A Computational Framework for Automated Nuclear Resonance Evaluation and Validation

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Noah Anthony Wy Walton

December 2024

To my loving wife, Bailey, who's given me courage and purpose.

To my parents and grandparents, who instilled in me the values of hard work, dedication, and the importance of enjoying life's journey.

Acknowledgments

I would like to express my deepest gratitude to my advisor, Vladimir Sobes, for first believing in me and then inspiring me. His guidance, insight, and support have been invaluable, shaping my growth as both a researcher and an individual. He has fostered an environment within our research group marked by camaraderie and thought-provoking discussions, and I am grateful for each of the members I've had the opportunity to know and collaborate with.

I am also grateful to my committee members and the many mentors who have guided my development, including Jesse Brown, Jason Hayward, Lawrence Heilbronn, Denise Neudecker, David Brown, Mike Grosskopf, and others. Their constructive feedback, encouragement, and expertise were instrumental in shaping this work.

This material is based upon work supported by the Department of Energy National Nuclear Security Administration through the Nuclear Science and Security Consortium under Award Number DE-NA0003996.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Abstract

Global and national efforts to deliver high-quality nuclear data to users have a wide-ranging impact, affecting applications in national security, reactor operations, basic science, medicine, and more. Cross section evaluation is a major part of this effort, combining theory and experimentation to produce recommended values and uncertainties for reaction probabilities. This thesis presents two major, novel methodological contributions to the field of nuclear data evaluation, with a focus on resonance region cross sections. The first is a methodology for automating resonance parameter inference, saving valuable time for evaluators and analysts, while also enhancing reproducibility. The second is a computational framework that leverages high-utility generative modeling to test, validate, and benchmark the performance of inferential methods. The integration of these two approaches enables a quantitative assessment of the automated algorithm's performance and establishes a framework that can be broadly applied to address a wide range of scientific questions. Several demonstrations highlight the computational experiments made possible by the framework, and the final results of the automated algorithm are compared against human evaluation of actual measurement data for ^{181}Ta.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

A primary objective of the nuclear data community is to produce recommended mean values and some measurement of confidence (often in the form of covariances) for neutron-induced reaction cross sections. This is a monumental effort that requires global collaboration among scientists, software developers, and experimental facilities, ultimately resulting in evaluated nuclear data libraries such as ENDF/B-VIII.0 [2] or JEFF 3.3 [1]. The data in these libraries represent the frontier of our knowledge in fundamental nuclear physics, encapsulating this information in a universal format. These libraries are essential inputs for predictive modeling and simulation of nuclear systems, supporting a wide range of applications in both engineering and fundamental science, including astrophysics, power reactor operations, stockpile stewardship, non-proliferation, various industrial processes, and isotope production for medical use. Additionally, the uncertainty, or covariance, associated with the mean values can be systematically propagated to simulated output quantities. This allows for the construction of confidence intervals on predicted quantities, which is a major consideration when making engineering or safety decisions based on predictive modeling. Recent advances in simulation capabilities have led to an increased demand for accuracy and completeness in the fundamental nuclear data that underpin predictive simulations [3].

Nuclear data evaluation is a crucial step in producing accurate nuclear data libraries, combining theoretical models and experimental measurements to provide the best estimates for mean values and their uncertainties. Often, the work is segmented into different areas of expertise. Complete documentation and consistency across this massive project are

recognized challenges, even with modern computational hardware. Furthermore, much of the existing knowledge, experimental data, and software were developed during the second half of the 20th century. As nuclear data scientists from that era retire, preserving their domain knowledge becomes essential, while also presenting an opportunity to modernize existing methods and tools. The need to seize this opportunity is well recognized within the international community of nuclear data evaluators and experimentalists, with the goal of addressing challenges such as workforce support, preservation of institutional knowledge, rigorous and trustworthy uncertainty quantification, and improved reproducibility, reportability, and reliability. These goals and their significance are formally addressed by several national and international collaborations. For example, the Working Party on International Nuclear Data Evaluation Co-operation [4], hosted by the Organization for Economic Co-operation and Development's Nuclear Energy Agency (OECD-NEA), has several active or recently active subgroups aimed at addressing these issues, including one specifically focused on reproducibility [5]. The United States' Nuclear Data Program currently hosts the Cross Section Evaluation Working Group (CSEWG), which was established over 50 years ago to address pertinent issues in the field.

Automation represents a promising, viable approach to enhancing reproducibility in the evaluation process. Automation refers to a standardized computational methodology or set of procedures, ideally a software package, that reduces manual input and ensures proper execution of each incremental step. The expertise of the evaluator will always be needed, but introducing automation into the workflow will reduce the manual workload on an evaluator and make it easier to document more nuanced decisions. The inference of physics model parameters from experimental data is a promising candidate for automation in the evaluation workflow. The big-picture for this thesis is to develop a computational framework for automation of the inference problem in evaluation and for quantitatively testing and benchmarking the performance of the automated tool. This framework provides a means for computational validation of methods for automation and uncertainty quantification, differing from the traditional association of 'validation' in nuclear data with the comparison of evaluated values to integral experimental data.

The evaluation of nuclear data is often divided into distinct areas of expertise, with each facing its own set of unique challenges. This thesis will focus on the evaluation of neutron-induced reaction cross sections in the resolved resonance region (RRR). Developing these methods specifically for RRR cross sections is supported in several ways. Firstly, many applications are particularly sensitive to the accuracy and uncertainty of cross sections in this energy regime. Also, accurately characterizing this region is foundational to the evaluation of other energy regimes. The importance of RRR evaluations is detailed in a number of comprehensive reviews [6, 7]. Secondly, the need for reproducibility in RRR evaluations is pertinent. The default method for fitting to experimental data – a Bayesian method which reduces to generalized linear least squares (GLLS) under certain conditions – suffers a number of shortcomings in practice. 1) It does not consider discrete variables or parameter cardinality which are key aspects of determining the most appropriate model. 2) It is well documented that the uncertainty quantification (UQ) is underpredictive [8, 9], the presumed cause being the violation of assumptions in the GLLS regression.

This thesis delivers a method and corresponding tool to automatically perform the fitting of theory to differential measurement data for RRR cross section evaluations. This differs from traditional approaches in two primary ways: (1) it does not rely on a prior evaluation or manually determined starting point, (2) if the resulting parameters are not satisfactory, the evaluator should change the input(s) to the automated methodology (curated data, algorithm hyperparameters, etc.) rather than directly altering the resonance parameters themselves. The latter results in an easy way to report reproducible resonance parameter estimates while the former eliminates what has been pointed out as one of the most challenging part of RRR evaluation [10].

The automated fitting approach is paired with a synthetic data method/tool to aid in the development of the automated evaluation and to benchmark its inferential capabilities. By generating synthetic data that is statistically indistinguishable from real measurement data, these benchmarks can be extrapolated to evaluations of real data. This provides a unique ability to quantify the limits of the evaluation community's inferential capabilities - allowing traditionally experience-based decisions and theoretical assumptions to be rigorously tested and justified. The relevance of this ability is demonstrated through an impactful study

quantifying the end of the resolved resonance region and beginning of the unresolved resonance region or URR. Traditionally, this has been determined in a heuristic manner based on expert judgement. Within the framework proposed here, the degradation of evaluation accuracy as a function of decreasing experimental resolution will be quantitatively assessed.

The following chapters will provide both a historical and contemporary overview of the field of RRR evaluation (2), outline the development and testing of a novel computational framework for evaluation (3 and 4), and present several demonstrations that showcase the engineering and scientific questions this framework can address (5). Lastly, a discussion of the impact this work has and the future directions for it will be had in 6.

Chapter 2

Background

2.1 Nuclear Cross-Sections

Reaction cross-sections describe the probability of a given reaction occurring when a particle is incident on some target nucleus. Generally, the notation for a microscopic[1] reaction cross section is σ_{rxn} and the units are barns (b) where 1b = 10^{-24}cm^2. There are two primarily relevant processes that occur, direct and compound-nucleus reactions. There are other reaction mechanisms such as pre-equilibrium and spallation, but these are less relevant to the discussion in this thesis. In direct reactions, an incident particle interacts mostly with the surface of the target nuclei in a quick manner. This reaction type dominates at high incident particle energies. The theoretical model used to describe direct reaction (high-energy) cross-sections is the Optical Model [11]. When the incident particle has less energy, direct reactions are less probable and instead it starts to interact with the target nucleus over longer time frames. This allows for the formation of a new, 'compound' nucleus. The incident particle distributes it mass/energy across the compound nucleus and equilibrium is reached. The new, compound-nucleus is now in an excited state and must de-excite. There are many paths for de-excitation via the emission of some particle like a gamma, neutron, proton, or other. Compound-nucleus reactions were originally described by the Hauser-Feshbach model in 1952 [12]. Compound nucleus reaction cross sections are the focus of this

[1]Microscopic is a naming convention relevant for neutron transport calculations to differentiate it from the effective reaction cross section for a specific volume of material (macroscopic). The microscopic cross section is the general physics property and will be the focus of this thesis.

thesis, though extension of the methods developed here to energy regimes where the optical model dominates could be considered in future work.

In the energy regime of compound nucleus reactions, elastic scattering and radiative capture reactions—denoted by σ_{el} and σ_{γ} respectively—are often the only energetically possible reactions. For fissionable isotopes, a fission reaction cross section σ_{fis} also exists. Each of the possible reaction cross sections are referred to as constituent reactions, and the total reaction cross section σ_{tot} is given by the sum of all constituent reaction cross sections.

2.1.1 Resonance Theory

When a compound nucleus is formed, the incident particle interacts with many nucleons within the target nucleus and distributes its energy across all nucleons. The result is the compound nucleus; a new, excited nucleus that must decay back to a stable state via the emission of some particle. The compound nucleus is most likely to form at specific, probable excited states or energy levels. At high incident particle energies, there are many such levels, allowing them to be represented statistically with the Hauser-Feshbach model [12]. However, at lower incident particle energies, representing the specific level becomes important as they begin to cause discrete, significant increases in the reaction cross-section called resonances. The R-matrix theory for compound-nucleus reactions was developed to describe these reaction states and represent the cross-sections in detail rather than as a statistical average. The theory was fully formulated in [13]. More recent reviews of the theory with relevant additions can be found in Refs. [14, 15]. The following will be a high-level overview based on the above references, discussing the relevant parameters and assumptions for the work presented in this thesis.

A foundational concept of R-matrix theory is the separation of the process into an internal and external region by a reaction channel radius (a_c) that is roughly larger than the geometrical radius of the nucleus. The external region is dominated by long-range forces and the wavefunction can be well-approximated by a free particle with the possibility for consideration of the coulomb force. The internal region is dominated by the short-range nuclear force, which is less understood, especially in regions where the fine details of the nuclear potential have significant impact such as the resonance range. Rather than

attempting to model this internal region via first principles, the phenomenological R-Matrix theory parameterizes it. The wavefunction for the confined internal region is expanded into discrete basis functions that follow conservation laws (momentum, angular spin, and parity) and the logarithmic derivative is matched to that of the known external region at the boundary (defined by the reaction channel radius) for consistency. The parameters of this model are then fit to experimental measurement data. The elements of the R-matrix itself are the inverse of the logarithmic derivative of the wavefunction at the boundary of the internal and external regions, given by

$$R_{cc'} = \sum_{\lambda} \frac{\gamma_{\lambda c}\gamma_{\lambda c'}}{E_\lambda - E}\delta_{J^\pi J^{\pi'}} \ , \tag{2.1}$$

where: λ = resonance level,
c/c' = incoming/outgoing channel,
γ = reduced width amplitude,
E_λ = energy of the resonance,
E = incident particle energy,
J^π = total angular momentum and parity π.

The subscripts c and c' represent the incoming and outgoing channels respectively, describing the state of the system before and after the interaction that occurs in the internal region. Here, 'state' characterizes the two particles involved by mass, charge, spin, threshold energy, and parity. It is convenient to define the relationship between the reduced width amplitudes (γ) and the partial widths (Γ) because these correspond to the observable width of the resonance shape:

$$\Gamma_{\lambda,c} = 2P_c(E)\gamma^2_{\lambda,c} \ , \tag{2.2}$$

where $P(E)$ is the penetrability factor. The term (E_λ) corresponds to the incident energy location of the resonance. The R-matrix mathematically relates to cross-section via the level (A) and scattering (U) matrices:

$$R \Longrightarrow A \Longrightarrow U \Longrightarrow \sigma. \tag{2.3}$$

With the full derivation left to the references, the general form of a reaction cross-section from channel c to c' is given by

$$\sigma_{cc'} = \sum_{J^\pi} \frac{\pi}{k_\alpha^2} g_\alpha |e^{2iw_c}\delta_{cc'} - U_{cc'}|^2, \tag{2.4}$$

where: α = definition of the two particles in the pair (mass/charge/spin),
J^π = total angular momentum and parity (π)of the pair,
k_α^2 = wave-number of the pair,
g_{J_α} = spin statistical factor,
w_c = coulomb phase shift (0 for non-coulomb channels),
U = scattering matrix.

The total cross section is calculated as sum over all incident and exit channels, c and c' respectively. Similarly, the capture cross section is calculated as a sum over all c' that corresponding to gamma channels.

The full R-matrix formalism is not often used in practice, especially for heavy nuclei with many resonances. The most common approximation is the Reich-Moore or eliminated channel approximation [16]. If the excited, compound nucleus decay via the emission of photons, many possible decay channels exist in the form of different gamma cascades down to the ground state. The Reich-Moore approximation (RM) aggregates these gamma channels, neglecting interference terms between individual gamma channels and between the aggregate gamma channel and other non-gamma channels. These interference terms are assumed to be negligible due to the nature of gamma channels as compared to particle channels. This is supported by a relatively constant radiation width observed in resonances of the same spin and parity and a generally symmetric resonance shape observed in radiative capture cross-section measurements. The RM formalism is recommended for modern evaluations [6, 8].

While resonance parameters are not known from first principles, they tend to fall on well characterized statistical distributions. For any given spin group (J^π), the channel widths tend to fall on the Porter-Thomas distribution [17], which generalizes to a χ^2 distribution with degrees of freedom equal to the number of open channels scaled by the average width. This is derived from the idea that the reduced width amplitudes γ_c are normally distributed

around zero with a finite variance given by $\langle\gamma_c^2\rangle$:

$$\left\langle\gamma_c^2\right\rangle - \left\langle\gamma_c\right\rangle^2 = \left\langle\gamma_c^2\right\rangle - 0 \equiv \left\langle\gamma_c^2\right\rangle \tag{2.5}$$

$$\gamma_c \sim N(0, \left\langle\gamma_c^2\right\rangle). \tag{2.6}$$

Normalizing by the standard deviation gives a standard normal

$$\frac{\gamma_c}{\sqrt{\left\langle\gamma_c^2\right\rangle}} \sim N(0,1), \tag{2.7}$$

and the square of a standard normal is known to follow a χ_k^2 distribution with one degree of freedom. Furthermore, the sum of several χ_k^2 distributions will follow a χ_K^2 distribution where $K = \sum k$, assuming that each random variable in the sum is independent. There is usually only one or two open neutron channels. As discussed earlier, the RM formalism aggregates what is presumed to be a large number of possible outgoing gamma channels. In this case, the resulting distribution would have a large number of degrees of freedom and a smaller relative variance[2]. As a result, the evaluated aggregate capture channel is often set to a single value for all resonances of a given isotope and only one capture channel is fit.

The energy spacing between neighboring resonances of the same spin group also tends to fall on a well characterized probability distribution called the Wigner distribution. This distribution can be derived from random matrix theory, particularly the Gaussian Orthogonal Ensemble (GOE), and our current understanding of quantum chaotic systems. The probability distribution for resonance spacing was first introduced in [18].

The applicability of these statistical distributions is supported by their use in representing the unresolved resonance range (URR) in continuous energy Monte Carlo transport calculations. The onset of the URR occurs where individual resonance structures cannot be fully resolved experimentally but still cause significant variance in the cross section. At this point, the resonances are treated in a statistical manner with a Hauser-Feshbach (HF) model of the average cross section. It is the responsibility of the evaluator to decide where the RRR should end and the URR should begin.

[2]This assumes that each channel contributing to the sum has the same or close to the same variance.

The HF average cross section in often insufficient for simulations because they involve non-linear transformations and the average of a non-linear function does not equal the non-linear function of the average $\langle f(x)\rangle \neq f(\langle x\rangle)$. This phenomenon is commonly referred to in the field of nuclear data and neutron transport as resonance self-shielding and can be mitigated by constructing probability tables which represent the variance in cross section in this energy regime and allow the proper average, $\langle f(x)\rangle$, to be recovered. Probability tables are constructed from realizations of resonance parameter sequences sampled from the Porter-Thomas and Wigner distributions. This is done in a number of processing codes [19, 20, 21]. This work leverages the high-fidelity implementation developed by Brookhaven National Laboratory (BNL) [22] distributed with the open source processing code FUDGE [21] developed by Lawrence Livermore National Laboratory (LLNL).

2.1.2 Neutron Time-of-Flight Measurements

The theoretical parameters used to calculate reaction cross sections in the resonance region are not known exactly, they must be empirically determined through comparison to experimental measurements. However, microscopic reaction cross sections are not directly measurable. Instead, reaction cross sections are inferred from other directly measured observables, or more precisely, from composite observables, which are functions of directly measured observables and other experimental conditions/parameters. The reaction cross sections—more specifically, their theoretical parameterizations—are related to the composite observables through a mathematical model, making them latent variables.

The observables, the descriptions of experimental conditions, and even sometimes the mathematical models have uncertainties associated with them. As a result, a key challenge in experimental physics is designing experiments that maximize sensitivity to the latent variables of interest while minimizing the impact of other uncertainties. There are a wide range of experimental data that have sensitivity to resonance cross sections which are generally separated into two categories: integral and differential measurement data. The former are measurements of nuclear systems that can be compared to cross sections via neutron transport calculations that are a function of many nuclear data and often involve an integral over the energy dimension of the cross sections (i.e., the sensitivity of

the response distributed to many nuclear data). Integral data have historically been held back for validation, while differential data are used for evaluation[3]. The term differential is somewhat of a misnomer because the units of the data collected are not actually differential[4]. Instead, the term is used to distinguish the data from the previous type, indicating that the energy dependence of the cross section has not been collapsed. Experimentally, the primary method by which the energy dependence of a cross section is measured is through neutron time-of-flight (TOF) experiments.

In TOF experiments, the time it takes for neutrons to travel across some distance is precisely measured, within a few nano-seconds. This is related to the kinetic energy of the neutron by the following, non-relativistic equation:

$$KE = \frac{1}{2}mv^2 = \frac{1}{2}m(\frac{L}{t})^2, \tag{2.8}$$

where: KE = kinetic energy of the particle,
m = mass of the particle ,
v = velocity of the particle ,
L = flight path ,
t = time-of-flight .

The measurement data are linearly separated in TOF. Because of the proportionality to $\frac{1}{ToF^2}$, the measurement data in energy become further apart at high energies. This effect is often referred to as the resolution.

The actual data collected are small electronic pulses that are then amplified, shaped, and converted to digital signals and analyzed. The resulting data are detector counts with respect to linearly separated time bins, usually collected over several different measurement cycles. Because detector counts are integer events, they can be combined by simple addition and the noise/uncertainty model is given by the Poisson probability distribution [25]. The experimentalist typically combines count data from multiple cycles and groups it into larger

[3]Recent work has looked at incorporating integral measurements in the evaluation of fundamental reaction cross sections [23] or adjusting them for application specific uses [24].
[4]The measurement data here are not rigorously/mathematically differential with respect to energy, that is, the cross section units are not barns per differential change in energy. Instead, each measurement point is the energy integral of the cross section over a small energy bin.

time bins, increasing the number of counts per time bin, therefore bettering the counting statistics, at the cost of a decreased resolution in TOF. While the analog signals truly are direct observables, the term "raw obseravables" or "raw data" generally refers to the combined, grouped detector counts.

Data reduction is the transformation of raw data into composite observables that are mathematically closer to $\sigma^{\mathrm{rxn}}(E)$. Much of the data reduction can be considered a correction for experimental conditions. For example, the measured signal must be corrected for the probability of creating a signal in the the detection system, called the detector response function, which changes with respect to the incident particle energy. For resonance cross sections, the closest object to $\sigma^{\mathrm{rxn}}(E)$ one can get with data reduction are transmission and reaction yield(s). Data reduction often simplifies data storage and facilitates parameter inference, but it requires meticulous uncertainty quantification. If only the reduced data are reported, it can be difficult to correct for biases in the parameters or methodologies used and can even render the data unusable [9, 26, 27]. The debate over the best approach to store and report measured data is ongoing in the field, with advantages and disadvantages to both raw and composite data reporting.

The uncertainty quantification associated with data reduction typically utilizes first-order linear error propagation, as outlined in both a general context and specifically for reaction cross-section measurements in Refs. [8, 25, 28]. The general form is

$$V = g^{\mathrm{T}} \mathrm{v}^{-1} g, \tag{2.9}$$

where V represents the $n \times n$ propagated covariance matrix for the n output parameters (composite observables) and v denotes the covariance matrix for the input parameters (raw observables and experimental parameters). The matrix g is the Jacobian relating input parameters to output parameters to first order. Because the raw observables are uncorrelated, it is convenient to decompose the calculation of V into a rank-m perturbation on a diagonal matrix,

$$V = \mathrm{s}^{-1} + g^{\mathrm{T}} \mathrm{m}^{-1} g, \tag{2.10}$$

where s is a diagonal matrix describing the independent variance component propagated from the raw observables and m is the $m \times m$ covariance matrix for the experimental parameters, often referred to as systematic uncertainty. Note that this reduces the dimensionality of g to $n \times m$ rather than $n \times (m + n)$. This error propagation approach can be related to the Bayesian framework for uncertainty quantification (more details Sec. 2.5.1) under the conditions that the parameters are all normally distributed[5] and their relationship to the output are all linear. The latter is not in fact true; however, in the limit of small variance, the non-linear relationships are well approximated by their first-order Taylor expansion, allowing Eqs. 2.9 and 2.10 to hold.

The following sections detail data reduction for transmission and capture yield experiments as these are the relevant experiments for the demonstrations in this thesis. These are also two of the most common types of experimental nuclear data available for many isotopes of interest. Data reduction can vary depending on the facility and the specific measurement procedure used. The measurement and reduction methodology on which this work is based is described in detail in [28] with measurements of ^{181}Ta at the Rensselaer Gaerttner Linear Accelerator, which will be analyzed later in this thesis.

Transmission Measurements

Transmission measurements are used to infer the total reaction cross-section, the relationship between the two is given by

$$T = e^{-n\sigma_{\text{tot}}}, \tag{2.11}$$

where T is the transmission, σ_{tot} is the total cross-sections in barns, and n is the target thickness through which neutrons are transmitting in units $\frac{\text{atoms}}{\text{barn}}$. The transmission is a composite observable, calculated via a data reduction methodology. It is essentially the ratio of neutrons that pass through a sample to those incident upon it. This is calculated by comparing the neutron counts with and without a target sample placed between the neutron beam and the detector system, with additional terms accounting for neutron beam

[5]This is not technically true for raw observables as detector counts are governed by the Poisson distribution. However, the Poisson distribution is well approximated by a normal distribution when the number of counts (λ) is large. See Appendix A.

consistency and background radiation. The full data reduction is described by Eqs. 2.12 and 2.13:

$$T = \frac{\text{sample in}}{\text{sample out}} = \frac{\alpha_1 \dot{c}_\text{s} - \alpha_2 k_\text{s} \dot{B} - \dot{B}0_\text{s}}{\alpha_3 \dot{c}_\text{o} - \alpha_4 k_\text{o} \dot{B} - \dot{B}0_\text{o}}, \tag{2.12}$$

$$\dot{c} = \frac{c}{b_\text{w}\,\text{trig}} \times dtcf, \tag{2.13}$$

where: T = transmission ,

$\dot{c}_{s/o}$ = dead time corrected count rates for sample in/out ,

α = monitor normalization factors ,

$k_{s/o}$ = normalization for time-dependent background for sample in/out ,

$\dot{B}$ = time dependent background function ,

$\dot{B}0$ = constant room background function ,

c = detector counts for either sample in or out ,

$dtcf$ = detector dead-time correction factor ,

b_w = time bin width ,

$trig$ = number of times that bin was open for detection .

The TOF-dependent characterization of background $\dot{B}$ is an analytical function that is fitted to black resonances; known strong resonances where the transmission is expected to be zero. There are several functional forms that can be used depending on the shape [29, 30], those relevant for this work are the exponential and power functions given by

$$\dot{B}(t_i) = ae^{-bt_i}, \tag{2.14}$$

and

$$\dot{B}(t_i) = at_i^{-b}, \tag{2.15}$$

respectively, where a and b are the fitted parameters and t_i is the TOF. Note that each quantity in Eq. 2.12 exists independently for a specific time bin. While the notation for time dependence has been omitted for conciseness, the actual calculation is performed using a vectorized representation, where the elements correspond to discrete TOF bins.

In the reference experiments, TOF bins for the detection system are about 6 ns. The data is dead time corrected and monitor normalized for each cycle and summed over all cycles. The monitor normalizations correct for changing intensities of the incident neutron beam between measurements. Then TOF bins are grouped at the discretion of the experimentalist. At this point, the relationships described in Eqs. 2.12–2.15 are used to calculate a transmission spectrum with respect to TOF. Uncertainties are propagated to the spectrum using Eq. 2.10, accounting for systematic errors from the background functions and non-systematic (statistical) errors from detector counting statistics.

The measured count spectrum, c, does not accurately represent the true spectrum incident on the detector due to the influence of the detector's response function. Estimating the detector response function and it's uncertainty can be one of the most challenging aspects of data reduction. For transmission measurements, the detector response is assumed to cancel out in the ratio as both the sample in and sample out measurements use the same detector system.

Capture Yield Measurements

Capture yield measurements are used to infer the capture reaction cross-section, the relationship between the two is given by

$$Y_\gamma = \frac{\sigma_\gamma}{\sigma_{\text{tot}}}(1-T) = \frac{\sigma_\gamma}{\sigma_{\text{tot}}}(1-e^{-n\sigma_{\text{tot}}}), \tag{2.16}$$

where Y is the yield, σ_γ is the capture cross section in barns, and T is the transmission. Data reduction for reaction cross sections is often more complicated than transmission measurements because the incident flux spectrum must be characterized. The general equation for a reaction yield is

$$Y_{\text{rxn}} = f_{\text{n}}\frac{\dot{c}_{\text{rxn}} - \dot{c}_{\text{bg}}}{\phi}, \tag{2.17}$$

where $\dot{c}_{\text{rxn}}$ is the count rate for the particular reaction of interest ($\dot{c}_\gamma$ for radiative capture), $\dot{c}_{\text{bg}}$ is the background count rate, ϕ is the un-normalized characterization of the incident neutron flux spectrum, and f_n is a normalization factor. The incident neutron flux is measured in a separate experiment where the flux shape is determined by measuring detector count rates,

subtracting background, and then normalizing by the yield of a well known sample, often B_4C as the $(n, \alpha\gamma)$ cross section for ^{10}B is well characterized and there are no competing photon reactions below 1 MeV. The data reduction for this flux measurement is given by

$$\phi = \frac{\dot{c}_{B_4C} - \dot{b}_{B_4C}}{Y_{B_4C}}, \tag{2.18}$$

where Y_{B_4C} is estimated using simulations. The complete data reduction equation for capture yield is then given by

$$Y_\gamma = f_n \frac{(\dot{c}_\gamma - \dot{c}_{bg}) Y_{B_4C}}{\dot{c}_{B_4C} - \dot{b}_{B_4C}}. \tag{2.19}$$

The following technicalities apply to yield measurements as they do to transmission measurements. Count rates are related to counts by Eq. 2.13. Rigorously, the parameters are vectorized corresponding to TOF. Lastly, uncertainties are propagated to the composite observable, capture yield (Y_γ), via Eq. 2.10.

Unlike transmission measurements, the detector response function does not cancel out in ratio. Therefore, the various count spectra must be corrected for this in a more detailed manner than in transmission measurements. This is partially addressed by the normalization to Y_{B_4C} in the flux measurement; however, there is significant uncertainty associated with this process as Y_{B_4C} is uncertain itself. Furthermore, radiative neutron capture events often emit multiple gamma particles at varying energies, forming a gamma cascade as the excited compound nucleus decays to the ground state. This phenomenon introduces additional challenges and uncertainties in accurately correlating the observable signal with the number of neutron capture events. Further details on the detection and uncertainty quantification techniques for radiative neutron capture events are left to Refs. [25, 28, 29, 31]

2.2 Resonance Evaluation

As discussed in Sec 2.1.2, it is impossible to measure, or transform measured data to, resonance parameters directly. Instead, experimental transformations must be made to the R-matrix theory in order to compare to composite observables such as transmission or capture yield. This requires (1) an initial guess for the resonance parameters and (2) an

experimental model for each measurement. The first step in an evaluation is then to review previous evaluations and compile the relevant measurement data. For most application-relevant isotopes, a prior evaluation exists and the new evaluation is likely to have access to new/recently measured data.

The consequence of 2 is that significant effort goes into analyzing experimental databases and corresponding publications in order to develop an accurate experimental model. Ideally, the experimentalist would be involved in this process; however, this is often not the case, especially for past measurement data. Expert judgement and discretion are invaluable as the evaluator decides which data to include and how to model the experimental conditions. There must be confidence in the experimental description, the measurements, and the reported uncertainties as imperfections in the measurement data are a known challenge in experimental nuclear data [9, 26, 32, 33]. Confidence is typically established by assessing the consistency of the data with the reported uncertainties, and it can be beneficial to model the experiment using previously evaluated resonance parameters during this process. If not addressed during the data curation stage, biases in the measurement data can carry over into the evaluation, affecting both the mean and covariance estimates of the resonance parameters.

Once there is a curated collection of data to be used in the evaluation and sufficient experimental models for each, resonance parameters must be inferred from the experimental data. Currently, the default procedure for resonance parameter estimation is to use a resonance modeling and analysis code, such as SAMMY [8] or REFIT [34]. These codes do three primary things; 1) calculate a theoretical cross section using an appropriate R-matrix formalism, 2) mathematically model the experimental conditions that affect measured data, and 3) adjust the resonance parameters of the R-matrix formalism to give a better fit to observations. Step 3, adjusting of resonances parameter estimates to fit observations, is done within the Bayesian framework. This framework for parameter estimation is default in nuclear data evaluation (and many other natural sciences) as it allows for the quantification of uncertainty and incorporation of prior knowledge. The Bayesian framework and it's implementation in SAMMY is discussed further in Sec. 2.5.1. However, for the purposes of the current discussion, the SAMMY code takes two different approaches to solving the

Bayesian problem: generalized linear least squares (GLLS) and Markov chain Monte Carlo (MCMC).

