

An optical-lensing inspired data thinning method for nuclear cross section data

M. Imbrišak,* A. E. Lovell, and M. R. Mumpower
Los Alamos National Laboratory, Los Alamos, NM, 87545, USA
(Dated: August 28, 2025)

In the study of nuclear cross sections, the computational demands of data assimilation methods can become prohibitive when dealing with large data sets. We have developed a novel variant of the data thinning algorithm, inspired by the principles of optical lensing, which effectively reduces data volume while preserving critical information. We show how it improves fitting through a toy problem and for several examples of total cross sections for neutron-induced reactions on rare-earth isotopes. We demonstrate how this method can be applied as an efficient pre-processing step prior to smoothing, significantly improving computational efficiency without compromising the quality of uncertainty quantification.

I. INTRODUCTION

When fitting physical models to experimental data, it is common to encounter large, and often correlated, datasets. This poses challenges not only in terms of computational cost—particularly when the models are themselves expensive to calculate—but also in data assimilation and uncertainty quantification. To address these issues, various data thinning techniques have been developed to reduce the volume of data used during optimization without significantly degrading model accuracy. In this paper, we introduce a novel data thinning method inspired by optical lensing. Such data thinning techniques are broadly applicable to challenges across nuclear and astrophysical modeling. In nuclear physics, evaluated data projects like ENDF/B [1] must assimilate thousands of correlated cross-section datasets. Surrogate reactions [2] also require efficient treatment of dense experimental data. In astrophysics, r-process nucleosynthesis [3, 4], supernova modeling [5], and EOS inference from neutron star mergers [6] are all data-intensive domains where thinning methods can mitigate computational bottlenecks in uncertainty propagation and model calibration.

The Hauser-Feshbach (HF) statistical model is used to compute nuclear reactions on medium- to heavy-mass nuclei. The HF model calculates energy-averaged nuclear reaction cross sections and incorporates a width fluctuation correction for overlapping resonances. Out of the several HF codes available on the market, including EMPIRE [7], TALYS [8], CCONE [9], we focus our attention on the LANL-developed CoH₃ code [10, 11]. These codes in general handle multi-particle evaporation from a compound nucleus and provide not just the reaction cross sections, but also a wealth of information on other observables such as the energy and angular distributions of secondary particles, γ -ray production cross sections, and the production of isomeric states. The CoH₃ code [10, 11] integrates the coupled-channels optical model to achieve precise nuclear reaction calculations within the

keV to tens of MeV energy range. The CoH₃ package is structured into modules that provide a comprehensive toolkit for nuclear reaction analysis: the one-body potential mean-field theory, the coupled-channels optical model, and the HF statistical decay model, modules that handle direct and semidirect radiative capture, pre-equilibrium processes, and prompt fission neutron emission.

The Experimental Nuclear Reaction Data (EXFOR) [12] database is often used to aggregate data of multiple nuclear reaction data, which includes cross section experiments. The well-known problem in such data aggregation is that it neglects experimental correlations, such as those between experiments [13]. To account for this, an iterative generalized least-squares algorithm, the Full Bayesian Evaluation Technique (FBET) [14, 15] was developed for nuclear data evaluation. FBET is a data assimilation method that evaluates nuclear data at a chosen energy grid and provides an estimate for correlated experimental uncertainties using a linearized Bayesian update procedure. These experimental uncertainties rank equally as important as uncertainties from model defects [16]. The FBET approach to treating correlated measurements is not unique to nuclear physics and belongs to a wider class of data assimilation methods which have been successfully applied in geophysical models [17, 18] and industrial applications [19]. A problem often encountered in these fields is computational complexity of large data sets needed for data assimilation. One of the solutions to this problem is reducing the amount of data in a statistically consistent way by finding an optimal subset of data that describes the wanted quantity equally well as the entire dataset. One class of such methods is data thinning, which accomplishes this using either clustering [20] or iteratively applying a selected statistic [21].

We have developed a data thinning algorithm inspired by optical lensing to improve the analysis of cross section data in the region with pronounced resonances. The data for many nuclei below ~ 1 MeV is abundant, but due to finite-energy sampling, it can often be hard to attribute a data point to a resonance peak versus the “smooth” underlying HF curve. While in principle, FBET can handle such data, it is often impractical to apply it to large datasets due to memory requirements. We propose our

* mimbrisak@lanl.gov

lensing method as a pre-processing step in order to identify the most impactful data points in a measurement set.

We describe the FBET approach, the CoH₃ code, and present the percentile-based and KDE-based data thinning methods in sections II A and II B. In Sec. IV, we then describe the testing of the lensing method and present our results in V.

II. LENSING

A. Percentile-based lensing estimates

Our lensing-inspired thinning method requires binning a dataset into N_b energy bins, $[x_1, x_2], \dots, [x_b, x_{b+1}], \dots, [x_{N_b}, x_{N_b+1}]$, and computing the percentiles corresponding to the minimum, $p = 0$, median, $p = 50$, and maximum $p = 100$ values of cross sections within the bin b , $y_{p,b}$. We then introduce the likelihood:

$$L_b(y, \sigma_y) = \prod_{p \in \{0, 50, 100\}} e^{-\frac{1}{2} \left(\frac{y - y_{p,b}}{\sigma_y} \right)^2}, \quad (1)$$

and find the index of the measurement from bin b that maximizes the likelihood of bin b' , $L_{b'}$,

$$i_{b,b'} = \operatorname{argmax}_{x_i \in [x_b, x_{b+1}]} L_{b'}(y_i, \sigma_{y_i}). \quad (2)$$

The procedure is as follows. We iterate over bins $b = 1, \dots, N_b$ and find the measurements $i_{b,b+1}$. We thereby get one measurement per bin. We then remove these measurements from the minimization and repeat the procedure to find the measurements $i'_{b,b+1}$. This procedure can be repeated until the desired limit on the number of measurements per bin is reached.

B. Kernel-density-based estimates

We also developed a modified version of the lensing method from Sec. II A that uses kernel density estimates (KDE) of the bin b'

$$K_h(y | h, b') = \sum_{x_j \in [x_{b'}, x_{b'+1}]} \frac{e^{-\frac{(y - y_j)^2}{2h^2}}}{\sqrt{2\pi} h n_{b'}}, \quad (3)$$

where $n_{b'}$ is the number of measurements in the bin b' . In this variant, the index of the measurement from bin b that maximizes the probability density of data from bin b' , $L_{b'}$,

$$i_{b,b'} = \operatorname{argmax}_{x_i \in [x_b, x_{b+1}]} K_h(y_i | h, b'). \quad (4)$$

This method has the advantage of taking into account more details of the measurement values in the bin b' and

is the limiting case of using all of the percentiles for computing the likelihood $L_b(y, \sigma_y)$.

