Combining Total Monte Carlo and Unified Monte Carlo: Bayesian nuclear data uncertainty quantification from auto-generated experimental covariances

P. Helgesson[a,b,*], H. Sjöstrand[a], A.J. Koning[b,a], J. Rydén[c], D. Rochman[d], E. Alhassan[a], S. Pomp[a]

[a] *Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden*
[b] *Nuclear Research and Consultancy Group NRG, Petten, The Netherlands*
[c] *Department of Mathematics, Uppsala University, Uppsala, Sweden*
[d] *Paul Scherrer Institute PSI, Villigen, Switzerland*

Abstract

The Total Monte Carlo methodology (TMC) for nuclear data (ND) uncertainty propagation has been subject to some critique because the nuclear reaction parameters are sampled from distributions which have not been rigorously determined from experimental data. In this study, it is thoroughly explained how TMC and Unified Monte Carlo-B (UMC-B) are combined to include experimental data in TMC. Random ND files are weighted with likelihood function values computed by comparing the ND files to experimental data, using experimental covariance matrices generated from information in the experimental database EXFOR and a set of simple rules. A proof that such weights give a consistent implementation of Bayes' theorem is provided. The impact of the weights is mainly studied for a set of integral systems/applications, *e.g.*, a set of shielding fuel assemblies which shall prevent aging of the pressure vessels of the Swedish nuclear reactors Ringhals 3 and 4.

In this implementation, the impact from the weighting is small for many of the applications. In some cases, this can be explained by the fact that the distributions used as priors are too narrow to be valid as such. Another possible explanation is that the integral systems are highly sensitive to resonance parameters, which effectively are not treated in this work. In other cases, only a very small number of files get significantly large weights, *i.e.*, the region of interest is poorly resolved. This convergence issue can be due to the parameter distributions used as priors or model defects, for example.

Further, some parameters used in the rules for the EXFOR interpretation have been varied. The observed impact from varying one parameter at a time is not very strong. This can partially be due to the general insensitivity to the weights seen for many applications, and there can be strong interaction effects. The automatic treatment of outliers has a quite large impact, however.

To approach more justified ND uncertainties, the rules for the EXFOR interpretation shall be further discussed and developed, in particular the rules for rejecting outliers, and random ND files that are intended to describe prior distributions shall be generated. Further, model defects need to be treated.

Keywords: Nuclear data, Uncertainty Propagation, Total Monte Carlo, Experimental Correlations, Unified Monte Carlo

1. Introduction and Total Monte Carlo

Nuclear data (ND) underpins all nuclear science and technology [1], and its accuracy is hence paramount. As for any scientific quantity, ND (and results derived from it) should be presented with both best estimates and with uncertainties.

Total Monte Carlo (TMC, [2]) is an ND uncertainty propagation method based on the idea of sampling nuclear reaction model parameters to a nuclear reaction code, typically the TALYS code system T6 [3], which complements TALYS [4] results with, *e.g.*, resonance data.

An overview of the methodology is seen in Fig. 1. By feeding T6 with n randomly sampled sets of parameters, n ENDF [5] libraries with ND are produced, referred to as random files in this text. By using each such random file in a simulation of a nuclear system of interest, n results for all output quantities are obtained. The output quantities could, *e.g.*, be grouped macroscopic cross sections, power distribution, k_{eff}, decay heat, dose rate, inventories, *etc.*, to mention a few possible examples. If the ND which is varied has any relevance to the system, the results will have a spread because of the varying ND. Using statistical inference one can then estimate the propagated ND uncertainty in any of the output quantities.

TMC has been applied to numerous different cases, ranging from shielding models [6], thoroughly studied pin

*petter.helgesson@physics.uu.se, Dpt. of Phys. and Astronomy, Box 516, 751 20 Uppsala, Sweden

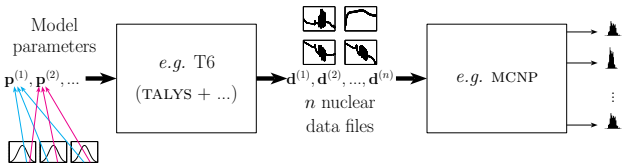

Figure 1: An overview of the TMC methodology.

cells [7] and a wide range of criticality safety benchmarks [8] to full core neutronics simulations [9] and even to models including thermo-hydraulics [10] and transients [11].

The methodology has a number of advantages compared to the conventional use of covariance matrices and sensitivities to propagate ND uncertainties; for example, non-Gaussian output distributions can be observed (examples of which can be seen in Refs. [2, 12]), it allows for non-linearities and also for more complete input distributions than simply central values and covariances. Another important advantage is the transparency compared to how the covariance data in the ND evaluations are produced. Finally, there is no need to process covariance matrices and to keep track of them in all codes in the entire chain of simulations.

One of the major drawbacks of TMC is the computational cost, which typically is most significant if Monte Carlo transport codes are used. Probably, a more important drawback of TMC of today is the determination of the distributions from which the model parameters are sampled, even though one can argue that conventional methods suffer from similar problems. Since the nuclear model parameters are limited by experiments, the experiments and their uncertainties should determine the distributions of these parameters. Substantial work *has* been done to adjust the distributions to be consistent with the experiments, but unfortunately without a rigorous statistical analysis, which has been pointed out in, *e.g.*, Refs. [13, 14]. This introduces some arbitrariness in these distributions, an arbitrariness which clearly propagates to the applications.

Many options for how to solve this for TMC or related methodologies have been discussed, for example Backward-Forward Monte Carlo [15] and Unified Monte Carlo (UMC-G [16] and UMC-B [17]). The use of "TMC+UMC" was first suggested in [16] to adress shortcomings of TMC, in particular the dismissal of experimental correlations, but the method was not implemented nor tested. A first attempt to combine TMC and UMC was made in Ref. [14] for evaluation of Prompt Fission Neutron Spectra, but the methodology was not clearly specified. A Bayesian Monte Carlo method with similarities to Ref. [15] was presented in Ref. [18]. In contrast to this work, both these methods explicitly estimate the posterior distribution of model parameters, and use certain approximations to do this. Also, different expressions for the likelihood function are used, and experimental correlations have not been taken into account in practice.