There are several shortcomings to the GLLS approach when applied to resonance analysis and nuclear data evaluation in general. These shortcomings arise primarily from the neglect of considerations and/or assumptions that underlie the statistical theory on which this approach is based. These include: linearity, Gaussian probability density functions, goodness of fit, model complexity, model or data defects, and incomplete prior information. Each of these can impact the accuracy of the Bayesian posterior model. Monte Carlo approaches provide greater flexibility in Bayesian modeling, accommodating non-Gaussian distributions, discrete parameters, as well as non-linear and hierarchical relationships. Their application to nuclear data evaluation in general has become a recent focus of research [31, 35, 36]. The MCMC implementation in SAMMY is a newer feature, developed to rigorously incorporate descriptions of imperfections in measurement data [37, 38]. Generally, data that would significantly bias parameter estimates is filtered out during the data curation stage; however, the referenced work addresses smaller data imperfections that may have a minimal impact on the parameters but can lead to a significant underestimation of the resonance parameter covariance. Historically, this issue has been addressed by artificially inflating the posterior covariance, which often leads to an overestimation.

Practical implementation challenges exist for both GLLS and MCMC approaches. Neither approach as implemented in SAMMY optimizes globally, over discrete parameters, or over model complexity. This means that the evaluator must specify the number of resonances to include, assign a quantum spin group to each resonance, and initialize the continuous resonance parameters. The resulting posterior will not alter the number of resonances or the spin group assignments, and the initialized energy location parameter is unlikely to change significantly. Furthermore, the aggregate capture width in the RM formalism discussed in Sec. 2.1.1 is often set to a constant average value for all resonances of a particular quantum spin group.

The spin group assignments do influence the resonance shape and can be informed by cross-section measurement data. However, the experimentally corrected cross-section can be relatively insensitive to spin assignments, with changes often falling below the

noise level in the data. Information about resonance parameter distributions are widely used in modern evaluation techniques, but they often lack a systematic approach for incorporating this information. Some brute force approaches to spin assignment have been explored/implemented in [10], and more recently a machine learning classification model in [39].

As the evaluation moves to higher energies, the experimental resolution degrades, as described by the relationship in Eq. 2.8. This makes determining individual resonance parameters challenging, and at a point determined by the evaluator, the model is changed to Hauser-Feshbach (HF) for the unresolved resonance region (URR). As discussed in Sec. 2.1.1, the HF model treats resonances in a statistical manner and reconstructs only the changes in average behavior parameterized by average resonance widths and level spacing(s). In theory, the average parameters fitted at the beginning of the URR should align with the empirical averages of the fitted resonances in the RRR. However, a consequence of the Porter-Thomas hypothesis is that the smallest resonances are the most probable. As a result, many resonances are expected to fall below the experimental noise level, making them undetectable in most measurement data. To maintain consistency with statistical distributions and align the average parameters in the RRR with those in the URR, evaluators may introduce very small 'fake' resonances in the high-energy range of the RRR, even though they may not be supported by the measurement data. Traditionally, there is not a systematic method for the evaluator to determine the point in energy at which a statistical resonance model should be used. When a threshold reaction channel opens, it can signal the clear onset of the URR, as there may not be sufficient measurement data to support an R-matrix evaluation with the additional channel. Otherwise, several heuristics are employed, such as a deviation from the expected trend in cumulative levels (Fig. 2.1) or a prior evaluation.

2.3 Demonstration on ^{181}Ta

As mentioned in Sec. 2.1.2, the demonstration of the methodology developed in this thesis is based on the ^{181}Ta measurements done at Rensselaer Polytechnic Institute (RPI) in [28]. ^{181}Ta is a relevant isotope for stockpile stewardship, various fission and fusion reactor

applications, and for experimental purposes as it is often used as a target for production of white-spectrum neutrons. Two quantum spin groups show up in the RRR of ^{181}Ta[6], $J^{\pi} = 3^+, 4^+$. The parameter averages for each were estimated from the JEFF-3.3 [1] evaluation and are detailed in Appendix B.

The U.S. evaluation in ENDF/B-VIII.0 [2] evaluation was very out-dated, with the RRR extending only up to about 330 eV. In contrast, the newer European evaluation, JEFF-3.3, defined resolved resonances up to 2.5 keV. The Japanese evaluation, JENDL-4.0 seems to have adopted identical resonance parameters as JEFF-3.3. An updated evaluation that includes the referenced measurement data will be introduced with the release of ENDF/B-VIII.1[7]. Key aspects of the evaluation will be covered here, with full details in Ref. [40]

Several data sets beyond those in [28] were incorporated into the ENDF/B-VIII.1 evaluation. The JEFF-3.3 evaluation was used as a prior, with two additional resonances added at 1.911 and 2.4596 keV. The spin assignments, however, were not taken from JEFF-3.3. Instead, the Atlas of Neutron Resonances [7] was used and those not reported were determined by randomly sampling combinations of spin assignments and selecting the assignment that best agreed with the expected resonance parameter distributions. From this point, prior parameter uncertainties were set to 0.01%, 2%, and 10% for the energy, capture, and neutron widths, respectively, and the data was fitted sequentially using SAMMY. The authors note that the results were sensitive to the order in which data sets were fitted sequentially. Lastly, 59 small 'fake' resonances were added to the evaluation above 700 eV. This process was carried out similarly to spin assignment, where random combinations were sampled, and the set that best matched the theoretical distributions was selected. A resonance parameter covariance matrix was generated using SAMMY. Several uncertain experimental parameters were propagated to the posterior resonance parameter covariance during the sequential fitting process and the final result was scaled by the square of the averaged χ^2 measure of fit for each sequentially fit data set.

The cumulative level spacing for the three different evaluations is shown in Fig. 2.1 where the sum of the integer number of resonances is visualized against incident neutron energy.

[6]Higher-order spin groups are negligible in the RRR as the penetrability factor ($P(E)$ in Eq. 2.2) becomes very small.
[7]The associated publication is currently in the review process.

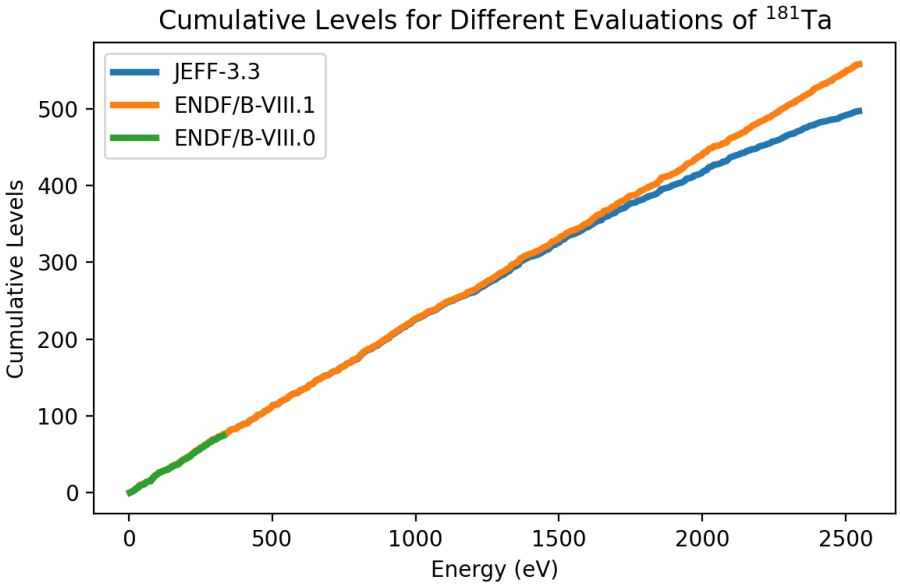

Figure 2.1: Cumulative level spacing for ENDF/B-VIII.0, JEFF-3.3, and ENDF/B-VIII.1 resonance region evaluation. The expected trend is a straight line with a slope of $\frac{1}{\langle D \rangle}$ where $\langle D \rangle$ is inferred from the fit itself.

The Wigner distribution implies that the cumulative sum should have a slope proportional to the inverse of the average level spacing, $\frac{1}{\langle D \rangle}$. In this case, the spin groups are combined, and $\langle D \rangle$ represents the average level spacing across all resonances. The use of 'fake' resonances in the ENDF/B-VIII.1 evaluation allows it to more closely follow the expected trend, even though only two additional true resonances were identified compared to JEFF-3.3.

2.4 Advanced Computational Methods in Evaluation

As computational resources expand, machine learning (ML) and other advanced computational methods have achieved notable success in various data science applications. This naturally fosters interest in integrating these techniques into the nuclear data pipeline, leading to new paradigms that address challenges such as uncertainty quantification (UQ), model deficiencies, reproducibility and automation, and database management in general.

On the experimental side, computational learning methods are being explored to optimize the design of experimental measurements (both integral and differential), thereby maximizing the benefits of the new measurement data [41, 42, 43]. Sparse inference and feature identification methods have been used to quantify unknown sources biases in experimental data [9] and relate them to specific experimental characteristics [44], thus uncovering the underlying physics causing bias. Of this type of development on the evaluation side, the TALYS project is perhaps the most relevant. TALYS is an evaluation code that is purposely structured to lend itself to an evaluation paradigm that includes automated parameter inference and opens the door to machine-learning methods [45, 46, 47].

A major theme throughout these efforts is reproducibility, which is formally addressed by the Working Party on International Nuclear Data Evaluation Co-operation (WPEC), subgroup 49—Reproducibility in Nuclear Data Evaluation [5]. Much of the recent work on advanced computational methods for nuclear data applications depends heavily on storing data in consistent and machine-readable databases. Significant efforts have been dedicated to enhancing these aspects of the experimental nuclear data bases, namely EXFOR [48, 49, 50]. Similar efforts for the storage of evaluated nuclear data have been made in the GNDS project [21, 51, 52]. GNDS is the new format for evaluated nuclear data, replacing the

ENDF-6 format [53], and has been expanded to include processed nuclear data for use in simulations.

The aforementioned efforts apply to the nuclear data pipeline as a whole. Major developments in machine learning (ML) and automation for evaluation, primarily associated with the TALYS project, focus on high-energy reaction evaluations. The RRR is largely left to analysis codes such as SAMMY [8] that remain manual and incremental in their approach to parameter inference. Efforts specifically in the RRR include the aforementioned ML classification of spin assignments [39] which has the potential to be combined with the automated fitting methodology developed in this thesis as demonstrated in [54]. Also, some preliminary and proof-of-concept work related to this thesis is presented in Refs. [55, 56].

The work in Ref. [57] proposes a ML-augmented evaluation pipeline for the RRR. Two ML regression models, k-nearest neighbors (KNN) and decision tree (DT), were trained on differential experimental data and validated against critical benchmarks. The major novelty of this work lies in training the model on multiple isotopes and energy regimes while incorporating additional features, such as atomic mass and neutron strength functions. This approach potentially allows for capturing trends across isotopes and energy regimes, a phenomenon that is only loosely understood through first-principles models. However, both machine learning models are entirely empirical. The omission of first-principles elements in the models likely reduces their accuracy and precision, and it definitely limits their applicability and interpretability.

2.5 Inferential Regression

Inferential regression is a statistical learning technique used to infer parameters from observational data. There are elements of Bayesian statistics, uncertainty, and optimization. This section will explore how these are related to nuclear data evaluation and, in particular, resonance parameter inference in the popular resonance analysis code SAMMY code [8].

2.5.1 Conjugate Bayes, GLLS, and SAMMY

In many areas of scientific analysis, observable data, D, is used to make inferences about the world. Those inferences are often tied to some theory or theoretical model, M, that allow scientists to characterize their inferences with a set of parameters; $M(\Theta)$. This is a useful practice as $M(\Theta)$ allows for interpretation and prediction. The Bayesian framework allows for the model and parameters to be represented as random variables with probability density functions (PDFs) that describe confidence in their values. Furthermore, it gives systematic procedure for including prior knowledge of beliefs about the likelihood of M and Θ. The following derivations are based Refs. [58, 59]. Bayes theorem is as follows:

$$P(\Theta|D,M) = \frac{P(D|\Theta,M)P(\Theta|M)}{P(D|M)} \tag{2.20}$$

For the purposes of this discussion, the model M will be considered determined, though there are cases where it may be useful to consider the probability of different models. The left-hand side, $P(\Theta|D,M)$, is what we are after. Often called the *a-posteriori* (posterior) PDF, this describes the updated probability of parameters, Θ, given the new data, D. The term $P(\Theta|M)$ represents the prior information about the parameters and is often called the *a-priori* (prior) PDF. The prior PDF (or knowledge) must be independent of the new data, D. The term $P(D|\Theta,M)$ describes the likelihood of observing the data, D, given determined parameters, Θ. The term in the denominator normalizes the output PDF, or posterior, to the quality of the model and is often called the evidence. This term quantifies how well the model, M, represents the data marginalized over all possible parameters, Θ, and is mathematically described as:

$$P(D|M) = \int P(D|\Theta,M)P(\Theta|M)d\Theta \tag{2.21}$$

Implementing shorthand notation for each distribution one can write the following for the posterior distribution:

$$P(\Theta|D,M) \equiv P(\Theta|D) = \frac{P(D|\Theta,M)P(\Theta|M)}{\int P(D|\Theta,M)P(\Theta|M)d\Theta} \equiv \frac{L(D|\Theta)\pi(\Theta)}{\int L(D|\Theta)\pi(\Theta)} \tag{2.22}$$

The posterior will be some middle ground between the prior and the likelihood, each weighted by their variance. From here, the posterior PDF representation of Θ can be used to make estimates, infer expectation values, and quantify uncertainty in parameters of interest.

Despite the simplicity of Bayes theorem and Eq. 2.22, mathematical solutions can be difficult as the PDFs can be complicated, multivariate functions that are either discrete or continuous. Brute force numerical solutions become infeasible with many parameters. Monte Carlo approaches exist and have found significant success. Also, there are a select few cases where analytic solutions exist. These cases leverage conjugate distributions as the prior and likelihood such that the posterior distribution can be derived in a closed form. Perhaps the most convenient example of this is the normal, or multivariate normal distributions. This family of distributions is conjugate with itself, meaning that if both the prior and likelihood are normal, the posterior will be a normal as well.

The Bayesian framework is very general, allowing for a wide range of applications. Working in constraints and assumptions, its relationship to well known regression and optimization techniques can be uncovered. The resonance analysis code SAMMY leverages assumptions of normal conjugate distributions and linearity (via a first order Taylor expansion) to derive the following, closed form matrix solution to Eq. 2.22.

$$M' = (M^{-1} + G^T V^{-1} G)^{-1} \tag{2.23}$$

$$P' = P + M' G^T V^{-1} (D - T) \tag{2.24}$$

The vectors $P, T,$ and P' and the matrices $M, V,$ and M' parameterize normal PDFs for the prior, likelihood, and posterior as follows:

$$\begin{aligned} \pi(\Theta) &\sim N(P, M) \\ L(D|\Theta) &\sim N(T, V) \\ P(\Theta|D, M) &\sim N(P', M') \end{aligned} \tag{2.25}$$

The matrix G is the Jacobian describing the relationship between T and P via a first order Taylor expansion. If the models were linear, standard notation would replace G with X and

$T = XP$. Lastly, D and V are the observed data and estimated data covariance respectively. Because of the approximate, first-order relationship between P and T, the solution's accuracy is dependent on the linearity in the space $P' - P$ and $D \pm V$. In attempt to improve accuracy, SAMMY has implemented internal iterations on the evaluation of G at intermediate values between P and P'. Ultimately, if the prior model is too far from the data then these equations will not converge.

2.5.2 Variable Selection

When inferring a model from empirical data, a common challenge is determining which explanatory variables are necessary. Having too many or not enough explanatory variables will impact bias and variance in the inference and often decrease predictive performance [60]. Furthermore, in natural science applications, explanatory variables often have a physical meaning and the goal of inference is to best describe the true underlying process. Appropriate variable selection in this case is vital for interpretability of results and the following physical conclusions.

One straightforward approach to variable selection is subset regression. This approach tries all possible combinations of models (subsets) from the complete set of explanatory variables (p). For each level containing $k = 1, 2, ...p$, explanatory variables, the best performing model, determined by a measure of fit such as χ^2, is identified. This approach is exhaustive and guaranteed to find the optimal model for each level of complexity; however, it can become infeasible for large models as the search space scales as 2^p. Stepwise selection is an approach that significantly narrows this search space by searching down a specific trajectory. In forward stepwise selection, all possible $k = 1$ models are fit and the explanatory variable that gives the most improvement determines the trajectory, i.e., the search space for the $k = 2$ models are restricted to those that contain the explanatory variable determined by the previous step. Backwards stepwise selection is similar, however, it starts at $k = p$ explanatory variables and iteratively removes the least impactful variable. In both cases, the search space scales by $1 + p(p + 1)/2$. While these approaches are not exhaustive—there is no guarantee that the best model for each k level of complexity is found—the guided search is much more computationally feasible for large models and tends to do well in practice [61].

2.5.3 Model Selection

In the variable selection approaches above, the goal is to find the optimal model for each level of complexity. The next step, selecting model complexity, is not as straightforward. The χ^2 measure of fit is expected to monotonically improve as model complexity increases; however, prediction error will begin to increase if the model is over-fit to the training data.

Information-theoretic criteria can be used to strike an appropriate balance between the goodness of fit and model complexity for the observed data. Some popular criteria for maximum likelihood or least squares regression problems are the Akaike information criterion (AIC) [62, 63] and Bayesian information criterion (BIC) [64].

Both AIC and BIC aim to approximate the "true" model with minimal information loss as conceptualized by the Kullback–Leibler (KL) distance — a measure of the statistical distance between models — with key differences discussed in [65, 66, 67, 68, 69]. To summarize, AIC aims to minimize prediction error, focusing on the accuracy of future predictions, while BIC seeks to identify the best model for explaining the underlying process that generated the data. BIC, which is based on Bayesian probability and inference, will generally place a greater penalty on model complexity, favoring simpler models, particularly when the sample size is large.

AIC and BIC are calculated using Eqs. 2.26 and 2.27 respectively, where p is the number of free model parameters, n is the number of observations, and $\widehat{L}$ is the maximized likelihood value for the model.

$$\mathrm{AIC} = 2p - 2\ln(\widehat{L}) \tag{2.26}$$

$$\mathrm{BIC} = p\ln(n) - 2\ln(\widehat{L}) \tag{2.27}$$

An additional term is added to the AIC criteria to adjust for small sample sizes giving the corrected AIC, or $\mathrm{AIC_c}$:

$$\mathrm{AIC_c} = 2p - 2\ln(\widehat{L}) + \frac{2p^2 + 2p}{n - p - 1}. \tag{2.28}$$

As can be seen from these equations, the first term will grow as more parameters are added to the model while the second term should monotonically decrease (assuming you have found

the optimal model for complexity p).This is the tradeoff between under-fitting and over-fitting that is quantified by these criteria. The model with the smallest criterion value is selected. Often criterion values will be reported in difference to minimum value, translating the selected model's criterion to a value of 0.0.

The statistical derivation of both AIC and BIC depend on p as an accurate estimate of model complexity. For linear models, this corresponds directly to the number of parameters; however, this is not the case for nonlinear models and instead p should be replaced with a better measure of model complexity [70]. For more complex/general models, estimating an effective number of parameters is not always reliable and these criteria may not give the best model selection. An alternative approach is to select the model complexity that minimizes prediction error as estimated by cross-validation (CV). Performing CV is more computationally intensive than calculating AIC or BIC; however, it relies on fewer assumptions about the model or data making it more general.

CV directly estimates prediction error using a testing subset of the data that is independent of the training set. When data is limited, k-fold CV can be employed to improve error estimates by iteratively splitting different parts of the data into training and testing sets, repeating the process k times. Each fold will give a score for the train and test set, generally given by a fit measure such as χ^2. The error estimates are averaged over the k folds, resulting in some mean and standard error. The simplest model with CV error within one standard error of the minimum is selected, a criterion known as the one-standard error rule [70]. The case $k = N - 1$, where N is the number of independent data points, is called leave-one-out-cross-validation (LOOCV) and gives the most unbiased estimate of generalization error; however, this can be computationally expensive depending on N and the cost of training the model. Generally, 5 or 10 folds are used, as results are not expected to significantly improve beyond this point. For additional information about this method, see reference [70].

2.5.4 Relating Bayes to Regression Techniques

As discussed in Sec. 2.5.1, the Bayesian framework is very general. The linear, Gaussian version of Bayes commonly used to estimate resonance parameters (given by Eqs. 2.23

and 2.24) is closely related to several other regression techniques. This section will expose these relationships in the continued notation of the previous section (given by the SAMMY code [8]).

Relating Conjugate Bayes and Non-Linear Regression

Consider a truly linear model with a noise term given by ϵ. The noise around the model is normally distributed with finite variance (note that an equivalent assumption is made to get to Eqs. 2.23 and 2.24). In this case, the maximum likelihood estimate (MLE) of the parameters P is given by minimizing the squared Mahalanobis distance between the model and observations D. The squared Mahalanobis length will be distributed as a chi-squared distribution with degrees of freedom equal to the dimension of the multivariate normal distribution describing ϵ. This squared distance is commonly referred to as the χ^2 statistic or goodness of fit and will be through the rest of this thesis.

$$\begin{aligned} D &= GP_{\text{true}} + \epsilon \\ \epsilon &\sim N(0, V) \end{aligned} \tag{2.29}$$

$$\chi^2 = (D - GP)^T V^{-1} (D - GP) \tag{2.30}$$

For the truly linear model, this optimization problem falls into the class of quadratic programming and has a closed form solution:

$$P_{\text{MLE}} = \arg\min_P \ \chi^2 \equiv (G^T V^{-1} G)^{-1} G^T V^{-1} (D - GP) \tag{2.31}$$

often called generalized least-squares (GLS). This a very general form, reducing to weighted least squares (WLS, ϵ is heteroskedastic but uncorrelated, $V_{i \neq j} = 0$) or ordinary least squares (OLS, ϵ is homoskedastic and uncorrelated, $V \propto \mathbf{I}$)[8].

[8]The GLS objective can be expressed identically as a weighted or ordinary least squares problem by transforming the residual vectors onto the linearly independent components of V. Because it is always a square symmetric matrix, V can be expressed by its eigenvector decomposition, $Q \Sigma Q^T$. Transforming the residuals using Q^T or $\Sigma^{1/2} Q^T$ yields a WLS or OLS objective whose minimum in the transformed basis corresponds to the minimum of the GLS objective in the original basis.

Notice the similar forms of Eqs. 2.31 and 2.24. If your model is linear – and therefore the χ^2 objective function is quadratic – Eq. 2.31 gives a closed form, least-squares solution for P. If the data and parameter uncertainties are distributed as Gaussian or multivariate normal, then the least-squares solution corresponds to the maximum likelihood estimate for P and the Bayesian posterior in the asymptotic limit of un-informative prior knowledge of the parameters (i.e., $M^{-1} = 0$).

In the case that the model is not linear with respect to the parameters, then the Jacobian matrix G becomes dependent on P and you lose access to a closed form solution. The local behavior of the non-linear response function $T(P)$, however, can be approximated via a first order Taylor expansion in some range of finite perturbation ΔP. This allows computable solutions to be had within finite perturbations that are carried out iteratively. The following derivations in this section are based on Ref. [71]. The derivative of the objective function (Eq. 2.30) is given by:

$$\begin{aligned} \frac{d\chi^2}{dP} &= 2(D-T)^{\mathrm{T}} V^{-1} \frac{d}{dP}(D-T) \\ &= -2(D-T)^{\mathrm{T}} V^{-1} \frac{dT}{dP} \\ &= -2(D-T)^{\mathrm{T}} V^{-1} G. \end{aligned} \tag{2.32}$$

One approach to solving this optimization problem may be to simply step down this gradient in an iterative fashion, (i.e., the gradient descent algorithm or one of it's many variations). In all following equations, G is evaluated at P^i. For this example, one gradient descent step would look like,

$$P^{i+1} = P^i + \alpha(D - T^i)^{\mathrm{T}} V^{-1} G, \tag{2.33}$$

with step size controlled by α. Alternatively, the Gauss-Newton (GN) method, shown in Eq. 2.34, rotates the gradient by $(G^{\mathrm{T}} V^{-1} G)^{-1}$ rather than setting a step size:

$$P^{i+1} = P^i + (G^{\mathrm{T}} V^{-1} G)^{-1}\ G^{\mathrm{T}} V^{-1} (D - T^i). \tag{2.34}$$

This is an approximation to the inverse of the Hessian and improves the speed of convergence particularly when the function is quadratic in ΔP. The Levenberg-Marquardt (LM) algorithm combines the GN algorithm with gradient descent by introducing a dampening

parameter, λ, that dynamically varies the weight of the GN rotation and step size:

$$P^{i+1} = P^i + (G^{\mathrm{T}} V^{-1} G + \lambda \mathbf{I})^{-1} \, G^{\mathrm{T}} V^{-1} (D - T^i). \tag{2.35}$$

If the λ is small, the step resembles a GN iteration. Conversely, if λ is large, the step resembles gradient descent. The general prescription is for the dampening parameter to start large and decrease as the objective function improves; gradient descent steps continuously move towards GN iterations as a local minimum is approached. Several variations on this algorithm exist, some replacing $\mathbf{I}$ with a more detailed weight.

The functional form here is identical to the GLLS equations (2.24) implemented in resonance analysis codes; however, the interpretation is different. M and P no longer representing prior knowledge; rather, they represent the previous location and step in an iterative optimization algorithm. Mathematically, setting $M^{-1} = \lambda \mathbf{I}$ makes Eqs. 2.24 and 2.35 exactly equivalent.

Chapter 3

Synthetic Data

3.1 Methodology

This section outlines the methodology used to generate synthetic resonance cross-section measurement data that is statistically indistinguishable from actual experimental data. The goal is to be able to produce unlimited, labeled data for use in the development and optimization of an automated evaluation algorithm. Labeled, in this case, means the solution is accessible and can be used to make decisions. Additionally, once the automated evaluation model has been optimized, its performance can be characterized and benchmarked. With the assumption that the synthetic data is sufficiently representative of real data, then benchmarks can be extrapolated to evaluations of real data. This will provide a unique capability to rigorously test and quantify the impact of evaluation decisions that have historically been based on assumption or experience.

Generating synthetic data is a common practice for many ML or statistical applications (e.g., healthcare analysis [72, 73], computer vision [74], software development [75]). The goal for many of these applications is to synthesize data that mimics the statistical properties of the observed data. The utility of the synthesized data is a measure of how well the statistical properties of the synthetic data match that of the observed data. A higher utility results in a better-trained model, a more accurate performance assessment, and higher confidence when applying trained methods to real data sets. The general process is to develop a generative

model for the observed data and then statistically sample realizations from that model to create a synthesized data set, primary steps are outlined below.

1. Assume some joint generative distribution, parameterized by θ, from which both observed data, X, and synthetic data, Y, are generated:

$$X, Y \sim f(\theta). \tag{3.1}$$

2. Estimate θ to approximate a distribution for Y

$$\hat{\theta} \doteq \theta, \tag{3.2}$$

3. then sample from the approximated generative distribution:

$$f(Y|X, \hat{\theta}). \tag{3.3}$$

4. The utility of the synthesized data describes how well the underlying distribution is approximated:

$$f(Y|X, \hat{\theta}) \sim f(\theta). \tag{3.4}$$

Step 2 is the most application-specific; there are many ways to attempt to characterize statistical distributions of observed data. The model and its parameters, $f(\theta)$, can be inferred using ML methods with no underlying assumptions or it can be completely physics-informed, not reliant on any observed data. Integration of a physics-based model can greatly improve utility while reducing the need for copious amounts of observed, labeled data. This method has been employed for manufacturing processes [76, 77], biological experiment design [78], and nuclear non-proliferation applications [79, 80]. When using a well-established physics-based model, step 3 often involves uncertainty sampling and/or noise sampling. If a process is accurately described by a physical model, the statistical variation in the data must be driven by uncertainties in the input parameters or noise in the observables.

To the best of the author's knowledge, a high-utility synthetic data approach has not been used to assist in developing methods for nuclear data evaluation. The most relevant

study is found in Ref. [35], where researchers used simulated experimental cross-section data to verify and compare classical and Monte Carlo (MC) Bayesian methods for uncertainty estimation. The 'fake' experimental data is not thoroughly explained; it was likely generated using a simple Gaussian noise model.

3.1.1 Physics Informed Generative Modeling

The generative model can be separated into two major parts. The first is a generative model for the resonance parameters. The second is a generative model for the measurement process, meaning the process by which a given set of determined resonance parameters generates observable data. These two models are independent, meaning that a different generative model for the measurement process could be used with the resonance parameter generative model, and vice versa.

The data flow diagram in Fig. 3.1 shows a high level overview of the data and process flow for this generation methodology. Generative modeling occurs moving right to left while reductive modeling occurs moving left to right. To draw samples from the generative model, uncertain parameters are randomly sampled with respect to a user-defined uncertainty distribution. When possible, established methods and/or software are used. This is delineated in the figure; processes that leverage established methods (and software) are shown in blue while processes that involve a novel method (and software) are shown in orange. The data-flow that falls within the gray border represents the current process for experimental measurement and evaluation.

