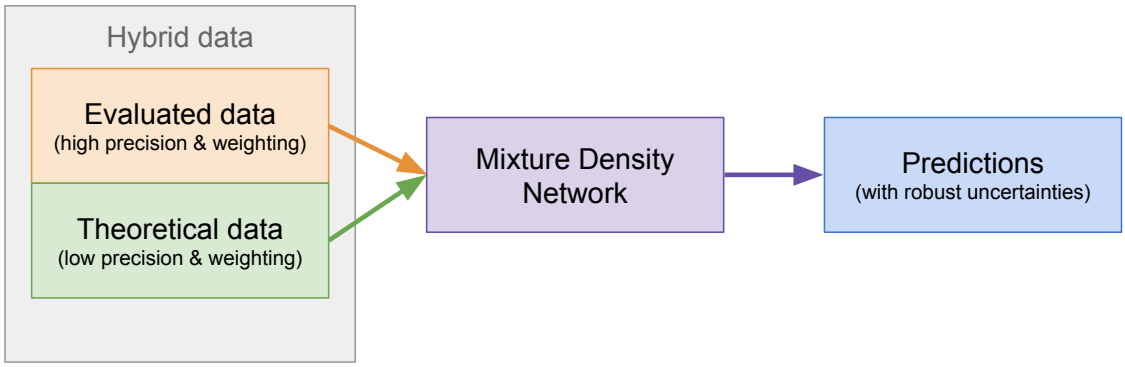

Figure 1. A schematic of our methodology. The procedure used in this work combines high precision evaluated data with a handful of less-precise theoretical data. This results in predictions with well-quantified uncertainties across the chart of nuclides.

A schematic of our methodology is shown in Figure 1 and encapsulated below. The modeling of masses begins by setting hyperparameters, summarized in Table 1, for the particular calculation. A random selection of hybrid data is made, as can be seen in Figure 2. The bulk of the masses selected for training comes from the AME (black squares) where high-precision evaluated data resides (gray squares). Only a handful of masses from theoretical models are taken for training (red squares).

After selection of hyperparameters and data, training begins which seeks to minimize the total loss, Eqn. 10. Training can take many epochs, and the data loss as well as the physics loss play important roles throughout this process, as discussed in Mumpower et al. [23]. Once the MDN has been trained on data, the results are assembled into predictions by sampling the posterior distribution several thousand times. The final output is a prediction of the mean value of the expected mass, M, and its associated uncertainty, $\sigma(M)$ for any provided nucleus defined by (Z, N).

3 RESULTS

In this section we present a MDN model trained on hybrid data. We analyze the performance with known data and discern the ability to extrapolate model predictions. We evaluate the impact of including theoretical data and the physics-based loss function.

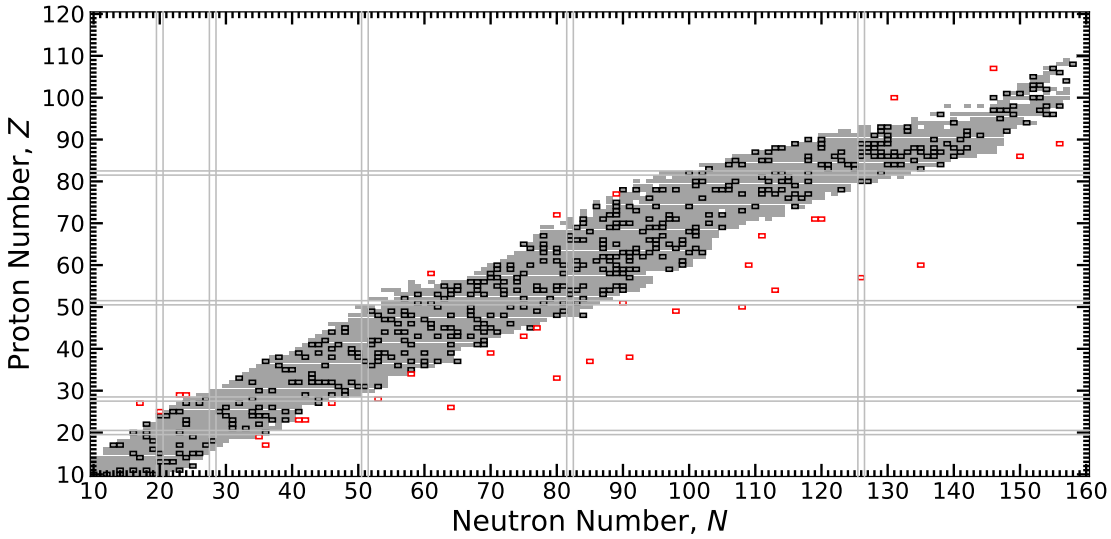

Figure 2. The chart of nuclides showing the extent of the 2020 Atomic Mass Evaluation (AME) indicated by grey squares, training nuclei part of the AME indicated by black squares, and the extra theoretical nuclei indicated by red squares. Closed proton and neutron shells are indicated by parallel lines.

3.1 Comparison with existing data

The final match to 506 training nuclei for our MDN model is $\sigma_{\rm RMS} = 0.279$ MeV. The total $\sigma_{\rm RMS}$ for the entire AME2020 is 0.395 MeV. This is an increase of roughly 0.116 MeV between training and verification data which is on the order of the accuracy of the GK relations. We limit these calculations to nuclei with $A \geq 50$. While the model can predict masses for nuclei lighter than $A = 50$, it generally performs worse in this region because there are fewer nuclei at lower mass numbers than in heavier mass regions. Therefore, there are fewer light nuclei selected in the random sample than heavy nuclei.

The absolute value of the mass residual, $\Delta M_{Z,N} = |M^{\rm AME}_{Z,N} - M^{\rm MDN}_{Z,N}|$, is one way to measure model performance. Figure 3 plots this quantity across the chart of nuclides versus the AME2020. The predictions of light nuclei tend to have an error on the order of several MeV with heavier nuclei around 0.3 MeV. The MDN model performance is on par with commonly used models in the literature.

In comparison to our previous results discussed in Mumpower et al. [23], the addition of light nuclei in training is found to relatively increase the discrepancy for heavier nuclei. The additional information modestly reduces the overall model quality as measured by $\sigma_{\rm RMS}$. On the other hand, the model is better positioned to describe the nuclear landscape more completely, insofar as the training process introduces information on the nuclear interaction that is uniquely captured in low-mass systems.