In the current authors' contribution to the International Workshop on Nuclear Data Covariances 2014 (CW-2014) [19], the problem of including experimental information in TMC was tackled similarly to what was suggested as an "augmented TMC" in the presentation of UMC-B [17]. That is, Bayes' theorem was implemented through the assignment of weights to random ND files, which were used to calibrate the central values and the ND uncertainties accounting for experimental information in a set of applications. It was observed that the method could have a significant impact on the ND uncertainties and possibly an even larger effect on the central values. It was, however, concluded that some improvements to the methodology were necessary. For example, it is important to generate random files that are intended to be "prior" files, *i.e.*, that they do not already include experimental information in such a way that significantly influences the results – this would lead to an inappropriate double counting of the experimental information and the original distribution would still have a strong impact on the results. Further, there were a number of arbitrary choices made in Ref. [19], for example, a 1 % uncertainty which was fully correlated over each reaction channel was added to all experimental uncertainties.

This study is an extension of the work in Ref. [19] where both the methodology and the results are described and discussed more carefully, and a few more integral systems are added. The perhaps most important supplement is that the sensitivities to the arbitrary choices made in Ref. [19] are analyzed by a variation of these parameters. The work primarily aims to further illustrate the methodology and to suggest where to focus its future development.

Sec. 2 covers the methodology, including a thorough explanation of the file weights and how experiments from the EXFOR database [20] are selected and interpreted to construct experimental covariance matrices. In Sec. 3.1, the integral systems studied in this work are briefly described and in Sec. 3.2 it is described how the sensitivity to arbitrary choices in the EXFOR interpretation is investigated. Results, with emphasis on the integral systems, follow in Sec. 4. Finally, discussion and conclusions are found in Secs. 5 and 6, respectively, followed by a rather extensive appendix with mathematical details and additional results. In particular, Appendix A includes a formal proof that the use of file weighs proportional to the likelihood, as implemented in UMC-B and in this paper, is a consistent implementation of Bayes theorem.

2. Methodology – including experimental data in TMC

2.1. Using file weights

The basic principle behind the Total Monte Carlo [2, 3, 21] methodology for propagation of nuclear data uncertainties is easy to understand. Instead of using one

nuclear data (ND) file for some nuclide, input parameters $\mathbf{p} = (p_1, p_2, ..., p_j, ..., p_N)^{\mathrm{T}}$ to a nuclear reaction code such as the TALYS code system [3, 22] are randomly sampled n times. This gives n ND files, each of which can be used in, *e.g.*, the computation of some quantity q in a nuclear system using a code such as MCNP [23]. Thus, one obtains n results for q which may be different due to the use of different ND files. The spread in the results can be used to quantify the uncertainty in q due to the considered ND, *cf.* Sec. 1 and Fig. 1. It is easy to implement TMC to assess ND uncertainties from various types of ND and a wide range of nuclides using freely available random ND files from the TENDL [3] homepage.

The spread of the investigated quantity q will depend strongly on which distribution the parameters $\mathbf{p}$ are sampled from. This distribution is quantified by a probability density function (PDF) $f(\mathbf{p})$. The parameters may in general be correlated to each other such that $f(\mathbf{p})$ should be multivariate instead of a product of N univariate distributions $f_j(p_j)$. The determination of $f(\mathbf{p})$ is difficult, since it is meant to quantify something as abstract as the current knowledge of the parameters. For the random ND files available at the TENDL homepage, the parameters have been sampled from distributions that are intended to quantify this knowledge, unfortunately without a strong theoretical foundation, but rather based on practical experience.

The nuclear reaction models that involve the parameters $\mathbf{p}$ are motivated by theory, but the actual values of the parameters are determined mainly from experiments. Therefore, the knowledge from experiments is regarded to dominate $f(\mathbf{p})$ in this work. In Sec. 2.2, it is described how Bayes' theorem is used to include experimental information into $f(\mathbf{p})$. In practice, for $k \in \{1, 2, ..., n\}$, the k'th ND file is given a weight which is proportional to the likelihood function

$$L\left(\mathbf{p}^{(k)}; \mathbf{x}\right) \propto e^{-\chi_k^2/2}, \tag{1}$$

where "the generalized χ^2",

$$\chi_k^2 = \left(\mathbf{x} - \boldsymbol{\tau}^{(k)}\right)^{\mathrm{T}} \mathbf{C}_{\mathrm{E}}^{-1} \left(\mathbf{x} - \boldsymbol{\tau}^{(k)}\right), \tag{2}$$

is a measure of the agreement between the considered experimental points $\mathbf{x} = (x_1, x_2, ..., x_m)^{\mathrm{T}}$ and the corresponding values in the k'th random file, $\boldsymbol{\tau}^{(k)} = \left(\tau_1^{(k)}, \tau_2^{(k)}, ..., \tau_m^{(k)}\right)^{\mathrm{T}} = \boldsymbol{\tau}\left(\mathbf{p}^{(k)}\right)$ (obtained with the k'th parameter set $\mathbf{p}^{(k)}$). The matrix $\mathbf{C}_{\mathrm{E}}$ is the "experimental covariance matrix", describing the uncertainties for the experimental points and also the correlations between them, arising from systematic uncertainties such as uncertainty in sample thicknesses, uncertainty in normalizing cross sections, *etc.* To more easily understand Eq. (2), one can note that if the experiments *would be* independent (so that $\mathbf{C}_{\mathrm{E}}$ is diagonal with $(\mathbf{C}_{\mathrm{E}})_{ii} = \sigma_i^2$ being the variance of the i'th experimental point),

$$\chi_k^2 = \sum_{i=1}^{m} \frac{\left(x_i - \tau_i^{(k)}\right)^2}{\sigma_i^2}, \tag{3}$$

which many people recognize as "the χ^2".