In an evaluation, the composite observable data (also known as reduced measurement data) make up the regression dataset used to infer the unknown latent parameters. Notice that this object is neither directly measured nor directly calculated from theoretical resonance parameters. Rather, the data reduction process is applied to transform the raw observable data to a composite observable (*e.g.,* transmission or capture yield). As discussed in Sec. 2.1.2, measured data in the resonance region cannot be 'reduced' back to the theoretical parameters. Therefore, the mathematical models for the physics theory and the experimental conditions must be applied to resonance parameters to get some estimated composite

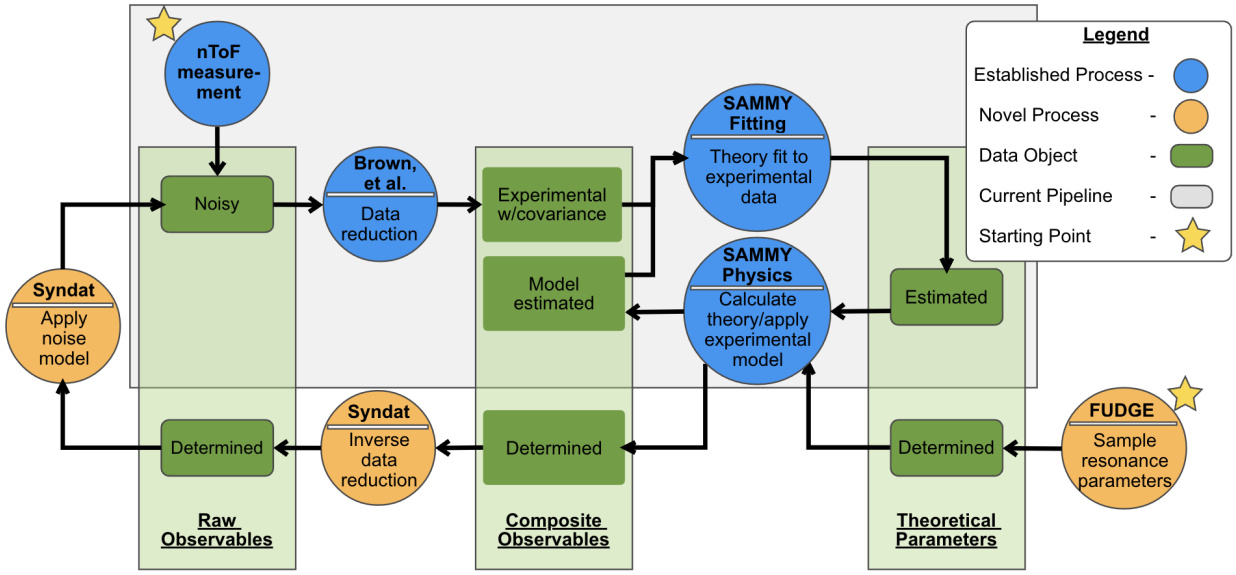

Figure 3.1: Data flow diagram for the data synthesis methodology. Circles represent processes, rectangles represent data. The gray border encompasses the current processes for experimental measurement and evaluation, within which established methodologies and software can be leveraged. The processes outside this border represent the novel methods developed in this thesis. Data objects that are part of the data generation flow are labeled determined because the true/exact value is known.

observable for comparison. In this way, the SAMMY code has an inherent generative model to facilitate comparison to composite observables.

The novelty of the data generation methodology can be seen in the data flow diagram in orange. The first generative model for resonance parameter sampling is based on the statistical theory of resonances. This generative model has been developed in the processing code FUDGE [21] to train ML models to characterize quantum spin group assignment [39] and is briefly discussed in Sec. 3.1.2. The second generative model utilizes established methods to compute the expected composite observables from a set of resonance and experimental parameters, along with a novel method for inverse data reduction to generate raw observables. This model is discussed in Sec. 3.1.3. It is implemented using the SAMMY code developed at Oak Ridge National Laboratory (ORNL) [8] and an open-source python code, Syndat, which was developed for this thesis [81].

The generative model for raw observable data (generative model two) can be highly specific to the experimental parameters and procedures. Currently, the methodology is developed and implemented for transmission and capture yield measurements with experimental methodologies corresponding to those in Refs. [28, 82] summarized in Sec. 2.1.2. The extension of this methodology to include other reaction measurements or experimental methodologies could be considered in future work. While a general set of experimental parameters and procedures may be sufficient, the synthesized data will be more representative of actual data—it will be higher utility—when more detailed experimental information is included. Consequently, the resulting studies and conclusions will be more meaningful and reliable.

3.1.2 Generative Model 1: Resonance Parameters

This section describes the methodology for sampling a set of resonance parameters corresponding to a theoretical cross-section. This will be the 'solution' to the synthesized experimental data. The concept of sampling resonance parameters is supported by the statistical theory of neutron resonances. Resonance structures in neutron cross-sections correspond to excited states of the compound nucleus formed when an incident neutron is absorbed and equilibrates in a target nucleus. Each of these levels belongs to a discrete

quantum spin group, characterized by the orbital angular momentum and parity of the particle-pair. For any given spin group, resonance parameters are observed to fall on well-characterized statistical distributions. These statistical distributions provide the basis for a generative probability model.

Resonance structures are parameterized by a location in energy and a total width made up of two or more constituents. For any given spin group, the width parameters tend to fall on the Porter-Thomas distribution, a χ^2 distribution with the number of degrees of freedom equal to the number of open channels. Similarly, the energy spacing between neighboring resonances of the same spin group tend to fall on the Wigner distribution[1]. This distribution can be derived from random matrix theory, particularly the Gaussian Orthogonal Ensemble (GOE), and our current understanding of quantum chaotic systems. The probability distributions for resonance widths and spacings were first introduced in [17] and [18] respectively, but are widely used in modern evaluation techniques [6, 7]. The applicability of these statistical distributions is further supported by their use in representing the unresolved resonance range (URR) in continuous energy Monte Carlo transport calculations. The onset of the URR occurs where individual resonance structures cannot be fully resolved experimentally but still cause large variance in the mean cross section. This energy region is evaluated as an average cross section. When accessing this cross section for Monte Carlo neutron transport simulations, however, probability tables can be used to better represent this energy regime. Probability tables are constructed from realizations of resonance parameter sequences sampled from the Porter Thomas and Wigner distributions. This is done in a number of processing codes [19, 20, 21]. This work leverages the high-fidelity implementation developed by Brookhaven National Laboratory (BNL) [22] distributed with the open source processing code FUDGE [21] developed by Lawrence Livermore National Laboratory (LLNL). Average resonance parameters are an input to this generative methodology and can be estimated from compilations such as the Atlas of Neutron Resonances [7] or analysis of current evaluation libraries such as ENDF/B [2].

[1]Properly, the distribution is given by GOE and the Wigner distribution is a simplification

3.1.3 Generative Model 2: Measurement Process

This section describes the methodology for modeling the process by which a set of theoretical resonance parameters produce raw observable data. These are detector count data, for which there is a known noise model (Poisson counting statistics). It is assumed that the generative model in Sec. 3.1.2 has been used to sample a realization of resonance parameters that are taken as a determined input[2]. In other words, this generative model assumes access to an exact theoretical cross section from which realizations of corresponding raw observable data can be generated.

The implementation of this generative model is divided into two components: (a) the theoretical and experimental models used to compute the expected composite observables, implemented in SAMMY, and (b) the inverse data reduction, implemented in the Syndat code. The first component (a) is a well established forward modeling methodology that is a routine part of any modern resonance evaluation. It is discussed in Sec. 2.2 with more information and defined inputs for SAMMY in Ref. [8]. For the generative process, any of the uncertain input parameters are randomly sampled with respect to user-defined distributions. The novelty lies in the interpretation and use of the SAMMY modeling code. The second generative component, which will be discussed in detail in the following two sections, introduces a more novel approach.

The two components of this generative model yield raw observable data. As aforementioned, generally the composite observable data is used for evaluation. Following the generative process, the synthetic data realizations are then reduced back to the composite observable via a standard data reduction methodology. If both the methodology and the parameter values/uncertainties in the reductive model align with those of the generative model, the measurement uncertainty quantification will be statistically consistent. However, by deliberately introducing flaws or approximations into the reductive model, their impact can be quantitatively assessed. For instance, the quantitative impact of omitting correlations in measurement data on the estimation of resonance parameters is analyzed in Sec. 3.3.3.

[2]As mentioned before, a different generative model could be used as the two are independent. A relevant example would be the extension of this framework to high energy evaluations where the first generative model changes but the second could remain the same.

RPI Transmission Measurement Model

If the composite observable of interest is transmission, the true underlying transmission is calculated from the sampled resonance and experimental parameters using SAMMY [8]. This expected composite observable is then used to calculate the corresponding expected raw observable data by inverting the data reduction process described in 2.1.2. This is shown in Eqs. 3.5 and 3.6, where, with a slight abuse of traditional notation, the bar denotes the determined true underlying value of a parameter:

$$\bar{\dot{c}}_s = \frac{\bar{T}(\bar{\alpha}_3\bar{\dot{c}}_\mathrm{o} - \bar{\alpha}_4\bar{k}_\mathrm{o}\bar{\dot{B}} - \bar{\dot{B}}0_\mathrm{o}) + \bar{\alpha}_2\bar{k}_\mathrm{s}\bar{\dot{B}} + \bar{\dot{B}}0_\mathrm{s}}{\bar{\alpha}_1}, \tag{3.5}$$

$$\bar{c}_s = \bar{\dot{c}}_s b_\mathrm{w}\mathrm{trig}\ . \tag{3.6}$$

Simulated true underlying values (indicated with bars) are randomly sampled from estimated uncertainty distributions, for example,

$$\bar{k}_\mathrm{o} \sim N(\bar{\bar{k}}_\mathrm{o}, \delta\bar{\bar{k}}_\mathrm{o}^2), \tag{3.7}$$

where the double bar indicates the population mean ($\bar{\bar{k}}_\mathrm{o}$) and variance ($\delta\bar{\bar{k}}_\mathrm{o}^2$) used to generate samples of the true underlying parameter for the generative model. This is done for all uncertainty parameters in the data reduction model detailed in Sec. 2.1.2. For a sampled realization of true underlying parameters—including resonance parameters that enter through $\bar{T}$ and the open count spectra that enters through $\bar{c}_o$[3]—the calculated raw observable $\bar{c}_s$ is also determined. Here we can confidently apply a noise model because radiation counting is an inherently stochastic process governed by Poisson counting statistics [25]:

$$c_s \sim P(\lambda = \bar{c}_s). \tag{3.8}$$

This is the end of the generative model as we have an approach to sample noisy, statistical realizations of raw observables in a neutron time-of-flight experiment. From here, the data reduction process in Sec. 2.1.2 is used to transform the synthetic raw observable data (c_s)

[3]See appendix for discussion on open count spectrum

into the composite observable data that is actually used in an evaluation. The result being a noisy realization of experimental transmission data (T) with an exact known solution ($\bar{T}$, which also corresponds to an exact set of resonance parameters). As discussed further in Sec. 3.2.1, if the reduction model is consistent with the generative model in parameter mean/variance estimate ($\hat{x} = \bar{\bar{x}}$ and $\hat{\delta x^2} = \bar{\delta x^2}$), then the uncertainty quantification will be statistically consistent with the noise for (T).

Two technicalities are briefly noted here. First, the expectation values for $\bar{c}_s$ are calculated using the integral of the cross-section (or transmission) over the represented time bin. This ensures that fine-structures in the cross-section will influence the data even if the structures are smaller than the bin-width of the reduced data. Second, the reference experiment and data reduction [28] applies a monitor normalization and dead-time correction at each cycle. To mimic this, the expected count rate is split into multiple cycles (35 in this case) before applying the noise model. Then, each cycle-separated subset is normalized to a monitor and dead time correction term that are sampled based on the reported value and uncertainty before applying the Poisson noise model. From there, each noisy, normalized set of data is recombined to get raw count data.

One could envision expanding the generative model to incorporate particle detection through ionization or scintillation, as well as signal amplification, shaping, and analysis. This could be modularized into a third generative model representing the detection system, where the $\bar{c}_s$ object is input to calculate another true realization of a detection system parameter, $\bar{x}$. This goes beyond the scope of this thesis, but will be referred to as the generative model 'stopping point'. This is an important concept, because the utility of the synthetic data is dependent on this stopping point. That is, modes of variance in actual data that are caused beyond the stopping point will not be reproduced in the synthetic data. Furthermore, the investigation of bias due to assumptions or mischaracterized uncertainties is not feasible for those assumptions or uncertainties that exist beyond the stopping point. Ultimately, for transmission measurements the current stopping point was deemed to be sufficient because (1) it is generally assumed that potential issues in the characterization of the detector system cancel out in the ratio and therefore can be neglected without impacting utility, and (2) the validation studies in Sec. 3.2 further support this intuition.

RPI Capture Yield Measurement Model

This section takes on the same methodology as in Sec. 3.1.3 except the composite observable of interest is capture yield. Inverting the capture yield reduction in Eq. 2.19 and again using bars to indicate true underlying parameters gives:

$$\bar{\dot{c}}_\gamma = \frac{\bar{Y}_\gamma(\bar{\dot{c}}_{\mathrm{B_4C}} - \bar{\dot{b}}_{\mathrm{B_4C}})}{\bar{f}_n \bar{Y}_{\mathrm{B_4C}}} + \bar{\dot{c}}_{bg}, \tag{3.9}$$

where the true underlying values (indicated with bars) are simulated by drawing random samples from estimated uncertainty distributions. Converting to counts

$$\bar{c}_\gamma = \bar{\dot{c}}_\gamma(b_{\mathrm{w}}\,\mathrm{trig}), \tag{3.10}$$

and sampling noise

$$c_\gamma \sim P(\lambda = \bar{c}_\gamma), \tag{3.11}$$

the generative capture yield model is complete up to the same stopping point as the transmission model. Unlike the transmission model, setting the the stopping point at detector counts for capture yield is likely to significantly reduce the utility. This is because the characterization of the detector system does not cancel out and therefore introduces additional modes of variance into c_γ beyond Eq. 3.11. Expanding the generative model for capture yield measurements is beyond the scope of this thesis; however, it represents a promising avenue for future work. Investigating and improving uncertainty quantification for the detection system in capture yield measurements remains an open area of research in the field.

3.2 Verification & Validation

The goal of the following sections is to give confidence in synthetic data realizations generated using the methodology developed in this thesis. Section 3.2.1 aims to verify the implementation by demonstrating that the generative and reductive models are self-consistent under specifically designed conditions. Section 3.2.2 aims to validate the

methodology by demonstrating that the synthetic data is of high utility, meaning it is statistically indistinguishable from actual measured data.

3.2.1 Verification of the Implementation

For a fixed set of resonance parameters (fixed theory), repeated samples drawn from the second generative model produce an empirical distribution that characterizes the complete noise model. This noise model can be viewed as a hierarchical distribution, comprising multiple normal distributions for measurement parameters and various nonlinear functions that relate them. Given that (1) parameter distributions used for data reduction—including uncertainty propagation—are consistent with those used for the generative model, (2) all distributions are normal with finite variance, and (3) parameters uncertainties δ propagated through non-linear functions are small enough to be within the linear regime[4], then the first order linear propagation of error detailed in Sec. 2.1.2 should accurately describe the noise model for the composite observable. That is, if

$$\sigma_{\text{exp}} = \sigma_{\text{theo}} + \epsilon, \tag{3.12}$$

then ϵ is accurately estimated by the uncertainty propagation as $N(0, V)$, where σ_{exp} is a vector of noisy data, σ_{theo} is a vector of the true theory given parameters, and V is the data covariance. This can be verified empirically by looking at the residual over many samples:

$$r = \sigma_{\text{exp}} - \sigma_{\text{theo}} \equiv \epsilon \sim N(0, V). \tag{3.13}$$

The χ^2 statistic over many samples should follow a χ^2_k distribution with degrees of freedom (k) equal to the number of data points (N_{D}) or length of the residual vector:

$$r^{\text{T}} V^{-1} r \equiv \chi^2 \sim \chi^2_{\text{D}}. \tag{3.14}$$

[4]The linear regime refers to a domain, δ, within which a nonlinear function can be accurately approximated by its first-order Taylor expansion.

If the correlations can be neglected, normalizing each residual by the variance should give a standard normal,

$$\frac{r}{\sqrt{\text{diag}(V)}} \sim N(0,1). \tag{3.15}$$

These properties are well known in statistical regression; nonetheless, the study validates the first-order linear approach to uncertainty quantification under the assumption that the listed conditions hold. This verification was conducted for the transmission models (with both background functions) and the capture yield model detailed in Sec. 3.3.1 to ensure that the mathematical implementation is correct. For all models: (1) the uncertainty on the data reduction parameters was reduced by two orders of magnitude, and the LINAC triggers for all detector count vectors were set to 10^8 to ensure linearity; (2) ten σ_{theo} values were randomly sampled between 0 and 1 (the full range of transmission or yield), and 10 corresponding energy points were sampled between 0 and 3000 (the full energy domain of the RRR); and (3) 1500 realizations were generated.

Empirical results are compared to expected probability density functions in Figs. 3.2–3.4. Table 3.1 shows mean of the residual (expected to be 0) and p-values for statistical significance tests comparing the empirical distributions to the expected parametric distributions. The normalized residual (NR) given by Eq. 3.15 is tested against a standard normal distribution using D'Agostino's K-squared test [83, 84]. The distribution of χ^2 statistics, or goodness of fit, given by Eq. 3.14 is tested against a χ^2_k distribution with degrees of freedom $k = 10$ using a one-sample Kolmogorov-Smirnov (KS) test [85]. In all cases, the empirical results visually agree with the expected distributions, and the statistical tests support—or, more rigorously, fail to reject—the null hypothesis that the population distributions underlying the empirical results are identical to the expected distributions they were tested against.

3.2.2 Validation of the Methodology

The utility of the first generative model for resonance parameter generation relies on the accuracy of prior knowledge regarding the average parameters and the validity of the theoretical distributions. Average parameters are inferred from evaluations, making them

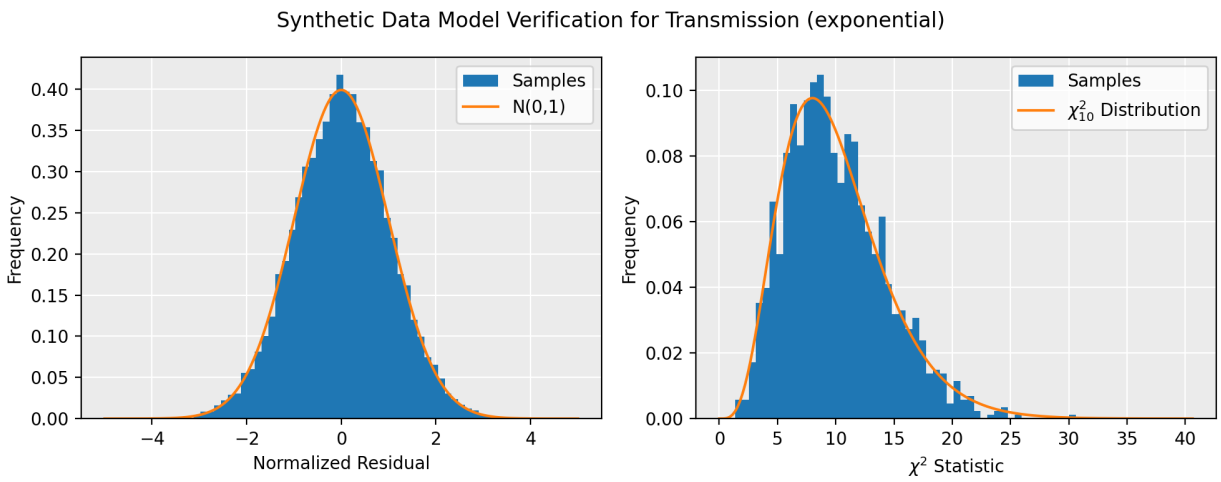

Figure 3.2: Verification of the generative model for transmission measurements with an exponential background function.

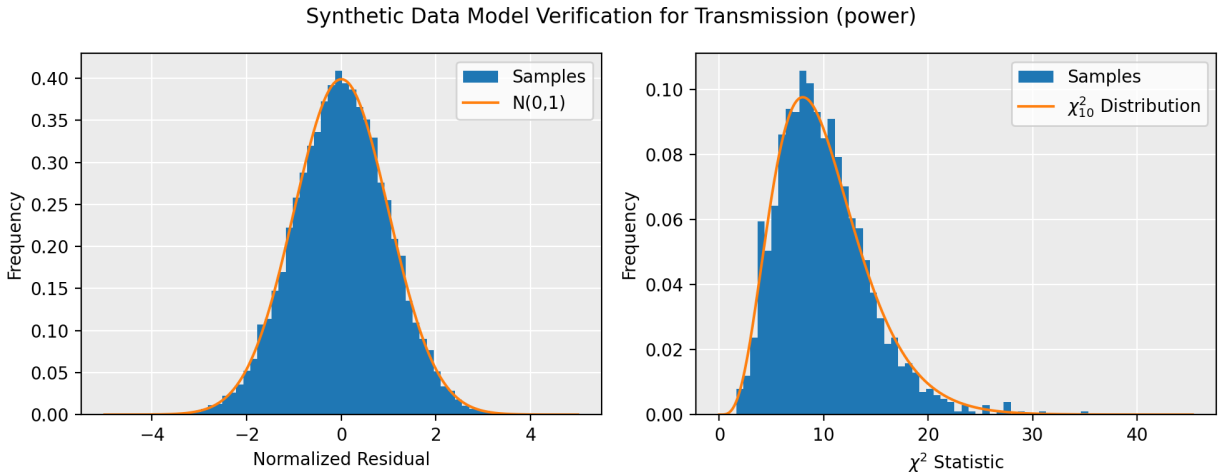

Figure 3.3: Verification of the generative model for transmission measurements with a power background function.

Table 3.1: Mean value of the residual and statistical significance tests for the normalized residuals (NR, given by Eq. 3.15) and χ^2 goodness of fit statistics (given by Eq. 3.14).

Model	Mean Residual	NR p-value	χ^2 p-value
Transmission (exponential)	6.2e-5	0.82	0.51
Transmission (power)	-3.5e-4	0.68	0.39
Yield	1.7e-3	0.43	0.50

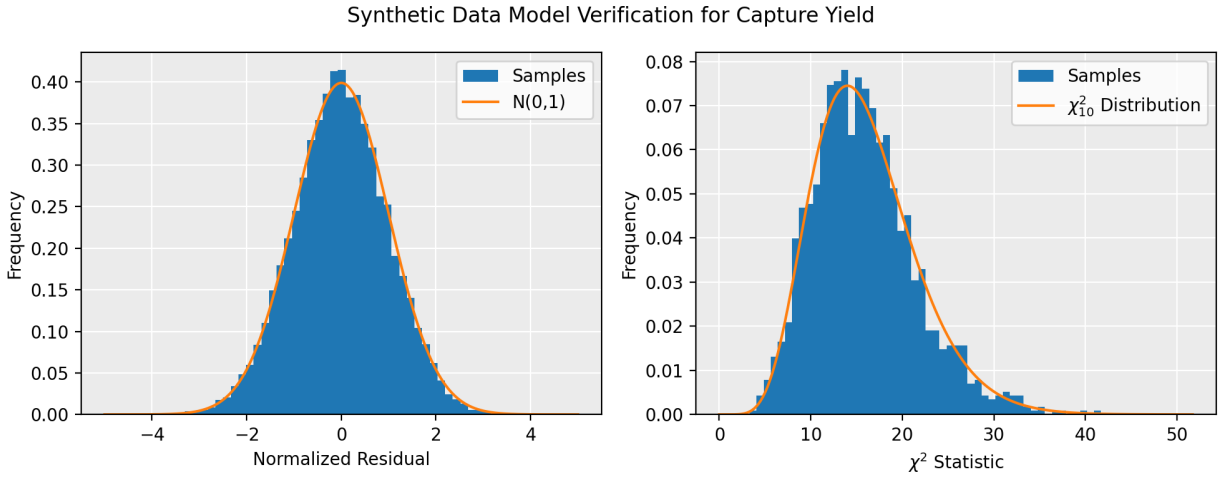

Figure 3.4: Verification of the generative model for capture yield measurements.

susceptible to bias. These can be randomly perturbed to avoid biasing the synthetic dataset to a particular set of incorrect average parameters. The fidelity of the theory behind the resonance parameter distributions is validated by the extent to which evaluations and statistical analysis of resonances to-date follow the proposed distributions. The caveat is that these theoretical parameter distributions are often considered in the evaluation process, making the resulting evaluation dependent on the theory. While investigating the fidelity of the statistical theory of resonances is beyond the scope of this thesis, it is important to note that the synthetic data framework developed here could be utilized for such a study.

The second generative model for generating noisy, experimental observables from a determined set of resonance parameters is the primary novel contribution of the methodology. Assessing the utility of this generative model is a non-trivial task. The primary challenge is that actual data with a labeled solution does not exist, and as a result, the second generative model cannot be decoupled from the first. The variance driven by the resonances (generative model one) dominates over the variance driven by the measurements (generative model two). Therefore, without access to the true underlying resonance parameters for ^{181}Ta, a robust statistical comparison of generative model two is not feasible.

The 12 mm ^{181}Ta transmission measurement data from the reference experiments [28, 82] was used for heuristic comparison. Evaluated ^{181}Ta resonance parameters from JEFF-3.3 [1] were used as a starting point and the maximum likelihood solution to the data set was found. These resulting parameters were fixed (not sampled) and used to generate synthetic datasets per the methodology of the second generative model described in this thesis. The result being synthetically generated statistical realizations of experimental data from an underlying set of determined resonance parameters. In order to compare synthetic and observed datasets, it must be assumed that the determined resonance parameter set underlies the reference data as well. This comparison is visualized in Fig. 3.5 where the assumed, underlying transmission is shown in green and the synthetically generated data can be visually compared to the reference data. Another point of comparison is the correlation structure of the uncertainty in the datapoints with respect to TOF. Figure 3.6 shows the correlation matrix for observed and synthetically generated data with no discernible differences in structure or magnitude. A more detailed heuristic comparison of the impact of monitor normalizations is provided

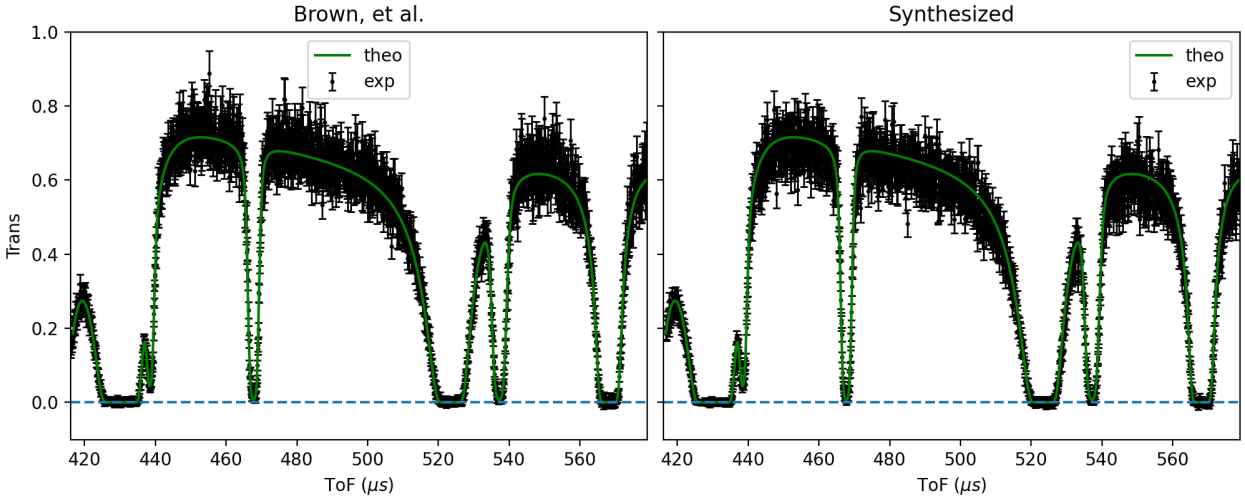

Figure 3.5: Observed experimental data compared to synthetically generated data assuming the GLLS solution to the observed data is the true underlying set of resonance parameters. Shifts in the distribution around the theoretical curve are expected as different samples come with different realizations of background or other reduction parameters.

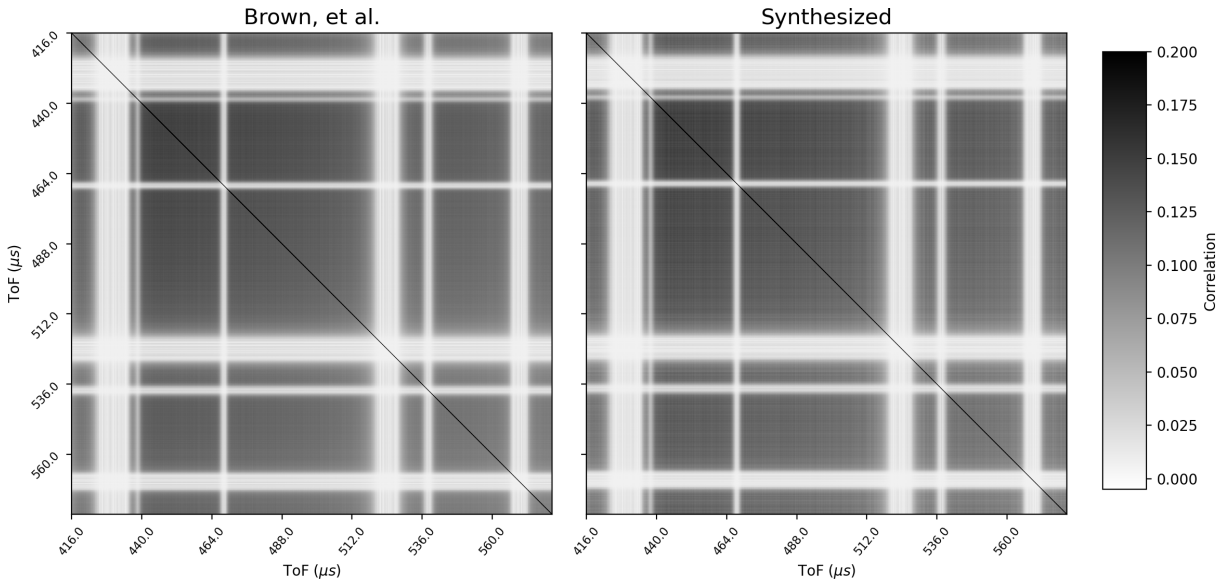

Figure 3.6: Time of flight correlation matrix for observed experimental data compared to synthetically generated data.

in Appendix A.1.

Figures 3.5 and 3.6 show data generated from fixed resonance parameters such that repeated samples would produce statistical realizations of experimental data from the same fixed cross section—emulating multiple experimental measurements of the same cross section. The claim for a high-utility generative model is that, assuming the resonance parameters are exactly correct, the observed data is one possible realization from the ensemble distribution of synthetically generated datasets. Any one realization can be characterized by a distribution of normalized residuals, calculated by,

$$\tau = \frac{T_{\text{exp}} - \text{E}[T]}{\sigma_{\text{exp}}} \tag{3.16}$$

where the T_{exp} is the experimental data point (synthetic or observed) and σ_{exp} is its estimated standard deviation. $\text{E}[T]$ is the expectation value for the transmission which is directly related to the theoretical cross-section that produced the data. In order to rigorously test the utility claim, access to the exact cross section that produced the observed reference data is needed. The Porter Thomas distribution suggests that the most abundant resonances are those with a very small width. Smaller resonances are often lost to experimental resolution and therefore missing from the evaluation and the determined cross section used to generate synthetic datasets. Without access to the true underlying resonance parameters of ^{181}Ta, it is likely that small resonance structures are contributing to the observed distribution.

