Constraining Fission Yields Using Machine Learning

Amy Lovell[1,2,*], *Arvind* Mohan[1,3,**], *Patrick* Talou[4,***], and *Michael* Chertkov[3,****]

[1]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[2]Nuclear Physics Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[3]Condensed Matter Physics and Complex Systems Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[4]Materials and Physical Data Group, X Computational Physics Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Abstract. Having accurate measurements of fission observables is important for a variety of applications, ranging from energy to non-proliferation, defense to astrophysics. Because not all of these data can be measured, it is necessary to be able to accurately calculate these observables as well. In this work, we exploit Monte Carlo and machine learning techniques to reproduce mass and kinetic energy yields, for phenomenological models and in a model-free way. We begin with the spontaneous fission of ^{252}Cf, where there is abundant experimental data, to validate our approach, with the ultimate goal of creating a global yield model in order to predict quantities where data are not currently available.

1 Introduction

It is important to reliably and consistently calculate fission observables for applications such as energy, non-proliferation, defense, and astrophysics. Often, fission observables are calculated independently of one another, leading to inconsistencies within evaluations (recently shown, for example, in [1]). However, now with tools such as CGMF [2], we can form a consistent picture of fission from scission to the emission of prompt neutrons and γ rays.

There are many models available to describe single prompt fission observables (e.g. emitted neutron energy spectra or neutron average multiplicity). Some of these models calculate fission yields from shape parameterizations of nuclei using fundamental interactions, i.e. [3–7]; this is typically computationally expensive. However, most are phenomenological and must be tuned individually to reproduce data. While there are many optimization schemes available (including Monte Carlo techniques), machine learning algorithms can, in principle, learn the complex mapping between observables of the same system and different systems, enhancing the calculation of correlated observables and giving more predictive power. Currently, machine learning is still in its infancy in nuclear physics, and only a few examples are available (i.e. [8, 9]).

*e-mail: lovell@lanl.gov
**e-mail: arvindm@lanl.gov
***e-mail: talou@lanl.gov
****e-mail: chertkov@lanl.gov

In this work, we begin with a study of ^{252}Cf observables, focusing first on mass and total kinetic energy yields. Because it spontaneously fissions, a large body of data have been collected, ranging from yields, to multiplicity distributions and correlated observables. It is therefore an ideal case to study a variety of optimization methods and understand the strengths of each before tackling systems where less experimental data is available. The standard optimization method that we discuss has also been studied for ^{252}Cf in a previous work [10], giving us a baseline with which to benchmark our results before exploring the more novel machine learning techniques.

This proceedings is organized as follows. In Section 2, we discussed the physical models that are used in this work, followed by the numerical methods in Section 3. We then summarize our results in Section 4 and conclude in Section 5.

2 Theory

Fission is a rich and complex process that can be described using a range of physical models. Here, we only focus on prompt fission, in particular, the emission of prompt neutrons and γ rays from the fission fragments.

2.1 CGMF

To model the decay of the fission fragments and resulting correlations, the Monte Carlo code CGMF, has been developed [2]. The Hauser-Feschbach statistical model is used to follow the decay of the two daughter fission fragments on an event-by-event basis. At each step of the decay, probabilities for emitting neutrons or γ rays are calculated. For each decay, the energy and angle of the fission fragments and emitted particles are recorded, enabling the calculation of correlations between various observables.

In order to initialize these decays, yields in mass, charge, total kinetic energy, spin, and parity are required, $Y(A, Z, TKE, J, P)$. Currently, the sum of three Gaussians is used for the mass distribution, the TKE distribution is described by a single Gaussian for each fragment mass, the charge is determined from Wahl systematics [11], the spin distribution is proportional to $(2J+1)\exp(-0.5J(J+1)\hbar^2/(\alpha \mathcal{I}_o(A,Z)T))$ ($\mathcal{I}_o(A,Z)$ being the ground-state moment of inertia of nucleus (A,Z) and α the spin cut-off parameter), and the parity is chosen to be positive or negative with equal probability. The total kinetic energy is shared between the two daughter fragments based on an effective temperature ratio between the light and heavy fragments, which can either be constant or mass-dependent.

2.2 Brosa Yield Model

Currently, the models for $Y(A)$ and $Y(TKE)$ in CGMF are uncorrelated. In 1990, Brosa [12] published a model for $Y(A, TKE)$ based on separate fission modes or channels which correspond to different families of shapes of the fissioning system: standard (shown to be comprised of three modes, S1, S2, and S3), superlong (SL), and superasymmetric (SX).