After a more thorough description of the background and details of Eqs. (1) and (2) in Sec. 2.2, the estimation of the experimental covariance matrix $\mathbf{C}_{\mathrm{E}}$ is discussed in Secs. 2.3 to 2.5.

2.2. Mathematical formalism

The experimental points $\mathbf{x} = (x_1, x_2, ..., x_m)^{\mathrm{T}}$ are observations of corresponding random variables $\mathbf{X} = (X_1, X_2, ..., X_m)^{\mathrm{T}}$. That is, $\mathbf{X}$ has a distribution describing the probability for the outcomes of experiments with identical setups as the performed experiments, and $\mathbf{x}$ are the observations that happened to be realized. It is possible for $\mathbf{x}$ to span over different energies, different types of ND and different nuclides. For example, x_1 could be a measurement of the (n,2n) cross section at 14 MeV for ^{235}U while x_2 is a measurement of the fission neutron yield, $\bar{\nu}$, at 25 meV for ^{239}Pu. In this study, the experiments are restricted to cover cross sections and only one nuclide is considered at a time.

Now, assume that $f_0(\mathbf{p})$ is the PDF for $\mathbf{p}$ *excluding* the experimental information in $\mathbf{x}$ (the prior). Bayes' theorem states that the PDF for $\mathbf{p}$ given $\mathbf{x}$ is

$$f(\mathbf{p}|\mathbf{x}) \propto L(\mathbf{p}; \mathbf{x}) f_0(\mathbf{p}), \tag{4}$$

where $\propto$ means "proportional to" and the likelihood function $L(\mathbf{p}; \mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|\mathbf{p})$ is the PDF for $\mathbf{X}$ given $\mathbf{p}$, evaluated at $\mathbf{x}$. That is, assuming that the particular $\mathbf{p}$ is the "true" set of parameters, $L(\mathbf{p}; \mathbf{x})$ is the value of the PDF for $\mathbf{X}$ at $\mathbf{x}$.

Assume that $\mathbf{p}^{(k)}$ are sampled from $f_0(\mathbf{p})$ for $k \in \{1, 2, ..., n\}$. Given such a sample, one can (*e.g.* using T6) compute theoretical values $\boldsymbol{\tau}^{(k)} = \left(\tau_1^{(k)}, \tau_2^{(k)}, ..., \tau_m^{(k)}\right)^{\mathrm{T}}$ corresponding to each experimental value in $\mathbf{x}$. Assuming that $\mathbf{p}^{(k)}$ is *true* for $k \in \{1, 2, ..., n\}$, *i.e.* that the k'th set of parameters provides a perfect representation of the physics, and that the errors in the experiments (systematic and random) are symmetrically distributed around zero, the expected value of $\mathbf{X}$ is $\langle \mathbf{X} \rangle = \boldsymbol{\tau}(\mathbf{p}^{(k)})$. If $\mathbf{X}$ is normally distributed with covariance matrix $\mathbf{C}_{\mathrm{E}}$, it follows directly from the definition of a multivariate normal distribution that the PDF for $\mathbf{X}$ given $\mathbf{p}^{(k)}$, *i.e.* $L\left(\mathbf{p}^{(k)}; \mathbf{x}\right)$, becomes [24]

$$L\left(\mathbf{p}^{(k)}; \mathbf{x}\right) = f_{\mathbf{X}}\left(\mathbf{x}|\mathbf{p}^{(k)}\right) = \frac{e^{-\chi_k^2/2}}{(2\pi)^{m/2}\sqrt{\det\left(\mathbf{C}_{\mathrm{E}}\right)}}, \tag{5}$$

where χ_k^2 is defined in Eq. (2). Since the denominator is independent of $\mathbf{p}^{(k)}$, Eq. (1) follows. To assume that $\mathbf{X}$ is Gaussian is conventional and to some extent supported

by the central limit theorem [24] and by the principle of maximum entropy [25, 26].

In Appendix Appendix A, it is seen that with weights defined as

$$w_k := \frac{L(\mathbf{p}^{(k)}, \mathbf{x})}{\sum_{\kappa=1}^{n} L(\mathbf{p}^{(\kappa)}, \mathbf{x})}, \tag{6}$$

(":=" denotes that w_k is *defined* by this equation) and if q is some integral or differential quantity (*e.g.*, neutron multiplicity k_{eff} or a cross section at a particular energy), then

$$\widehat{q^j(\mathbf{p})} := \sum_{k=1}^{n} w_k q^j\left(\mathbf{p}^{(k)}\right) \tag{7}$$

is a consistent estimate of the j^{th} moment of q given $\mathbf{x}$, *i.e.* of $\left\langle q^j(\mathbf{p})\right\rangle_{f(\mathbf{p}|\mathbf{x})}$, where the subscript denotes that the expected value is taken with respect to the PDF $f(\mathbf{p}|\mathbf{x})$. Note that the *full* posterior PDF, including correlations and higher moments, is implicitly taken into account in contrast to, *e.g.*, Refs. [15] and [18]. In these cases, either a multivariate Gaussian or zero correlations are assumed for $\mathbf{p}$ (which may have practical advantages).

With $j = 1$, $\widehat{q^j(\mathbf{p})} = \widehat{q(\mathbf{p})}$ simply estimates the expected value of q, and using the estimates of both the first and second moment one can obtain an estimate for the variance of q, namely,

$$\sigma^2_{\text{observed}}(q) = \widehat{q^2(\mathbf{p})} - \left(\widehat{q(\mathbf{p})}\right)^2. \tag{8}$$

Since the estimates of both moments are consistent, Slutsky's theorem (Eq. (A.5)) yields that this is a consistent estimate of the variance.