To address this, the study can be filtered to only the 'black resonances', meaning resonances that allow approximately no neutrons to be transmitted. At these sections of the ToF measurement, the expectation value for the transmission is known even if the theoretical cross section is not specified exactly (assuming, of course, that the presumed theoretical cross section is similar enough to produce a black resonance in the first place). The detector response in these ToF windows comes only from the background spectrum, therefore, the observed transmission data is a function of only the second generative model and is not affected by the potential presence of small, missed resonances in the evaluation. This is visualized in Fig. 3.7, where the 'black resonances' in the 410–580 μs ToF window are identified. In order to improve the power of the statistical comparisons through sample size,

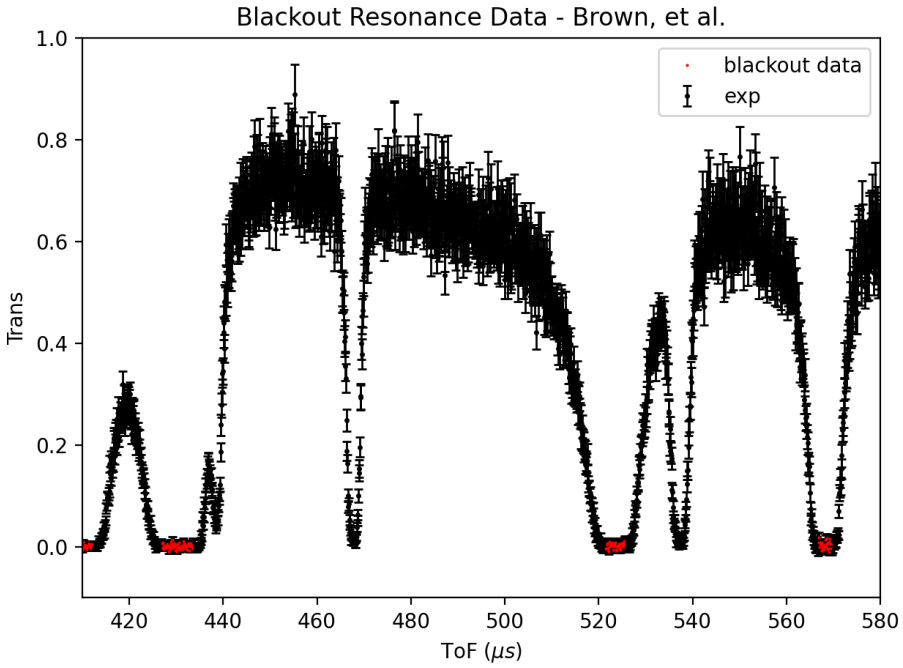

Figure 3.7: Blackout resonances and highlighted data where the expectation value can be well-estimated in the 410-580 μs ToF window.

all obviously blacked-out resonances from the RRR were considered. The resonance energy locations that were selected are given in Table 3.2. The total number of blackout residuals per RRR realization (synthetic and observed) was 572.

By filtering the data to include only black resonances, the distributions are isolated to the second generative model, where the variance is driven by counting statistics and uncertainties in the data reduction process rather than the underlying theoretical resonance parameters.

Given the hierarchical nature of the generating process, the objective of the generative model is to reproduce the ensemble distribution, while the actual data represent only a single realization. In other words, the actual data correspond to a single true set of parameters that describe the generating process, while the reductive model parameters serve as fixed estimates of these. The generative modeling methodology employed here randomly samples the true parameters. As a result, the ensemble distribution of residuals will not be directly comparable to the actual measured residual distribution. The claim for high-utility synthetic data is that the ensemble distribution should fully capture the observed distribution and have a nonzero probability of realizing statistically identical data. This claim is supported by the results shown in Figs. 3.8 and 3.9 and Table 3.3.

The histograms (Figs. 3.8 and 3.9) shows the ensemble distribution and the observed realization of normalized residuals and data uncertainty. The lines indicate the minimum and maximum of each, demonstrating that the ensemble distribution fully captures the observed realization. For each realization that contributed to the synthetically generated ensemble distribution, statistical tests were performed to compare with the observed realization. The two-sample Kolmogorov–Smirnov and Anderson–Darling tests [85, 86] were used to quantify the statistical similarity of the two distributions. The first tests the null hypothesis that the two sample sets have identical underlying continuous distribution functions. The second tests the null hypothesis that the two sample sets are drawn from the same population. The results of these statistical similarity tests can be represented using a significance level, or p-value. For different significance levels, the percentage of realizations for which the null hypothesis failed to be rejected is given in Table 3.3. These results show that a significant percentage of the synthetic data realizations are statistically indistinguishable from the observed data.

Table 3.2: JEFF 3.3 evaluated resonance energies nearest to the selected blackout locations and exact ToF ranges over which the expectation and residual were calculated.

Middle Energy (eV)	ToF Range (μs)
4.280	1175 - 1290
10.36	785 - 803
13.95	684.5 - 688.75
20.29	566.75 - 569.5
23.92	521.5 - 525.5
35.90	427.0 - 433.5
39.12	406.75 - 412.00
99.20	258 - 259
174.9	195.1 - 195.7

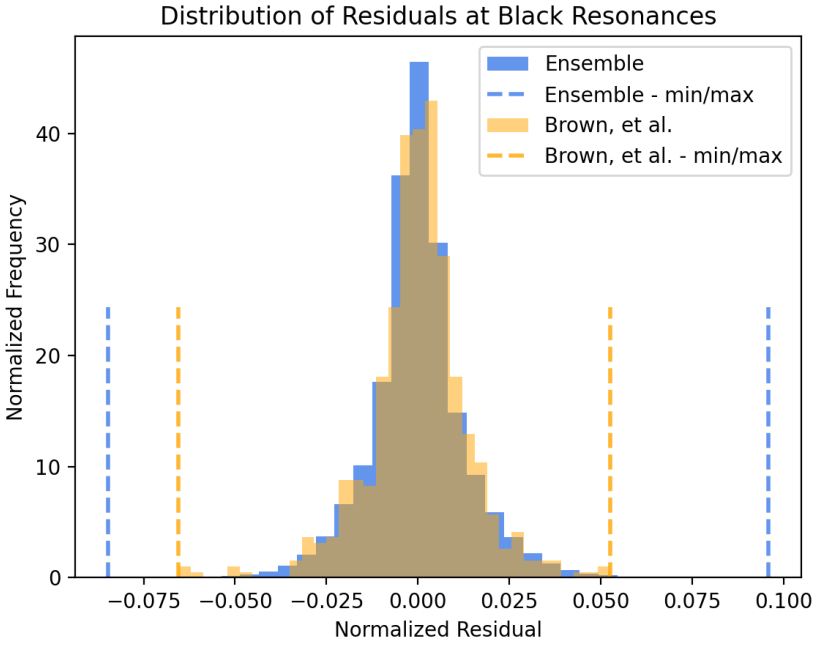

Figure 3.8: Histogram of residuals for the observed realization and synthetic ensemble. The synthetic ensemble fully captures the observed and is similar in shape.

Table 3.3: Percent realizations that failed to reject the null hypothesis for Kolmogorov-Smirnof and Anderson-Darling 2-sample statistical similarity tests at different significance levels.

p-value	KS test	AD test
0.05	54.5%	58.6%
0.10	47.0%	49.6%
0.15	41.0%	41.9%
0.20	36.6%	35.8%
0.25	32.2%	32.8%

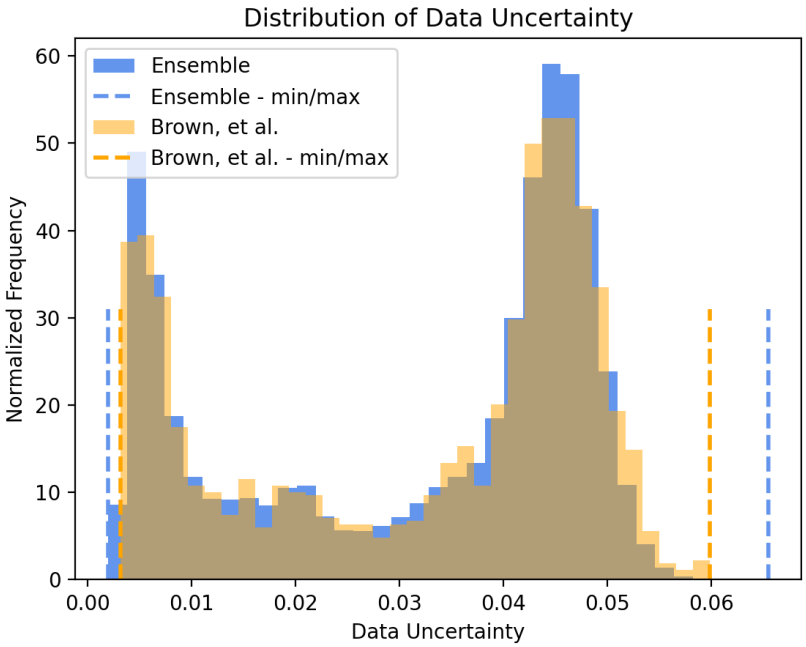

Figure 3.9: Histogram of transmission uncertainties for observed realization and synthetic ensemble. In this case, the data uncertainties for the entire window shown in Fig. 3.7 are considered rather than just those at black resonance locations. The synthetic ensemble fully captures the observed and is similar in shape.

3.3 Testing & Results

3.3.1 Implemented Models

As discussed in Sec 2.3, the demonstrations in this thesis effectively analyze the measurement data detailed in [82], emulating an evaluation of ^{181}Ta using only the referenced data. This includes four transmission measurements at 1, 3, 6, and 12 mm sample thicknesses as well as two capture yield measurements at 1 and 2 mm sample thicknesses. The primary reason the demonstration does not incorporate additional measurement data, as a full evaluation would, is that the novel framework presented in this thesis relies on high-utility generative synthetic data models for all measurements included in the analysis. The development and validation of these models requires significant time and a detailed knowledge of the measurements; thus, the scope was limited to these measurements alone.

Generative models for all six measurements were developed using the methodologies in Sec. 3.1.3 and are available in the online repository for the ATARI/Syndat python code. The four generative transmission models are considered full-fidelity for the following reasons: (1) the generative methodology is considered to be fully developed, (2) methodology was tested/validated against actual data, (3) they have access to all necessary measurement parameters including the open neutron flux spectra. The capture yield measurements are not considered full-fidelity for the reasons discussed in Sec. 3.1.3. While the limited-utility of the capture yield data may prevent some studies, the generative model is considered representative enough for the investigations in this thesis.

In Ref. [82], the 12 mm transmission measurements were to be used for validation (i.e., not used in the determination of resonances parameters but held back for comparison to estimated parameters). The same approach is taken for the following demonstrations such that only 3 transmission measurements and 2 capture yield measurements are used for inferring resonance parameters.

3.3.2 Performance Metrics

What is the ultimate objective of a resonance evaluation? While it might seem obvious that the goal is 'to be as close to the truth as possible,' this answer becomes ambiguous upon closer examination. Traditionally, the best method for assessing the accuracy of a resonance evaluation is through experimental validation. This involves using new observable measurements—such as differential or integral measurements—that were not considered in the evaluation. However, this approach is limited because these new observables are influenced by both the resonance evaluation and other uncertain input parameters (experimental parameters, noise, criticality simulations, etc.). Without isolated sensitivity to the evaluation in question, these accuracy measures are imprecise.

Access to the synthetic data resource developed in this thesis provides a unique capability to precisely measure evaluation accuracy. Within the synthetic data-testing framework, an estimated model can be directly compared to the solution. The comparison to the truth will be referred to as the performance metric. In an abstract sense, improving this performance metric is the central objective of this thesis, while the various methods employed may have their own specific internal objectives.

The synthetic data framework gives access to the true set of resonance parameters. However, it is still not exactly clear the most appropriate way to compare these to a given set of estimated parameters. Immediately, several competing methods for quantifying error come to mind: relative, absolute, squared, or other. Each of these loss functions will be more or less sensitive to a different kind of deviation from the solution and may drive decisions about the best approach in different directions. Furthermore, should this error can be calculated for the model parameters? The reduced model parameters?[5] What about the point-wise reconstruction of the cross section itself?

The current analysis has focused on error calculated for the point-wise reconstruction of each constituent reaction cross section, Doppler broadened to room temperature (300K). The reason for this is twofold: (1) broadened cross sections are what is used for neutron transport simulations, (2) the potential for disagreement in the cardinality of the estimated and true

[5]Reduced model parameters would be reduced widths (γ) instead of (Γ). See u vs p parameters in the SAMMY derivations of R-matrix theory for further details [8].

models complicates the comparison of resonance parameters themselves. This will be referred to as the theoretical cross section, in contrast to the experimental cross section, which undergoes further corrections to align with measured observables. The theoretical cross section is a much more precise comparison than experimental or measured cross sections, and the point-wise nature allows for a number of statistical accuracy measures to be applied.

The point-wise residual for one constituent reaction (rxn) will be defined as

$$r^{\text{rxn}}(E) = \sigma_{\text{fit}}^{\text{rxn}}(E) - \sigma_{\text{true}}^{\text{rxn}}(E) \ , \tag{3.17}$$

where $r^{\text{rxn}}(E)$ is calculated on a fine grid in energy[6]. The mean squared error across all reactions and energies is then calculated as

$$\text{MSE}_{\text{S}} = \frac{1}{N_E N_{\text{rxn}}} \sum_{\text{rxn}} \sum_{i} \left(r^{\text{rxn}}(E_i)\right)^2 \ , \tag{3.18}$$

and represents a combination of both bias and variance in the estimate. The square root can be taken to maintain interpretability and get the standard error in the same units as the residual:

$$\text{RMSE}_{\text{S}} = \sqrt{\frac{1}{N_E N_{\text{rxn}}} \sum_{\text{rxn}} \sum_{i} \left(r^{\text{rxn}}(E_i)\right)^2} \ . \tag{3.19}$$

Another, potentially more intuitive, metric is the mean absolute error (MAE) calculated as

$$\text{MAE}_{\text{S}} = \frac{1}{N_E N_{\text{rxn}}} \sum_{\text{rxn}} \sum_{i} \left|r^{\text{rxn}}(E_i)\right| \ . \tag{3.20}$$

Similarly, the mean absolute percent error is given by

$$\text{MAPE}_{\text{S}} = 100 \times \frac{1}{N_E N_{\text{rxn}}} \sum_{\text{rxn}} \sum_{i} \left| \frac{r^{\text{rxn}}(E_i)}{\sigma_{\text{true}}^{\text{rxn}}(E_i)} \right| \ . \tag{3.21}$$

Per the synthetic data framework, performance is assessed over many samples (N_{s}), or generated realizations of true resonance parameters and corresponding experimental data. The subscript $_{\text{S}}$ in Eqs. 3.18– 3.21 indicates that the metric is averaged over the energy

[6]For this work, it was found that 100 points per eV was sufficient enough to capture the sharpest resonance peaks at room temperature.

dimension and can be calculated on a per-sample basis. The metrics above can be averaged over samples as well, *e.g.,*

$$\text{MSE} = \frac{1}{N_\text{S}} \sum_{N_\text{S}} \text{MSE}_\text{S} \equiv \frac{1}{N_\text{S} N_\text{E} N_\text{rxn}} \sum_{N_\text{S}} \sum_{\text{rxn}} \sum_{i} \left(r^\text{rxn}(E_i)\right)^2 \ , \tag{3.22}$$

to give a single value representing the total performance across all samples, energies, and reactions. Minimizing the total MSE corresponds to maximizing the likelihood of σ_fit under Gaussian priors. Again, the square root can be taken to get the standard error in the same units as the residual:

$$\text{RMSE} = \sqrt{\text{MSE}} \ . \tag{3.23}$$

While the MSE is selected as the primary metric for comparing the quality of several estimators, the other metrics will frequently be reported based on their interpretability. Moreover, these different metrics tend to yield similar conclusions about which estimator is 'the best'.

3.3.3 Surrogate Objective Function for Regression

Ideally, one of the metrics discussed in Sec. 3.3.2 would be selected as a loss function and resonance parameter estimates would be optimized directly to it. That is, the deviation from the true cross section $\sigma^\text{rxn}_\text{true}(E)$ would be minimized. In reality, $\sigma^\text{rxn}_\text{true}(E)$ is not accessible, nor can it be measured or observed. As discussed in Sec. 2.1.2, what is actually measured are composite observables that are a function of $\sigma^\text{rxn}_\text{true}(E)$ and other experimental parameters, making $\sigma^\text{rxn}_\text{true}(E)$ a latent variable. These observables can be far removed from the latent variable of interest, meaning they have large sensitivity to the other observables or latent variables. Data reduction attempts to transform observational data to bring it closer to $\sigma^\text{rxn}_\text{true}(E)$; however, for resonance cross sections the closest object to $\sigma^\text{rxn}_\text{true}(E)$ one can get with data reduction is transmission and reaction yield(s). Data reduction generally makes parameter optimization easier, but can be problematic if there are biases in the other experimental parameters used in the transformations. This work has maintained the standard approach of optimizing parameters to transmission and reaction yield(s), though

the following study and discussion could be expanded to include optimization to direct observables or any in-between transformation.

For ^{181}Ta, the resonance evaluation has access to transmission and capture yield measurement data that is noisy and uncertain. An objective function with respect to this data must be selected as a surrogate for the ultimate objective that is a measure of accuracy with respect to $\sigma^{\text{rxn}}_{\text{true}}(E)$. In the following study, the synthetic data framework developed in this thesis is used to test various surrogate objective functions and quantitatively evaluate their performance. This simple demonstration highlights the primary use-case and novelty of the framework: the ability to test various decisions, assumptions, or errors in the evaluation process and quantitatively assess their impact in relation to the ultimate objective of accurately estimating $\sigma^{\text{rxn}}_{\text{true}}(E)$.

The study is done using 500 random samples from the generative models detailed in Sec. 3.3.1 in the energy domain 200–220 eV[7]. The conclusions resulting from this study are assumed to hold true for the real set of measurement data in [28, 82] because the utility of the generative models. The surrogate objective functions tested are the least squares objective,

$$\text{LS} = (D-T)^T(D-T), \tag{3.24}$$

the weighted least squares objective,

$$\begin{aligned} &\text{WLS} = (D-T)^T V^{-1}(D-T), \\ &\text{where}: \ V_{i\neq j} = 0 \end{aligned} \tag{3.25}$$

the generalized least squares objective,

$$\chi^2_{\text{D}} = (D-T)^T V^{-1}(D-T), \tag{3.26}$$

and the generalized least squares objective with a dynamic data covariance matrix calculated at the theory values,

$$\chi^2_{\text{T}} = (D-T)^T V(T)^{-1}(D-T). \tag{3.27}$$

[7]This energy domain was chosen arbitrarily for initial demonstrations; additional energy domains will be tested later in this thesis.

For each equation, D, T, and V represent the data, theory, and data covariance as detailed in Sec. 2.5. In a Bayesian sense, the progression from LS to wLS to χ^2_{D} represents the inclusion of additional information in the objective function. The difference between the χ^2_{D} and χ^2_{T} objectives comes about to address what the nuclear data community refers to as Peele's Pertinent Puzzle (PPP) [87]. This phenomenon comes about because the structure of V is known ahead of time, but the actual values are generally estimated using the measured data. Updating the V estimate to reflect the current estimate T is one of the suggested solutions in [8], though the difference between is said to only be an issue for discrepant data.

For each surrogate objective function, the true resonance parameters are used as the starting point. This guarantees the correct model complexity, spin assignment, and convergence, ensuring that factors such as the choice of iterative solver or its hyperparameters, aside from the convergence criteria, have minimal impact on the final solution[8]. The convergence criteria was set to be $\Delta \frac{F_{\mathrm{obj}}(T)}{N_{\mathrm{D}}} < 0.001$, where N_{D} is the number of data points.

Several total performance metrics are reported in Table 3.4, that is, accuracy measures averaged over all energies, reactions, and samples. The distribution of the $\mathrm{RMSE}_{\mathrm{S}}$ metric over 500 samples is shown in Fig. 3.10 for both constituent reactions. Note that distributions shown in Fig. 3.10 are a more detailed representation of the total RMSE metric in Table 3.4.

The accuracy with respect to the true cross section improves as additional information is included into the objective function, this result is expected. The significant improvement with the use of the χ^2_{T} objective is surprising as literature on the PPP phenomenon generally indicates that it is negligible if data are non-discrepant and without strong correlation—which seems to be the case for this data based on the investigations in Secs. 2.3 and 3.2 and Refs. [28, 82].

This brief study illustrates how the generative modeling framework can be leveraged to quantitatively assess which surrogate objective function most accurately estimates the true cross section. Alternatively, these findings provide a quantitative measure of the impact of potential errors in the evaluation process. For instance, the difference in MAE between χ^2_{T} and χ^2_{D} suggests that failing to correct for PPP in the evaluation would result in

[8]The LM algorithm with an internal SAMMY update step was used. See Sec. 4.2.1.

Table 3.4: Performance metrics for different objective functions.

Surrogate Objective	RMSE (bn)	MAE (bn)	MAPE (%)
χ^2_{T}	6.68	1.40	3.60
χ^2_{D}	16.37	3.22	21.76
WLS	23.40	4.97	24.76
LS	30.88	6.74	146.69

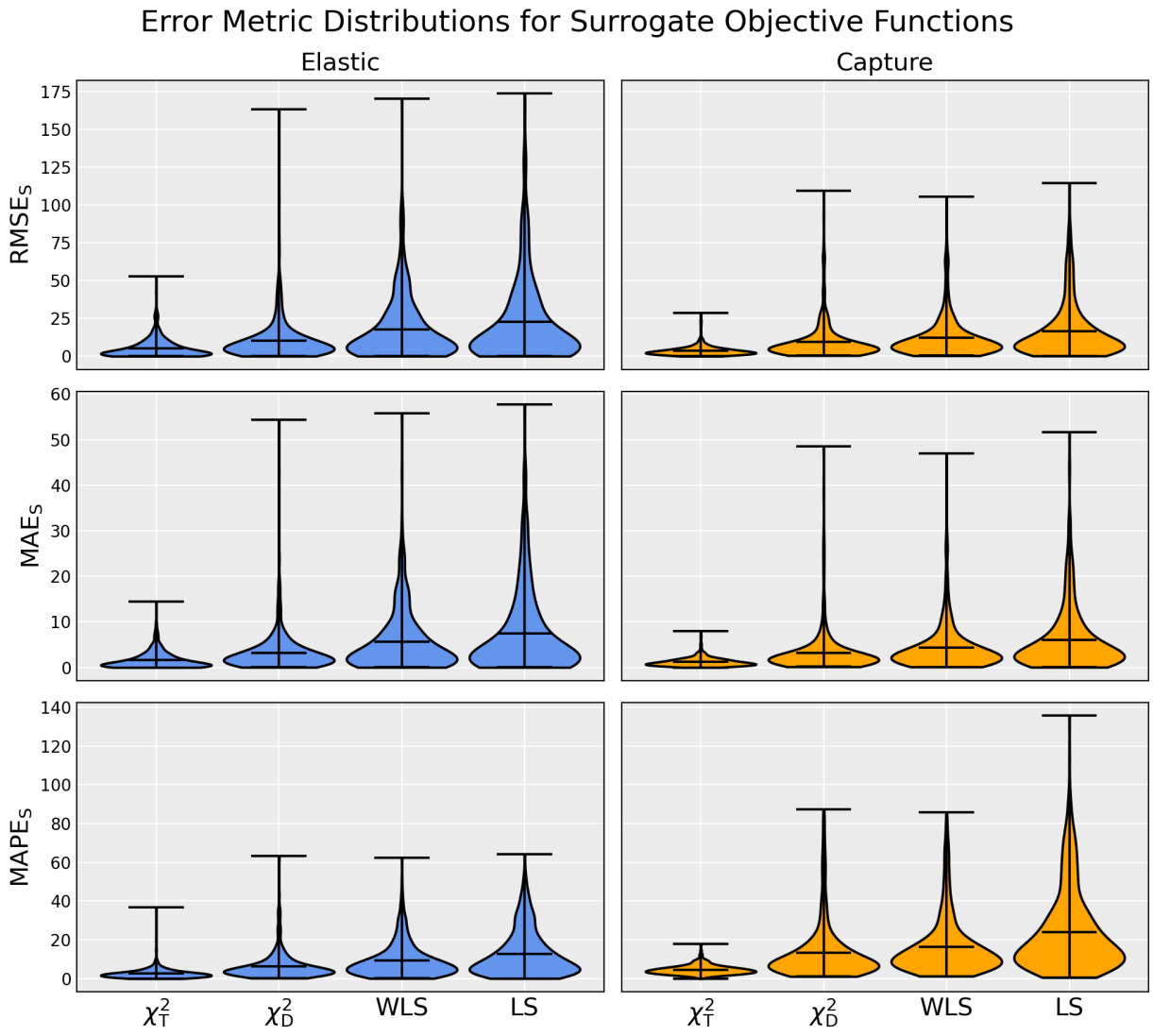

Figure 3.10: Distribution of Root Mean Squared Error metric across samples and energy for capture and elastic reactions.

predictions that are, on average, 2 barns further from the true cross section. Likewise, the difference between WLS and the proper GLS objective quantifies the impact of neglecting data correlations. Consider a similar study that emulates misreported data covariance, a common challenge for evaluators looking to incorporate past/historical data. This framework could be used to quantitatively assess the net impact of including a potentially discrepant or misreported dataset in an evaluation. This is, of course, a 'best case scenario' because the cardinality and spin assignment is known exactly. Studies reported later in this thesis will expand this concept to fully automated resonance analysis where the cardinality and spin assignments are not known ahead of time. In the subsequent studies, the results from this fit-from-true (FFT) analysis will be reported as the reference minimum achievable performance, as constrained by the choice of surrogate regression objective.

Chapter 4

Automated Fitting

A major contribution of this thesis is a methodology for automated resonance parameter inference. The goal is to automate this aspect of resonance evaluation, which is grounded in a well-defined objective based on statistical theory. The target for automation is (a) determination of number of resonances, (b) determination of the discrete quantum spin groups for each resonance, and (c) determination of optimal continuous resonance parameters and their uncertainty, given a and b. Item c is not novel, but traditional methods require a good starting point for continuous resonance parameters. Traditionally, a and b are determined subjectively by the evaluator, who may make these decisions manually, base them on previous evaluations, or use a combination of both methods. This approach introduces human bias and lacks reproducibility. The goal of the automated methodology presented here is to address these issues in evaluation while also providing a tool useful for the general analysis of experimental resonance data.

It should be noted that it is *not* the goal of this methodology to replace the evaluator. Firstly, many steps in the evaluation process are beyond the scope of this automated methodology (e.g., data curation, where expert judgment is invaluable). Secondly, this methodology is not foolproof, and the output should always be carefully checked by the evaluator. The methodology presented here is meant to augment the evaluator by taking a set of easily-reportable inputs and outputting reliable and consistent model estimates. If the evaluator is not satisfied with the model estimate, they can modify the algorithm's inputs

rather than manually adjusting the model parameters. In this way, automation enhances reproducibility.

4.1 Problem Formulation

As exposed in Sec. 2.5.4, from an optimization perspective, the Bayesian prior (P and M) serves as a good starting point and provides guidance on step size and direction. Furthermore, in the case where there is no *a priori* information about the parameters, parameter uncertainty estimates will depend only on the data covariance V and the linear approximation of the derivative of the model with respect to the parameters at the maximum likelihood estimate:

$$\begin{aligned} M' &= (M^{-1} + G^T V^{-1} G)^{-1} \\ \text{where : } M^{-1} &= 0 \\ G &= \frac{\delta T(P)}{\delta P}\Big|_{P=P_{\text{MLE}}} \end{aligned} \tag{4.1}$$

In this case, M' can be evaluated at any time with access to V and P_{MLE}[1] . Ignoring the calculation of M' until later, the goal is to solve the optimization problem

$$P_{\text{MLE}} = \arg\min_P \ \chi^2, \tag{4.2}$$

without knowing the size P or good initial values. In other words, this is a non-convex, non-linear optimization problem with unknown cardinality/model complexity.

The most challenging aspect of this problem is the non-convexity. This is addressed by designing a starting point that falls within convex regions[2]. This starting point is a dense set of resonance features, or feature bank, composed of many resonances (details discussed in 4.1.1). Then, the model is optimized using a non-linear iterative optimization (details discussed in 4.2.1). The solution from the initial feature bank is at one extreme of model complexity. Here, model complexity has a convenient interpretation as each

[1]This version of uncertainty quantification does depend on Gaussian and Linear approximations. If these are not good approximations, Bayesian Monte Carlo methods can be used. These methods, however, still require a good starting point for sampling.
[2]This approach is inspired by feature engineering in machine learning applications.

resonance corresponds to a discrete level in the compound nucleus, that is, model complexity is synonymous with the number of resonances.

The result given by the optimized initial feature bank should perform very well with respect to the regression objective (i.e., it will have a very small χ^2). However, theoretical knowledge about resonance level spacing [18] and the general philosophy of Occam's razor that underlies scientific inference tells us that this result is likely over-fitting and the number of resonances should be reduced. Thus, a variable/feature selection scheme is implemented to reduce model complexity. The goal of variable selection is to find the optimal model for each level of model complexity. From a physics perspective, each level of model complexity corresponds to an integer number of resonances. The variable selection scheme starts from a many-resonance model and reduces down to zero resonances[3]. Further details on approaches to variable selection are given in 4.2.2. During the variable selection scheme, optimal spin group assignments are also determined as the initial feature bank is composed of all possible spin assignments[4].

Algorithm 1 shows the overall workflow for the optimization problem. The result is a set of optimal model(s) for each level of model complexity (number of resonances). The appropriate number of resonances can then be selected by some model selection criteria as discussed in Sec. 2.5.3 or by some other approach determined by the evaluator. The author recognizes that evaluators may consider additional data (e.g., gamma multiplicity, resonance integral) or additional likelihood models (e.g., Porter Thomas & Wigner parameter distributions) to aid in the determination of spin assignments and/or number of resonances[5]. Currently, these are not included in the automated methodology presented here, but they are a natural extension for future work. Also, step 1 in Algorithm 1, the initial feature bank optimization is tailored to ^{181}Ta, given that only neutron and gamma channel widths are present. Section 4.1.1 will address the modifications required to extend this approach to other isotopes.

[3]Or some reasonably low number of resonances.
[4]Optimal here means with respect to the χ^2 regression objective. In the future, resonance statistics could be incorporated into the objective function for spin assignment
[5]For example, in Ref. [54] each level of model complexity was iterated with a separate tool for spin assignment developed at Brookhaven National Laboratory [39].

