Deep Neural Network Driven Simulation Based Inference Method for Pole Position Estimation under Model Misspecification

Daniel Sadasivan,[1,*] Isaac Cordero,[1,†] Andrew Graham,[1,‡] Cecilia Marsh,[1,§] Daniel Kupcho,[1,¶] Melana Mourad,[1,**] and Maxim Mai[2,3,††]

[1] *Ave Maria University, Ave Maria, FL 34142, USA*
[2] *Albert Einstein Center for Fundamental Physics, Institute for Theoretical Physics, University of Bern, Sidlerstrasse 5, Bern, 3012, Switzerland*
[3] *Department of Physics, The George Washington University, 725 21st St, NW, Washington, 20052, District of Columbia, USA*

The method of Simulation Based Inference is shown to lead to a more accurate resonance parameter estimation than traditional χ^2 minimization in certain cases of model misspecification in a case-study of $\pi\pi$ scattering and the $\rho(770)$-resonance. Models fit to certain data sets using χ^2 minimization can make inaccurate predictions for the pole position of the $\rho(770)$. SBI is shown to make a more robust predictions for the pole positions. This is significant, both as a proof of concept that the SBI method can be used in cases of model misspecification, and because models of $\pi\pi$ scattering are a crucial part to many physical systems of contemporary interest ($a_1(1260)$, $\omega(782)$ etc.).

arXiv:2507.18824v2 [hep-ph] 17 Jun 2026

I. INTRODUCTION

A traditional and widely-used method for model parameter determination given data is to find the parameters that minimize χ^2 statistic. Simulation Based Inference (SBI) is an alternative method, developed and used in Refs. [1–4] and applied to a number of fields including particle physics [5], astrophysics [6, 7], the Mpemba Effect [8], neuroscience [9], climate science and evolutionary biology [10]. A pedagogical introduction in this matter can be found in Ref. [11]. In a nutshell, this method involves using a model to generate synthetic data, also referred to as pseudodata, and then using this pseudodata to construct a probability distribution for the model parameters given the data. The probability distribution is often modeled using deep neural networks.

Recent research has focused on understanding how SBI performs in cases referred to as misspecification, in which either the data or the model is flawed. Some works have shown challenges with the SBI method [12], and some have used SBI as a method to test for model misspecification [13] while others develop versions of SBI that are more robust to model misspecification [14, 15]. The present paper uses a version of SBI that is robust under model misspecification to predict parameters relevant to hadronic physics, first for a toy model and then for the actual data. It should be noted that in an ideal case in which (a) data are normally distributed around a model that is considered and (b) the minimization routine can find the global minimum, the χ^2 method is guaranteed to give more probable fit parameters than the SBI method. However, in certain non-ideal cases, the SBI method might be able to outperform the method of χ^2 minimization. Thus, the aim of this paper is not to replace the standard practice of obtaining parameters through χ^2 minimization, but rather to introduce an alternative that can be useful in certain cases, specifically in the presence of model misspecification.

The specific focus of this work is universal parameters of resonances, unstable particles that can be formed in scattering or decay processes. These parameters are encoded in the pole positions of the transition amplitudes. They are typically obtained by providing a suitable model for the latter respecting principles of S-matrix theory (possibly accommodating further symmetries) and fitting the model-parameters to the data. Then the amplitude (an analytic function of the complex-valued energy) is analytically continued to the unphysical Riemann sheets where poles are allowed and are associated with resonances. For a pedagogical review, see Ref. [16].

The vector, isovector $\rho(770)$-meson, has quantum numbers that can overlap with those of a $\pi\pi$ system. Thus, its parameters can be extracted from the scattering data of two pions. Indeed, phase-shift data for this process are available [17, 18]. However, inconsistencies have been reported in the pole positions extracted from these data and the PDG values, obtained using data from a larger number of experiments, see e.g., Refs. [19–23]. Thus, an open question remains whether a method different from the χ^2-minimization can be used to mitigate possible inconsistencies in the data extracting the pole positions of the $\rho(770)$. We note that the $\rho(770)$-resonance has also been studied in lattice QCD [24–36]. While beyond the scope of the current work, the connection of the SBI methodology to input from lattice QCD is not unthinkable, see, e.g., recent work [37] also

* daniel.sadasivan@avemaria.edu
† isaac.cordero@my.avemaria.edu
‡ andrew.graham@my.avemaria.edu
§ cecilia.marsh@my.avemaria.edu
¶ daniel.kupcho@my.avemaria.edu
** melana.mourad@my.avemaria.edu
†† maxim.mai@faculty.unibe.ch

building on alternative statistical tools.

The applications of this are not limited to the $\rho(770)$-resonance. Experimental data on $K^-p \to K^-p$ and other final states in the low-energy regime are suffering under large ambiguities [38, 39] while also building a large bulk of an input required to access the universal parameters of the famous $\Lambda(1405)$ two-pole structure [40]. Similarly, ambiguities in the meson-electroproduction data can lead to incorrect extraction of baryon resonance parameters [41]. Furthermore, models for $\pi\pi$ scattering in the kinematic range covering the $\rho(770)$-resonance are also of high relevance for multi-hadron interaction. For example, understanding of the three-pion interaction requires knowledge of the two-pion subsystem. For pertinent application ($a_1(1260)$ and $\omega(782)$) within the IVU approach [42] see Refs. [43–47] and for further related works Refs. [42, 48–65].

This work is not the first to apply methods involving neural networks to study resonance phenomena. Methods similar to SBI are used in: Ref. [66] to predict the type of resonance; Ref. [67] to determine the pole position and pertinent Riemann sheets of the Pc(4312)+; Ref. [68] to predict the pole structure of PψN(4312)+; Ref. [69] to study the $\Lambda(1405)$ resonance; Ref. [68] to obtain the scattering length and effective range of the $X(3872)$ and the T_{cc}+; Ref. [70] to predict meson mass and width based only on their quantum numbers; Refs. [71, 72] to analyze hidden charm pentaquarks. Refs. [73–85] provide additional general methods for applying neural networks or other machine learning methods to the study of pole positions or analytic continuation. What distinguishes the present paper from previous work in applying deep learning to the study of resonances is an explicit focus on making predictions in the case of model misspecification.

This paper is structured as follows. Section II A describes the specific method of SBI that is applied in this work. Section II B applies this method to a contrived example, for which the correct answer is known, in order to demonstrate the effectiveness of the method in comparison to χ^2 minimization. The method is then applied to $\pi\pi$ scattering data. Section III A applies two separate models to two separate data sets, isolating effects of each model and data set. Section III B uses an SBI method to determine what model to use with the data and then to make a prediction for both data sets. Section IV discusses reasons why the SBI method might outperform the method of χ^2 minimization. The work is concluded in Sec. V.

II. SIMULATION BASED INFERENCE APPLIED TO MODEL MISSPECIFICATION

A. Methodology

This paper presents an alternative method for fitting a function, which can be written as $f(x;\vec{a})$ with x and $\vec{a}$ being an independent variable and model parameter list, respectively. The problem of fitting involves comparing this function to data to determine the most probable free parameters. This is often done by minimizing a loss function such as the χ^2 statistic, which (if correlations in data are neglected) is given by

$$\chi^2 = \sum_i^{N_{data}} \frac{(f(x_i;\vec{a}) - y_i)^2}{\sigma_i^2}\,, \tag{1}$$

where x_i are independent measurements, and y_i denote the dependent measurement with uncertainties σ_i. This well-known formula can be used to find the most probable parameters $\vec{a}$ given the measured data y_i under the assumption that the data are normally distributed about the function for some values of $\vec{a}$, and the prior uniform probability distribution of fit parameters.

In the case of model misspecification, meaning that no values of $\vec{a}$ exist for which the data are normally distributed about the function, the most accurate parameters (given the data) may not be the parameters corresponding to the minimum χ^2. SBI relies on different assumptions and therefore, in certain situations, could give more accurate probability distributions in these situations. Because there are multiple ways that SBI can be implemented, we describe the method employed in this work below, also visualizing the data pipeline in Fig. 1.

1. Sets of parameters $\vec{a}_j$ are randomly generated. The range and probability distribution for this must be carefully chosen to minimize introduced bias. If most of the generated $\vec{a}_j$ lead to a certain feature, the SBI method might predict this feature, whether or not the data contain evidence for this feature. The distribution used to generate these parameters can be thought of as the Bayesian prior distribution for the parameters.

2. For each of these parameter sets, the function f is used to generate a set of pseudodata, x^p, y^p and σ^p, where $x_i^p = x_i$ and $\sigma_i^p = \sigma_i$ for provided independent variable and corresponding uncertainties. In other words, the independent variable and the uncertainty are the same for the pseudodata as for the actual data. The dependent variable is determined as $y_i^p = R(f(x_i), \sigma_i)$, i.e., a normally distributed random variable with a mean $f(x_i)$ and standard deviation σ_i. When done this way, the method works for a fixed set of data. If new data is added, new values of x_i must be used. In principle, this method can also be extended to assess the uncertainty on the independent variable, x_i. This can be of large relevance, e.g., for negative-strangeness meson-baryon systems [38, 86–93], where experimental data not only has large error bars on the measured cross sections but also on the energy variable. In these circumstances, pseudodata for the independent variable can be generated with $x_i^p = R(x_i, \sigma_{xi})$, where σ_{xi} is the uncertainty on x_i.