A straightforward generalization of the KDE method involves taking errors σ_y instead of a fixed bandwidth h . This σ -based KDE lensing approach naturally adapts the KDE method to the measurement uncertainty of each data point, thereby preserving the probabilistic structure of the input distribution. Unlike traditional KDE lensing, which applies a fixed kernel width regardless of the measurement precision, the σ -based variant ensures that more precise data contribute sharper likelihood features, while noisier data are appropriately smoothed. This makes the method especially advantageous when the data exhibit heteroscedasticity, i.e., when uncertainties vary significantly across samples.

III. METHODS

Our lensing method can be used as a pre-processing step when analyzing dense data, such as nuclear cross-section data in the resonance region. We demonstrate the entire procedure by applying the FBET data smoothing procedure and fitting the CoH₃ model. Therefore, in sections III A, III B, and III C we briefly describe the standard FBET techniques, the CoH₃ optical potential model and the bootstrap technique used for testing.

A. FBET

In nuclear data evaluation the FBET method is often used to incorporate prior information by a linearized approximation to the Bayes rule [see, e.g., 14, for details]. The experimental nuclear cross sections \mathbf{y}_m are modeled as functions of a random variable, \mathbf{y} that describes cross sections evaluated at chosen energy grid points, subject to model M , i.e., $\mathbf{y}_m = f_M(\mathbf{y})$. This procedure assumes the multivariate normal distribution for the prior values of \mathbf{y} :

$$p(\mathbf{y} | M) = N e^{-\frac{1}{2}(\mathbf{y} - \mathbf{y}_0)^T A_0^{-1}(\mathbf{y} - \mathbf{y}_0)}, \quad (5)$$

where vector \mathbf{y}_0 is the *a priori* mean value of the parameter vector \mathbf{y} , A_0 is the a-priori covariance matrix of the parameters, and N the normalization factor. By further assuming that experimental data have uncertainties described by the experimental covariance matrix B - typically containing measurement uncertainties on the diagonal - one obtains the posterior distribution

$$p(\mathbf{y} | M) = N' \exp \left(-\frac{1}{2}(\mathbf{y} - \mathbf{y}_0)^T A_0^{-1}(\mathbf{y} - \mathbf{y}_0) \right) \quad (6)$$

$$- \frac{1}{2}(f(\mathbf{y}) - \mathbf{y}_m)^T B^{-1}(f_M(\mathbf{y}) - \mathbf{y}_m), \quad (7)$$

where N' is a new normalization constant that absorbs the contributions from both the prior and likelihood terms. The procedure introduces the sensitivity matrix

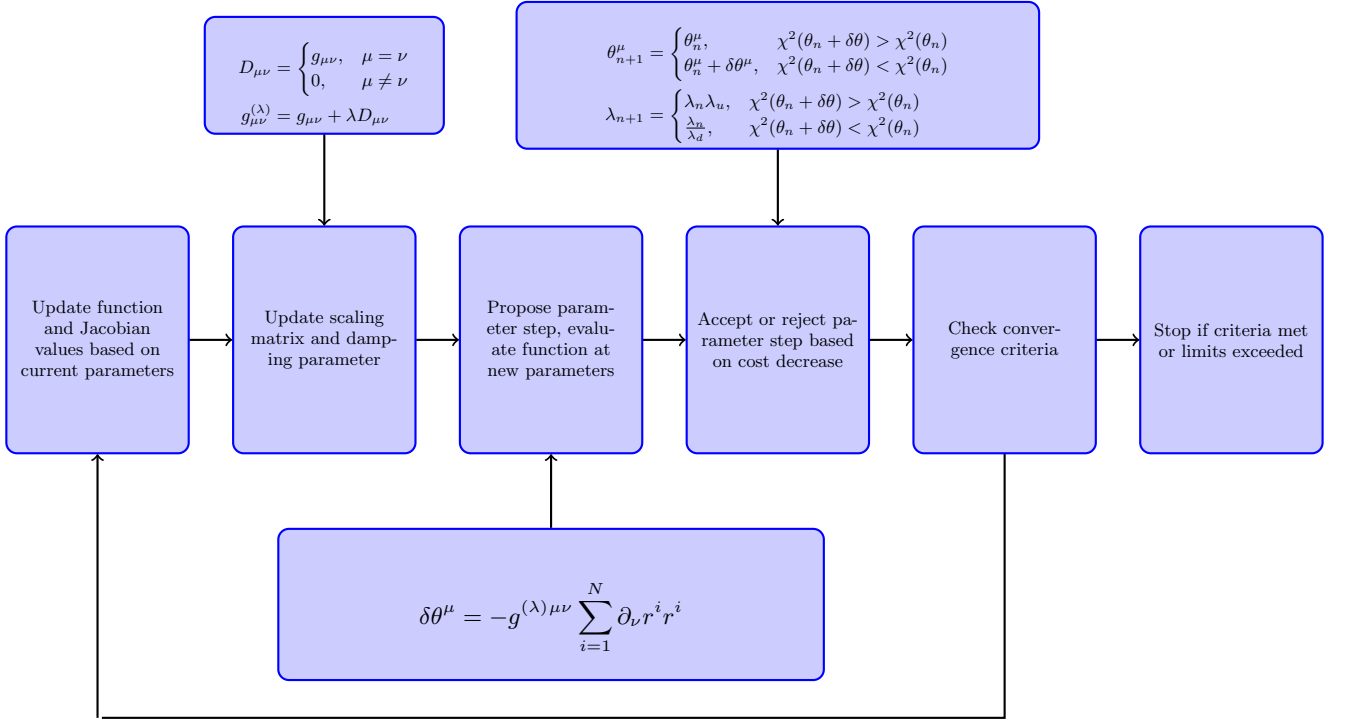


FIG. 1: Schematic representation of the Levenberg-Marquardt procedure.

Lensing method	slope	intercept	$t(\text{slope})$	$P(\text{slope})$	$t(\text{intercept})$	$P(\text{intercept})$	Chow's F	$P(\text{Chow's } F)$	Evaluation time [ms]
Percentile-based	0.98 ± 0.31	2.62 ± 0.16	3.33	0.02	3.85	0.01	0.12	0.9965	11.62 ± 0.95
KDE-based	1.91 ± 0.09	2.20 ± 0.05	1.02	0.35	4.07	0.01	0.18	0.9965	150 ± 30
KDE σ -based	2.02 ± 0.09	2.21 ± 0.06	0.23	0.82	3.46	0.01	0.02	0.9966	148 ± 32

TABLE I: Comparison of linear model simulations' statistics. The columns slope and intercept show the best-fitting linear model parameters to a particular lensed dataset. The t -labeled columns show their respective t -statistics, while P -labeled columns show the corresponding 2-sided P -values. The F statistic and its corresponding P -value was computed to compare the best-fitting values to the best-fitting values of the bootstraps. The median and standard deviation of a particular lensing method's computation time were estimated from bootstrap simulations' computation times.

to linearize the relationship between \mathbf{y}_1 and $f_M(\mathbf{y}_1)$ as $f_M(\mathbf{y}_1) = S\mathbf{y}_1$, resulting in the following linearized estimates for the mean *a posteriori* value \mathbf{y}_1 , and the corresponding *a posteriori* covariance matrix A_1 :

$$\mathbf{y}_1 = \mathbf{y}_0 + A_0 S^T (S A_0 S^T + B)^{-1} (\mathbf{y}_m - f_M(\mathbf{y}_0)) \quad (8)$$

$$A_1 = A_0 - A_0 S^T (S A_0 S^T + B)^{-1} S A_0. \quad (9)$$

The matrix A_1 contains correlations between all experimental data, which can then be used in the fitting process.

