

To the Graduate Council:

I am submitting herewith a thesis written by Jacob Forbes entitled “A Computational Framework for Repeatable and Verifiable Nuclear Resonance Uncertainty Quantification.” I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Data Science.

Vladimir Sobes, Major Professor

We have read this thesis
and recommend its acceptance:

Vladimir Sobes

Jesse Brown

Denise Neudecker

Lawrence Heilbronn

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

A Computational Framework for Repeatable and Verifiable Nuclear Resonance Uncertainty Quantification

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Jacob Forbes
December 2025

© by Jacob Forbes, 2025
All Rights Reserved.

To my beautiful wife. Thank you for your unwavering love and encouragement.

Acknowledgements

I am humbled by the caliber of people I have been privileged to work with during my time at Rocky Top. This work was only possible because of the support of my advisor, committee, and fellow researchers.

I first wish to thank my advisor, Vladimir Sobes, for his encouragement and mentorship. He has cultivated a culture of teamwork, intellectual curiosity, and excellence among our research group. I am honored to be a part of his team.

I am also grateful to my committee members. Thank you Denise Neudecker, Jesse Brown, and Lawrence Heilbronn for welcoming me into the nuclear data community and for your steady support.

I am especially grateful to Noah Walton who was extremely generous with his time and energy in supporting this research. Noah's explanations of nuclear data concepts, guidance on nuclear data software, and his feedback on my research were essential to the production of this thesis. I would also like to thank Oleksii Zivenko, Cole Fritsch, Amanda Lewis, and Justin Loring for the many ways they supported this research.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of the Air Force, Department of Defense, or the U.S. Government.

Abstract

Accurate uncertainty quantification for neutron cross sections is essential for credible transport simulations and decision support. In the resolved resonance region, generalized least-squares (GLS) fits implemented in SAMMY produce evaluated cross sections and covariances that are widely used, yet the associated uncertainties can be underestimated when ideal assumptions are violated. Manual variance inflation is common but undermines reproducibility. This dissertation develops a reproducible, evaluator-in-the-loop pipeline that measures and corrects those distortions. The method couples SYNDAT with a scripted R -matrix fitter (AutoFit) to produce labeled benchmarks and empirical evaluated cross-section covariance matrices (ECSCMs) as ground truth.

Controlled studies show that, under ideal GLS conditions, SAMMY matches empirical covariances. Otherwise, the dominant errors arise from missing resonances and using single-level Breit–Wigner where multilevel coupling matters; moderate mis-scaling of correlated experimental terms is comparatively benign. We learn per-energy variance corrections from AutoFit features calibrated to empirical coverage. Random forests with a log-normal residual family best balance accuracy and conservatism on held-out ladders. We then extend to full energy- and channel-covariances via a block-feature design and nearest-PSD projection, validating correlations with rotated coverage. The result is a transparent, data-driven workflow that delivers calibrated, traceable ECSCMs for the resolved-resonance region and

supports in-the-loop use, with a path to on-the-loop quality control across evaluation campaigns.

Table of Contents

1	Introduction	1
2	Background	5
2.1	Cross-Sections and Resonance Theory	5
2.1.1	Microscopic Cross-Sections	5
2.1.2	Compound Nucleus and Resonances	7
2.1.3	Multilevel R -Matrix Theory	9
2.2	Measuring Resonance Parameters	11
2.2.1	Time-of-Flight Experiments	12
2.2.2	Fitting Procedure for Resonance Parameters	15
2.2.3	Uncertainty Quantification for Evaluated Cross Sections	16
2.3	Physics-Informed Synthetic Data	18
3	Measuring Accuracy of Evaluated Cross-Section Covariances	20
3.1	Procedure for computing empirical ECSCM	21
3.2	SAMMY vs. Empirical uncertainty quantification example	22
3.3	Single-value metric for measuring SAMMY ECSCM accuracy	23
3.4	Validating SAMMY ECSCM estimates when GLS assumptions are met	26
3.5	SAMMY ECSCM accuracy case studies	29
3.5.1	Case Study 1: Impact of the size of a missing resonance	29
3.5.2	Case Study 2: Impact of choice of R-matrix formalism	33
3.5.3	Case Study 3: Impact of misreported systematic data covariance	37

3.6	Lessons learned	41
4	Automated Corrections for SAMMY Total Cross-Section Covari-	
	ances	43
4.1	Baseline setup and empirical targets	44
4.1.1	Synthetic ladder ensemble	44
4.1.2	Essentials of the AutoFit routine	44
4.1.3	Developing target values	45
4.2	Choosing and validating a probability model	47
4.3	Feature construction	57
4.3.1	Block 1: GLS covariance diagonals	59
4.3.2	Block 2: squared bias between cardinalities	60
4.3.3	Block 3: squared fitted cross section	60
4.3.4	Block 4: LOESS-predicted variance	62
4.3.5	Intercept term and design-matrix dimensions	62
4.4	Machine-learning corrections	63
4.4.1	Ordinary least squares (OLS)	63
4.4.2	Regularized linear models: Ridge, Lasso, and Elastic Net . . .	64
4.4.3	Dimension-reduced linear models: Principal Components and Partial Least Squares	65
4.4.4	Gradient boosting regression	66
4.4.5	Support vector regression	66
4.4.6	Random-forest regression	67
4.4.7	Gaussian-mixture correction	68
4.5	Results: Machine-Learned Variances	69
4.5.1	Coverage as a function of cross-section magnitude	73
4.5.2	Recommendation and outlook	75
5	Automated Corrections for Energy and Channel Correlations	77
5.1	Empirical two-channel ECSCM targets	78

5.2	Feature construction for full covariances	79
5.2.1	Block 1: GLS ECSCMs (S)	82
5.2.2	Block 2: Cardinality-bias outer products (B)	82
5.2.3	Block 3: Outer products of fitted cross sections (F)	83
5.2.4	Block 4: LOESS-predicted scale (L)	85
5.2.5	Intercept, diagonal-offset, and design-matrix assembly	86
5.3	Reassembly and nearest-PSD projection	87
5.4	Mapped variance performance learned from full ECSCMs	89
5.5	Coverage in a rotated basis	89
6	Conclusions and Future Work	97
6.1	Future Work	99
	Bibliography	101
	Vita	108

List of Tables

3.1	Summary of convergence test input parameters and their values . . .	29
3.2	Summary of missing resonance case study input parameters and their values	31
3.3	Summary of R-matrix formalism case study input parameters and their values	34

List of Figures

2.1	Measured total neutron cross-section for ^{56}Fe between 305 and 600 keV (bars) and the corresponding one-channel Reich–Moore R -matrix fit (solid line). Peaks are labelled by the orbital angular-momentum quantum number ℓ of the incoming neutron: $\ell = 0$ (s -waves) give the broad, asymmetric humps; $\ell = 1$ and $\ell = 2$ (p - and d -waves) appear as much narrower, more symmetric spikes because higher- ℓ neutrons must tunnel through larger centrifugal barriers.* Adapted from Perey <i>et al.</i> [1] (reproduced in JEFF Report 18 [2]).	8
3.1	A synthetic sample of Ta-181 from a user-defined ladder of fifteen resonances.	24
3.2	In this example, SAMMY is used to fit an underspecified model. Not only does SAMMY miss the resonance at 51 eV, it significantly overestimates the confidence in the fit at 51 eV. Building an empirical ECSCM using the method presented in section 3.1 results in an accurate uncertainty quantification and allows us to assess the impact of underspecifying the model on the accuracy of the SAMMY UQ. . .	25
3.3	Convergence test for ECSCMs; F-norm difference between two empirical ECSCMs decreases with sample size N ; assumptions per GLS. . .	30
3.4	The missing resonance case study shows SAMMY ECSCM inaccuracy is roughly proportional to the size of the missing resonance in the underspecified model.	32

3.5	Data is generated using a Reich-Moore (XCT) model for resonances close in energy. The data are fit using both XCT and Single-Level Breit-Wigner (SLBW) models. SLBW does not model interactions between the resonances and results in a poor fit.	35
3.6	The R-matrix formalism case study generated XCT samples from two resonances 2 eV apart for different locations of the first resonance. The case study confirms SAMMY ECSCMs are inaccurate when using an SLBW approximation to fit the XCT data at low energies. The SLBW approximation becomes adequate at higher energies.	36
3.7	For the ^{238}U benchmark, the difference in the standard deviation of k between empirical and SAMMY parameter covariances, $\Delta\sigma_k$, is plotted against the multiplicative factor w applied to $gm g^\top$ during fitting. Points show the mean across repetitions; whiskers denote one standard deviation.	39
3.8	Empirical relationship between the Frobenius norm $\Phi = \ C_E - C_S\ _F$ and $\Delta\sigma_k$ for the ^{238}U benchmark. The trend indicates that increasing matrix-level error translates predictably into added uncertainty on k . Here $\Delta\sigma_k$ is the absolute difference in the standard deviation of k_{eff} ; no pcm conversion is applied.	40
4.1	Left: The empirical ECSCM for ladder 244. Right: The true total cross-section and $n = 50$ fitted cross-sections using AutoFit on i.i.d. SYNDAT samples.	46
4.2	Empirical coverage $\hat{\gamma}(c)$ for the Normal residual model, evaluated at confidence levels 0.60–0.95. The dashed gray line marks perfect calibration ($\hat{\gamma} = c$). Points above the line indicate conservative intervals, while points below it correspond to over-confident bands. This baseline serves as a visual base case for the reader before comparing more complex probability models.	50

4.3	Empirical coverage $\hat{\gamma}(c)$ for three candidate residual models (Normal, Cauchy ($\nu = 1$), and log-Normal) evaluated at confidence levels $c \in \{0.60, 0.65, \dots, 0.95\}$. The dashed gray line indicates perfect calibration ($\hat{\gamma} = c$). The Cauchy model (green) is uniformly conservative; the Normal (blue) and log-Normal (orange) agree near $c \approx 0.70$ but diverge at the extremes, with the log-Normal remaining slightly conservative as $c \rightarrow 0.95$	51
4.4	Empirical coverage $\hat{\gamma}(c)$ for the Student- t residual family at selected degrees of freedom $\nu \in \{0.5, 1, 2, 5, 10, \infty\}$. The dashed line denotes perfect calibration ($\hat{\gamma} = c$). Heavy-tailed models ($\nu \leq 2$) are uniformly conservative, while the Normal limit ($\nu = \infty$) still over-covers for $c \gtrsim 0.70$, indicating that all t models in this range produce intervals that are wider than empirically required.	53
4.5	Empirical coverage $\hat{\gamma}(c)$ for the generalized Gaussian residual model at shape parameters $\beta \in \{0.25, 0.5, 1, 2, 5, 10, 1000\}$. The dashed line marks perfect calibration ($\hat{\gamma} = c$). Small values $\beta \leq 1$ yield under-coverage (intervals too narrow), while large values yield over-coverage. $\beta = 2$ corresponding to the Normal model is not strictly outperformed by any other $\beta \neq 2$	55
4.6	Piece-wise linear inverse mapping $c \mapsto \hat{\beta}(c)$ constructed on the training data. A logarithmic y -axis reveals that lower confidence levels require markedly heavier tails (small $\hat{\beta}$) than high confidence levels.	56
4.7	Coverage curves obtained by matching the generalized Gaussian shape parameter to the nominal confidence, $\beta = \hat{\beta}(c)$ from Figure 4.6. The mapping is learned on the first 200 synthetic ladders (left) and then applied unchanged to the 50 test ladders (right). The dashed line indicates perfect calibration.	58

4.8	Scatter of empirical standard deviation $s = \sqrt{v_{\text{emp}}}$ against fitted cross section $\hat{\sigma}$ for the 200 training ladders (log-log axes) over all energies. The red curve is a LOESS smoother (span 0.2). Beyond the small- $\hat{\sigma}$ local minimum, larger cross sections systematically predict larger sample-to-sample fluctuations, motivating both the fitted cross-section precitor described in Block 3 and the non-parametric variance estimate in Block 4.	61
4.9	Empirical coverage of 95 % log-normal bands for four residual families.	70
4.10	Comparison of log-normal confidence bands based on empirical variances, SAMMY variances, and four machine learned variances on a single resonance ladder. These results visually depict how some methods are overly precise (e.g., SAMMY), some overly conservative (e.g., OLS), and some relatively well calibrated.	72
4.11	Logistic curves showing the relationship between coverage performance and fitted cross-section of log-normal curves for various methods. The shape and location of the logistic curves indicate the accuracy of variance predictions over energy regions far from resonances, near small resonances, and near large resonances.	74
5.1	Left: empirical two-channel ECSCM $\hat{C}_E = \frac{1}{N} R^\top R$ with elastic (bottom left) and capture (upper right) within-channel blocks and the two cross-channel blocks off the diagonal. Right: corresponding true and fitted cross sections on $\mathcal{E}$	80
5.2	Empirical standard deviation $s = \sqrt{v_{\text{emp}}}$ versus fitted cross section $\hat{\sigma}$ on log-log axes, stacked by channel. Top: capture; bottom: elastic. Each point aggregates over energies, ladders, and cardinality levels; the red curve is a LOESS smoother fit separately in each channel. The clear relationship between s and $\hat{\sigma}$ motivates the multiplicative the use of $F = \hat{\sigma}\hat{\sigma}^\top$ as a regressor for full-matrix covariance prediction. . . .	84

5.3	Coverage curves of the variances (main-diagonals): empirical hit ratio versus nominal confidence for Normal, Cauchy, and Log-Normal families, split by train/test. Because the diagonal is learned jointly with off-diagonals, OLS is less conservative than in Chapter 4, while Random Forest performance is relatively consistent.	90
5.4	Coverage after rotating to the method’s principal-component basis. Normal (top), Cauchy (middle), and Log-Normal (bottom) families; train (left) and test (right). OLS benefits from learning the full ECSCM and attains markedly better rotated coverage than when trained on variances only. RF continues to be conservative across the board.	93
5.5	Elastic cross-section bands at 95% nominal confidence for four methods (Empirical, SAMMY, OLS, Random Forest) on a representative ladder. Green dashed curves are Log-Normal bands derived from the learned full covariance; blue solid is the true cross section; red dash-dot is the fitted curve.	95
5.6	Capture cross-section bands at 95% nominal confidence (same four methods and ladder as in Figure 5.5). RF remains conservative; OLS strikes a favorable accuracy–efficiency balance.	96

Chapter 1

Introduction

Safe reactor operation, precise isotope production, and credible non-proliferation analysis all begin with quantitative knowledge of how neutrons interact with matter. That knowledge can be expressed through the neutron cross-section: an energy-dependent area that represents the probability that an incoming neutron will induce a specified reaction. Because engineering decisions depend just as strongly on the uncertainty in that probability, each cross-section curve is often accompanied by a covariance matrix that records both the uncertainty at every energy and the correlations that couple uncertainties across the spectrum.

Raw experimental results, theory, and expert judgement are combined into comprehensive data libraries such as ENDF/B-VIII.0 and JEFF-3.3 [3, 4]. These libraries supply the neutron-interaction inputs for every modern transport calculation. Their influence extends from reactor fuel-cycle design and industrial irradiation scheduling to national-security stockpile stewardship and medical-isotope production [5, 6, 7, 8]. Contemporary Monte-Carlo codes propagate the covariances that accompany the cross sections, thereby producing statistically defensible confidence intervals on key observables such as k_{eff} or isotope yield [9]. As such, accurate uncertainty quantification of cross section estimates is as important as the accuracy of the cross section estimates themselves.

Producing and maintaining an evaluated library is a multi-decade endeavour that draws on experimentalists, theorists, evaluators, and software developers distributed across several institutions. Legacy software, incomplete documentation, and the imminent retirement of senior experts threaten both consistency and institutional memory [10, 11]. These difficulties become particularly acute in the resolved resonance region (RRR), where sharp, energy-localized structures dominate the cross section. In the RRR, evaluators fit multilevel R -matrix models—typically in the Reich–Moore approximation—to time-of-flight data using software tools such as SAMMY, a generalized least-squares code developed at Oak Ridge National Lab specifically for resonance analysis (its mathematical framework is outlined in Chapter 2). When the statistical assumptions behind that fit are violated, the resulting cross section covariances have long been regarded as unrealistically small [12, 13]. Evaluators often compensate by manually inflating these uncertainties according to local practice [13, 14]. Although such inflation protects downstream calculations, it obscures provenance and undermines reproducibility, limiting the value of subsequent sensitivity analyses.

In this context, *automation* denotes a documented, version-controlled workflow that executes routine tasks while leaving scientific judgement firmly in the evaluator’s hands. Walton and co-workers have demonstrated the power of such workflows: their physics-informed synthetic-data generator, SYNDAT, coupled with an automated, stepwise R -matrix fitter, produces resonance evaluations that are numerically identical whenever the same inputs are supplied and can generate unlimited labelled benchmarks for validation [15, 16, 17, 18, 19].

A useful vocabulary for levels of autonomy appears in the *U.S. Air Force Doctrine Note 25-1 on Artificial Intelligence* [20]. Systems that keep a human *in the loop* provide recommendations that require explicit approval; those that place the user *on the loop* act autonomously unless intervention is requested. The methods developed in this dissertation are designed for in-the-loop operation: evaluators select data, hyperparameters, and quality criteria, and each decision is stored in a machine-readable

log. As confidence grows, the same framework could one day assume an on-the-loop role, scanning large evaluation campaigns and flagging only the cases that warrant expert attention.

This work examines how specific violations of generalized least-squares assumptions distort propagated cross-section covariances and proposes an automated uncertainty-quantification pipeline that restores statistical fidelity without manual variance inflation. The study directly addresses reproducibility goals set by Cross Section Evaluation Working Group (CSEWG) and by the Working Party on International Nuclear Data Evaluation Co-operation. Although the immediate focus is the resolved resonance region—where application sensitivities are highest—the modular design invites extension to unresolved-energy regimes and to charged-particle reactions. The ultimate objective is to deliver evaluations whose covariances are both defensible and reproducible, thereby strengthening confidence in the nuclear data that underpin critical technological and policy decisions for decades to come.

The goal of this dissertation is to deliver a reproducible, evaluator-in-the-loop methodology for producing statistically defensible cross-section covariances in the resolved resonance region. The work proceeds in three stages. First, it develops a physics-based methodology for quantifying the accuracy of evaluated cross-section covariance matrices (ECSCMs) by generating synthetic data to compute accurate, empirical covariance against which covariances computed using the status quo methodology can be compared. Second, it uses these targets to select and calibrate probability models whose nominal confidence levels match empirical coverage, enabling machine learning of improved covariance estimates. Third, it extends this mapping from variances to the full covariance structure, incorporating correlations across energies and reaction channels by learning from patterns in the evaluated cross sections themselves. The resulting pipeline produces calibrated ECSCMs without manual inflation and retains traceability to all inputs and hyper-parameters.

The intended near-term use is “in the loop.” Evaluators can run the pipeline alongside existing SAMMY analyses to (a) diagnose under- or over-coverage in

the diagonal ECSCM elements, and (b) produce accurate, repeatable cross-section covariances. Over time, the same framework can support “on-the-loop” quality control across many evaluations by flagging cases where empirical coverage or correlation patterns deviate from nominal targets. Although the present work focuses on neutron total, elastic, and capture channels in the resolved resonance region, the modular design invites extension to unresolved-energy regimes, other reaction types, and broader multichannel covariance applications.

Chapter 2

Background

2.1 Cross-Sections and Resonance Theory

Readers who already work comfortably with neutron-reaction theory may wish to skip ahead and re-join the narrative at Section 2.3. This section builds just enough nuclear-physics background so that readers from other technical disciplines have sufficient context to understand the methods presented in this thesis and to adapt them to their own problems.

2.1.1 Microscopic Cross-Sections

Imagine directing a pencil-thin beam of neutrons at a foil that is only a few micrometres thick. By counting how many neutrons undergo a specific reaction—elastic scattering, radiative capture, fission, and so on—you obtain an experimental estimate of the single-pass interaction probability for that reaction. In nuclear science we report this using the microscopic cross section $\sigma_x(E)$, where the subscript x labels the reaction and E is the neutron’s kinetic energy. A cross section isn’t the likelihood of an interaction, although it’s mathematically directly related to it. It is an effective area (in barns); for a thin target with areal density η (atoms cm⁻²), the single-pass reaction probability satisfies $P_x(E) \approx \sigma_x(E) \eta$.

The formal connection between likelihood and count rate is

$$R_x = \int_0^\infty \sigma_x(E) \phi(E) dE \quad [\text{s}^{-1}], \quad (2.1)$$

with $\phi(E)$ the differential neutron flux ($\text{neutrons cm}^{-2} \text{s}^{-1} \text{eV}^{-1}$). You can read Eq. (2.1) as a weighted average: the integrand pairs “how often neutrons at energy E arrive” with “how likely they are to react,” and the integral adds up contributions from every energy because real-world neutron fields never consist of a single energy alone.

Although σ_x carries the dimensions of an area, the number is effective, not geometric. Quantum mechanics assigns a complex scattering amplitude to every collision; the cross-section is the squared magnitude of that amplitude expressed in area units. A large value therefore signals “high probability,” not “large target nucleus.” Because the areas are unimaginably small on an everyday scale, the field uses its own unit, the *barn* ($1 \text{ b} = 10^{-24} \text{ cm}^2$) [21]. The thermal fission cross-section of ^{235}U , roughly 585 b, corresponds to an area smaller than the head of a pin by many orders of magnitude.

A single incoming neutron may leave the interaction zone in many different ways. If the neutron escapes the nucleus after interacting with the nucleus, we call the process elastic scattering; if the neutron sticks and the nucleus later releases a γ ray, we call the process radiative capture; if the nucleus splits, we call the process fission; other reactions eject an alpha-particle or a proton. The cross-sections associated with each reaction type, or channel, sum to the total cross-section (analogous to how the probabilities of mutually exclusive, collectively exhaustive components must sum to 1). Mathematically that statement appears as

$$\sigma_{\text{tot}}(E) = \sum_x \sigma_x(E), \quad (2.2)$$

where $\sigma_{\text{tot}}(E)$ is the total microscopic cross section. In the thin-sample limit of Eq. (2.6), the product $n\sigma_{\text{tot}}(E)$ is a dimensionless interaction probability per traversal, so σ_{tot} can be interpreted as setting the scale for “how likely a neutron of energy E is to interact with the target in any way.”

The identity offers a built-in consistency check for experimentalists: if one group measures σ_{tot} while another measures σ_γ and σ_n on different equipment, the three data sets must respect the sum rule within their quoted uncertainties; any mismatch is a diagnostic clue that something in the experimental chain—perhaps a detector efficiency or a normalization constant—needs re-examination.

2.1.2 Compound Nucleus and Resonances

In the resolved resonance region (RRR) cross-sections do not increase or decrease monotonically as a function of energy. Most nuclides display cross-section curves that look like a city skyline, full of narrow peaks rising above a lower, smoother background (see Figure 2.1). The physical picture behind those peaks was supplied by Niels Bohr in 1936 through his compound-nucleus hypothesis [22]. For incident energies below a few hundred keV the neutron is not ejected immediately after the initial collision. Instead it fuses with the target to create a highly excited, metastable composite system—the compound nucleus. During its brief life the excitation energy is shuffled among all the nucleons, and the system loses memory of the detailed entrance channel. Only conserved quantities such as total spin and parity (J^π), together with the incident neutron’s direction, remain correlated with the formation process. Eventually the compound system decays by emitting whatever particle liberates enough energy to reach a final state. Quantum mechanics links the mean lifetime τ of that state to the width of the resonance: the full width at half maximum is $\Gamma \approx \hbar/\tau$, where $\hbar \equiv h/2\pi = 1.055 \times 10^{-34}$ Js is the (reduced) Planck constant.

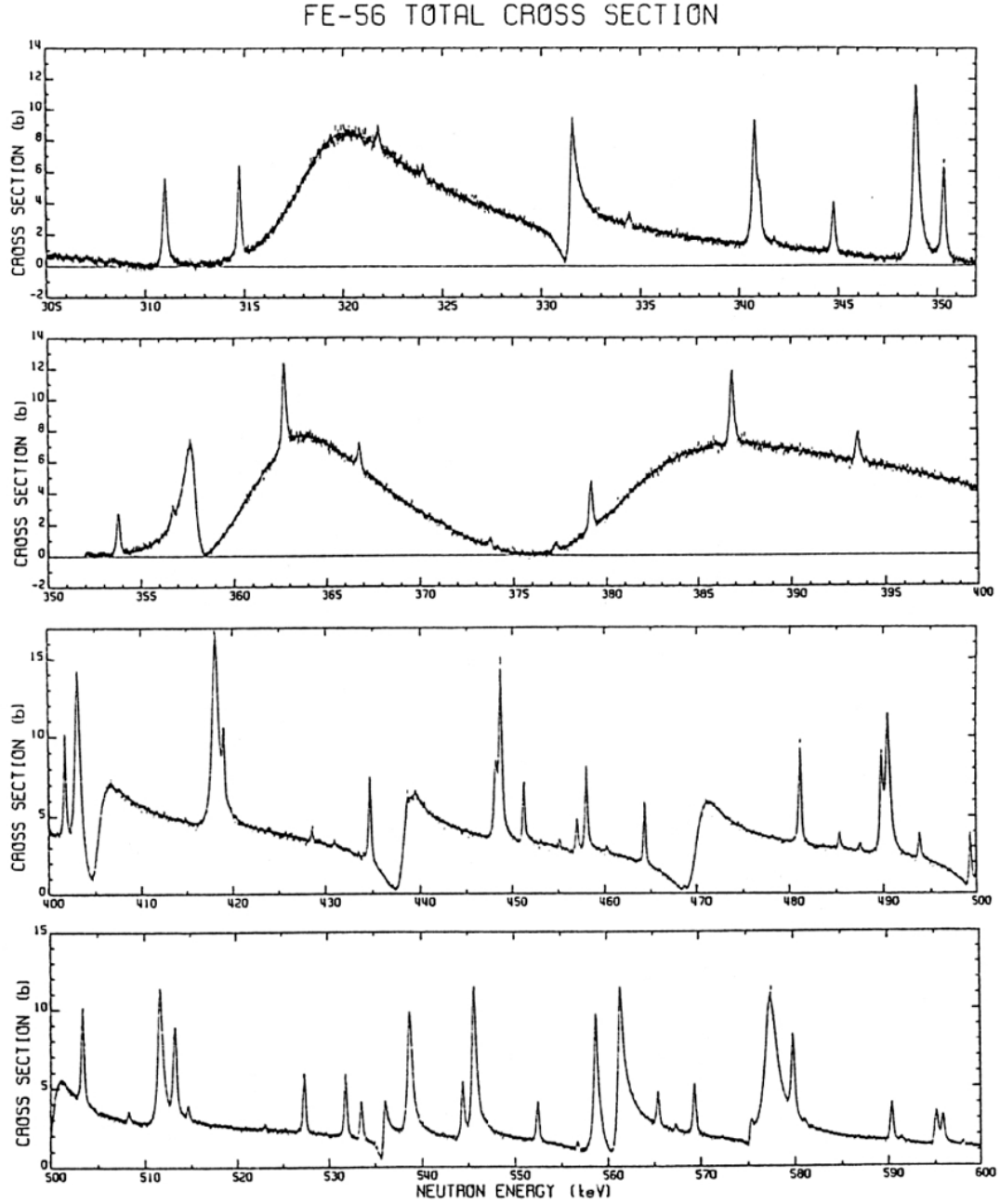


Figure 2.1: Measured total neutron cross-section for ^{56}Fe between 305 and 600 keV (bars) and the corresponding one-channel Reich–Moore R -matrix fit (solid line). Peaks are labelled by the orbital angular-momentum quantum number ℓ of the incoming neutron: $\ell = 0$ (s -waves) give the broad, asymmetric humps; $\ell = 1$ and $\ell = 2$ (p - and d -waves) appear as much narrower, more symmetric spikes because higher- ℓ neutrons must tunnel through larger centrifugal barriers.* Adapted from Perey *et al.* [1] (reproduced in JEFF Report 18 [2]).

When adjacent peaks are separated by gaps larger than their widths, each peak can be treated independently. The single-level Breit–Wigner expression [23]

$$\sigma_x(E) \approx \frac{\pi\lambda_n^2}{g} \frac{\Gamma_{n,\lambda}\Gamma_{x,\lambda}}{(E - E_\lambda)^2 + \left(\frac{1}{2}\Gamma_\lambda\right)^2}, \quad (2.3)$$

makes that idea precise. The prefactor $\pi\lambda_n^2$ sets the geometric scale via the neutron’s de Broglie wavelength λ_n —the area that a slow neutron of energy E occupies; g is a statistical factor that adjusts the peak height to reflect how many equally likely internal configurations the neutron can occupy when it merges with the nucleus; E_λ centers the resonance in energy; the partial widths $\Gamma_{n,\lambda}$ and $\Gamma_{x,\lambda}$ give the decay probabilities into the neutron or channel x ; and the sum Γ_λ fixes the peak width.