The behavior of the model with respect to select isotopic chains are shown in Figure 4. In regions of the chart where the MDN model is confident in its predictions such as in the $Z = 79$ isotopic chain, the uncertainties are very well constrained. The converse is also true, as is the case with higher uncertainties along the $Z = 43$ isotopic chain. The tin ($Z = 50$) isotopic chain highlights an intermediate case.

Inspection of this figure shows that the model has a preference for evaluated data in this region and does not follow the trends of HFB32, despite HFB32 masses being provided for training. This result reveals a unique feature of our modeling: evaluated data, due to its low uncertainties, is highly favored while theoretical points, with relatively larger uncertainties, are used as guideposts for how nuclei behave where

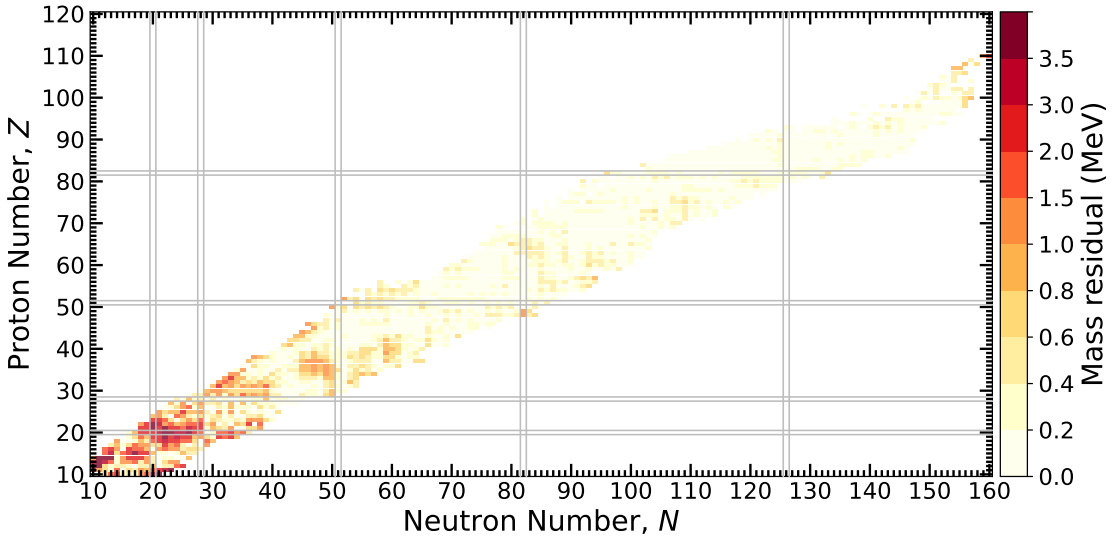

Figure 3. The absolute value of mass residuals across the chart of nuclides using an MDN model and the AME2020. Heavier nuclei are generally well described by the model while lighter nuclei exhibit larger discrepancies. See text for details.

there is no data. How much a particularly model is favored farther from stability depends on how much weighting we provide it with the choice of model uncertainty. The trends of the MDN predictions are discussed in the next section.

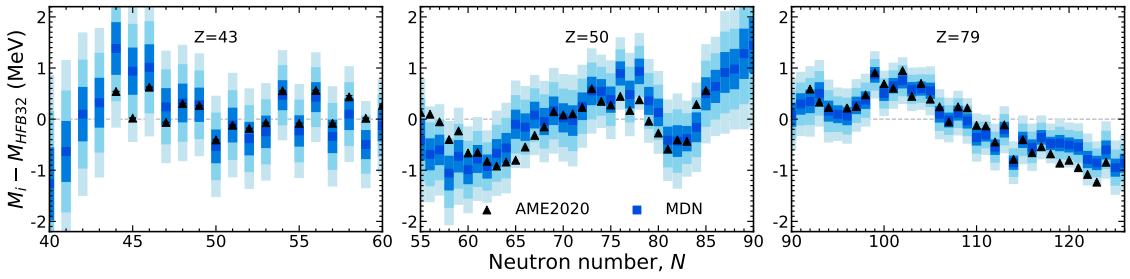

Figure 4. The prediction of masses along three isotopic chains in comparison to AME2020 data. Masses are plotted in reference to HFB32. The MDN model captures the trends exhibited in data and furnishes individual uncertainties (the one, two, and, three sigma confidence intervals are shown by blue shading).

3.2 Trends away from measured data

The extrapolation quality of atomic mass predictions is an important problem, especially for astrophysical applications where this information is needed for thousands of unmeasured species [41, 42]. The formation of the elements in particular requires robust predictions with well-quantified uncertainties [43]. The MDN supplies such information, and we now gauge the quality of the extrapolations by comparison with other theoretical models.