For each mode, the yields are described by

$$Y(A, TKE) = \sum_m Y_m(A) Y_m(TKE|A), \tag{1}$$

where

$$Y_m(A) = \frac{w_m}{\sqrt{8\pi\sigma_m^2}} \left[\exp\left(-\frac{(A - \bar{A}_m)^2}{2\sigma_m^2} \right) + \exp\left(-\frac{(A - A_{cn} + \bar{A}_m)^2}{2\sigma_m^2} \right) \right], \tag{2}$$

and

$$Y_m(TKE|A) = \left(\frac{200}{TKE}\right)^2 \exp\left[2\frac{d_m^{max} - d_m^{min}}{d_m^{dec}} - \frac{T_m(A)}{d_m^{dec}} - \frac{(d_m^{max} - d_m^{min})^2}{T_m(A)d_m^{dec}}\right]. \quad (3)$$

Here,

$$T_m(A) = \frac{(Z_{cn}/A_{cn})^2(A_{cn} - A)Ae^2}{TKE} - d_m^{min}, \quad (4)$$

where Z_{cn} and A_{cn} are the charge and mass of the fissioning system. Each mode, m, has six free parameters, w_m, $\bar{A}_m$, σ_m, d_m^{max}, d_m^{min}, d_m^{dec}, which are, respectively, the weight of the mode, the mean mass of the heavy fragment, the width of the mass distribution, the most favorable semilength for fission, the minimum semilength below which fission will not occur, and the the length scale for the decay of the exponential. Since the weights describe the relative probabilities of each mode being populated, they must sum to one.

3 Methods

We have been exploring two paths to construct these yields. The first uses more traditional optimization techniques, in particular a Markov Chain Monte Carlo (MCMC). The second uses a machine learning algorithm. These are described below.

3.1 Monte Carlo Methods

First, a Markov Chain Monte Carlo [13] is used to find a best fit for the free parameters in Equations 2 and 3. This is done by exploring parameter space through randomly generated sets of parameters. Each randomly generated parameter set, i, depends on the previous one, $i-1$, through a Gaussian distribution $x_i \sim \mathcal{N}(x_{i-1}, \epsilon x_0)$, where ϵ is a scaling factor, and x_0 is a fixed parameter set. Here, x_0 is the initial parameter set as given by Brosa in [12]. This combination allows flexibility in the step size while keeping the step size appropriate for the scale of each parameter.

Each new parameter set is accepted or rejected based on the criteria

$$R < \frac{\exp(-\chi_i^2/2)}{\exp(-\chi_{i-1}^2/2)}, \quad (5)$$

where R is a random number sampled uniformly in $[0, 1]$, and χ_i^2 is the χ^2-value of the i^{th} parameter set, defined as

$$\chi^2 = \sum_j \left(\frac{Y_j^{\text{th}} - Y_j^{\text{exp}}}{\sigma_{Y_j}^{\text{exp}}}\right)^2. \quad (6)$$

Here, Y_j^{th} is a calculated yield for a given value of mass and kinetic energy as in Eq. 1, Y_j^{exp} is the experimental yield for the same mass and kinetic energy, and $\sigma_{Y_j}^{\text{exp}}$ is the experimental error on Y_j^{exp}.

In this work, we used three values of ϵ to ensure that we were not stuck in a local minimum due to the choice of x_0 and to find the minimum χ^2 value (or equivalently, the maximum likelihood). These values were 20ϵ, 5ϵ, and ϵ (with $\epsilon = 0.001$), in sequence from largest to smallest, to find the parameter set that minimized Eq. 6.

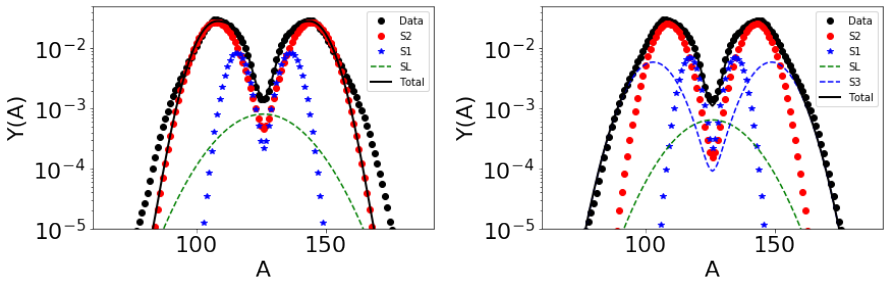

Figure 1. Mass yields when (left) three Brosa modes were fit using the MCMC and when (right) four Brosa modes were fit. Black solid lines show the total fit compared to the experimental data from [15].