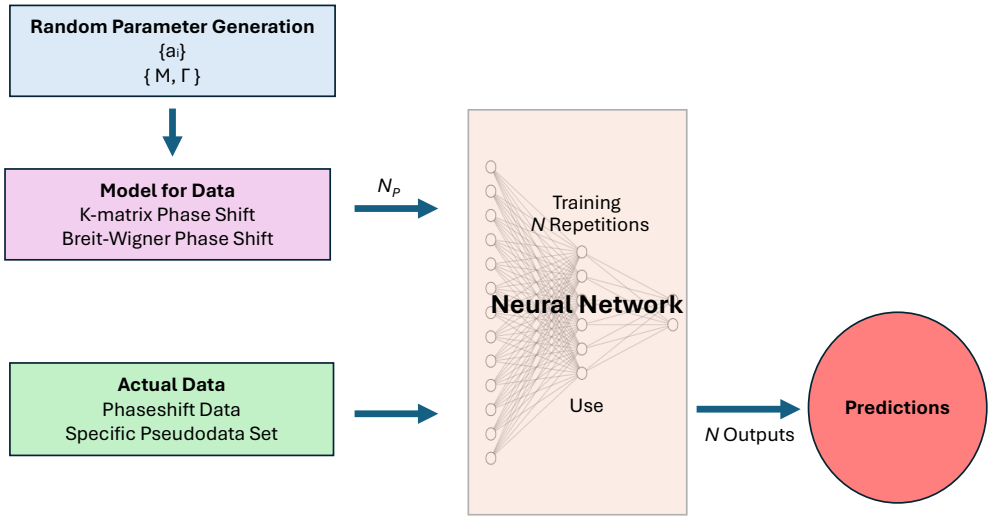


FIG. 1. A diagrammatic depiction of the data pipeline in the SBI method employed in this work. In each of the boxes on the left the middle/bottom row refer to the studied case of actual $\pi\pi$ scattering data and the toy example, respectively.

3. If necessary, only data points with the feature being studied are retained. For instance, in our example of $\rho(770)$-resonance, we only retain data points with physical poles in a realistic range. Clearly, this step can introduce bias similar to the bias described in Step 1. For instance, if a neural network that is only trained on data corresponding to reasonable ρ poles returns a ρ pole that is no more accurate than the average of the training data, this is not evidence that the neural network has learned the data set. This will be discussed in detail in Sec. III C. The number of the remaining pseudodata sets is referred to in the following as N_P.

4. The pseudodata is used to train a neural network. The values y_i^p are used as the inputs of the neural network and the randomly generated $\vec{a}$ used as outputs. A variety of loss functions could be used for this training, however, the most common is the mean squared error, which is also employed in this work.

5. Once the neural network is trained, y_i from the actual data are given as inputs and the outputs, predictions for $\vec{a}$, are recorded.

6. Steps 4. and 5. are repeated N times. This is necessary because there is some randomness in the selection of batches of $\{y^p, \vec{a}\}$ while training the neural network. The output predictions for $\vec{a}$ are stored.

7. The average values of the model parameter list obtained in step 6 give the best fit predictions $\bar{a}$.

Assuming a symmetric normal distribution of the values, the uncertainty of the average of the neural networks is obtained through standard deviation as $\bar{a}_i \pm \sigma(a_i)/\sqrt{N}$ for each model parameter a_i. This term accounts only for the uncertainty of the average of the neural network in the limit of a large number of trials, not for the bias that might occur if this average does not converge to the true value. To our knowledge, there is no general method that accurately estimates the value of this bias in all cases. This can present a real challenge for the use of SBI. Due to this concern, confidence regions for the SBI predictions should be treated with less certainty than equivalent confidence regions obtained through a χ^2 minimization. However, in this paper, we have focused on the case of model misspecification, for which, χ^2 minimization also returns biased results with no clear way to estimate the magnitude of the bias. Thus, this challenge is not unique to SBI.

B. Toy Examples

We begin with two toy examples to demonstrate the SBI. In the first, model misspecification is created by intentionally modifying several data points to prevent a reasonable fit, and in the second, model misspecification is created by generating synthetic data with non-Gaussian uncertainty but not including information about this probability distribution in the χ^2 minimization or SBI fit. In both examples we use a Breit-Wigner like parametrization of the scattering amplitude

$$T(E) = \frac{\sqrt{k}}{(E^2 - M^2) - iM\Gamma} \tag{2}$$

with a resonance mass M, width Γ and energy E. Using this parametrization, we generate unmodified synthetic

TABLE I. A comparison of the free parameters obtained through χ^2 minimization and SBI. The upper half shows the values using modified data and the lower half shows the values for the non-Gaussian synthetic data. The parameters used to generate the unmodified data are given in the second column.

Modified	Generating	χ^2	SBI
M [MeV]	857.0	889.2 ± 6.5	861.3 ± 5.0
Γ [MeV]	119.0	482.6 ± 17.2	293.5 ± 6.5
Non-Gaussian	Generating	χ^2	SBI
M [MeV]	150.0	291.6 ± 62.0	148.7 ± 1.2
Γ [MeV]	150.0	56.7 ± 52.3	207.7 ± 2.8

data, i.e., phase-shift $\delta(E) = \arctan(\mathrm{Im}\, T(E)/\mathrm{Re}\, T(E))$ that would be plausible for a specific value of M and Γ.

1. *Example 1 (outliers)*

The first toy example is intentionally chosen to provide an extreme case to visualize the SBI method when dealing with outliers. This example is performed selecting the parameters $M = 857\,\mathrm{MeV}$ and $\Gamma = 119\,\mathrm{MeV}$ and then intentionally changing the data (modified data) such that no values of M and Γ could plausibly have generated the data. Specifically, 11 unmodified points are generated for values of E ranging from 400 MeV to 1200 MeV in increments of 80 MeV, adding a normally-distributed random variable with a standard deviation of 0.05 to $\delta(E)$. After the unmodified data is calculated, the points are modified such that $\delta(E = 560\,\mathrm{MeV}) = 2.7$ and $\delta(E = 1040\,\mathrm{MeV}) = 0.5$ as depicted in Fig. 2. Note that this stark example is chosen as a first test contrasting χ^2 vs SBI methodology, only. While another case is discussed below, we note that dealing with outliers is, indeed, an important aspect in many modern hadron spectroscopy. Partially, this is due to outdated, old data sets or simple typos in data basis, see, e.g., the discussion in Refs. [41, 48, 94, 95]. Methods mitigating this issue range from manual checks to data pruning [80, 96, 97].

The probability that the modified data could have been generated from any values of M and Γ is vanishingly small. The usual χ^2 minimization can, still, be applied to the modified data. The best fit returns values of M and Γ provided in the upper half of Tab. I while the corresponding phase-shift is shown in the left side of Fig. 2.

Simulation Based Inference is also employed, with its predictions shown the right column of Tab. I. There, we randomly generate $N_P = 10^6$ points of training data with values of $M/\,\mathrm{MeV}$ uniformly distributed between 400 and 1200 and values of $\Gamma/\,\mathrm{MeV}$ uniformly distributed between 0 and 500. The pseudodata is passed to a neural network. Specifically, they include 4 hidden layers, with 40, 30, 20, and 10 nodes, respectively. Each node in each layer has a `relu` activation function. The output layer has two nodes with the leaky `relu` activation function. The mean squared error loss function is used. The neural networks are run for 1000 epochs or until they have been run for 20 epochs without improving the validation loss. In almost all cases, the latter occurs, indicating that further epochs would not improve the validation loss. The batch size chosen for the training process is 2000. The neural network is trained with the Adam optimizer. We note that the `relu` activation function, the Adam optimizer, and the mean squared error loss function are standard and have not been fine-tuned for a specific problem at hand. The leaky `relu` for the output layer was chosen because more common activation functions cannot return negative values. The number of nodes and layers was chosen somewhat arbitrarily. To assess this possible source of bias, we have performed numerical tests described in the App. A, showing that the size and architecture of the neural network is reasonable. We employ the same neural network architecture across all cases.

Taking a closer look on the results in Tab. I, we note that the SBI and χ^2 method give an M that is similarly close to the original M used to generate the unmodified data. At the same time, the value of Γ in the SBI method is closer by more than an order of magnitude than the value from the χ^2 method to the true value. The confidence region of $\delta(E)$ predicted by SBI shown in the blue shaded area of Fig. 2. While this region does not cover all of the data points, it is much closer than the $\delta(E)$ calculated with the parameters obtained through the χ^2 minimization. This toy model demonstrates that, under model misspecification the SBI can outperform χ^2 minimization providing more accurate predictions of the underlying parameters.

Note that, the neural network with a mean squared error loss function is not minimizing the same quantity as the χ^2. Even though both quantities are a sum of squares, the χ^2 statistic is the sum of the squared difference between the predicted and measured values of the phaseshift, δ, whereas the mean squared error is the squared difference between the predicted and actual values of the fit parameters, $\vec{a}$. Thus, the SBI prediction for the phase-shift should not be expected to converge with the χ^2 prediction.

In this toy example, the exact nature of the model misspecification is known. Thus, a reasonable alternative strategy could include, removing the visibly inaccurate data points as outliers or using different loss function, rather than the χ^2 function to reduce reliance on a few points. However, the strength of using SBI is that it did not have to identify why the data or model was flawed. In many cases of real application, the flaws that cause model misspecification are not identifiable. SBI might be used in some of these situations to obtain more accurate results even without identifying them.

