Quantified uncertainties in fission yields from machine learning

A.E. Lovell[1,*], *A.T.* Mohan[1], *P.* Talou[1], and *M.* Chertkov[1,2]

[1]Los Alamos National Laboratory, Los Alamos, New Mexico, 87545, USA
[2]University of Arizona, Tucson, Arizona, 85721, USA

Abstract. As machine learning methods gain traction in the nuclear physics community, especially those methods that aim to propagate uncertainties to unmeasured quantities, it is important to understand how the uncertainty in the training data - coming either from theory or experiment - propagates to the uncertainty in the predicted values. Gaussian Processes and Bayesian Neural Networks are being more and more widely used, in particular to extrapolate beyond measured data. However, studies are typically not performed on the impact of the experimental errors on these extrapolated values. In this work, we focus on understanding how uncertainties propagate from input to prediction when using machine learning methods. We use a Mixture Density Network (MDN) to incorporate experimental error into the training of the network and construct uncertainties for the associated predicted quantities. Systematically, we study the effect of the size of the experimental error, both on the reproduced training data and extrapolated predictions for fission yields of actinides.

1 Introduction

Post-scission fission fragment mass yields are a necessary input for a variety of applications, including energy and stockpile stewardship, fission decay codes such as `CGMF` [1, 2], `FREYA` [3, 4], and `BeoH` [5], and the synthesis of nuclei in astrophysical environments, e.g., [6]. Even for applications that only require fission yields for major actinides, experimental data are sorely lacking for yields as a function of mass, charge, and fragment kinetic energy for all reactions and incident energies. The challenges are compounded when inputs are required for astrophysical calculations, where fission yields across the nuclear chart are required, in particular out toward the neutron dripline where measurements are currently impossible. Here, theoretical predictions or systematics derived from data are needed to make these calculations possible.

There has been much progress recently in calculating mass yields from both microscopic [7, 8] and microscopic-macroscopic models [9, 10]. Although these methods are very promising, these calculations are not yet at the precision required for many applications. As a complement, systematics have been developed by fitting to experimental data [11–13]. In addition, machine learning techniques, mainly the Bayesian Neural Network, have recently begun to be used to make predictions for mass yields [14]. Here, we continue along those

---

[*]e-mail: lovell@lanl.gov

lines, proposing to use a rather novel machine learning algorithm, the Mixture Density Network [15], to fill in the sparse set of fission yield data with quantified uncertainties.

The outline of this work is as follows: in Section 2, we briefly discuss the Mixture Density Network and the numerical details that are needed for the algorithm, results are presented and discussed in Section 3, and we conclude in Section 4.

2 Mixture Density Networks

To fill in the sparse landscape of nuclear data, we propose using a rather novel machine learning technique, the Mixture Density Network (MDN) [15]. While standard feed-forward Neural Networks (NN) aim to map input, x, directly to output, y, through a non-linear function, $y = f(x)$, the MDN describes the output as a mixture of Gaussian functions,

$$f(x) = \sum_{i=1}^{N} = \alpha_i(x)\mathcal{N}(\mu_i(x), \sigma_i(x)). \tag{1}$$

The neural network is then used to determine the weights, means, and standard deviations of each Gaussian mixture, α_i, μ_i, and σ_i. In contrast to standard deterministic NNs where predictions typically consist of a mean value and possibly a standard deviation (leaving it to the user to assume a shape for the posterior distribution), the MDN predicts the full posterior distribution for each output. Through the construction of the full posterior, we are able to fully quantify the uncertainty on each predicted value. A mean and standard deviation can be calculated, but we still have access to the exact shape of the distribution and any correlations between the predicted outputs.

2.1 Numerics for the MDN

For our convergence studies, we use a single Gaussian mixture. To determine the mean, width, and standard deviation of this Gaussian, the NN has four layers with ten nodes per layer. The training set consists of 3000 $[A, Y(A)]$ pairs where $Y(A)$ is sampled from a Gaussian distribution centered on a fit to experimental data with a predefined standard deviation. (The convergence of the number of points in the training set is shown in Sec. 3.1.) For the testing set, we randomly sample 10000 mass values in the range $[A_c/2 - 50, A_c/2 + 50]$, where A_c is the mass of the compound nucleus. Algorithms, training sets, or testing sets that deviate from these specifications are detailed in the text.

3 Results

In this proceedings, we discuss briefly the convergence studies performed for the MDN, focusing on how the uncertainties propagate from the training set to the MDN prediction. We then show results of training on a set constructed from experimental data, and finally discuss the potential of the MDN for interpolating between training sets.