3.2 Probabilistic Neural Networks

Neural networks (NN) are a machine learning algorithm that try to learn a complex relationship between input and output using a large-scale, data-driven optimization over hundreds of thousands of parameters. The base unit of the network is a neuron, several of which can be arranged in layers (the input and output each comprise a layer), and each neuron has a weight and a bias that are learned. The combination of these layers form the NN. The structure of the layers is driven by the specific application.

In a typical NN, the weights and biases are optimized based on a maximum likelihood estimation (MLE), which is a deterministic approach and cannot take into account uncertainties in the training data. Inherently, nuclear data contain uncertainties which should be included. In addition, a standard MLE will perform an average over discrepant data sets which may not reflect the confidence in each individual set.

To take this into account, we use a probabilistic machine learning approach - the Mixture Density Network (MDN) [14]. This approach predicts the complex mapping between input and output, $y = f(x)$, as a mixture of Gaussians,

$$f(x) = \alpha_1 \mathcal{N}(\mu_1, \sigma_1) + ... + \alpha_n \mathcal{N}(\mu_n, \sigma_n). \tag{7}$$

The values of α_i, μ_i, and σ_i are learned by the NN, and the user has control over the number of Gaussians that are included (as well as the underlying architecture of the NN).

4 Results and Discussion

We began with a MCMC to constrain the Brosa mode parameters for ^{252}Cf spontaneous fission yields. First, data from [15] were used to constrain, simultaneously, the Brosa parameters for the S2, S1, and SL modes. We then introduced the S3 mode, rerunning the MCMC for the four modes, and finally included the SX mode as well. The results for $Y(A)$ using three and four modes are shown in Fig. 1, along with the contributions of from each of the modes. (The contribution from the SX mode was effectively negligible.)

The $Y(A, TKE)$ distributions from these three parameterizations were then used as input for CGMF, from which we compared the resulting prompt fission observables to those that were calculated from the current models for $Y(A)$ and $Y(TKE)$. This is shown in Table 1. It is clear that the newly calculated values of $\bar{\nu}$ are significantly higher than the precise experimental value. This is due to the fitted value of $\langle TKE \rangle$ - as $\langle TKE \rangle$ and $\bar{\nu}$ have been shown to be highly anti-correlated, e.g. [16]. To remedy this, we instead fit the five Brosa

N_{modes}	$\langle TKE \rangle$ (MeV)	σ_{TKE} (MeV)	$\bar{\nu}$
O	185.77	8.97	8.758
3	184.46	10.20	3.956
4	184.16	10.74	3.946
5	184.27	10.71	3.948
5_O	185.85	8.85	3.767

Table 1. Observables calculated from CGMF from the default parameterization (O), three, four, and five modes, as well as when five modes were fit to the default yields of CGMF (5_O).

Mode	w	$\overline{A}$	σ	d^{min}	d^{max}	d^{dec}
S2	0.638	142.8	5.04	14.6	18.0	0.226
S1	0.111	135.3	3.34	9.31	17.6	0.095
SL	0.021	126.0	13.2	13.9	17.8	0.376
S3	0.229	148.8	7.02	16.7	18.7	0.299
SX	0.001	152.1	0.032	14.7	21.0	0.172

Table 2. Brosa parameters corresponding to the CGMF calculation in the last row of Table 1.

modes to $Y(A, TKE)$ sampled from the current Gaussian implementation within CGMF. This gives much more accurate results for $\langle TKE \rangle$ and $\bar{\nu}$ (last row of Table 1). The corresponding Brosa parameters are listed in Table 2.

Once $\langle TKE \rangle$ and $\bar{\nu}$ were better reproduced, a sensitivity study was performed to determine the linear response of several observables calculated with CGMF to the parameters of the Brosa modes. This response is calculated as

$$S_{ij} = \frac{x_0^i}{R_0^j} \left(\frac{\partial R^j}{\partial x^i} \right)_{x_0}, \tag{8}$$

where x_0^i are the best-fit parameters, and R_0^j are the calculated observables at x_0. The derivative in Eq. 8 is calculated at x_0 using the three-point midpoint formula. The ratio x_0^i/R_0^j is included for normalization. Responses were calculated for $\langle TKE \rangle$, σ_{TKE}, $\bar{\nu}$, $\langle \nu(\nu-1) \rangle$ (second factorial moment), $\langle \nu(\nu-1)(\nu-2) \rangle$ (third factorial moment), $\bar{\nu}_\gamma$, $\langle \nu_\gamma(\nu_\gamma - 1) \rangle$, and $\langle \nu_\gamma(\nu_\gamma-1)(\nu_\gamma-2) \rangle$; x_0 were the parameters from the five mode fit, excluding $\bar{A}_{SL}$, fixed at 126, and w_{SL}, used to enforce the normalization condition described in Sec. 2.2.