If the incoming neutron energy E exactly matches the resonance energy E_λ , the term $(E - E_\lambda)^2$ vanishes and the cross section reaches its maximum. As soon as E moves away from E_λ , the denominator grows quadratically, so the peak falls off rapidly on either side; the total width Γ_λ sets the energy span over which the cross-section stays appreciable. The numerator simply says that two things must happen in sequence: the compound nucleus must first form (governed by $\Gamma_{n,\lambda}$) and then decay through the specific channel of interest (governed by $\Gamma_{x,\lambda}$). Multiply the probabilities of those two steps, divide by the “distance from resonance” supplied by the denominator, and scale by the geometric area $\pi\lambda^2$. In that sense the Breit–Wigner curve is something akin to “probability in, probability out, divided by how far the incident neutron’s energy is from the resonance energy.”

2.1.3 Multilevel R -Matrix Theory

*Quantum mechanics classifies an incoming neutron by its orbital angular momentum ℓ relative to the target nucleus. The lowest class, $\ell = 0$ or *s-waves*, feels no centrifugal barrier, so even slow neutrons reach the nuclear surface and the resulting resonances are wide and asymmetric. Higher classes— $\ell = 1$ (*p-waves*), $\ell = 2$ (*d-waves*), and so on—must tunnel through increasingly thick barriers, narrowing the peaks and making them more symmetric. Because many ℓ values are open at once in the 300–600 keV band, their overlapping tails form the “city-skyline” pattern that ultimately motivates multilevel R -matrix theory.

For nuclei in which neighbouring resonances overlap, the single-level Breit–Wigner expression is no longer adequate. A more general treatment was provided by Lane and Thomas in the form of multilevel R -matrix theory [24]. The method introduces a notional spherical surface of radius a , called the channel radius. Inside this surface the strong nuclear force dominates and determines the complicated many-body dynamics of the compound system; outside, the neutron encounters only the centrifugal potential that arises from its orbital motion and, for charged ejectiles, the long-range Coulomb field[†].

The time-independent Schrödinger equation (the quantum analogue to $F = ma$) is solved on both sides of the surface, and the two solutions are matched so that the total wave-function is continuous and differentiable. All interior dynamics are condensed into an energy-dependent matrix $\mathbf{R}(E)$ whose element $R_{cc'}(E)$ measures the amplitude with which a resonance that is born in channel c' reaches the boundary in channel c . Here a channel is an exit mode distinguished by fundamental quantum properties of the ejected particle such as orbital angular momentum ℓ and total spin J . Transmission through the exterior barrier is quantified by a diagonal penetration matrix $\mathbf{P}(E)$; its entry $P_c(E)$ varies from nearly zero at low energies, when the barrier is thick, to increasing and unbounded values at higher energies, when the barrier thins.

The resulting quantity is the collision matrix

$$\mathbf{U}(E) = (\mathbf{I} - i\mathbf{R}(E)\mathbf{P}(E))^{-1}(\mathbf{I} + i\mathbf{R}(E)\mathbf{P}(E)), \quad (2.4)$$

from which each reaction cross-section is obtained by squaring the absolute value of the corresponding matrix element, e.g. $|U_{ab}(E)|^2$ for a transition from entrance channel a to exit channel b . For a detailed derivation and physical interpretation

[†]The centrifugal potential is the quantum analogue of the outward push you feel when swinging a ball on a string. The Coulomb field is the electrostatic repulsion between positive charges; it acts between the protons in the target and any charged particle (such as an α or emitted proton) that tries to leave the nucleus.

of the collision-matrix expression, the reader is referred to Frohner’s comprehensive JEFF Report 18 [2].

If only one resonance and one reaction channel are open, both $\mathbf{R}$ and $\mathbf{P}$ become ordinary numbers and the entire expression reduces to the single-level Breit–Wigner formula discussed earlier. In the general multichannel case, the shared interior parameters couple all channels, so an adjustment to one partial width automatically reshapes the others, maintaining the probability balance required by Eq. (2.2).

In practice one rarely attempts a fully coupled treatment of the hundreds of γ -ray cascades that can follow neutron capture. Capture channels differ from particle channels in three decisive ways: photons face no centrifugal barrier, they do not interfere appreciably with one another, and once emitted they cannot re-enter the reaction. The Reich–Moore approximation [25, 26] takes advantage of those facts by merging all capture modes into a single, non-interfering channel whose only role is to add its radiative width $\bar{\Gamma}_{\gamma,\lambda}$ to the total width Γ_λ . The resulting R -matrix is written only for the particle channels, conserving probability while keeping the matrix dimension small enough for routine evaluations.

The conceptual distinction between single-level and multilevel formalisms is central to one of the case studies presented later in this dissertation. By applying the single-level model in an energy region where channel coupling is significant, the study quantifies the distortion introduced when that coupling is ignored—an effect that becomes particularly visible in the propagated covariance matrices. Readers interested in further discussion of these formalisms and their specialized limits will find a detailed treatment in JEFF Report 18 [2].

2.2 Measuring Resonance Parameters

Many technological and scientific applications—from power-reactor design to medical-isotope production—depend on reliable neutron cross-section data and on the uncertainty bands that accompany those data (see Chapter 1).

Unfortunately, those cross sections cannot be measured directly in the resolved-resonance region. A single resonance may change by orders of magnitude over an energy span of only a few electron-volts, yet the probability that a neutron of exactly the right energy will react in a laboratory-thickness foil is tiny. If the foil is made thicker to catch more reactions, most neutrons are removed near the entrance surface and never probe the resonant probability deeper inside—a phenomenon known as self-shielding. Even with an optimally thin sample, the neutron beam emerging from a spallation source[‡] is not mono-energetic; it arrives as a broad burst whose individual energies must be reconstructed from flight time, leaving the experimenter with transmission ratios or capture yields, not cross sections themselves.

The remainder of this section explains how transmission and capture experiments are conducted, how the resulting spectra are converted via Bayesian generalized least squares into resonance energies, widths, and their covariance matrix, and finally how those parameter uncertainties are propagated into the cross-section space used by transport codes.

2.2.1 Time-of-Flight Experiments

A time-of-flight (TOF) experiment converts distance and clock time into a map of neutron energies [28, 29]. Each proton pulse striking the spallation target releases a flash of neutrons that travels a known flight path L toward a downstream detector. Because the distance is fixed, recording the arrival time t for every neutron allows its kinetic energy to be reconstructed through $E = \frac{1}{2}m_n(L/t)^2$. This reconstruction is appropriate for non-relativistic energies; only non-relativistic energies will be considered in this dissertation.

The experiment then inserts a thin sample foil to probe how those neutrons interact. With the foil removed the open-beam run records a reference spectrum

[‡]A spallation source produces neutrons by directing a high-energy proton beam onto a heavy metal such as tungsten or lead; the impact knocks dozens of neutrons out of each target nucleus, creating short, intense pulses with a broad energy spectrum [27].

$N_0(t)$. Re-inserting a foil of areal number density n (atoms cm⁻²) yields the sample-in spectrum $N(t)$.

After background subtraction and dead-time correction (using live-time scalers), the ratio

$$T_i = \frac{N(E_i)}{N_0(E_i)} \quad (2.5)$$

defines the experimental transmission in energy bin E_i . In the thin-sample limit the transmission is linked to the total microscopic cross section by

$$T_i = e^{-n\sigma_{\text{tot}}(E_i)}, \quad (2.6)$$

the Beer–Lambert law written for neutrons [30, 31]. Equation (2.6) shows why a direct cross-section measurement is impractical: for a resonance where σ_{tot} may jump by three orders of magnitude over a few eV, the exponential either leaves almost every neutron untouched—or removes almost all of them near the front surface, a self-shielding effect that hides the very probability one wishes to observe. Working with the ratio (2.5) retains the experimental information in a form that the R -matrix model can reproduce directly. In this formulation self-shielding is not deconvolved from the data; rather, it is accounted for within the forward R -matrix calculation (as implemented in SAMMY). Capture experiments follow the same beam geometry but replace the transmission ratio with the absolute γ -ray yield emitted by the sample. In both transmission and capture modes, the raw observable is calculated from R -matrix theory, closing the loop between measurement and evaluation.

Experimental uncertainties fall into two broad categories: statistical, which arise from the random nature of counting, and systematic, which stem from factors that bias many energy bins in the same direction. If N and N_0 were raw counts, they would each follow Poisson statistics with variance equal to their mean. After background subtraction and dead-time correction, however, the reported $N(E_i)$ and $N_0(E_i)$ are corrected quantities whose uncertainties inherit additional contributions from those

procedures. Propagating Poisson variances through Eq. (2.5) yields

$$(\delta T_i)^2 \simeq \frac{1}{N_0(E_i)} + \frac{1}{N(E_i)}, \quad (2.7)$$

which we take as the statistical variance occupying the diagonal of the experimental covariance matrix. More complete treatments can include correlated contributions from background and dead-time corrections, but this form captures the leading statistical behavior.

Systematic effects act across many bins at once. A drift in proton current can raise or lower the entire open-beam spectrum; an error in sample thickness scales every point of the sample-in spectrum; background radiation raises or lowers uncertainties across all energies from each channel. Such correlations are captured compactly by writing the total covariance in an implicit form

$$\mathbf{V} = v + g m g^\top, \quad (2.8)$$

where v is the diagonal matrix of statistical variances, g is a column vector whose i -th entry g_i is the sensitivity of bin i to a chosen systematic parameter (for example, the normalization constant), and m is the variance of that parameter. The outer product $g m g^\top$ therefore spreads the systematic uncertainty across all energies in the same pattern that the underlying systematic parameter would imprint, without the memory cost of storing a fully dense matrix.

The transmission vector $\mathbf{T} = (T_1, T_2, \dots, T_n)^\top$, together with its covariance $\mathbf{V}$, provides the experimental input to the Bayesian generalized least-squares (GLS) fit introduced in the next sub-section; the R -matrix model supplies a matching vector of calculated transmissions and the sensitivity matrix that links changes in resonance parameters to changes in $\mathbf{T}$.

2.2.2 Fitting Procedure for Resonance Parameters

The measured transmission vector $\mathbf{T}$ and its covariance matrix $\mathbf{V}$ are compared with R -matrix calculations to infer the underlying resonance parameters. For any trial parameter set $\mathbf{P} = (E_\lambda, \Gamma_{n,\lambda}, \Gamma_{\gamma,\lambda}, \dots)^\top$, the model produces a calculated transmission $\mathbf{T}_{\text{calc}}(\mathbf{P})$. A Bayesian generalized least-squares (GLS) algorithm, implemented in SAMMY [32], adjusts $\mathbf{P}$ until theory and experiment agree within their quoted uncertainties.

The fit starts from a prior parameter vector $\mathbf{P}_0$ —usually taken from a previous evaluation—and its prior covariance matrix $\mathbf{M}_0$. Linearizing the model about $\mathbf{P}_0$ introduces the sensitivity matrix

$$G_{ij} = \left. \frac{\partial T_i}{\partial P_j} \right|_{\mathbf{P}_0},$$

whose rows show how strongly each energy bin reacts to infinitesimal changes in each parameter. In this linear approximation the posterior covariance $\mathbf{M}'$ is obtained first, and the posterior mean $\mathbf{P}'$ is then evaluated with that result [12]:

$$\mathbf{M}' = (\mathbf{M}_0^{-1} + G^\top \mathbf{V}^{-1} G)^{-1}, \quad (2.9)$$

$$\mathbf{P}' = \mathbf{P}_0 + \mathbf{M}' G^\top \mathbf{V}^{-1} (\mathbf{T} - \mathbf{T}_{\text{calc}}(\mathbf{P}_0)). \quad (2.10)$$

Equation (2.9) updates the prior covariance by adding the information supplied by the experiment: the matrix $G^\top \mathbf{V}^{-1} G$ is large for parameters to which the data are sensitive and small for those they scarcely affect. Equation (2.10) then shifts the prior mean toward the measured value; the discrepancy $\mathbf{T} - \mathbf{T}_{\text{calc}}(\mathbf{P}_0)$ is weighted by the experimental precision $\mathbf{V}^{-1}$ and mapped back into parameter space through the sensitivity matrix, with the posterior covariance $\mathbf{M}'$ determining how far that shift can go.

SAMMY iterates Eqs. (2.9)–(2.10) until successive parameter vectors differ by less than a user-defined tolerance and the reduced χ^2 approaches unity.[§] The converged $\mathbf{P}'$ is the best estimate of the resonance energies and widths, and $\mathbf{M}'$ is their covariance matrix. Because the experimental covariance $\mathbf{V}$ enters explicitly in Eq. (2.9), any underestimated correlations—or approximations that the linear expansion cannot capture—propagate directly into $\mathbf{M}'$ and, in turn, into the cross-section covariances that transport simulations rely on.

2.2.3 Uncertainty Quantification for Evaluated Cross Sections

Many modern transport calculations can be run with only the fitted cross-section curve, but meaningful uncertainty quantification on transport observables requires a covariance matrix that accompanies the evaluated curve. Starting from the posterior parameter covariance $\mathbf{M}'$, evaluators obtain the evaluated cross-section covariance by a first-order Taylor expansion. For a given reaction channel x the evaluated cross section is the smooth function $\sigma_x(E; \mathbf{P}')$, and its gradient with respect to the resonance parameters is the row vector

$$\frac{\partial \sigma_x(E; \mathbf{P}')}{\partial \mathbf{P}}.$$

Propagating the small, correlated variations encapsulated in $\mathbf{M}'$ gives

$$C_x(E, E') = \frac{\partial \sigma_x(E; \mathbf{P}')}{\partial \mathbf{P}} \mathbf{M}' \frac{\partial \sigma_x(E'; \mathbf{P}')}{\partial \mathbf{P}}^\top, \quad (2.11)$$

the quantity written to ENDF File 33. Because resonance peaks are narrow but isolated, the cross section is locally almost linear in each parameter, so the first-order approximation captures nearly all of the evaluated variance in the resolved region.

[§]The reduced χ^2 is the weighted sum of squared residuals per degree of freedom, $\chi_{\text{red}}^2 \equiv (\mathbf{T} - \mathbf{T}_{\text{calc}})^\top \mathbf{V}^{-1} (\mathbf{T} - \mathbf{T}_{\text{calc}}) / \nu$, where ν is the number of data points minus the number of fitted parameters. A value near 1 indicates that the model reproduces the data within their stated uncertainties.

The simplicity of Eq. (2.11) masks several ways in which the propagated covariance can underestimate reality—issues documented in detail in the SAMMY guide [12]. First, the calculation of $\mathbf{M}'$ rests on ideal assumptions: the R -matrix model is complete, all resonances are present, every spin is correct, experimental corrections are perfect, etc. Second, Bayes’ equation for the covariance contains no measure of the quality of the fit; a poor fit leaves $\mathbf{M}'$ unchanged. Third, the experimental data set is not unique: different evaluators may choose or re-normalize different subsets of measurements, yet the uncertainty associated with that choice is not represented in $\mathbf{M}'$.

A missing resonance produces an especially misleading picture of uncertainty. When no parameters are available to describe a real level at energy E_λ , the fitted cross section at that energy is forced to lie on the smooth “background” formed by the wings of nearby peaks. The evaluated value $\sigma_x(E_\lambda; \mathbf{P}')$ is therefore far smaller than it should be. Because the first derivatives $\partial\sigma_x/\partial P_j$ that appear in Eq. (2.11) scale roughly with the magnitude of the cross section itself, the propagated covariance turns out to be proportional to σ_x^2 . A resonance that is absent from the parameter list thus contributes twice to overconfidence: its peak height is missing, and the associated uncertainty—tied to that height by the propagation formula—shrinks in tandem. The evaluated cross-section covariance is smallest precisely where the true uncertainty is largest, an effect that will be demonstrated quantitatively in Chapter 3.

The nuclear-data community is well aware of these shortcomings. The SAMMY input manual lists keywords that let an evaluator enlarge the posterior covariance by hand, either by scaling every element of $\mathbf{M}'$ or by adding explicit normalization-error and background-error terms.[¶] Such adjustments do broaden the File-33 uncertainties, but they depend on expert judgement and are hard to reproduce exactly. The methods

[¶]A normalization error reflects uncertainty in the overall scale factor that converts detector counts to absolute probability; if the factor is high or low by, say, 1%, every point in the spectrum shifts by the same 1%. A background error reflects uncertainty in the spectrum that has been subtracted to remove ambient radiation and scatter from structural materials. Both errors therefore introduce correlated uncertainties that affect the entire spectrum together.

developed in this dissertation aim to replace that ad-hoc inflation with a transparent, data-driven procedure that quantifies the missing variance and restores it in a way anyone can repeat.

2.3 Physics-Informed Synthetic Data

Large, labelled data sets have transformed uncertainty analysis in many fields, but in neutron time-of-flight work they are difficult to obtain [33, 34, 35, 36, 37, 38, 39]. Beam time is limited, thin samples yield few counts, thick samples suffer from self-shielding, and the exact resonance parameters are never observed directly. A physics-based generator of synthetic data circumvents these limits: one can create as many virtual experiments as needed, each accompanied by the exact parameter set and cross sections that produced the counts.

Walton and co-workers have released an open-source generator that recreates this process in two controlled stages [15]. The first stage samples a complete, internally consistent vector of resonance parameters $\mathbf{P} = (E_\lambda, \Gamma_{n,\lambda}, \Gamma_{\gamma,\lambda}, \dots)$. Gaps between resonance energies are drawn from the Wigner distribution, whose mild “level-repulsion” prevents two states of the same spin and parity from sitting on top of each other and reproduces the spacing statistics observed in heavy nuclei. For each level, the neutron width is then sampled from the Porter–Thomas distribution—a chi-squared distribution with one degree of freedom that reflects the random mixing of many compound-nucleus configurations and therefore produces many narrow levels and a few broad ones. Finally, the radiative (gamma) width is selected from a log-normal spread centred on the systematics compiled by Brown *et al.* [40], ensuring that the total width satisfies $\Gamma_\lambda = \Gamma_{n,\lambda} + \Gamma_{\gamma,\lambda}$ while allowing realistic variation from level to level. The result is a synthetic resonance ladder statistically indistinguishable from real resolved-resonance data.

In the second stage the parameter vector is passed through a detailed forward model of the RPI transmission experiment. The R -matrix is evaluated exactly,

Doppler and resolution broadening are applied with the same kernels used for real data, multiple-scattering corrections are added, and Poisson noise appropriate to the specified beam intensity is folded in. The result is a spectrum of detector counts—open beam and sample in—together with a bookkeeping file that records the ground-truth $\mathbf{P}$ and the corresponding evaluated cross sections $\sigma_x(E; \mathbf{P})$. Walton verified the fidelity of the synthetic spectra by repeating Brown’s ^{181}Ta analysis with both real and synthetic data; the two produced statistically indistinguishable residual distributions and nearly identical posterior covariances [40, 15].

With this capability in hand, an evaluator can estimate by direct sampling how often a given method recovers the true parameters within their quoted uncertainties, explore failure modes such as missing resonances under controlled conditions, and train machine learning models that require large samples.

Chapter 3

Measuring Accuracy of Evaluated Cross-Section Covariances

The preceding chapters argued that how much confidence we place in a neutron-interaction library matters just as much as the best-estimate values themselves. Chapter 2 showed, however, that the covariance matrices produced by the widely used code SAMMY can be badly under-estimated whenever the ideal generalised least-squares (GLS) assumptions are not met. The natural next step—taken up in this chapter—is to measure how large those errors become for specific, realistic departures from ideality and, in doing so, to identify the modelling choices that deserve an evaluator’s closest attention.

To that end, we introduce a fully reproducible workflow that couples the physics-informed synthetic data generator SYNDAT with an empirical evaluation of the cross-section covariance [15]. In practice, we generate an unlimited number of labeled time-of-flight spectra from synthetic resonance parameters, use those samples to build an empirical evaluated covariance matrix that is correct by design, and then compare it to the covariance that SAMMY would report for the same experimental configuration. After presenting the workflow in detail, we apply it to three case

studies, closing with the main lessons for evaluators and for the automated pipeline developed in Chapter 4.

3.1 Procedure for computing empirical ECSCM

Suppose we wish to assess the accuracy of a SAMMY evaluated cross-section covariance estimate (ECSCM), C_S , for some set of experimental conditions and model assumptions $\mathcal{A}$. We begin by computing an empirical ECSCM, C_E , against which C_S can be compared. Empirical ECSCM, C_E , is computed as follows.

- (1) Define $\mathcal{A} = \{P, \bar{P}, M, V^*, V, \mathcal{E}, \dots\} \equiv$ the set of all experimental conditions and model assumptions relevant to the evaluation of C_S including but not limited to

$P^* \equiv$ the true resonance parameters from which the data are sampled,

$\bar{P} \equiv$ the prior resonance parameters means used for fitting the data,

$M \equiv$ the prior resonance parameter covariances used for fitting the data,

$V^* \equiv$ the true data covariances from which the data are sampled,

$V \equiv$ the estimated data covariances used for fitting the data, and

$\mathcal{E} = \{E_j\}_1^J \equiv$ a user-defined energy grid of J energies.

The purpose of this step is to ensure that the conditions for computing C_E are identical to those for computing C_S . As such, $\mathcal{A}$ should contain all variables for controlling data generation, parameter fitting, and ECSCM calculation. Defining $\mathcal{A}$ represents setup of the ECSCM comparison study.

- (2) Compute $\sigma(P^*, \mathcal{E}) = \{\sigma_j\}_1^J \equiv$ true cross sections evaluated with true resonance parameters P^* on energy grid $\mathcal{E}$.

- (3) With sufficiently large N , generate a set of N synthetic, transmission data samples $\mathcal{S} = \{\mathcal{D}_n\}_1^N$, where $\mathcal{D} = \{(\epsilon_w, \tau_w)\}_1^W \equiv$ a sample of W observed transmissions $\tau_{1:W}$ at energies $\epsilon_{1:W}$. It need not be the case that $W = J$.
- (4) For each $\mathcal{D}_n \in \mathcal{S}$, compute posterior resonance parameters $\bar{P}'_n$.
- (5) For each set of posterior resonance parameters $\bar{P}'_n \in \bar{P}'_{1:N}$, compute $\hat{\sigma}^s(\bar{P}'_n, \mathcal{E}) = \{\hat{\sigma}_j^n\}_1^J \equiv$ fitted cross sections for sample $\mathcal{D}_n$ with posterior resonance parameters $\bar{P}'_n$ on energy grid $\mathcal{E}$.
- (6) Construct a matrix $R_{N \times J}$ of cross section residuals by computing the residuals for each $\mathcal{D}_n \in \mathcal{S}$ over energy grid $\mathcal{E}$. The residual for the n -th sample at the j -th energy is given by

$$r_{n,j} = \sigma_j - \hat{\sigma}_j^n. \quad (3.1)$$

R can be built with the double loop: $\forall \mathcal{D}_n \in \mathcal{S}, \forall E_j \in \mathcal{E}$, compute $r_{n,j}$.

- (7) Finally, compute $C_E(\mathcal{A}) = \frac{1}{N} R^T R \equiv$ a $J \times J$ empirical ECSCM for setup $\mathcal{A}$.

In the limit of infinite data samples N , empirical ECSCM C_E is the analytically correct ECSCM for setup $\mathcal{A}$.

3.2 SAMMY vs. Empirical uncertainty quantification example

The impact of underspecifying the model on SAMMY uncertainty quantification (UQ) can be expressed visually. Consider a sample of Ta-181 from a user-defined ladder of $k = 15$ resonances as depicted in Figure 3.1. Now suppose that an underspecified model is used to fit the data. Specifically, only $\hat{k} = 12$ sets of resonance parameters are modeled for fitting. Suppose prior parameter estimates are set to correspond to the $\hat{k} = 12$ largest resonances. Under this setup, not only does SAMMY predictably miss the small resonance located at 51 eV, it significantly underestimates the evaluated

cross section uncertainty at 51 eV. This is because C_S is calculated by first-order propagation of uncertainty from M' , the posterior RPCM, and M' in this case includes no information about the resonance at 51 eV. Computing C_E using the empirical method presented in Section 3.1 accurately characterizes the uncertainty in the region of the missed resonance. Figure 3.2 includes standard error bands computed from the main diagonals of the SAMMY and empirical ECSCMs. In this example, the theoretical cross-section is well outside the ± 2 SAMMY standard error bands and is within the ± 2 empirical standard error bands. SAMMY reports relative uncertainty $< 2\%$ on the evaluated cross section at 51 eV, when in fact the reported relative uncertainty should be $\approx 16\%$.

This simple example demonstrates the potential utility of synthetic data and empirical UQ. Namely, the proposed methodology allows the nuclear data community to study the impact of common, real-world assumption violations on the accuracy of SAMMY ECSCM estimates when performing evaluations on real data. How does the size of a missed resonance, the degree to which systematic variances are misreported, or the choice of R-matrix formalism impact SAMMY ECSCM accuracy? By testing these questions with representative synthetic data, the evaluator can have confidence about the effects these assumption violations will have on their real-world evaluations.

3.3 Single-value metric for measuring SAMMY ECSCM accuracy

In the section 3.2 example, we used error bands to visualize how underspecifying the model can impact SAMMY UQ. Comparing SAMMY and empirical evaluated cross section relative uncertainties allowed us to quantify the impact of underspecifying the model on SAMMY UQ at a particular energy, but did not provide a holistic assessment of the accuracy of a SAMMY ECSCM over the entire energy region. Additionally, the relative uncertainties contain information only from the ECSCM main diagonal.

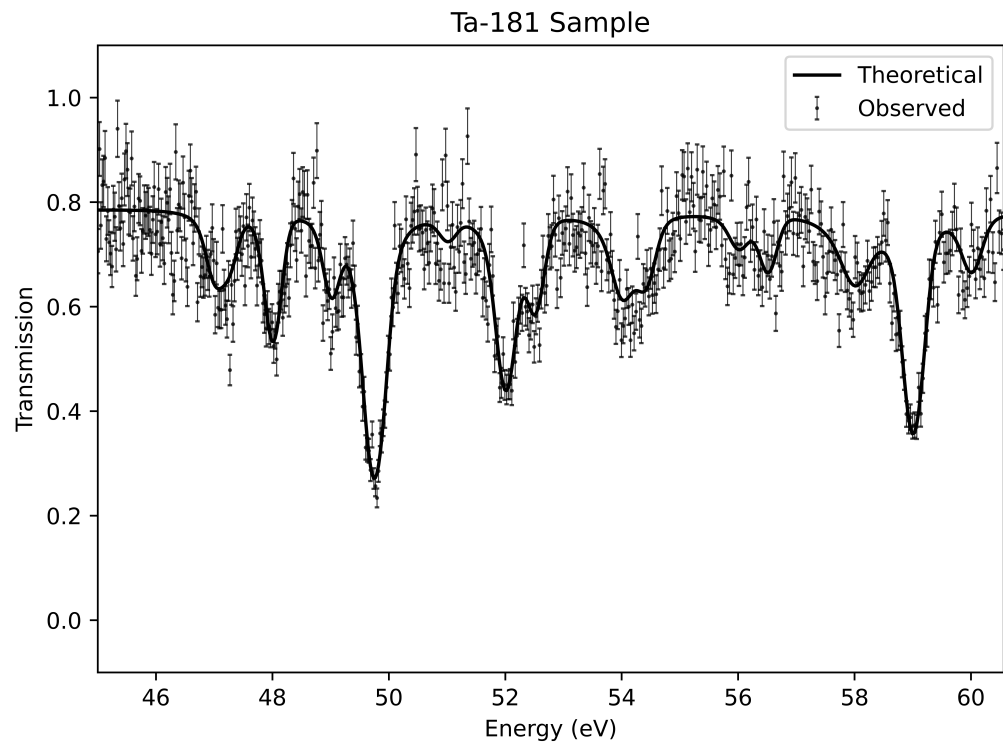


Figure 3.1: A synthetic sample of Ta-181 from a user-defined ladder of fifteen resonances.