B. CoH₃ optical model fitting via LM algorithm

The CoH₃ code computes nuclear reaction cross sections in the fast energy range, including total, shape elastic, and direct inelastic scattering, as well as direct/semidirect capture, pre-equilibrium emission, and

particle and γ -ray emission processes [11]. Importantly, CoH₃ computes particle transmission coefficients internally and does not rely on external optical model solvers. For deformed nuclei, the code employs a coupled-channels formalism extended with rotational and vibrational models. Throughout this work, we use the global Koning–Delaroche optical potential [22], as implemented in CoH₃, for all neutron-induced reaction calculations.

To determine the effectiveness of our data thinning method, we employed a standard damped Levenberg–Marquardt (LM) algorithm [23] to optimize multiplicative scaling factors (tweaks) of the six optical model parameters of the neutron Koning–Delaroche optical potential [22]. These scaling factors tweak the real potential depth, t_{OV} , the imaginary surface potential depth, t_{OW} , the real potential radius, t_{ORV} , the imaginary surface potential radius, t_{ORW} , the real potential diffuseness, t_{OAV} , and the imaginary surface potential diffuseness, t_{OAW} .

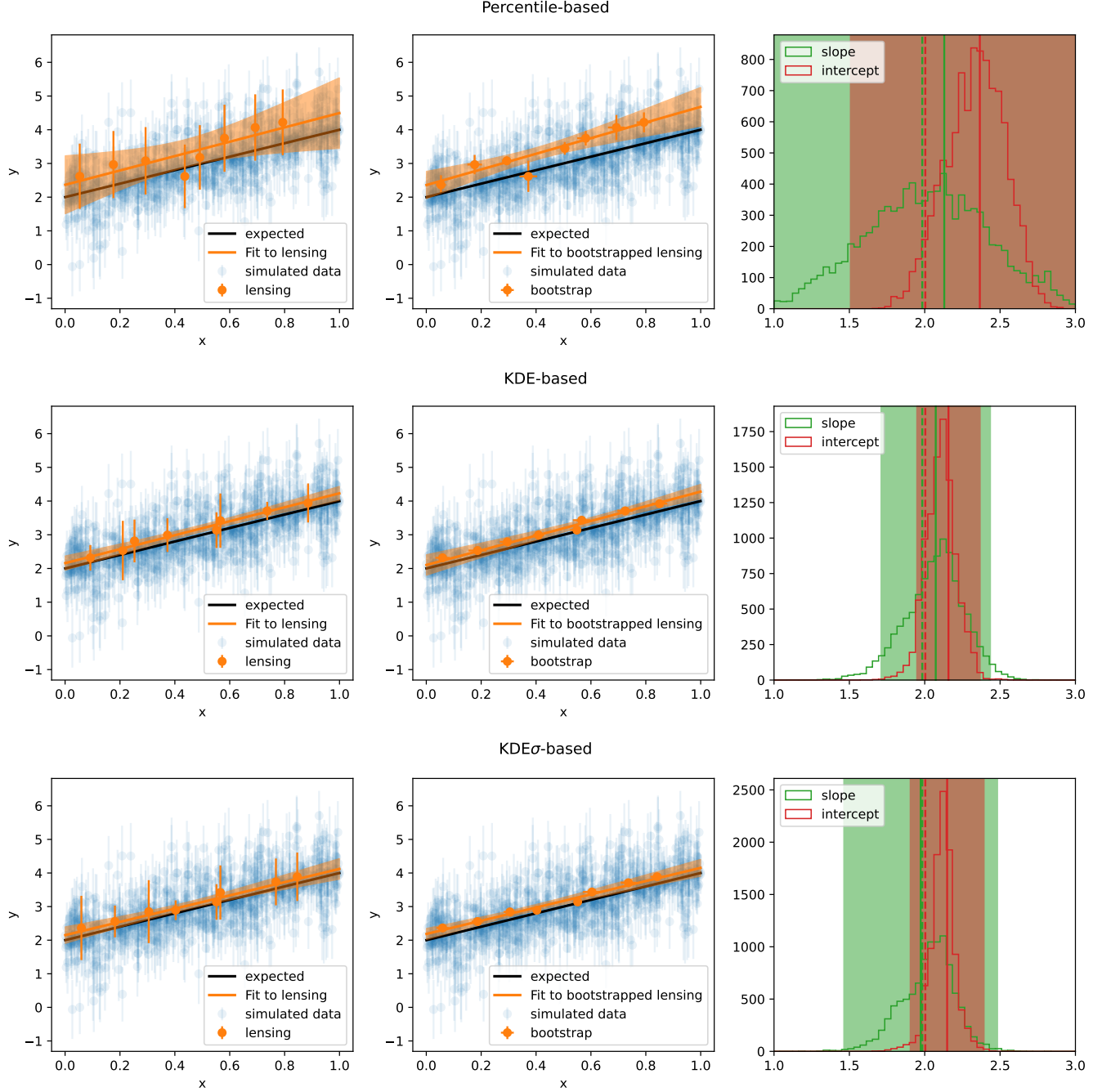


FIG. 2: Lensing of linear model simulations (blue error bars) using the percentile-based (top row), KDE-based (middle row) and KDE σ -based methods (bottom row). Orange points show the result of the lensing procedure (left panels) and their bootstrapped estimates (middle panels). The right panels show the histograms of the slope (green) and intercept (red) of a linear model fitted to the bootstrap samples as well as the vertical lines marking the prediction intervals, computed from the histograms, where dashed lines are for the best fits and shaded intervals for the prediction intervals. Solid lines represent the non-bootstrapped best-fits.

A schematic overview of the fitting procedure is provided in Fig. 1. The fits were performed with experimental total cross-section data, covering neutron energies from approximately 0.01 MeV to 10 MeV.