In many cases, such as in this work, particle transport codes are Monte Carlo codes, giving rise to a random uncertainty in the results for q aside from the ND uncertainty. As discussed in Refs. [21, 7], the random variance can be estimated either by using the uncertainty estimates from the transport code or by performing a number of runs without varying any ND (in this work, the latter method is used to avoid the risk of a bias in the estimate of the statistical uncertainty in Monte Carlo transport codes pointed out in, *e.g.*, Ref. [27]). Denoting this estimate $\sigma^2_{\text{stat}}(q)$, one standard deviation due to ND uncertainty can be estimated using

$$\sigma_{\text{ND}}(q) = \sqrt{\sigma^2_{\text{observed}}(q) - \sigma^2_{\text{stat}}(q)}, \tag{9}$$

as is also discussed in Refs. [21, 7], in these cases without any weights in the definition of $\sigma^2_{\text{observed}}(q)$, however.

The uncertainty of the uncertainty itself, *i.e.* the uncertainty due to finite sample sizes of the estimates obtained using Eq. (9), is estimated by repeatedly dividing the observations into two groups as described in [7]. This estimate should be interpreted with care if only a small number of files have significantly large weights.

In this study, the distributions inherent in the files obtained from the TENDL homepage are used as prior distributions $f_0(\mathbf{p})$ for each considered nuclide, determining the distribution of the theoretical values $\boldsymbol{\tau}$. Since these files are intended to include the experimental information, the distributions are not appropriate priors. This should be kept in mind when interpreting the results, and in future studies other random files should be used.

2.3. Interpreting the EXFOR database

In this study, experimental cross section data for energies below and above the resonance range, taken from the EXFOR database [20] in both the original format and in the `xc4` format is used. The latter format is more convenient for computational applications but lacks some of the original information. Estimates of experimental covariance matrices are rarely available in EXFOR [28], and the *full* experimental covariance matrix $\mathbf{C}_{\text{E}}$ (*i.e.* including cross-experimental correlations), necessary for computing χ^2 using Eq. (2), is certainly not available. However, with information on random and systematic uncertainties, $\mathbf{C}_{\text{E}}$ can be estimated as described in Sec. 2.4. Some of this information can be found in EXFOR, but the information is not complete and the interpretation of the database is not straightforward. Therefore, the set of rules below are used to obtain estimates of random and systematic uncertainties. Observe that the rules are preliminary and quite coarse at this stage, and that they in future studies shall be refined with an extraction of more details from EXFOR and more justified assumptions.

1. EXFOR entries with a root-mean-square uncertainty of less than 0.1 % are, coarsely, considered unrealistic and are discarded.

2. If only one uncertainty estimate is given for each experimental point, the point is assigned with both a random and a systematic uncertainty of this magnitude. In this way, experiments with a less rigorous uncertainty treatment are conservatively penalized. If there are more than one uncertainty estimate per point given, uncertainties denoted `ERR-S` are interpreted as random, and all others are considered systematic. If no uncertainty denoted `ERR-S` is found, the uncertainty according to the `xc4` format is used as random.

3. The neutron energy in the i'th experimental point ($i \in \{1, 2, ..., m\}$) comes with uncertainties of two different kinds:

   (a) The "energy spectrum width" $\Delta E'_i$ of the neutron flux (typically of a neutron beam), giving an uncertainty in the energy of *each* incoming neutron, such that the measured cross section becomes an average weighted with the distribution of the neutrons.

   (b) The mean energy $E_{0,i}$ has an uncertainty $\Delta E_{0,i}$.

   In this study, it is assumed that energy resolutions quoted in the `xc4` formatted EXFOR files refer to

the first type, $\Delta E_i'$. If no energy resolution is given, it is assumed that $\Delta E_i' = 1\,\%$. Further, it is assumed that $\Delta E_{0,i} = 0.5\,\%$ for all experimental points. With these assumptions, $\Delta E_i'$ and $\Delta E_{0,i}$ are "translated" to uncertainties in the cross section of the random files $\tau_i^{(k)}$ ($k \in \{1, 2, ..., n\}$). This is done by assuming a normal distribution for the PDF for the individual neutron's deviation from the average energy $E_{0,i}$, written as $f_{E_i'}(E' - E_{0,i})$, and also for the PDF for $E_{0,i}$, denoted $f_{E_{0,i}}(E)$. The distributions have variances $(\Delta E_i')^2$ and $(\Delta E_{0,i})^2$ and expected values $\langle E_i' - E_{0,i}\rangle = 0$ and $\langle E_{0,i}\rangle = E_i$, respectively, where E_i is the quoted energy for the i'th experimental point.

If the pointwise theoretical value at energy E' is $\varsigma(E')$, the cross section one would observe at $E_{0,i}$, with an energy spectrum described by $f_{E_i'}(E' - E_{0,i})$, is

$$\varsigma_{\text{observed}}(E_{0,i}) = \int_{-\infty}^{\infty} f_{E_i'}(E' - E_{0,i})\varsigma(E')\,\mathrm{d}E', \quad (10)$$

using the above definition of $f_{E_i'}(E' - E_{0,i})$, describing the spread in energy of the individual neutrons. Thus, also taking the uncertainty of the peak energy $E_{0,i}$ into account (quantified by the PDF $f_{E_{0,i}}(E)$), the variance in $\tau_i^{(k)}$ due to the energy resolution is

$$V_{\Delta E}(\tau_i) = \int_{-\infty}^{\infty} f_{E_{0,i}}(E) \left(\int_{-\infty}^{\infty} f_{E_i'}(E' - E_{0,i})\varsigma(E')\right)^2 \mathrm{d}E'\,\mathrm{d}E - \left(\int_{-\infty}^{\infty} f_{E_{0,i}}(E) \int_{-\infty}^{\infty} f_{E_i'}(E' - E_{0,i})\varsigma(E')\,\mathrm{d}E'\,\mathrm{d}E\right)^2, \quad (11)$$

using that $V\left(\varsigma_{\text{observed}}(E_{0,i})\right) = \left\langle \varsigma^2_{\text{observed}}(E_{0,i})\right\rangle - \left\langle \varsigma_{\text{observed}}(E_{0,i})\right\rangle^2$. In practice, the average theoretical (T6) value at the energy E' is used for $\varsigma(E')$. Numerically, the integration is carried out using the trapezoidal rule using the grid provided in the random ACE file with the greatest index (see Sec. 3.1), and with integration limits at the grid points closest to, but outside, the range $E_i \pm \sqrt{(\Delta E_i')^2 + (\Delta E_{0,i})^2}$ (the probability densities are first normalized with respect to these integration limits). In practice, this variance is added to the random variance of the experimental point. It may be a good idea to consider this uncertainty as systematic in future work.