3.1 Convergence studies

First, we test the convergence of the number of points needed in the training set using mass yields for ^{252}Cf spontaneous fission. To do this, for each mass in the range $[A_c/2 - 50, A_c/2 + 50]$, we randomly sample the mean of $Y(A)$ from a Gaussian with a standard deviation of 20% of the mean. While a 20% error is much larger than what is typically measured for the
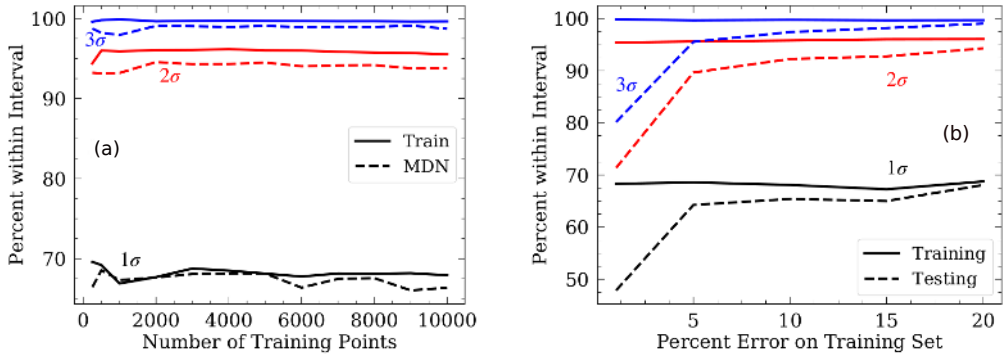
Figure 1. Percentage of the training set (solid lines) and MDN validation (dashed lines) that fall within 1σ (black), 2σ (red), and 3σ (blue) of the mean value for (a) the number of samples in the training set and (b) percentage error on mass yields used to construct the training set of 3000 samples.

mass yields, if the convergence holds for this large uncertainty, then the training set should be more than large enough for smaller values of the uncertainty as well.

For each convergence study, we compare the training set and the MDN prediction by showing the percentage of the training samples and MDN pulls that fall within 1, 2, and 3 standard deviations of the known mean value (e.g. the value sampled to construct the training set). If the posterior distribution of the MDN is the same as the distribution of the training set, ~68% of samples should fall within 1σ, 95% within 2σ, and 99% within 3σ. This should also be reflected in the distribution of the training set, as long as there is an adequate number of samples in the training set.

In Fig. 1(a), we show the convergence of the percentage of training samples and MDN predictions that fall within 1, 2, and 3 σ as a function of the number of samples in the training set. We see that even for the training set, below about 1000 samples, the statistics are not enough to consistently construct these intervals. For the MDN validation, around 3000 samples, the predictions have mostly converged, although we note that the 1σ intervals are less stable (flat) than the other two. In each case, we see that a slightly lower percentage of MDN predicted values of $Y(A)$ fall within each standard deviation than are within the training set, meaning that the MDN somewhat increases the uncertainty during the propagation.

In Fig. 1(b), we show the percentage within each standard deviation as a function of the error that was assumed on the training set. Immediately, we notice that the uncertainties are propagated poorly when a 1% error on the training set is taken, as uncertainties here are much larger than the errors included on the training set. It is also worth noting that, in this case, we needed four Gaussian mixtures for the neural network to converge. Our recent studies show that this is due to orders of magnitude spanned by the mass yields, and when the tails of the distribution are removed, only one Gaussian is needed in the mixture. For the larger values of the uncertainty on the training set (5%, 10%, 15%, and 20%), the MDN still produces uncertainties that are slightly larger than what was put on the training set but not as significantly as with the 1% errors.

3.2 Reproducing experimental data

We next show how the MDN can constrain several data sets simultaneously, using experimentally measured mass yields for ^{252}Cf(sf) [16–19]. To construct the training set, we draw
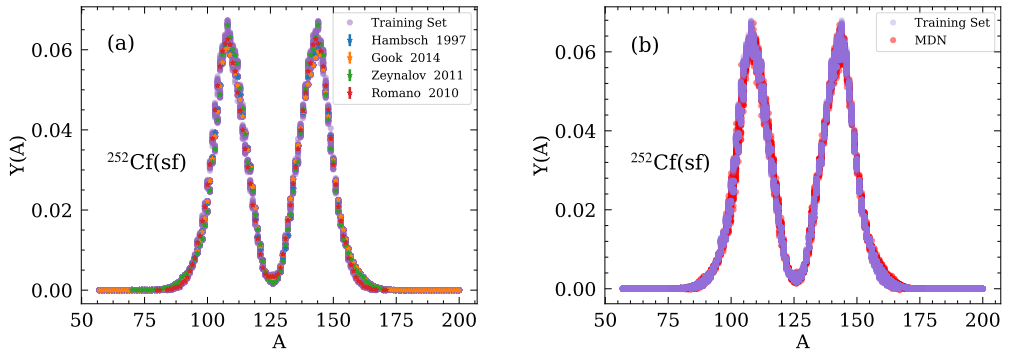
Figure 2. (a) Comparison between experimental mass yields for ^{252}Cf(sf) from [16–19] (colored stars) and the constructed training set for the MDN (purple circles). (b) MDN prediction (red) underlaid behind the training set (purple).