The LM algorithm is an iterative nonlinear least-

squares optimizer modulated by a varying damping parameter. At each iteration, a proposed parameter update is accepted or rejected based on the reduction of the objective function, defined here as the χ^2 between the experimental data and model predictions. The fit-

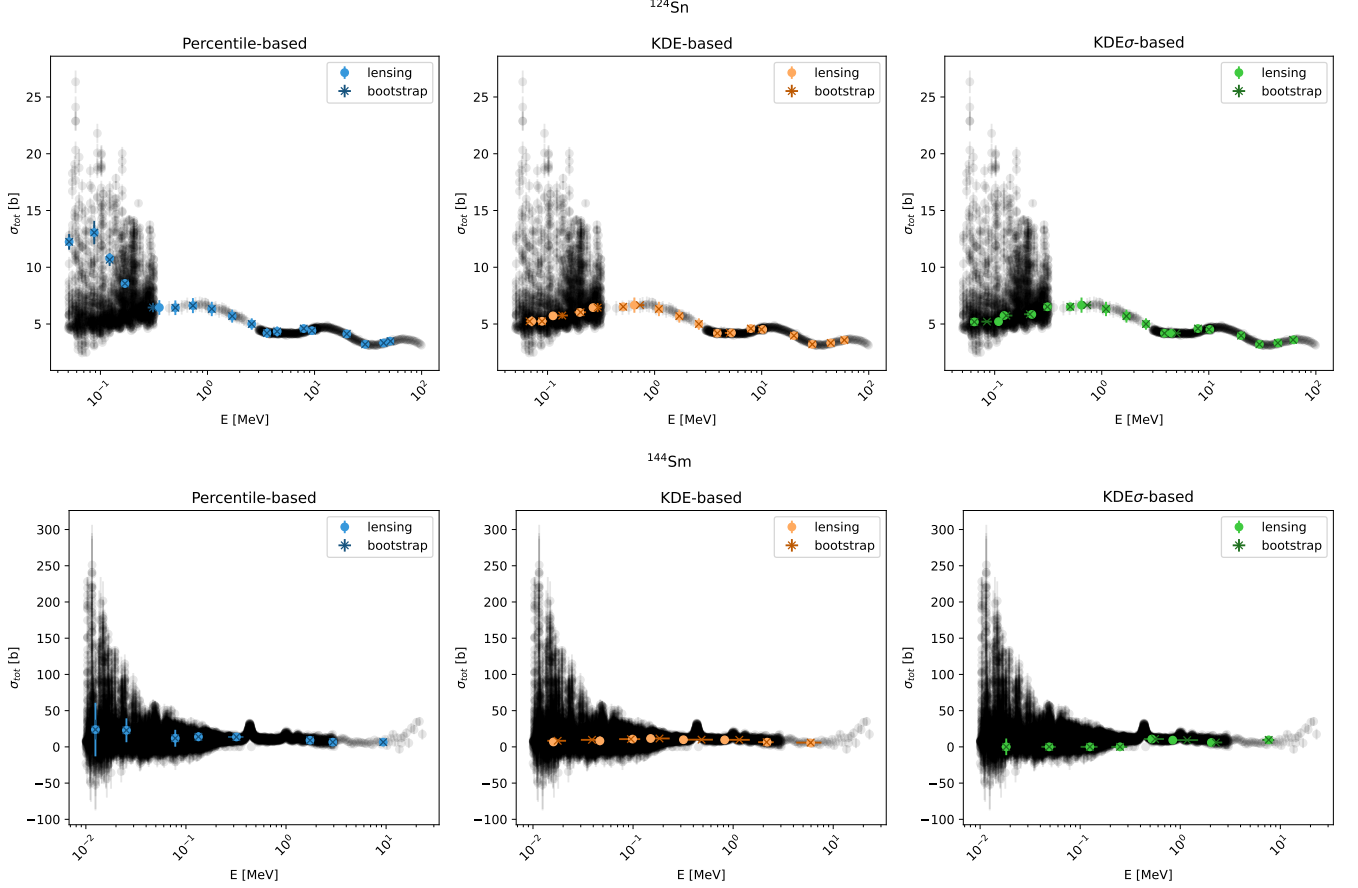


FIG. 3: Total cross section dataset and the lensed subset. The left panels display the percentile-based estimates, the middle panels show the KDE-based estimates, and the right panels present KDE estimates incorporating heteroscedastic errors. Black points represent the full dataset, while blue, orange, and green points indicate the lensed subsets. Crosses denote the bootstrap estimates for the lensing bin values. The data presented is total cross section for $n + {}^{124}\text{Sn}$ (top) and $n + {}^{144}\text{Sm}$ (bottom).

ting procedure uses the residuals $\mathbf{r}(\theta)$ between the FBET-smoothed data \mathbf{y}_1 and the model predictions $\mathbf{f}(\theta)$,

$$\mathbf{r}(\theta) = A_1^{-1/2} (\mathbf{y}_1 - \mathbf{f}(\theta)), \quad (10)$$

along with the corresponding model Jacobian matrix

$$J = \nabla_{\theta}^T \mathbf{r}(\theta) = -A_1^{-1/2} \nabla_{\theta}^T \mathbf{f}(\theta), \quad (11)$$

to compute the parameter update step $\delta\theta$ via the standard Gauss–Newton formula [23, 24]:

$$\delta\theta = -(J^T J)^{-1} J^T \mathbf{r}. \quad (12)$$

To quantify uncertainty in the fitted parameters, we estimate the parameter covariance matrix Σ_{θ} using the Cramér–Rao bound:

$$\Sigma_{\theta} = (J^T J)^{-1}. \quad (13)$$

This matrix is then propagated to the model output to obtain the cross section covariance matrix,

$$\Sigma_{\mathbf{y}} = \nabla_{\theta}^T \mathbf{f}(\theta) \Sigma_{\theta} \nabla_{\theta} \mathbf{f}^T(\theta). \quad (14)$$

The cross section uncertainties on the predicted cross sections, $\sigma_{\mathbf{y}}$, are defined as the square roots of the diagonal entries of $\Sigma_{\mathbf{y}}$ and represent one standard deviation around the model prediction. These are used to define confidence intervals on the energy grid:

$$[\mathbf{f}(\theta) - \sigma_{\mathbf{y}}(\theta), \mathbf{f}(\theta) + \sigma_{\mathbf{y}}(\theta)]. \quad (15)$$

C. Bootstrap

To assess the variability and robustness of our estimates, we employ the bootstrap method [25], a non-parametric resampling technique that approximates the sampling distribution of a statistic by repeatedly drawing samples with replacement from the observed data. This approach enables estimation of standard errors, confidence intervals, and the stability of fitted parameters in both the percentile- and kernel-density-based methods.