Algorithm 1 Non-convex, non-linear optimization with variable selection

Step 1: Initial feature bank optimization

initialize a dense set of resonance energies all with $\langle \Gamma_\gamma \rangle_\lambda$ and small $\Gamma_{n,\lambda}$

$\vec{\mathbf{P}} = \{P_\lambda \mid \lambda = 1, 2, \ldots, N\}$

$P_\lambda = (E_\lambda, \Gamma_{\gamma\lambda}, \Gamma_{n,\lambda})$

optimize over E_λ and Γ_n and eliminate resonances that remained small

while un-converged **do**

    $\min_{\vec{\Gamma}_n} \chi^2$

    $\min_{\vec{E}_\lambda, \vec{\Gamma}_n} \chi^2$

    $\vec{\mathbf{P}} \leftarrow \{P_\lambda \mid P_\lambda \in \vec{\mathbf{P}} \text{ and } \Gamma_{n,\lambda} \geq \epsilon\}$

end while

Step 2: Backwards stepwise variable selection

iteratively remove resonances with the least impact on the objective until number of resonances (N) is 0

while $N \geq$ target **do**

    for P_λ in $\vec{\mathbf{P}}_N$ **do**

        $\vec{\mathbf{P}}_i \leftarrow \vec{\mathbf{P}} \setminus P_i$ (remove resonance i from model)

        $s_i \leftarrow \min_{\vec{\mathbf{P}}_i} \chi^2$ (optimize and get score)

    end for

    $\vec{\mathbf{P}} \leftarrow \vec{\mathbf{P}}_i$ for best s_i

end while

4.1.1 Initial Feature Bank Design

Implementation of a robust non-linear solver will give solutions to local minima for the non-linear objective function; however, it does not address the non-convexity of the problem. Global minimization of non-convex problems is challenging, often requiring stochastic and/or heuristic algorithms with no guarantee that the global optimum will be found [88]. This becomes more difficult when the cardinality of the problem is not known. To address this, domain knowledge is used to engineer a starting point that forces convexity in the objective function. This starting point is considered a "feature bank" where each feature corresponds to a resonance structure at some energy, E_λ. Naturally, the exact regions of convexity will depend on the specific data and model, making the initial feature bank settings a user-defined input.

The toy examples given in Appendix A.2 demonstrate several important characteristics of the objective function that motivate the design of the initial feature bank. Firstly, there is usually a narrow, convex region around the optimal energy location (E_λ) of a resonance. As E_λ moves away from the minimum, the objective function becomes non-convex with respect to the width parameters. When this happens, the neutron width (Γ_n) develops two local minima, one at a smaller value and the other at a very large and improbably value. Secondly, the objective function is largely insensitive to the capture width (Γ_γ) regardless of the other two parameters. This is not always the case as the sensitivity to Γ_γ depends on the ratio of Γ_γ to Γ_n. For this reason, along with other factors such as the possibility of additional channels, the initial feature bank design should be tailored to the specific evaluation at hand.

Generally, a dense, uniform grid of resonance energy locations (e.g., $>$ 1 per eV) for each possible spin group (J^π) will ensure convexity with respect to E_λ. However, if the average level spacing is very small, a more dense grid may be necessary and spacing can be determined by some lower quantile of the respective Wigner distribution [18].

An initial feature bank energy grid could be set using a peak-finding algorithm like find_peaks in SciPy [89], but this step was deemed unnecessary for two reasons. First, these algorithms require thresholds to distinguish signal from noise, which vary with experimental

data properties (e.g., noise level, resonance width, resolution), and thus don't reduce the user-defined input. Second, multiple measurements of the same resonance signal (e.g., different observables or thicknesses). This presents the challenge of combining peaks found in each of the different measurements. The most robust approach would be to simply add all of them to the model, but this would result in a very large and expensive initial feature bank.

The initializing values for Γ_n should generally be small to ensure they do not wander off to a large, unlikely minimum (shown in Fig. A.6 in the appendix). This can be given by some lower quantile of the respective Porter Thomas distributions [17]. The choice of initializing values for Γ_γ is less obvious; convexity in this parameter is highly sensitive to the other parameters E_λ and Γ_n, while the objective function itself is not very sensitive to it. Furthermore, the Γ_γ parameter is theorized to deviate only slightly from its average value and, at the evaluator's discretion, may not be updated from the prior-estimated average value at all. Consequently, the prior estimated average value is used to populate the initial resonance feature bank with Γ_γ values. If the evaluator chooses to optimize this parameter, then it can be fitted at the last step of the variable selection scheme for each model complexity (i.e., after near-optimal values have been found for E_λ and Γ_n).

4.2 Methods

Section 4.1 formulated the resonance parameter inference problem as an optimization and detailed an algorithmic procedure to solve it. Within that algorithm, there are a three modular parts for which a variety of methods could be used. The first is the iterative, non-linear solver used to solve the optimization problem given by Eq. 4.2 for a fixed model complexity. The second is the variable selection method used to reduce model complexity or number of resonances. Lastly, a model selection approach is needed to determine the most appropriate model complexity. For each of these parts, several different approaches were developed, implemented, and tested.

4.2.1 Iterative Solvers

For the iterative, non-linear optimization routine, the goal is to minimize the χ^2 regression objective by continuous optimization of the resonance parameters $E_\lambda, \Gamma_\gamma,$ and Γ_n for all resonances in the model. Several algorithms are briefly discussed in Sec. 2.5.4; here, those that are implemented and tested in this thesis will be explicitly defined by the mathematical equation for a parameter update step. That is, the equation for P^{i+1} given P^i where P^i represents the vector of continuous resonance parameters are step i in the iterative algorithm. Parameters D, T, and G represent the data, theory, and Jacobian as described in Sec. 2.5.1. The Jacobian G is always evaluated at P^i, with notation dropped for brevity.

A step in the Gauss-Newton (GN) algorithm is described by

$$P^{i+1} = P^i + (G^{\mathrm{T}}V^{-1}G)^{-1}\ G^{\mathrm{T}}V^{-1}(D - T^i), \tag{4.3}$$

and it does not incorporate a hyperparameter to determine the step size. A step in the gradient descent (GD) algorithm step is described by

$$P^{i+1} = P^i + \alpha G^{\mathrm{T}}V^{-1}(D - T), \tag{4.4}$$

where α controls the step size. The α parameter is actually a vector that is related to a user define scalar, α_0 by

$$\alpha_j = \begin{cases} \alpha_0, & \text{if } P_j \notin E_\lambda \\ \frac{\alpha_0}{100}, & \text{if } P_j \in E_\lambda \end{cases} \tag{4.5}$$

to limit the step of the resonance energy parameter E_λ. This is done because the objective function is known to be highly non-linear and non-convex with respect to E_λ; furthermore, the dense initial feature bank ensures that resonance energies require minimal adjustment.

Two versions of the Levenberg-Marquardt (LM) algorithm were implemented, denoted $\mathrm{LM_a}$ and $\mathrm{LM_b}$. The $\mathrm{LM_a}$ is a traditional implementation, with a parameter update step given by

$$P^{i+1} = P^i + \left(G^{\mathrm{T}}V^{-1}G + \lambda\mathbf{I}\right)^{-1}\ G^{\mathrm{T}}V^{-1}(D - T^i), \tag{4.6}$$

where $\mathbf{I}$ is the identity vector and λ is the LM dampening parameter. To facilitate notation and consistency across methods, λ can be set by $\frac{1}{\alpha}$, maintaining the interpretation of α_0 controlling step size. LM_b is a modified implementation that corresponds to the internal SAMMY implementation with the relative prior parameter covariance[6]. The parameter update step is given by

$$P^{i+1} = P^i + \left(G^\text{T} V^{-1} G + \lambda Q^i \mathbf{I}\right)^{-1} \; G^\text{T} V^{-1} (D - T^i), \tag{4.7}$$

where

$$Q^i_\lambda = \begin{cases} P^i_j, & \text{if } P_j \notin E_\lambda \\ P^i_\Gamma, & \text{if } P_j \in E_\lambda \end{cases}. \tag{4.8}$$

This effectively weights the step size and dampening for individual parameters by their magnitude if that parameter is a channel width. For E_λ parameters, the dampening is weighted by the corresponding total resonance width.

Part of the LM algorithm is a prescription for dynamically changing λ with each parameter update step. The concept of a dynamic step size can also be applied to the gradient descent algorithm. The high-level procedures for dynamic step size updating and convergence criteria used for the GD, LM_a, and LM_b algorithms are outlined in Algorithm 2 where the ϵ, u, d, and α_0 are user-input hyperparameters. For the subsequent studies in this thesis, the hyperparameter values used for the iterative non-linear solver are: $\epsilon = 0.01$, $\alpha_0 = 0.1$, $u = 1.5$, and $d = 5.0$.

The final iterative non-linear optimization algorithm implemented/tested in this thesis is stochastic gradient descent (SGD). The primary benefit of SGD arises when evaluating the objective function and/or it's gradient with respect to all available data is infeasible or costly; in such cases, random subsets of the data can be selected for each parameter update step making the objective function stochastic at each iteration[7]. SGD algorithms often incorporate some form of momentum in both the step size and direction to prevent an unlikely data subset from deviating the parameter estimates significantly from the global

[6] See Sec. 4.2.4

[7] In some literature, SGD refers specifically to the use of a single data point per step, while mini-batch GD employs subsets of the overall data. This thesis will use the term SGD to denote mini-batch gradient descent.

Algorithm 2 Dynamic step size procedure

$P \leftarrow P^0$
$L \leftarrow \chi^2(P^0)$
$\alpha \leftarrow \alpha_0$
while un-converged **do**
    $P' = P + \text{step}(\alpha)$
    $L' = \chi^2(P')$
    if $L - L' > \epsilon$ **then** (if objective improves: increase alpha, update parameters)
        $\alpha \leftarrow \alpha \times u$
        $P \leftarrow P'$
        $L \leftarrow L'$
    else if $L - L' < 0$ **then** (if objective deteriorates: decrease alpha)
        $\alpha \leftarrow \alpha/d$
    else
        converged
    end if
end while

minimum of the overall objective function. The SGD algorithm implemented in this thesis is ADAM [90], which leverages adaptive estimates of higher order gradient moments to give stable, efficient steps down a stochastic objective surface. The implementation of the ADAM algorithm is straightforward and uses the recommended hyperparameters; thus, it will not be repeated here.

For the resonance parameter optimization problem, the cost of evaluating the objective function and it's gradient with respect to all available data (D) is not a limiting factor. However, introducing stochasticity into the objective function and momentum into the steps helps to avoid shallow local minima and inflection points. This is done by taking random subsets of linearly-independent components of data, that is, principal components of the data covariance matrix V [91, 92]. Section 3.3.3 indicated that the optimal choice for the regression objective function choice is χ^2_{T}, where the data covariance estimate is updated to reflect the current parameter estimates. In this scenario, components of V are calculated at each step. These components will remain independent from one another during the current step; however, they may change or interact in subsequent steps. While this may not be optimal, it is ultimately acceptable because the purpose of the SGD implementation here is to introduce stochasticity into the objective function, helping to avoid local minima or inflection points.

4.2.2 Variable Selection

The initial feature bank solve will give an overfit model with many resonances (p). This result should perform very well with respect to the optimization object (i.e., it will have a very small χ^2); however, domain knowledge says that this is not correct and the number of resonances should be reduced. Starting with a full set of p explanatory variables naturally lends itself to backward stepwise selection which seeks to find the optimal model for each integer number of resonances, starting from a many-resonance model and reducing down to zero. Additionally, this is preferable as compared to forward stepwise selection because the first steps (few-resonance models) will likely fall outside of the region of convexity in the objective function and result in non-physical models.

The backwards stepwise approach is computationally greedy[8] compared to the exhaustive best-subset selection approach; however, the exhaustive approach quickly becomes computationally infeasible as the number of resonances increases. Additional options for more-or-less greedy selection trajectories can be included in the implementation of this backwards selection algorithm. For example; by the nature of R-Matrix theory, as the neutron channel width (Γ_n) of a resonance goes to zero, the resonance's impact on all reaction cross sections also goes to zero. This is convenient as, in many cases, several resonances will have a negligible Γ_n after the initial feature bank is optimized, allowing them to be removed from the model with effectively no impact on the objective function. A threshold can be applied to either (1) automatically exclude negligibly small resonance features during continuous optimization, or (2) skip fitting all $p-1$ models at a given level if the deterioration in the objective for any model falls below a specified threshold. These options are left to the user to decide based on domain knowledge. For the subsequent studies in this thesis, (1) is not done and (2) is done if the objective deterioration (normalized by N_{data}) is less than 0.001. In each case, the less-greedy approach gives more confidence that a global optima has been found. However, empirical observations during the development of this algorithm reveal the following insight: the level of greediness tends to influence the models encountered at high complexities, particularly when many of the resonance features are merely fitting noise in the data; however, as the selection algorithm progresses toward more reasonable model complexities (as indicated by model selection criteria and *a priori* knowledge about resonance level spacing), different levels of greediness converge to nearly identical models, irrespective of the path taken at higher complexity.

4.2.3 Model Selection

The variable selection routine gives optimized models for level of complexity, but does not identify which complexity should be selected. Information-theoretic (IT) criteria and cross-validation approaches to model selection were discussed in Sec. 2.5.3. For this problem, IT criteria are expected to perform poorly due to the complex non-linear model. However, these criteria essentially describe a functional trade-off between the χ^2 goodness-of-fit and model

[8]Greedy implies trading robustness for efficiency.

complexity, with the AIC that is

$$\mathrm{AIC} = 2p + \chi^2 \ , \tag{4.9}$$

where p is an appropriate measure of model complexity. The decision to select the p_i model over p_{i-1} is determined by

$$\mathrm{AIC}_i < \mathrm{AIC}_{i-1}, \tag{4.10}$$

which can be re-written as

$$\chi^2_{i-1} - \chi^2_i + 2(p_i - p_{i-1}) < 0, \tag{4.11}$$

where i represents the current model complexity and corresponds to an integer number of resonances. Ensuring that the variable selection routine gives a model every integer number of resonances (a natural feature of stepwise selection) means that the term $2(p_i - p_{i-1})$ is the same for all i. The challenge of estimating p for the non-linear model can now be reframed as a hyperparameter selection problem, defining a threshold for acceptable deterioration in χ^2 as model complexity is reduced. The criteria for selecting model complexity i and halting the variable selection routine is then given by

$$\frac{\chi^2_{i-1} - \chi^2_i}{N} > \Delta\chi^2, \tag{4.12}$$

where the χ^2 scores are normalized to the number of data points N so that $\Delta\chi^2$ can be applied for different sized datasets.

The k-fold cross-validation (CV) approach can be applied as long as the data can be split into independent subsets, or folds. In this context, the only hyperparameter is the number of folds, representing a different class of hyperparameter. Ultimately, the best k-fold estimate of CV error is LOOCV, as discussed in Sec. 2.5.3; any fewer folds are a trade-off between data limitations and computational efficiency. Systematic uncertainties in the data reduction process introduce energy correlations between data. The most attractive way to ensure independent folds would be to draw random samples of linearly independent components of the data covariance matrix, V. However, the optimal regression objective updates V with each parameter update[9]. Therefore, as described in the SGD implementation in Sec. 4.2.1,

[9]Supporting study detailed in Sec. 3.3.3.

the principle components of V must be recalculated at each step, introducing correlations between folds. Over the course of many steps and with significantly different estimates of V, it is likely that the folds will become highly correlated. Correlated folds can lead to underestimates of the generalization error, ultimately diminishing the effectiveness of this approach for model selection.

An alternative data-segmentation approach is to split the data by measurement, placing each measurement in its own fold. The different measurements are assumed to be mostly independent, except for potential correlations which could arise from the use of the same experimental equipment—such correlations are not quantified in the data reduction and would therefore also pose a challenge when segmenting over components of V. The measurement segmentation is ultimately preferred; however, it has several unsatisfactory characteristics. First, the approach is limited by the number of available measurement datasets, allowing for only 5-fold cross-validation in this case[10]. Second, this creates an imbalance in the magnitude of variance across the folds, a problem that could be mitigated through stratified sampling [93] if the independent components of V could be preserved throughout the entire regression process. Measurement-based segmentation is used throughout the studies in this thesis. The folds and underlying concept of the cross-validation approach are illustrated in Fig. 4.1.

4.2.4 Implementation

Internal to SAMMY

Sec. 2.5.4 exposed the relationship between the linear Bayes equations solved by the SAMMY code and other regression techniques. SAMMY has a user option to solve the generalized least squares (GLS) problem by setting the inverse of prior parameter covariance matrix to exactly zero ($M^{-1} = 0$). Another user option in this GLS mode will reduce the parameter step size by a factor of 10 such that each time SAMMY is called it solves the following equation:

$$P' = P + \frac{1}{10}\,(G^T V^{-1} G)^{-1}\, G^T V^{-1}(D - T). \tag{4.13}$$

[10]Fewer folds could be used, but generally 5 or 10 folds are recommended.

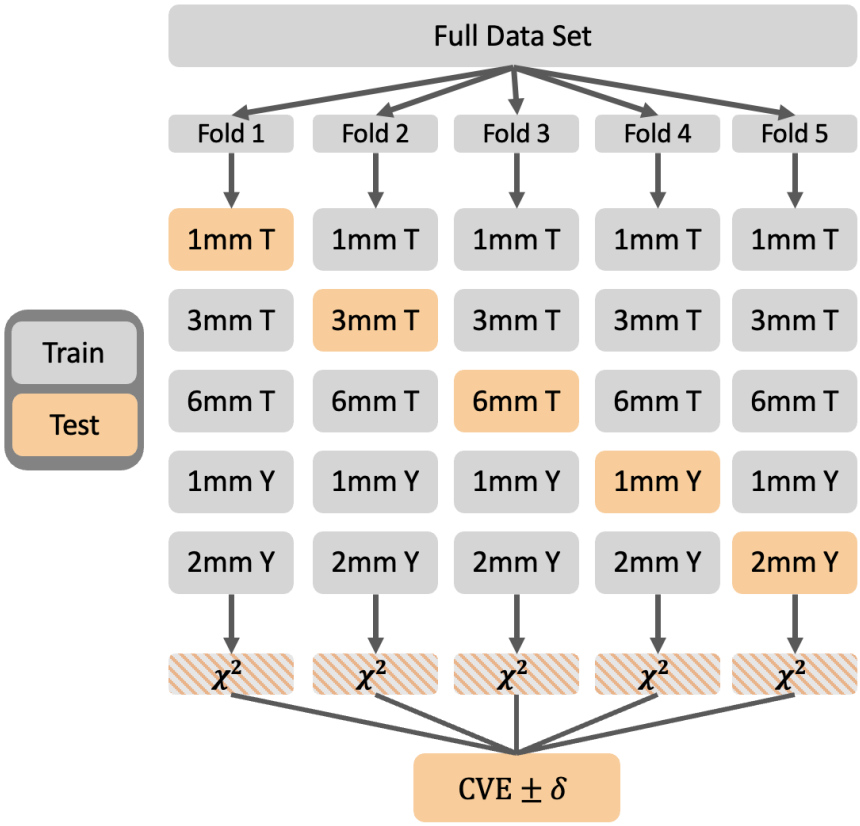

Figure 4.1: Measurement-based data segmentation for cross validation with the reference five ^{181}Ta measurements.

Thus, the Gauss-Newton (GN) algorithm can be implemented by iteratively calling SAMMY in this mode. This is an intentional option within the SAMMY code and further details can be found in the manual [8].

Alternatively, SAMMY has an option for a relative prior parameter covariance scaled by a single user input, ζ. If this option is used, each time it is called, SAMMY solves the following equation:

$$P' = P + (\text{diag}(\zeta P)^{-1} + G^T V^{-1} G)^{-1}\ G^T V^{-1}(D - T) \tag{4.14}$$

Replacing ζ with $\frac{1}{\lambda}$, this becomes identical to a LM step with a dampening factor that is weighted by the magnitude of each parameter[11]. By iteratively calling SAMMY, capturing the output, and dynamically varying the ζ input, the LM_b algorithm discussed in Sec. 4.2.1 can be implemented internal to SAMMY.

In both examples above, SAMMY solves the reported equation with user-controlled internal iterations to improve the linear assumptions on G. Unless stated otherwise, no internal iterations were used. This facilitates comparison to other solution methods and the implementations external to SAMMY (Sec. 4.2.4). For most evaluations, multiple datasets or measurements are available; for the demonstrations in this thesis, five are used. SAMMY offers two approaches for this: (1) sequentially fitting each dataset, using Bayes' theorem to update parameters with each new dataset while retaining information from the previous one, or (2) fitting the data simultaneously using the 'YW' solution scheme. Unless otherwise specified, the latter approach is preferred, as it is more robust to outliers and provides a straightforward maximum likelihood objective, as discussed in Sec. 4.1.

External to SAMMY

The SAMMY code includes a user option to return the derivatives of the experimentally corrected theory (T) with respect to the resonance parameters (P). These derivatives form the Jacobian matrix (G), which is crucial for any derivative-based optimization method. Extracting these derivatives from the SAMMY code enables a broader range of problems

[11] Note that, for indices of $\text{diag}(\zeta P)$ corresponding to a resonance energy parameter, values are instead set to $\frac{1}{10}$ of the broadened resonance width.

to be addressed using various approaches. Throughout this thesis, the term 'external' will refer to implementations where SAMMY is used solely for populating the G matrix and calculated D and T, while all other matrix operations are handled directly in the Python code ATARI [81].

4.3 Testing & Results

4.3.1 Iterative Solvers

For the iterative, non-linear optimization method, the goal is to minimize the regression objective for each level of model complexity. The study in Sec. 3.3.3 indicates that the best regression objective is the generalized least squares objective with the data covariance matrix calculated at the theory, χ^2_{T}. Using the same 500 synthetic data realizations in the 200–220 eV window, the GD, $\mathrm{LM_a}$, $\mathrm{LM_b}$, and SGD method were tested for efficiency and robustness by fitting the χ^2_{T} regression objective from the dense initial feature bank, then following the backwards stepwise regression path down to zero resonances. For a fair comparison, the results for each method were generated using the external implementation and identical hyperparameter settings where possible.

Since robustness is evaluated relative to the regression objective, synthetic data is not strictly necessary. In principle, this study could be conducted using real data across the 200–220 eV range, or across multiple windows at different energies. However, the synthetic data framework employed here provides a controlled experimental setup with sufficient samples to ensure statistical significance.

The results of this study are summarized in Fig. 4.2; however, the results from the GN solver are not included, as this algorithm struggled to converge. As expected, the GD algorithm is the least robust. At higher model complexities, SGD appears to be the most robust method but is outperformed by both LM variations as the true model complexity is approached. Although initially surprising, this can be explained by the stochastic nature of the SGD algorithm. SGD is generally effective at avoiding local minima and excels at finding the *region* of the global minimum. However, it struggles to reach the precise minimum

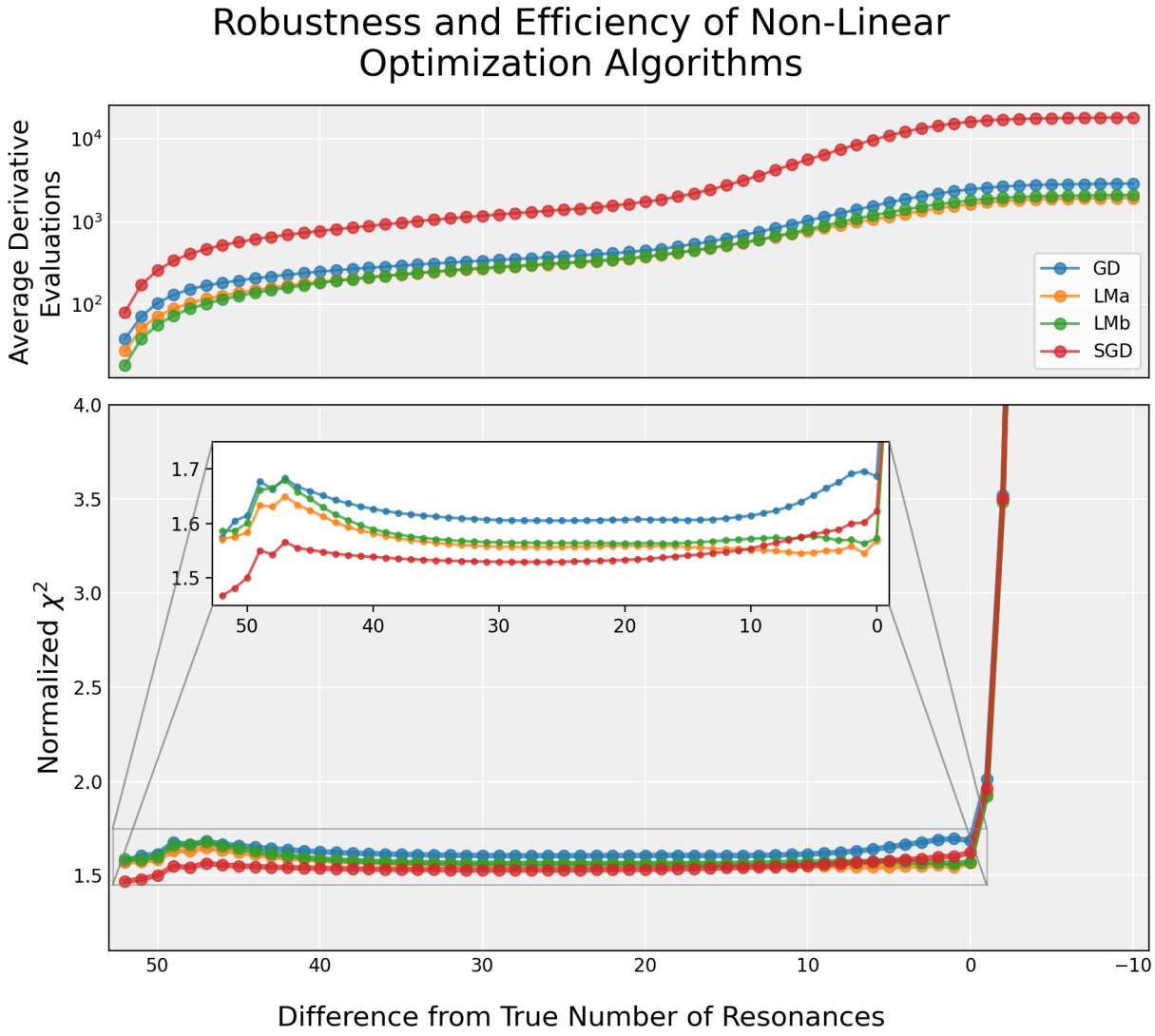

Figure 4.2: Comparison of the robustness and efficiency for different non-linear optimization algorithms—gradient descent (GD), Levenberg-Marquardt variations (LMa and LMb), and stochastic gradient descent (SGD)—solving the selected regression objective function for 500 synthetic data samples in the 200–220eV domain. Results are shown starting from the initial feature bank and through the stepwise variable selection, the horizontal axis represents the stepwise path through model complexity in difference to the true model complexity (number of resonances). The average derivative evaluations are shown cumulatively with respect to the stepwise regression and averaged over the 500 samples. The regression objective is averaged over the number of data points and the 500 samples.

because it relies on stochastic samples of the overall objective function. In contrast, the inclusion of a Hessian approximation in the LM algorithm is known to enhance speed and precision of convergence as the algorithm approaches a minimum (global or local). This is further reflected by the derivative evaluations, indicating that the LM variations are the most efficient. The poor efficiency of SGD is also expected as the stochastic path takes more steps to converge. As discussed in Sec. 4.2.1, the efficiency of SGD is primarily related to memory usage, as evaluating the overall objective function may not be feasible in certain applications.

4.4 Scaling

The methods discussed earlier should theoretically work for problems of any size, but they encounter challenges as the energy window increases. Over a range of several thousand eV, the initial feature bank may contain thousands of resonances per spin group. This presents two main challenges. First, there is the computational cost of calculating the theory and derivatives for this many resonances and data points. This is handled by SAMMY, which efficiently manages this by truncating the derivatives for distant resonances. Parallelization is beneficial in terms of wall time; however, in terms of computational time, the improvement is minimal. The primary limitation is the stepwise variable selection routine as the number of models that must be fitted scales as $1 + \frac{p(p+1)}{2}$. This is addressed by decomposing the problem into smaller overlapping windows.

The approach is visualized in Fig. 4.3. In the first stage, overlapping windows are defined by their energy domain, with the measurement data truncated accordingly. A buffer is added on either side of each window, allowing the initial feature bank to extend beyond the window and represent resonances that may influence the model within it (shown by the dotted line). Each window in stage 1 is solved independently, allowing for parallelization. The resulting resonances are then combined, often leading to duplicates in the overlapping regions. In stage 2, windows that bound these overlapping regions are defined by a new energy domain (shown in red). However, in this stage, the initial feature bank is based on the output from stage 1, and variable selection is performed only for resonances that fall within the overlapping regions

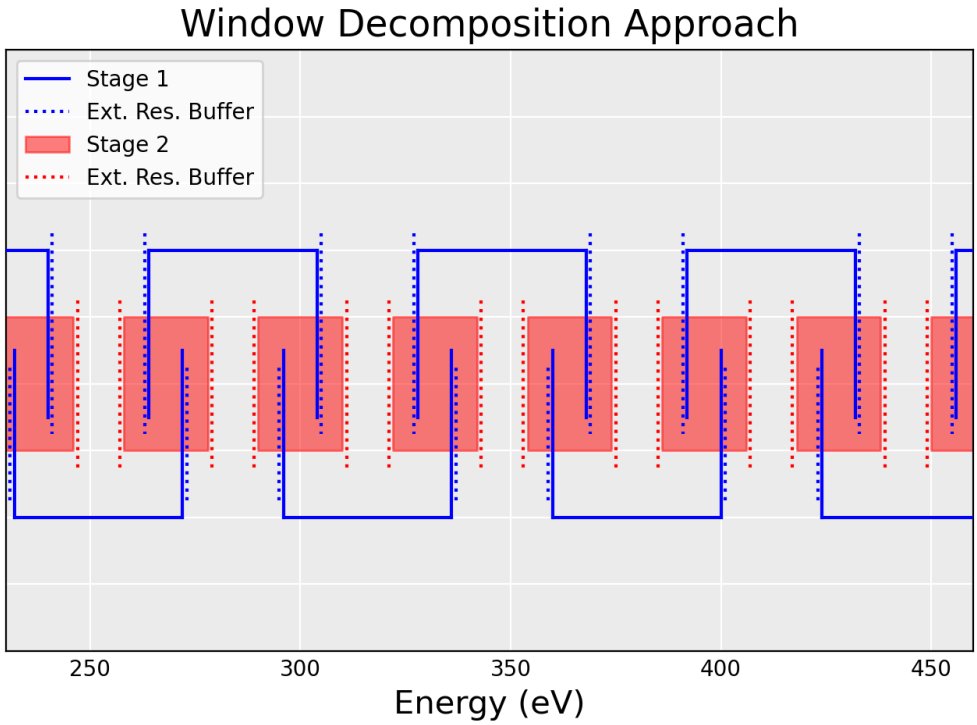

Figure 4.3: Window decomposition scheme for automated fitting of large energy domains.

identified in the previous stage. If the stage 2 windows overlap with other stage 2 windows, a 3rd iteration is performed and so on. The sizes of the windows, overlapping regions, and external resonance buffers are all user-defined parameters, with specific conditions in place to ensure that the overlapping regions and external resonance buffers do not exceed appropriate limits for the specified window size. For the subsequent studies in this thesis, if window decomposition was used the parameters are: stage 1 window = 11.5 eV, stage 1 window overlap = 3/11.5 eV, stage 2 window = 6.4 eV, external resonance buffer = 1.0 eV.