4. If the total systematic uncertainty is less than $\sigma_{\text{sys,min}} = 2\,\%$ (relative to the average *theoretical* value), a systematical uncertainty is added such that this value is reached. The same procedure but with $\sigma_{\text{stat,min}} = 1\,\%$ is used for the random uncertainty.

5. Extra uncertainties of $\sigma_{\text{sys,extra abs}} = \sigma_{\text{stat,extra abs}} = 1\,\text{mb}$ are added to the systematic and random uncertainty, respectively, for each experimental point as well as an extra random uncertainty of $\sigma_{\text{stat,extra rel to }\sigma} = 20\,\%$ relative to the random uncertainty and an extra systematic uncertainty of $\sigma_{\text{sys,extra rel}} = 0.5\,\%$ relative to the average theoretical value. These additions are due to a belief that experimental uncertainties often are underestimated, *e.g.*, only random uncertainty due to Poisson statistics may be reported. Also, adding an absolute uncertainty decreases the importance of smaller cross sections which typically are less important in applications since larger cross sections dominate.

6. An uncertainty of $\sigma_{\text{fully correlated}} = 1\,\%$ which is fully correlated for all experimental points within the same reaction channel is used. This is a temporary attempt to take cross-experimental correlations, arising from *e.g.* relative measurements, into account.

The above introduced EXFOR interpretation parameters (EI parameters) $\sigma_{\text{fully correlated}}$, $\sigma_{\text{stat,extra rel to }\sigma}$, $\sigma_{\text{sys,extra rel}}$, $\sigma_{\text{stat,extra abs}}$, $\sigma_{\text{sys,extra abs}}$, $\sigma_{\text{stat,min}}$, $\sigma_{\text{sys,min}}$ as well as P_{tol} (defined in Sec. 2.5) are varied in Sec. 3.2.

2.4. Building the experimental covariance matrix $\mathbf{C}_{\text{E}}$

Given the random and systematic uncertainties from Sec. 2.3, we construct the element on the i'th row and in the j'th column of the experimental covariance matrix according to

$$(\mathbf{C}_{\text{E}})_{ij} = \delta_{ij}\sigma_i^2 + \sum_{\ell=1}^{\nu} \sigma_{i\ell}\sigma_{j\ell}, \quad (12)$$

where σ_i^2 is the random variance of the i'th experimental point, $\sigma_{i\ell}$ is the uncertainty in the i'th experimental point due to the ℓ'th systematic contribution, ν is the number of systematic contributions (for all the considered experimental points) and the Kronecker delta, δ_{ij}, equals unity if $i = j$ and zero otherwise. A systematic contribution could be, *e.g.*, some particular target's thickness, the detector efficiency of a particular detector in some energy range or a particular normalizing cross section.

The expression in Eq. (12) results from modeling the random variable X_i, describing the i'th experimental point, as

$$X_i = Y_i + \sum_{\ell=1}^{\nu} \sigma_{i\ell}\mathcal{E}_\ell, \quad (13)$$

where Y_i is a random variable defined such that $\langle Y_i\rangle$ is the true value for the desired quantity and $V(Y_i) = \sigma_i^2$, $\mathcal{E}_\ell$ is a random variable describing the ℓ'th systematic contribution, with $\langle \mathcal{E}_\ell\rangle = 0$ and $V(\mathcal{E}_\ell) = 1$ (an observation of the product $\sigma_{i\ell}\mathcal{E}_\ell$ is the *error* due to the ℓ'th systematic contribution). $\mathcal{E}_\ell$ and Y_i are all mutually independent.

Under these assumptions, taking the covariance between the random variables describing the i'th and j'th experimental points, X_i and X_j, yields Eq. (12), since all other terms become zero and since $V(\mathcal{E}_\ell) = 1$.

2.5. Treating outliers and discarding energy zones

Just as in Ref. [19], outlying experiments (and "energy zones", see below) are in this study identified based on their deviation from the used nuclear data files. This procedure has the risk of confirming erroneous nuclear data because the evidence against it is rejected due to the same particular error. Nevertheless, this methodology is presently used because of its simplicity, but it shall be improved in future work.

As described below, an approximate p-value is computed for each experiment, which under certain (very strong) assumptions describes the probability for an observation *at least as extreme* as the observed experimental result. This p-value is compared to a tolerance $P_{\text{tol}} = 5\,\%$, and if the p-value is less than P_{tol}, the experiment is considered to be an outlier and is excluded from the random file weighting. The p-value is estimated from the difference between the respective experiments and the corresponding mean values of the random files taking both the experimental covariance and the covariance in the random files into account. At the end of the section, it is also described how the experiments are grouped into "energy zones", which also can be rejected using similar p-values.