Figure 2 shows the results of these calculations. The most influential parameters are d_{S2}^{max}, d_{S3}^{max}, $\bar{A}_{S2}$, and $\bar{A}_{S3}$. The two d^{max} parameters strongly influence the distributions of TKE and number of neutrons, while $\bar{A}$ has a reduced - but noticeable - effect on the distribution of the number of γ rays. This is consistent with what was seen in [10] where $\bar{A}$ and d^{max} of the most prominent mode (S2) had the largest response. When all five modes were included, the S2 and S3 modes contributed to approximately 80% of the yields; these are also the two modes that cause the most impact on the observables calculated here.

Using a MCMC works well when we have an adequate amount of data with which to constrain the model. However, many of the systems that we are interested in describing do not have measured $Y(A, TKE)$ distributions, and it is not clear that we will find the same level of agreement between theory and experiment when this is not available. In addition, to extrapolate to systems where even less is known, trends in energy and mass have to be

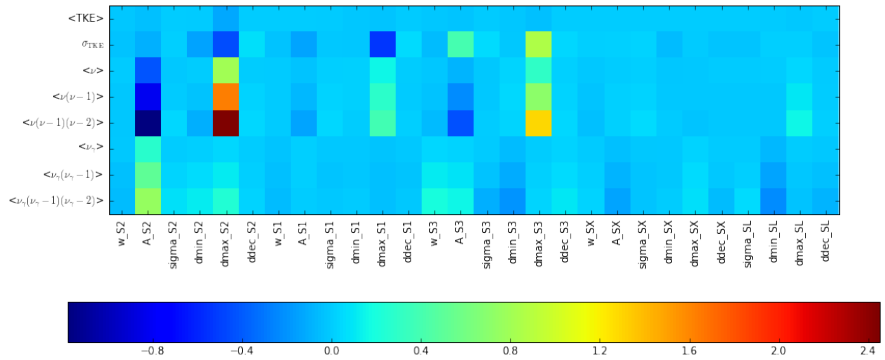

Figure 2. Responses of various observables calculated with CGMF to small changes in the Brosa mode parameters, as defined by Eq. 8.

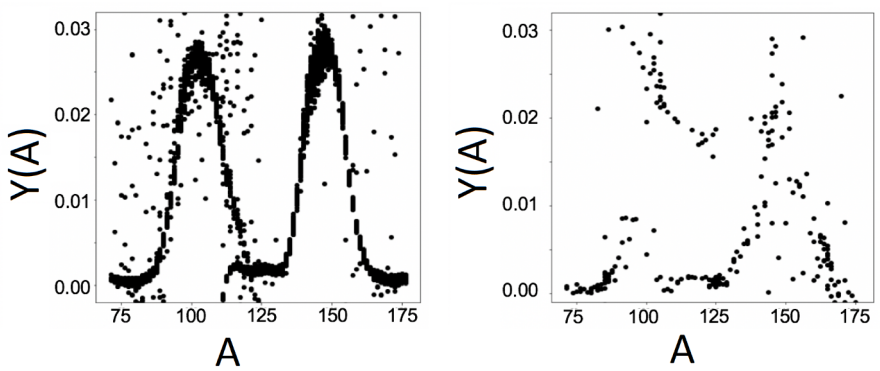

Figure 3. Results of the MDN attempt to learn $Y(A)$ for a sparse data set of ^{238}Pu(sf) (left) using transfer learning with ^{252}Cf(sf) (right) without transfer learning.

explicitly determined for these parameters. For these reasons, we have also begun to explore machine learning techniques.

In this work, we also explore the suitability of a MDN to estimate fission observables. Although the ultimate goal is to calculate all of the prompt fission observables, we began with mass yields. First, we used a three-layer MDN and train 20 Gaussians on CGMF calculations for $Y(A)$ of ^{252}Cf spontaneous fission. This method was able to learn the weights and Gaussian parameters to reproduce the inputted simulations.