100 samples for each mass value from a Gaussian centered at the experimentally measured yield with a standard deviation equal to the reported uncertainty. Each mass is assumed to be independent from the others within the same experiment, and we assume zero correlations between the four experiments. This assumption is a poor one: we know that there are correlations within and between experiments. Studies to include these correlations are underway.

In Fig. 2(a), we show the training set sampled from the experimental mass yield data compared to the experimental data. Although ^{252}Cf(sf) mass yields have been measured very precisely over the last several decades, some small discrepancies between the four data sets can still be seen, in particular at the peaks of the distribution and in the symmetric region. This gives us an opportunity to investigate how the MDN handles overlapping but not identical data sets - without having to struggle with completely discrepant data.

Here, we again use a four-layer neural network with ten nodes per layer. In addition, we choose a mixture of four Gaussians to allow for the possibility that each data set will need to be described by a separate Gaussian. In panel (b) of Fig. 2, we then shown the resulting MDN prediction in red compared to the training set in purple. The MDN is qualitatively able to reproduce the training set, including the spread of the data (for instance around $A \approx 160$, the peaks of the distribution, and the symmetric mass region).

We can construct the mean value and standard deviation of $Y(A)$ from the MDN prediction. The mean values fall within the spread of the experimental data, and although the standard deviation is larger than the experimental errors, preliminary studies show that the spread can be reduced by introducing physical constraints into the training set (e.g. enforcing the symmetry between the light and heavy peaks). Additionally, a study of the posterior distribution of the MDN predictions shows that the MDN is not simply reproducing the distribution of the training set but gives some insight into the credibility of each data set.

To give more detail about the distribution of the MDN, we can also look at the mean yield values for each mass and the spread of the MDN prediction. These fall within the spread of the data, and although the resulting standard deviation of the MDN is larger than the experimental errors on the data, it reflects the spread in the data. In addition, looking at the distribution of the MDN predictions for specific values of the mass, we could see that the MDN doesn't just average over the four data sets, but it gives a distribution based on the strength of each set.
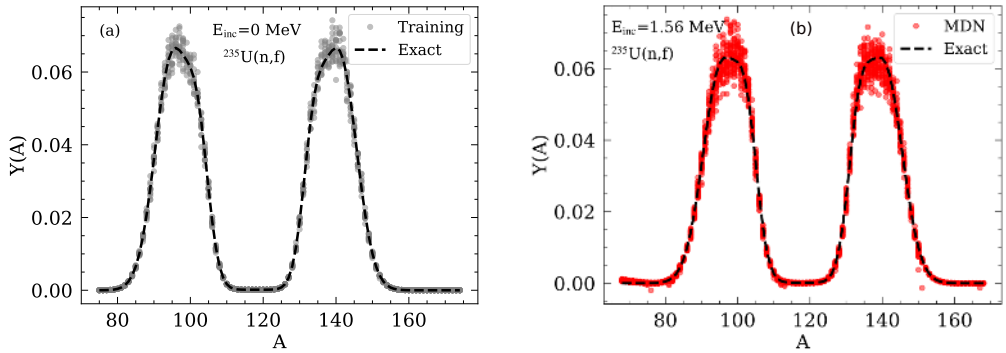
Figure 3. Mass yields for ^{235}U(n,f). (a) Training set at $E_{inc} = 0$ MeV (thermal) compared to the exact parameterization from CGMF and (b) MDN prediction for $E_{inc} = 1.56$ MeV compared to the exact CGMF parameterization.

3.3 Interpolating the energy dependence

While being able to reproduce the training set is a good test of the MDN, ultimately to be useful, we need to be able to interpolate between the training sets. To do this, we use mass yields for ^{235}U(n,f) as a function of incident neutron energy. We train on whole number energies from 0 (to denote thermal) to 10 MeV, and interpolate between energies on this grid. Here, we assume a 5% uncertainty for each mass yield and take the mean values to be the energy-dependent fit to experimental data. By default, this parameterization does not take into account multi-chance fission.