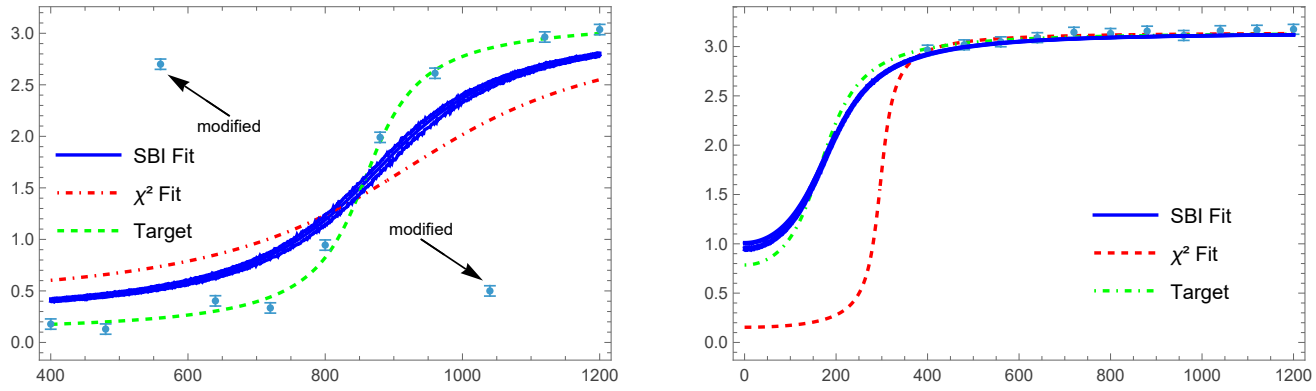


FIG. 2. Left: Phase-shift data with two points that are intentionally modified to create a case of extreme model misspecification in comparison to the function that gives the minimum χ^2 and the SBI result including confidence region. The SBI curve lies much closer to the unmodified data than the curve computed through χ^2 minimization. Right: A similar comparison of the SBI method to the neural network method. In this case the synthetic data is not modified but instead is generated with a non-Gaussian distribution. The SBI neural network is trained on only Gaussian-generated pseudo data so that it does not have an advantage.

2. *Example 2 (non-Gaussian errors)*

As a second toy example we consider a case where the data used to obtain final results from the neural network are not modified but are generated from a non-Gaussian probability distribution. Specifically, random noise is added to the phase-shift generated from an exponential distribution given by $e^{-(x-1)\lambda}\lambda$ for $x > 0$, with $\lambda = 0.05$. This distribution has mean of 0 and a standard deviation of 0.05 just as in the previous toy model. However, due to the asymmetry of this distribution, more data points lie above the curve than below however, the points that do lie below are more likely to be far below, which gives them a disproportionate influence on the χ^2. Thus, this is a situation of model misspecification even though the data have not been modified.

In this case, we choose the "true" parameters used to generate the data to be $M = 150$ and $\Gamma = 150$. The data are generated for the same values of E as the previous case. We have chosen different generating parameters for this toy model and the previous one because after trying several non-Gaussian distributions for the generating pole parameters of the previous toy-model (chosen to provide a good visualization of the SBI method) we were unable to create a case of model misspecification. This does not, however, mean that in general such distribution does not exists. We note that this situation still has important physical relevance, e. g., for subthreshold resonances, such as $\Lambda(1405)$ [40] or $T_{cc}^{+}(3875)$ mentioned above.

The pseudodata used to train the neural network is still Gaussian, because otherwise the knowledge of target distribution is needed. Of course, the SBI method would obtain more accurate results if it were trained on data generated with the same probability distribution as the one used for the data in the prediction, but we choose to train the neural network on the Gaussian data to make the comparison more even. The χ^2 statistic is constructed without knowledge of the probability distribution of the data generated. We work with an SBI method that has no additional knowledge. The parameters for this method are given in Tab. I and the curves are shown in Fig. 2. As in the case of the previous toy example, the SBI method predicts parameters that are closer to the ones used to generate the data than the ones predicted by the χ^2 method.

III. APPLICATION TO $\pi\pi$ SCATTERING AND THE $\rho(770)$

As discussed before, the properties of the $\rho(770)$-resonance can be accessed through $\pi\pi$-scattering data. As a representative example, we employ in this paper fits to the $\pi\pi$-scattering data using a K-matrix like approach reading

$$T(E) = \tilde{v}(k_{\rm cm})\frac{1}{K_n^{-1}(E) - \Sigma(E)}\tilde{v}(k_{\rm cm})\,,$$
$$K_n^{-1}(E) = M_\pi^2 \sum_{i=0}^{n-1} a_i \left(\frac{E^2}{M_\pi^2}\right)^i\,,$$
$$\Sigma(E) = \int_0^\infty \frac{dk\,k^2}{(2\pi)^3}\,\frac{1}{2E_k}\left(\frac{E^2}{4E_k^2}\right)^n \frac{\tilde{v}^2(k)}{\sigma - 4E_k^2 + i\epsilon}\,, \qquad (3)$$

used, e.g., in Ref. [44]. Here, $k_{\rm cm} = \sqrt{E^2/4 + M_\pi^2}$, $\tilde{v}(k) = \sqrt{16\pi/3}k$, $E_k = \sqrt{k^2 + M_\pi^2}$, while M_π denotes the pion mass. These models can use any number of subtractions (n) which defines the number of free parameters of the model (a_i). When the free parameters of these models (e.g., $n = 2$ or $n = 3$ subtractions) are fit to the experimental data from Refs. [17] and [18] using χ^2 minimization the predicted pole positions differ

TABLE II. The pole positions of the $\rho(770)$ from SBI and χ^2 methods for data sets from Ref. [17] (Protopopescu) and Ref. [18] (Estabrooks). Two and three subtractions (n) are used in both methods. Several reference values from the PDG [98] are provided in the lower table for comparison. All pole positions are visualized in Fig. 3.

		SBI				χ^2				
Data	n	$a_0 \cdot 10^3$	$a_1 \cdot 10^3$	$a_2 \cdot 10^3$	E^*[MeV]	$a_0 \cdot 10^3$	$a_1 \cdot 10^3$	$a_2 \cdot 10^3$	χ^2_{dof}	E^*[MeV]
Estabrooks	2	−444	+14.2		$766.3(1.8) - i\,79.0(0.7)$	−460	+15.6		12.8	$750.3(0.6) - i\,69.8(0.5)$
Protopopescu	2	−454	+14.9		$763.7(1.0) - i\,75.5(0.3)$	−467	+15.7		3.7	$755.6(0.8) - i\,70.4(0.6)$
Estabrooks	3	−188	−3.22	+0.41	$769.6(2.9) - i\,71.8(0.3)$	−223	+0.93	+0.39	3.6	$766.8(1.5) - i\,74.0(0.7)$
Protopopescu	3	−368	+8.90	+0.20	$772.3(2.1) - i\,78.7(0.6)$	−319	+6.30	+0.25	0.4	$762.8(1.2) - i\,78.5(1.1)$
All	3	−187	−3.17	+0.42	$772.4(2.3) - i\,75.4(0.4)$	−290	+4.10	+0.29	3.1	$763.7(0.9) - i\ 76.9(0.6)$

Reference/PDG values

	$E^* = M - i\,\Gamma/2$ [MeV]
e^+e^-	$775.3(0.3) - i\,73.9(0.5)$
Photoproduction	$769.0(1.0) - i\,75.9(1.3)$
Other Reactions	$769.0(0.9) - i\,75.5(0.9)$

significantly from the PDG values [98] including larger data sets. This problem is not unique to the K-matrix but was also discussed in a broader context [19–23] indicating the presence of certain inconsistencies in these data. This section covers the application of SBI to $\pi\pi$ scattering to make predictions for the pole position of the $\rho(770)$. This is done for each data set and model separately in Sec. III A and all together in Sec. III B.

A. Predictions for Individual Data Sets and Models

In this section, we apply two and three subtraction models (c.f. n in Eq. (3)) to data from Refs. [17] or [18] to which we refer to as Protopopescu data and Estabrooks data after the respective first author of each publication. This defines four different cases, allowing us to assess the effects that might be specific to one model/data set.

The best fit parameters with χ^2 minimization are determined and shown including the extracted pole positions in the right part of the Tab. II. Left part of the same table collects the results of the corresponding SBI method. For the χ^2 method, the uncertainties are calculated with bootstrap resampling, while the pseudodata are generated following the method described in Sec. II A with some details relegated to Sec. III C. The bottom of Tab. II gives the reference values taken from the PDG tables [98] where larger range of data sets and factors were used. We emphasize that our goal here is not to improve on theoretical tools (see, e.g., Refs. [20, 99–101]) but rather to study possible alternatives to the usually employed χ^2 methodology. The positions for each case for both the χ^2 method and the SBI method, as well as the PDG pole positions are visualized in Fig. 3.

In all four cases, the SBI pole position is closer to the pole position from the most precisely determined e^+e^- data.

Overall, it is reasonable to claim that in all four cases, the SBI method leads to pole predictions that are as close or closer to the reference values than the method of χ^2 minimization. Note, however, that because the different models are partly related and because there are only two models and two data sets, this result should not be considered to be as strong as if four completely independent cases were employed.