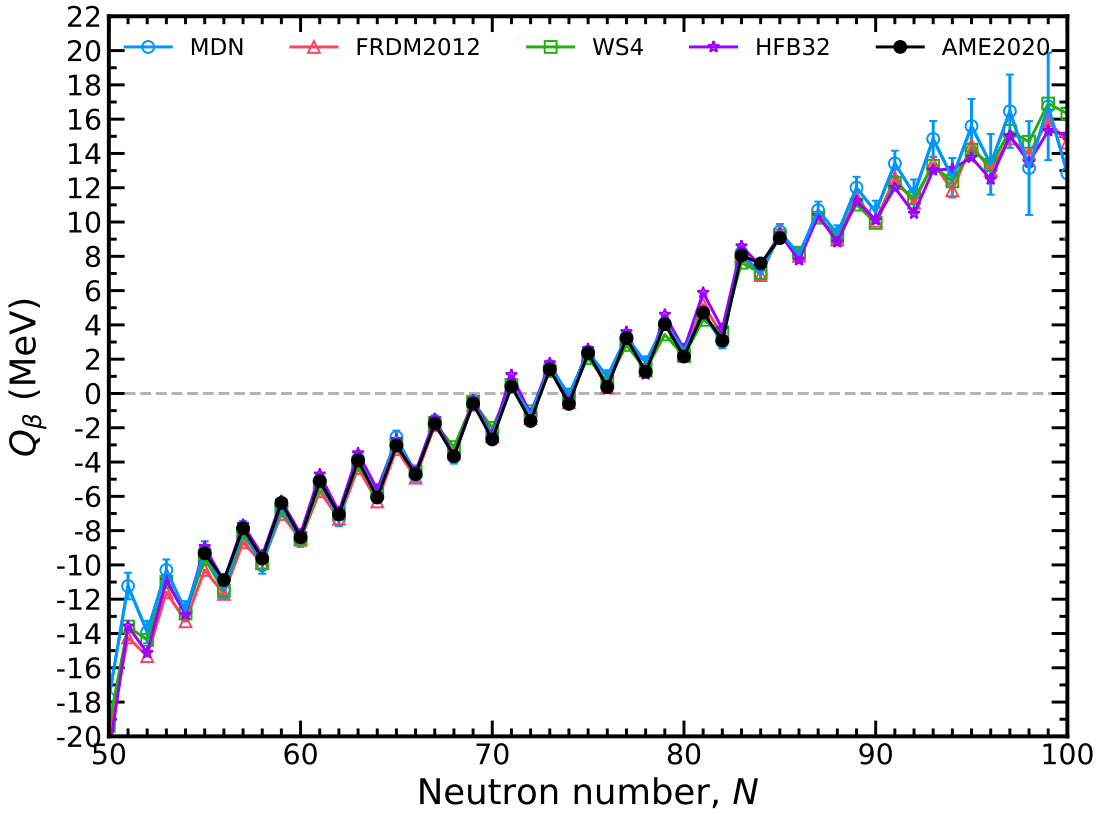

Figure 5. The total available energy for nuclear β^- decay, Q_β, along the tin isotopic chain ($Z = 50$) with a 1-σ confidence interval. The MDN model (blue) reproduces known data (black) and continues reasonably physical behavior when extrapolating. Theoretical models used in training are plotted for comparison.

The $Q_\beta(Z,N) = M_{Z,N} - M_{Z+1,N-1}$ is plotted in Figure 5 for the tin ($Z = 50$) isotopic chain. Additionally plotted is the uncertainty, δQ_β, which is calculated via propagation of error

$$\delta Q_\beta(Z,N) = \sqrt{\sigma^2(M_{Z,N}) + \sigma^2(M_{Z+1,N-1}) - 2\sigma(M_{Z,N}, M_{Z+1,N-1})}\,. \tag{11}$$

The mass uncertainties $\sigma(M_{Z,N})$, $\sigma(M_{Z+1,N-1})$ are outputs of the MDN. The correlation between the masses, $\sigma(M_{Z,N}, M_{Z+1,N-1})$, is assumed to be zero. The model has excellent performance where data is known and this result can be considered representative for other isotopic chains. Predicted uncertainties grow with decreasing and increasing neutron number outside of measured data, underscoring the Bayesian nature of our approach.

Also shown in Figure 5 are the theoretical models used in training. Comparison with these models shows that the MDN continues to retain physical behavior when extrapolating to neutron deficient or neutron rich regions. Intriguingly, the MDN model does not preferentiate one specific model when extrapolating. Instead, where there begins to be discrepancies among the theoretical models, the uncertainties begin to increase. For Q_β, the predictions along the tin isotopic chain begin to be dominated by uncertainties roughly ten units from the last available AME2020 data point.

In Figure 6 we show the extrapolation quality of one-neutron separation energies, $S_{1n}(Z,N) = M_n + M_{Z,N-1} - M_{Z,N}$. The propagation of error formula, Eqn. 11, is again employed to calculate δS_{1n} since

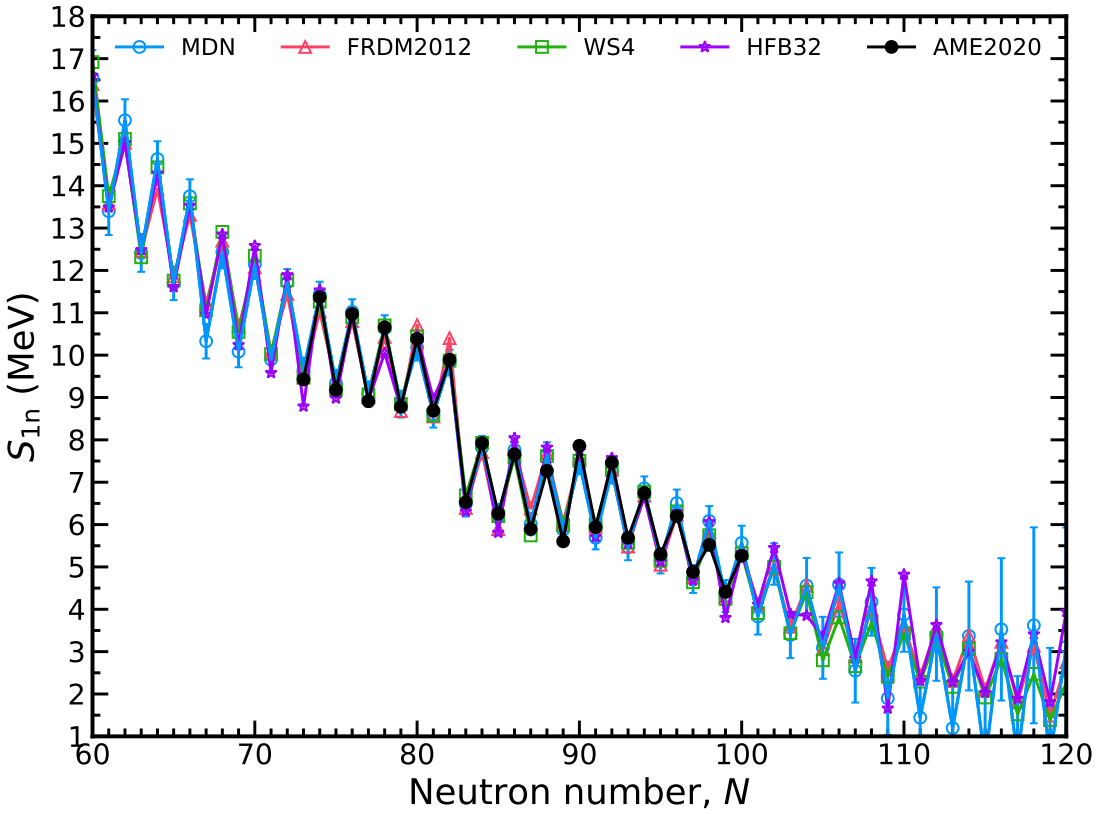

Figure 6. The one-neutron separation energy, S_{1n}, along the promethium isotopic chain ($Z = 61$) with a 1-σ confidence interval. The MDN model (blue) reproduces known data (black) and continues reasonably physical behavior when extrapolating. Theoretical models used in training are plotted for comparison.

this quantity also depends on mass differences. The uncertainty of the mass of the neutron, M_n, can be safely ignored. The qualitative behavior of S_{1n} is well described. No unphysical inversions of S_{1n} are found with the MDN model in contrast to the HFB32 model where this behavior can arise; observe around $N = 105$. Again, we find that roughly ten units from the last measured isotope, uncertainties begin to rise substantially.