For $j \in \{1, 2, ..., M\}$, where M is the number of experiments, let $\mathbf{x}_{(j)}$ be the vector with the experimental results in experiment j (a part of the vector $\mathbf{x}$) and define $\overline{\boldsymbol{\tau}}_{(j)}$ as the mean value of the corresponding values in the random ND files. Further assume that $\mathbf{C}_{\mathrm{E}(j)}$ is the experimental covariance matrix for the j'th experiment, *i.e.* the block in $\mathbf{C}_{\mathrm{E}}$ with indices corresponding to the experimental points in $\mathbf{x}_{(j)}$. Finally, let $\mathbf{C}_{\tau(j)}$ denote the covariance for the corresponding theoretical values inherent in the random files, simply the matrix with sample covariances over the random files. Now define

$$\chi^2_{(j),\text{central}} := \left(\mathbf{x}_{(j)} - \overline{\boldsymbol{\tau}}_{(j)}\right)^{\mathrm{T}} \left(\mathbf{C}_{\mathrm{E}(j)} + \mathbf{C}_{\tau(j)}\right)^{-1} \left(\mathbf{x}_{(j)} - \overline{\boldsymbol{\tau}}_{(j)}\right) \tag{14}$$

and compute

$$P_j := 1 - F_{\chi^2(m_j)}\left(\chi^2_{(j),\text{central}}\right), \tag{15}$$

where m_j is the number of experimental points in experiment j and $F_{\chi^2(m_j)}$ is the distribution function for a χ^2-distributed random variable with m_j degrees of freedom. Then, the experiment is rejected if $P_j < P_{\text{tol}} = 5\,\%$.

The motivation for Eqs. (14) and (15) follows. Analogously to how $\mathbf{x}$ is an observation of $\mathbf{X}$, $\mathbf{X}_{(j)}$ is defined such that $\mathbf{x}_{(j)}$ is an observation of $\mathbf{X}_{(j)}$. Define $\mathbf{M}_{(j)}$ as the random vector describing the *true* (unknown) values for the physical quantities corresponding to $\mathbf{X}_{(j)}$. Consider the j'th experiment given that $\mathbf{M}_{(j)} = \mathbf{m}_{(j)}$ for some (non-random) vector $\mathbf{m}_{(j)}$, denoted $\mathbf{X}_{(j)}|(\mathbf{M}_{(j)} = \mathbf{m}_{(j)})$. It is natural to assume that $\mathbf{X}_{(j)}$ is distributed around the true values, *i.e.*, $\left\langle \mathbf{X}_{(j)}|(\mathbf{M}_{(j)} = \mathbf{m}_{(j)})\right\rangle = \mathbf{m}_{(j)}$, and it is already assumed that $\mathbf{X}$ is Gaussian with covariance matrix $\mathbf{C}_{\mathrm{E}}$ such that $\mathbf{X}_{(j)}$ is Gaussian with covariance matrix $\mathbf{C}_{\mathrm{E}(j)}$ [29], giving $\mathbf{X}_{(j)}|(\mathbf{M}_{(j)} = \mathbf{m}_{(j)}) \sim N\left(\mathbf{m}_{(j)}, \mathbf{C}_{\mathrm{E}(j)}\right)$. Further assume that $\mathbf{M}_{(j)} \sim N\left(\overline{\boldsymbol{\tau}}_{(j)}, \mathbf{C}_{\tau(j)}\right)$. This can be interpreted as if the true values are distributed according to the information in the random files which is the desired interpretation of the random files; however, this is a circular argument since we want to determine this distribution. To summarize, we have

$$\begin{gathered}\mathbf{X}_{(j)}|(\mathbf{M}_{(j)} = \mathbf{m}_{(j)}) \sim N\left(\mathbf{m}_{(j)}, \mathbf{C}_{\mathrm{E}(j)}\right) \\ \text{with } \mathbf{M}_{(j)} \sim N\left(\overline{\boldsymbol{\tau}}_{(j)}, \mathbf{C}_{\tau(j)}\right),\end{gathered} \tag{16}$$

cf., Ref. [24, Sec. II.3]. It can be seen from Ref. [30] that this implies

$$\mathbf{X}_{(j)} \sim N\left(\overline{\boldsymbol{\tau}}_{(j)}, \mathbf{C}_{\mathrm{E}(j)} + \mathbf{C}_{\tau(j)}\right). \tag{17}$$

Assuming that $\overline{\boldsymbol{\tau}}_{(j)}$ really is a fixed vector (that it is not adapted to fit observations of $\mathbf{X}_{(j)}$), this gives

$$\mathbf{X}_{(j)} - \overline{\boldsymbol{\tau}}_{(j)} \sim N\left(\mathbf{0}_{m_j}, \mathbf{C}_{\mathrm{E}(j)} + \mathbf{C}_{\tau(j)}\right), \tag{18}$$

where $\mathbf{0}_{m_j}$ is the zero column vector with m_j elements. Thus [24],

$$\left(\mathbf{X}_{(j)} - \overline{\boldsymbol{\tau}}_{(j)}\right)^{\mathrm{T}} \left(\mathbf{C}_{\mathrm{E}(j)} + \mathbf{C}_{\tau(j)}\right)^{-1} \left(\mathbf{X}_{(j)} - \overline{\boldsymbol{\tau}}_{(j)}\right) \sim \chi^2(m_j), \tag{19}$$

yielding that P_j of Eq. (15) is the p-value for the observation of $\mathbf{x}_j$.

For some reaction channels in some energy ranges ("energy zones"), the used random files do not agree well with experimental data. If there are many different experiments available, it is unlikely that this is due to unknown systematic errors in the experiments, and the deviation can be due to either an erroneous parameter distribution or the fact that the models are imperfect. To ensure that such zones do not have a large impact on the weights, an attempt is made to automatically identify such zones and remove them from the weight assigning. First, let "energy zone" refer to a combination of a particular reaction channel and one of the following energy ranges: $[0, 0.1)$ eV, $[0.1, 1)$ eV, $[1, 5)$ eV, $[1, 5)$ MeV, $[5, 10)$ MeV and $[10, 20)$ MeV. Now, the experiments are divided into experimental subsets defined by which energy zone they belong to. Then, for j spanning over all experimental subsets which is in one energy zone, P_j is obtained from Eqs. (14) and (15) and subset j is determined to be an outlying subset if $P_j < P_{\text{tol}}$. If more than half of the experimental subsets in one energy zone is determined to be an outlier, the whole energy zone is discarded. In practice, the rejection of energy zones is carried out before the rejection of experiments, such that the subsets of an experiment (covering several energy zones) in the surviving energy zones may "survive" even if the experiment as a whole would be rejected.