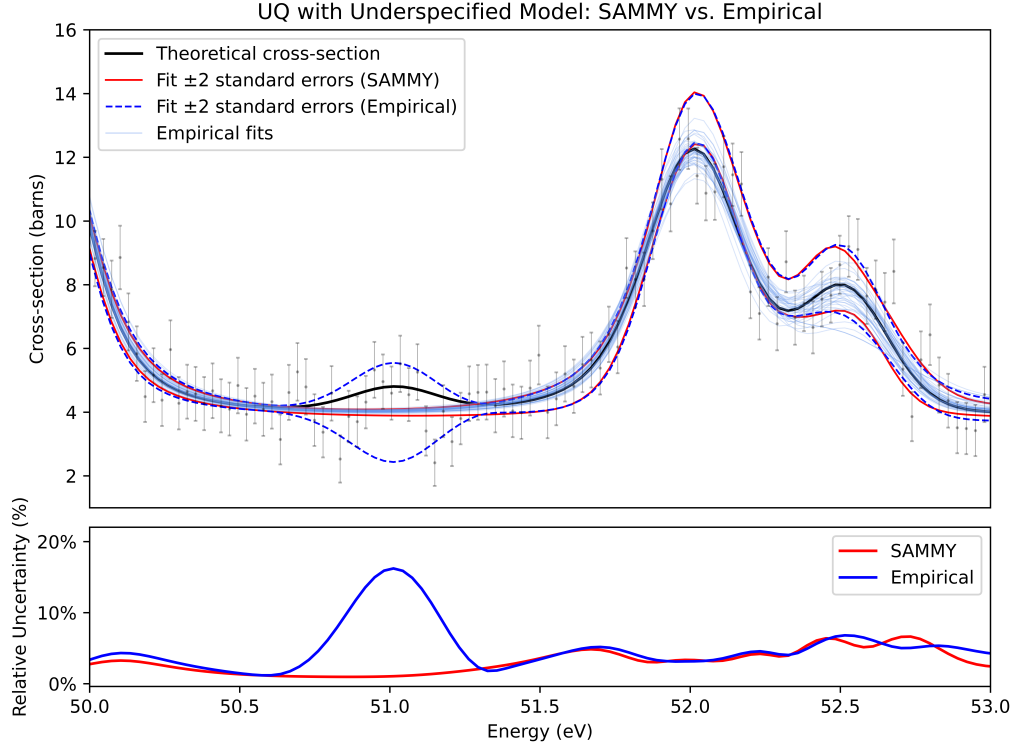


Figure 3.2: In this example, SAMMY is used to fit an underspecified model. Not only does SAMMY miss the resonance at 51 eV, it significantly overestimates the confidence in the fit at 51 eV. Building an empirical ECSCM using the method presented in section 3.1 results in an accurate uncertainty quantification and allows us to assess the impact of underspecifying the model on the accuracy of the SAMMY UQ.

A single-value metric which quantifies the accuracy of an entire SAMMY ECSCM for a given setup, $\mathcal{A}$, would be preferred. To this end, we propose the Frobenius norm (F-norm) of the difference between SAMMY ECSCM, $C_S(\mathcal{A})$, and empirical ECSCM, $C_E(\mathcal{A})$:

$$\|C_E - C_S\|_F = \sqrt{\sum_{i=1}^J \sum_{j=1}^J (C_{E_{i,j}} - C_{S_{i,j}})^2} \quad (3.2)$$

where J is equal to the number of elements in user-defined energy-grid $\mathcal{E} = \{E_j\}_1^J$. The F-norm of the difference matrix represents the Euclidean distance between the two matrices when considered as points in high dimensional space. It is sensitive to element-wise differences between C_S and C_E , is always non-negative, is symmetric with respect to C_E and C_S , has a meaningful zero value (i.e., $\|C_E - C_S\|_F = 0 \implies C_S = C_E$), and is computationally efficient. Other metrics were considered, but the natural interpretation and other desirable properties of $\|C_E - C_S\|_F$ make it preferable to its alternatives. It is the metric we use in all forthcoming case studies for assessing the accuracy of $C_S(\mathcal{A})$ for various setups $\mathcal{A}$.

3.4 Validating SAMMY ECSCM estimates when GLS assumptions are met

In the limit of infinite data samples N , empirical ECSCM $C_E(N)$ is accurate for setup $\mathcal{A}$. If C_S were fixed for any particular setup $\mathcal{A}$, then a simulation-based convergence test to demonstrate the accuracy of SAMMY under ideal conditions would be straightforward. We would compute $\|C_E(N) - C_S\|_F$ for increasing values of N to show that

$$\mathcal{L} \equiv \lim_{N \rightarrow \infty} \|C_E(N) - C_S\|_F = 0. \quad (3.3)$$

Such an approach is not possible. SAMMY computes posterior RPCM M' and posterior parameter estimates $\bar{P}'$ with the equations

$$(M')^{-1} = M^{-1} + G^t V^{-1} G \quad (3.4)$$

and

$$\bar{P}' = \bar{P} + M' G^t V^{-1} (D - \bar{T}) \quad (3.5)$$

where, $\bar{P}$ are the prior mean parameter values, V is the estimated data covariance matrix, G is the sensitivity matrix, D are the observed cross sections, and $\bar{T}$ are the theoretical cross sections evaluated at $\bar{P}$. We draw attention to the fact that M' is a function of M , G , and V but not of the data itself. Therefore, one might expect $C_S(\mathcal{D}_n, V^*, P, M)$ to be fixed $\forall \mathcal{D}_n \in \mathcal{S}$ if each dataset $\mathcal{D}_n \sim \text{i.i.d. } \mathcal{N}(T(P^*), V^*)$.

Nonetheless, $C_S(\mathcal{D}_n, V^*, P, M)$, unlike M' , is in fact sensitive to the input $\mathcal{D}_n$. This is because

$$C_S = \mathcal{J} M' \mathcal{J}^t \quad (3.6)$$

where each element of the Jacobian $\mathcal{J}$ for energy j and resonance parameter k is defined as

$$\mathcal{J}_{j,k} = \left. \frac{\partial \hat{\sigma}_j}{\partial P_k} \right|_{P=\bar{P}'} \quad \forall j \in \{1, \dots, J\}, \forall k \in \{1, \dots, K\}. \quad (3.7)$$

Notice that the partial derivative of the fitted cross section at some energy j with respect to some resonance parameter k is evaluated at average posterior resonance parameters $\bar{P}'$ which *are* a function of the data. Therefore, $C_S(\mathcal{D}_n, V^*, P, M)$ does depend on $\mathcal{D}_n$, C_S is not fixed for any particular setup $\mathcal{A}$, and $\mathcal{L} \neq 0$ even when GLS assumptions are met.

Then for any setup $\mathcal{A}$, there are two sources of variation that contribute to $\mathcal{L}$: (1) the average difference between $\lim_{N \rightarrow \infty} C_E(N)$ and C_S due to systematic inaccuracies

in C_S , and (2) the average difference between $\lim_{N \rightarrow \infty} C_E(N)$ and C_S due to statistical variation in C_S . In the case of all GLS assumptions being met, the first source of variation disappears, and

$$\mathcal{L} \equiv \lim_{N \rightarrow \infty} \|C_E(N) - C_S\|_F = E [\|C_S(\mathcal{D}_i^b) - C_S(\mathcal{D}_j^b)\|_F] \quad (3.8)$$

where $E [\|C_S(\mathcal{D}_i^b) - C_S(\mathcal{D}_j^b)\|_F]$ is the mean F-norm of the difference of two SAMMY ECSCMs computed with i.i.d. samples $\mathcal{D}_i$ and $\mathcal{D}_j$. The mean value $E [\|C_S(\mathcal{D}_i^b) - C_S(\mathcal{D}_j^b)\|_F]$ can be accurately and efficiently approximated by

(1) generating B pairs of samples $\{\mathcal{D}_i^b, \mathcal{D}_j^b\}_1^B$,

(2) computing the F-norms

$$\mathcal{F}_b = \|C_S(\mathcal{D}_i^b) - C_S(\mathcal{D}_j^b)\|_F \quad \forall b \in \{1, \dots, B\}, \quad (3.9)$$

(3) and computing the sample mean F-norm

$$\bar{\mathcal{F}} = \frac{1}{B} \sum_{b=1}^B \mathcal{F}_b \approx E [\|C_S(\mathcal{D}_i^b) - C_S(\mathcal{D}_j^b)\|_F]. \quad (3.10)$$

In summary, a simulation-based validation of SAMMY ECSCMs when all GLS assumptions are met is achieved by showing

$$\lim_{N \rightarrow \infty} \|C_E(N) - C_S\|_F = E [\|C_S(\mathcal{D}_i^b) - C_S(\mathcal{D}_j^b)\|_F] \approx \bar{\mathcal{F}}. \quad (3.11)$$

A convergence test was performed for numbers of samples ranging from $N = 100$ to $N = 50,000$. The test used two-resonance Ta-181 samples generated from SYNDAT. Table 3.1 summarizes key input parameters. To ensure alignment with GLS assumptions, transmission values for the convergence test were sampled using a simple normal noise model with user-defined V^* , and the data were fit with $V = V^*$.

*Here “ppeV” denotes points per electron-volt of energy.

Table 3.1: Summary of convergence test input parameters and their values

Input parameters	Value
Resonance 1 parameters: $\{E_{\lambda_1}, \Gamma_{\gamma_1}, \Gamma_{n_1}\}$	$\{44.0, 1\text{e-}3, 20\}$
Resonance 2 parameters: $\{E_{\lambda_2}, \Gamma_{\gamma_2}, \Gamma_{n_2}\}$	$\{46.0, 1\text{e-}3, 20\}$
SAMMY prior mean parameters (did not fit Γ_γ): $\{E_{\lambda_1}, \Gamma_{n_1}, E_{\lambda_2}, \Gamma_{n_2}\}$	$\{44.0, 20, 46.0, 20\}$
Energy grid parameters: $\{E_0, \text{ppeV}, J\}$	$\{40.0, 40.0, 498\}$
Energy grid: $\mathcal{E}$	$\left\{E_n = E_0 + \frac{n-1}{\text{ppeV}}\right\}_{n=1}^J$

Figure 3.3 summarizes the results. $\bar{\mathcal{F}} \approx E [||C_S(\mathcal{D}_i^b) - C_S(\mathcal{D}_j^b)||_F]$ was approximated using $B = 60$ pairs of samples. The test confirms that

$$\lim_{N \rightarrow \infty} ||C_E(N) - C_S||_F = E [||C_S(\mathcal{D}_i^b) - C_S(\mathcal{D}_j^b)||_F] \approx \bar{\mathcal{F}}. \quad (3.12)$$

That is, SAMMY ECSCMs are in fact accurate when all GLS assumptions are met.

3.5 SAMMY ECSCM accuracy case studies

3.5.1 Case Study 1: Impact of the size of a missing resonance

The section 3.2 example shows how underspecifying the model can cause inaccuracies in SAMMY UQ. We can quantify the impact of the size of a missed resonance on the accuracy of corresponding SAMMY ECSCM estimates with the following simulation experiment.

- (1) Define a series of B true resonance ladders $\mathcal{P} = \{P_b^*\}_1^B$ with one of the resonances varying in size from ladder to ladder. E.g., vary Γ_n for one of the resonances in each ladder while keeping all other parameters the same.

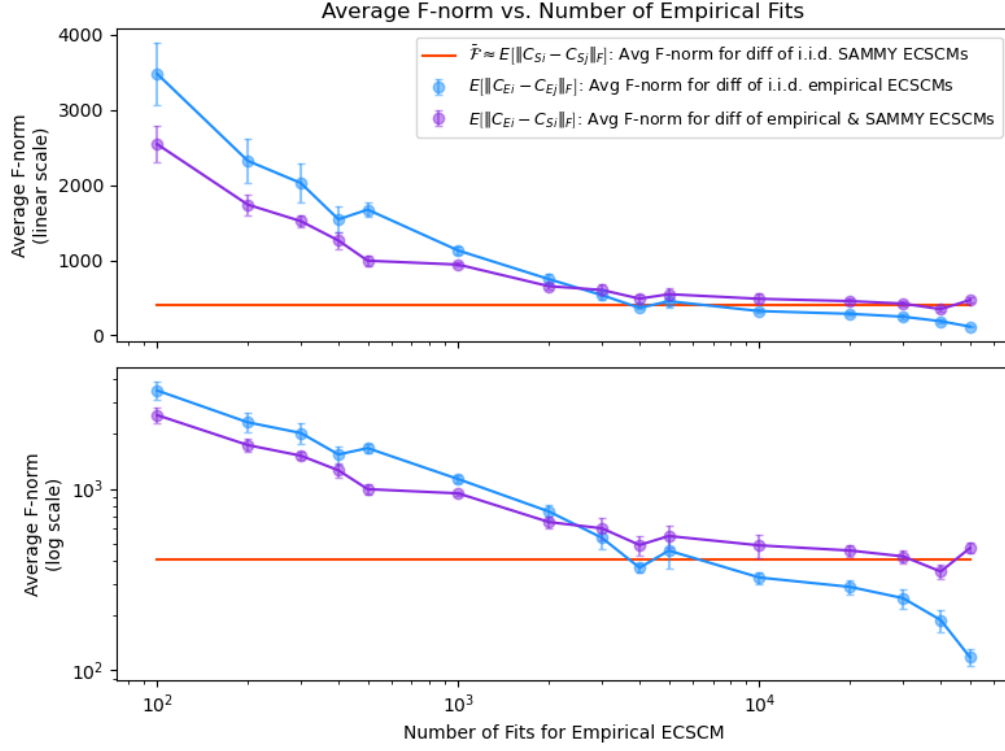


Figure 3.3: The convergence test confirms $\lim_{N \rightarrow \infty} \|C_E(N) - C_S\|_F = E[\|C_S(\mathcal{D}_i^b) - C_S(\mathcal{D}_j^b)\|_F] \approx \bar{\mathcal{F}}$ (i.e., SAMMY ECSCMs are accurate) when all GLS assumptions are met. Additionally, the plot confirms the F-norm of the difference of two independent empirical ECSCMs approaches zero for increasing number of samples N .

- (2) Define prior resonance parameters P to exclude parameters for the missing resonance included in true resonance ladders $\mathcal{P}$.
- (3) For each $P_b^* \in \mathcal{P}$, deploy the procedure outlined in section 3.1 to compute $\|C_E(P_b^*) - C_S(P_b^*)\|_F$.
- (4) Examine the relationship between F-norms $\|C_E(P_b^*) - C_S(P_b^*)\|_F$ and the resonance size variable (e.g., Γ_n) for each $P_b^* \in \mathcal{P}$.

A missing resonance case study was performed with values of Γ_n of the missing resonance corresponding to missing resonances between zero and approximately four times the magnitude of measurement uncertainties of observed transmissions near the missing resonance. The case study used two-resonance Ta-181 samples generated from SYNDAT, and fit only the first resonance. Table 3.2 summarizes key input parameters.

Table 3.2: Summary of missing resonance case study input parameters and their values

Input parameter	Value
Resonance 1 parameters: $\{E_{\lambda_1}, \Gamma_{\gamma_1}, \Gamma_{n_1}\}$	$\{44.0, 1\text{e-}1, 10\}$
Missing resonance fixed parameters: $\{E_{\lambda_2}, \Gamma_{\gamma_2}\}$	$\{46.0, 1\text{e-}1\}$
Missing resonance varied parameters: $\{\Gamma_{n_2}\}$	$\{0, 0.005, 0.010, 0.015, 0.020, 0.025, 0.030, 0.035, 0.040\}$
SAMMY prior mean parameters (did not fit Γ_γ): $\{E_{\lambda_1}, \Gamma_{n_1}, E_{\lambda_2}, \Gamma_{n_2}\}$	$\{44.0, 10, 46.0, 10\}$
Energy grid parameters: $\{E_0, \text{ppeV}, J\}$	$\{40.0, 40.0, 498\}$
Energy grid: $\mathcal{E}$	$\left\{E_n = E_0 + \frac{n-1}{\text{ppeV}}\right\}_{n=1}^J$

Figure 3.4 summarizes the results. The case study indicates

$$\|C_E(\Gamma_n) - C_S(\Gamma_n)\|_F \propto \Gamma_n, \quad (3.13)$$

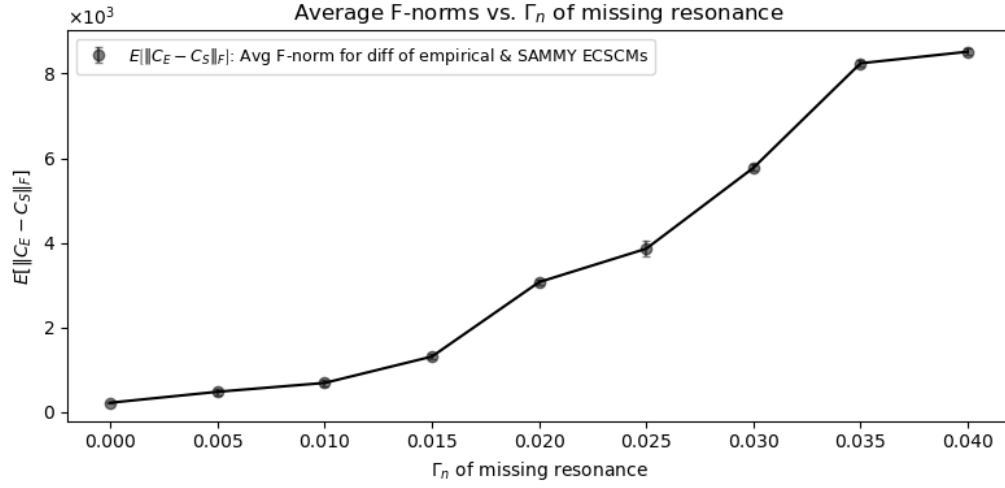


Figure 3.4: The missing resonance case study shows SAMMY ECSCM inaccuracy is roughly proportional to the size of the missing resonance in the underspecified model.

where Γ_n defines the size of the missing resonance. That is, the inaccuracy of SAMMY ECSCM estimates in the underspecified model is roughly proportional to the size of the missing resonance.

3.5.2 Case Study 2: Impact of choice of R-matrix formalism

As discussed previously, the Single-Level Breit-Wigner (SLBW) approximation does not model interactions between closely spaced resonances. Applying this approximation when fitting cross-sections can lead to poor fits in the regions between resonances, particularly at low energies (see Figure 3.5). Given changes in resonance widths and resonance level densities, SLBW becomes a more adequate approximation at higher energies. We can quantify the impact of using the SLBW approximation at various energies on the accuracy of SAMMY ECSCM estimates with the following simulation experiment.

- (1) Define a series of B resonance ladders $\mathcal{P} = \{P_b^*\}_1^B$ of two nearby resonances with the location of the first resonance varying from ladder to ladder. Maintain the absolute distance between the resonances from ladder to ladder.
- (2) For each $P_b^* \in \mathcal{P}$, deploy the procedure outlined in section 3.1 to compute $\|C_E(P_b^*) - C_S(P_b^*)\|_F$, using a Reich-Moore (XCT) model for data generation and an SLBW model for fitting.
- (3) Examine the relationship between F-norms $\|C_E(P_b^*) - C_S(P_b^*)\|_F$ and the location of the first resonance for each $P_b^* \in \mathcal{P}$.

The R-matrix formalism case study was performed with first resonance locations ranging from $E_\lambda = 50$ eV to $E_\lambda = 400$ eV. The case study used two-resonance Ta-181 samples generated from SYNDAT using a Reich-Moore model, and fit the data with both Reich-Moore and SLBW models. Table 3.3 summarizes key input parameters.

Figure 3.6 summarizes the results. The case study indicates SAMMY ECSCMs are innaccurate when using an SLBW approximation to fit XCT data samples from

Table 3.3: Summary of R-matrix formalism case study input parameters and their values

Input parameter	Value
Resonance 1 shape parameters: $\{\Gamma_{\gamma_1}, \Gamma_{n_1}\}$	$\{1\text{e-1}, 10\}$
Resonance 2 shape parameters: $\{\Gamma_{\gamma_2}, \Gamma_{n_2}\}$	$\{1\text{e-1}, 10\}$
Resonance 1 location parameters: $\{E_{\lambda_1}\}$	$\{50, 60, 70, 80, 90, 100, 200, 300, 400\}$
Resonance 2 location parameters: $\{E_{\lambda_2}\}$	$E_{\lambda_1} + 2$
SAMMY prior mean parameters (did not fit Γ_γ): $\{E_{\lambda_1}, \Gamma_{n_1}, E_{\lambda_2}, \Gamma_{n_2}\}$	$\{E_{\lambda_1}, 10, E_{\lambda_2}, 10\}$
Energy grid parameters: $\{E_0, \text{ppeV}, J\}$	$\{E_{\lambda_1} - 4, 40.0, 498\}$
Energy grid: $\mathcal{E}$	$\left\{E_n = E_0 + \frac{n-1}{\text{ppeV}}\right\}_{n=1}^J$

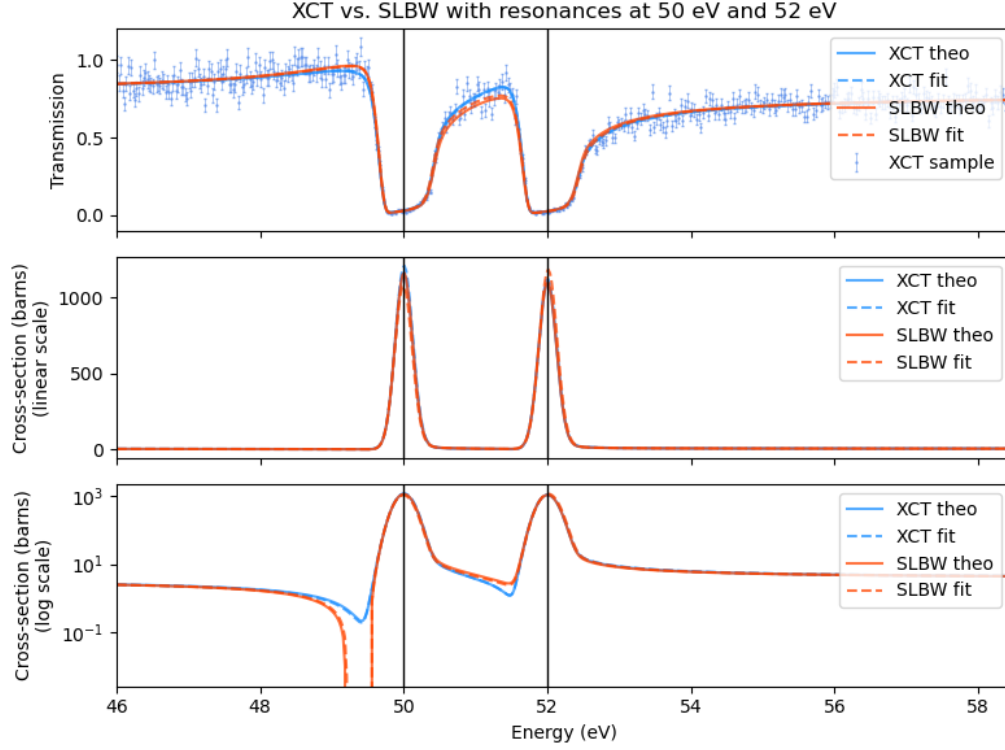


Figure 3.5: Data is generated using a Reich-Moore (XCT) model for resonances close in energy. The data are fit using both XCT and Single-Level Breit-Wigner (SLBW) models. SLBW does not model interactions between the resonances and results in a poor fit.

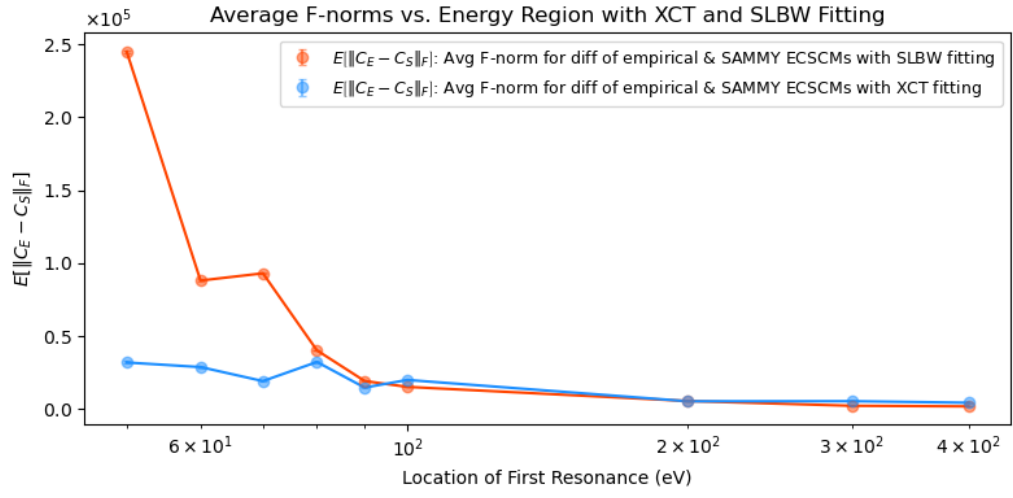


Figure 3.6: The R-matrix formalism case study generated XCT samples from two resonances 2 eV apart for different locations of the first resonance. The case study confirms SAMMY ECSCMs are inaccurate when using an SLBW approximation to fit the XCT data at low energies. The SLBW approximation becomes adequate at higher energies.

resonance ladders with resonances close enough to generate interference effects. The SLBW approximation performs just as well as the XCT formalism at higher energies.

3.5.3 Case Study 3: Impact of misreported systematic data covariance

SAMMY Guide section IV.D.3 [41] represents the experimental covariance as

$$V = v + g m g^T, \quad (3.14)$$

where v is the diagonal matrix of statistical counting variances, m is the covariance of data-reduction parameters, and g collects the partial derivatives of the calculated observable with respect to those parameters. To isolate how misreporting the correlated term $g m g^T$ affects the accuracy of evaluated cross-section covariances (ECSCMs) and, ultimately, application uncertainty, we switch here from the Ta-181 setup of the first two case studies to the single-resonance ^{238}U analytic benchmark Joshi et al. [42]. That benchmark provides closed-form derivatives for the k -eigenvalue with respect to resonance cross sections and parameters for a system built from the lowest s-wave resonance of ^{238}U (capture) and the lowest resonance of ^{239}Pu (fission). We use those analytic sensitivities to propagate resonance-parameter covariance directly to σ_k , the standard deviation of the effective neutron multiplication factor. This choice lets us map matrix-level errors such as $\|C_E - C_S\|_F$ onto a quantity of immediate engineering interest, $\Delta\sigma_k$.

Let $\mathcal{W} = \{w_b\}$ be a set of scale factors that deliberately under- or over-state the systematic term in (3.14). For each w_b we repeat the empirical-vs-SAMMY comparison of Section 3.1 but with the ^{238}U benchmark as the forward model:

- (1) Adopt the one-level ^{238}U capture resonance from the ^{238}U benchmark and its surrounding analytic setup. Generate synthetic transmission data with the true data covariance $V^* = v + g m g^T$ (fixed across w_b) and fit each dataset

in SAMMY using the misreported covariance

$$V(w_b) = v + w_b(gmg^\top).$$

- (2) For each w_b , form the SAMMY posterior resonance-parameter covariance $M'_S(w_b)$ and its empirical counterpart $M'_E(w_b)$ (built from sample-to-sample fit residuals as in Section 3.1).
- (3) Using the analytic benchmark's Jacobian with respect to resonance parameters, $d \equiv \partial k / \partial P$, propagate each parameter covariance to the standard deviation of k via

$$\sigma_k(M') = \sqrt{d^\top M' d}. \quad (3.15)$$

The vector d follows directly from the closed-form sensitivities reported in Eqs. (4)–(7) and Table IV of the benchmark Joshi et al. [42].