B. Combined Fit and Classifier Network

In this section, we discuss how to incorporate information for all four cases. The two data sets can be combined into a single data set, but both different models cannot be combined – one or another must be chosen. The method we use to determine the best model to use to fit the data is similar to the SBI method used in this work, see Sec. II A. A large number of pseudodata sets are generated from models with various number of subtractions (c.f., n in Eq. (3)). Only pseudodata points that have realistic pole positions are retained as discussed in Sec. III C. A neural network is then trained to predict the number of subtractions that generated the pseudodata. Finally, the experimental data is given to the neural network returning an estimate of the probability that this data was generated from a given model assuming it was generated from such model types. This classifier neural network uses the softmax activation function for the final node. The network is trained using a cross-entropy loss function. The estimated probabilities

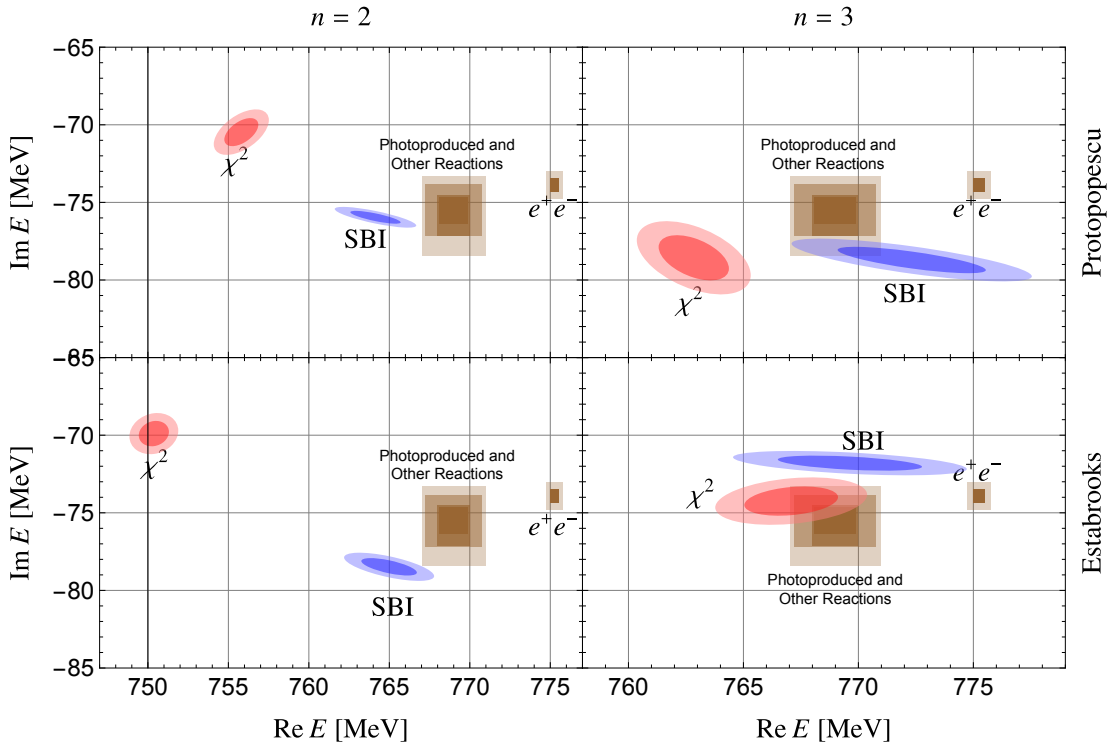


FIG. 3. The pole position confidence regions of the $\rho(770)$ for data sets from Ref. [17] (Protopopescu) and [18] (Estabrooks), using two and three subtractions. The red rectangles (1 and 2 standard deviations) give the confidence region obtained through χ^2 minimization and the blue that of the SBI. The brown rectangles give the PDG values for various types of experiments which serve as a benchmark. In these cases, the SBI method makes predictions for the ρ-pole that as close or closer to the reference values than the χ^2 method. It should, however, be noted that each case is related to two others so this result is not as strong as if all four cases were completely independent. Furthermore, the confidence regions given for the SBI and χ^2 do not account for bias in the prediction and thus should not be treated with the same credence as predictions obtained in the absence of model mispecification.

$P(n)$ with uncertainties in the parenthesis read

$$\begin{aligned} P(1) &= 0.00067(00003)\,, \\ P(2) &= 0.06482(00095)\,, \\ P(3) &= 0.86657(00122)\,, \\ P(4) &= 0.06791(01004)\,, \\ P(5) &= 0.00001(00000)\,. \end{aligned} \tag{4}$$

These results indicate that, if the actual data were generated from a K-matrix model, three subtractions is most likely to have generated it, which we use in the following. Note that these predictions are not for a specific set of a_i but for any values of a_i.

The SBI predictions using both data sets are given in bottom row of Tab. II together with the χ^2 parameters, while the predicted pole positions are also depicted in the left panel of Fig. 4. Right panel of Fig. 4 shows the model phase-shift predictions in comparison to the data. The degree of model misspecification in this case is much lower than in the Toy Example which was specifically chosen as an extreme case. It may not be "visible by eye," however, the $\chi^2_{\rm dof}$ for this case is 3.1 (given in Tab. II). For this situation, the χ^2 corresponds to a p-value of 0.17×10^{-3} indicating that model misspecification is clearly present. The red χ^2 curve lies closer to the data than the SBI method. This fact is not surprising because the χ^2 statistic is a measure of how far the curve is from the data and, therefore, the curve corresponding to the minimum χ^2 will be the one closest to the data. Nevertheless, the SBI curve makes a prediction for the ρ pole position that is in better agreement with PDG data sets (as shown in the left of Fig. 4). This situation can be compared to the toy example in which the χ^2 curve was closer (as measured by squared distance) to the modified data, but the SBI method made a prediction for the pole position that is in closer agreement with the model used to generate the majority of the data points. This indicates that possible ambiguities in the data are systematically better handled in the SBI method.

C. Training Data Generation and Tests for Bias

We generate pseudodata as described in Step 1 in Sec. II A. The parameters a_0, a_1, a_2 are generated randomly with a uniform distribution within the following ranges $[-\,\mathrm{GeV}^2/M_\pi^2, \mathrm{GeV}^2/M_\pi^2]$, $[-1, 1]$, $[-M_\pi^2/\,\mathrm{GeV}^2, M_\pi^2/\,\mathrm{GeV}^2]$, respectively. These ranges are

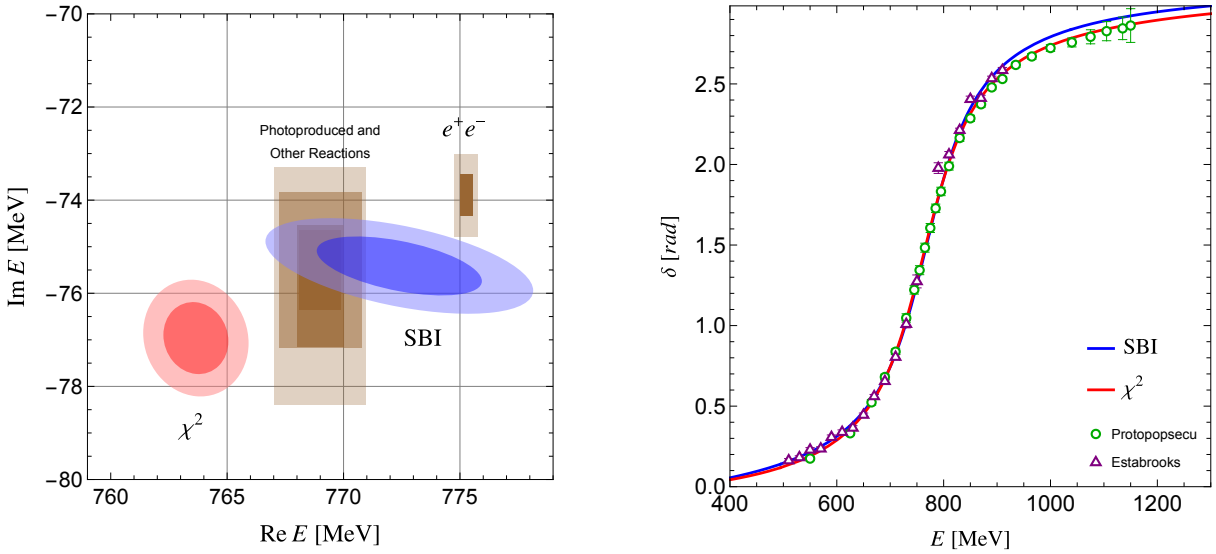


FIG. 4. Left: A comparison of the SBI method (blue) and the χ^2 method (red) for the pole positions. The SBI pole position is closer to all PDG data sets than the χ^2 minimization. Right: A comparison of the phase-shifts for the SBI method and the χ^2 method with the Protopopsecu [17] and Estabrooks [18] data. The best fit $\chi^2_{\rm dof}$ value and further details can be found in Tab. II.

well beyond the values that would generate reasonable $\rho(770)$-resonance poles.

Next, as described in Step 3 of Sec. II A, we include only parameters which lead to a pole position within the largest possible square centered on the pole-position (obtained from the χ^2 method) that is below the real axis. In other words, if the χ^2 determined pole from the actual data has a position $E^* = a - ib$, we retain pseudodata points if the corresponding χ^2 determined pole $E^*_P = x - iy$ fulfills $a-b < x < a+b$ and $2b < y < 0$. Such a method centers the range of allowed pole positions on the E^* but does not require addition choices of ranges. This pseudodata is generated until $N_P = 10^6$ points of training data are obtained. This step is necessary because most randomly generated parameters do not lead to any physical pole and only approximately 2% lead to a pole in the range in which we retain data. If this step were not performed, the neural network with the lowest loss function would be one that made the best predictions on the majority of the training data, which did not have poles. We emphasize again that the goal of this work is not to provide a universal neural network trained to assess any scenario but rather to provide an alternative method (SBI) to the usual χ^2 minimization for extracting pole positions from data which includes ambiguities. In principle, anchoring the training data to the χ^2 can effect the final predictions. However, this step ensures that, if the SBI method makes a more accurate prediction than the χ^2 method, it is not because it is trained on data that is closer to known position.