Chapter 5

Final Results

5.1 Performance Testing Framework

By leveraging the synthetic data framework, the automated fitting algorithm's performance can be characterized, enhanced, and benchmarked. This process mirrors the iterative solver study in Sec. 4.3, where the comparison combined efficiency and robustness relative to the surrogate regression objective, χ^2. In this case, however, the focus shifts to a metric that directly assesses performance based on the true cross section. These performance metrics are described in Sec. 3.3.2 and applied to determine the optimal surrogate regression objective in Sec. 3.3.3. As established in the previous sections of this thesis, the best-performing iterative solver and the optimal regression objective are utilized in all subsequent studies in this section unless otherwise specified.

5.1.1 Model Selection approach

The two model selection approaches described in Sec. 4.2.3 are compared over 500 synthetic data realizations in the 200–220 eV window. The $\Delta\chi^2$ approach provides a correction to traditional information-theoretic criteria and requires selecting the $\Delta\chi^2$ threshold as a hyperparameter. There is some intuition regarding the optimal threshold, as the true model is expected to have $\chi^2/N \approx 1$, where N is the number of data points. However, this serves only as a ballpark estimate, and it remains uncertain whether it is better to favor simpler or

more complex models. The synthetic data framework facilitates hyperparameter selection through a supervised learning approach, where the optimal hyperparameter is determined by comparing the results to the true solution cross section using the performance metrics. This is shown in Fig. 5.1, where the $\Delta\chi^2$ hyperparameter is varied across a range of values, and the RMSE, MAE, and MAPE performance metrics are monitored. Both RMSE and MAE suggest an optimal $\Delta\chi^2$ of approximately 0.05, whereas MAPE indicates a value of 0.1. This discrepancy is likely due to MAPE's greater sensitivity to misplaced resonances rather than missed ones, as the denominator, $\sigma_{\text{true}}^{\text{capture}}$ can approach zero. As a result, the optimal $\Delta\chi^2$ tends to be larger, favoring simpler models.

In contrast, k-fold cross-validation (CV) does not require a hyperparameter to be learned within the synthetic data framework, which is a key advantage of this approach. This eliminates the risk of learning a suboptimal hyperparameter from flawed synthetic data and enables more confident application to real data. Moreover, the results in Fig. 5.1 indicate that CV outperforms even the optimal choice for $\Delta\chi^2$ across all performance metrics. The CV approach appears to be more sensitive to the specific data set (sample) at hand, whereas the $\Delta\chi^2$ hyperparameter must be optimal across all samples.

5.1.2 Performance Characterization

Taking the CV-selected model for each case, further performance characterization can be conducted to quantitatively benchmark the overall automated fitting approach's ability to accurately estimate resonance cross sections. Additionally, this section tests the window decomposition method detailed in Sec. 4.4 by comparing the performance of the 20 eV window solved directly to it solved across 3 decomposed windows.

Performance metrics are reported in Table 5.1. Results are close to the minimum achievable (given by FFT) for RMSE and MAE and nearly identical between the direct and decomposed solves. In fact, the decomposed solve gives slightly better performance for these metrics. The two approaches are not necessarily identical. The larger window will have more information about neighboring resonances and interference effects and therefore generally provide spin assignments that better fit the data. On the other hand, cross validation for model selection will tend towards more complex models in the smaller window due to the

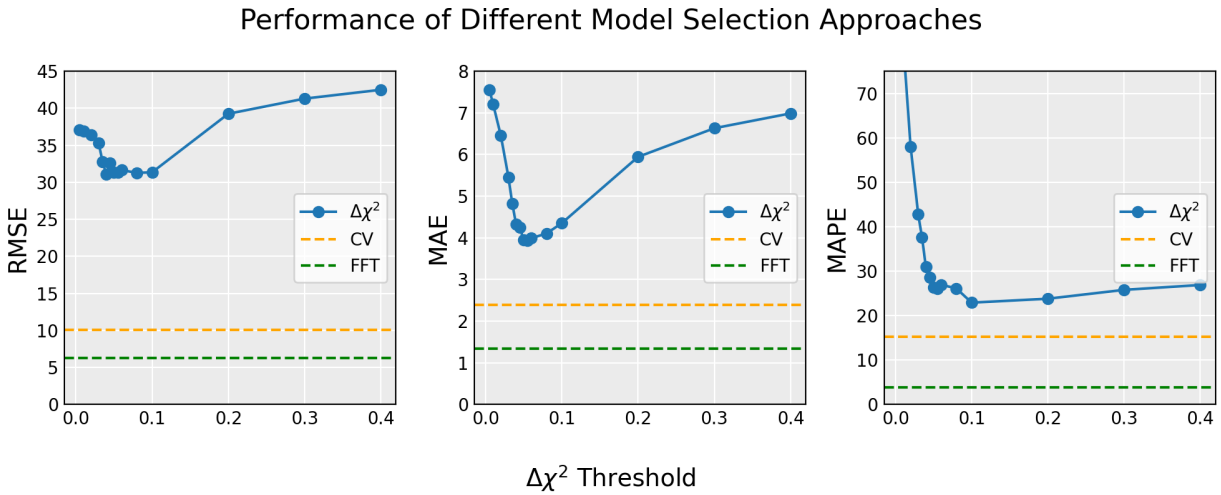

Figure 5.1: Various performance metrics for $\Delta\chi^2$ and k-fold CV approaches to model selection for 500 synthetic data samples in the 200–220 eV window. The x-axis shows the varied $\Delta\chi^2$ hyperparameter values and the optimal occurring around 0.05. The performance of the CV approach is represented by the orange line, while the theoretical minimum achievable performance, limited by the surrogate regression objective, is indicated by the green line labeled FFT (fit-from-true).

Table 5.1: Performance metrics for automated fitting (AutoFit) over 500 samples in the 200–220 eV energy domain both with and without using window decomposition (AutoFit and AutoFit Decomp. respectively). Corresponding fit from true (FFT) results are shown as a reference for the best achievable given the limits of the surrogate regression objective. Additional statistics are reported showing on the average total number of resonances in the window ($\mathrm{N_{res}}$) and the average number of resonances in the window with neutron widths above the 20th quantile of the Porter-Thomas distribution ($\mathrm{N_{res}} > Q_{0.2}$).

Fitting Approach	RMSE (bn)	MAE (bn)	MAPE (%)	$\mathrm{N_{res}}$	$\mathrm{N_{res}} > Q_{0.2}$
FFT	6.68	1.40	3.60	4.65	3.59
AutoFit	10.64	2.46	14.80	3.66	3.39
AutoFit Decomp.	10.35	2.29	19.50	4.06	3.46

application of the one standard error rule[1]. In this case, the benefit of increased mdoel complexity outweighs the benefit of capturing neighboring interference effects in both the RMSE and MAE metric. As previously discussed, the MAPE will more harshly penalize a misplaced resonance than a missed resonance due to the denominator $\sigma_{\text{true}}^{\text{capture}}$ approaching zero. This corresponds to a deterioration in performance for the decomposed solve. The correlation between each metric is shown in Fig. 5.2. RMSE and MAE are strongly correlated while MAPE is only loosely correlated to the other two metrics and has large outliers.

As discussed in Sec. 3.3.2, directly comparing resonance parameters is not straightforward. However, high-level statistics regarding parameter estimation can still be extracted from the data. For example, Table 5.1 also reports the average number of resonances in the window per sample (N_{res}). This is also reported with N_{res} filtered to those with neutron widths above the 20th quantile of the Porter-Thomas distribution ($\text{N}_{\text{res}} > Q_{0.2}$).

5.1.3 Start of the URR

Section 5.1.2 benchmarked the automated fitting algorithm's performance in 200–220 eV window and demonstrated that the performance of a single window aligns closely with the performance of the finalized results after the larger problem has been decomposed and subsequently reconstructed. Here, the impact of deteriorating experimental resolution on performance will be quantified by sweeping the 20 eV window across energies of the RRR and into the URR. This investigation aims to quantitatively aid in the determination of the end of the RRR as performance metrics are expected to degrade with the experimental resolution.

The distribution of RMSE_{s} is shown in Fig. 5.3 along with distributions of error in the total number of resonances ($N_{\text{fit}} - N_{\text{true}}$) and percent error in energy-integrated transmission, $\langle T \rangle$. As expected, more resonances are missed as you move up in energy. However, a deterioration of RMSE proportional to the deterioration of the experimental resolutions is not seen. In fact, the RMSE improves in the middle energy regions of the RRR, decreases around the URR boundary, then slightly improves well beyond the URR. Moreover, the distributions themselves remain largely unchanged; instead, a few outlying cases appear to

[1]This phenomenon is explored further in Appendix A.3

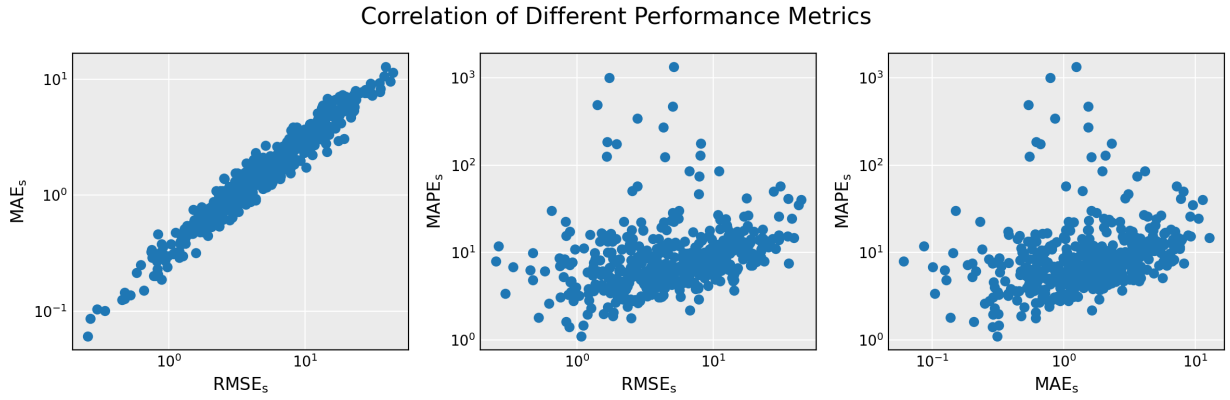

Figure 5.2: Correlation between different performance metrics.

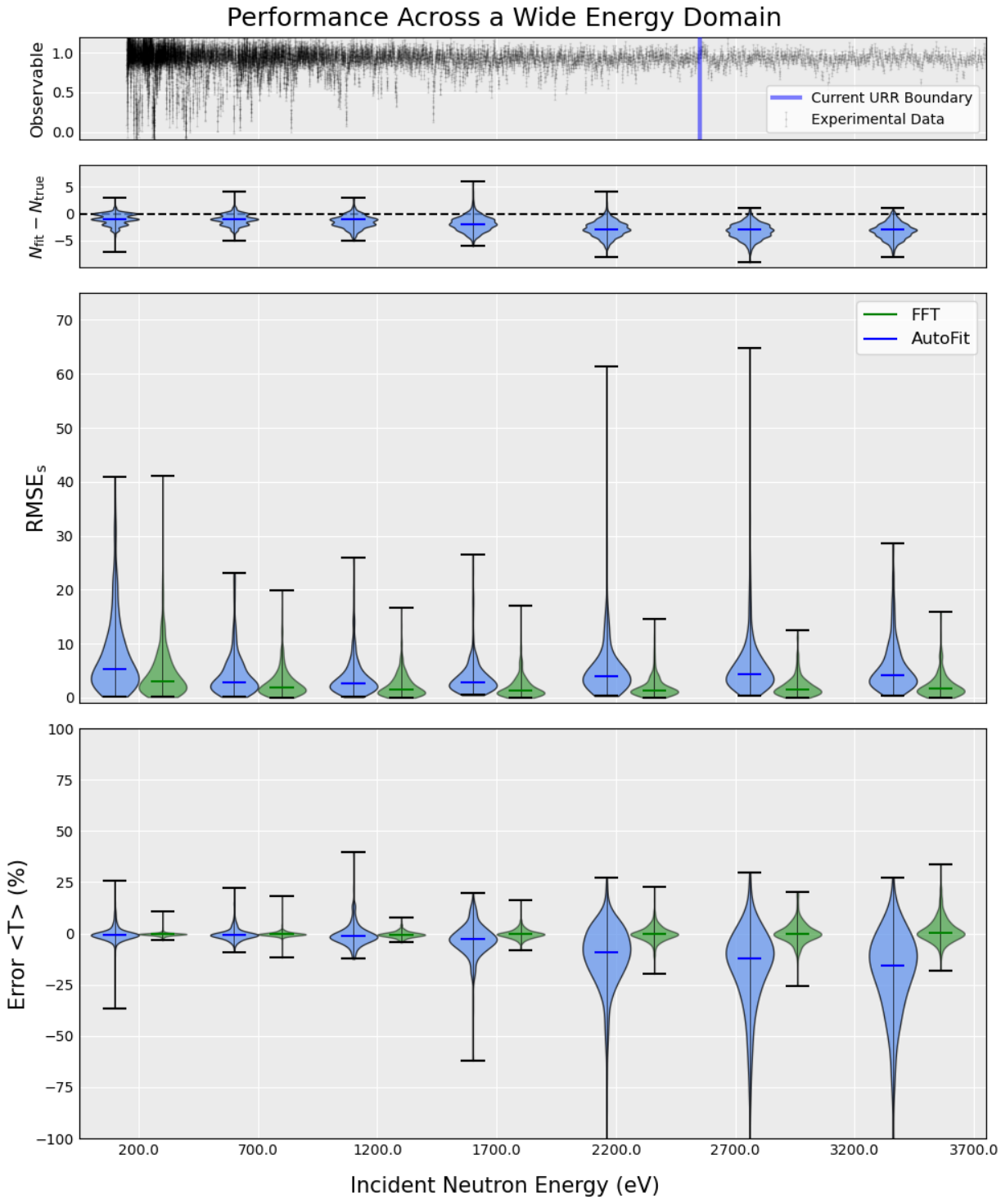

Figure 5.3: Distribution of error in number of resonances, RMSE performance metric, and percent error in energy averaged transmission for a 20 eV window at different incident neutron energies over 500 synthetic data samples. Light blue indicates results for the automated fitting methodology while green indicates the fit from true (FFT) minimum achievable performance.

be driving the increase in RMSE. This suggests that there is more to consider than the experimental resolution alone. Experimentally, this could be differences in the background or flux spectra. Another consideration is that the resonance penetrability increases at higher energies, making the resonances broader and less peaked. Therefore, even if resonances are poorly estimated or missing, the RMSE becomes an easier target. This is supported by the near constant improvement in the RMSE objective for fit from true (FFT) as the window moves to higher energies.

A major motivation for ending the RRR and starting the URR is that it is typically assumed that if too many resonances are missed, simulations won't accurately reproduce self-shielding effects (see Sec. 2.1.1) because the resonance evaluation will no capture all of the variance in the cross section. This can also be quantified within this framework, as shown by the percent error in $\langle T \rangle$. This is meant to represent a general application, and is calculated as

$$100 \times \frac{\langle T_{\text{true}} \rangle - \langle T_{\text{fit}} \rangle}{\langle T_{\text{true}} \rangle}, \tag{5.1}$$

where

$$\langle T \rangle = \int \exp(-n\sigma_{\text{tot}}(E))dE, \tag{5.2}$$

and n is a dynamically selected via a grid search between 10^{-5} and 1.0 to produce the maximum error for a given σ_{true} and σ_{fit}. This metric tells a somewhat different story as the deterioration is more apparent and the energy integrated transmission starts to be overestimated well before the current URR boundary.

While these results do not indicate a clear onset of the URR, the quantitative analysis will help justify the evaluator's decision. The contrast between the RMSE metric and the error in $\langle T \rangle$ suggests that if the variance in the cross section could be captured even with poorly estimated resonances, the reconstruction of the pointwise cross section may still be accurate. Moreover, if this methodology were extended to include models of the URR, a more informed transition point could be identified where the error metrics for the Hauser-Feshbach treatment and the R-matrix treatment intersect.

5.2 Analysis of Actual Data

The automated resonance inference methodology presented in this thesis was demonstrated on the five sets of actual ^{181}Ta measurement data taken at Rensselaer Polytechnic Institute [28, 82]. These measurements correspond to the five generative models detailed in Sec. 3.3.1, therefore, the performance seen in Sec. 5.1.2 can be considered a statistical description of the performance on the actual data.

To corroborate the performance characterization, results are compared to the recent ENDF/B-VIII.1 evaluation of ^{181}Ta. In the energy domain 200–3000 eV, the ENDF/B-VIII.1 evaluation used the measurement data used here and only one additional transmission measurement. Therefore, similar results are expected, perhaps with a slightly worse χ^2 fit to the data. The AutoFit methodology was allowed to vary the capture width at the final step by just 2% from the average to match what was done in the ENDF/B-VIII.1 evaluation.

Figure 5.4 shows the cumulative levels as a function of energy for both evaluations across the 200–3000 eV energy domain. The JEFF-3.3 evaluation is also shown for comparison, though it did not have access to the measurement data used here. Good agreement is seen for the combined levels until around 700 eV. Beyond 700 eV, the ENDF/B-VIII.1 evaluators recognized the possibility of missing increasingly large resonances as experimental resolution deteriorates and started introducing 'fake' resonances to better match the expected average level spacing. This fake-resonance evaluation technique is not incorporated into the automated methodology, leading to discrepancies. Even in the absence of these fake resonances, AutoFit favors simpler models, as shown in the rightmost column of Table 5.2 and by the non-linear behavior in Fig. 5.4 around 800 and 1700 eV. Other than these two 'bends', the expected constant slope is seen in the AutoFit results.

Splitting the levels by spin assignment reveals significant discrepancies between AutoFit and ENDF/B-VIII.1 and is shown in Fig. 5.5. These differences arise from the distinct approaches to spin assignment. The automated methodology considers only the fit to the measurement data, whereas the ENDF/B-VIII.1 evaluation considers only the agreement with the resonance parameter distributions. The latter approach is acceptable in part because ^{181}Ta has only two spin groups, both of which are s-wave and exhibit very similar

Table 5.2: Total number of resonances identified above 200 eV and below the energy specified in the column header.

Evaluation Approach	Number of Resonances		
	< 700 eV	< 2550 eV	< 2550 eV (no fake)
JEFF-3.3	109	453	453
ENDF/B-VIII.1	109	514	455
AutoFit	106	448	448

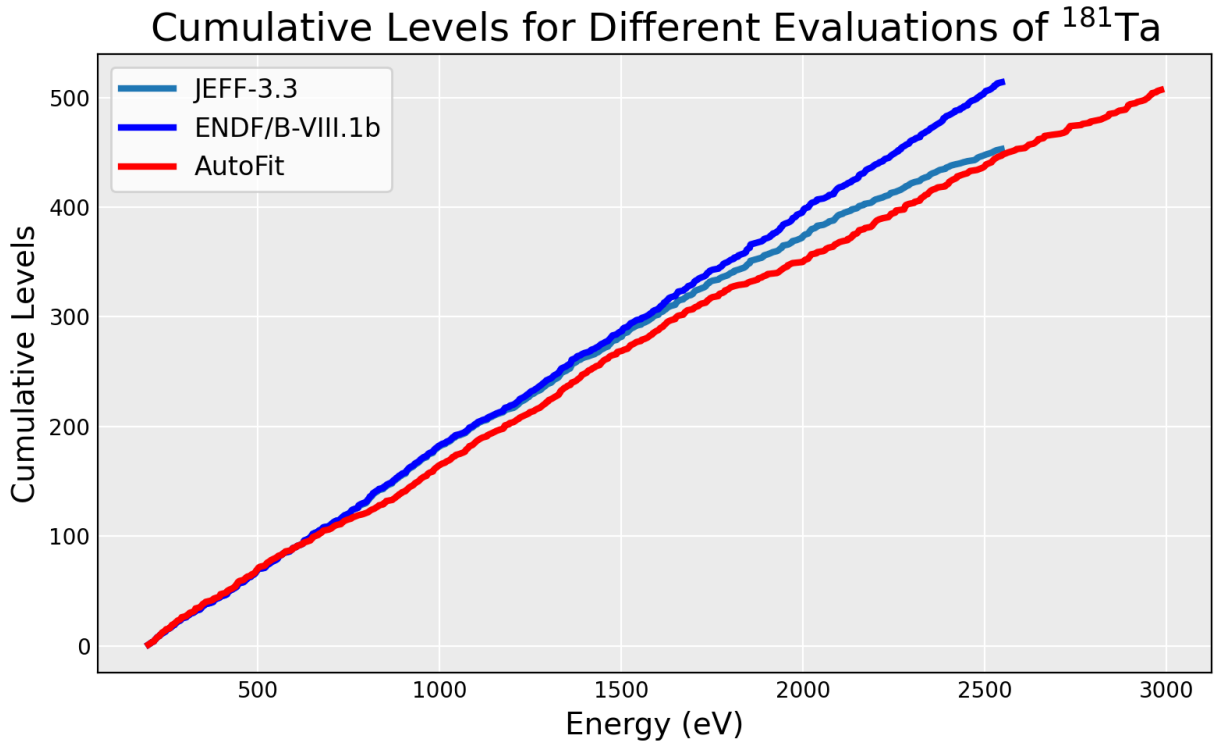

Figure 5.4: Cumulative resonance levels across the RRR of the actual measurement data. Results are shown for JEFF-3.3 and ENDF/B-VIII.1 evaluations and for the automated approach developed in this thesis.

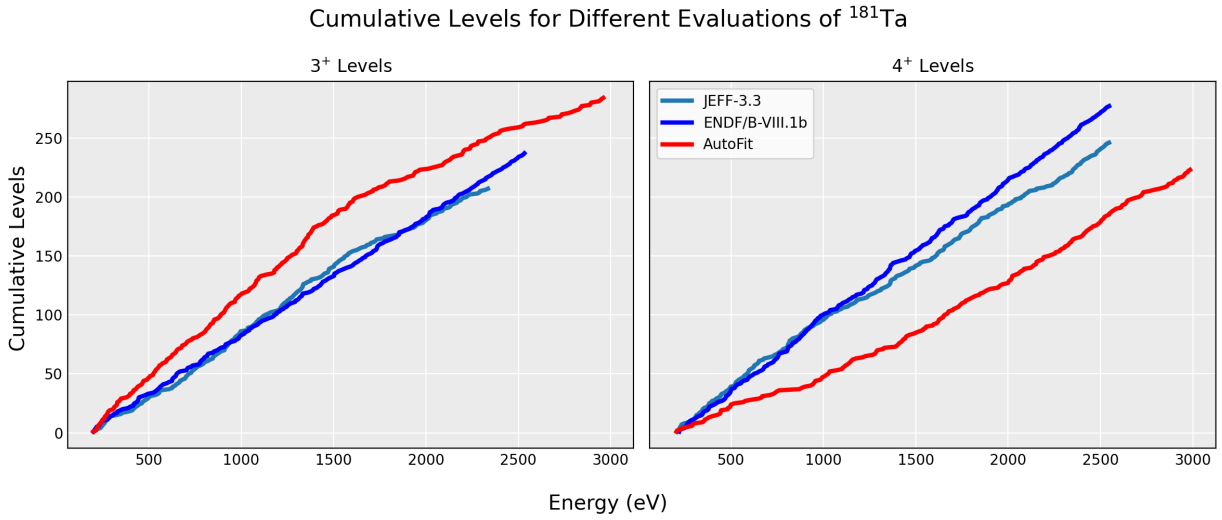

Figure 5.5: Cumulative resonance levels for each spin group across the RRR of the actual measurement data. Results are shown for JEFF-3.3 and ENDF/B-VIII.1 evaluations and for the automated approach developed in this thesis.

shapes. Consequently, the automated approach extracts limited information about resonance spin assignments solely from the measurement data. However, the data do exhibit some sensitivity to spin assignments, as indicated by the higher χ^2 in ENDF/B-VIII.1 as compared to AutoFit (reported in Table 5.3), suggesting that measurement data should perhaps be considered when determining spin assignments even in such cases where it is traditionally assumed not to matter.

While the estimated number of levels in AutoFit begins to deviate from ENDF/B-VIII.1, the quality of the fit to the data remains consistent. This is illustrated by the normalized residuals in Fig. 5.6. Excellent agreement is observed between the two evaluations until around 1500 eV. Beyond 1500 eV, there is good agreement in the overall spread other than in a few places where there are seemingly non-statistical features in the AutoFit residuals as compared to ENDF/B-VIII.1. This is expected to some degree as the AutoFit approach favors more parsimonious models, possibly involving larger but less probable resonance widths.

Table 5.3: Chi2 comparison between different fitting methods. Reported χ^2 is calculated for data below 2550 eV as the ENDF/B-VIII.1 evaluation stops identifying resonances beyond this. Despite the additional dataset used in the ENDF/B-VIII.1 evaluation, the normalized χ^2 value should not change significantly as long as the data are not discrepant. The reported *ENDF/B-VIII.1 Fit* method takes the ENDF/B-VIII.1 evaluation and fits it to these 5 data sets with the same convergence criteria as AutoFit to remove ambiguity about the impact of the additional dataset used.

Fitting Method	χ^2/N_{data}
AutoFit	1.358
ENDF/B-VIII.1	1.704
ENDF/B-VIII.1 Fit	1.581

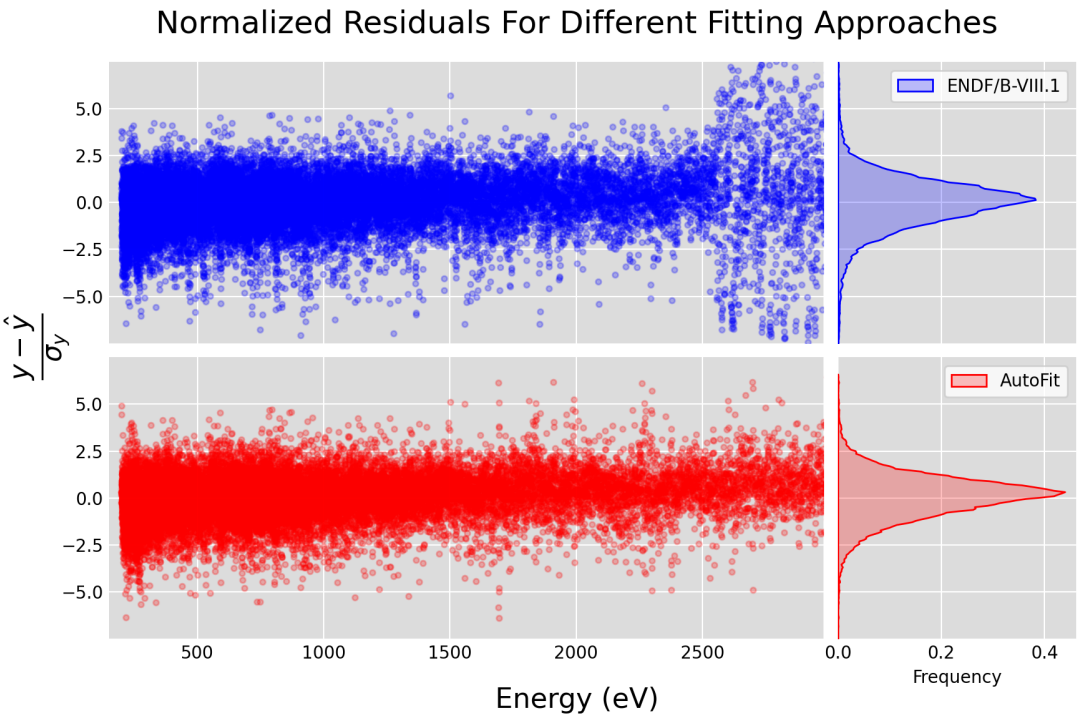

Figure 5.6: Normalized residuals, calculated as the difference between data and fit divided by the data uncertainty for the actual measurement data. Results are shown for the ENDF/B-VIII.1 evaluation and the automated approach developed in this thesis.

Chapter 6

Conclusion

This thesis makes two major methodological contributions to the field of nuclear data evaluation that complement each other, though each can stand independently. The first is a methodology for the automated identification and characterization of neutron-induced resonant cross sections in experimental data. Fully automating the resonance inference process not only saves valuable time for evaluators and analysts but also enhances reproducibility and reportability within the evaluation process. The second contribution is a computational framework that leverages high-fidelity generative modeling to test, validate, and benchmark inferential methods. As demonstrated in this thesis, integrating the automated evaluation approach within this framework allows for statistical characterization of performance across numerous realizations, which can be used to inform evaluator decisions, bound uncertainty, and test inherent assumptions.

The generative modeling methodology provides a way to generate high-utility, synthetic realizations of cross section measurement data in the resolved resonance range (RRR). This is a multi-step process requiring: 1) a generative model for true underlying resonance parameter realizations, 2) a generative model for the process and noise associated with the measurement of raw observables that are influenced by the true underlying resonance parameters. In order to generate high-utility data, both generative models must produce datasets that are statistically indistinguishable from real or observed data. For resonance parameter generation, the utility of the data is driven by the statistical theory of resonances and a rigorous analysis is outside the scope of this thesis. The focus here is primarily on

the development of the second generative model for two measurement types, which is more relevant for subsequent studies.

The methodology for high-utility transmission data was fully developed and validated both heuristically and rigorously. To this end, the decision to set the generative model stopping point at cycle-separated detector counts was sufficient. The stopping point for the capture yield data is set to be the same; however, this results in medium-utility synthetic data as the detector response characterization is not included. This was deemed sufficient for the studies in this thesis; however, an expansion of the generative modeling methodology for capture yield data would enable impactful studies on the assumptions and uncertainty quantification associated with this type of measurement data. For both transmission and capture yield models, the implementation was numerically verified.

One of the key challenges in developing these generative models is that accurately modeling the experiment requires detailed knowledge of the setup. While access to this information is ideal and often necessary for evaluation anyway, intentionally hypothesizing uncertainties and/or correlations and excluding them from the evaluation can provide insight into the impact of unquantified uncertainties and the evaluation's robustness in handling them. Paired with recent work on templates for measurement uncertainty, which provides standardized estimates for various experimental uncertainties, this approach could address significant challenges in RRR evaluation, representing a natural extension and application of the framework presented in this thesis.