A consequence of the growth of uncertainties is that the prediction of the neutron dripline, $S_{1n} = 0$ is also largely uncertain for any isotopic chain. We conclude that hybrid data does not presently place stringent constraints on this quantity, which is widely recognized as an open problem by the community [44, 45, 46].

Another quantity that can be used to gauge the quality of extrapolations is the two-neutron separation energy, $S_{2n}(Z, N) = 2M_n + M_{Z,N-2} - M_{Z,N}$. The two-neutron separation exhibits less odd-even staggering than S_{1n} because the subtraction always pairs even-N or odd-N nuclei. The behavior of the MDN model is shown in Figure 7 for the lutetium ($Z = 71$), tungsten ($Z = 74$), iridium ($Z = 71$), and mercury ($Z = 80$) isotopic chains. All experimental data falls within the 1-σ confidence intervals except for ^{206}Hg. A relatively robust shell closure is predicted at $N = 126$, though there is some weakening at the smaller proton numbers.

Finally, we consider the behavior of the physics of ground-state masses across the nuclear chart using the predictions of the MDN model. Whether or not the GK relations are preserved is yet another test of the extrapolation quality of the MDN. The calculation of the left hand side of Eqns. 6 and 7 is shown in

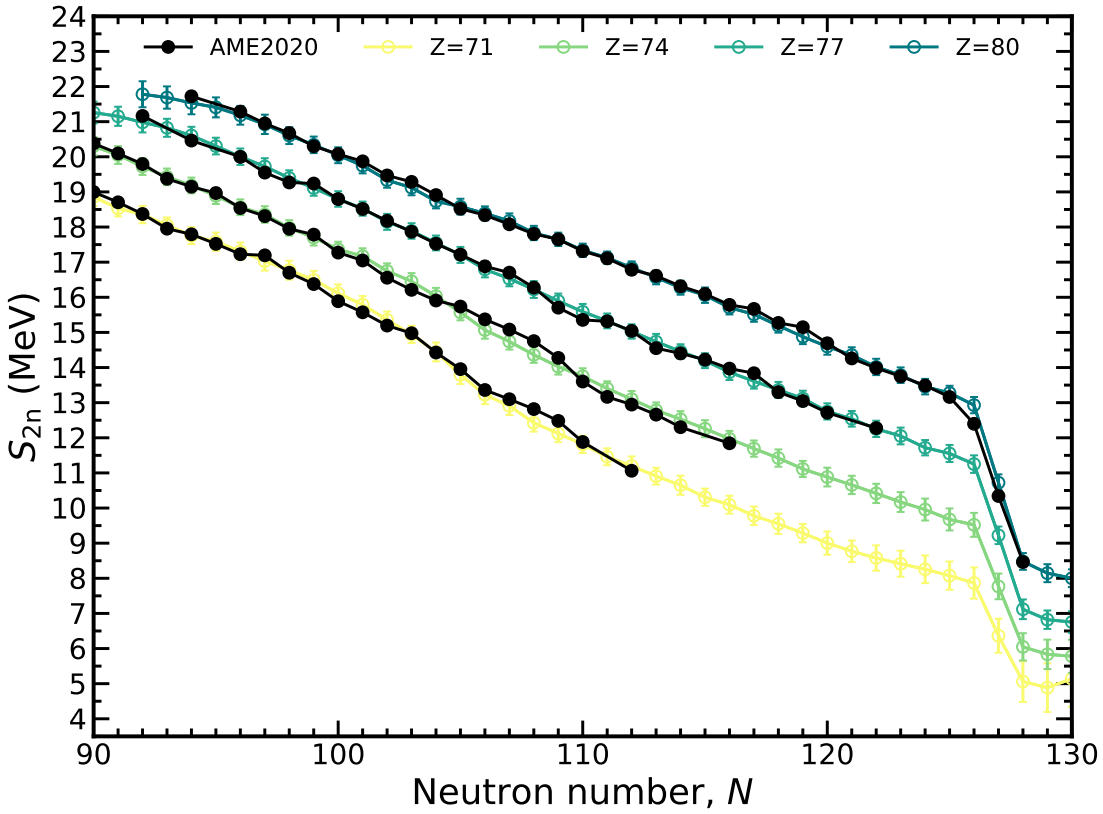

Figure 7. The two-neutron separation energy, S_{2n}, along several isotopic chains plotted with a 1-σ confidence interval. The MDN model (colors) reproduces known data (black) and continues reasonably physical behavior when extrapolating.

Figure 8 across the chart of nuclides. Yellow shading indicates the GK relations are satisfied while orange and red shading indicate potential problem areas where the relations have broken down. Given the behavior seen in the previous figures, it is clear that towards the neutron dripline, the uncertainties have grown so large that the model is unsure of the preservation of the GK relations. To emphasize this point, we calculate the nuclei for which the mass uncertainty, δM, is larger than the one-neutron separation energy, S_{1n}, and designate this as a region bound in black. We observe that this bounded region is precisely where the orange and red regions are located. Figure 8 suggests that a potential modification of the loss function that encodes the GK relations could be made to include uncertainties obtained from the MDN. This line of reasoning is the subject of future work.

3.3 Impact of theoretical data and physics constraint

We now assess the impact of the inclusion of theoretical data and the physics loss on the predictive capabilities of the MDN. Figure 9 shows four different training sets in the context of S_{1n} values. The line labeled MDN is the network shown throughout this manuscript that includes both hybrid data and physical constraint. A1 is a MDN model trained only with experimental data, lacking information about theory or the physical loss defined by the GK relations; A2 is a MDN model trained with the physics loss but without theoretical data; and A3 is a MDN model trained with theory data but without any physics loss.