- (4) Record (i) the matrix-space discrepancy $\Phi(w_b) \equiv \|C_E(w_b) - C_S(w_b)\|_F$ on the evaluation grid $\mathcal{E}$ and (ii) the application-space discrepancy

$$\Delta\sigma_k(w_b) \equiv \sigma_k(M'_E(w_b)) - \sigma_k(M'_S(w_b)). \quad (3.16)$$

Figure 3.7 summarizes how $\Delta\sigma_k$ responds to the scale applied to the systematic term. The curve is characteristically U-shaped: severe under-reporting ($w \ll 1$) and severe over-reporting ($w \gg 1$) both enlarge the gap between the empirically correct and GLS-propagated application uncertainties, while a broad plateau ($10^{-2} \lesssim w \lesssim 10^1$ in this setup) leaves $\Delta\sigma_k$ comparatively small. In practical terms, moderate mis-scalings in $gmg^\top$ tend to be absorbed by the fit, but aggressive mis-scalings drive noticeable bias in the k -uncertainty inferred from M'_S . These findings are consistent with prior evaluation studies that explicitly treated experimental covariance. In particular, Neudecker, Capote, and Leeb (2013) showed that including correlated experimental uncertainties is as important as including model-defect terms for

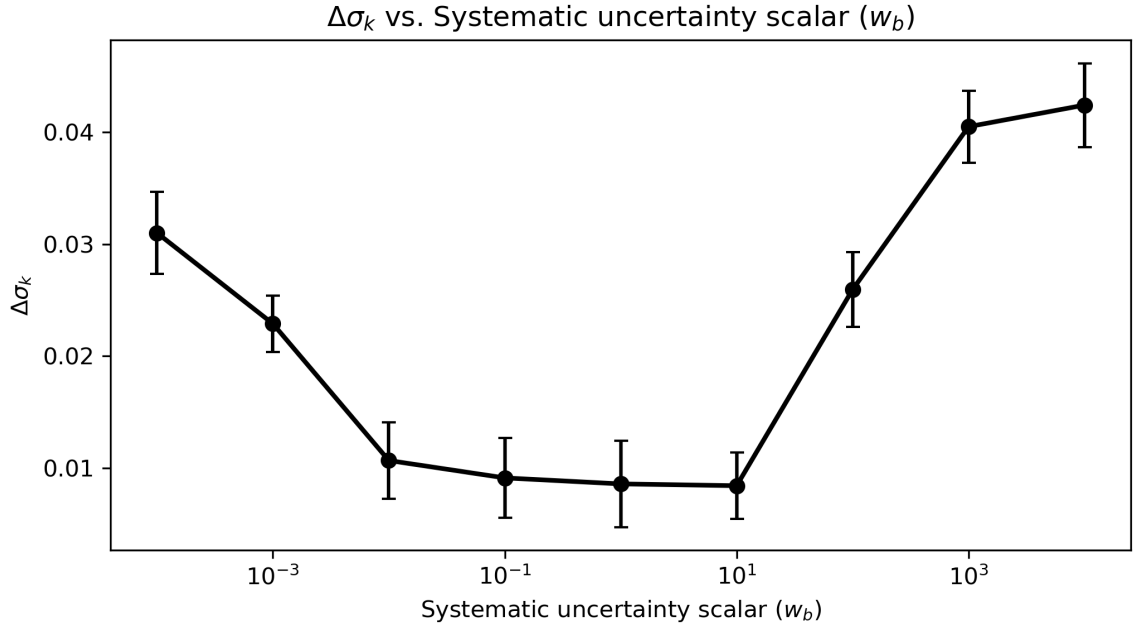


Figure 3.7: For the ^{238}U benchmark, the difference in the standard deviation of k between empirical and SAMMY parameter covariances, $\Delta\sigma_k$, is plotted against the multiplicative factor w applied to gmg^T during fitting. Points show the mean across repetitions; whiskers denote one standard deviation.

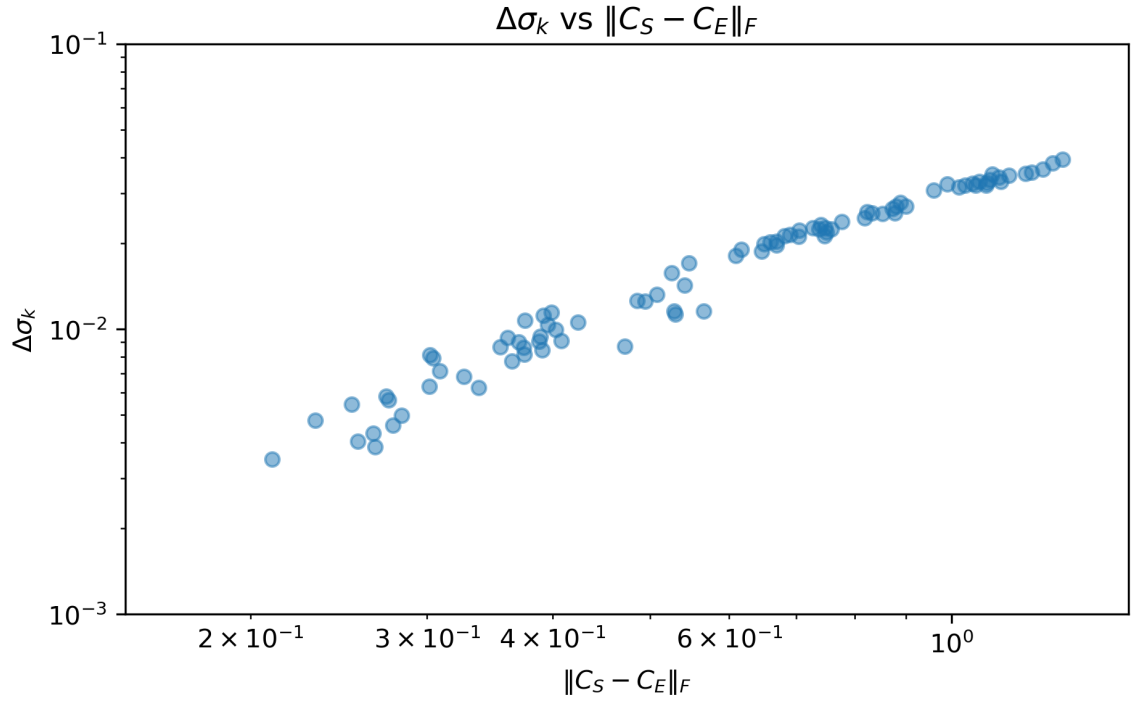


Figure 3.8: Empirical relationship between the Frobenius norm $\Phi = \|C_E - C_S\|_F$ and $\Delta\sigma_k$ for the ^{238}U benchmark. The trend indicates that increasing matrix-level error translates predictably into added uncertainty on k . Here $\Delta\sigma_k$ is the absolute difference in the standard deviation of k_{eff} ; no pcm conversion is applied.

producing realistic evaluated uncertainties, and that neglecting or approximating experimental correlations leads to unreasonably small evaluated uncertainties.[\[43\]](#)

Together with Case Studies 1–2, these ^{238}U results refine the earlier lesson: GLS propagation is comparatively robust to moderate misspecification of correlated experimental errors, but once the systematic term is badly misreported the ensuing bias in M'_S is large enough to matter for integral responses. The analytic Jacobian supplied by the benchmark makes that consequence explicit by relating ECSCM matrix discrepancies to $\Delta\sigma_k$.

3.6 Lessons learned

The three cases offer several lessons for our consideration. The convergence experiment of [Figure 3.3](#) establishes a useful baseline. With perfect knowledge of the data covariance, a complete resonance ladder, and the correct multilevel formalism, the evaluated cross-section covariance produced by SAMMY is statistically indistinguishable from empirical truth. Any subsequent discrepancy therefore signals a violation of those premises rather than a flaw in the underlying GLS machinery.

Case Study 1 shows that omitting even a single low-strength resonance inflates the Frobenius error almost linearly with the neutron width Γ_n of the missing level ([Figure 3.4](#)). Because the first-order propagation formula in SAMMY cannot account for a peak that is not parameterized, the local uncertainty collapses and the global Φ grows by orders of magnitude. In practice, then, diligent resonance identification is perhaps the most significant task for maintaining the accuracy of evaluated cross-section uncertainties.

When interference between neighbouring resonances is important, fitting with the single-level Breit–Wigner approximation forces the wings of adjacent levels to meet in the middle and drives a systematic mis-estimate of the between-resonance valley ([Figure 3.6](#)). The penalty fades when the two resonances are translated to a higher

energy, consistent with the expectation that the SLBW approaches the multilevel theory at sufficiently high E .

By contrast, misreporting the correlated experimental term $gm g^\top$ tends to be second-order until the scaling is extreme; the ^{238}U benchmark makes that statement quantitative by mapping $\|C_E - C_S\|_F$ to $\Delta\sigma_k$.

While this chapter focused on quantifying inaccuracies, the same synthetic-benchmark framework allows for corrective action. Because every run records the input configuration $\mathcal{A}$ and the resulting error $\Phi(\mathcal{A})$, the data form a labelled training set for machine-learning models that can predict the missing-variance contribution from easily computed features (level density, goodness-of-fit statistics, shape of residuals, *etc.*). Chapter 4 develops this idea: we train models that learn to augment SAMMY covariance matrices, restoring statistical fidelity without manual tuning.

Chapter 4

Automated Corrections for SAMMY Total Cross-Section Covariances

Chapter 3 used synthetic experiments to show that the generalized least-squares machinery in SAMMY is accurate only when every resonance in the analysis window is identified, the multilevel formalism matches the physics, and the data covariance is specified correctly. The purpose of the present chapter is to develop an automated, reproducible workflow that computes accurate diagonal elements of that covariance so that stated confidence levels agree with empirical coverage when matched with an appropriate probability model. The work here focuses on the main diagonal of covariance matrices, deferring cross-energy and cross-channel correlations to the next chapter.

4.1 Baseline setup and empirical targets

4.1.1 Synthetic ladder ensemble

The energy window for samples in the present study is restricted to 200–220 eV in ^{181}Ta , an interval wide enough to contain between one and seven resolved resonances and narrow enough that the entire covariance diagonal can be represented compactly. Using SYNDAT we generated 250 independent resonance ladders with statistical properties matching the resolved region of tantalum. The first 200 ladders form the training set; the remaining 50 are reserved for the test set.

4.1.2 Essentials of the AutoFit routine

A full GLS resonance fit is a non-linear, non-convex optimization problem in which the objective surface contains many shallow minima corresponding to different choices of resonance energies, widths, and counts. AutoFit, introduced by Walton et al. [17], is a repeatable, robust methodology for fitting resonances in this context.

The algorithm starts with an intentionally dense grid of candidate resonance energies, each supplied with a small neutron width and an average radiative width. With so many degrees of freedom the objective surface is convex and a Levenberg–Marquardt search converges reliably to a local minimum. Resonances whose neutron widths remain near zero after this step are discarded.

The remaining resonances are removed one at a time, each time re-optimising the continuous parameters. This procedure produces, in a single pass, a sequence of models of descending cardinality $k_{\max}, k_{\max} - 1, \dots, 1, 0$, each one locally optimal for its cardinality.

Finally the set of candidate models is scored. AutoFit scores model cardinalities using five-fold cross-validation (CV) on a user-specified goodness of fit metric. In the energy window considered here, CV yields the best predictive performance and therefore determines the accepted cardinality k_{final} .

The outcome is a locally optimal set of resonance parameters for the chosen model cardinality as well as the complete history of higher-cardinality fits. The intermediate solutions capture resonance structures that AutoFit identified but ultimately rejected under cross-validation. Rather than discard that information, this chapter leverages it directly to improve the accuracy of covariance estimates.

4.1.3 Developing target values

The first step in our reproducible work flow for computing accurate, evaluated cross-section variance estimates is to compute target values toward which we can train. This is achieved by computing empirical covariances using the procedure outlined in Section 3.1.

All covariance estimates in what follows are evaluated on a fixed covariance grid of fifty-one energies,

$$\mathcal{E} = \{E_j\}_{j=1}^{51}, \quad E_j = 200 + 0.4(j - 1) \text{ eV}, \quad (4.1)$$

which is fine enough to resolve the structure of every ladder yet small enough to keep the learning problem numerically tractable. The raw time-of-flight spectrum is sampled on a separate energy grid; the grid $\mathcal{E}$ specified here is used for post-fit ECSCM calculations inside SAMMY.

In Chapter 3 an empirical covariance was built by taking many independent transmission samples for one ladder, and fitting each sample once with a single, fixed-cardinality SAMMY model. Empirical covariances were then computed by

$$C_E(\mathcal{A}) = \frac{1}{N} R^T R \quad (4.2)$$

where C_E is the empirical ECSCM, $\mathcal{A}$ is the set of all experimental conditions and model assumptions relevant to the evaluation of the covariances, N is the number of i.i.d. transmission samples generated, and R be the residual matrix.

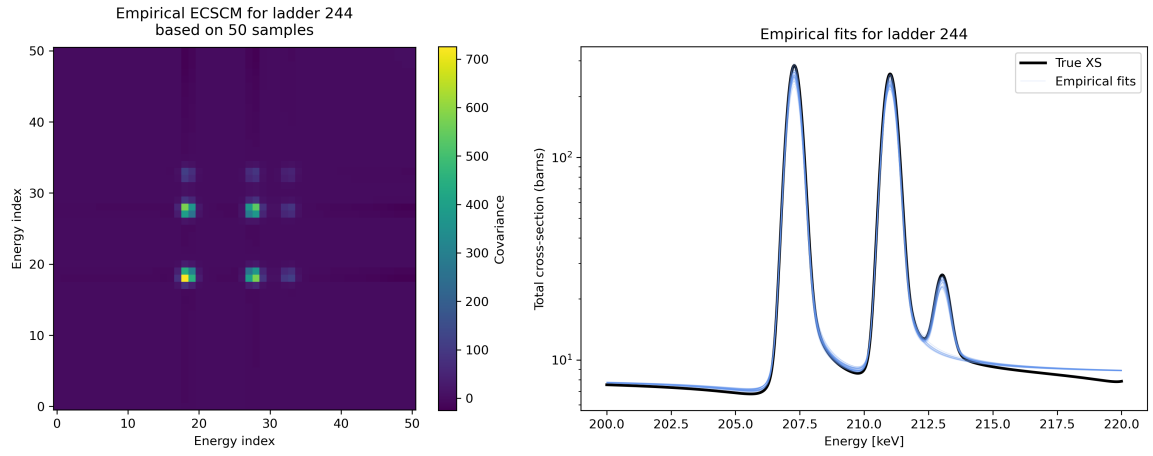


Figure 4.1: Left: The empirical ECSCM for ladder 244. Right: The true total cross-section and $n = 50$ fitted cross-sections using AutoFit on i.i.d. SYNDAT samples.

The present study follows the same variance-from-residuals procedure but replaces the single fit with a complete AutoFit cycle. For every ladder fifty independent transmission spectra are generated; each spectrum passes through the full three-stage AutoFit routine with dense feature bank, backward elimination, and five-fold cross-validation, to obtain a self-consistent posterior parameter vector. Evaluating those fifty vectors on the covariance grid $\mathcal{E}$ produces a residual matrix whose column-wise variance is a rank ≤ 50 estimate of the true diagonal covariance for that ladder. Using fifty samples ensures stable estimates of the first and second moments of the distribution of residuals at each energy index of each ladder while balancing the computational costs of producing more samples. Repeating the procedure for all 250 ladders yields a collection of 250 empirical ECSCMs for total cross-section. We extract the main diagonals of the ECSCMs as vectors that serve as the target values during training and as the ground truth for the final test-set comparison.

4.2 Choosing and validating a probability model

Having access to $n = 50$ residuals per ladder allows us to inductively explore suitable probability models. It's important to note that the probability model of these residuals is dependent on the fitting procedure (e.g., AutoFit) and cannot be safely generalized as a fundamental description of cross-section uncertainties independent of the fitting procedure. In fact, identifying an appropriate probability model is an essential step each time any part of the upstream workflow is modified (e.g., synthetic data generator, fitting procedure, etc.).

For every ladder $l \in \{1, \dots, L\}$ and every grid energy $E_j \in \mathcal{E}$ the AutoFit procedure yields a fitted cross-section $\hat{\sigma}_{jl}$ and an empirical variance $v_{jl} = C_{E,jj}^l$ extracted from the diagonal of that ladder's empirical ECSCM. What remains unspecified is the shape of the sampling distribution for $\tilde{\sigma}_{jl}$, the fitted cross-section that would be obtained if the experiment were repeated under identical conditions.

Let

$$\tilde{\sigma}_{jl} \mid \hat{\sigma}_{jl}, v_{jl}, D \sim D(\mu = \hat{\sigma}_{jl}, \varsigma^2 = v_{jl}),$$

where D is a parametric family (Normal, Student- t , log-Normal, *etc.*) whose location and scale parameters are fixed to the AutoFit mean and empirical variance. A good choice of D is one whose central prediction intervals achieve the nominal frequentist coverage simultaneously for every confidence level $c \in (0, 1)$.

This criterion can be tested as follows. Let D be a parametric family with cumulative distribution function $F_D(\cdot; \boldsymbol{\vartheta})$ and vector of free parameters $\boldsymbol{\vartheta}$. For every energy-ladder pair (j, l) we choose $\boldsymbol{\vartheta}_{jl}$ so that D reproduces the first two moments,

$$\mathbb{E}_D[\tilde{\sigma}_{jl}] = \hat{\sigma}_{jl}, \quad \text{Var}_D[\tilde{\sigma}_{jl}] = v_{jl}. \quad (4.3)$$

For the Normal family this reduces to $\boldsymbol{\vartheta}_{jl} = (\mu, v) = (\hat{\sigma}_{jl}, v_{jl})$; for the log-Normal one obtains $\boldsymbol{\vartheta}_{jl} = (\mu_{\ln}, \sigma_{\ln}^2)$ with

$$\sigma_{\ln}^2 = \ln(1 + v_{jl}/\hat{\sigma}_{jl}^2), \quad \mu_{\ln} = \ln \hat{\sigma}_{jl} - \frac{1}{2}\sigma_{\ln}^2.$$

Other families or finite mixtures are handled analogously.

Given a two-sided confidence level $c \in (0, 1)$ define

$$p_{\text{lo}} = \frac{1}{2}(1 - c), \quad p_{\text{hi}} = 1 - p_{\text{lo}}.$$

The $s = \lfloor 100c \rfloor\%$ prediction band for $\tilde{\sigma}_{jl}$ is then

$$I_{jl}^{(D, c)} = \left[F_D^{-1}(p_{\text{lo}}; \boldsymbol{\vartheta}_{jl}), F_D^{-1}(p_{\text{hi}}; \boldsymbol{\vartheta}_{jl}) \right], \quad (4.4)$$

where F_D^{-1} is the quantile function of D . Equation (4.4) collapses to the familiar form $\hat{\sigma}_{jl} \pm q\sqrt{v_{jl}}$, where q is the $(1 - \frac{\alpha}{2})$ quantile of the standardized error distribution, whenever D is symmetric about its mean.

Introduce the indicator

$$\delta_{jl}^{(D,c)} = \mathbf{1}\{\sigma_{jl}^{\text{true}} \in I_{jl}^{(D,c)}\},$$

and aggregate over all $M = L \times |\mathcal{E}|$ combinations to obtain the empirical coverage

$$\hat{\gamma}^{(D)}(c) = \frac{1}{M} \sum_{l=1}^L \sum_{j=1}^{|\mathcal{E}|} \delta_{jl}^{(D,c)}. \quad (4.5)$$

An ideal probability model satisfies $\hat{\gamma}^{(D)}(c) \xrightarrow{P} c$ for every $c \in (0, 1)$; in finite samples we seek $\hat{\gamma}^{(D)}(c)$ as close as possible to the identity line $y = x$. For the present study, we evaluate (4.5) on the grid $c \in \{0.60, 0.65, \dots, 0.90, 0.95\}$ for each candidate family $D \in \mathcal{D}$.

The Normal model serves as a suitable base case. Figure 4.2 shows that the Normal model is imperfect. It is overly conservative for confidence levels $c \geq 0.70$ and overly precise for confidence levels $c \leq 0.65$. Other distribution families should be included in $\mathcal{D}$ for testing. Next we might include distributions such as the Cauchy and the Lognormal to explore how changing tail weights or forcing positive values impacts the coverage performance. Figure 4.3 extends the calibration comparison to two families that alter the tails in opposite directions. The heavy-tailed Cauchy model is uniformly conservative, producing hit ratios well above the ideal $y = x$ line across the entire confidence spectrum. By contrast, the log-Normal model behaves much like the Normal, providing the user a strictly-positive alternative. These results illustrate how tail weight (Cauchy) and support restrictions (log-Normal) manifest in coverage, motivating further exploration of candidate probability models.

The Student- t family introduces a single scale-free parameter that controls tail thickness, the degrees of freedom $\nu > 0$. In a location-scale parameterization the probability density function is

$$f_t(x \mid \mu, \sigma^2, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma} \left[1 + \frac{1}{\nu} \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}}, \quad (4.6)$$

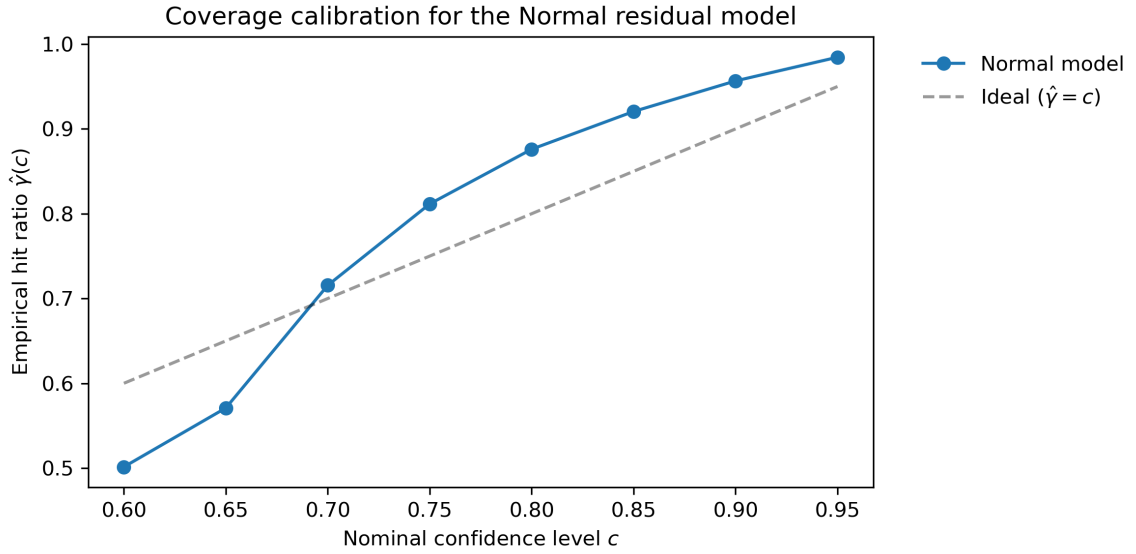


Figure 4.2: Empirical coverage $\hat{\gamma}(c)$ for the Normal residual model, evaluated at confidence levels 0.60–0.95. The dashed gray line marks perfect calibration ($\hat{\gamma} = c$). Points above the line indicate conservative intervals, while points below it correspond to over-confident bands. This baseline serves as a visual base case for the reader before comparing more complex probability models.

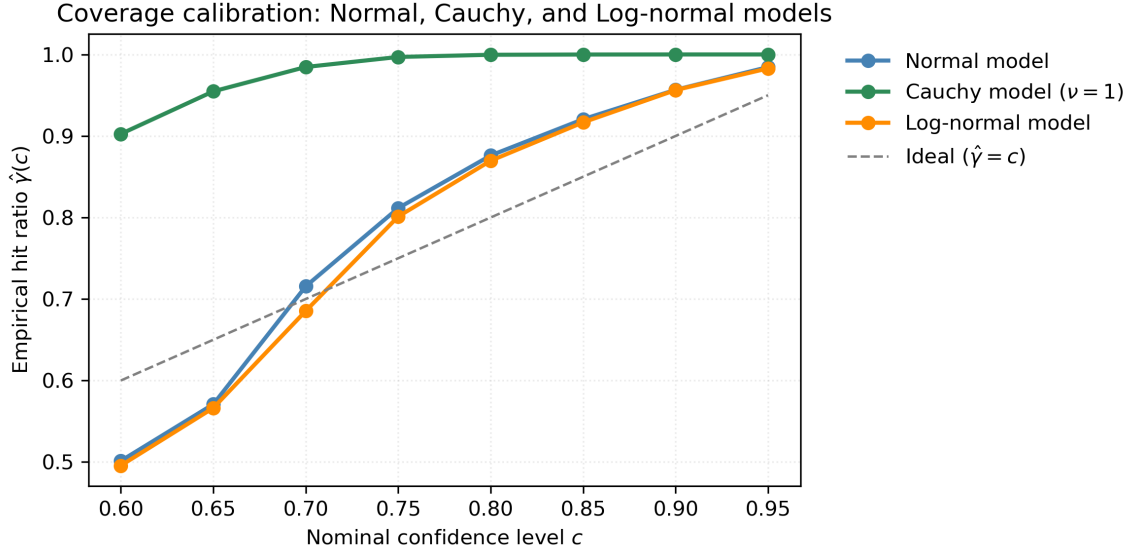


Figure 4.3: Empirical coverage $\hat{\gamma}(c)$ for three candidate residual models (Normal, Cauchy ($\nu = 1$), and log-Normal) evaluated at confidence levels $c \in \{0.60, 0.65, \dots, 0.95\}$. The dashed gray line indicates perfect calibration ($\hat{\gamma} = c$). The Cauchy model (green) is uniformly conservative; the Normal (blue) and log-Normal (orange) agree near $c \approx 0.70$ but diverge at the extremes, with the log-Normal remaining slightly conservative as $c \rightarrow 0.95$.

which reduces to the Cauchy distribution at $\nu = 1$ and converges to the Normal distribution as $\nu \rightarrow \infty$. Given a fitted cross-section $\hat{\sigma}_{jl}$ and empirical variance v_{jl} , moment matching sets the location to $\mu = \hat{\sigma}_{jl}$ and, for $\nu > 2$, the scale to

$$\sigma = \sqrt{v_{jl} \frac{\nu - 2}{\nu}}.$$

Two-sided 100c % prediction bands follow the quantile recipe of Eq. (4.4),

$$I_{jl}^{(t,c)} = [\hat{\sigma}_{jl} \pm t_{\nu}^{-1}\left(\frac{1+c}{2}\right) \sigma], \quad (4.7)$$

where t_{ν}^{-1} is the standard t quantile function with ν degrees of freedom.

Figure 4.4 traces the empirical hit ratio for $\nu \in \{0.5, 1, 2, 5, 10, \infty\}$. Two clear patterns emerge: (i) heavy-tailed models ($\nu \leq 2$) are uniformly conservative, clustered near $\hat{\gamma}(c) \approx 1$ for every c ; and (ii) even the Normal limit $\nu = \infty$ remains *overly conservative* once $c \gtrsim 0.70$. In other words, widening the tails only exacerbates the mismatch, while shrinking them (increasing ν) never narrows the intervals enough to match the empirical scale.

The persistence of over-coverage across the entire t family indicates that tail thickness alone cannot reconcile fitted and empirical variability. We therefore turn next to the generalized Gaussian family, whose shape parameter controls peakedness; this allows intervals narrower than Normal bands and may better match the observed residual spread.

The generalized Gaussian (also called the exponential-power) family is parameterised by a location μ , a scale $\alpha > 0$, and a shape parameter $\beta > 0$ that controls both peakedness and tail weight:

$$f_{\text{GG}}(x \mid \mu, \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left\{-\left|\frac{x-\mu}{\alpha}\right|^{\beta}\right\}. \quad (4.8)$$

Special cases include the Laplace distribution ($\beta = 1$) and the Normal distribution ($\beta = 2$); as $\beta \rightarrow \infty$ the density collapses to a uniform distribution over a fixed support.

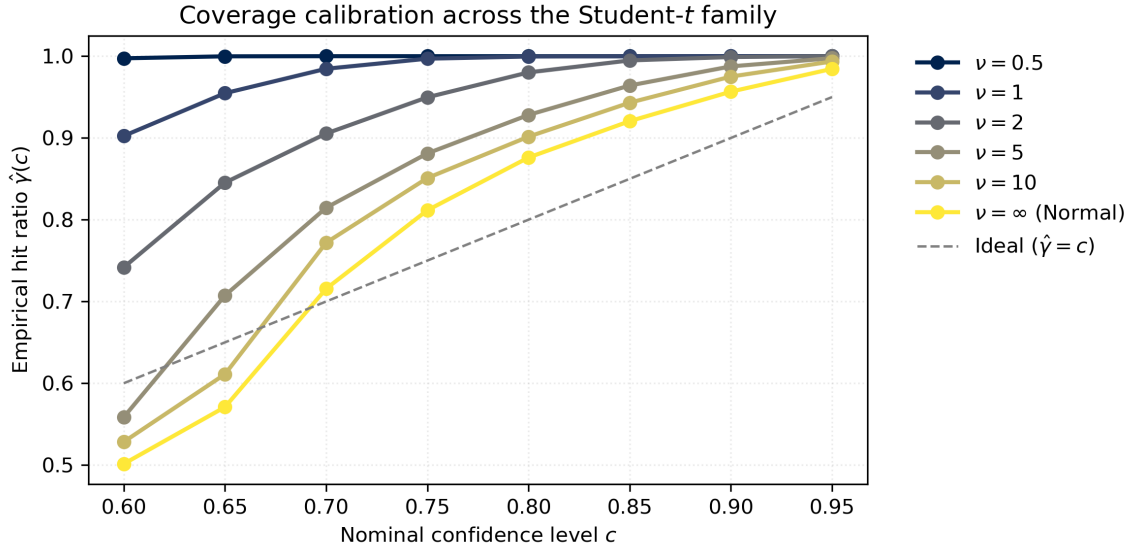


Figure 4.4: Empirical coverage $\hat{\gamma}(c)$ for the Student- t residual family at selected degrees of freedom $\nu \in \{0.5, 1, 2, 5, 10, \infty\}$. The dashed line denotes perfect calibration ($\hat{\gamma} = c$). Heavy-tailed models ($\nu \leq 2$) are uniformly conservative, while the Normal limit ($\nu = \infty$) still over-covers for $c \gtrsim 0.70$, indicating that all t models in this range produce intervals that are wider than empirically required.