The training data is given to a neural network with exactly the same details as the neural network used in the toy example of Sec. II B. For each case, the neural networks are run $N = 100$ times. The value of 100 is chosen somewhat arbitrarily, however, we performed numerical tests described in App. B indicating that the sample size is enough for convergence. Randomness in optimization such as the batch selection leads to different final results. These $N = 100$ different results are used to calculate the uncertainty as described in Step 7. in Sec. II A.

A common concern to many applications of neural networks and AI in general is that a possible bias or prior in the pseudodata might lead to unnaturally accurate results. For example, even though the range of values in which the poles are retained is centered around the χ^2 pole, it is possible that the pole values for randomly generated parameters that fall in this range tend to be clustered closer to the reference PDG results. To address this, we calculate the pole positions for each point of the training data (10^6) and take the average as provided in Tab. III. These positions can also be compared to the pole positions obtained previously for the two methods in Tab. II. Quantitatively, this is encoded in the provided distances in Tab. III between the poles from the two methods and the average training data poles for each model and data set. For $n = 3$ and all data combinations, the average of the poles from the training data is quite far from the predictions of either method. The distance to either method is more than three times as large as the that between the methods themselves. This demonstrates that the particular choice of the training data cannot have been the reason for the better performance of the SBI method. For the $n = 2$ cases, the average of the training data pole positions is closer to SBI result. This indicates that the data used to train the neural network might slightly influence the accuracy of the SBI method. However, the fact that the

TABLE III. The average of the pole position calculated from each point in the pseudodata used to train the neural network. The right three columns give the distance in MeV between the pole positions in this table and the pole positions predicted by the SBI method and χ^2 minimization method in Tab. II.

Data	n	pseudodata E^* [MeV]	Distance χ^2 to pseudodata	Distance pseudodata to SBI	Distance SBI to χ^2
Estabrooks	2	$757.4(1.2) - i\,80.8(0.9)$	13.2(1.6)	9.1(1.3)	18.5(1.1)
Protopopescu	2	$757.3(1.2) - i\,81.0(0.8)$	10.6(1.5)	8.3(1.3)	9.5(1.0)
Estabrooks	3	$766.7(1.5) - i\,45.0(0.8)$	29.0(2.3)	26.8(2.0)	3.2(1.7)
Protopopescu	3	$764.8(1.2) - i\,44.4(1.0)$	34.2(2.3)	35.1(2.3)	9.5(1.8)
All	3	$766.1(1.8) - i\,44.1(0.9)$	32.9(2.2)	32.3(2.0)	8.8(1.3)

average training data pole is roughly equidistant from the predictions from the two methods in both cases indicates that this bias is not the primary cause of the success of the SBI method. Furthermore, we note that the distance from the SBI result to the reference values is much smaller than that to the training data, or between the χ^2 result to the reference values as well, see Fig. 3.

IV. DISCUSSION

In the previous sections we have demonstrated that SBI can be more accurate than χ^2 minimization in certain cases of model misspecification. In this section, we discuss how the neural network's predictions can be interpreted. To some extent, we are limited to speculation, because, in general neural networks are *"black boxes"* which behavior cannot be easily inferred from the details of the network [102].

We can however remark that this process can be understood as doing something related to Bayesian inference. In model fitting (whether or not misspecification is present) the desired result is a probability distribution for the fit parameters, $\vec{a}$. This distribution is conditional on the measured data $\vec{y}$. In other words, the distribution to be estimated is $P(\vec{a}|\vec{y})$. As mentioned in Sec. II A, the probability distribution used to generate the training parameters $\vec{a}_j$ can be interpreted as the prior probability distribution. Each point of training data is generated in proportion to the likelihood of values for a given $\vec{a}$, thus, if the set is large enough, it contains an implicit probability distribution for $P(\vec{y}|\vec{a})$. Thus, the neural network has access to enough information to perform an approximate Bayesian inference. Due to the complex nature of neural networks it cannot be verified if this is exactly what is being done, but the success of the neural network supports this claim.

V. CONCLUSION

We have presented a method of Simulation Based Inference and shown it to be a viable alternative method to the χ^2 minimization in case of model misspecification both for a toy example and for the case of the actual $\pi\pi$ phase-shift data used to predict the $\rho(770)$-resonance parameter. In the latter case, the SBI, indeed provides resonance parameters closer to the reference values for two different models and two different data sets indicating the robustness of the method. Additionally, a classifier neural network was employed to determine which model is most likely to have generated both sets of data simultaneously. SBI was then employed to determine the best parameters for this model. These parameters can be applied, beyond determination of the $\rho(770)$ to study of three-body states such as the $a_1(1260)$.

While it is difficult to exactly determine the reason why the SBI method outperforms the χ^2 minimization method in the described cases of model misspecification, we have speculated that the effectiveness of the SBI is due to the additional information the method has access to.

The SBI cannot outperform χ^2 minimization in cases that are close to ideal, and for this reason, we do not believe that it will replace the latter method. However, the strength of the method presented here gives strong reason to believe that application to other situations where data (including possible ambiguities) can lead to model misspecification would allow for a more robust determination of pole parameters.

ACKNOWLEDGMENTS

We are grateful to Dany Lane, Saverio Perugini, Alex Tsai, Dave Ireland, Denny Lane Sombillo and Anastasios Belias for useful discussions regarding machine learning applications. We thank James Carlini, Christian Pichay, Antonio Iijima and Maria Voce for contributing to this research. We thank Michael Döring for helping in developing useful algorithms used in parts of this work. We also acknowledge the constructive suggestions made by the reviewers of the paper. The work of MM was funded through the Heisenberg Programme by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 532635001.

[1] Simon Tavaré, David J Balding, Robert C Griffiths, and Peter Donnelly, "Inferring coalescence times from DNA sequence data," Genetics **145**, 505–518 (1997).

[2] Donald B Rubin, "Bayesianly justifiable and relevant frequency calculations for the applied statistician," The Annals of Statistics **12**, 1151–1172 (1984).

[3] Peter J Diggle and Richard J Gratton, "Monte Carlo methods of inference for implicit statistical models," Journal of the Royal Statistical Society: Series B (Methodological) **46**, 193–212 (1984).

[4] Mark A Beaumont, Wenyang Zhang, and David J Balding, "Approximate Bayesian computation in population genetics," Genetics **162**, 2025–2035 (2002).

[5] Johann Brehmer and Kyle Cranmer, "Simulation-based inference methods for particle physics," (2020), arXiv:2010.06439 [hep-ph].

[6] Siddharth Mishra-Sharma, "Inferring dark matter substructure with astrometric lensing beyond the power spectrum," Mach. Learn. Sci. Tech. **3**, 01LT03 (2022), arXiv:2110.01620 [astro-ph.CO].

[7] Ronan Legin, Yashar Hezaveh, Laurence Perreault Levasseur, and Benjamin Wandelt, "Simulation-Based Inference of Strong Gravitational Lensing Parameters," (2021), arXiv:2112.05278 [astro-ph.CO].

[8] Felipe Amorim, Joey Wisely, Nathan Buckley, Christiana DiNardo, and Daniel Sadasivan, "Predicting the mpemba effect using machine learning," Physical Review E **108** (2023), 10.1103/physreve.108.024137.

[9] Pedro J Gonçalves, Jan-Matthis Lueckmann, Michael Deistler, Marcel Nonnenmacher, Kaan Öcal, Giacomo Bassetto, Chaitanya Chintaluri, William F Podlaski, Sara A Haddad, Tim P Vogels, David S Greenberg, and Jakob H Macke, "Training deep neural density estimators to identify mechanistic models of neural dynamics," eLife **9**, e56261 (2020).

[10] Grace Avecilla, Julie Chuong, Fangfei Li, Gavin Sherlock, David Gresham, and Yoav Ram, "Simulation-based inference of evolutionary parameters from adaptation dynamics using neural networks," (2021).

[11] Astro Automata, "A tutorial on simulation-based inference," (2020), accessed: 2025-05-23.

[12] Patrick Cannon, Daniel Ward, and Sebastian M. Schmon, "Investigating the impact of model misspecification in neural simulation-based inference," (2022), arXiv:2209.01845 [stat.ML].

[13] Noemi Anau Montel, James Alvey, and Christoph Weniger, "Tests for model misspecification in simulation-based inference: From local distortions to global model checks," Phys. Rev. D **111**, 083013 (2025), arXiv:2412.15100 [astro-ph.IM].

[14] Daolang Huang, Ayush Bharti, Amauri Souza, Luigi Acerbi, and Samuel Kaski, "Learning robust statistics for simulation-based inference under model misspecification," (2023), arXiv:2305.15871 [stat.ML].

[15] Antoine Wehenkel, Juan L. Gamella, Ozan Sener, Jens Behrmann, Guillermo Sapiro, Marco Cuturi, and Jörn-Henrik Jacobsen, "Addressing misspecification in simulation-based inference through data-driven calibration," (2024), arXiv:2405.08719 [stat.ML].

[16] Maxim Mai, "Theory of resonances," (2025), arXiv:2502.02654 [hep-ph].

[17] S. D. Protopopescu, M. Alston-Garnjost, A. Barbaro-Galtieri, Stanley M. Flatte, J. H. Friedman, T. A. Lasinski, G. R. Lynch, M. S. Rabin, and F. T. Solmitz, "$\pi\pi$ Partial Wave Analysis from Reactions $\pi^+ p \to \pi^+\pi^-\Delta^{++}$ and $\pi^+ p \to K^+K^-\Delta^{++}$ at 7.1-GeV/c," Phys. Rev. **D7**, 1279 (1973).

[18] P. Estabrooks and Alan D. Martin, "$\pi\pi$ Phase Shift Analysis Below the$K\bar{K}$ Threshold," Nucl. Phys. **B79**, 301–316 (1974).