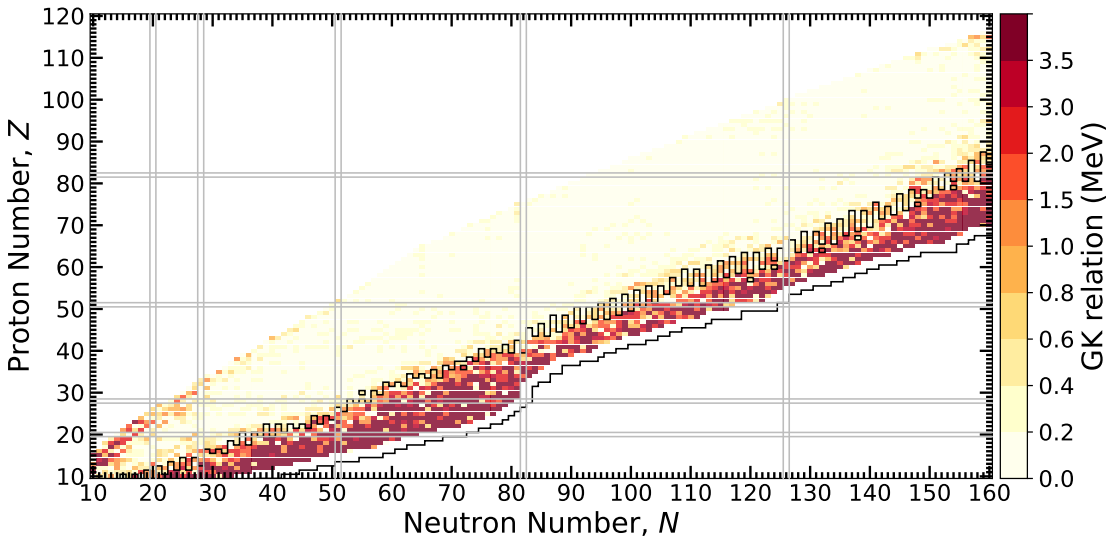

Figure 8. A test of how well the GK relations are maintained throughout the chart of nuclides. Lower values indicate predictions inline with GK, which is found nearly everywhere, except for the most extreme cases where the model is uncertain at the limits of bound nuclei. The black outlined nuclei have $\delta M > S_{1n}$, indicating where mass uncertainties are large.

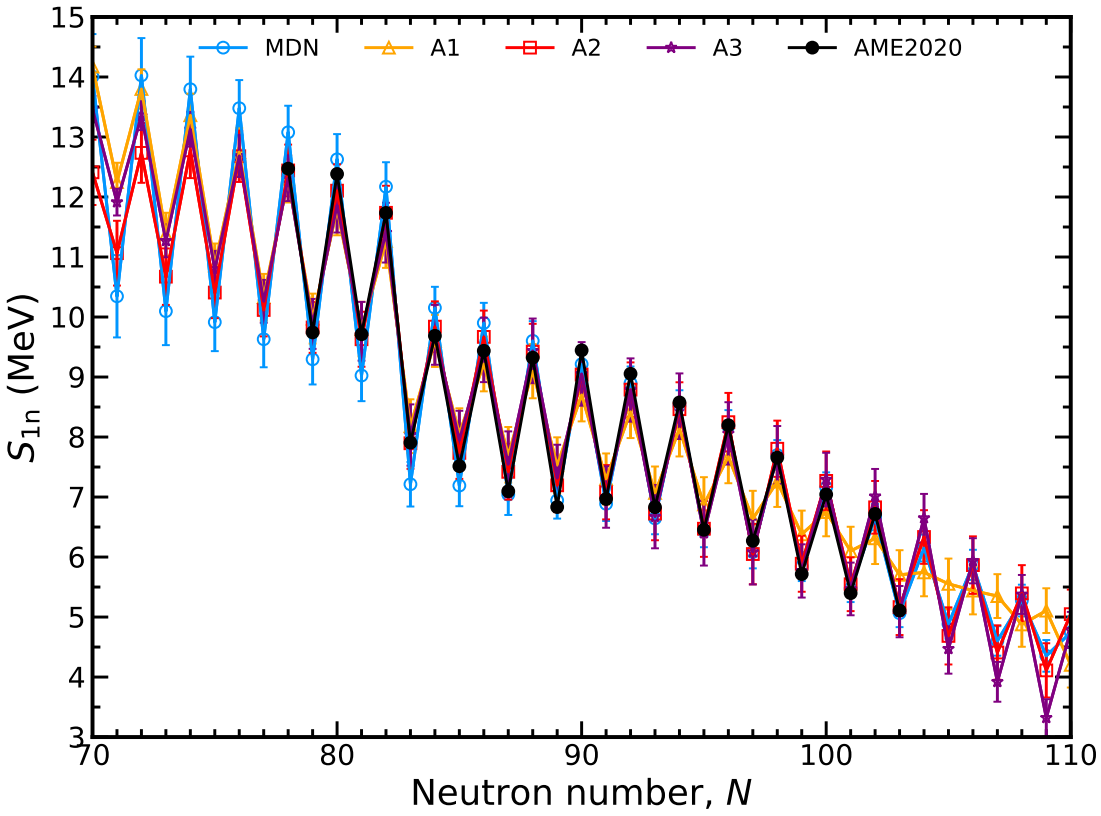

Figure 9. Comparison of MDN models with different assumptions for input data and training loss along the dysprosium ($Z = 66$) isotopic chain. See text for details.

Table 2. The parameters found using a least-square fit of Eqn. 12 to the MDN uncertainties across the chart of nuclides.

p_0	p_1	p_2	p_3	p_4	p_5
-7.035×10^{-1}	-1.037×10^{-1}	-2.997×10^{-3}	-3.729×10^{-4}	8.782×10^{-7}	1.109×10^{0}

From these four sets, it is clear that both hybrid data *and* the physics-based loss are necessary to provide quality extrapolations into unknown regions. Training with the lack of theory and GK (A1) exhibits a less desirable preference for a smooth extrapolation of $S_{1\mathrm{n}}$ values. The addition of the physics loss (A2) improves the situation by restoring the odd-even behavior observed in measured nuclei. The expected behavior in $S_{1\mathrm{n}}$ is also restored by run A3, where the hybrid data includes theory but training is not informed of the GK relations.