The variance is

$$\text{Var}(X) = \alpha^2 \frac{\Gamma(3/\beta)}{\Gamma(1/\beta)},$$

so moment matching with an empirical variance v_{jl} sets

$$\alpha_{jl} = \sqrt{v_{jl}} \left[\frac{\Gamma(1/\beta)}{\Gamma(3/\beta)} \right]^{1/2}, \quad \mu_{jl} = \hat{\sigma}_{jl}.$$

Two-sided 100c% prediction bands follow the same quantile rule as before:

$$I_{jl}^{(\beta,c)} = \left[\mu_{jl} + \alpha_{jl} q_{\text{GG}}^-(c, \beta), \mu_{jl} + \alpha_{jl} q_{\text{GG}}^+(c, \beta) \right], \quad (4.9)$$

where $q_{\text{GG}}^\pm$ are the lower and upper $(1-c)/2$ quantiles of the unit-scale generalized Gaussian with shape β .

Figure 4.5 shows that very small β values ($\beta \leq 0.5$) are severely under-calibrated, a consequence of their sharply peaked cores. As β increases the intervals widen; $\beta \approx 5$ –10 essentially eliminates the systematic over-coverage observed with the Normal and Student- t families for $c \gtrsim 0.75$ while remaining slightly conservative near the extremes. Beyond $\beta \approx 10$ the curves are overly conservative for all confidence levels.

A single value of the generalized Gaussian shape parameter β cannot simultaneously achieve nominal coverage across the full confidence spectrum (Figure 4.5). However, inspection of the same figure shows that for every $c \in (0.60, 0.95)$ there exists at least one β for which the intervals are correctly calibrated. We therefore invert the mapping $c \mapsto \hat{\gamma}(c; \beta)$ on the training ensemble: for each confidence level we solve for the smallest $\hat{\beta}(c)$ such that $\hat{\gamma}(c; \hat{\beta}) \geq c$. Linear interpolation on a dense β -grid yields a smooth, monotone function $c \mapsto \hat{\beta}(c)$ (Figure 4.6).

Using this confidence-specific shape parameter, prediction bands are built exactly as in Eq. (4.4) but with $\beta = \hat{\beta}(c)$ chosen per confidence level. Figure 4.7 compares the resulting empirical coverage on the 200 training ladders and on the 50 unseen test ladders. By construction the training curve lies almost perfectly on the ideal

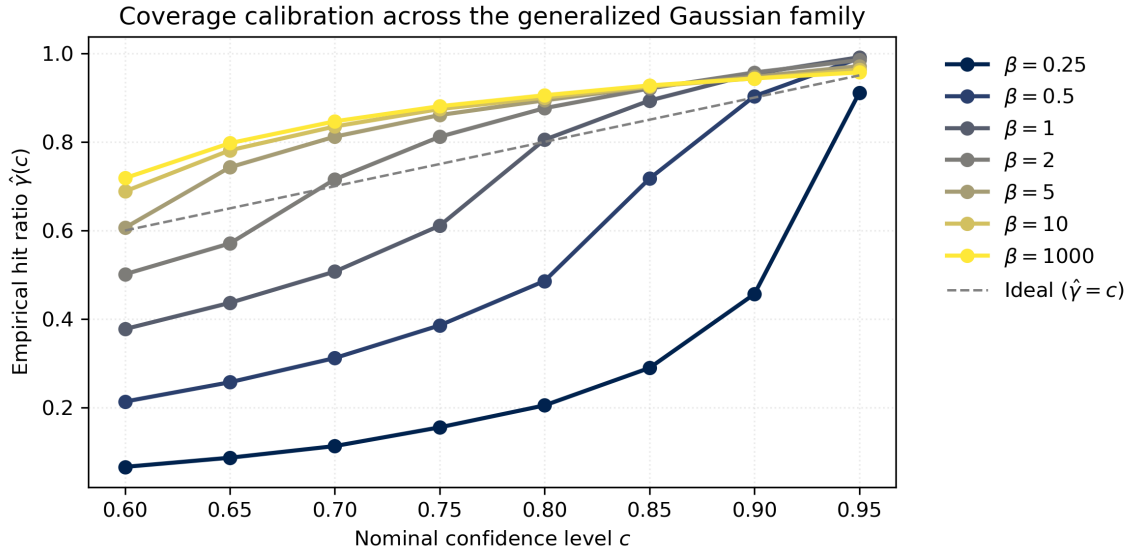


Figure 4.5: Empirical coverage $\hat{\gamma}(c)$ for the generalized Gaussian residual model at shape parameters $\beta \in \{0.25, 0.5, 1, 2, 5, 10, 1000\}$. The dashed line marks perfect calibration ($\hat{\gamma} = c$). Small values $\beta \leq 1$ yield under-coverage (intervals too narrow), while large values yield over-coverage. $\beta = 2$ corresponding to the Normal model is not strictly outperformed by any other $\beta \neq 2$.

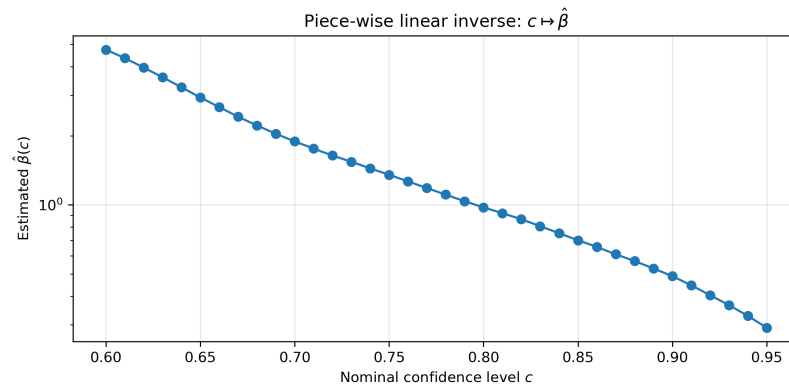


Figure 4.6: Piece-wise linear inverse mapping $c \mapsto \hat{\beta}(c)$ constructed on the training data. A logarithmic y -axis reveals that lower confidence levels require markedly heavier tails (small $\hat{\beta}$) than high confidence levels.

$y = x$ line; encouragingly, the test curve also remains within a few percentage points of the diagonal over the entire range $0.60 \leq c \leq 0.95$. The modest degradation in the 0.70–0.80 region is attributable to sampling noise and the fact that the inverse mapping was not re-estimated on the test data. Although the piece-wise linear scheme is admittedly ad-hoc—it treats β as a deterministic function of the requested confidence rather than a property of the residual distribution. It demonstrates that peakedness is a tunable lever for aligning empirical and nominal coverage.

Ideally, however, the residual model would be confidence-invariant. A shape parameter that changes with c may be difficult to propagate in forward-sampling codes, and there is no guarantee that the inverse mapping calibrated on empirical variances will carry over to derived quantities such as mapped variance elements (as we will see in the forthcoming analysis). Future work therefore seeks a single, globally calibrated distribution—or a hierarchical formulation whose hyper-parameters can be sampled alongside the physics model—to avoid this confidence-specific patch.

4.3 Feature construction

The ultimate goal of the mapping step is to learn, at each evaluation energy $E_j \in \mathcal{E}$, a deterministic function

$$\hat{v}_{jl} = g(\mathbf{x}_{jl}), \quad j = 1, \dots, |\mathcal{E}|, \quad l = 1, \dots, L,$$

that transforms observable or easily computed quantities $\mathbf{x}_{jl}$ into a prediction of the true variance v_{jl} . The choice of explanatory variables reflects two kinds of information embedded in the AutoFit sequence of higher-cardinality fits that precede the selection of the final model cardinality k_{final} . Some of this information is explicit: the GLS covariance matrices that SAMMY outputs at each cardinality. Other information is implicit, appearing in the way the fitted cross section changes as resonances are

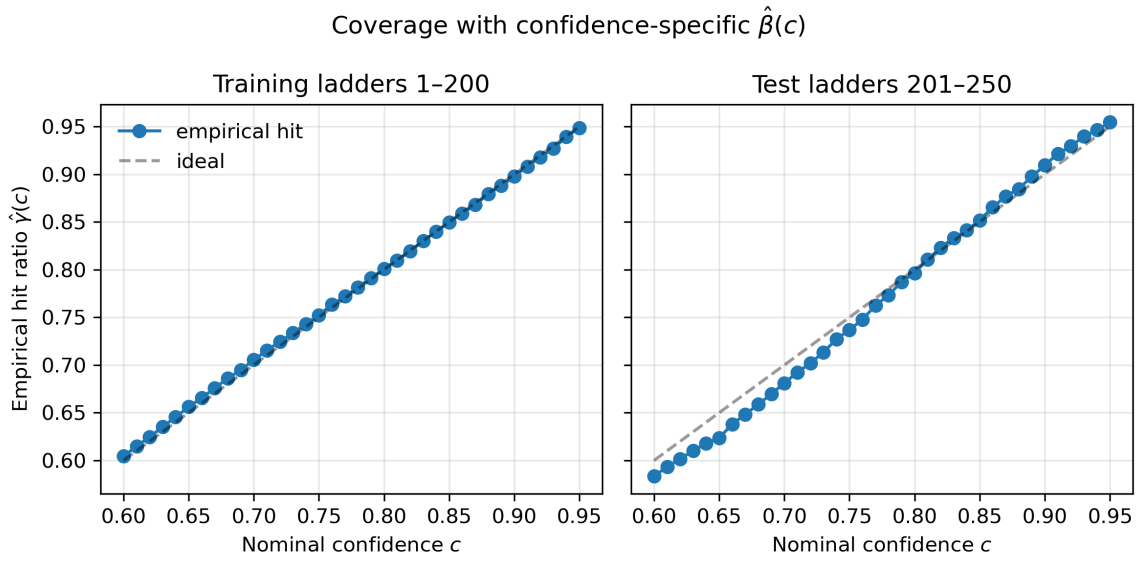


Figure 4.7: Coverage curves obtained by matching the generalized Gaussian shape parameter to the nominal confidence, $\beta = \hat{\beta}(c)$ from Figure 4.6. The mapping is learned on the first 200 synthetic ladders (left) and then applied unchanged to the 50 test ladders (right). The dashed line indicates perfect calibration.

successively removed. Our feature matrix incorporates both types through four distinct blocks, which we formalize below beginning with the GLS covariance terms.

4.3.1 Block 1: GLS covariance diagonals

For a given ladder l AutoFit delivers a monotone sequence of six model cardinalities,

$$\kappa_5 = k_{\text{final}} + 5, \kappa_4 = k_{\text{final}} + 4, \dots, \kappa_0 = k_{\text{final}},$$

each accompanied by a full 51×51 covariance matrix $C^{(l, \kappa_r)}$ computed inside SAMMY. Let

$$\mathbf{v}^{(l, \kappa_r)} = \text{diag}(C^{(l, \kappa_r)}) \in \mathbb{R}^{51}, \quad r = 0, \dots, 5, \quad (4.10)$$

denote the main diagonal of that matrix, evaluated on the fixed grid $\mathcal{E}$. Component j of this vector, $v_{jl}^{(r)} := [\mathbf{v}^{(l, \kappa_r)}]_j$, is the formal GLS variance (in barns²) assigned by SAMMY to the fitted cross-section value at energy E_j when the model contains κ_r resonances.

The first feature block simply concatenates these six variance vectors:

$$\mathbf{x}_{jl}^{\text{GLS}} = (v_{jl}^{(5)}, v_{jl}^{(4)}, v_{jl}^{(3)}, v_{jl}^{(2)}, v_{jl}^{(1)}, v_{jl}^{(0)})^\top \in \mathbb{R}^6. \quad (4.11)$$

Stacking over ladders and energies produces a $M \times 6$ matrix, $M = |\mathcal{E}| \times L$, whose columns capture how SAMMY's internal uncertainty estimate evolves as the resonance count is pruned.

Two observations motivate the remaining blocks. First, the GLS based covariance estimates inherit all of the idealized assumptions built into the Reich-Moore formalism; empirically they undershoot the true variance whenever those assumptions are violated. Second, the pattern of how $v_{jl}^{(r)}$ changes with r is itself correlated with the squared bias incurred by omitting a resonance, suggesting that higher-order functions of the fitted cross section may provide additional explanatory power.

4.3.2 Block 2: squared bias between cardinalities

An implicit source of information about the evaluated cross-section uncertainty is contained in the difference of fitted cross-sections between model cardinalities. We use these differences as regressors. For each ladder l and energy index j let

$$\hat{\sigma}_{jl}^{(r)} \quad r = 5, 4, \dots, 0$$

denote the sequence of fitted cross sections produced by AutoFit as the cardinality is stepped from κ_5 down to $\kappa_0 = k_{\text{final}}$. Define the pointwise bias incurred by pruning one extra resonance,

$$b_{jl}^{(r)} = \hat{\sigma}_{jl}^{(r)} - \hat{\sigma}_{jl}^{(0)}, \quad r = 5, 4, \dots, 1, \quad (4.12)$$

and collect its square $(b_{jl}^{(r)})^2$ into a five-component feature block,

$$\mathbf{x}_{jl}^{\text{bias}} = (b_{jl}^{(5)2}, b_{jl}^{(4)2}, b_{jl}^{(3)2}, b_{jl}^{(2)2}, b_{jl}^{(1)2})^\top \in \mathbb{R}^5. \quad (4.13)$$

The square, rather than the raw difference, is used because the target quantity is a variance. Additionally, notice that only differences with respect to the final model appear, so the block length is five instead of six.

4.3.3 Block 3: squared fitted cross section

Figure 4.8 makes plain that fitted cross sections encode information about their own uncertainty: once $\hat{\sigma}$ exceeds a modest threshold, the empirical standard deviation $s_{jl} = \sqrt{v_{jl}}$ grows monotonically with $\hat{\sigma}_{jl}$. To make this information available for the machine learning mapping we include, for every pruning step,

$$\mathbf{x}_{jl}^{\sigma^2} = (\hat{\sigma}_{jl}^{(5)2}, \hat{\sigma}_{jl}^{(4)2}, \hat{\sigma}_{jl}^{(3)2}, \hat{\sigma}_{jl}^{(2)2}, \hat{\sigma}_{jl}^{(1)2}, \hat{\sigma}_{jl}^{(0)2})^\top. \quad (4.14)$$

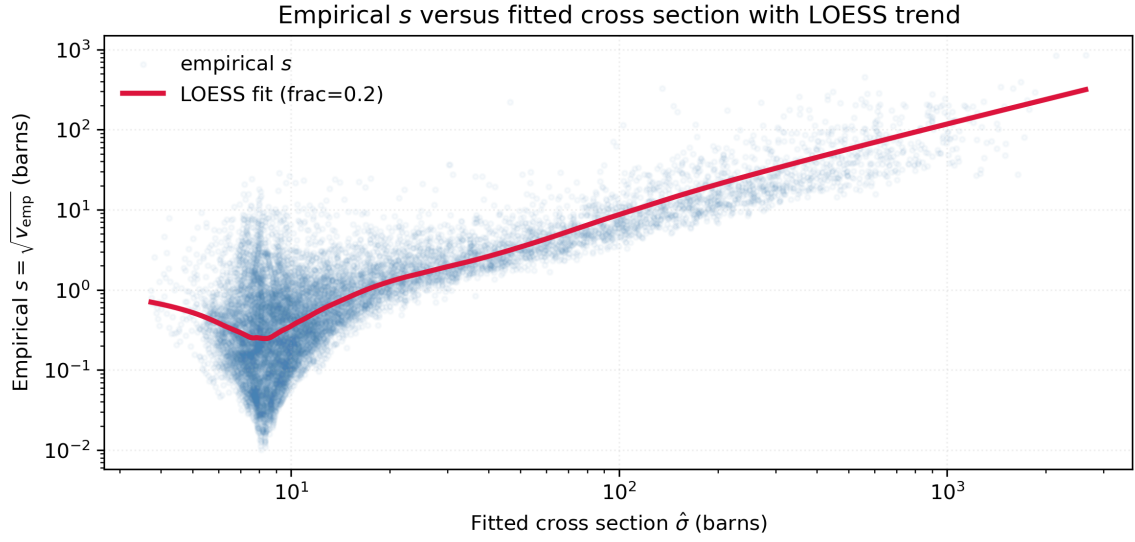


Figure 4.8: Scatter of empirical standard deviation $s = \sqrt{v_{\text{emp}}}$ against fitted cross section $\hat{\sigma}$ for the 200 training ladders (log-log axes) over all energies. The red curve is a LOESS smoother (span 0.2). Beyond the small- $\hat{\sigma}$ local minimum, larger cross sections systematically predict larger sample-to-sample fluctuations, motivating both the fitted cross-section precitor described in Block 3 and the non-parametric variance estimate in Block 4.

Notice we square the fitted cross-sections to match the units of the regressors with their targets.

4.3.4 Block 4: LOESS-predicted variance

Having noted a relationship between fitted-cross section and empirical uncertainty, we can use a basic locally weighted regression (LOESS) to characterize the relationship and use the corresponding, crude predictions as regressors. We train a locally weighted regression on the training ladders,

$$\hat{s} : \hat{\sigma} \mapsto \hat{s}(\hat{\sigma}) \approx s_{jl},$$

using a tricube kernel and span 0.2. The smoother, plotted as the red curve in Figure 4.8, is then evaluated at the six fitted cross-section values produced by AutoFit:

$$\hat{v}_{jl,\text{loess}}^{(r)} = [\hat{s}(\hat{\sigma}_{jl}^{(r)})]^2, \quad r = 5, 4, \dots, 0. \quad (4.15)$$

Squaring restores the units of variance, and collecting the results produces

$$\mathbf{x}_{jl}^{\text{loess}} = (\hat{v}_{jl,\text{loess}}^{(5)}, \hat{v}_{jl,\text{loess}}^{(4)}, \hat{v}_{jl,\text{loess}}^{(3)}, \hat{v}_{jl,\text{loess}}^{(2)}, \hat{v}_{jl,\text{loess}}^{(1)}, \hat{v}_{jl,\text{loess}}^{(0)})^\top. \quad (4.16)$$

Notice the LOESS model is trained solely on the 200 training ladders. As such, it can be treated as a crude machine learning model against which the other models can be compared.

4.3.5 Intercept term and design-matrix dimensions

A column vector of ones,

$$\mathbf{x}_{jl}^{\text{const}} = 1,$$

is appended to permit an explicit intercept in methods that do not add one internally.

Putting Eqs. (4.11), (4.13), (4.14), (4.16), and the constant together yields, for each (j, l) pair, a twenty-four-component feature vector

$$\mathbf{x}_{jl} = (\mathbf{x}_{jl}^{\text{GLS}}, \mathbf{x}_{jl}^{\text{bias}}, \mathbf{x}_{jl}^{\sigma^2}, \mathbf{x}_{jl}^{\text{loess}}, 1)^\top \in \mathbb{R}^{24}.$$

Stacked over grids and ladders the training design matrix has $10\,200 \times 24$ rows and columns, while the test matrix has $2\,550 \times 24$. All regressors are on the natural scale of squared barns, so no further normalization is applied before model fitting.

The next section turns to the learning algorithms themselves and documents how each estimator converts the twenty-four predictors into a variance prediction at every energy point.

4.4 Machine-learning corrections

This section introduces each learning model evaluated in the present study. For every model the discussion sketches the general technique and how it is applied to the covariance-correction task, without attempting a full tutorial. Readers seeking comprehensive background and formal derivations are directed to Hastie, Tibshirani, and Friedman’s classic text *The Elements of Statistical Learning* [44].

4.4.1 Ordinary least squares (OLS)

The starting point is the simplest possible mapping: an ordinary linear regression from the feature vector $\mathbf{x}_{\ell j} \in \mathbb{R}^{24}$ to the empirical variance $y_{\ell j} = C_E^{(\ell)}(E_j, E_j)$, where the indices $\ell = 1, \dots, 200$ enumerate ladders in the training set and $j = 1, \dots, 51$ enumerate energies on the covariance grid. Concatenating ladders and energies produces the design matrix $X \in \mathbb{R}^{10200 \times 24}$ and the target vector $\mathbf{y} \in \mathbb{R}^{10200}$. The ordinary-least-squares solution

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = (X^\top X)^{-1} X^\top \mathbf{y}$$

minimises the mean squared error between the linear prediction $X\boldsymbol{\beta}$ and the empirical variances. Although OLS makes no attempt to temper over-fitting, it provides a useful baseline: any more elaborate model must demonstrate an improvement in coverage on the held-out test ladders if its extra complexity is to be justified.

Once $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ is fixed, the diagonal variance predicted for a new ladder is obtained by evaluating the same 24 regressors on that ladder, multiplying by $\hat{\boldsymbol{\beta}}_{\text{OLS}}$, and enforcing non-negativity. Confidence bands at the desired credibility level are then formed in the same way as for SAMMY: plus or minus the nominal number of standard deviations, but using the OLS-predicted variance in place of the GLS value.

4.4.2 Regularized linear models: Ridge, Lasso, and Elastic Net

A linear regression can be stabilized by adding a penalty that discourages large coefficients. Such regularisation reduces variance at the cost of introducing a small bias and often yields better predictive performance when regressors are noisy or partially collinear, as is the case here.

Next we describe Ridge regression. Ridge augments the ordinary least-squares objective with an ℓ_2 -norm penalty,

$$\hat{\boldsymbol{\beta}}_{\text{ridge}} = \arg \min_{\boldsymbol{\beta}} \|X\boldsymbol{\beta} - \mathbf{y}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2,$$

where $\lambda > 0$ is a hyper-parameter chosen by cross-validation. Closed-form solution remains available and every coefficient is shrunk smoothly toward zero but none is forced exactly to zero.

Next we describe Lasso regression. Replacing the quadratic penalty with an ℓ_1 term gives the lasso,

$$\hat{\boldsymbol{\beta}}_{\text{lasso}} = \arg \min_{\boldsymbol{\beta}} \|X\boldsymbol{\beta} - \mathbf{y}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1,$$

which not only shrinks coefficients but can set some of them exactly to zero, thereby performing implicit feature selection.

Next we describe Elastic-Nets. The elastic-net combines both penalties,

$$\hat{\boldsymbol{\beta}}_{\text{EN}} = \arg \min_{\boldsymbol{\beta}} \|X\boldsymbol{\beta} - \mathbf{y}\|_2^2 + \lambda \left[\frac{1}{2}(1 - \alpha) \|\boldsymbol{\beta}\|_2^2 + \alpha \|\boldsymbol{\beta}\|_1 \right],$$

interpolating between ridge ($\alpha = 0$) and lasso ($\alpha = 1$). In all experiments a grid search over λ and α is nested inside the same five-fold cross-validation used for ridge and lasso.

These three regularized linear models share the interpretability of OLS while offering a controlled trade-off between variance reduction and bias.

4.4.3 Dimension-reduced linear models: Principal Components and Partial Least Squares

The twelve regressors extracted from each ladder are not guaranteed to be linearly independent; several carry nearly the same information (e.g. the SAMMY variance for two adjacent cardinalities). Principal component regression (PCR) and partial least squares (PLS) address this multicollinearity by projecting the original feature space onto a smaller set of orthogonal directions before fitting the response.

Next we describe PCR. PCR begins with an uncentered singular-value decomposition, $X = U\Sigma V^T$, and retains only the first $q < 24$ columns of V . The design matrix is replaced by its rank- q projection, $X_q = XV_{(:,1:q)}$, and an ordinary least-squares fit is performed in that reduced space. Choosing q trades bias for variance: a small q discards noise but risks underfitting, whereas a large q approaches the OLS solution. Five-fold cross-validation on the training ladders selects q_{opt} , and the resulting coefficients are mapped back to the full 24-dimensional space for interpretability.

Next we describe PLS. PLS also constructs orthogonal directions but, unlike PCR, each new direction is chosen to maximize the covariance between the transformed features and the response vector $\mathbf{y}$. The algorithm produces a sequence of latent variables, $T_1, T_2, \dots, T_q$, along with corresponding weight vectors. Fitting proceeds by regressing $\mathbf{y}$ on the matrix $T = [T_1 \dots T_q]$. The number of latent variables is again selected by five-fold cross-validation. In practice PLS often achieves the same predictive accuracy as PCR with fewer components because it uses the response information while constructing the projection.

4.4.4 Gradient boosting regression

Gradient boosting builds a non-linear function by adding shallow decision trees one at a time, each new tree trained to predict the residual left by the ensemble so far. At iteration m the current model is $f_{m-1}(X)$; a regression tree $h_m(X)$ is fitted to the negative gradient of the loss (for squared-error loss this is simply the current residual) and the ensemble is updated,

$$f_m(X) = f_{m-1}(X) + \eta h_m(X),$$

where the learning rate $0 < \eta \leq 1$ scales the contribution of the new tree. Repeating for M iterations yields a flexible predictor that can approximate complex, non-linear response surfaces while controlling overfitting through the choice of M , η , and the depth of each. We determine these hyperparameters with cross-validation.

4.4.5 Support vector regression

Support vector regression (SVR) fits a function that deviates from each training target by no more than a chosen margin ε while remaining as flat as possible. In its linear

form the goal is to find coefficients $\mathbf{w}$ and intercept b that minimize

$$\frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*),$$

subject to $-\varepsilon - \xi_i \leq \mathbf{w}^\top \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^*$, with slack variables $\xi_i, \xi_i^* \geq 0$. The trade-off between model flatness and margin violations is controlled by the penalty parameter C . Introducing a kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ allows the same optimization to be carried out in an implicit high-dimensional feature space, giving the non-linear SVR predictor $f(\mathbf{x}) = \sum_{i=1}^N \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b$.

A radial basis function kernel $K(\mathbf{x}, \mathbf{z}) = \exp(-\gamma\|\mathbf{x} - \mathbf{z}\|^2)$ is used. The hyperparameters (C, ε, γ) are selected by five-fold cross-validation on the training ladders.

4.4.6 Random-forest regression

A random forest constructs an ensemble of regression trees, $\{T^{(m)}\}_{m=1}^M$, each one grown on a different bootstrap replica of the training set. At every split in every tree, the candidate split variable is chosen from a randomly selected subset of the full predictor list. These two sources of randomness—row sampling and column sampling—decorrelate the individual trees so that their average prediction achieves low variance without the high bias that would result from aggressive pruning.

For any feature vector $\mathbf{x}$ the ensemble prediction is the simple mean

$$\hat{y}_{\text{RF}}(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^M \hat{y}^{(m)}(\mathbf{x}), \quad (4.17)$$

where $\hat{y}^{(m)}(\mathbf{x})$ is the piecewise-constant output of tree $T^{(m)}$. Because each tree is a non-parametric partition of the predictor space, the forest as a whole can approximate highly non-linear response surfaces and capture interactions among predictors without explicitly including interaction terms.

4.4.7 Gaussian-mixture correction

This method leverages the information provided by AutoFit for all tested cardinalities $k = 0, 1, \dots, k_{\text{final}} + 5$. Each candidate model provides a mean curve $\mu_k(E_j)$ —the fitted total cross section—and an associated diagonal covariance $c_k(E_j)$ obtained by propagating the posterior resonance-parameter variance through SAMMY. The mixture approach treats these $K := k_{\text{final}} + 6$ pairs $\{\mu_k, c_k\}_{k=0}^{K-1}$ as the components of a discrete Gaussian ensemble and then averages them with weights based on the quality of fit of each model.

For every ladder the five-fold cross-validation routine reports a χ^2 statistic χ_k^2 for each cardinality. That statistic measures, on the natural scale of residual sum of squares, how well model k predicts spectra that were not used during its fit. A smaller χ_k^2 therefore implies higher plausibility. We convert the scores to positive weights with a Gaussian kernel,

$$w_k = \exp[-(\chi_k^2 - \chi_{\min}^2)^2/2], \quad (4.18)$$

where $\chi_{\min}^2$ is the best (smallest) score observed on that ladder.