The automated fitting methodology addresses three key technical challenges in resonance parameter inference: unknown cardinality, non-convexity, and non-linearity. This approach can be divided into four main components: initial feature bank engineering, iterative optimization, variable selection, and model selection. While the initial feature bank engineering is domain-specific, the other components rely on well-established techniques from regression analysis. Integrating the development of the automated fitting approach with the synthetic data-based computational testing framework allowed for a comparison of various methods.

The Levenberg-Marquardt (LM) algorithm for iterative non-linear optimization proved to be the most effective. This is especially convenient, as LM can be integrated directly into

the resonance analysis code SAMMY, facilitating its immediate adoption into evaluators' workflows. The k-fold cross-validation (CV) approach significantly outperformed the $\Delta\chi^2$ method for model selection. This is preferable because it makes the automated fitting process independent of the synthetic data framework. In contrast, the $\Delta\chi^2$ approach requires high-utility synthetic data to determine the optimal hyperparameter, which may not always be available. While CV is ideal, it cannot always be performed due to limited access to independent data. In theory, principal component decomposition of the data covariance matrix (V) would enable any dataset to be split into independent subsets, allowing k-fold CV for any collection of data and any k. However, this approach is no longer feasible if V changes with the parameter values—a known solution to the issue of Peele's Pertinent Puzzle in nuclear data evaluation.

A limitation of this methodology arises from the backward stepwise variable selection approach. Firstly, this method is not guaranteed to discover the global minima for any given model complexity. A solution to this challenge is best-subset selection; however, its practicality diminishes rapidly with an increasing number of resonances. Transitioning from best-subset to stepwise represents a shift toward a more greedy algorithm for variable selection, that is, trading optimality and robustness for efficiency. Even within the stepwise algorithm, there exist variations in greediness. Empirical observations during the development of this algorithm reveal the following insight: the level of greediness tends to influence the models encountered at high complexities, particularly when many of the resonance features are merely fitting noise in the data; however, as the selection algorithm progresses toward more reasonable model complexities (as indicated by model selection criteria and *a priori* knowledge about resonance level spacing), different levels of greediness converge to nearly identical models, irrespective of the path taken at higher complexity. As the energy window of interest expands, even the more aggressive versions of the stepwise algorithm face computational limitations. This issue is effectively addressed by the window decomposition approach discussed in Sec. 4.4, where it was demonstrated that the performance after reconstruction remains nearly identical.

The finalized automated fitting algorithm can be easily generalized to other measurement data and/or other isotopes. The only change to the algorithm lies in the design of the initial

feature bank, for which an evaluator should have strong intuition (i.e., rough approximation of level density, number of spin groups present, desired precision for non-linear optimization). The design of the initial feature bank also provides a straightforward way to integrate existing evaluator workflows or additional knowledge into the algorithm. For instance, *a priori* known resonances can be incorporated into the initial feature bank and held fixed throughout the algorithm—that is, excluded from variable and model selection routines. While the automated fitting algorithm does not require the synthetic data testing framework, the studies in this thesis demonstrate how the evaluation could benefit from the framework as a whole. By leveraging the synthetic data-based computational framework, virtually infinite sets of 'training' data can be generated for testing. Although the automated fitting algorithm does not actually 'learn' from this training data (except in the case of tuning the $\Delta\chi^2$ hyperparameter), it is valuable for development and decision making. For instance, the window decomposition approach used to scale the analysis to a domain of thousands of eV was successfully validated. In the approach to the unresolved resonance region (URR) study, the degradation of performance due to experimental resolution was quantified, allowing the evaluator to make an informed decision about the end of the RRR. Additionally, this study raises an important question: What is the ultimate objective of an evaluation?

The RMSE between the true and estimated Doppler-broadened constituent reaction cross sections was chosen as the ultimate objective of this work. This choice was made primarily because the Doppler-broadened cross section is the actual value used in most applications. Additionally, it is not clear how to compare resonance parameters directly between models of different cardinalities. This choice impacts the interpretation of the results in the URR study as the degradation of RMSE is not as expected even though more and more resonances are missed. This is due to the fact that the automated fitting algorithm is selecting more parsimonious models with wider resonances rather than many small resonances and the RMSE does not seem to be very sensitive to this. However, the error in energy-integrated transmission was shown to be sensitive to this as self-shielding effects come into play. Future developments of the automated fitting algorithm could explore incorporating likelihood with respect to parameter distributions into the regression objective to further improve RMSE at high energies and introduce the variance in the cross section necessary to accurately capture

self-shielding. The synthetic data-based testing framework could facilitate a quantitative comparison of performance with any additions to the algorithm.

Overall, the combined methodologies presented in this thesis give a foundation for an automated evaluation framework, reducing the manual workload and improving the reproducibility associated with analyzing and evaluating experimental resonance data. This automation does not replace an evaluator's responsibility to carefully curate experimental data to be used in an evaluation. However, it offers a means to eliminate potential biases, enhance uncertainty quantification, and test inherent assumptions by using the generative modeling methodologies. This thesis focuses on the evaluation of RRR cross sections for non-fissionable isotopes. Three natural extensions of this work are: (1) developing a generative model for fission measurements—which would closely align with improving the capture yield model, (2) extending the automated fitting tool to high-energy cross sections that use the Hauser-Feshbach or Optical Model, and (3) developing and testing uncertainty quantification methods within the computational framework for validation pioneered in this thesis. Ultimately, the fundamental concepts presented here aim to establish a general framework for algorithm development, uncertainty quantification, and computational experimentation.

Bibliography

[1] A. J. M. Plompen, O. Cabellos, C. De Saint Jean, M. Fleming, A. Algora, M. Angelone, P. Archier, E. Bauge, O. Bersillon, A. Blokhin, F. Cantargi, A. Chebboubi, C. Diez, H. Duarte, E. Dupont, J. Dyrda, B. Erasmus, L. Fiorito, U. Fischer, D. Flammini, D. Foligno, M. R. Gilbert, J. R. Granada, W. Haeck, F. J. Hambsch, P. Helgesson, S. Hilaire, I. Hill, M. Hursin, R. Ichou, R. Jacqmin, B. Jansky, C. Jouanne, M. A. Kellett, D. H. Kim, H. I. Kim, I. Kodeli, A. J. Koning, A. Yu. Konobeyev, S. Kopecky, B. Kos, A. Krása, L. C. Leal, N. Leclaire, P. Leconte, Y. O. Lee, H. Leeb, O. Litaize, M. Majerle, J. I. Márquez Damián, F. Michel-Sendis, R. W. Mills, B. Morillon, G. Noguère, M. Pecchia, S. Pelloni, P. Pereslavtsev, R. J. Perry, D. Rochman, A. Röhrmoser, P. Romain, P. Romojaro, D. Roubtsov, P. Sauvan, P. Schillebeeckx, K. H. Schmidt, O. Serot, S. Simakov, I. Sirakov, H. Sjöstrand, A. Stankovskiy, J. C. Sublet, P. Tamagno, A. Trkov, S. van der Marck, F. Álvarez Velarde, R. Villari, T. C. Ware, K. Yokoyama, and G. Žerovnik. The joint evaluated fission and fusion nuclear data library, JEFF-3.3. *European Physics Journal A*, 56(181), 2020. x, 1, 20, 46, 124

[2] D. A. Brown, M. B. Chadwick, R. Capote, A. C. Kahler, A. Trkov, M. W. Herman, A. A. Sonzogni, Y. Danon, A. D. Carlson, M. Dunn, D. L. Smith, G. M. Hale, G. Arbanas, R. Arcilla, C. R. Bates, B. Beck, B. Becker, F. Brown, R. J. Casperson, J. Conlin, D. E. Cullen, M. A. Descalle, R. Firestone, T. Gaines, K. H. Guber, A. I. Hawari, J. Holmes, T. D. Johnson, T. Kawano, B. C. Kiedrowski, A. J. Koning, S. Kopecky, L. Leal, J. P. Lestone, C. Lubitz, J. I. Márquez Damián, C. M. Mattoon, E. A. McCutchan, S. Mughabghab, P. Navratil, D. Neudecker, G. P. A. Nobre, G. Noguere, M. Paris, M. T. Pigni, A. J. Plompen, B. Pritychenko, V. G. Pronyaev, D. Roubtsov, D. Rochman, P. Romano, P. Schillebeeckx, S. Simakov, M. Sin, I. Sirakov, B. Sleaford, V. Sobes, E. S. Soukhovitskii, I. Stetcu, P. Talou, I. Thompson, S. van der Marck, L. Welser-Sherrill, D. Wiarda, M. White, J. L. Wormald, R. Q. Wright, M. Zerkle, G. Žerovnik, and Y. Zhu. ENDF/B-VIII.0: The 8th major release of the nuclear reaction data library with cielo-project cross sections, new standards and thermal scattering data. *Nuclear Data Sheets*, 148:1–142, 2018. Special Issue on Nuclear Reaction Data. 1, 20, 37

[3] L. A. Bernstein, D. A. Brown, A. J. Koning, B. T. Rearden, C. E. Romano, A. A. Sonzogni, A. S. Voyles, and W. Younes. Our future nuclear data needs. *Ann. Rev. Nuc. Part. Sci.*, 69:109–136, 2019. 1

[4] E. Dupont, M. B. Chadwick, Y/ Danon, C. De Saint Jean, M. Dunn, U. Fischer, R. A. Forrest, T. Fukahori, Z. Ge, H. Harada, et al. Working party on international nuclear data evaluation cooperation (WPEC). *Nuclear Data Sheets*, 120:264–267, 2014. 2

[5] WPEC Subgroup 49 (SG49). Reproducibility in Nuclear Data Evaluation [online]. https://oecd-nea.org/download/wpec/sg49/index.htm, 2020. [cited 09/01/2024]. 2, 22

[6] F. H. Fröhner. Evaluation and analysis of nuclear resonance data, JEFF report 18. Technical report, OECD-NEA, 2000. 3, 8, 37

[7] S. F. Mughabghab. *Atlas of Neutron Resonances: Resonance Parameters and Thermal Cross Sections Z=1-100*. Elsevier Science, 5th edition, 2006. 3, 20, 37

[8] N. M. Larson. Updated users' guide for SAMMY: Multilevel R-matrix fits to neutron data using Bayes' equations. Technical Report ORNL/TM-9179/R8, ORNL, ORNL, Oak Ridge, TN, 2008. 3, 8, 12, 17, 23, 29, 36, 38, 39, 54, 58, 75

[9] R. Capote, S. Badikov, A. D. Carlson, I. Duran, F. Gunsing, D. Neudecker, V. G. Pronyaev, P. Schillebeeckx, G. Schnabel, D. L. Smith, and A. Wallner. Unrecognized sources of uncertainties (USU) in experimental nuclear data. *Nuclear Data Sheets*, 163:191–227, 2020. 3, 12, 17, 22

[10] L. Leal. Data analysis and evaluation with SAMMY. https://www.osti.gov/biblio/755652, 2000. [cited 9/01/2024]. 3, 19

[11] H. Feshbach. The optical model and its justification. *Ann. Rev. Nuc. Part. Sci.*, 8:49–104, 1958. 5

[12] W. Hauser and H. Feshbach. The inelastic scattering of neutrons. *Phys. Rev.*, 87:366–373, Jul 1952. 5, 6

[13] A. M. Lane and R. G. Thomas. R-matrix theory of nuclear reactions. *Rev. Mod. Phys.*, 30:257–353, Apr 1958. 6

[14] P Descouvemont and D Baye. The r-matrix theory. *Reports on Progress in Physics*, 73(3):036301, feb 2010. 6

[15] I. J. Thompson and F. M. Nunes. *Nuclear Reactions for Astrophysics: Principles, Calculation and Applications of Low-Energy Reactions*. Cambridge University Press, 2009. 6

[16] C. W. Reich and M. S. Moore. Multilevel formula for the fission process. *Phys. Rev.*, 111:929–933, Aug 1958. 8

[17] C. E. Porter and R. G. Thomas. Fluctuations of nuclear reaction widths. *Phys. Rev.*, 104:483–491, Oct 1956. 8, 37, 66

[18] E. P. Wigner. Results and theory of resonance absorption. In *Proceedings of Conf. on Neutron Physics by Time-of-Flight*, pages 59–71, Gatlinburg, Tennessee, 1957. Oak Ridge National Laboratory, Physics Division. 9, 37, 63, 65

[19] R. E. MacFarlane and A. C. Kahler. Methods for processing ENDF/B-VII with NJOY. *Nuclear Data Sheets*, 111(12):2739–2890, 2010. Nuclear Reaction Data. 10, 37

[20] M. E. Dunn. Production of probability tables for the unresolved-resonance region using the AMPX cross-section processing system. Technical report, ORNL, 2001. 10, 37

[21] B. R. Beck. FUDGE: A program for performing nuclear data testing and sensitivity studies. *AIP Conference Proceedings*, 769(1):503–506, 2005. 10, 22, 36, 37

[22] D. A. Brown. A tale of two tools: mcres.py, a stochastic resonance generator, and grokres.py, a resonance quality assurance tool. 10 2018. 10, 37

[23] V. Sobes. *Coupled differential and integral data analysis for improved uncertainty quantification of the ^{63,65}Cu cross section evaluations*. PhD thesis, Massachusetts Institute of Technology, 2014. 11

[24] WPEC Subgroup 33 (SG33). Assessment of Existing Nuclear Data Adjustment Methodologies [online]. https://www.oecd-nea.org/upload/docs/application/pdf/2020-01/nsc-wpec-doc2010-429.pdf, 2010. [cited 09/01/2024]. 11

[25] N. Tsoulfanidis and S. Landsberger. *Measurement and Detection of Radiation*. CRC Press, Taylor & Francis Group, 2015. 11, 12, 16, 39

[26] D. Neudecker, B. Hejnal, F. Tovesson, M. C. White, D. L. Smith, D. Vaughan, and R. Capote. Template for estimating uncertainties of measured neutron-induced fission cross-sections. *EPJ Nuclear Sci. Technol.*, 4:21, 2018. 12, 17

[27] D. Neudecker, D. L. Smith, F. Tovesson, R. Capote, M. C. White, N. S. Bowden, L. Snyder, A. D. Carlson, R. J. Casperson, V. Pronyaev, S. Sangiorgio, K. T. Schmitt, B. Seilhan, N. Walsh, and Younes W. Applying a template of expected uncertainties to updating 239pu(n,f) cross-section covariances in the neutron data standards database. *Nuclear Data Sheets*, 163:228–248, 2020. 12

[28] J. M. Brown. *Measurements, evaluation, and validation of Ta-181 resolved and unresolved resonance regions*. PhD thesis, Rensselaer Polytechnic Institute, 2019. 12, 13, 16, 19, 20, 36, 40, 46, 57, 58, 88

[29] H. Postma and P. Schillebeeckx. *Neutron Resonance Capture and Transmission Analysis*. John Wiley Sons, Ltd, 2009. 14, 16

[30] D. B. Syme. The black and white filter method for background determination in neutron time-of-flight spectrometry. *Nuclear Instruments and Methods in Physics Research*, 198(2):357–364, 1982. 14

[31] P. Schillebeeckx, B. Becker, Y. Danon, K. Guber, H. Harada, J. Heyse, A. R. Junghans, S. Kopecky, C. Massimi, M. C. Moxon, N. Otuka, I. Sirakov, and K. Volev. Determination of resonance parameters and their covariances from neutron induced reaction cross section data. *Nuclear Data Sheets*, 113(12):3054–3100, 2012. Special Issue on Nuclear Reaction Data. 16, 18

[32] A. D. Carlson, V. G. Pronyaev, R. Capote, G. M. Hale, Z. P. Chen, I. Duran, F. J. Hambsch, S. Kunieda, W. Mannhart, B. Marcinkevicius, R.O. Nelson, D. Neudecker, G. Noguere, M. Paris, S. P. Simakov, P. Schillebeeckx, D. L. Smith, X. Tao, A. Trkov, A. Wallner, and W. Wang. Evaluation of the neutron data standards. *Nuclear Data Sheets*, 148:143–188, 2018. Special Issue on Nuclear Reaction Data. 17

[33] D. Neudecker, A. M. Lewis, E. F. Matthews, J. Vanhoy, R. C. Haight, D. L. Smith, P. Talou, S. Croft, A. D. Carlson, B. Pierson, A. Wallner, A. Al-Adili, L. Bernstein, R. Capote, M. Devlin, M Drosg, D. L. Duke, S. Finch, M. W. Herman, K. J. Kelly, A. Koning, A. E. Lovell, P. Marini, K. Montoya, G. P. A. Nobre, M. Paris, B. Pritychenko, H. Sjöstrand, L. Snyder, V. Sobes, A. Solders, and J. Taieb. Templates of expected measurement uncertainties. *EPJ Nuclear Sci. Technol.*, 9:35, 2023. 17

[34] M. C. Moxon, T. C. Ware, and C. J. Dean. Refit-2009: A least-square program for resonance analysis of neutron transmission, capture, fission and scattering data users' guide for REFIT-2009-2009-10. Technical Report ORNL/TM-9179/R8, Nuclear Energy Agency, ORNL, Oak Ridge, TN, 2010. 17

[35] C. De Saint Jean, P. Archier, E. Privas, and G. Noguere. On the use of Bayesian Monte-Carlo in evaluation of nuclear data. *EPJ Web Conf.*, 146:02007, 2017. 18, 34

[36] C. De Saint Jean, P. Archier, E. Privas, and G. Noguere. On the use of bayesian monte-carlo in evaluation of nuclear data. *EPJ Web Conf.*, 146:02007, 2017. 18

[37] J. Brown, G. Arbanas, D. Wiarda, and Holcomb A. Bayesian optimization framework forimperfect data or models. Technical report, ORNL, 2022. 18

[38] J. M. Brown, G. Arbanas, D. Wiarda, and A. Holcomb. Bayesian optimization framework for imperfect data or models. Technical Report ORNL/TM-2022/2448, ORNL, 2022. 18

[39] G. P. A. Nobre, D. A. Brown, S. J. Hollick, S. Scoville, and P. Rodríguez. Novel machine-learning method for spin classification of neutron resonances. *Phys. Rev. C*, 107:034612, Mar 2023. 19, 23, 36, 63

[40] D. P. Barry, M. T. Pigni, J. M. Brown, A. M. Lewis, T. H. Trumbull, K. H. Guber, B. J. McDermott, R. C. Block, and Y. Danon. A new Ta-181 neutron resolved resonance region evaluation. *Annals of Nuclear Energy*, 208:110778, 2024. 20

[41] D. Neudecker, O. Cabellos, A. R. Clark, M. J. Grosskopf, W. Haeck, M. W. Herman, J. Hutchinson, T. Kawano, A. E. Lovell, I. Stetcu, P. Talou, and S. A. Vander Wiel. Informing nuclear physics via machine learning methods with differential and integral experiments. *Phys. Rev. C*, 104:034611, Sep 2021. 22

[42] J. D. Hutchinson, J. L Alwin, A. R. Clark, T. E. Cutler, M. J. Grosskopf, W. Haeck, M. W. Herman, N. A. Kleedtke, J. R. Lamproe, R. C. Little, I. J. Michaud, D. Neudecker, M. E. Rising, T. A. Smith, N. W. Thompson, and S. A. Vander Wiel. EUCLID: A new approach to improve nuclear data coupling optimized experiments with validation using machine learning [slides]. 7 2022. 22

[43] D. Siefman, C. Percher, J. Norris, A. Kersting, and D. Heinrichs. Constrained bayesian optimization of criticality experiments. *Annals of Nuclear Energy*, 151:107894, 2021. 22

[44] M. Grosskopf. Using machine learning to explore, diagnose, and correct for bias in nuclear data [online]. https://www-nds.iaea.org/index-meeting-crp/CM-AI-ML-2020-12/, 2020. [cited 09/01/2024]. 22

[45] A. J. Koning and D. Rochman. Modern nuclear data evaluation with the TALYS code system. *Nuclear Data Sheets*, 113(12):2841–2934, 2012. 22

[46] G. Schnabel, H. Sjöstrand, J. Hansson, D. Rochman, A. Koning, and R. Capote. Conception and software implementation of a nuclear data evaluation pipeline. *Nuclear Data Sheets*, 173:239–284, 2021. 22

[47] A. Koning, S. Hilaire, and S. Goriely. TALYS: modeling of nuclear reactions. *The European Physical Journal A*, 59(6):131, 2023. 22

[48] A. Lewis, D. Neudecker, A. Koning, D. Barry, J. Brown, and G. Schnabel. WPEC Subgroup 50 (SG50): Developing an automatically readable, comprehensive and curated experimental nuclear reaction database. *EPJ Web of Conf.*, 284:18003, 2023. 22

[49] V. V. Zerkin and B. Pritychenko. The experimental nuclear reaction data (EXFOR): Extended computer database and web retrieval system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 888:31–43, 2018. 22

[50] D. Barry, J. Brown, A. Carlson, Y. Danon, M. Fleming, D. Foligno, M. Grosskopf, R. Haight, A. Holcomb, K. J. Kelly, A. Koning, A. Lewis, D. Neudecker, S. Okumura, N. Otsuka, B. Pritychenko, G. Schnabel, H. Sjöstrand, D. Smith, S. A. Van Der Wiel, N. A. W. Walton, and J. Wang. WPEC Subgroup 50 (SG50): Data-format requirement document for an automatically readable, comprehensive and curated experimental reaction database MEDUSAL, 2024. 22

[51] C. M. Mattoon, B. R. Beck, N. R. Patel, N. C. Summers, G. W. Hedstrom, and D. A. Brown. Generalized nuclear data: A new structure (with supporting infrastructure) for handling nuclear data. *Nuclear Data Sheets*, 113(12):3145–3171, 2012. Special Issue on Nuclear Reaction Data. 22

[52] NEA (2023). Specifications for the generalised nuclear database structure version 2.0. https://www.oecd-nea.org/jcms/pl_85822/specifications-for-the-generalised-nuclear-database-structure-version-2-0?details=true. [cited 09/01/2024]. 22

[53] D. A. Brown. ENDF-6 formats manual—data formats and procedures for the evaluated nuclear data files ENDF/B-VI, ENDF/B-VII and ENDF/B-VIII. 9 2023. 23

[54] N. A. W. Walton, O. Zivenko, W. Fritsch, J. Forbes, A. Lewis, J. Brown, and V. Sobes. Automated resonance identification in nuclear data evaluation, 2024. 23, 63

[55] N. A. W. Walton, J. Armstrong, H. Medal, and V. Sobes. Automated resonance evaluation; non-convex decomposition method for resonance regression and uncertainty quantification. *EPJ Web of Conf.*, 284:16004, 2023. 23

[56] J. L. Armstrong. *Decomposition Approach to Parametric Nonconvex Regression; Nuclear Resonance Analysis*. PhD thesis, University of Tennessee, 2021. 23

[57] P. Vicente-Valdez, L. Bernstein, and M. Fratoni. Nuclear data evaluation augmented by machine learning. *Annals of Nuclear Energy*, 163:108596, 2021. 23

[58] J. S. Speagle. A conceptual introduction to markov chain monte carlo methods, 2020. 24

[59] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, London, 2013. 24

[60] G. Heinze, C. Wallisch, and D. Dunkler. Variable selection—a review and recommendations for the practicing statistician. *Biometrical Journal*, 60(3):431–449, 2018. 26

[61] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning : with applications in R*. New York : Springer, [2013] ©2013, 2013. 26

[62] H. Akaike. *Information Theory and an Extension of the Maximum Likelihood Principle*, pages 199–213. Springer New York, New York, NY, 1998. 27

[63] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974. 27

[64] G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461 – 464, 1978. 27

[65] H. Akaike. Information measures and model selection. *Int Stat Inst*, 44:277–291, 1983. 27

[66] K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Verlag, 2002. 27

[67] J. Ding, V. Tarokh, and Y. Yang. Model selection techniques: An overview. *IEEE Signal Processing Magazine*, 35(6):16–34, 2018. 27

[68] P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47, 2004. 27

[69] K. P. Burnham and D. R. Anderson. Multimodel inference: Understanding aic and bic in model selection. *Sociological Methods & Research*, 33(2):261–304, 2004. 27

[70] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition, 2009. 28

[71] H. P. Gavin. The levenberg-marquardt method for nonlinear least squares curve-fitting problems c ©. 2013. 30

[72] H. Prasanna Das, R. Tran, J. Singh, X. Yue, G. Tison, A. Sangiovanni-Vincentelli, and C. J. Spanos. Conditional synthetic data generation for robust machine learning applications with limited pandemic data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(11):11792–11800, Jun. 2022. 32

[73] R. J. Chen, M. Y. Lu, R. Y. Chen, Drew F. K. Williamson, and F. Mahmood. Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, 5(6):493–497, 2021. 32

[74] K. Wei, Y. Fu, Y. Zheng, and J. Yang. Physics-based noise modeling for extreme low-light photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021. 32

[75] M. A. Whiting, J. Haack, and C. Varley. Creating realistic, scenario-based synthetic data for test and evaluation of information analytics software. In *Proceedings of the 2008 Workshop on beyond time and errors: novel evaluation methods for information visualization*, pages 1–9, 2008. 32

[76] M. W. Hopwood, J. S. Stein, J. L. Braid, and H. P. Seigneur. Physics-based method for generating fully synthetic IV curve training datasets for machine learning classification of PV failures. *Energies*, 15(14), 2022. 33

[77] M. Duquesnoy, C. Liu, D. Z. Dominguez, V. Kumar, E. Ayerbe, and A. A. Franco. Machine learning-assisted multi-objective optimization of battery manufacturing from synthetic data generated by physics-based simulations, 2022. 33

[78] H. Rajakaruna and V. V. Ganusov. Mathematical modeling to guide experimental design: T cell clustering as a case study. *Bulletin of Mathematical Biology*, 84(10):103, 2022. 33

[79] T. Burr, M. Hamada, T. Graves, and S. Myers. Augmenting real data with synthetic data: An application in assessing radio-isotope identification algorithms. *Quality and Reliability Engineering International*, 25:899 – 911, 12 2009. 33

[80] A. D. Nicholson, D. E. Peplow, J. M. Ghawaly, M. J. Willis, and D. E. Archer. Generation of synthetic data for a radiation detection algorithm competition. *IEEE Transactions on Nuclear Science*, 67(8):1968–1975, 2020. 33

[81] N. A. W. Walton. ATARI [online]. https://github.com/Naww137/ATARI, 2024. [cited 10/01/2024]. 36, 76

[82] J. M. Brown, D. P. Barry, R. C. Block, A. Youmans, H. Choun, A. Ney, E. Blain, M. J. Rapp, and Y. Danon. New measurements to resolve discrepancies in evaluated model parameters of Ta-181. *Nuclear Science and Engineering*, 2023. 36, 46, 53, 57, 58, 88, 117

[83] R. B. D'Agostino. An omnibus test of normality for moderate and large size samples. *Biometrika*, 58(2):341–348, 1971. 43

[84] R. B. D'Agostino and E. S. Pearson. Tests for departure from normality. empirical results for the distributions of b2 and b1. *Biometrika*, 60(3):613–622, 1973. 43

[85] F. J. Massey Jr. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951. 43, 50

[86] T. W. Anderson and D. A. Darling. A test of goodness of fit. *Journal of the American statistical association*, 49(268):765–769, 1954. 50

[87] R. W/ Peelle. Peelle's pertinent puzzle, informal memorendum dated october 13, 1987. *ORNL, USA*, 1987. 58

[88] P. Jain and P. Kar. Non-convex optimization for machine learning. *Foundations and Trends in Machine Learning*, 10(3-4):142–363, 2017. 65

[89] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, R. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, E. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. 65

[90] Diederik P. K. and Jimmy B. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. 70

[91] Andrzej M. and Waldemar R. Principal components analysis (PCA). *Computers & Geosciences*, 19(3):303–342, 1993. 70

[92] V. Klema and A. Laub. The singular value decomposition: Its computation and some applications. *IEEE Transactions on Automatic Control*, 25(2):164–176, 1980. 70

[93] S. Szeghalmy and A. Fazekas. A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. *Sensors*, 23(4), 2023. 73

Appendices

Appendix A

Investigations

A.1 Monitor Normalizations for Count Data

During inspection of the synthetic data model for transmission in Sec. 3.1.3, a clustering phenomenon was observed at low transmission values. The phenomenon is also observed in the actual measurement data, as show in Fig. A.1 particularly at 200 and 800 μs.

This is a results of the Poisson noise distribution at low counts and the monitor normalizations over each cycle. The raw count data is monitor normalized with each cycle, and then the reported data is a sum over around 35 cycles. For a single cycle, all TOF bins will have some number of counts in them (0 is possible) governed by the Poisson distribution and the monitor normalization will apply to the entire set. This is done for each of about 35 cycles, making the monitor corrected number of counts in any one tof bin

$$\sum_{i=1}^{\text{cycles}} \alpha_i c_i \ , \tag{A.1}$$

where α_i is the monitor correction for cycle i, a random variable distributed as $N(1, \sigma_\alpha^2)$ and c_i is the number of counts for cycle i, a Poisson distributed variable. For any given cycle, the probability density function for the monitor corrected counts is then given by the product:

$$P_Z(z) = \sum_{c=0}^{\infty} P_c(c) P_M(\frac{z}{c}) \frac{1}{c} \tag{A.2}$$

The derivation of Eq. A.2 is done using the product rule for algebra of random variables and is confirmed using Monte Carlo in Fig. A.2. The probability density associated with the data clustering is visible, and aside from the clusters, the distribution is well approximated by a normal distribution with a variance equal to the mean. However, this approximation deteriorates as the number of counts increases, since the variance of the true distribution grows more rapidly than that of the approximate normal distribution. This is shown in Fig. A.3. The deterioration of the approximation onsets earlier (at lower average counts) as the uncertainty in the monitor normalization random variable increases. This means that the statistical uncertainty assigned to raw count data could potentially be underestimated depending on how/if uncertainty is propagated through the combination of counts over cycles.

A.2 Objective Function Characterization

A.2.1 Initial Feature Bank Design

Figure A.4 shows some observational data and a prior model. Both are fake data with a realistic resonance signal and a simple Gaussian noise model generated for this example (*i.e.,* not the high-utility synthetic data developed for this thesis). The resonance model is parameterized with energy location (E_λ) and width parameters (Γ_γ and Γ_n). A grid search over the parameters can be done to map the behavior of the regression objective function. Figure A.5 shows the regression objective function with respect to E_λ and Γ_n in 3D. For any given value of width parameters, the convex region exists only narrowly around E_λ. Furthermore, as E_λ moves away from the minima of the convex region, the regression objective function becomes non-convex with respect to the other two parameters, as shown in Figures A.6 and A.7. While this is a simplified example, it captures the general relationship between resonance parameter and regression objective function. These non-convex and non-linear relationships are the motivation for the algorithmic approach to automated resonance inference described in Sec. 4.2.