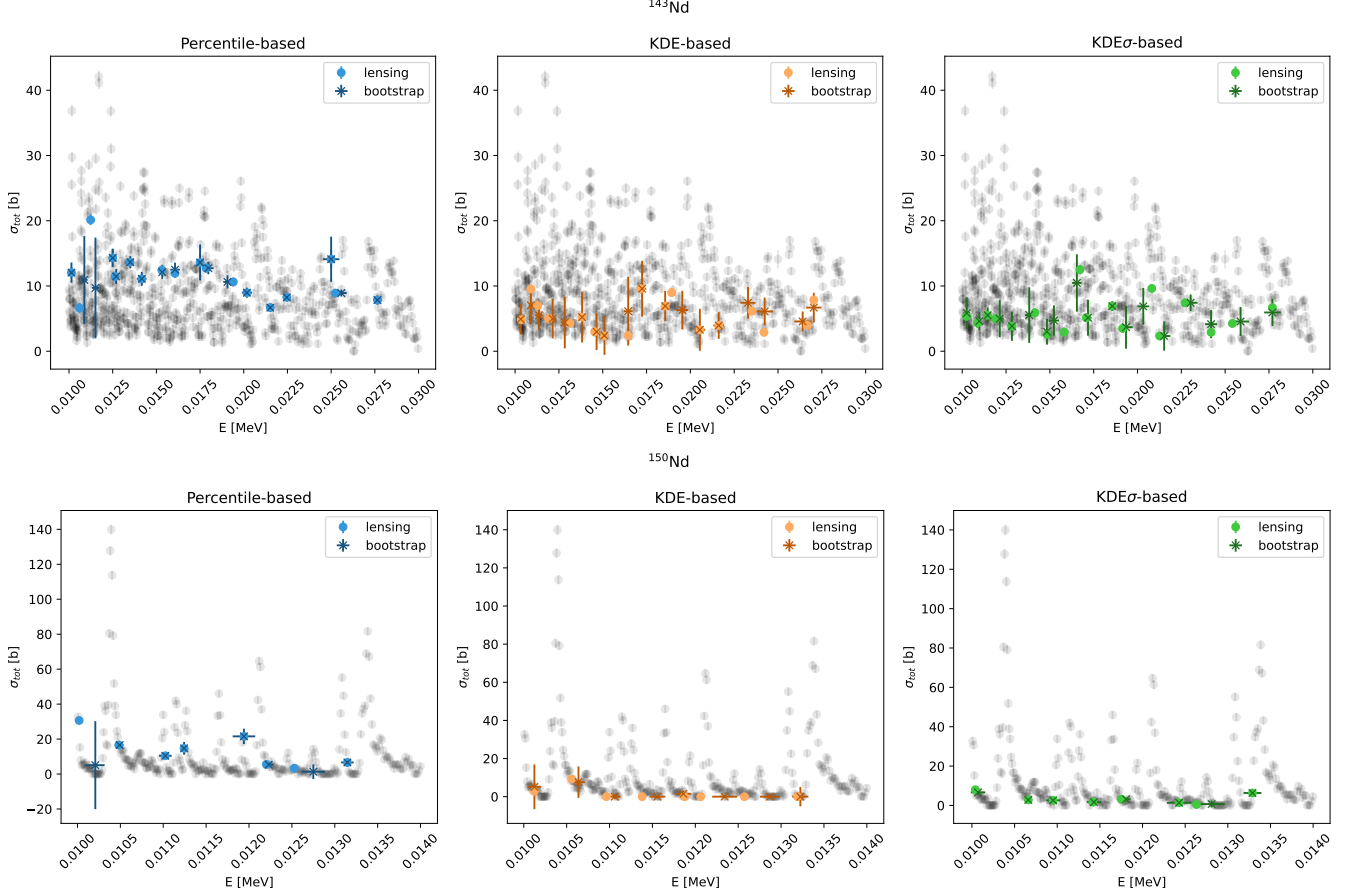


FIG. 4: Same as Fig. 3, the data presented is total cross section dataset and the lensed subset for $n + {}^{143}\text{Nd}$ (top) and $n + {}^{150}\text{Nd}$ (bottom). The left panels display the percentile-based estimates, the middle panels show the KDE-based estimates, and the right panels present KDE estimates incorporating heteroscedastic errors. Black points represent the full dataset, while blue, orange, and green points indicate the lensed subsets. Crosses denote the bootstrap estimates for the lensing bin values.

D. Classical thinning

The thinning through estimation method, as described in [21], iteratively reduces a dataset by removing the least informative points with respect to a chosen estimator. The procedure is designed to preserve predictive power while reducing redundancy in large spatial or multidimensional datasets. The method starts with a dataset of N pairs of values, (x_i, y_i) indexed by $i \in P_0 = \{1, \dots, N\}$, and a kernel regression estimator $\hat{y}(x, S)$, that takes a subset $S \subset P_0$ and evaluates a chosen kernel K_h for a particular point, x :

$$\hat{y}(x, S) = \frac{\sum_{j \in S} K_h(x, x_j) y_j}{\sum_{j \in S} K_h(x, x_j)}, \quad (16)$$

where K_h is the Gaussian kernel symmetric kernel function with bandwidth h :

$$K_h(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2h^2}\right). \quad (17)$$

For each point i the associated mean squared error, $e(x, S)$ is computed

$$e(x_i, S) = (y_i - \hat{y}(x_i, S))^2. \quad (18)$$

The algorithm computes $e(x_i, P_0 \setminus \{i\})$ for all $i \in P_0$ and proceeds by ranking the points based on $e(x_i, P_0 \setminus \{i\})$ for all $i \in P_0$ and removes the point with the smallest error, yielding a subset P_1 . The procedure is then iterated, for a step k the expression can be written as

$$P_{k+1} = P_k \setminus \arg \min_{i \in P_k} (e(x_i, P_k \setminus \{i\})) \quad (19)$$

This process continues until a desired target mean squared error is reached. The resulting reduced dataset