We note that the improvement in extrapolation behavior resulting from the hybrid data and physics-based loss is generally independent of any hyperparameters that otherwise appear in the MDN. In particular, we have preliminarily verified this result against systematic variations in relevant hyperparameters, including network size, training size, different nuclei, different input features and different blends of theoretical models. All of these variations demonstrate a similar propensity for improvement when both physics loss *and* hybrid data are used. These results lead us to reaffirm our previous observation that the addition of theory data serves as guideposts for the network solutions while the GK relations are used to ensure a refined solution.

3.4 Estimated growth of uncertainties with neutron number

The behavior of mass uncertainties as one traverses the chart of nuclides can be ascertained using the MDN predictions. Here we consider the evolution of uncertainties provided by an MDN model as a function of neutron excess, δN. We take the definition of δN to mean the number of neutrons from the last stable isotope and use the NuBase (2020) data to make this determination [47]. For each isotopic chain, δN may reference a slightly different neutron number for the particular isotope. The choice of this variable provides a relatively straight forward way to observe how mass uncertainties grow far from known data. The average and standard deviation of the MDN model's uncertainties are shown in Fig. 10.

The functional form of the average uncertainty growth as a function of neutron excess is well approximated by the following relation,

$$\sigma(\delta N) \approx p_0 + p_1 \delta N + p_2 \delta N^2 + p_3 \delta N^3 + p_4 \delta N^4 + p_5^{\delta N} \ , \tag{12}$$

where the parameters, p_i, are fit numerically (least squares) and are given in Table 2. This functional form may be readily used in simulations of nucleosynthesis to approximate uncertainties in masses with models that do not provide this information.

4 CONCLUSION

We present a Bayesian averaging technique that can be used to study the ground-state masses of atomic nuclei with corresponding uncertainties. In this work, we combine high-precision evaluated data, weighted strongly, with theoretical data for nuclei which are further from stability, more poorly understood, and therefore weighted more weakly. Training of a probabilistic neural network is used to construct the posterior distribution of ground-state masses. Along with a loss function for matching data, a second, physics-based loss function is employed in training to emphasize the relevant local behavior of masses.

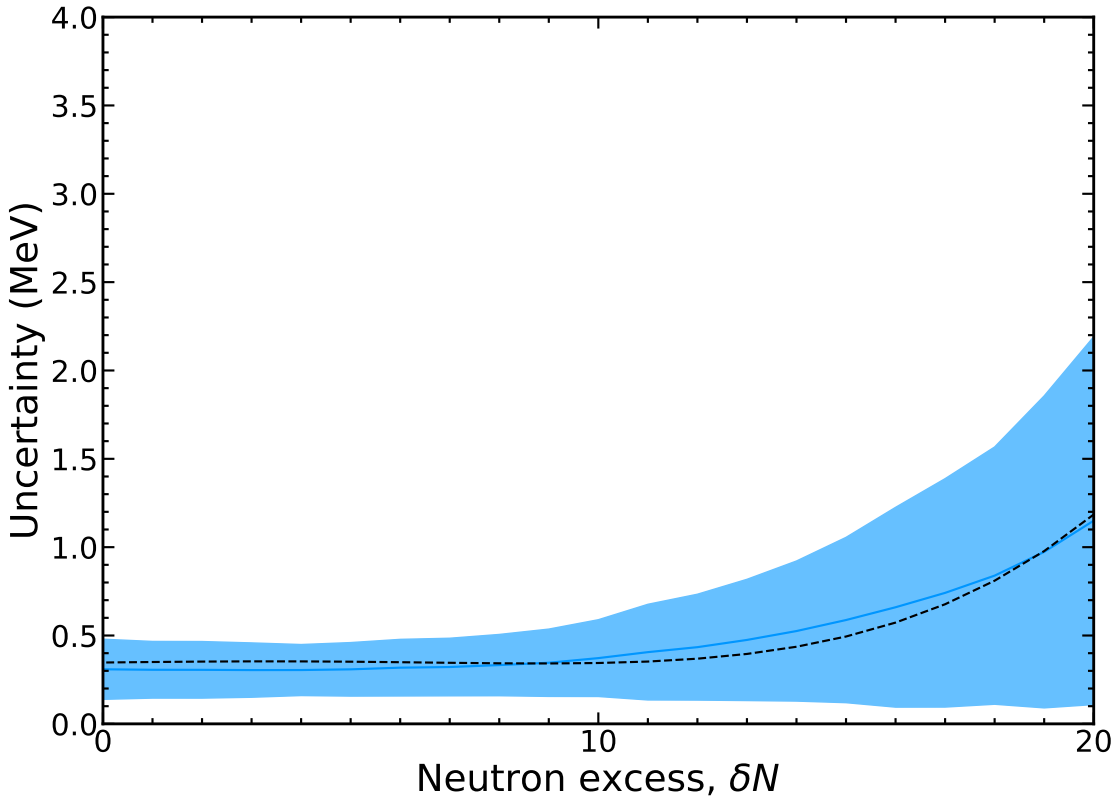

Figure 10. The average and standard deviation of the growth of mass uncertainties as a function of neutron excess. The functional form of the average can be well modeled by Eqn. 12 as plotted with a dashed black line.

Excellent performance is obtained with comparison to known data, on the order of $\sigma_{\mathrm{RMS}} \sim 0.3$ MeV, and the physical behavior of solutions is maintained when extrapolating. It is found that available data from experiment and theory are not, at this time, sufficient to resolve the relatively large uncertainties towards the limits of bound nuclei using the framework developed in this work. We emphasize the continuing need for advances in nuclear experiment and theory to reduce these uncertainties.

Our Bayesian averaging procedure enables the rapid construction of a mass model using any combination of precise and imprecise data through adjustable stochastic weights of the hybrid training inputs. For instance, if a particular theoretical model is favored over another, its sampling can be adjusted accordingly to emphasize its importance. Similarly, new high-precision data may be incorporated in the future from measurement campaigns at radioactive beam facilities. At the same time, our technique also enables freedom in the exploration of the relevant physics of ground-state masses. This can be achieved by probing a variety of physics-based features, for example, or by introducing alternative physics-based loss functions in training.