At every grid energy E_j the predictive distribution is taken to be a one-dimensional Gaussian mixture,

$$p(\sigma_{\text{tot}} | E_j) = \sum_{k=0}^{K-1} w_k \mathcal{N}(\mu_k(E_j), c_k(E_j)).$$

We leverage Monte-Carlo sampling for subsequent results analysis for this method. $N = 10\,000$ draws are taken from the mixture at each energy and the desired quantiles are read directly from the sample. The q -credibility band is thus the interval between the empirical $\frac{1}{2}(1 - q)$ and $\frac{1}{2}(1 + q)$ quantiles of the sampled values.

4.5 Results: Machine-Learned Variances

Section 4.2 introduced the coverage indicator $\delta_{jl}^{(D,c)}$ for a residual family D and confidence level c . That definition is repeated here with the empirical variance v_{jl} replaced by a machine-learning prediction $\hat{v}_{jl}$:

$$\tilde{\sigma}_{jl} \mid \hat{\sigma}_{jl}, \hat{v}_{jl}, D \sim D(\mu = \hat{\sigma}_{jl}, \varsigma^2 = \hat{v}_{jl}), \quad (4.19)$$

$$I_{jl}^{(D,c)} = [F_D^{-1}(\tfrac{1}{2}(1-c); \hat{\sigma}_{jl}, \hat{v}_{jl}), F_D^{-1}(\tfrac{1}{2}(1+c); \hat{\sigma}_{jl}, \hat{v}_{jl})], \quad (4.20)$$

$$\delta_{jl}^{(D,c)} = \mathbf{1}\{\sigma_{jl}^{\text{true}} \in I_{jl}^{(D,c)}\}. \quad (4.21)$$

Averaging over every ladder and every grid energy gives the empirical hit ratio

$$\hat{\gamma}(c; D, g) = \frac{1}{L|\mathcal{E}|} \sum_{l=1}^L \sum_{j=1}^{|\mathcal{E}|} \delta_{jl}^{(D,c)}, \quad (4.22)$$

where g denotes the variance-mapping model that produced $\hat{v}_{jl}$. Coverage curves in Figure 4.9 plot $c \mapsto \hat{\gamma}(c; D, g)$ on the grid $c \in \{0.60, 0.65, \dots, 0.95\}$ for four residual families and the eight most informative variance maps.

The pink SAMMY curve confirms the deficiency already documented in Chapter 3: ECSCMs propagated directly from the RPCM are systematically precise. No adjustment of the residual probability model, not even by using a conservative model like the Cauchy distribution, restores calibration because the reported variance in energies corresponding to missed resonances are orders of magnitude too small.

OLS and the feature reduction linear variants suffer from the opposite problem. Namely, they are overly conservative. Shrinking coefficients (Ridge), removing coefficients (Lasso), or rotating the predictor space (PCR, PLS) changes the slope of the curve only marginally. For the sake of keeping the plot interpretable, only ElasticNet was included as a representative for the family of feature reduction methodologies.

Coverage curves — ML-mapped variance across four residual families

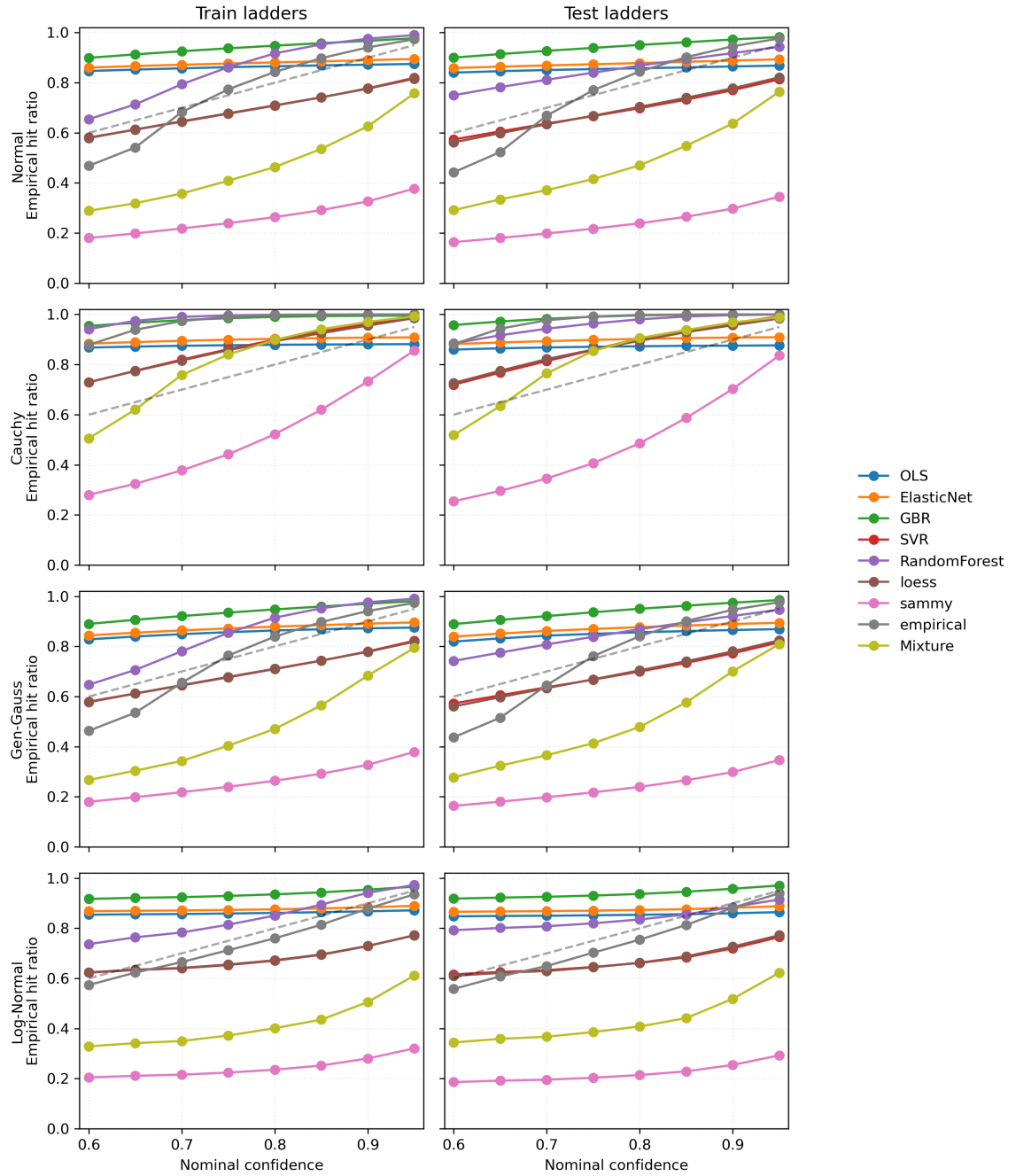


Figure 4.9: Empirical coverage of 95 % log-normal bands for four residual families.

Interestingly, LOESS and SVR produce nearly identical curves. It’s not entirely clear why this is, as the LOESS model is based exclusively on fitted cross-sections and SVR has all regressors available to it. Regardless, as a pair they indicate that computationally expensive models like SVR can be matched or even outperformed by relatively simple ones.

Random Forest shows promise in its coverage performance. It’s coverage curve generally tracks the coverage = confidence reference line while remaining consistently on the conservative side, which is preferred. The ensemble of trees partitions the predictor space finely enough to recognise at which energies additional variance should be predicted while keeping variance inflation moderate between resonances. For use cases that prioritize coverage more than absolute interval width, particularly coverage of missed resonances, this trade-off is preferable to the under-coverage of methods like SVR and SAMMY and to the excessive over-coverage of the linear group.

Figure 4.10 zooms in on a single ladder in the test set. Two weak resonances were missed from the AutoFit final fitted model, one at approximately 203 eV and a second at approximately 215 eV. Viewing the performance of a selection of variance prediction methods on a single ladder allows us to compare the qualitatively. The empirical plot provides a visual description of an accurate uncertainty quantification. As expected, the SAMMY variances are overly precise, especially at the energies near the missing resonances. OLS and Elastic-Net predicted variances are conservative enough to capture the missing resonances, but at the expense of being extremely conservative over the entire energy regions between resonances. SVR does predict increased variance at the energies near the missing resonances, but the predictions are insufficient. Random Forest adequately increases variance for the missing resonances, and it does this without being nearly as conservative as OLS and other linear models over the valleys between resonances.

These observations are consistent with Figure 4.9. Methods that cannot detect the weak resonances miss them and appear under-calibrated; methods that correct everywhere overshoot in the valleys and appear conservative. Random Forest

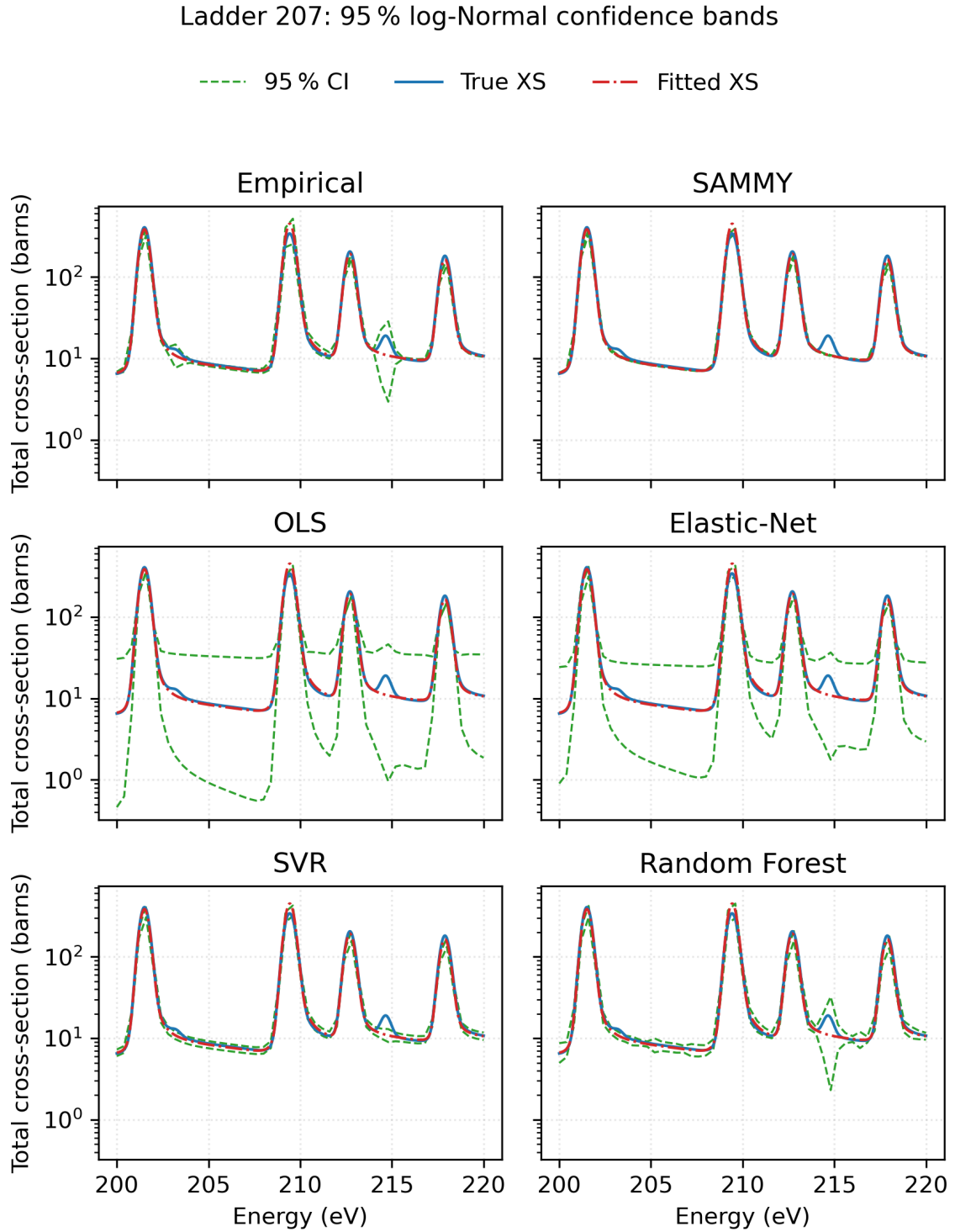


Figure 4.10: Comparison of log-normal confidence bands based on empirical variances, SAMMY variances, and four machine learned variances on a single resonance ladder. These results visually depict how some methods are overly precise (e.g., SAMMY), some overly conservative (e.g., OLS), and some relatively well calibrated.

commends itself as a model that adjusts amplitude where bias is largest without the cost of being overly conservative elsewhere. The qualitative observations we make viewing the results from a single ladder can be summarized quantitatively. In the next section, we propose a method to distinguish variance-prediction methods based on their relative performance over energies near resonances and between resonances.

4.5.1 Coverage as a function of cross-section magnitude

To quantify regional performance we compute, for every test ladder, a kernel-smoothed estimate of the true cross section,

$$\sigma_{jl}^{\text{smooth}} = \frac{\sum_{j'} K_h(E_j - E_{j'}) \sigma_{jl}^{\text{true}}}{\sum_{j'} K_h(E_j - E_{j'})}, \quad K_h(\Delta) = \exp[-\tfrac{1}{2}(\Delta/h)^2], \quad (4.23)$$

with bandwidth $h = 0.6$ eV. The indicator $\delta_{jl}^{(D,0.95)}$ from Eq. (4.22) is then plotted against $\sigma_{jl}^{\text{smooth}}$ for all 51×50 energy points. Fitting a logistic regression,

$$\Pr\{\delta = 1 \mid x\} = \frac{1}{1 + \exp[-(\alpha + \beta x)]}, \quad (4.24)$$

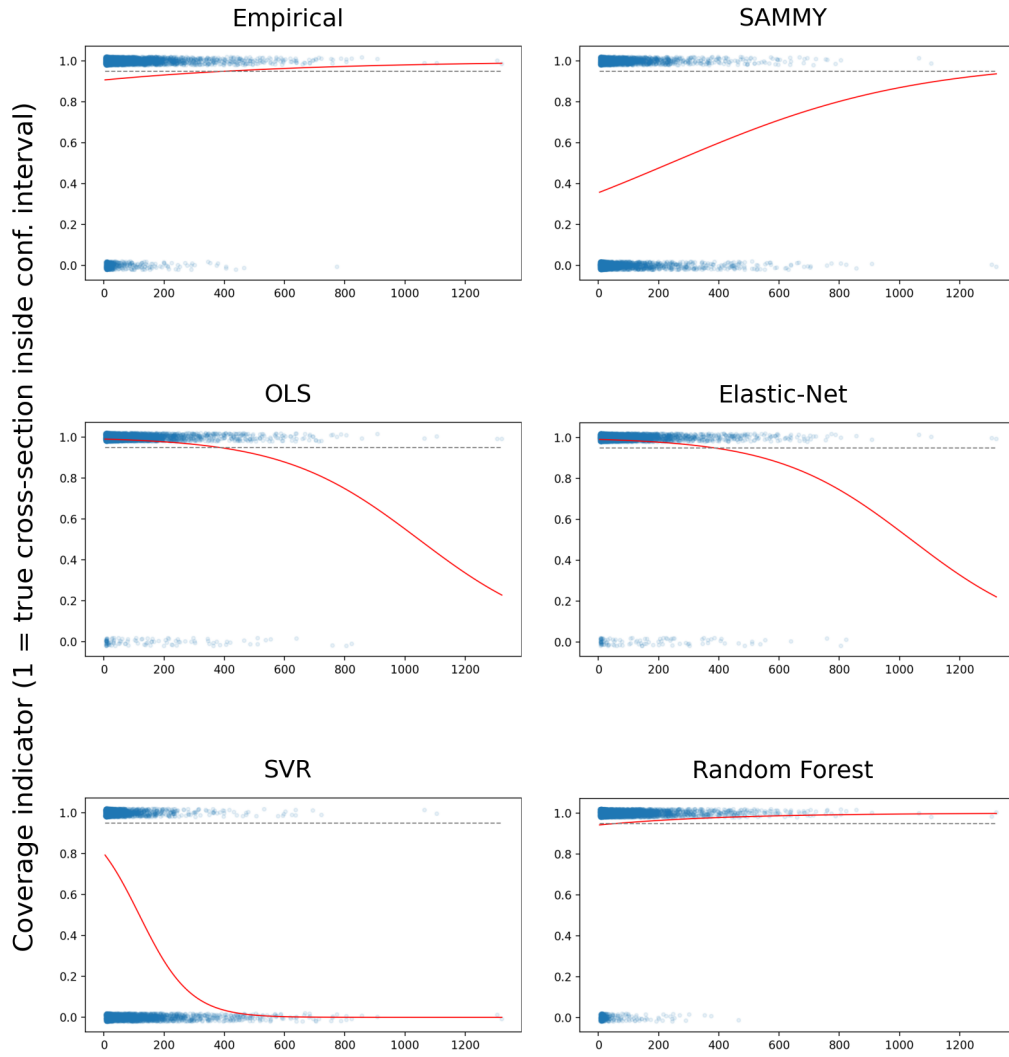
yields the red curves in Figure 4.11. Here x is the smoothed cross section and (α, β) are maximum-likelihood estimates.

By smoothing the cross-section, we have a metric that signals at which energies we are near large resonances or multiple small resonances, near a small (and perhaps missable) resonance, and when we are firmly between resonances. As such, the logistic curve is a way of visualizing and quantifying the performance of the various variance-prediction methods over regions near to and far from resonances.

We would expect the empirical baseline logistic curve to be relatively flat, because empirical variances are computed around the true-cross section and should therefore capture the true cross-section over the appropriate proportion of energies independent of proximity to resonances. The results match this expectation, with empirical baseline nearly equal to the horizontal reference line. SAMMY, on the other hand,

Logistic fit of 95 % coverage vs. true cross-section for all test ladders

● Unique energy+ladder hits/misses — Logistic fit - - - Target coverage 0.95



Smoothed true cross-section (barns)

Figure 4.11: Logistic curves showing the relationship between coverage performance and fitted cross-section of log-normal curves for various methods. The shape and location of the logistic curves indicate the accuracy of variance predictions over energy regions far from resonances, near small resonances, and near large resonances.

produces an increasing logistic curve. This can be explained by at least two factors. As previously discussed at length, SAMMY is overly precise over regions of small, missed resonances. As such, SAMMY’s coverage performance over relatively small to medium smoothed cross-sections is unsurprisingly biased low. Additionally, even over energies between resonances where the bias between fitted cross-section and true cross-section are relatively small, SAMMY bands are so precise that the confidence bands can run in parallel entirely above or below the true cross-section for large stretches of energy. OLS and Elastic Net start above the nominal line, signalling over-conservatism in valleys, but fall toward the nominal line for larger cross-sections, indicating that their excessive variance is most pronounced in energy regions between resonances or near smaller resonances. The SVR logistic curve descends even more steeply. This is consistence with its reluctance to expand its variance sufficiently at small resonances.

The Random Forest logistic curve, like the empirical logistic curve, is relatively flat. It sits slightly above the nominal confidence line, indicating that it is conservative. The absence of a strong slope to the logistic curve means that it performs adequately both in resonance valleys and near resonances. These results indicate that the Random Forest predicted variances will be moderately conservative across all energies.

4.5.2 Recommendation and outlook

For applications where only per-energy variances are needed and energy–energy correlations can be neglected, the combination of Random Forest variance mapping and a log-normal residual family offers the best compromise uncovered in this study. Random Forest coverage tracks the nominal coverage well relative to competing methods, and the model remains conservative independent of resonance proximity. Linear shrinkage models are usable when computational simplicity is paramount, but the user must be willing to accept highly inflated variances. No residual family tested

here achieves perfect calibration, so the log-normal choice is made for physical rather than statistical reasons: its support is $[0, \infty)$ and an evaluator need not clip negative draws.

The present chapter treated each energy independently. Correlations across energy and across reaction channel were left untouched. The next chapter develops a framework for adding those correlations into the various prediction methodologies, completing the path from AutoFit output to full cross-section covariance predictions.

Chapter 5

Automated Corrections for Energy and Channel Correlations

Chapter 4 developed an automated, reproducible workflow that corrects the SAMMY evaluated cross-section covariance diagonal so that stated confidence levels align with empirical coverage when paired with an appropriate probability model. There, we learned a mapping from AutoFit outputs to per-energy variance $s^2(E)$ and showed that both model-based regressors and a simple mixture construction can produce accurate diagonal uncertainty bands.

The present chapter extends that program from marginal (per-energy) variance to the full evaluated cross-section covariance matrix (ECSCM), including cross-energy and cross-channel correlations. Specifically, for each synthetic ladder we seek to predict the 102×102 covariance across two reactions (elastic and capture) sampled on the same 51-point covariance grid used in Chapter 4. This richer target captures (i) correlation of the cross section across energies within a channel and (ii) correlation between the elastic and capture channels at (possibly different) energies. We use the same ensemble of 250 ladders, the same AutoFit workflow, and the same train/test split as before (ladders 1–200 for training and 201–250 for testing).

5.1 Empirical two-channel ECSCM targets

All covariance estimates in this chapter are evaluated on the fixed 51-point grid

$$\mathcal{E} = \{E_j\}_{j=1}^{51}, \quad E_j = 200 + 0.4(j-1) \text{ eV}, \quad (5.1)$$

which matches Chapter 4. For each ladder we generate N independent transmission samples ($N = 50$ in the results shown), and each sample is processed by the full three-stage AutoFit routine. Evaluating both the fitted and the true cross sections on $\mathcal{E}$ for the two reactions $r \in \{\text{el}, \text{cap}\}$ produces residuals

$$r_r^{(n)}(E_j) = \sigma_r^{\text{true}}(E_j) - \hat{\sigma}_r^{(n)}(E_j), \quad n = 1, \dots, N, \quad j = 1, \dots, 51. \quad (5.2)$$

We assemble, for each sample n , a length-102 residual vector by concatenating elastic (first) and capture (second) components,

$$\mathbf{r}^{(n)} = (r_{\text{el}}^{(n)}(E_1), \dots, r_{\text{el}}^{(n)}(E_{51}), r_{\text{cap}}^{(n)}(E_1), \dots, r_{\text{cap}}^{(n)}(E_{51}))^\top \in \mathbb{R}^{102}. \quad (5.3)$$

Stacking these row-wise yields the residual matrix

$$R = \begin{bmatrix} (\mathbf{r}^{(1)})^\top \\ \vdots \\ (\mathbf{r}^{(N)})^\top \end{bmatrix} \in \mathbb{R}^{N \times 102}. \quad (5.4)$$

Consistent with Chapter 4, the empirical two-channel ECSCM for a ladder is defined by the inner product

$$\hat{C}_E = \frac{1}{N} R^\top R \in \mathbb{R}^{102 \times 102}. \quad (5.5)$$

(For $N = 50$ the difference between $1/N$ and $1/(N-1)$ is immaterial at the level of precision reported here.)

By construction, $\widehat{C}_E$ has a 2×2 block structure with 51×51 submatrices:

$$\widehat{C}_E = \begin{bmatrix} \widehat{C}_{\text{el,el}} & \widehat{C}_{\text{el,cap}} \\ \widehat{C}_{\text{cap,el}} & \widehat{C}_{\text{cap,cap}} \end{bmatrix}, \quad (\widehat{C}_{r,r'})_{ij} = \frac{1}{N} \sum_{n=1}^N r_r^{(n)}(E_i) r_{r'}^{(n)}(E_j). \quad (5.6)$$

The on-diagonal blocks $\widehat{C}_{\text{el,el}}$ and $\widehat{C}_{\text{cap,cap}}$ are the within-channel cross-energy covariances (the ECSCMs for each channel individually), while the off-diagonal blocks $\widehat{C}_{\text{el,cap}}$ and $\widehat{C}_{\text{cap,el}}$ capture cross-channel covariance between elastic and capture. Each ladder therefore contributes a full, symmetric 102×102 matrix that we will use as the supervised-learning target in the remainder of the chapter.

Figure 5.1 (left panel) shows a typical empirical $\widehat{C}_E$ with its two-by-two block pattern; the right panel plots the corresponding true and fitted cross sections for the same ladder and sample on the grid $\mathcal{E}$. As in Chapter 4, heat-map structure reflects both resonant features and model inadequacies; here those patterns also appear in the off-diagonal blocks, indicating energy- and channel-based correlations.

5.2 Feature construction for full covariances

The mapping step in this chapter parallels Chapter 4, but now the target is the full evaluated cross-section covariance matrix (ECSCM) across energies and across two reaction channels (elastic, capture). Let $\mathcal{E} = \{E_j\}_{j=1}^{51}$ be the fixed covariance grid and let $c \in \{\text{el, cap}\}$ index the channel. For each ladder l and each model cardinality in an AutoFit-generated sequence (defined below), SAMMY returns fitted cross sections on $\mathcal{E}$. We stack the two channels to form

$$\widehat{\boldsymbol{\sigma}}^{(l,\kappa_r)} = \begin{bmatrix} \widehat{\boldsymbol{\sigma}}_{\text{el}}^{(l,\kappa_r)} \\ \widehat{\boldsymbol{\sigma}}_{\text{cap}}^{(l,\kappa_r)} \end{bmatrix} \in \mathbb{R}^{102}, \quad \widehat{\boldsymbol{\sigma}}_{\text{el}}^{(l,\kappa_r)}, \widehat{\boldsymbol{\sigma}}_{\text{cap}}^{(l,\kappa_r)} \in \mathbb{R}^{51}.$$

Here $\kappa_r = k_{\text{final}} + r$ with $r = 0, 1, \dots, 5$ denotes the final cardinality selected by AutoFit together with the five higher-cardinality (“overfit”) solutions retained from

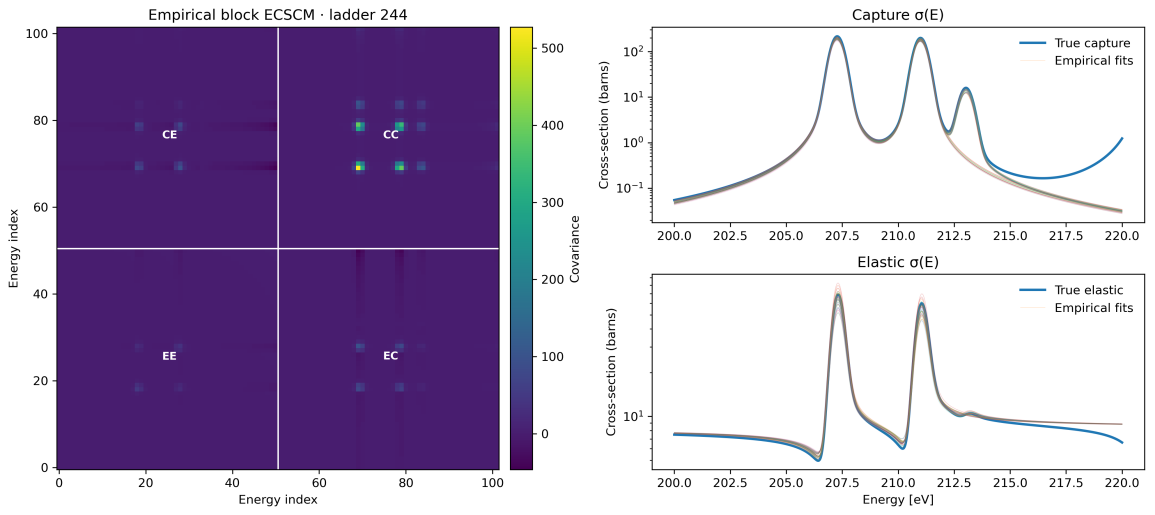


Figure 5.1: Left: empirical two-channel ECSCM $\hat{C}_E = \frac{1}{N}R^\top R$ with elastic (bottom left) and capture (upper right) within-channel blocks and the two cross-channel blocks off the diagonal. Right: corresponding true and fitted cross sections on $\mathcal{E}$.

the backward-elimination history, exactly as in Section 4.3. The empirical target for ladder l is the 102×102 covariance

$$C_{\text{emp}}^{(l)} = \frac{1}{N} R_l^\top R_l,$$

where R_l is the $N \times 102$ residual matrix whose columns are the elastic and capture residuals evaluated on $\mathcal{E}$ (Section 4.1). This matrix has the natural 2×2 block structure

$$C_{\text{emp}}^{(l)} = \begin{bmatrix} C_{\text{emp,el-el}}^{(l)} & C_{\text{emp,el-cap}}^{(l)} \\ C_{\text{emp,cap-el}}^{(l)} & C_{\text{emp,cap-cap}}^{(l)} \end{bmatrix},$$

with nonzero off-diagonal blocks capturing cross-channel correlations.