We then used simulations of $Y(A)$ from ^{252}Cf to learn the mass yields for spontaneous fission of ^{238}Pu through transfer learning. To do this, we first trained the MDN on ^{252}Cf yields to learn the weights and biases. We transfered these weights and biases to another similar MDN network, which was then trained on a sparse set of ^{238}Pu mass yields containing only 3% of the generated yields. Figure 3 shows the results of this transfer learning from ^{252}Cf(sf) to ^{238}Pu(sf). Even though only a sparse data set was used for ^{238}Pu, the resulting yields show good agreement with the actual yields. However, if the sparse data set is used without the pretraining on ^{252}Cf, the ^{238}Pu yields are nonsensical, as shown in the right panel of Fig. 3.

5 Conclusions

In conclusion, we explored two ways of constraining fission fragment yields, using both parametric models as well as machine learning techniques. Using a Markov Chain Monte Carlo, we were able to reproduce fission fragment yields for the spontaneous fission of ^{252}Cf as a function of mass and total kinetic energy. These yields were implemented into the Hauser-Feschbach Monte Carlo code, CGMF, in order to calculate correlated fission observables. Although the experimental value of $\bar{\nu}$ was not reproduced when fitting to the experimental data of [15], this value was reproduced when $Y(A, TKE)$ sampled from the default version CGMF was used to constrain the Brosa mode parameters. Using not only the yields, but also neutron and γ-ray observables, it should be possible to optimize the Brosa parameters across all of these observables. We took a step towards this global optimization by performing a sensitivity study of the parameters to some of these observables calculated with CGMF.

In parallel, we have been using Mixture Density Networks, a type of probabilistic neural network, to construct fission yields in a non-parametric way. Using a network of three layers and 20 Gaussians, we were able to reproduce simulated $Y(A)$ from CGMF for ^{252}Cf(sf). Furthermore, we showed that by pretraining on ^{252}Cf mass yields, we were able to reproduce sparse ^{238}Pu(sf) mass yields through transfer learning. Although this small demonstration is just a proof-of-principle calculation, it gives an idea of the power of machine learning, not only for predicting the yields of unmeasured isotopes, but also ultimately to make predictions for fissioning systems where there is sparse or non-existent data.

The authors would like to thank Samuel Jones, Ionel Stetcu, and Harsha Nagarajan for useful discussions. This work was supported by the Office of Defense Nuclear Nonproliferation Research & Development (DNN R&D), National Nuclear Security Administration, US Department of Energy. It was performed under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract DEAC52-06NA25396. We gratefully acknowledge the support of the U.S. Department of Energy through the LANL/LDRD Program and the Center for Non Linear Studies.

References

[1] P. Jaffke, Nucl. Sci. Eng. **190** 258 (2018)
[2] P. Talou, T. Kawano, I. Stetcu, P. Jaffke, M.E. Rising, and A.E. Lovell, Comp. Phys. Comm. *in preparation*
[3] P. Möller and T. Ichikawa, Eur. Phys. J. A **51** 173 (2015)
[4] A. Sierk, Phys. Rev. C **96** 034603 (2017)
[5] M. Verriere, N. Schunck, and T. Kawano, arXiv:1811.05568v1 [nucl-th] 13 Nov 2018
[6] N. Schunck, D. Duke, H. Carr, and A. Knoll, Phys. Rev. C, **90** 054305 (2014)
[7] D. Regnier, N. Dubray, N. Schunck, and M. Verriere, Phys. Rev. C **93** 054611 (2016)
[8] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Phys. Rev. C, **98** 034318 (2018)
[9] L. Neufcourt, Y. Cao, W. Nazarewicz, E. Olsen, and F. Viens, arXiv:1901.07632v1 [nucl-th] 22 Jan 2019
[10] A. Carter, *Sensitivity Analysis and Optimization of Cf-252(sf) Observables in CGMF*, LA-UR-17-31022, Master's Thesis, University of Michigan (2018)
[11] A.C. Wahl, Technical Report LA-13928 (2002)
[12] U. Brosa, S. Grossmann, and A. Müller, Phys. Rep. **197**, 167-262 (1990)
[13] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087-1092 (1953)
[14] C.M. Bishop, Neural Computing Research Group Report NCRG/94/004 (1994)

[15] A. Göök, F. Hambsch, and M Vidali, Phys. Rev. C **90**, 064611 (2014)
[16] P. Talou, B. Becker, T. Kawano, M.B. Chadwick, and Y. Danon, Phys. Rev. C **83**, 064612 (2011)