The methodology outlined here can be generalized to describe any nuclear physics property of interest, particularly when reliable extrapolations are necessary. This technique opens new avenues into Machine Learning research in the context of nuclear physics through the unification of data, theory, and associated physical constraints to empower predictions with quantified uncertainties. We look forward to extensions of this work to model nuclear decay properties, such as half-lives and branching ratios, as particularly promising opportunities in the near future.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

M. R. Mumpower led the study, overseeing all aspects of the project, including writing the code used in this study. M. Li ran and analyzed combinations of hybrid input data as well as hyperparamters. T. M. Sprouse, A. E. Lovell, and A. T. Mohan provided technical expertise for Machine Learning methods. All authors contributed to the writing of the manuscript, and read and approved the submitted version.

FUNDING

This research was supported by the Los Alamos National Laboratory (LANL) through its Center for Space and Earth Science (CSES). CSES is funded by LANL's Laboratory Directed Research and Development (LDRD) program under project number 20210528CR.

ACKNOWLEDGMENTS

M.R.M., M.L., T.M.S., A.E.L., A.T.M. were supported by the US Department of Energy through the Los Alamos National Laboratory (LANL). LANL is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). M. L. and B. S. M. were supported by NASA Emerging Worlds grant 80NSSC20K0338.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

[1] Horowitz CJ, et al. r-process nucleosynthesis: connecting rare-isotope beam facilities with the cosmos. *Journal of Physics G Nuclear Physics* **46** (2019) 083001. doi:10.1088/1361-6471/ab0849.

[2] Kajino T, Aoki W, Balantekin A, Diehl R, Famiano M, Mathews G. Current status of r-process nucleosynthesis. *Progress in Particle and Nuclear Physics* **107** (2019) 109–166. doi:10.1016/j.ppnp.2019.02.008.

[3] Thoennessen M. Plans for the facility for rare isotope beams. *Nuclear Physics A* **834** (2010) 688c – 693c. doi:https://doi.org/10.1016/j.nuclphysa.2010.01.125. The 10th International Conference on Nucleus-Nucleus Collisions (NN2009).

[4] Möller P, et al. New calculations of gross β-decay properties for astrophysical applications: Speeding-up the classical r process. Phys. Rev. C **67** (2003) 055802. doi:10.1103/PhysRevC.67.055802.

[5] Geesaman D. Reaching for the Horizon: The 2015 NSAC Long Range Plan. *APS DNP* (2015), *APS Meeting Abstracts*, vol. 2015, AA1.001.

[6] Negele JW. The mean-field theory of nuclear structure and dynamics. *Rev. Mod. Phys.* **54** (1982) 913–1015. doi:10.1103/RevModPhys.54.913.

[7] Goan E, Fookes C. Bayesian neural networks: An introduction and survey. *Case Studies in Applied Bayesian Data Science* (Springer International Publishing) (2020), 45–87. doi:10.1007/978-3-030-42553-1_3.

[8] Mohebali B, Tahmassebi A, Meyer-Baese A, Gandomi AH. Chapter 14 - probabilistic neural networks: a brief overview of theory, implementation, and application. Samui P, Tien Bui D, Chakraborty S, Deo RC, editors, *Handbook of Probabilistic Models* (Butterworth-Heinemann) (2020), 347–367. doi:https://doi.org/10.1016/B978-0-12-816514-0.00014-X.

[9] Clark JW, Gazula S. *Artificial Neural Networks that Learn Many-Body Physics* (Boston, MA: Springer US) (1991), 1–24. doi:10.1007/978-1-4615-3686-4_1.

[10] Clark JW, Gazula S, Bohr H. Nuclear phenomenology with neural nets. Taylor JG, Caianiello ER, Cotterill RMJ, Clark JW, editors, *Neural Network Dynamics* (London: Springer London) (1992), 305–322.

[11] Gazula S, Clark J, Bohr H. Learning and prediction of nuclear stability by neural networks. *Nuclear Physics A* **540** (1992) 1–26. doi:https://doi.org/10.1016/0375-9474(92)90191-L.

[12] Gernoth KA, Clark JW, Prater JS, Bohr H. Neural network models of nuclear systematics. *Phys. Lett. B* **300** (1993) 1–7. doi:10.1016/0370-2693(93)90738-4.

[13] Boehnlein A, Diefenthaler M, Sato N, Schram M, Ziegler V, Fanelli C, et al. Colloquium: Machine learning in nuclear physics. *Reviews of Modern Physics* **94** (2022). doi:10.1103/revmodphys.94.031003.

[14] Hutchinson JD, et al. Euclid: A new approach to improve nuclear data coupling optimized experiments with validation using machine learning [slides] **0** (2022). doi:10.2172/1898108.

[15] Verriere M, Schunck N, Kim I, Marević P, Quinlan K, Ngo MN, et al. Building surrogate models of nuclear density functional theory with Gaussian processes and autoencoders. *Frontiers in Physics* **10** (2022) 1028370. doi:10.3389/fphy.2022.1028370.

[16] Adams C, Carleo G, Lovato A, Rocco N. Variational monte carlo calculations of $a \leq 4$ nuclei with an artificial neural-network correlator ansatz. *Phys. Rev. Lett.* **127** (2021) 022502. doi:10.1103/PhysRevLett.127.022502.

[17] Gnech A, Adams C, Brawand N, Lovato A, Rocco N. Nuclei with up to a = 6 nucleons with artificial neural network wave functions. *Few-Body Systems* **63** (2022) – , 7. 10.1007/s00601-021-01706-0.

[18] Utama R, et al. Nuclear mass predictions for the crustal composition of neutron stars: A Bayesian neural network approach. Phys. Rev. C **93** (2016) 014311. doi:10.1103/PhysRevC.93.014311.

[19] Neufcourt L. Neutron drip line from bayesian model averaging. *PRL* **122** (2019) 062502. doi:10.1103/PhysRevLett.122.062502.

[20] Lovell AE, Mohan AT, Sprouse TM, Mumpower MR. Nuclear masses learned from a probabilistic neural network. Phys. Rev. C **106** (2022) 014305. doi:10.1103/PhysRevC.106.014305.