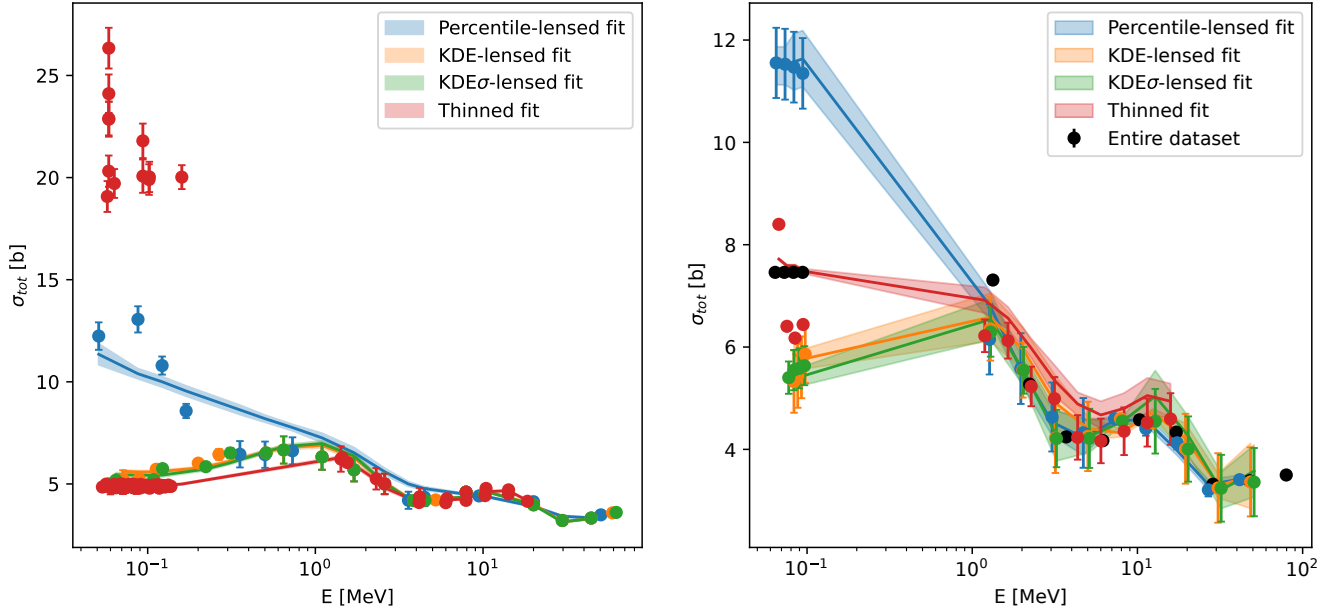


FIG. 5: The reduced ^{124}Sn datasets and the corresponding CoH₃ model fits for the total cross section channel. Left panel shows the lensed and thinned data for ^{124}Sn , while the right panel shows the FBET-smoothed lensed data and the FBET-smoothed thinned data. The black dots show the FBET smoothing applied to the entire dataset.

P_k retains the key predictive characteristics of P_0 while substantially reducing its size.

IV. TESTING ON SIMULATED DATA

We demonstrate our method on simulated data taken from a normal distribution, following a linear model with heteroscedastic, uncorrelated errors. To this end, we constructed a random sample of a linear model with set values of slope and intercept, computed the fit of a linear model to the lensed subsample, and estimated the reliability of these fits using bootstrap sampling of the simulated dataset.

The data are simulated using uniform distributions for a variable, x , and its standard deviation, σ . Samples of a dependent variable, y , are then drawn from a normal distribution.

$$x \sim U(0, 1) \quad (20)$$

$$\sigma \sim U(.01, 1) \quad (21)$$

$$y(x, \sigma) \sim \mathcal{N}(2x + 2, \sigma^2 \mid x, \sigma). \quad (22)$$

For this example a simulated sample size of $N_s = 1000$.

The results of fitting a linear model to the lensed subsets of the simulated data are shown in the left panels of Fig. 2. In Table I, we show the comparison of the lensing methods based on the t-tests on the slope and intercept, computed with respect to the expected slope and intercept values, using the simulated relation $y = 2x + 2$.

To evaluate the stability and reliability of the lensing procedure, we conducted bootstrap subsampling $N_B = 10000$ times. In the middle panels of Fig. 2, we show the bootstrapped values - the median and standard deviation of the bootstrap subsamples for each x -bin. The orange line is a fit to the orange points.

We then fit each bootstrap of the lensed subsample with a linear model and compare the resulting slope and intercept distributions in Fig. 2. We find that the prediction intervals ($\sqrt{N_s + 1} \times$ estimated parameter error of the simulated sample, and $\sqrt{N_B + 1} \times$ estimated parameter error of the lensed subsample), shown as vertical lines, agree well with the bootstrap distributions.

The Chow F statistics [26] and their corresponding P -values were computed to compare the best-fitting model's evaluation to the best-fitting evaluations of the bootstraps. In brief, the method computes the F -statistic based on the sum of squares of the residuals of the best-fitting model's residual sum of squares, RSS_m , the bootstrap best-fitting model's residual sum of squares RSS_b and the residual sum of squares of a union of the lensing dataset and the bootstrapped lensing dataset, RSS_u as

$$F = \frac{(RSS_u - (RSS_m + RSS_b))/k}{(RSS_m + RSS_b)/(n_m + n_b - 2k)}, \quad (23)$$

where k is the number of model parameters, i.e. $k = 2$, and n_m and n_b are the sizes of the lensed dataset and its bootstrap, respectively. We list the Chow F statistic values and the corresponding p -values (using the $F(k, n_m + n_b - 2k)$ distribution) in table I. We find that there is no statistically significant difference between the

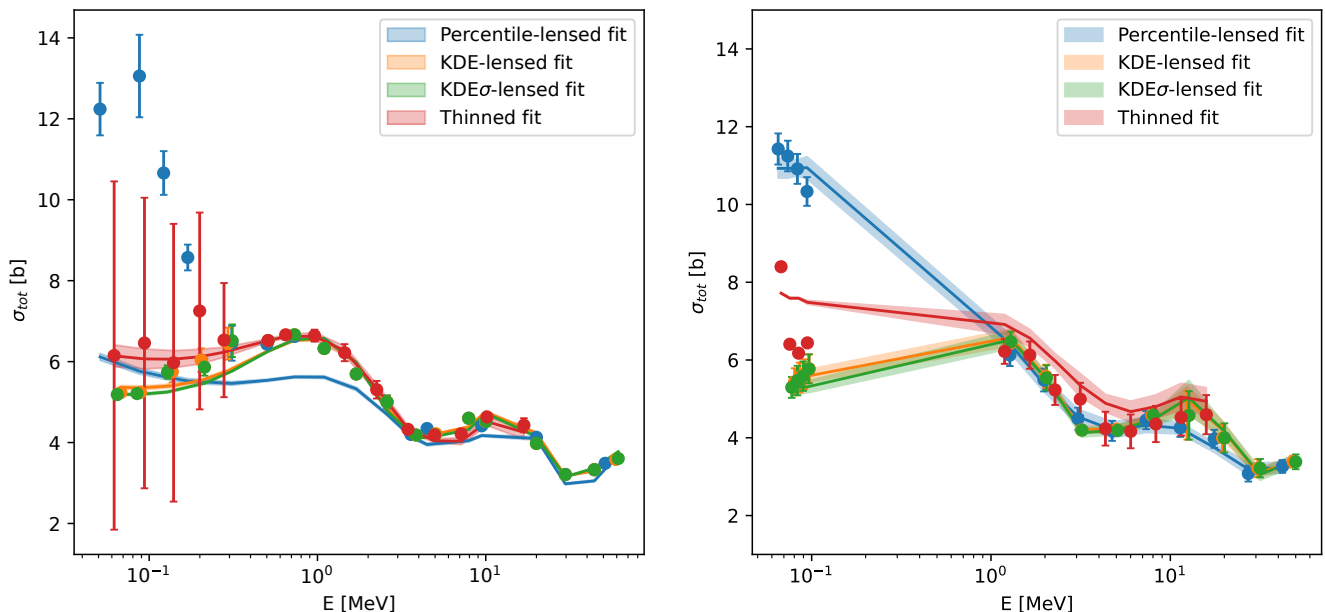


FIG. 6: The bootstrapped and reduced ^{124}Sn datasets and the corresponding CoH₃ model fits for the total cross section channel. Left panel shows the lensed and thinned data for ^{124}Sn , while the right panel shows the FBET-smoothed lensed data and the FBET-smoothed thinned data.

lensed dataset and its bootstrap ($P > 0.1$). The method therefore gives reliable results for all three lensing variants.