[19] B. Ananthanarayan, G. Colangelo, J. Gasser, and H. Leutwyler, "Roy equation analysis of pi pi scattering," Phys. Rept. **353**, 207–279 (2001), arXiv:hep-ph/0005297 [hep-ph].

[20] G. Colangelo, J. Gasser, and H. Leutwyler, "$\pi\pi$ scattering," Nucl. Phys. **B603**, 125–179 (2001), arXiv:hep-ph/0103088 [hep-ph].

[21] R. Garcia-Martin, R. Kaminski, J. R. Pelaez, J. Ruiz de Elvira, and F. J. Yndurain, "The Pion-pion scattering amplitude. IV: Improved analysis with once subtracted Roy-like equations up to 1100 MeV," Phys. Rev. D **83**, 074004 (2011), arXiv:1102.2183 [hep-ph].

[22] Gilberto Colangelo, Martin Hoferichter, and Peter Stoffer, "Two-pion contribution to hadronic vacuum polarization," JHEP **02**, 006 (2019), arXiv:1810.00007 [hep-ph].

[23] J. R. Pelaez, A. Rodas, and J. Ruiz De Elvira, "Global parameterization of $\pi\pi$ scattering up to 2 GeV," Eur. Phys. J. C **79**, 1008 (2019), arXiv:1907.13162 [hep-ph].

[24] S. Aoki *et al.* (CS), "ρ Meson Decay in 2+1 Flavor Lattice QCD," Phys. Rev. D **84**, 094505 (2011), arXiv:1106.5365 [hep-lat].

[25] Jozef J. Dudek, Robert G. Edwards, and Christopher E. Thomas (Hadron Spectrum), "Energy dependence of the ρ resonance in $\pi\pi$ elastic scattering from lattice QCD," Phys. Rev. D **87**, 034505 (2013), [Erratum: Phys.Rev.D 90, 099902 (2014)], arXiv:1212.0830 [hep-ph].

[26] Gunnar S. Bali, Sara Collins, Antonio Cox, Gordon Donald, Meinulf Göckeler, C. B. Lang, and Andreas Schäfer (RQCD), "ρ and K^* resonances on the lattice at nearly physical quark masses and $N_f = 2$," Phys. Rev. D **93**, 054509 (2016), arXiv:1512.08678 [hep-lat].

[27] David J. Wilson, Raul A. Briceno, Jozef J. Dudek, Robert G. Edwards, and Christopher E. Thomas, "Coupled $\pi\pi, K\bar{K}$ scattering in P-wave and the ρ resonance from lattice QCD," Phys. Rev. D **92**, 094502 (2015), arXiv:1507.02599 [hep-ph].

[28] Ziwen Fu and Lingyun Wang, "Studying the ρ resonance parameters with staggered fermions," Phys. Rev. D **94**, 034505 (2016), arXiv:1608.07478 [hep-lat].

[29] Constantia Alexandrou, Luka Leskovec, Stefan Meinel, John Negele, Srijit Paul, Marcus Petschlies, Andrew Pochinsky, Gumaro Rendon, and Sergey Syritsyn, "P-wave $\pi\pi$ scattering and the ρ resonance from lattice QCD," Phys. Rev. D **96**, 034525 (2017), arXiv:1704.05439 [hep-lat].

[30] Christian Andersen, John Bulava, Ben Hörz, and Colin Morningstar, "The $I = 1$ pion-pion scattering amplitude and timelike pion form factor from $N_\mathrm{f} = 2+1$ lattice QCD," Nucl. Phys. B **939**, 145–173 (2019), arXiv:1808.05007 [hep-lat].

[31] Markus Werner *et al.* (Extended Twisted Mass), "Hadron-Hadron Interactions from $N_f = 2 + 1 + 1$ Lattice QCD: The ρ-resonance," Eur. Phys. J. A **56**, 61 (2020), arXiv:1907.01237 [hep-lat].

[32] Dehua Guo, Andrei Alexandru, Raquel Molina, Maxim Mai, and Michael Döring, "Extraction of isoscalar $\pi\pi$ phase-shifts from lattice QCD," Phys. Rev. D **98**, 014507 (2018), arXiv:1803.02897 [hep-lat].

[33] B. Hu, R. Molina, M. Döring, M. Mai, and A. Alexandru, "Chiral extrapolations of the $\boldsymbol{\rho(770)}$ meson in $\mathbf{N_f = 2 + 1}$ lattice QCD simulations," Phys. Rev. D **96**, 034520 (2017), arXiv:1704.06248 [hep-lat].

[34] Bin Hu, Raquel Molina, Michael Doring, Andrei Alexandru, and Dehua Guo, "Two-flavor simulations of the $\rho(770)$ and the role of the $K\overline{K}$ channel," PoS **LATTICE2016**, 096 (2016).

[35] Dehua Guo, Andrei Alexandru, Raquel Molina, and Michael Döring, "Rho resonance parameters from lattice QCD," Phys. Rev. D **94**, 034501 (2016), arXiv:1605.03993 [hep-lat].

[36] Matthias Fischer, Bartosz Kostrzewa, Maxim Mai, Marcus Petschlies, Ferenc Pittler, Martin Ueding, Carsten Urbach, and Markus Werner (Extended Twisted Mass, ETM), "The ρ-resonance from Nf=2 lattice QCD including the physical pion mass," Phys. Lett. B **819**, 136449 (2021), arXiv:2006.13805 [hep-lat].

[37] Miguel Salg, Fernando Romero-López, and William I. Jay, "Bayesian Analysis and Analytic Continuation of Scattering Amplitudes from Lattice QCD," (2025), arXiv:2506.16161 [hep-lat].

[38] D. Sadasivan, M. Mai, and M. Döring, "S- and p-wave structure of $S = -1$ meson-baryon scattering in the resonance region," Phys. Lett. B **789**, 329–335 (2019), arXiv:1805.04534 [nucl-th].

[39] Ferenc Pittler, Maxim Mai, Ulf-G. Meißner, Ryan F. Ferguson, Peter Hurck, David G. Ireland, and Bryan McKinnon, "Universal parameters of the $\Lambda(1380)$, the $\Lambda(1405)$ and their isospin partners from a combined analysis of Lattice QCD and experimental results," (2025), arXiv:2507.14283 [hep-ph].

[40] Maxim Mai, "Review of the $\Lambda(1405)$ A curious case of a strangeness resonance," Eur. Phys. J. ST **230**, 1593–1607 (2021), arXiv:2010.00056 [nucl-th].

[41] Maxim Mai, Michael Döring, Carlos Granados, Helmut Haberzettl, Ulf-G. Meißner, Deborah Rönchen, Igor Strakovsky, and Ron Workman (Jülich-Bonn-Washington), "Jülich-Bonn-Washington model for pion electroproduction multipoles," Phys. Rev. C **103**, 065204 (2021), arXiv:2104.07312 [nucl-th].

[42] M. Mai, B. Hu, M. Doring, A. Pilloni, and A. Szczepaniak, "Three-body Unitarity with Isobars Revisited," Eur. Phys. J. A **53**, 177 (2017), arXiv:1706.06118 [nucl-th].

[43] D. Sadasivan, M. Mai, H. Akdag, and M. Döring, "Dalitz plots and lineshape of $a_1(1260)$ from a relativistic three-body unitary approach," Phys. Rev. D **101**, 094018 (2020), [Erratum: Phys.Rev.D 103, 019901 (2021)], arXiv:2002.12431 [nucl-th].

[44] Daniel Sadasivan, Andrei Alexandru, Hakan Akdag, Felipe Amorim, Ruairí Brett, Chris Culver, Michael Döring, Frank X. Lee, and Maxim Mai, "Pole position of the a1(1260) resonance in a three-body unitary framework," Phys. Rev. D **105**, 054020 (2022), arXiv:2112.03355 [hep-ph].

[45] Maxim Mai, Andrei Alexandru, Ruairí Brett, Chris Culver, Michael Döring, Frank X. Lee, and Daniel Sadasivan (GWQCD), "Three-Body Dynamics of the a1(1260) Resonance from Lattice QCD," Phys. Rev. Lett. **127**, 222001 (2021), arXiv:2107.03973 [hep-lat].

[46] Yuchuan Feng, Fernando Gil, Michael Döring, Raquel Molina, Maxim Mai, Vanamali Shastry, and Adam Szczepaniak, "A unitary coupled-channel three-body amplitude with pions and kaons," Phys. Rev. D **110**, 094002 (2024), arXiv:2407.08721 [nucl-th].

[47] Haobo Yan, Maxim Mai, Marco Garofalo, Ulf-G. Meißner, Chuan Liu, Liuming Liu, and Carsten Urbach, "ω Meson from Lattice QCD," Phys. Rev. Lett. **133**, 211906 (2024), arXiv:2407.16659 [hep-lat].

[48] Michael Döring, Johann Haidenbauer, Maxim Mai, and Toru Sato, "Dynamical coupled-channel models for hadron dynamics," (2025), arXiv:2505.02745 [nucl-th].

[49] Marco Garofalo, Maxim Mai, Fernando Romero-López, Akaki Rusetsky, and Carsten Urbach, "Three-body resonances in the φ^4 theory," JHEP **02**, 252 (2023), arXiv:2211.05605 [hep-lat].

[50] Maxim Mai, Michael Döring, and Akaki Rusetsky, "Multi-particle systems on the lattice and chiral extrapolations: a brief review," Eur. Phys. J. ST **230**, 1623–1643 (2021), arXiv:2103.00577 [hep-lat].