[21] Niu Z, Liang H. Nuclear mass predictions based on bayesian neural network approach with pairing and shell effects. *Physics Letters B* **778** (2018) 48–53. doi:https://doi.org/10.1016/j.physletb.2018.01.002.

[22] Perez RN, Schunck N. Controlling extrapolations of nuclear properties with feature selection. *arXiv preprint arXiv:2201.08835* (2022).

[23] Mumpower MR, Sprouse TM, Lovell AE, Mohan AT. Physically interpretable machine learning for nuclear masses. Phys. Rev. C **106** (2022) L021301. doi:10.1103/PhysRevC.106.L021301.

[24] Bishop CM. Mixture Density Networks. Tech. rep., Aston University (1994).

[25] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. Wallach H, Larochelle H, Beygelzimer A, d'Alche Buc F,

Fox E, Garnett R, editors, *Advances in Neural Information Processing Systems 32* (Curran Associates, Inc.) (2019), 8024–8035.

[26] Casten RF, Brenner DS, Haustein PE. Valence p-n interactions and the development of collectivity in heavy nuclei. *Phys. Rev. Lett.* **58** (1987) 658–661. doi:10.1103/PhysRevLett.58.658.

[27] Casten R, Zamfir N. The evolution of nuclear structure: The npnn scheme and related correlations. *Journal of Physics G: Nuclear and Particle Physics* **22** (1999) 1521. doi:10.1088/0954-3899/22/11/002.

[28] Wang M, Huang W, Kondev F, Audi G, Naimi S. The AME 2020 atomic mass evaluation (II). tables, graphs and references. *Chinese Physics C* **45** (2021) 030003. doi:10.1088/1674-1137/abddaf.

[29] Nesterenko DA, Eronen T, Kankainen A, Canete L, Jokinen A, Moore ID, et al. Phase-Imaging Ion-Cyclotron-Resonance technique at the JYFLTRAP double Penning trap mass spectrometer. *European Physical Journal A* **54** (2018) 154. doi:10.1140/epja/i2018-12589-y.

[30] Clark J, Savard G, Mumpower M, AKankainen. Precise mass measurements of radioactive nuclides for astrophysics. *EPJA* (2023).

[31] Duflo J, Zuker A. *Phys. Rev. C* **52** (1995) R23. doi:10.1103/PhysRevC.52.R23.

[32] Möller P, Sierk A, Ichikawa T, Sagawa H. *At. Data Nucl. Data Tables* **109-110** (2016) 1. doi:10.1016/j.adt.2015.10.002.

[33] Wang N, Liu M, Wu X, Meng J. Surface diffuseness correction in global mass formula. *Physics Letters B* **734** (2014) 215–219. doi:https://doi.org/10.1016/j.physletb.2014.05.049.

[34] Kortelainen M, McDonnell J, Nazarewicz W, Reinhard PG, Sarich J, Schunck N, et al. Nuclear energy density optimization: Large deformations. *Phys. Rev. C* **85** (2012) 024304. doi:10.1103/PhysRevC.85.024304.

[35] Goriely S, Chamel N, Pearson JM. Further explorations of skyrme-hartree-fock-bogoliubov mass formulas. xvi. inclusion of self-energy effects in pairing. *Phys. Rev. C* **93** (2016) 034337. doi:10.1103/PhysRevC.93.034337.

[36] Lovell AE, Mohan AT, Talou P. Quantifying uncertainties on fission fragment mass yields with mixture density networks. *Journal of Physics G Nuclear Physics* **47** (2020) 114001. doi:10.1088/1361-6471/ab9f58.

[37] [Dataset] Kingma DP, Ba J. Adam: A method for stochastic optimization (2017).

[38] Garvey GT, Gerace WJ, Jaffe RL, Talmi I, Kelson I. Set of nuclear-mass relations and a resultant mass table. *Rev. Mod. Phys.* **41** (1969) S1–S80. doi:10.1103/RevModPhys.41.S1.

[39] Barea J, Frank A, Hirsch JG, Isacker PV, Pittel S, Velázquez V. Garvey-kelson relations and the new nuclear mass tables. *Phys. Rev. C* **77** (2008) 041304. doi:10.1103/PhysRevC.77.041304.

[40] Jänecke J, Masson P. Mass predictions from the garvey-kelson mass relations. *Atomic Data and Nuclear Data Tables* **39** (1988) 265–271. doi:https://doi.org/10.1016/0092-640X(88)90028-9.

[41] Mumpower M, Surman R, McLaughlin G, Aprahamian A. The impact of individual nuclear properties on r-process nucleosynthesis. *Progress in Particle and Nuclear Physics* **86** (2016) 86–126. doi:https://doi.org/10.1016/j.ppnp.2015.09.001.

[42] Martin D, Arcones A, Nazarewicz W, Olsen E. Impact of nuclear mass uncertainties on the r process. *Phys. Rev. Lett.* **116** (2016) 121101. doi:10.1103/PhysRevLett.116.121101.

[43] Sprouse TM, Navarro Perez R, Surman R, Mumpower MR, McLaughlin GC, Schunck N. Propagation of statistical uncertainties of skyrme mass models to simulations of r-process nucleosynthesis. *Phys. Rev. C* **101** (2020) 055803. doi:10.1103/PhysRevC.101.055803.

[44] Erler J, Birge N, Kortelainen M, Nazarewicz W, Olsen E, Perhac AM, et al. The limits of the nuclear landscape. Nature **486** (2012) 509–512. doi:10.1038/nature11188.

[45] Xia X, Lim Y, Zhao P, Liang H, Qu X, Chen Y, et al. The limits of the nuclear landscape explored by the relativistic continuum hartree–bogoliubov theory. *Atomic Data and Nuclear Data Tables* **121-122** (2018) 1–215. doi:10.1016/j.adt.2017.09.001.

[46] Neufcourt L, Cao Y, Giuliani SA, Nazarewicz W, Olsen E, Tarasov OB. Quantified limits of the nuclear landscape. *Physical Review C* **101** (2020). doi:10.1103/physrevc.101.044307.

[47] Kondev F, Wang M, Huang W, Naimi S, Audi G. The NUBASE2020 evaluation of nuclear physics properties. *Chinese Physics C* **45** (2021) 030001. doi:10.1088/1674-1137/abddae.