To learn a predictor for the full covariance, we build four regressor blocks, each reflecting a different source of information about uncertainty, and append two bookkeeping columns (a constant and a diagonal-offset index). As before, we learn a deterministic map that acts elementwise on the covariance surface:

$$\widehat{C}_{ij}^{(l)} = g(\mathbf{x}_{ij}^{(l)}), \quad 1 \leq i \leq j \leq 102, \quad (5.7)$$

where $\mathbf{x}_{ij}^{(l)}$ collects the regressors associated with the (i, j) element. Since all matrices here are symmetric, we work on the upper triangle ($i \leq j$) and reuse the values for (j, i) .

Following is how we handle vectorization of the full matrices. Let $n = 2|\mathcal{E}| = 102$ and $m = n(n+1)/2 = 5253$. Define the upper-triangle vectorization operator

$$\text{vecu}(A) = (A_{11}, A_{12}, \dots, A_{1n}, A_{22}, A_{23}, \dots, A_{nn})^\top \in \mathbb{R}^m, \quad A \in \mathbb{S}^n, \quad (5.8)$$

which lists $\{A_{ij} : 1 \leq i \leq j \leq n\}$ in lexicographic order. For an upper-triangle index (i, j) we also define the (nonnegative) diagonal offset

$$\xi_{ij} = |j - i|, \quad 0 \leq \xi_{ij} \leq n - 1, \quad (5.9)$$

which counts separation in grid steps. Because the two channels are stacked (elastic entries occupy indices 1 : 51, capture 52 : 102), offsets near 51 correspond to cross-channel pairs at similar energies, while ξ_{ij} near 0 (or small positive values) corresponds to near-diagonal, same-channel pairs.

5.2.1 Block 1: GLS ECSCMs (S)

For each ladder l and cardinality κ_r , we compute channelwise ECSCMs with SAMMY and place them on the diagonal of a 102×102 block matrix:

$$S^{(l, \kappa_r)} = \text{blkdiag}\left(C_{\text{GLS,el}}^{(l, \kappa_r)}, C_{\text{GLS,cap}}^{(l, \kappa_r)}\right) \in \mathbb{S}^{102}. \quad (5.10)$$

By construction the off-diagonal channel blocks of $S^{(l, \kappa_r)}$ are zero. The six vectors $\text{vecu}(S^{(l, \kappa_r)})$, $r = 5, 4, \dots, 0$, form the first regressor block and pass into the map $g(\cdot)$ as distinct columns, allowing the learner to exploit how SAMMY's internal uncertainty evolves with overfitting.

5.2.2 Block 2: Cardinality-bias outer products (B)

Let $\hat{\boldsymbol{\sigma}}^{(l, \kappa_0)}$ denote the final (five-fold CV selected) fit. For each overfit cardinality define the fitted-cross-section bias vector

$$\mathbf{d}^{(l, \kappa_r)} = \hat{\boldsymbol{\sigma}}^{(l, \kappa_0)} - \hat{\boldsymbol{\sigma}}^{(l, \kappa_r)}, \quad r = 5, 4, \dots, 1, \quad (5.11)$$

and form the rank-one outer product

$$B^{(l, \kappa_r)} = \mathbf{d}^{(l, \kappa_r)} \mathbf{d}^{(l, \kappa_r)\top} \in \mathbb{S}^{102}, \quad r = 5, 4, \dots, 1. \quad (5.12)$$

The $r = 0$ case would yield the zero matrix and is omitted. These matrices capture how systematic shifts induced by pruning resonances co-vary across both energy and

channel; large components in $\mathbf{d}^{(l, \kappa_r)}$ translate directly into broad covariance features through (5.12).

5.2.3 Block 3: Outer products of fitted cross sections (F)

The stacked LOESS trends in Fig. 5.2 show, for both channels, that the empirical standard deviation $s = \sqrt{v_{\text{emp}}}$ increases monotonically with the fitted cross section $\hat{\sigma}$ once outside a small- $\hat{\sigma}$ neighborhood. This channelwise scaling motivates a multiplicative template for covariance structure: where the fitted cross section is large, both variances and covariances tend to be large.

To make this information available across energies and channels, we include, for each ladder l and each cardinality κ_r , the rank-one outer product of the fitted cross-section vector evaluated on the 102-point, two-channel grid (elastic followed by capture):

$$F^{(l, \kappa_r)} = \hat{\boldsymbol{\sigma}}^{(l, \kappa_r)} \hat{\boldsymbol{\sigma}}^{(l, \kappa_r)\top} \in \mathbb{S}^{102}, \quad r = 5, 4, \dots, 0. \quad (5.13)$$

If we write $\hat{\boldsymbol{\sigma}}^{(l, \kappa_r)} = [\hat{\boldsymbol{\sigma}}_{(l, \kappa_r)}^{\text{el}}; \hat{\boldsymbol{\sigma}}_{(l, \kappa_r)}^{\text{cap}}]$ with each subvector in $\mathbb{R}^{51}$, then $F^{(l, \kappa_r)}$ has the block structure

$$F^{(l, \kappa_r)} = \begin{bmatrix} \hat{\boldsymbol{\sigma}}^{\text{el}} \hat{\boldsymbol{\sigma}}^{\text{el}\top} & \hat{\boldsymbol{\sigma}}^{\text{el}} \hat{\boldsymbol{\sigma}}^{\text{cap}\top} \\ \hat{\boldsymbol{\sigma}}^{\text{cap}} \hat{\boldsymbol{\sigma}}^{\text{el}\top} & \hat{\boldsymbol{\sigma}}^{\text{cap}} \hat{\boldsymbol{\sigma}}^{\text{cap}\top} \end{bmatrix}, \quad (5.14)$$

so that (i) the two diagonal quadrants provide cross-energy templates within each channel and (ii) the off-diagonal quadrants carry cross-channel templates. Every entry has the correct physical units (barns²), and elements at indices (a, b) scale as $\hat{\sigma}_a \hat{\sigma}_b$, matching the monotone empirical behavior in Fig. 5.2. In the design matrix we use the upper-triangular elements of $F^{(l, \kappa_r)}$ for each r , yielding six F -columns per (l) that the learning algorithm can weight differently across cardinalities.

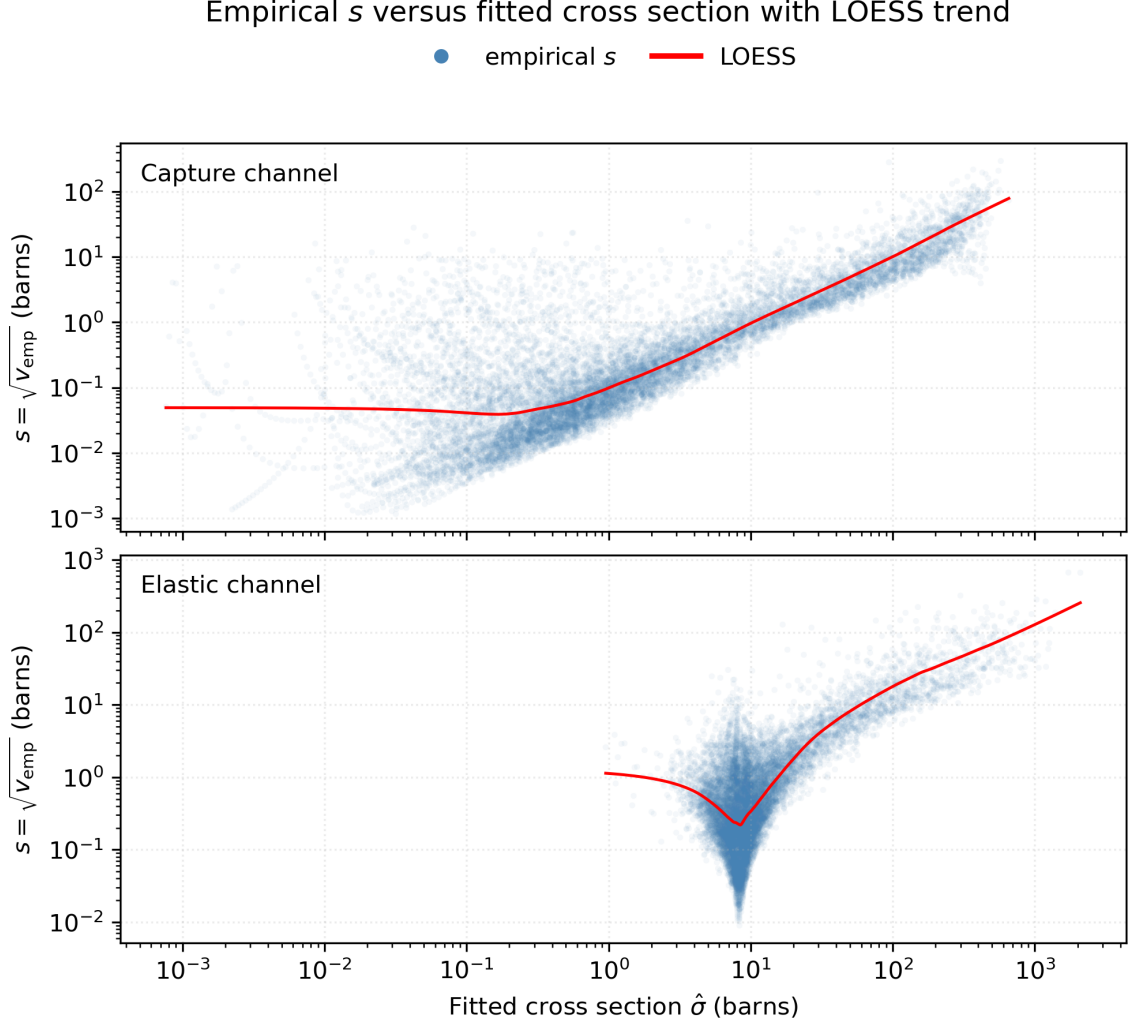


Figure 5.2: Empirical standard deviation $s = \sqrt{v_{\text{emp}}}$ versus fitted cross section $\hat{\sigma}$ on log-log axes, stacked by channel. Top: capture; bottom: elastic. Each point aggregates over energies, ladders, and cardinality levels; the red curve is a LOESS smoother fit separately in each channel. The clear relationship between s and $\hat{\sigma}$ motivates the multiplicative use of $F = \hat{\sigma}\hat{\sigma}^\top$ as a regressor for full-matrix covariance prediction.

5.2.4 Block 4: LOESS-predicted scale (L)

Figure 5.2 shows that, in both channels, the empirical standard deviation $s = \sqrt{v_{\text{emp}}}$ grows monotonically with the fitted cross section $\hat{\sigma}$ once outside a small- $\hat{\sigma}$ neighborhood. To encode this channelwise, nonparametric scale information we fit, on the training ladders only, two LOESS smoothers (one per channel) mapping fitted cross section to standard deviation,

$$\hat{s}_{\text{el}} : \hat{\sigma} \mapsto \hat{s}_{\text{el}}(\hat{\sigma}), \quad \hat{s}_{\text{cap}} : \hat{\sigma} \mapsto \hat{s}_{\text{cap}}(\hat{\sigma}).$$

For ladder l and cardinality κ_r we evaluate the smoothers elementwise on the 51-point fitted curves in each channel and then stack the results on the two-channel grid (elastic followed by capture),

$$\hat{\mathbf{s}}^{(l, \kappa_r)} = \begin{bmatrix} \hat{\mathbf{s}}_{(l, \kappa_r)}^{\text{el}} \\ \hat{\mathbf{s}}_{(l, \kappa_r)}^{\text{cap}} \end{bmatrix} = \begin{bmatrix} \hat{s}_{\text{el}}(\hat{\boldsymbol{\sigma}}_{(l, \kappa_r)}^{\text{el}}) \\ \hat{s}_{\text{cap}}(\hat{\boldsymbol{\sigma}}_{(l, \kappa_r)}^{\text{cap}}) \end{bmatrix} \in \mathbb{R}_{\geq 0}^{102}, \quad r = 5, 4, \dots, 0. \quad (5.15)$$

The LOESS block is then the rank-one outer product

$$L^{(l, \kappa_r)} = \hat{\mathbf{s}}^{(l, \kappa_r)} \hat{\mathbf{s}}^{(l, \kappa_r)\top} \in \mathbb{S}^{102}, \quad (5.16)$$

which supplies a scale-only template in the correct physical units (barns²). Writing $\hat{\mathbf{s}}^{(l, \kappa_r)} = [\hat{\mathbf{s}}_{(l, \kappa_r)}^{\text{el}}; \hat{\mathbf{s}}_{(l, \kappa_r)}^{\text{cap}}]$ with each subvector in $\mathbb{R}_{\geq 0}^{51}$, $L^{(l, \kappa_r)}$ inherits the same 2×2 block structure as $F^{(l, \kappa_r)}$:

$$L^{(l, \kappa_r)} = \begin{bmatrix} \hat{\mathbf{s}}^{\text{el}} \hat{\mathbf{s}}^{\text{el}\top} & \hat{\mathbf{s}}^{\text{el}} \hat{\mathbf{s}}^{\text{cap}\top} \\ \hat{\mathbf{s}}^{\text{cap}} \hat{\mathbf{s}}^{\text{el}\top} & \hat{\mathbf{s}}^{\text{cap}} \hat{\mathbf{s}}^{\text{cap}\top} \end{bmatrix}, \quad (5.17)$$

so the diagonal quadrants provide cross-energy scale templates within each channel while the off-diagonal quadrants carry cross-section scale templates. Because $\hat{\mathbf{s}}$ is non-negative, $L^{(l, \kappa_r)}$ captures amplitude (variance) patterns implied by the stacked LOESS curves of Fig. 5.2 without imposing any particular correlation shape. As in

Block 3, the design matrix contains the upper-triangular elements of $L^{(l, \kappa_r)}$ for each r , contributing six L -columns per ladder that the learner can combine with the S , B , and F blocks to recover full covariance structure.

5.2.5 Intercept, diagonal-offset, and design-matrix assembly

For each ladder l we build a design matrix $X^{(l)} \in \mathbb{R}^{m \times p}$ by vectorizing the upper triangle of each block at each cardinality and stacking the resulting columns together with a constant and the offset index:

$$X^{(l)} = \begin{bmatrix} \mathbf{1}, \text{vecu}(S^{(l, \kappa_5)}), \text{vecu}(S^{(l, \kappa_4)}), \dots, \text{vecu}(S^{(l, \kappa_0)}), \\ \text{vecu}(B^{(l, \kappa_5)}), \text{vecu}(B^{(l, \kappa_4)}), \dots, \text{vecu}(B^{(l, \kappa_1)}), \\ \text{vecu}(F^{(l, \kappa_5)}), \dots, \text{vecu}(F^{(l, \kappa_0)}), \\ \text{vecu}(L^{(l, \kappa_5)}), \dots, \text{vecu}(L^{(l, \kappa_0)}), \boldsymbol{\xi} \end{bmatrix}, \quad (5.18)$$

where $\boldsymbol{\delta} = \text{vecu}(\Xi)$ with $\Xi_{ij} = \xi_{ij}$ from (5.9). The total column count is

$$p = 1 + 6(S) + 5(B) + 6(F) + 6(L) + 1 = 25.$$

The ladder-specific target vector is

$$\mathbf{y}^{(l)} = \text{vecu}(C_{\text{emp}}^{(l)}) \in \mathbb{R}^m. \quad (5.19)$$

Finally, stacking over ladders produces the global design and target used by the learning algorithms,

$$X = \begin{bmatrix} X^{(1)} \\ \vdots \\ X^{(L)} \end{bmatrix} \in \mathbb{R}^{(Lm) \times p}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(L)} \end{bmatrix} \in \mathbb{R}^{Lm}. \quad (5.20)$$

Following are some final remarks on the construction of regressors. All regressors are on their natural (barns²) scale and no further normalization is applied prior to fitting. The GLS block $S^{(l, \kappa_r)}$ encodes within-channel covariances; cross-channel structure is learned by combining B , F , and L with the diagonal-offset column δ , which allows the model to learn the empirical decay of correlations with increasing energy separation. Because $B^{(l, \kappa_0)} \equiv 0$, the bias block contributes five, not six, columns. Working on the upper triangle avoids duplication and yields $m = 5\,253$ rows per ladder; with $L = 200$ training ladders this gives 1,050,600 training samples for the elementwise map (5.7).

5.3 Reassembly and nearest-PSD projection

The learning step returns, for each ladder l , a prediction for every upper-triangular covariance element on the two-channel grid. Let $n = 102$ denote the total number of evaluation points (51 elastic energies followed by 51 capture energies) and $m = n(n+1)/2$ the number of upper-triangular entries. Write the predicted entries as a vector

$$\hat{\mathbf{c}}^{(l)} \in \mathbb{R}^m, \quad \hat{\mathbf{c}}^{(l)} = \{\hat{C}_{ij}^{(l)}\}_{1 \leq i \leq j \leq n}.$$

We reconstruct a full symmetric matrix by the inverse half-vectorization followed by explicit symmetrization,

$$\hat{C}_{\text{raw}}^{(l)} = \text{Sym}(\text{vech}^{-1}(\hat{\mathbf{c}}^{(l)})) = \frac{1}{2}(\text{vech}^{-1}(\hat{\mathbf{c}}^{(l)}) + \text{vech}^{-1}(\hat{\mathbf{c}}^{(l)})^\top) \in \mathbb{S}^n. \quad (5.21)$$

Indexing places the elastic points first and capture points second, so $\hat{C}_{\text{raw}}^{(l)}$ has the 2×2 block structure

$$\hat{C}_{\text{raw}}^{(l)} = \begin{bmatrix} \hat{C}_{\text{el,el}}^{(l)} & \hat{C}_{\text{el,cap}}^{(l)} \\ \hat{C}_{\text{cap,el}}^{(l)} & \hat{C}_{\text{cap,cap}}^{(l)} \end{bmatrix}, \quad \hat{C}_{\text{cap,el}}^{(l)} = (\hat{C}_{\text{el,cap}}^{(l)})^\top, \quad (5.22)$$

where each quadrant is 51×51 . The diagonal quadrants describe within-channel covariances across energy; the off-diagonal quadrants carry cross-channel covariances between elastic and capture.

Even though (5.21) enforces symmetry, generic regression or machine-learning predictions are unconstrained and need not satisfy the positive-semidefinite (PSD) requirement of covariance matrices: there may exist vectors $\mathbf{z}$ for which $\mathbf{z}^\top \hat{C}_{\text{raw}}^{(l)} \mathbf{z} < 0$. This can occur when some predicted off-diagonal entries are inconsistently large relative to the corresponding variances, or due to small extrapolation and numerical errors. To ensure our machine-learned quantities are valid covariance matrices, we project each reassembled two-channel matrix to the nearest PSD matrix [45].

For each ladder l we solve the convex projection problem

$$\hat{C}_{\text{psd}}^{(l)} = \arg \min_{X \in \mathbb{S}_+^n} \|X - \hat{C}_{\text{raw}}^{(l)}\|_F, \quad (5.23)$$

where $\mathbb{S}_+^n$ is the cone of $n \times n$ PSD matrices. The solution is obtained in closed form by an eigenvalue thresholding of the symmetric part:^{*}

$$B = \frac{1}{2} \left(\hat{C}_{\text{raw}}^{(l)} + \hat{C}_{\text{raw}}^{(l)\top} \right), \quad (\text{symmetrize}), \quad (5.24)$$

$$B = Q \Lambda Q^\top, \quad (\text{eigendecomposition}), \quad (5.25)$$

$$\Lambda_+ = \text{diag}(\max(\lambda_1, \varepsilon), \dots, \max(\lambda_n, \varepsilon)), \quad (\text{clip negatives}), \quad (5.26)$$

$$\hat{C}_{\text{psd}}^{(l)} = Q \Lambda_+ Q^\top. \quad (\text{reconstruct}) \quad (5.27)$$

Here λ_i are the eigenvalues of B , Q is orthogonal, and $\varepsilon \geq 0$ is a small numerical floor ($\varepsilon \approx 10^{-10}$ in our implementation) that can be set to zero for a merely semidefinite matrix or made positive to ensure strict positive definiteness. This projection alters the matrix as little as possible in Frobenius norm while enforcing all covariance constraints. By applying (5.24)–(5.27) to the full two-channel matrix $\hat{C}_{\text{raw}}^{(l)}$ —rather

^{*}This is the orthogonal projection onto $\mathbb{S}_+^n$ under the Frobenius norm. The initial “symmetrize” step below is included for completeness: in our pipeline $\hat{C}_{\text{raw}}^{(l)}$ is assembled by mirroring the predicted upper triangle, hence it is already exactly symmetric and $\frac{1}{2}(\hat{C}_{\text{raw}}^{(l)} + \hat{C}_{\text{raw}}^{(l)\top}) = \hat{C}_{\text{raw}}^{(l)}$.

than to the quadrants separately—we preserve cross-channel information during the PSD correction.

5.4 Mapped variance performance learned from full ECSCMs

When the learning target is the full two-channel covariance, rather than only the marginal variances as in Chapter 4, the predicted diagonal necessarily changes. The elementwise map $g(\cdot)$ in (5.7) is trained to explain the entire upper triangle, so its coefficients are shaped jointly by both on-diagonal and off-diagonal structure. Intuitively, large or coherent off-diagonal features (e.g., broad within-channel energy correlations or cross-channel bands) pull the learned diagonal toward values that make the whole matrix plausible under the regressors. This is in contrast to training on variances alone, where the diagonal is learned in isolation.

Figure 5.3 summarizes the performance of learned variances on the main diagonal for Normal, Cauchy, and Log-Normal families. Two qualitative changes are visible relative to Chapter 4: (i) some models previously identified as overly conservative like OLS are less conservative than they were when fit to variances only, reflecting the influence of off-diagonal structure during training; and (ii) other models, such as Random Forest remain consistently conservative, as before. The “Empirical” and “SAMMY” references behave as expected, with SAMMY under-covering and the empirical target alignment varying depending on the chosen probability model.

5.5 Coverage in a rotated basis

While diagonal coverage reveals how well a method calibrates its variances, it does not test whether the correlations are correct. To assess correlation accuracy, we propose

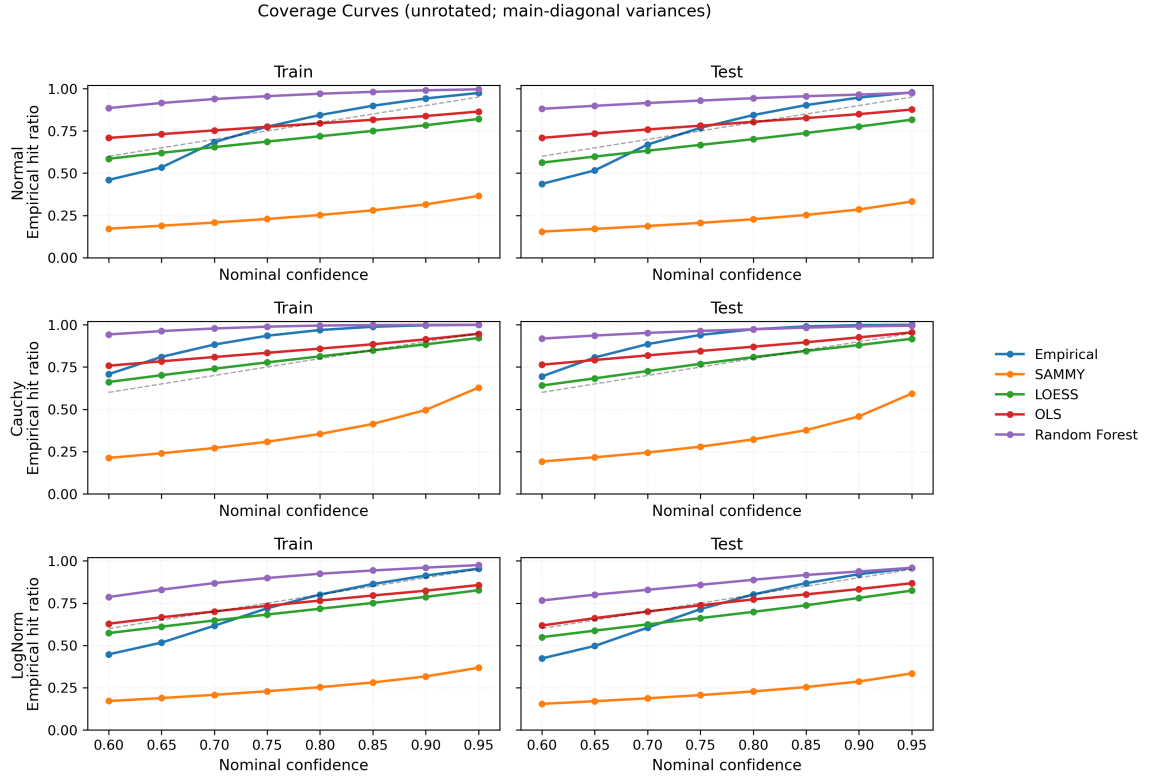


Figure 5.3: Coverage curves of the variances (main-diagonals): empirical hit ratio versus nominal confidence for Normal, Cauchy, and Log-Normal families, split by train/test. Because the diagonal is learned jointly with off-diagonals, OLS is less conservative than in Chapter 4, while Random Forest performance is relatively consistent.

a methodology which measures coverage after rotating residuals into a basis that decorrelates the residuals.

For a ladder l and sample n , let $\mathbf{r}^{(n)} \in \mathbb{R}^{102}$ be the residual vector defined in Section 5.1. Let $\hat{C}^{(l)} \in \mathbb{S}^{102}$ be the method's predicted two-channel covariance for that ladder, and compute its eigendecomposition

$$\hat{C}^{(l)} = Q^{(l)} \Lambda^{(l)} Q^{(l)\top}, \quad \Lambda^{(l)} = \text{diag}(\lambda_1^{(l)}, \dots, \lambda_{102}^{(l)}),$$

with $Q^{(l)}$ orthogonal and eigenvalues $\lambda_k^{(l)} \geq 0$. Define the rotated residuals

$$\mathbf{u}^{(n,l)} = Q^{(l)\top} \mathbf{r}^{(n)} \in \mathbb{R}^{102}. \quad (5.28)$$

If the model family is correct and the predicted covariance is accurate, then in this principal-component (PC) basis each coordinate u_k should behave like an independent univariate error with scale parameter set by a function of λ_k .

Under the multivariate Normal model $\mathbf{r}^{(n)} \sim \mathcal{N}(\mathbf{0}, \hat{C}^{(l)})$, the rotated coordinates are independent $u_k \sim \mathcal{N}(0, \lambda_k)$. For a nominal two-sided confidence level $1 - \alpha$, the (symmetric) interval for u_k is

$$I_k^N(1-\alpha) = \left[-z_{1-\alpha/2} \sqrt{\lambda_k}, z_{1-\alpha/2} \sqrt{\lambda_k} \right], \quad (5.29)$$

where z_p is the standard-Normal p -quantile. We declare a hit when $u_k \in I_k^N(1-\alpha)$ and report the empirical hit ratio over all ladders, samples, and PC indices k . When both the shape (Q) and the scale (Λ) of the covariance are correct, the observed hit ratio should track the nominal level across α .

This methodology can be adapted for non-Normal probability models. For users wishing to model heavy-tailed errors, one could use the multivariate Cauchy with scatter matrix $\hat{C}^{(l)}$. The PC rotation (5.28) diagonalizes the scatter, yielding independent univariate Cauchy coordinates $u_k \sim \text{Cauchy}(0, \gamma_k)$ with $\gamma_k = \sqrt{\lambda_k}$. The

central $(1 - \alpha)$ interval is

$$I_k^C(1-\alpha) = \left[-h_\alpha \gamma_k, h_\alpha \gamma_k \right], \quad h_\alpha = \tan\left(\frac{\pi}{2}(1 - \alpha)\right), \quad (5.30)$$

and hits are defined analogously.

The methodology can also be extended to non-negative probability models, such as the Log-Normal. Let $\hat{\boldsymbol{\sigma}} \in \mathbb{R}_{>0}^{102}$ be the fitted two-channel curve on $\mathcal{E}$ and define the log-error vector

$$\boldsymbol{\ell}^{(n)} = \log(\boldsymbol{\sigma}^{\text{true}}) - \log(\hat{\boldsymbol{\sigma}}) \approx D^{-1} \mathbf{r}^{(n)}, \quad D = \text{diag}(\hat{\boldsymbol{\sigma}}), \quad (5.31)$$

where the approximation follows from a first-order (delta-method) linearization. We therefore construct a log-scale covariance by congruence transformation,

$$\hat{C}_{\log} = D^{-1} \hat{C} D^{-1}, \quad (5.32)$$

rotate $\boldsymbol{\ell}^{(n)}$ using the eigenvectors of $\hat{C}_{\log}$, and apply Normal intervals (5.29) with the eigenvalues of $\hat{C}_{\log}$.