In Fig. 3(a), we show one example of the training set compared to the exact parameterization from CGMF for incident thermal neutrons. Instead of directly mapping each mass to a yield, the MDN maps pairs of inputs (mass and incident neutron energy) to a single output (mass yield). To test the ability of the MDN to interpolate between the energies in the training set, we randomly sample real numbers in the energy range of 0 to 10 MeV. In Fig. 3(b), we show one example from the testing set at an incident neutron energy of 1.56 MeV. The MDN prediction is again compared to the exact energy-dependent parameterization, in black. Although this energy is nearly exactly in the middle of two of the training points (1 MeV and 2 MeV), the MDN prediction is still centered around the exact parameterization. The spread in the MDN prediction is about twice as large as the 5% uncertainty included on the training set; preliminary studies indicate that the large uncertainties are driven by the tails of the distributions.

4 Conclusion

In summary, we are using a novel machine learning technique, the Mixture Density Network, to reproduce fission fragment mass yields with well-quantified uncertainties. We test the convergence of the MDN as a function of the number of data points needed in the training set and confirm that the MDN can reproduce mass yields for the spontaneous fission of ^{252}Cf with a range of uncertainties from 1% to 20%. We also use the MDN to simultaneously train on four different sets of experimental data for the mass yields of ^{252}Cf(sf) to provide a mean and posterior distribution for each mass value. In addition, we show how then MDN can be used to interpolate between training sets using energy-dependent yields for ^{235}U(n,f). In each

case, the MDN slightly increases the uncertainty when propagating errors from training set to prediction, performing worse for smaller errors.

Many of these features of the MDN are still under investigation, including why the uncertainties are overpredicted for small errors on the training set and why more Gaussian mixtures are needed for the stability of the network when small (~1%) errors are included in the training set. Preliminary studies suggest that this is due to the several orders of magnitude covered by the fission yields in the tails of the distributions; if this is the case, this challenge can be mitigated by removing this part of the distribution during training or by using a logarithmic scaling. After seeing the ability of the MDN to interpolate between training sets, we can next investigate the power of the MDN to extrapolate beyond the given training set. Once the numerics are stable and we have a clear understanding of how well the MDN can extrapolate, this method will be extended to make predictions across a region of the nuclear chart. Efforts along all of these paths are currently underway.

References

[1] P. Talou, T. Kawano, I. Stetcu, P. Jaffke, M.E. Rising, A.E. Lovell, Comp. Phys. Comm. (in progress)

[2] B. Becker, P. Talou, T. Kawano, Y. Danon, I. Stetcu, Phys. Rev. C **87**, 014617 (2013)

[3] J. Verbeke, J. Randrup, R. Vogt, Computer Physics Communications **191**, 178 (2015)

[4] J. Verbeke, J. Randrup, R. Vogt, Computer Physics Communications **222**, 263 (2018)

[5] S. Okumura, T. Kawano, P. Talou, P. Jaffke, S. Chiba, J. Nucl. Sci. Tech. **55**, 1009 (2018)

[6] N. Vassh, R. Vogt, R. Surman, J. Randrup, T.M. Sprouse, M.R. Mumpower, P. Jaffke, D. Shaw, E.M. Holmbeck, Y. Zhu et al., Journal of Physics G: Nuclear and Particle Physics **46**, 065202 (2019)

[7] N. Schunck, D. Duke, H. Carr, A. Knoll, Phys. Rev. C **90**, 054305 (2014)

[8] D. Regnier, N. Dubray, N. Schunck, M. Verrière, Phys. Rev. C **93**, 054611 (2016)

[9] A.J. Sierk, Phys. Rev. C **96**, 034603 (2017)

[10] P. Möller, T. Ichikawa, Eur. Phys. J. A **51**, 173 (2015)

[11] A.C. Wahl, Tech. rep., Los Alamos Nat. Lab. LA-13928 (2002)

[12] U. Brosa, S. Grossmann, A. Müller, Physics Reports **197**, 167 (1990)

[13] K.H. Schmidt, B. Jurado, C. Amouroux, Tech. Rep. NEA/DB/DOC(2014)1, Nuclear Energy Agency (2014)

[14] Z.A. Wang, J. Pei, Y. Liu, Y. Qiang, Phys. Rev. Lett. **123**, 122501 (2019)

[15] C.M. Bishop, Tech. rep., Aston University, Department of Computer Science and Applied Mathematics (1994)

[16] A. Göök, F.J. Hambsch, M. Vidali, Phys. Rev. C **90**, 064611 (2014)

[17] C. Budtz-Jørgensen, H.H. Knitter, Nuclear Physics A **490**, 307 (1988)

[18] C. Romano, Y. Danon, R. Block, J. Thompson, E. Blain, E. Bond, Phys. Rev. C **81**, 014607 (2010)

[19] S. Zeynalov, F.J. Hambsch, S. Obertstedt, J. Korean Phy. Soc. **59**, 1369 (2011)