[51] Andrei Alexandru, Ruairí Brett, Chris Culver, Michael Döring, Dehua Guo, Frank X. Lee, and Maxim Mai, "Finite-volume energy spectrum of the $K^-K^-K^-$ system," Phys. Rev. D **102**, 114523 (2020), arXiv:2009.12358 [hep-lat].

[52] Chris Culver, Maxim Mai, Ruairí Brett, Andrei Alexandru, and Michael Döring, "Three pion spectrum in the $I = 3$ channel from lattice QCD," Phys. Rev. D **101**, 114507 (2020), arXiv:1911.09047 [hep-lat].

[53] M. Mai, M. Döring, C. Culver, and A. Alexandru, "Three-body unitarity versus finite-volume $\pi^+\pi^+\pi^+$ spectrum from lattice QCD," Phys. Rev. D **101**, 054510 (2020), arXiv:1909.05749 [hep-lat].

[54] Maxim Mai, Chris Culver, Andrei Alexandru, Michael Döring, and Frank X. Lee, "Cross-channel study of pion scattering from lattice QCD," Phys. Rev. D **100**, 114514 (2019), arXiv:1908.01847 [hep-lat].

[55] C. Culver, M. Mai, A. Alexandru, M. Döring, and F. X. Lee, "Pion scattering in the isospin $I = 2$ channel from elongated lattices," Phys. Rev. D **100**, 034509 (2019), arXiv:1905.10202 [hep-lat].

[56] Maxim Mai, "3-body quantization condition in unitary isobar formalism," PoS **LATTICE2018**, 050 (2018), arXiv:1810.00604 [hep-lat].

[57] Maxim Mai and Michael Doring, "Finite-Volume Spectrum of $\pi^+\pi^+$ and $\pi^+\pi^+\pi^+$ Systems," Phys. Rev. Lett. **122**, 062503 (2019), arXiv:1807.04746 [hep-lat].

[58] M. Döring, H. W. Hammer, M. Mai, J. Y. Pang, § A. Rusetsky, and J. Wu, "Three-body spectrum in a finite volume: the role of cubic symmetry," Phys. Rev. D **97**, 114508 (2018), arXiv:1802.03362 [hep-lat].

[59] Maxim Mai, Bin Hu, Michael Doring, Alessandro Pilloni, and Adam Szczepaniak, "Three-body scattering in isobar ansatz," PoS **Hadron2017**, 140 (2018).

[60] M. Mai and M. Döring, "Three-body Unitarity in the Finite Volume," Eur. Phys. J. A **53**, 240 (2017), arXiv:1709.08222 [hep-lat].

[61] Peng Guo and Michael Döring, "Lattice model of heavy-

light three-body system," Phys. Rev. D **101**, 034501 (2020), arXiv:1910.08624 [hep-lat].
[62] Peng Guo, Michael Döring, and Adam P. Szczepaniak, "Variational approach to N-body interactions in finite volume," Phys. Rev. D **98**, 094502 (2018), arXiv:1810.01261 [hep-lat].
[63] M. Mikhasenko *et al.* (JPAC), "Dalitz-plot decomposition for three-body decays," Phys. Rev. D **101**, 034033 (2020), arXiv:1910.04566 [hep-ph].
[64] M. Albaladejo, D. Winney, I. V. Danilkin, C. Fernández-Ramírez, V. Mathieu, M. Mikhasenko, A. Pilloni, J. A. Silva-Castro, and A. P. Szczepaniak (JPAC), "Khuri-Treiman equations for 3π decays of particles with spin," Phys. Rev. D **101**, 054018 (2020), arXiv:1910.03107 [hep-ph].
[65] M. Mikhasenko, Y. Wunderlich, A. Jackura, V. Mathieu, A. Pilloni, B. Ketzer, and A. P. Szczepaniak, "Three-body scattering: Ladders and Resonances," JHEP **08**, 080 (2019), arXiv:1904.11894 [hep-ph].
[66] L. Ng, L. Bibrzycki, J. Nys, C. Fernandez-Ramirez, A. Pilloni, V. Mathieu, A. J. Rasmusson, and A. P. Szczepaniak (Joint Physics Analysis Center, JPAC), "Deep learning exotic hadrons," Phys. Rev. D **105**, L091501 (2022), arXiv:2110.13742 [hep-ph].
[67] L. Ng, Ł. Bibrzycki, J. Nys, C. Fernández-Ramírez, A. Pilloni, V. Mathieu, A. J. Rasmusson, and A. P. Szczepaniak, "Machine learning exotic hadrons," Nuovo Cim. C **47**, 207 (2024).
[68] Leonarc Michelle Santos, Vince Angelo A. Chavez, and Denny Lane B. Sombillo, "Pole structure of PψN(4312)+ via machine learning and uniformized S-matrix," J. Phys. G **52**, 015104 (2025), arXiv:2405.11906 [hep-ph].
[69] Vince Angelo A. Chavez and Denny Lane B. Sombillo, "Line shape analysis of $\Lambda(1405)$ in $\gamma p \to K^+\Sigma^-\pi^+$ reaction using convolutional neural network," in *21st International Conference on Hadron Spectroscopy and Structure* (2025) arXiv:2506.04622 [hep-ph].
[70] M. Malekhosseini, S. Rostami, A. R. Olamaei, R. Ostovar, and K. Azizi, "Meson mass and width: Deep learning approach," Phys. Rev. D **110**, 054011 (2024), arXiv:2404.00448 [hep-ph].
[71] Darwin Alexander O. Co and Denny Lane B. Sombillo, "Analysis of hidden-charm pentaquarks as triangle singularities via deep learning," PoS **QNP2024**, 031 (2025), arXiv:2411.14044 [hep-ph].
[72] Aurélien Dersy, Matthew D. Schwartz, and Alexander Zhiboedov, "Reconstructing S-matrix Phases with Machine Learning," JHEP **05**, 200 (2024), arXiv:2308.09451 [hep-th].
[73] Adam B. Mapa and Denny Lane B. Sombillo, "Feature extraction in partial wave analysis using K-matrix approach," in *21st International Conference on Hadron Spectroscopy and Structure* (2025) arXiv:2506.04628 [hep-ph].
[74] Darwin Alexander O. Co, Vince Angelo A. Chavez, and Denny Lane B. Sombillo, "Deep learning framework for disentangling triangle singularity and pole-based enhancements," Phys. Rev. D **110**, 114034 (2024), arXiv:2403.18265 [hep-ph].
[75] Denny Lane Sombillo, "Extraction of S-matrix pole structure using deep learning," PoS **PANIC2021**, 175 (2022).
[76] Denny Lane B. Sombillo, Yoichi Ikeda, Toru Sato, and Atsushi Hosaka, "Classifying Near-Threshold Enhancement Using Deep Neural Network," Few Body Syst. **62**, 52 (2021), arXiv:2106.03453 [hep-ph].
[77] Denny Lane B. Sombillo, Yoichi Ikeda, Toru Sato, and Atsushi Hosaka, "Model independent analysis of coupled-channel scattering: A deep learning approach," Phys. Rev. D **104**, 036001 (2021), arXiv:2105.04898 [hep-ph].
[78] Denny Lane B. Sombillo, Yoichi Ikeda, Toru Sato, and Atsushi Hosaka, "Unveiling the pole structure of S-matrix using deep learning," Rev. Mex. Fis. Suppl. **3**, 0308067 (2022), arXiv:2104.14182 [hep-ph].
[79] Denny Lane B. Sombillo, Yoichi Ikeda, Toru Sato, and Atsushi Hosaka, "Classifying the pole of an amplitude using a deep neural network," Phys. Rev. D **102**, 016024 (2020), arXiv:2003.10770 [hep-ph].
[80] J. Landay, M. Mai, M. Döring, H. Haberzettl, and K. Nakayama, "Towards the Minimal Spectrum of Excited Baryons," Phys. Rev. D **99**, 016001 (2019), arXiv:1810.00075 [nucl-th].
[81] R. Molina, Justin Landay, Bin Hu, Michael Doring, and Cesar Fernández-Ramírez, "Model selection for pion photoproduction," PoS **Hadron2017**, 135 (2018).
[82] Dimitrios Petrellis and Dalibor Skoupil, "Ridge regression for minimizing the couplings of hyperon resonances in K+Λ photoproduction," Phys. Rev. C **107**, 045206 (2023), arXiv:2212.14305 [nucl-th].
[83] Petr Bydžovský, Aleš Cieplý, Dimitrios Petrellis, Dalibor Skoupil, and Nicholas Zachariou, "Model selection for K+$\Sigma-$ photoproduction within an isobar model," Phys. Rev. C **104**, 065202 (2021), [Erratum: Phys.Rev.C 110, 029901 (2024)], arXiv:2106.15199 [nucl-th].
[84] T. Alghamdi *et al.*, "Toward a generative modeling analysis of CLAS exclusive 2π photoproduction," Phys. Rev. D **108**, 094030 (2023), arXiv:2307.04450 [hep-ph].
[85] Daniele Binosi, Alessandro Pilloni, and Ralf-Arno Tripolt, "Study for a model-independent pole determination of overlapping resonances," Phys. Lett. B **839**, 137809 (2023), arXiv:2205.02690 [hep-ph].
[86] Daniel Sadasivan, Maxim Mai, Michael Döring, Ulf-G. Meißner, Felipe Amorim, John Paul Klucik, Jun-Xu Lu, and Li-Sheng Geng, "New insights into the pole parameters of the $\Lambda(1380)$, the $\Lambda(1405)$ and the $\Sigma(1385)$," Front. Phys. **11**, 1139236 (2023), arXiv:2212.10415 [nucl-th].
[87] Feng-Kun Guo, Yuki Kamiya, Maxim Mai, and Ulf-G. Meißner, "New insights into the nature of the $\Lambda(1380)$ and $\Lambda(1405)$ resonances away from the SU(3) limit," Phys. Lett. B **846**, 138264 (2023), arXiv:2308.07658 [hep-ph].
[88] P. C. Bruns, A. Cieplý, and M. Mai, "Testing chiral unitary models for the $\Lambda(1405)$ in K+$\pi\Sigma$ photoproduction," Phys. Rev. D **106**, 074017 (2022).
[89] Maxim Mai, "Status of the $\Lambda(1405)$," Few Body Syst. **59**, 61 (2018).
[90] Ales Cieplý and Maxim Mai, "Theoretical approaches to low energy $\bar{K}N$ interactions," EPJ Web Conf. **130**, 02001 (2016), arXiv:1610.06444 [nucl-th].
[91] A. Cieplý, M. Mai, Ulf-G. Meißner, and J. Smejkal, "On the pole content of coupled channels chiral approaches used for the $\bar{K}N$ system," Nucl. Phys. A **954**, 17–40 (2016), arXiv:1603.02531 [hep-ph].