Note that experiments and energy zones are rejected before the rules in Sec. 2.3 are applied, in contrast to what is done in Ref. [19].

3. This study: applications, considered nuclear data and sensitivities

Since the methodology primarily focuses on finding uncertainties due to nuclear data in simulated integral systems ("applications"), a selection of applications has been studied for this paper. In Sec. 3.1, the reader finds short descriptions of each application along with which data has been varied and which output quantities that have been studied.

In all cases the files are weighted using the (n,tot), (n,el), (n,2n), (n,f) and (n,γ) cross sections where they apply; for ^{56}Fe, the (n,n$'$) cross section is also included. The weights are obtained from experiments with neutron energies $E < 5\,\text{eV}$ or $E \geq 1\,\text{MeV}$, depending on whether the system is judged to have a thermal spectrum or a fast spectrum, respectively. These energy limits are chosen to include the most important energies except the resonance range. In principle, one can use both energy ranges simultaneously, but the files are randomized with high and low energies independently of each other, why this most likely will lead to a loss of precision without much gain in accuracy.

The theoretical values $\boldsymbol{\tau}$ are obtained from ACE files (format used for nuclear data in, *e.g.*, MCNP) that are obtained from processing by NJOY 99.336 [31] at a temperature of 300 K. The TMC methodology is applied twice, both using the full random ND files and using random data only for the cross section data in the energy ranges that are used in the weighting ($E < 5\,\text{eV}$ or $E \geq 1\,\text{MeV}$); the latter is referred to as partial variation in the following. For the partial variation, n copies are made of the default TENDL-2012 ACE file (processed using the appropriate temperature as in the reference of each application) and then the cross sections of each random ACE file is inserted into the corresponding copy in the desired energy range, using linear interpolation to match the energy grid. Angular distributions, fission neutron yield ($\bar{\nu}$) and cross sections at other energies, for example, are thus kept fixed in these cases. This is to isolate the effects of the nuclear data and the energy ranges for which experiments have been used. The results will focus on the runs using the partial variation, but the results using the variation of the full random files are also provided since they are the more complete estimates of the ND uncertainty.

3.1. Applications

Below, the different applications are described. The information is summarized in Table 1.

- A PWR pin cell modeled with depletion in Serpent [32] as in Ref. [7], either with UO_2 or MOX fuel. In the UO_2 case, ^{235}U, ^{238}U and ^{239}Pu have been varied (separately). In the MOX case, only ^{239}Pu has been varied. The neutron multiplication factor k_∞ at End Of Life (EOL) is studied (any output from Serpent can easily be studied but the amount of information becomes huge).
- A Serpent model of the lead cooled conceptual reactor design ELECTRA (European LEad Cooled ReActor) as in Refs. [33, 34], only at beginning of life. ^{239}Pu data is varied and the uncertainty in k_{eff} is studied.
- An MCNP-6 [35] model of a 14 MeV neutron source shielded by a concrete wall at the FREIA facility at Uppsala University, as in Ref. [6] but with ENDF/B-VI.2 and older libraries replaced by ENDF/B-VII.1. The concrete has a high content of iron, and the impact of ^{56}Fe data on the dose per source neutron is studied. In Ref. [6], the uncertainty due to ^{56}Fe data is found to be 4.2 %.
- An MCNP-6 model of the ^{239}Pu Jezebel benchmark [36] (plutonium-metal-fast-001 in the International Handbook of Evaluated Criticality Safety Benchmark Experiments [37]), an unreflected, critical, plutonium sphere. In the actual experiment, k_{eff} was close to 1, but due to an uncertainty in the actual critical mass (for example including corrections for an imperfect sphere), the "experimental uncertainty" is estimated to give $k_{\text{eff}} = 1.000 \pm 0.002$ [36].
- An MCNP-6 model used in Ref. [38] to compute the shielding factor obtained from using shielding fuel assemblies (SFA). The study in Ref. [38] is part of the design of the second generation of the SFA in Vattenfall's Ringhals 3 and 4 reactors in southern Sweden. The SFA reduce the reactor vessel's exposure to high-energy neutron flux – the shielding factor is defined as the fast flux ($E > 1\,\text{MeV}$) near a sensitive weld in a reference case without SFA's divided by that obtained using the SFA's. Here, the ^{56}Fe data uncertainty in the flux of the 29th cycle of Ringhals 3 is studied. The results quoted throughout this text except in Sec. 4.4 concern the flux at the point at the weld where $y = 0$ (in the studied quarter of the reactor), see Fig. 10.

3.2. Sensitivity to arbitrary choices

In Sec. 2.3, it is described how the EXFOR database is interpreted in this study. At several stages in the interpretation, quite arbitrary choices have been made, and it is relevant to ask how these choices may affect the results. To this end, many of the EXFOR interpretation parameters (EI parameters) and the rejection parameters have been varied according to the ranges in Table 2. P_{tol}, defined in Sec. 2.5, is also included among the EI parameters. In each case, all other parameters have been fixed to their default values, given in Secs. 2.3 and 2.5 and in boldface font in Table 2. The resulting weights as well as their impact on both the cross sections and on the applications have been studied.

The variations range over rather extreme values (*e.g.*, $\sigma_{\text{fully correlated}}$ ranges up to 10^4 %). This is to study limit-

Table 1: Summation of the application runs in this study. "Quantity" refers to the one most thoroughly studied output quantity.