Because the rotation uses the method's own Q and Λ , it simultaneously probes (i) whether the predicted correlation directions are correct (misaligned Q mixes signal across PCs and degrades coverage), and (ii) whether the scales are right (too small/large λ_k lead to under/over-coverage). The resulting hit-ratio curves therefore serve a desirable metric for assessing the relative performance of machine learned covariances when predicting correlations between energies and channels.

Figure 5.4 reports coverage after PC rotation, for Normal, Cauchy, and Log-Normal. With over one million training rows (upper-triangular elements across 200 ladders), the linear map $g(\cdot)$ has ample data to learn how the $S/B/F/L$ blocks combine into full ECSCMs. Relative to training on variances alone (Chapter 4), linear models such as OLS show measurable gains in rotated coverage indicating that its learned correlation structure is substantially better calibrated.

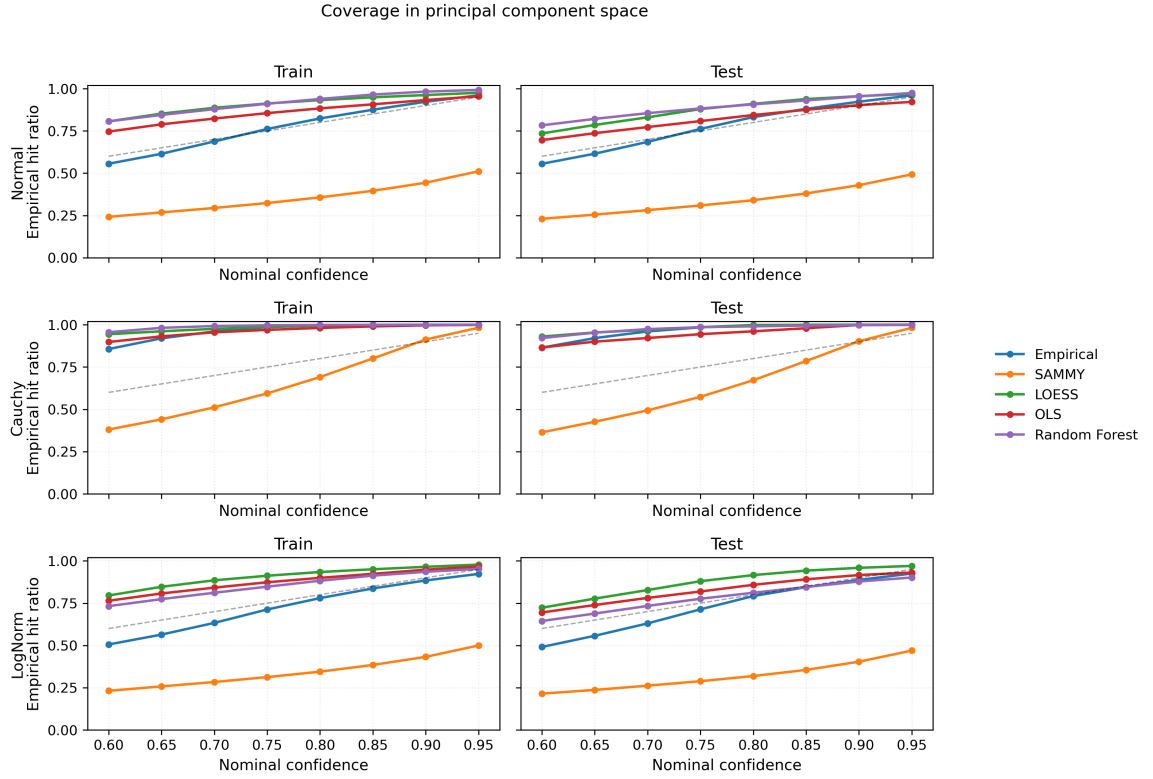


Figure 5.4: Coverage after rotating to the method’s principal-component basis. Normal (top), Cauchy (middle), and Log-Normal (bottom) families; train (left) and test (right). OLS benefits from learning the full ECSCM and attains markedly better rotated coverage than when trained on variances only. RF continues to be conservative across the board.

If conservatism is paramount (e.g., downstream tolerances require systematic over-coverage), Random Forests remains a balanced option that are adequately conservative. If efficiency is important, OLS delivers accuracy close to the nominal line without the computational cost of Random Forests training/inference, but at the expense of not capturing as many missed resonances as methods such as Random Forests. Figures 5.5 and 5.6 are illustrative of this relative performance. OLS is notably better when trained on the full covariance regressor set.

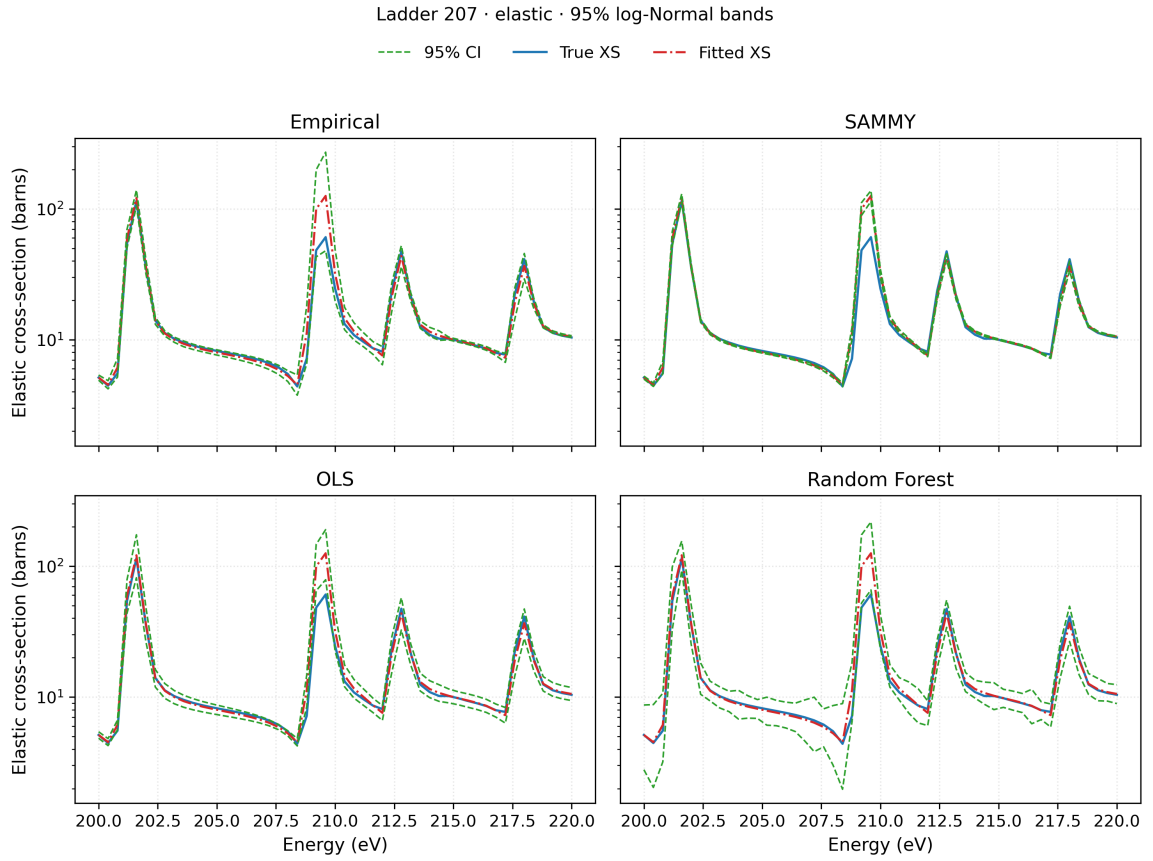


Figure 5.5: Elastic cross-section bands at 95% nominal confidence for four methods (Empirical, SAMMY, OLS, Random Forest) on a representative ladder. Green dashed curves are Log-Normal bands derived from the learned full covariance; blue solid is the true cross section; red dash-dot is the fitted curve.

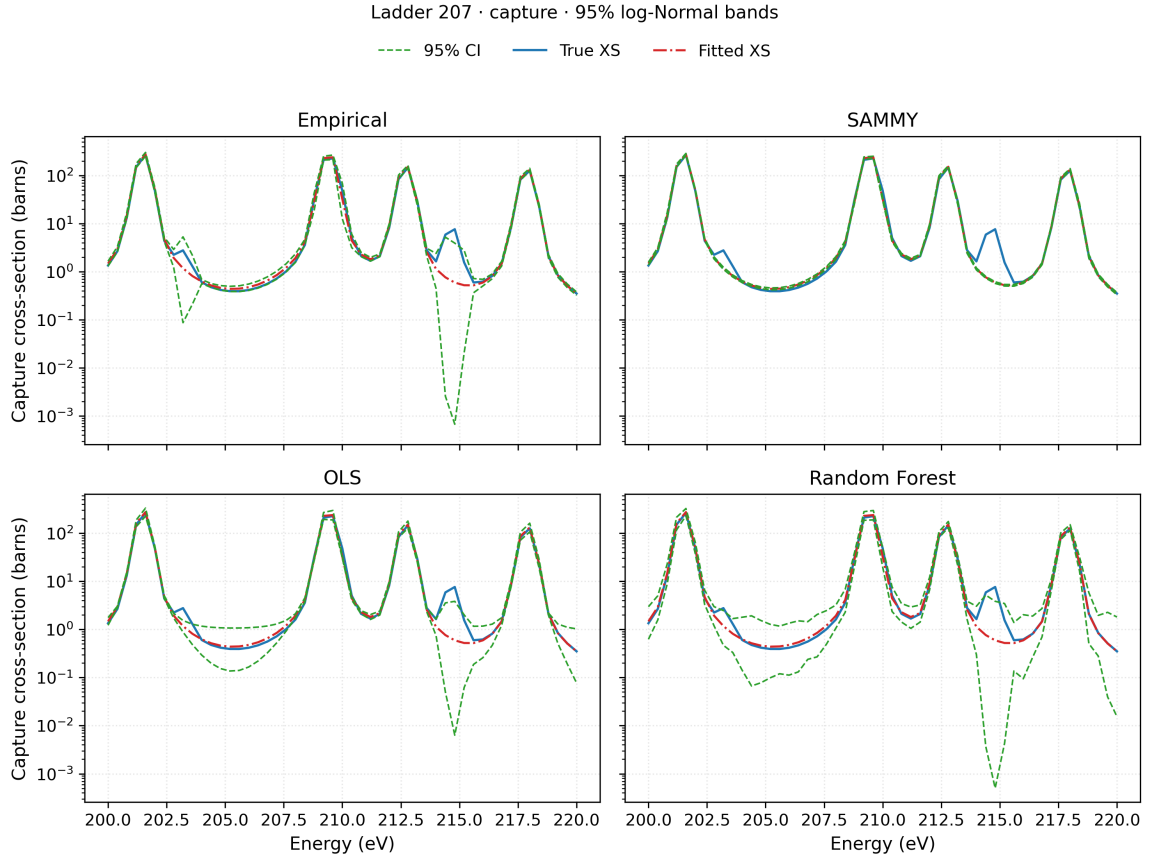


Figure 5.6: Capture cross-section bands at 95% nominal confidence (same four methods and ladder as in Figure 5.5). RF remains conservative; OLS strikes a favorable accuracy–efficiency balance.

Chapter 6

Conclusions and Future Work

This thesis asked three practical questions for resonance-region uncertainty quantification. (i) How do violations of generalized least-squares (GLS) assumptions distort evaluated cross-section covariances (ECSCMs)? (ii) How large are those distortions for realistic departures from ideality? (iii) Can we restore statistical fidelity by an automated post-processing step that requires no manual tuning? Using physics-based synthetic data and a reproducible fitting workflow, we quantified the failures, measured their impact, and built correctives that are calibrated by confidence interval coverage.

The central contribution is a complete, reusable methodology. Given (a) access to synthetic data such as SYNDAT and (b) an approved, scripted fitting procedure such as AUTOFIT, we showed how to (1) measure the effect of controllable conditions on ECSCM accuracy, (2) diagnose and select an adequate probability model for residuals, and (3) use information across multiple resonance cardinalities to learn cross-section covariances that capture marginal variances, correlations across energies, and correlations across channels.

Chapter 3 used controlled synthetic studies to separate the effects of three fragile assumptions in GLS resonance analysis: complete level identification, a correct reaction formalism, and an accurate data covariance. The results show a steep cost

to ECSCM accuracy when a resonance is missed (wrong cardinality) or when an inadequate formalism is used; both induce significant, measurable error in SAMMY-generated ECSCMs. In contrast, reasonable misreporting of the systematic data covariance was comparatively benign; ECSCM accuracy was fairly robust to such errors over several orders of magnitude.

Chapter 4 established that covariance information in higher-order models uncovered by AUTOFIT can be used in machine learning methods to correct the diagonal of the ECSCM. Across 60–95% nominal targets, empirical coverage on an independent test set saw measurable improvement from the GLS predicted variances. Among the methods tried, random forests (RF) demonstrated the best balance between accuracy and coverage conservatism.

Chapter 5 extended the methodology beyond marginal variances to the full two-channel covariance on a fixed energy grid per channel (elastic followed by capture). We constructed a design matrix that acts elementwise on the covariance surface: GLS-provided channelwise ECSCMs (S), cardinality-bias outer products (B), cross-section outer products (F), and LOESS-predicted scale outer products (L), augmented by a diagonal-offset regressor. After predicting the upper-triangle entries we reassembled the full matrix and applied a nearest-PSD projection to guarantee a valid covariance. A practical implication is worth noting: because the learner operates on the full matrix, the recovered variances need not match those obtained when training on the diagonal alone. Indeed, our unrotated coverage curves showed that simple models such as ordinary least squares (OLS) is less conservative in this setting than in Chapter 4, while RF remained consistently conservative across data sets.

A key contribution of Chapter 5 is a way to judge the learned correlations, not just variances. We introduced a hit-ratio diagnostic in rotated coordinates: evaluate coverage after orthogonally rotating the residual vector and the predicted covariance (e.g., to principal-component space). For multivariate normal errors, if $\hat{C}$ matches the true covariance C , then the standardized PC coordinates $z = \Lambda^{-1/2}U^\top r$ are

i.i.d. $N(0, 1)$ ($C = U\Lambda U^\top$), so nominal one-dimensional intervals must achieve the nominal coverage. Departures in the rotated curves therefore diagnose correlation misspecification. We demonstrated how to extend this methodology to non-Normal probability models such as the Cauchy and Log-Normal. Applying this diagnostic showed that, with the abundance of training data available for the full-matrix task, OLS closed much of the gap to RF in rotated space: coverage is closer to nominal while maintaining substantially lower computational costs. If strict conservatism is paramount, RF remains the best choice among the tested models; if efficiency is important, OLS provides materially improved accuracy for significantly less computational cost.

Across calibration tests the log-normal model was an adequate and physically coherent probability model: it matched the empirical hit ratios as well as the alternatives while guaranteeing non-negative bands. The final figures in Chapter 5 demonstrate that the learned full covariances produce realistic, log-normal confidence bands for both channels that track true cross sections and widen near resonances that were missed by the AutoFit final cardinality model.

6.1 Future Work

Several opportunities remain, each naturally integrated into the present framework.

First, we propose extending the model adequacy study beyond the models tested here. This thesis provides a framework for measuring the accuracy of any probability model of interest. Future work may be able to identify a probability model that more accurately characterizes empirical residuals without the need to change distribution parameters as a function of nominal confidence.

We recommend future research explore the use of neural networks as a candidate model to be compared with the machine learning methods tested in this thesis. The framework presented in this thesis can be used to test a variety of neural network architectures and hyperparameter sets.

We propose adapting the workflow to uncertainty quantification of angular distributions. The framework presented in this thesis could be easily extend to uncertainties on measures other than cross-sections. Extending these methods to assess the accuracy of angular distribution covariances would be a meaningful contribution to the nuclear data community. Likewise, the methodology presented in this thesis could be extended to additional nuclides and to different resonance windows.

The methodology could also be applied to resonance windows at higher or lower energies. Doing so would not only provide calibrated covariance estimates in those regions, but would also enable a direct comparison between two approaches: (i) applying parameters learned in one energy region to another (extrapolation), and (ii) relearning parameters directly in the new region. The degree to which these two agree would indicate how confidently the methodology can be generalized across energy ranges, and where it may require re-training. This comparison would offer a principled way to test the robustness of the learned mappings beyond the specific window studied here.

The emphasis of this thesis has been methodological. We established a transparent, reproducible path from synthetic benchmarks to calibrated uncertainties for evaluated cross sections. These tools represent a fully reproducible framework for accurate, repeatable uncertainty quantification of evaluated cross-sections in the resolved resonance region.

Bibliography

- [1] CM Perey, FG Perey, JA Harvey, NW Hill, and NM Larson. 56 fe resonance parameters for neutron energies up to 850 kev. Technical report, Oak Ridge National Lab., 1990. [xi](#), [8](#)
- [2] FH Frohner. Evaluation and analysis of nuclear resonance data. 2000. [xi](#), [8](#), [11](#)
- [3] 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. ISSN 0090-3752. doi: <https://doi.org/10.1016/j.nds.2018.02.001>. URL <https://www.sciencedirect.com/science/article/pii/S0090375218300206>. Special Issue on Nuclear Reaction Data. [1](#)

- [4] Arjan JM Plompen, O Cabellos, Cyrille de Saint Jean, Michael Fleming, A Algora, Maurizio Angelone, P Archier, E Bauge, O Bersillon, A Blokhin, et al. The joint evaluated fission and fusion nuclear data library, jeff-3.3. *The European Physical Journal A*, 56:1–108, 2020. [1](#)
- [5] Friederike Bostelmann, William A. Wieselquist, and Matthew A. Jessee. Impact of the endf/b-viii.0 library on advanced reactor simulations. Technical report, Oak Ridge National Laboratory, 2019. URL <https://www.osti.gov/servlets/purl/1607092>. [1](#)
- [6] National Nuclear Data Center. Who uses endf data and how?, n.d. URL <https://www.nndc.bnl.gov/endl/usage.html>. [1](#)
- [7] Arjan J. Koning, Aruna Korde, Dmytro Fedorchenko, Hanna Skliarova, Mark De Jong, Melissa Denecke, Stephan Heinitz, Sergey Chemerisov, Valeriia Starovoitova, Valery I. Radchenko, Vitaly Khryachkov, and William Diamond. Photonuclear production of radioisotopes. Technical Report IAEA-TECDOC-2051, International Atomic Energy Agency, Vienna, Austria, 2024. URL <https://www-pub.iaea.org/MTCD/publications/PDF/TE-2051web.pdf>. [1](#)
- [8] Lee A Bernstein, Friederike Bostelmann, Mike Carpenter, Mark Chadwick, Max Fratoni, Ayman Hawari, Lawrence Heilbronn, Calvin Howell, Jo Ressler, Cynthia Keppel, et al. Second report of the nuclear data subcommittee of the nuclear science advisory committee. Technical report, Lawrence Livermore National Laboratory (LLNL), Livermore, CA (United States), 2023. [1](#)
- [9] Oscar Cabellos and Luca Fiorito. Examples of monte carlo techniques applied for nuclear data uncertainty propagation. In *EPJ Web of Conferences*, volume 211, page 07008. EDP Sciences, 2019. [1](#)
- [10] E Dupont, MB Chadwick, Y Danon, C De Saint Jean, M Dunn, U Fischer, RA Forrest, T Fukahori, Z Ge, Hiroshi Harada, et al. Working party on

- international nuclear data evaluation cooperation (wpec). *Nuclear data sheets*, 120:264–267, 2014. 2
- [11] WPEC Subgroup 49 (SG49). Reproducibility in nuclear data evaluation. <https://oecd-nea.org/download/wpec/sg49/index.htm>, 2020. Accessed: 2025-03-04. 2
- [12] Nancy M Larson. Updated user’s guide for sammy: Multilevel r-matrix fits to neutron data using bayes’ equations. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2008. Section IV.E.6. 2, 15, 17
- [13] Roberto Capote, Sergei Badikov, Allan D Carlson, Ignacio Duran, Frank Gunsing, Denise Neudecker, Vladimir G Pronyaev, Peter Schillebeeckx, Georg Schnabel, Donald L Smith, et al. Unrecognized sources of uncertainties (usu) in experimental nuclear data. *Nuclear Data Sheets*, 163:191–227, 2020. 2
- [14] Denise Neudecker, Donald L Smith, Fredrik Tovesson, Roberto Capote, Morgan Curtis White, Nathaniel S Bowden, Lucas Snyder, Allan D Carlson, Robert J Casperson, Vladimir Pronyaev, et al. Applying a template of expected uncertainties to updating ^{239}Pu (n, f) cross-section covariances in the neutron data standards database. *Nuclear Data Sheets*, 163:228–248, 2020. 2
- [15] Noah Walton, Jesse Brown, William Fritsch, Dave Brown, Gustavo Nobre, and Vladimir Sobes. Methodology for physics-informed generation of synthetic neutron time-of-flight measurement data. *Computer Physics Communications*, 294:108927, 2024. 2, 18, 19, 20
- [16] Noah AW Walton, Oleksii Zivenko, William Fritsch, Jacob Forbes, Amanda Lewis, Jesse Brown, and Vlad Sobes. Automated resonance identification in nuclear data evaluation. *arXiv preprint arXiv:2402.14122*, 2024. 2
- [17] Noah AW Walton, Oleksii Zivenko, Amanda M Lewis, William Fritsch, Jacob Forbes, Jesse M Brown, David A Brown, Gustavo PA Nobre, and Vladimir Sobes.

- Automated resonance fitting for nuclear data evaluation. *Nuclear Science and Engineering*, pages 1–16, 2025. 2, 44
- [18] Noah Walton. *A Computational Framework for Automated Nuclear Resonance Evaluation and Validation*. Ph.d. dissertation, University of Tennessee, Knoxville, TN, 2024. URL https://trace.tennessee.edu/utk_graddiss/11394. 2
- [19] Oleksii Zivenko, Noah AW Walton, William Fritsch, Jacob Forbes, Amanda M Lewis, Aaron Clark, Jesse M Brown, and Vladimir Sobes. Validating automated resonance evaluation with synthetic data. *Annals of Nuclear Energy*, 212:111081, 2025. 2
- [20] Department of the Air Force. Air force doctrine note 25-1: Artificial intelligence. Doctrine note, Department of the Air Force, Maxwell Air Force Base, AL, April 2025. URL https://www.doctrine.af.mil/Portals/61/documents/AFDN_25-1/AFDN%2025-1%20Artificial%20Intelligence.pdf. 2
- [21] Steven Goldfarb and Katarina Anthony. Broad as a barn. *Nature Physics*, 15(4):414–414, 2019. 6
- [22] Niels Bohr. Neutron capture and nuclear constitution. *Uspekhi Fizicheskikh Nauk*, 16(4):425–435, 1936. 7
- [23] Gregory Breit and Eugene Wigner. Capture of slow neutrons. *Physical review*, 49(7):519, 1936. 9
- [24] AM Lane and RG Thomas. R-matrix theory of nuclear reactions. *Reviews of Modern Physics*, 30(2):257, 1958. 10
- [25] CW Reich and MS Moore. Multilevel formula for the fission process. *Physical Review*, 111(3):929, 1958. 11

- [26] Nancy M Larson. Updated user’s guide for sammy: Multilevel r-matrix fits to neutron data using bayes’ equations. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2008. [11](#)
- [27] John R. Helliwell. *A Beginner’s Guide to Neutron Reactor and Spallation Sources*, pages 13–14. Springer Nature Switzerland, Cham, 2025. ISBN 978-3-031-80181-5. doi: 10.1007/978-3-031-80181-5_2. [12](#)
- [28] Dan Gabriel Cacuci. *Handbook of Nuclear Engineering: Vol. 1: Nuclear Engineering Fundamentals; Vol. 2: Reactor Design; Vol. 3: Reactor Analysis; Vol. 4: Reactors of Generations III and IV; Vol. 5: Fuel Cycles, Decommissioning, Waste Disposal and Safeguards*, volume 1. Springer Science & Business Media, 2010. [12](#)
- [29] John RD Copley and Terrence J Udovic. Neutron time-of-flight spectroscopy. *Journal of research of the National Institute of Standards and Technology*, 98(1): 71, 1993. [12](#)
- [30] Johann Heinrich Lambert. *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Eberhard Klett, Augsburg, 1760. [13](#)
- [31] August Beer. Bestimmung der absorption des rothen lichts in farbigen flüssigkeiten. *Annalen der Physik und Chemie*, 86:78–88, 1852. [13](#)
- [32] Nancy M Larson. Updated user’s guide for sammy: Multilevel r-matrix fits to neutron data using bayes’ equations. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2008. Section IV.A. [15](#)
- [33] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76:243–297, 2021. [18](#)

- [34] Barbara Zaparoli Cunha, Christophe Droz, Abdel-Malek Zine, Stéphane Foulard, and Mohamed Ichchou. A review of machine learning methods applied to structural dynamics and vibroacoustic. *Mechanical Systems and Signal Processing*, 200:110535, 2023. [18](#)
- [35] Xizhao Wang and Yulin He. Learning from uncertainty for big data: future analytical challenges and strategies. *IEEE Systems, Man, and Cybernetics Magazine*, 2(2):26–31, 2016. [18](#)
- [36] Marília Barandas, Lorenzo Famiglini, Andrea Campagner, Duarte Folgado, Raquel Simão, Federico Cabitza, and Hugo Gamboa. Evaluation of uncertainty quantification methods in multi-label classification: A case study with automatic diagnosis of electrocardiogram. *Information Fusion*, 101:101978, 2024. [18](#)
- [37] Curtis Northcutt, Lu Jiang, and Isaac Chuang. Confident learning: Estimating uncertainty in dataset labels. *Journal of Artificial Intelligence Research*, 70: 1373–1411, 2021. [18](#)
- [38] Arthur Hoarau, Constance Thierry, Arnaud Martin, Jean-Christophe Dubois, and Yolande Le Gall. Datasets with rich labels for machine learning. In *2023 IEEE International Conference on Fuzzy Systems (FUZZ)*, pages 1–6. IEEE, 2023. [18](#)
- [39] Reihaneh H Hariri, Erik M Fredericks, and Kate M Bowers. Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big data*, 6(1): 1–16, 2019. [18](#)
- [40] Jesse M Brown. *Measurements, evaluation, and validation of ta-181 resolved and unresolved resonance regions*. Rensselaer Polytechnic Institute, 2019. [18](#), [19](#)
- [41] Nancy 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. Section IV.D.3. [37](#)

- [42] Ketaki Joshi, Nicholas Branam, Isaac Meyer, Ben Forget, Abdulla Alhajri, and Vladimir Sobes. An analytic benchmark for neutron boltzmann transport with downscattering—part iii: Uncertainty propagation and multigroup covariance matrices. *Nuclear Science and Engineering*, 197(7):1356–1363, 2023. [37](#), [38](#)
- [43] Denise Neudecker, Roberto Capote, and Helmut Leeb. Impact of model defect and experimental uncertainties on evaluated output. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 723:163–172, 2013. doi: 10.1016/j.nima.2013.05.005. [41](#)
- [44] Trevor Hastie. The elements of statistical learning: data mining, inference, and prediction, 2009. [63](#)
- [45] Nicholas J Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear algebra and its applications*, 103:103–118, 1988. [88](#)

Vita

Jacob Forbes was born in Vista, California and spent his formative years in Baton Rouge, Louisiana and Kingwood, Texas. He graduated from the U.S. Air Force Academy in 2013 and commissioned as an Operations Research Analyst in the U.S. Air Force. Jacob's previous assignments include: analyst in the Analysis and Lessons Learned directorate, Headquarters Air Force Global Strike Command, Barksdale AFB, LA; analyst in the Strategy Division, 609 Combined Air Operations Center, Al Udeid Air Base, Qatar; and instructor at the Department of Mathematical Sciences, U.S. Air Force Academy, CO. Jacob's doctoral research is focused on applying computational methods and machine learning to challenges in nuclear physics. Following graduation, he will continue his research as an Air Force Fellow at Oak Ridge National Laboratory. He is profoundly grateful for the support of his family, friends, and mentors, whose encouragement sustained him throughout his graduate studies.