Computation times for the lensing methods are shown in Table I, reported as the median and standard deviation over bootstrap simulations. We find that the percentile-based lensing method is significantly faster than the KDE-based approach. For comparison, the classical thinning method required over 5 minutes to process the same dataset, due to the higher computational cost at each iteration.

V. RESULTS

We demonstrate our method on total cross section data for four neutron-induced reactions— $n+^{124}\text{Sn}$ [27–32], $n+^{144}\text{Sm}$ [33, 34], $n+^{143}\text{Nd}$ [35, 36], and $n+^{150}\text{Nd}$ [35]—chosen to provide a diverse sample that includes both resonance-rich and smooth, non-resonant measurements. These results are intended to showcase the efficacy of our data thinning approach; a comprehensive multi-channel evaluation [37] and extensions for highly unbalanced datasets [38] will be presented in follow-up work.

Figs. 3 and 4 display the full datasets alongside the subsets extracted using our lensing algorithm. Left panels show percentile-based estimates, middle panels show KDE results, and right panels depict KDE estimates with heteroscedastic error treatment. In all panels, black points denote the complete dataset; colored points (blue,

orange, green) represent the lensed subsets; and crosses indicate bootstrap-based estimates for the lensing-bin values.

The selected datasets span both narrow and wide energy ranges, allowing us to evaluate the robustness of the thinning procedure across different experimental regimes. In particular, ^{143}Nd and ^{150}Nd offer dense low-energy data, suitable for benchmarking statistical consistency, but lack high-energy measurements (above ~ 1 MeV) required for meaningful CoH₃ model fits. In contrast, ^{124}Sn and ^{144}Sm datasets extend into the fast neutron regime and enable full model optimization. To illustrate this difference, we perform fits for ^{124}Sn (Sec. V).

Our results indicate that the lensing method tends to exclude points within pronounced resonance peaks, especially in the ^{124}Sn and ^{144}Sm cases, thereby favoring smoother samples more representative of the average cross section. However, for reactions with limited energy ranges (notably ^{143}Nd and ^{150}Nd), the percentile-based bootstrap estimates occasionally show offset from the lensed subsets, likely due to uneven data distribution in the energy–cross-section plane. KDE-based estimates, which take local data density into account, help mitigate this bias.

Table II reports computation times for the lensing methods, including standard errors from 100 bootstrap simulations. The percentile-based variant is substantially faster than KDE-based lensing. All of the lensing methods are faster than the thinning method. The number of bootstrap samples was chosen to strike a balance between reliable uncertainty estimation and total compu-

Lensing method	^{124}Sn	^{150}Nd	^{143}Nd	^{144}Sm
Percentile-based	0.0312 ± 0.0014	0.01268 ± 0.00091	0.0146 ± 0.0013	0.0121 ± 0.0023
KDE-based	0.237 ± 0.018	0.0898 ± 0.0078	0.1218 ± 0.0044	0.546 ± 0.013
KDE σ -based	0.196 ± 0.016	0.0715 ± 0.0064	0.1167 ± 0.0035	0.446 ± 0.015
Thinning	1500 ± 390	9.0 ± 1.1	133.5 ± 7.0	1.98×10^{5a}

^a Evaluation for the ^{144}Sm thinning method was omitted due to long runtimes.

TABLE II: Comparison of computation times (in seconds) for the lensing methods and the thinning method across the selected nuclei. Uncertainties were estimated using 100 bootstrap samples.

tation times for bootstrap simulations.

Application to model fitting to $n + ^{124}\text{Sn}$

To further assess the impact of our data reduction techniques, we applied the FBET data assimilation approach to the $n + ^{124}\text{Sn}$ lensed subset (see Fig. 3), followed by a fit of a 6-parameter CoH₃ model. Figure 5 contrasts the FBET-smoothed CoH₃ fit from the lensed dataset (blue points) with the outcome of the standard data thinning method, which was applied to the full initial dataset before undergoing FBET processing. The comparison reveals that the lensing approach yields estimates with closer agreement to the FBET-smoothed expectations than does the traditional thinning procedure. Notably, while the standard thinning method selects points based on clustering of the most-distant (influential) measurements, our lensing algorithm selectively eliminates resonance-associated points, thereby enhancing the fidelity of the reduced dataset.

The left panel of Fig. 5 illustrates the raw lensed and thinned data prior to FBET smoothing. These results underscore that both thinning and lensing methods serve best as pre-processing steps in tandem with FBET for subsequent CoH₃ fits. In Fig. 5, we also present the results of applying the FBET procedure to the lensing and thinning reduced total cross section dataset. Notably, the thinning procedure (red points) exhibits a strong correspondence with the FBET-smoothed complete dataset. This agreement stems from the fact that thinning preferentially selects points that lie far from the mean—predominantly capturing the resonance peaks. By contrast, the lensing methods consider the overall distribution of data within an energy bin, thus weighting points closer to the mean cross section. This fundamental difference in data selection explains the observed discrepancies between the two approaches.

By condensing the dataset prior to applying FBET, our approach markedly decreases the computational burden traditionally associated with smoothing and parameter fitting, thereby expediting the overall analysis workflow and enabling more rapid exploration of parameter spaces in large-scale nuclear data applications.

In Fig. 6, we present the bootstrap estimates for the reduced datasets shown in Fig. 5 and fit these estimates

with the CoH₃ model. This analysis is not intended as an alternative to the lensing methods; rather, it illustrates that the bootstrap estimates derived from the lensing approaches yield results consistent with the expected lensed datasets, even after model fitting. These bootstrap estimates are not actual measurements, and both the bootstrap and smoothed bootstrap fits should be regarded as purely illustrative. This can be seen in the behavior of thinning bootstraps (red points in Fig. 6) whose medians, in contrast to the actual thinned subsets in Fig. 5, below 1 MeV are aligned with the lensing methods only because the thinning bootstrap estimates appear symmetrically above and below lensing bootstraps, with significantly larger uncertainties. We find that our method provides robust results since the FBET smoothed bootstraps in the right panel of Fig. 6 are in agreement with the non-bootstrapped smoothed data in the right panel of Fig. 5.

VI. CONCLUSION

We have presented a novel data thinning algorithm inspired by optical lensing, designed to improve the pre-processing of nuclear cross section data for subsequent statistical analysis. Our approach prioritizes the selection of structurally informative data points, particularly in resonance-rich regions, and is compatible with standard Bayesian assimilation workflows such as FBET.

By applying our method to both simulated and experimental nuclear data, we have shown that the optical lensing method offers a practical compromise between information retention and computational efficiency. In cases where full dataset processing is prohibitively expensive, our method allows analysts to work with a representative subset that yields statistically comparable outcomes.

While our main focus has been on nuclear cross sections, the generality of the approach suggests broader applicability to other fields that involve dense experimental datasets with correlated structures.