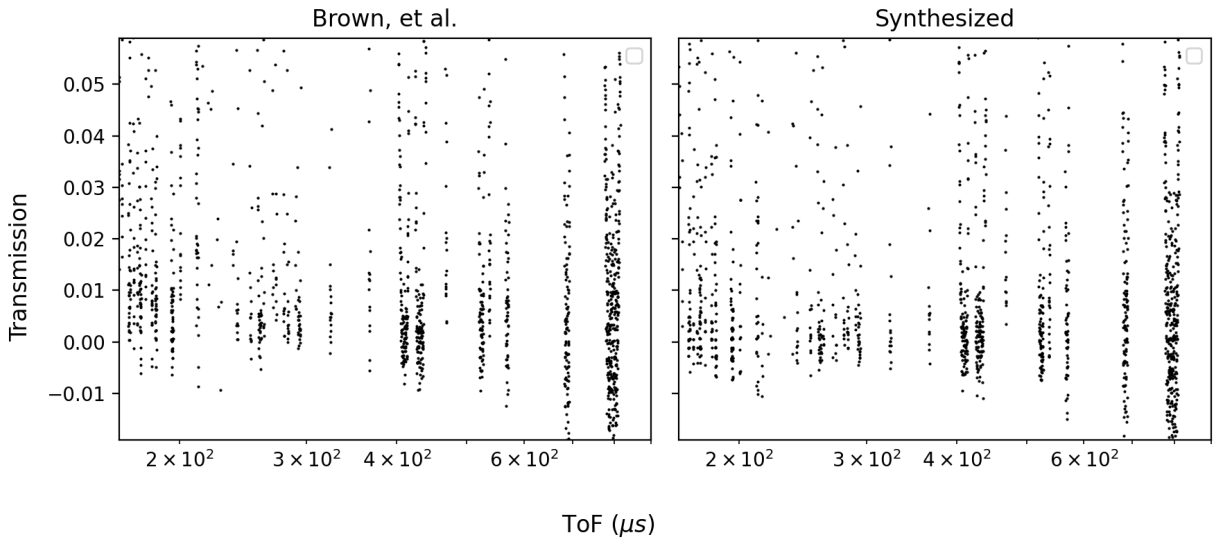

Figure A.1: Heuristic comparison of the transmission data clustering phenomenon at regions of low-count rates for synthetic and actual data.

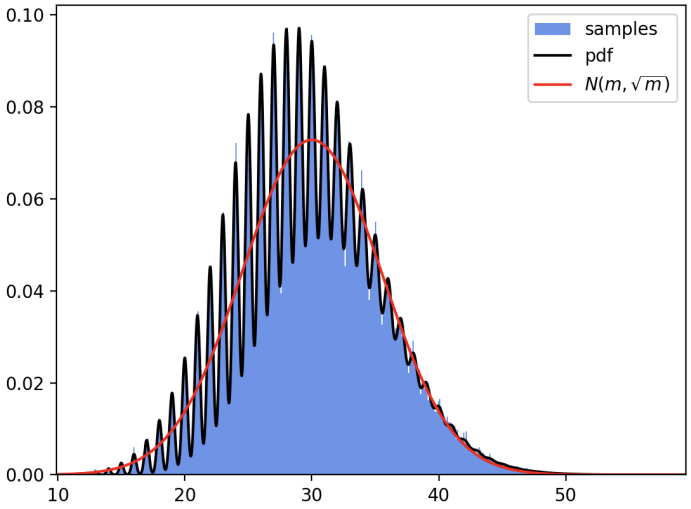

Figure A.2: Probability density function for the product of a Poisson and Normal distributed random variables.

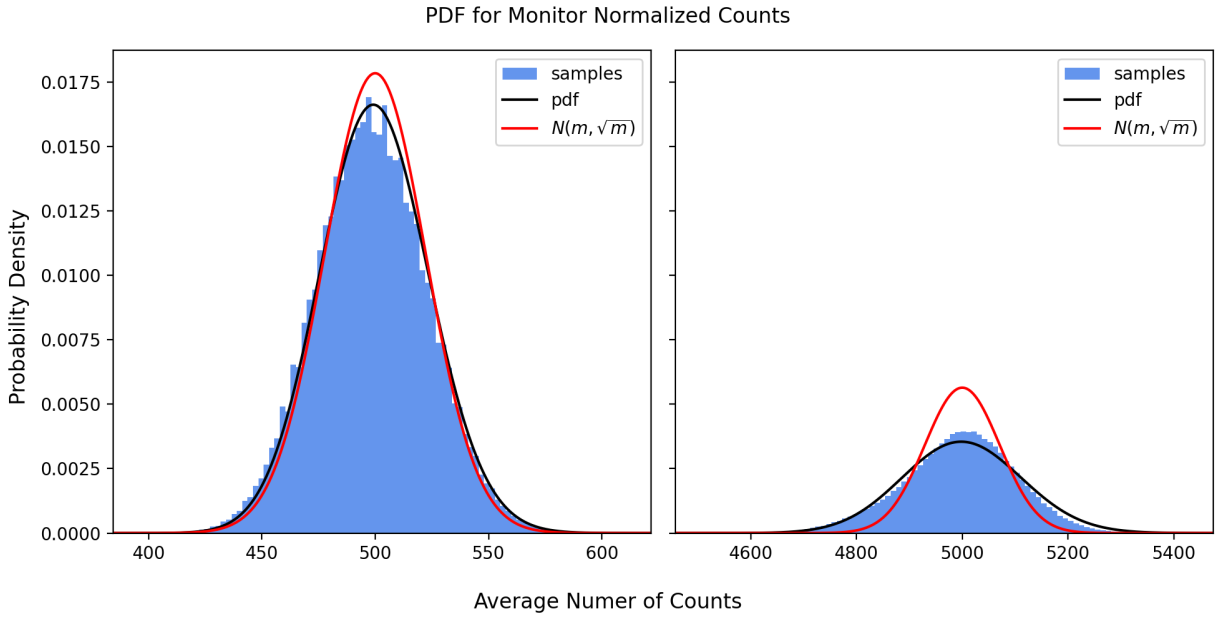

Figure A.3: Probability density function for the product of a Poisson and Normal distributed random variables at different average counts. The presumed approximation of a normal distribution with mean=variance is show for comparison.

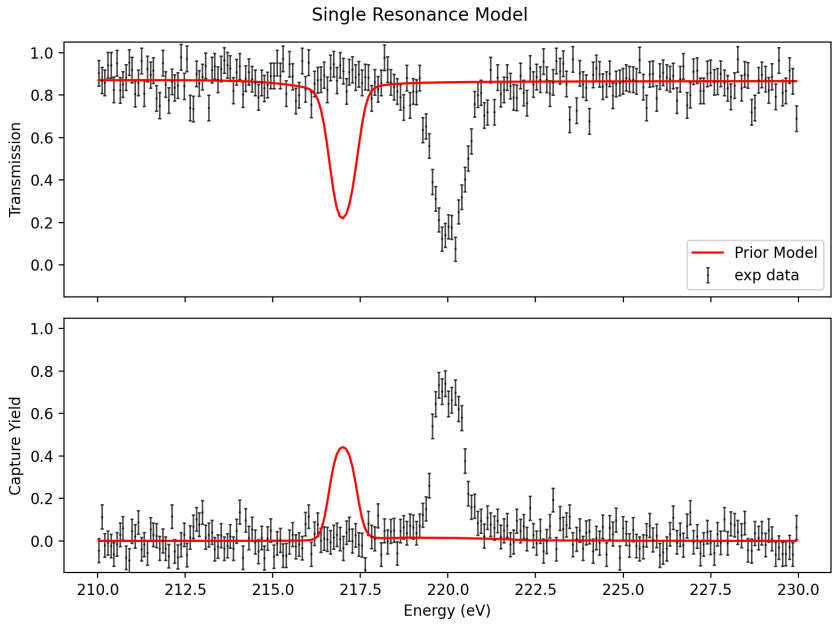

Figure A.4: Fake model/observational data for mapping the objective function.

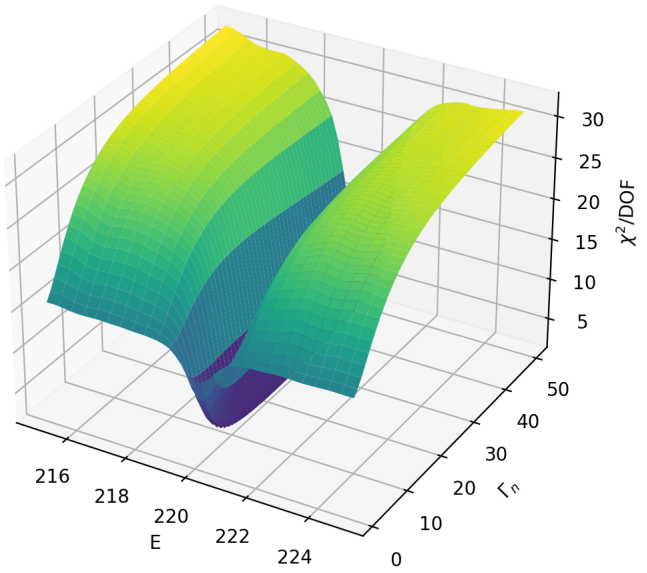

Figure A.5: χ^2 objective function surface over resonance energy and neutron width parameter.

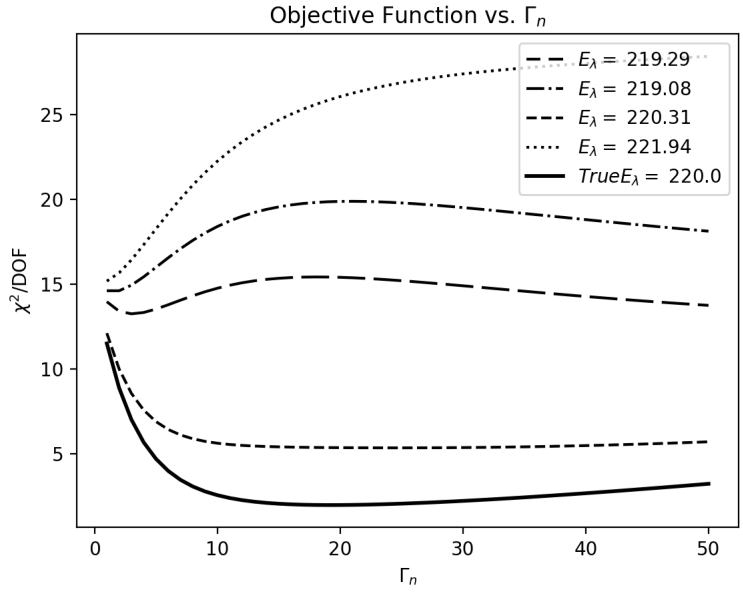

Figure A.6: χ^2 objective function value over neutron width parameter.

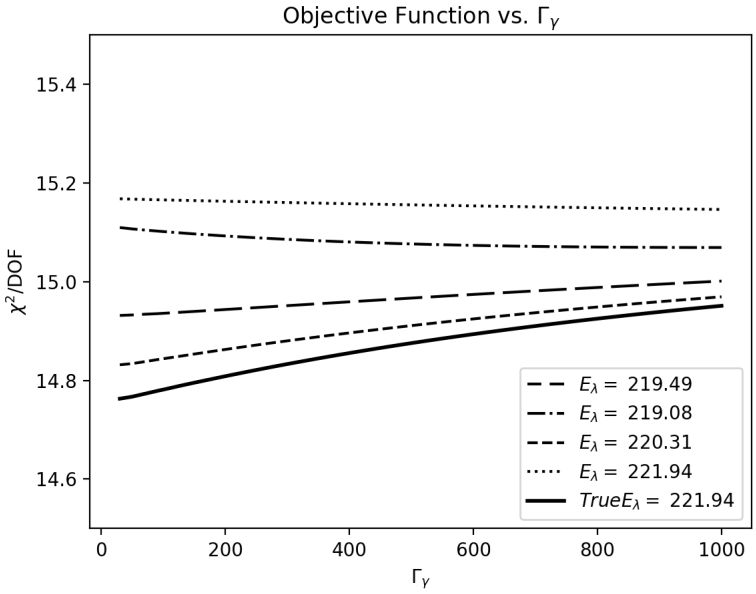

Figure A.7: χ^2 objective function value over capture width parameter.

A.2.2 Peelle's Pertinent Puzzle

Upon further exploration of the regression objection function, the need for the Peelle's Pertinent Puzzle (PPP) fix to the χ^2 regression objective discussed in Sec. 3.3.3 was identified. The regression objective with and without the PPP fix is denoted as χ^2_{T} and χ^2_{D} respectively. Before the fix was implemented, a recurring phenomenon was noticed. A minima in the χ^2_{D} objective would occur with resonance widths notably smaller than expected. This phenomenon occurs in both the synthetic data generated for this thesis and the actual data from [82] and will be detailed here using the actual data. This investigation will focus on transmission data alone for two reasons. First, transmission data is generally assumed not to have correlations large enough to induce PPP phenomena. Second, the SAMMY code includes built-in capabilities to correct for PPP, which are applicable to most capture data reduction methodologies due to a common approach to normalization, including the method detailed in this thesis. However, these capabilities are not compatible with the transmission data reduction methodology used here and in many other measurements; instead, the implicit data covariance method is recommended for this type of data.

Considering the very small window of incident neutron energies, 203–212 eV. In the actual measurement data, there are likely only two resonances. The resonance energies and widths were estimated by hand to ensure a starting point/prior within the local convex region of the data, then the data were fit with the Levenberg-Marquardt algorithm and SAMMY. The results are shown in Fig. A.8 where the optimized fit (blue) is much further from the experimental data than the starting point (red). Despite this visually misleading result, the optimization significantly improves the regression objective as χ^2/N_{data} starts at 2.40 and ends at 0.47. The latter value is suspiciously low as the true model should have an expected value of 1.0. If the resonance energies and capture widths are fixed, the regression objective can be visualized as a function of the two neutron widths ($\Gamma_{\mathrm{n},1}$, $\Gamma_{\mathrm{n},2}$)as shown in Fig. A.9, where the global minima is, in fact, observed at small neutron widths. Although a shallow local minimum exists at larger, more realistic neutron widths, relying on this minimum is not advisable as there is always the possibility of the non-linear optimization algorithm finding the global minimum particularly when starting with small resonances in the initial feature

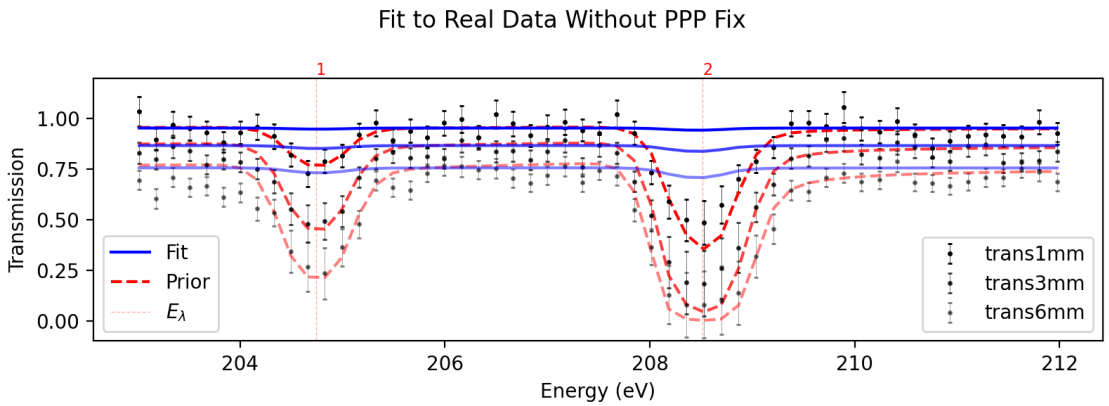

Figure A.8: Starting (red:Prior) and optimized (blue:Fit) model using the χ^2_{D} regression objective on three transmission measurement data sets from the actual data.

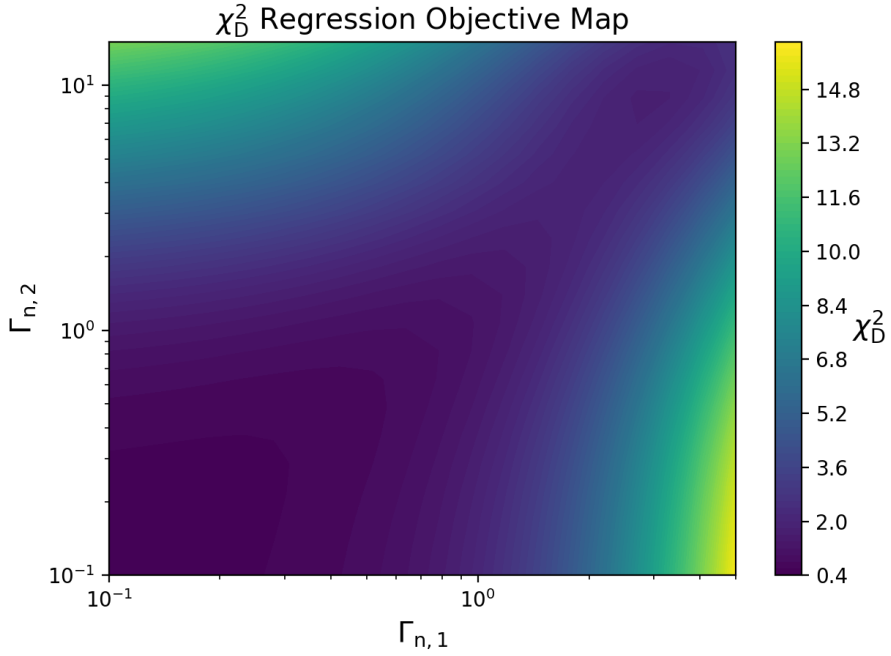

Figure A.9: Contour map of the χ^2_{D} regression objective with respect to neutron width for the two resonances shown in Fig. A.8.

bank as justified by he study in Appendix A.2. When analyzing this example, intuition suggests that the global minimum of the regression objective may not provide the best estimate of the underlying model. This is confirmed in examples within the synthetic data framework, where the true underlying model is known. A much more intuitive objective map emerges after applying the PPP fix to the transmission data, as shown in Fig. A.10.

Peele's Pertinent Puzzle (PPP) is generally not considered an issue for transmission data. However, for this specific data, the PPP phenomenon introduces a non-physical global minimum and appears to exert a constant pull toward lower values of Γ_n, even if constrained to a local minimum by additional data (such as capture data). This effect is further supported by the study in Sec. 3.3.3, which shows that poor performance is driven not only by outliers but by a significant shift in the probability density. Ultimately, that study demonstrates how this framework can be leveraged to assess whether the PPP fix is necessary for evaluating a given dataset.

A.3 Window Size & Model Selection

In Sec. 5.1.2, it was observed that the k-fold cross validation (CV) for model selection identified more resonances in the same data when the 20 eV window was decomposed into two 10 eV windows. In other words, the larger the data set used to calculate cross validation error (CVE) the more the model selection approach tended towards simpler models. This can be explained by the relative increase in variance of the CVE estimates with the size of the data set coupled with the application of the one standard error rule.

The CVE of each fold is calculated as

$$\mathrm{CVE}_k = \chi^2_k / N_k, \tag{A.3}$$

where N is the number of data in fold k. This is a generalization of mean squared error (MSE) which is traditionally used for CV test error when the residuals are uncorrelated.

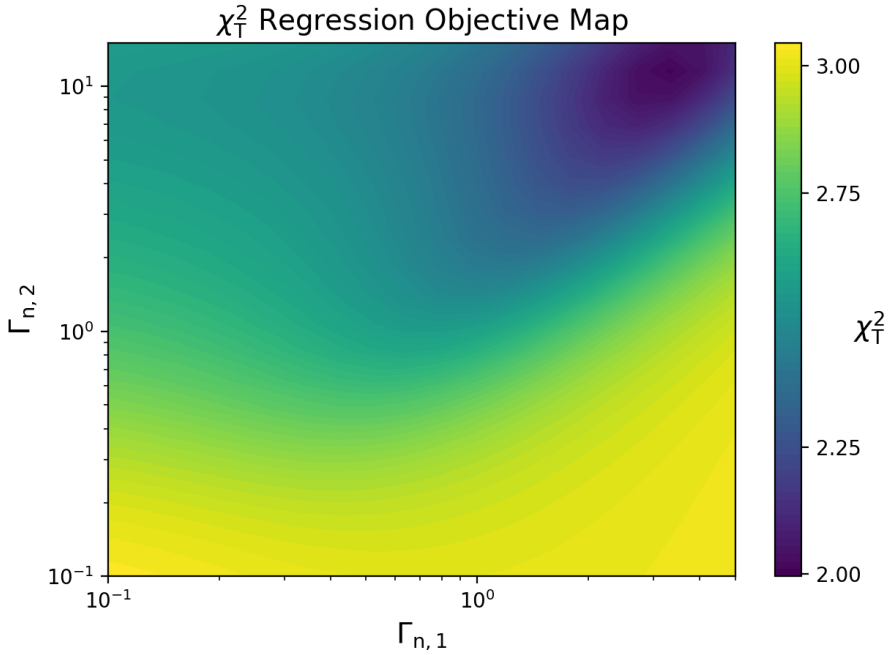

Figure A.10: Caption

The final score is calculated as the mean of all test errors,

$$\mathrm{CVE} = \frac{1}{k}\sum_{k} \mathrm{CVE}_k, \tag{A.4}$$

with standard error given by

$$\delta_{\mathrm{CVE}} = \frac{\sigma_{\mathrm{CVE}_k}}{\sqrt{k}}, \tag{A.5}$$

where σ_{CVE_k} is the sample variance of test errors calculated as

$$\sigma_{\mathrm{CVE}_k} = \sqrt{\frac{1}{k-1}\sum_{k}(\mathrm{CVE}_k - \mathrm{CVE})^2}. \tag{A.6}$$

The one standard error rule says to selected the simplest model within one standard error of the minimum CVE. This will occur where the increase in CVE is greater that the standard error, δ_{CVE}. For the correct model, CVE_k is expected to follow a χ^2 distribution with degrees of freedom N_k, scaled by $\frac{1}{N_k}$. The variance of the χ^2 distribution is known to be $2N_k$, meaning it will grow with the number of data points. The variance of the scaled distribution actual decreases with N_k, but the competing deterioration in CVE also gets scaled. This can obscure the visualization, so in the following results, N_k is fixed to be the same across all folds. This ensures that the $\frac{1}{N_k}$ scaling becomes a constant factor, eliminating the need to normalize CVE_k by N_{data} (i.e., $\mathrm{CVE}_k = \chi^2_k$). The impact of increasing σ_{CVE_k} due to additional data is shown in Fig A.11, where k-fold CV chooses a different model complexity for different window sizes. For this study, data was generated in the 200–260 eV range with three true resonances at 230, 232, and 238 eV. The k-fold cross validation approach for model selection from an initial feature bank was then performed using windows of 228–240, 225–245, and 220–250 eV. The true resonance signals were always within the window. The additional data on the wings is well described without resonances so it should not impact the quality of fit where the resonance signals are.

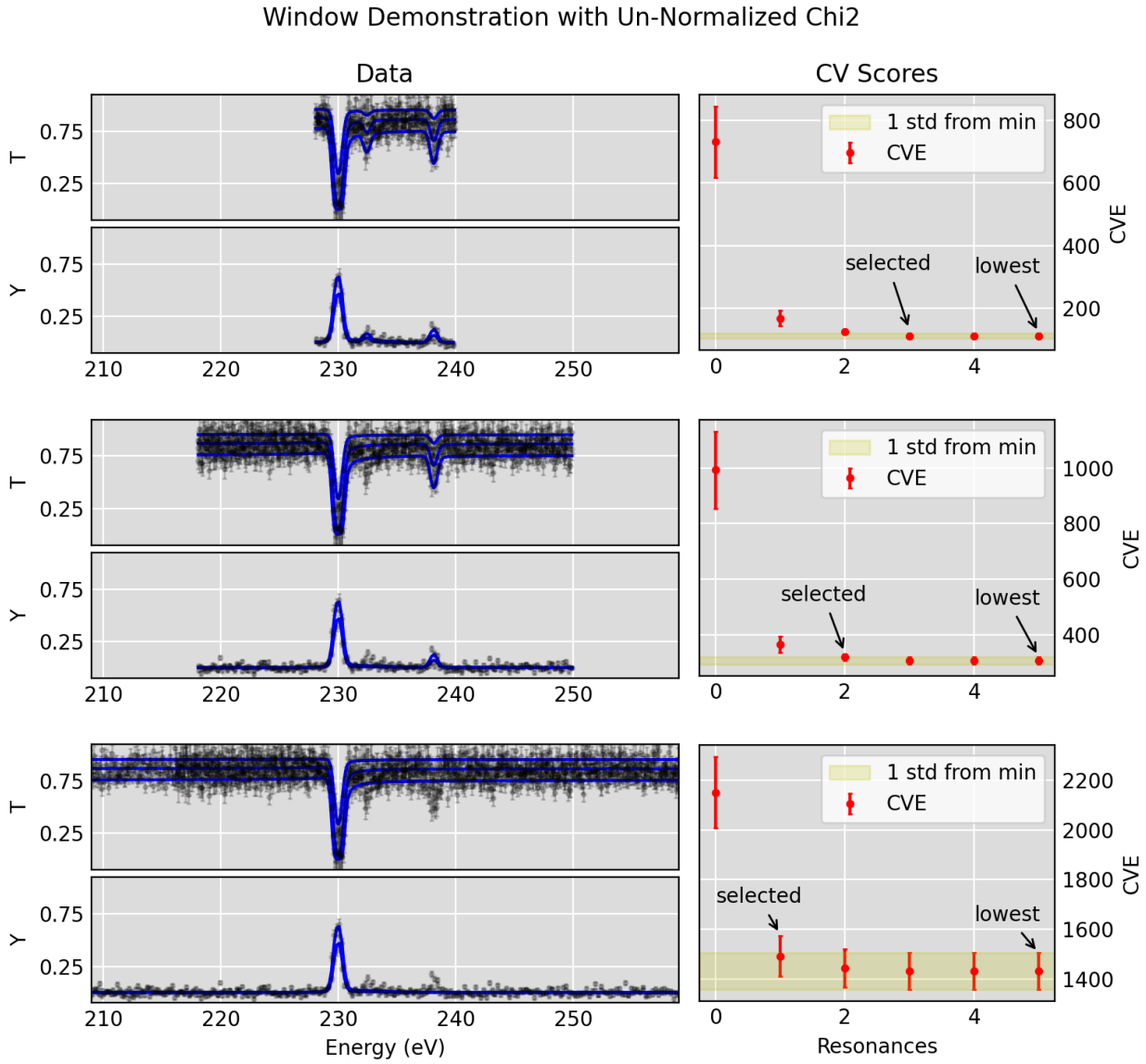

Figure A.11: k-fold cross validation for model selection window size study.

Appendix B

Parameter Reporting

B.1 Resonance Parameters

Table B.3 gives parameter notation and spin group definitions used throughout this thesis and Table B.2 gives the estimated average parameters used for resonance ladder generation and the initial feature bank design.

B.2 Hyperparameters

AutoFit algorithm hyperparameters required by the user along with settings used to generate results in this thesis. Hyperparameters are separated by the section of the algorithm that they apply to. For example, the hyperparameters for the nested Levenberg-Marquardt (LM) algorithm are under the Iterative Solver section. Informed selections may use *a priori* knowledge about resonance parameter distributions; in this case, the notation for a given quantile of the respective distribution will be denoted with a Q.

Table B.1: Parameter notation used throughout this thesis along with units.

Parameter	Description
E_λ	Resonance energy in eV
Γ_γ	Eliminated capture width in meV evaluated at resonance energy (E_λ)
Γ_n	Neutron width in meV evaluated at resonance energy (E_λ)
γ_γ^2	Reduced eliminated capture width in meV, ($\frac{\Gamma_\gamma}{2P(E_\lambda)}$)
γ_n^2	Reduced neutron width in meV, ($\frac{\Gamma_n}{2P(E_\lambda)}$)
SG	s-wave spin group with total angular momentum & parity: J^π
$\langle D \rangle_{J^\pi}$	Average level spacing for spin group J^π
$\langle \gamma_\gamma^2 \rangle_{J^\pi}$	Average reduced eliminated capture width for spin group J^π
$\langle \gamma_n^2 \rangle_{J^\pi}$	Average reduced neutron width for spin group J^π

Table B.2: Average resonance parameters used as *a priori* knowledge throughout this thesis. Parameters were estimated from the JEFF-3.3 evaluation [1].

Parameter	Value
$\langle D \rangle_{3^+}$	9.0 eV
$\langle \gamma_\gamma^2 \rangle_{3^+}$	32.0 meV
$\langle \gamma_n^2 \rangle_{3^+}$	452.6 meV
$\langle D \rangle_{4^+}$	8.3 eV
$\langle \gamma_\gamma^2 \rangle_{4^+}$	32.0 meV
$\langle \gamma_n^2 \rangle_{4^+}$	332.2 meV

Table B.3: User selected hyperparameters for AutoFit algorithm along with defaults.

Algorithm Section	Setting	Description
Initial Feature Bank	$\Delta E_\lambda = 0.625$	Spacing between resonances energies
	$\Gamma_\gamma^{J^\pi} = \langle \Gamma_\gamma^{J^\pi} \rangle$	Initializing values for capture widths
	$\Gamma_{\mathrm{n}}^{J^\pi} = 100 \cdot Q_{0.01}^{J^\pi}$	Initializing values for neutron widths
Iterative Solver	$\epsilon = 0.01$	Obj. threshold for convergence
	$\alpha_0 = 0.1$	Initial step size parameter
	$u = 1.5$	LM dampening parameter increase
	$d = 5.0$	LM dampening parameter decrease
Variable Selection	0.0	Γ_{n} threshold to eliminate resonance
	0.001	Obj. threshold to skip fitting for level
Window Decomposition	11 eV	Stage 1 window size
	3/11 eV	Stage 1 window overlap
	6.4 eV	Stage 2 window size
	1.0 eV	Resonance buffer

Vita

Noah Walton was born in Indiana and spent his formative years in various locations, including Idaho, Kentucky, and Tennessee. He graduated from Central High School in Knoxville, Tennessee, and enrolled at the University of Tennessee in 2016, where he pursued a major in nuclear engineering. In his final undergraduate year, Noah began research under Dr. Vladimir Sobes, whose mentorship inspired him to pursue a Doctor of Philosophy degree at the University of Tennessee. Noah's doctoral research is focused on applying advanced computational methods and machine learning to challenges in nuclear physics. Following graduation, he will continue his research as a Postdoctoral Fellow at Los Alamos National Laboratory. He is profoundly grateful for the support of his family, friends, and mentors, all of whom have contributed significantly to his achievements.