[92] Maxim Mai and Ulf-G. Meißner, "Constraints on the chiral unitary $\bar{K}N$ amplitude from $\pi\Sigma K^+$ photoproduction data," Eur. Phys. J. A **51**, 30 (2015), arXiv:1411.7884 [hep-ph].

[93] Albert Feijoo, Volodymyr Magas, and Angels Ramos, "$S=-1$ meson-baryon interaction and the role of isospin filtering processes," Phys. Rev. C **99**, 035211 (2019), arXiv:1810.07600 [hep-ph].

[94] Maxim Mai, Michael Döring, Carlos Granados, Helmut Haberzettl, Jackson Hergenrather, Ulf-G. Meißner, Deborah Rönchen, Igor Strakovsky, and Ron Workman (Jülich-Bonn-Washington), "Coupled-channels analysis of pion and η electroproduction within the Jülich-Bonn-Washington model," Phys. Rev. C **106**, 015201 (2022), arXiv:2111.04774 [nucl-th].

[95] M. Mai, J. Hergenrather, M. Döring, T. Mart, Ulf-G. Meißner, D. Rönchen, and R. Workman (Jülich–Bonn–Washington), "Inclusion of $K\Lambda$ electroproduction data in a coupled channel analysis," Eur. Phys. J. A **59**, 286 (2023), arXiv:2307.10051 [nucl-th].

[96] R. Navarro Perez, J. E. Amaro, and E. Ruiz Arriola, "Statistical error analysis for phenomenological nucleon-nucleon potentials," Phys. Rev. C **89**, 064006 (2014), arXiv:1404.0314 [nucl-th].

[97] R. Navarro Pérez, J. E. Amaro, and E. Ruiz Arriola, "Coarse-grained potential analysis of neutron-proton and proton-proton scattering below the pion production threshold," Phys. Rev. C **88**, 064002 (2013), [Erratum: Phys.Rev.C 91, 029901 (2015)], arXiv:1310.2536 [nucl-th].

[98] S. Navas *et al.* (Particle Data Group), "Review of particle physics," Phys. Rev. D **110**, 030001 (2024).

[99] Martin Hoferichter, Jacobo Ruiz de Elvira, Bastian Kubis, and Ulf-G. Meißner, "Nucleon resonance parameters from Roy–Steiner equations," Phys. Lett. B **853**, 138698 (2024), arXiv:2312.15015 [hep-ph].

[100] R. Garcia-Martin, R. Kaminski, J. R. Pelaez, and J. Ruiz de Elvira, "Precise determination of the f0(600) and f0(980) pole parameters from a dispersive data analysis," Phys. Rev. Lett. **107**, 072001 (2011), arXiv:1107.1635 [hep-ph].

[101] J. R. Pelaez, "Light scalars as tetraquarks or two-meson states from large N(c) and unitarized chiral perturbation theory," Mod. Phys. Lett. A **19**, 2879–2894 (2004), arXiv:hep-ph/0411107.

[102] Joel Grus, *Data Science from Scratch: First Principles with Python*, 2nd ed. (O'Reilly Media, Sebastopol, CA, 2019).

Appendix A: Additional Numerical Tests: Network Architecture

The architecture of the neural network is in general somewhat arbitrary. To provide some level of objectivity in the context of the current application, we perform a numerical test demonstrating that, while there is not one ideal choice, the architecture is at least credible.

The neural network employed in this paper (except for the ones used in this test) chooses 4 hidden layers with 40, 30, 20 and 10 nodes respectively. To test whether the number of nodes is sufficient we perform the following procedure. We vary the total number of nodes in the neural network, starting by fixing the network at 4 layers with the first layer having 4 times the number of nodes in the final layer, the second having 3 times, and the third layer having 2 times. We generate 10 sets of M and Γ with uniformly distributed values between 400-1200 MeV and 0-500 MeV, respectively. With these values, we generate corresponding sets of pseudodata using the Breit-Wigner-like function described in the toy model. We then train the neural network 10 separate times. For each time, we calculate the Mean Squared Error (MSE) , which is the chosen loss function in this work, between the model's prediction and the parameters used for generation. The average MSE for each neural network is given in the left half of Tab. IV. We find that, for a very small network, increasing the number of nodes significantly increases the accuracy of prediction, however, beyond a certain size, increasing the number of nodes has a vanishingly small effect. The neural network we work with in the main part, with 100 nodes, is close to the point at which increasing the size has only a small effect on the accuracy. Since increasing the size of the neural network comes with an increased computational cost, we choose to work with a neural network of 100 nodes.

The right side of Tab. IV holds the number of nodes fixed at 100 and varies the number of layers. The nodes are distributed among the layers as closely as possible to the distribution in which each layer decreases by the number of nodes contained in the final hidden layer. We find that a neural network with only one layer is already reasonably accurate, but increasing the number of layers provides a modest improvement up to around 4 layers. Beyond this, increasing the number of layers starts to provide a sharp increase in the MSE.

TABLE IV. A numerical test about the ideal architecture of the neural network. The left two columns test the number of nodes used. The MSE referenced is the average mean squared error for 10 different parameters. In all rows in these columns the nodes are distributed in four layers, with nodes decreasing in integer multiples of the number of nodes in the final layer. The left two columns test the number of hidden layers used. 100 nodes is distributed among the given number of layers. In both sides of the figure, the neural network used in the main text of this paper with 100 nodes and four hidden layers is put in bold font.

# of nodes	MSE
20	9444.05
50	4737
80	59.84
100	**53.5**
150	53.77
500	53.2

# of hidden layers	MSE
1	61.05
2	61.42
3	56.83
4	**53.5**
5	62.35
9	28987

Appendix B: Additional Numerical Tests: Convergence

Here, we perform a numerical test to determine whether 100 samples of the neural network are sufficient and whether the result converges with the sample size. Our method is as follows: We use the data set employed in the first toy example in which two points have been intentionally altered. We generate a list of 2000 neural network predictions. This list is then randomly broken into samples of a given size, N_s. The average and standard deviation of each sample is computed. From these a quantity known as the Standard Error of the Standard Deviation (SESD) is computed. This quantity consists of taking the standard deviation of the list of the standard deviations for each set. The Mean of Standard Deviations (MSD) is also computed. To obtain a more reliable value, these quantities are averaged over 1000 random partitions of the list of data into samples of size N_s. The ratio of these two quantities gives an estimate of what fractional or percentage variation that might occur between the standard deviation of a single sample of size N_s and the standard deviation of an arbitrarily large sample. These values are shown for M and Γ in the left two columns of Tab. V. We find that the standard deviation of a typical sample of size $N_s = 5$ might be off by 40% but that reduces to 8-11% for a sample of $N_s = 100$ as is employed in this work.

Additionally, we compute the $\bar{\chi}^2_{\rm dof}$ by partitioning the list into samples of size N_s as described in the preceding paragraph. The χ^2 statistic is calculated using the average $\overline{x_i^s}$ ($x \in \{M, \Gamma\}$) and standard deviation σ_i^s of sample

$i \in \{1, 2000/N_s\}$ as well as the average from the whole list excluding that sample $\overline{x_i}$ as

$$\bar{\chi}^2_{\text{dof}} := \sum_i (\overline{x_i^s} - \overline{x_i})^2 / (\sigma_i^s / \sqrt{N_s})^2 \,. \tag{B1}$$

Assuming that the list is large enough that differences between its average quantities and the average quantities in an infinite list are negligible, and assuming that error is distributed normally and thus the uncertainty is $\sigma_i^s/\sqrt{N_S}$, this statistic should follow the $\bar{\chi}^2$ distribution. If this distribution is divided by the number of degrees of freedom—in this case the number of samples in the list—it has an expected value of 1. Table V contains this measure for both M and Γ averaged over many random re-partitions of the list. It shows that for a sample of $N_s = 5$, the sample size is not large enough to obtain a reasonable convergence and uncertainty may not be sufficient. However, with a sample of $N_s = 100$, the size employed in this work, the $\bar{\chi}^2_{dof}$ approaches 1, indicating that this uncertainty can sufficiently account for uncertainties due to fluctuations in the neural network.

TABLE V. The average $\bar{\chi}^2_{\text{dof}}$ and average SESD/MSD ratio for M and Γ

N_s	χ^2_{dof}	SESD/MSD for M	SESD/MSD for Γ
5	1.87	0.40	0.39
10	1.27	0.27	0.24
25	1.08	0.18	0.18
50	1.04	0.16	0.12
100	1.03	0.11	0.08