ACKNOWLEDGMENTS

This work was performed under the auspice of the U.S. Department of Energy by Los Alamos National Labora-

tory under 89233218CNA000001. Research reported in this publication was supported by the U.S. Department

of Energy LDRD program at Los Alamos National Laboratory (Project No. 20240004DR).

-
- [1] M. B. Chadwick *et al.*, Nuclear Data Sheets **112**, 2887 (2011).
 - [2] J. E. Escher, J. T. Burke, F. S. Dietrich, N. D. Scielzo, I. J. Thompson, and W. Younes, Reviews of Modern Physics **84**, 353 (2012).
 - [3] T. Kajino, G. J. Mathews, and W. Aoki, Progress in Particle and Nuclear Physics **107**, 109 (2019).
 - [4] M. R. Mumpower, R. Surman, G. C. McLaughlin, and A. Aprahamian, Progress in Particle and Nuclear Physics **86**, 86 (2016).
 - [5] K. Langanke and G. Martínez-Pinedo, Reviews of Modern Physics **75**, 819 (2003).
 - [6] M. Oertel, M. Hempel, T. Klähn, and S. Typel, Reviews of Modern Physics **89**, 015007 (2017).
 - [7] M. Herman, R. Capote, M. Sin, A. Trkov, B. Carlson, P. Obložinský, C. Mattoon, H. Wienke, S. Hoblit, Y. Cho, G. Nobre, V. Plujko, and V. Zerkin, Indc(nds)-0603, International Atomic Energy Agency (2013).
 - [8] A. J. Koning and D. Rochman, Nuclear Data Sheets **113**, 2841 (2012).
 - [9] O. Iwamoto, Journal of Nuclear Science and Technology **44**, 687 (2007).
 - [10] T. Kawano, P. Talou, M. B. Chadwick, and T. Watanabe, Journal of Nuclear Science and Technology **47**, 462 (2010).
 - [11] T. Kawano, arXiv e-prints, arXiv:1901.05641 (2019), arXiv:1901.05641 [nucl-th].
 - [12] N. Otuka, E. Dupont, V. Semkova, B. Pritychenko, A. I. Blokhin, M. Aikawa, S. Babykina, M. Bossant, G. Chen, S. Dunaeva, R. A. Forrest, T. Fukahori, N. Furutachi, S. Ganesan, Z. Ge, O. O. Gritzay, M. Herman, S. Hlavač, K. Katō, B. Lalremruata, Y. O. Lee, A. Makinaga, K. Matsumoto, M. Mikhaylyukova, G. Pikulina, V. G. Pronyaev, A. Saxena, O. Schwerer, S. P. Simakov, N. Soppera, R. Suzuki, S. Takács, X. Tao, S. Taova, F. Tárkányi, V. V. Varlamov, J. Wang, S. C. Yang, V. Zerkin, and Y. Zhuang, Nuclear Data Sheets **120**, 272 (2014), arXiv:2002.07114 [nucl-ex].
 - [13] D. Neudecker, P. Talou, T. Kawano, and F. Tovesson, Nuclear Data Sheets **118**, 353 (2014).
 - [14] H. Leeb, D. Neudecker, and T. Srdinko, Nuclear Data Sheets **109**, 2762 (2008).
 - [15] H. Leeb, S. Gundacker, D. Neudecker, T. Srdinko, and V. Wildpaner, Journal of Korean Physical Society **59**, 959 (2011).
 - [16] D. Neudecker, R. Capote, and H. Leeb, Nuclear Instruments and Methods in Physics Research A **723**, 163 (2013).
 - [17] L. M. Stewart, S. L. Dance, and N. K. Nichols, Tellus A: Dynamic Meteorology and Oceanography **65**, 19546 (2013).
 - [18] C. Cardinali, Observation influence diagnostic of a data assimilation system, in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)*, edited by S. K. Park and L. Xu (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013) pp. 89–110.
 - [19] S. Cheng, J.-P. Argaud, B. Iooss, D. Lucor, and A. Ponçot, Stochastic Environmental Research and Risk Assessment **33**, 2033 (2019).
 - [20] T. Ochotta, C. Gebhardt, D. Saupe, and W. Wergen, Quarterly Journal of the Royal Meteorological Society **131**, 3427 (2005).
 - [21] T. Ochotta, C. Gebhardt, D. Saupe, and W. Wergen, Quarterly Journal of the Royal Meteorological Society **131**, 3427 (2005).
 - [22] A. J. Koning and J. P. Delaroche, Nuclear Physics A **713**, 231 (2003).
 - [23] D. W. Marquardt, Journal of the society for Industrial and Applied Mathematics **11**, 431 (1963).
 - [24] K. Levenberg, Quarterly of Applied Mathematics **2**, 164 (1944).
 - [25] B. Efron, Annals of Statistics **7**, 1 (1979).
 - [26] G. C. Chow, Econometrica **28**, 591 (1960).
 - [27] R. F. Carlton, J. A. Harvey, and N. W. Hill, Phys. Rev. C **54**, 2445 (1996).
 - [28] R. W. Harper, T. W. Godfrey, and J. L. Weil, Phys. Rev. C **26**, 1432 (1982).
 - [29] R. Musaelyan, V. Skorkin, I. Surkova, N. Sklyar, V. Ovdienko, M. Fedorov, and T. Yakovenko, Yad.Fiz. **50** (1989).
 - [30] C. D. Pruitt, R. J. Charity, L. G. Sobotka, J. M. Elson, D. E. M. Hoff, K. W. Brown, M. C. Atkinson, W. H. Dickhoff, H. Y. Lee, M. Devlin, N. Fotiades, and S. Mosby, Phys. Rev. C **102**, 034601 (2020), arXiv:2006.00024 [nucl-ex].
 - [31] J. Rapaport, M. Mirzaa, H. Hadizadeh, D. E. Bainum, and R. W. Finlay, Nuclear Physics A **341**, 56 (1980).
 - [32] Y. V. Dukarevich, A. N. Dyumin, and D. M. Kaminker, Nuclear Physics A **92**, 433 (1967).
 - [33] R. L. Macklin, N. W. Hill, J. A. Harvey, and G. L. Tweed, Phys. Rev. C **48**, 1120 (1993).
 - [34] A. N. Dyumin, A. I. Egorov, G. N. Popova, and V. A. Smolin, Bulletin of the Russian Academy of Sciences: Physics **37**, 91 (1973).
 - [35] H. Tellier, *Properties of levels induced in stable isotopes of neodymium by neutrons in the resonance region*, Technical Report Note No. 1459 (Centre d’Études Nucléaires, 1971).
 - [36] K. Wisshak, F. Voss, F. Käppeler, L. Kazakov, and G. Reffo, Phys. Rev. C **57**, 391 (1998).
 - [37] M. Imbrišak (), in prep.
 - [38] M. Imbrišak (), in prep.