Run #	Appl.	Quantitiy	# of random files ^{239}Pu	^{235}U	^{238}U	^{56}Fe	weights from E $< 5\,\text{eV}$	$\geq 1\,\text{MeV}$
1	UO_2 pin	k_∞ at EOL	683	-	-	-	✓	-
2	UO_2 pin	k_∞ at EOL	-	740	-	-	✓	-
3	UO_2 pin	k_∞ at EOL	-	-	345	-	✓	-
4	MOX pin	k_∞ at EOL	683	-	-	-	✓	-
5	Jezebel	k_{eff}	683	-	-	-	-	✓
6	ELECTRA	k_{eff}	683	-	-	-	-	✓
7	FREIA	Dose/n	-	-	-	896	-	✓
8	SFA	$\phi(E > 1\,\text{MeV})$	-	-	-	200	-	✓

Table 2: The used values of the EXFOR-interpretation parameters and rejection parameters that are varied in the sensitivity study. Definitions in Sec. 2.3 and Sec. 2.5, respectively. Default values in boldface.

Parameter	Range
$\sigma_{\text{fully correlated}}$ [%]	$0, 0.5, \mathbf{1}, 3, 5, 10, 100, 1000, 10^4$
$\sigma_{\text{stat,extra rel to }\sigma}$ [%]	$0, 5, 10, \mathbf{20}, 50, 500$
$\sigma_{\text{sys,extra rel}}$ [%]	$0, 0.25, \mathbf{0.5}, 1, 5, 100$
$\sigma_{\text{stat,extra abs}}$ [mb]	$0, 0.5, \mathbf{1}, 2, 5, 1000$
$\sigma_{\text{sys,extra abs}}$ [mb]	$0, 0.5, \mathbf{1}, 2, 5, 1000$
$\sigma_{\text{stat,min}}$ [%]	$0, 0.1, 0.5, \mathbf{1}, 2, 5, 10, 100$
$\sigma_{\text{sys,min}}$ [%]	$0, 0.2, 1, \mathbf{2}, 5, 10, 100$
P_{tol} [%]	$0, 1, \mathbf{5}, 10, 50$

ing behavior for verification purposes and theoretical discussion.

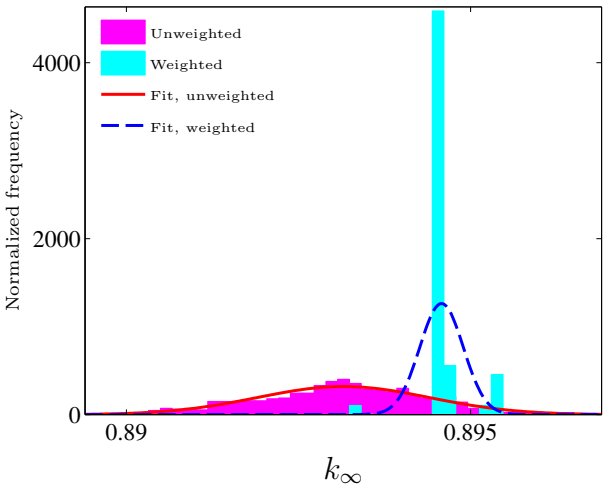

Figure 2: The unweighted and default weighted distribution of k_∞ at EOL for the UO_2 pin cell, varying ^{235}U data.

4. Results

In this section, the central values and ND uncertainties for the applications described in Sec. 3.1 are found. In Sec. 4.1, the results are presented with and without applying the default file weights, and Sec. 4.2 covers the main findings from varying the different EI parameters according to Sec. 3.2. Weights for the corresponding ND from other ND libraries such as ENDF/B-VII.1 are found in Sec. 4.3. Finally, the results for the Shielding fuel assemblies (SFA) are studied a bit more thoroughly in Sec. 4.4, since this application have not been studied before with respect to ND uncertainties.

4.1. Integral results with default weights

Table 3 views the central values and estimates of the ND uncertainty using partial variations, with and without using the weights for the random files, obtained using the default settings for the EI parameters. The table also views the unweighted results that are obtained using the full random files in the variation.

The uncertainties vary depending on the applications and on which sets of ND that are varied, and the unweighted uncertainties obtained with the full random files agree with other studies using TMC [7, 39, 6], except for the SFA, which have not previously been studied with respect to ND uncertainties. For FREIA, there is some discrepancy compared to Ref. [6], but this can be explained by that newer nuclear data libraries are used in this study. For the SFA (which are discussed more in Sec. 4.4), the unweighted ND uncertainty is (perhaps surprisingly) greater for partial variations than for variation of the full random files.

In most cases, the partial variation contributes to a large part of the uncertainty obtained from full variation, if not even the majority. The exceptions are ELECTRA and, more distinguishing, the UO_2 pin varying ^{238}U. Since the unweighted uncertainty obtained with partial variations in the latter of these two cases is practically zero, there is not much that can be said from the weighted result in this case; since it is limited by a distribution giving such a small uncertainty, the weighted uncertainty will also be very small.

In the other cases with thermal systems, varying ^{239}Pu or ^{235}U, the unweighted uncertainties differ from zero while the weighted uncertainties become zero. The central values are significantly shifted by the weights, too. A closer

Table 3: Central values and ND uncertainties for the main quantities of the integral systems described in Sec. 3.1 with the full random files and with the partial random files, with and without using the default file weights. The quoted uncertainties denote estimates of one standard deviation.

Exp. from	Application	ND	Quantity	Full var. Unw.	Partial variation Unw.	Partial variation Weighted
$E < 5\,\mathrm{eV}$	UO_2 pin	^{239}Pu	k_∞	.8952(2)	.8955(2)	.8985(4)
			σ_{ND} [%]	.54(2)	.46(1)	$.(3)\cdot 10^{-2}$
		^{235}U	k_∞	.89154(8)	.89315(5)	.8946(3)
			σ_{ND} [%]	.240(6)	.117(5)	$.(2)\cdot 10^{-2}$
		^{238}U	k_∞	.8969(2)	.89333(4)	.89337(6)
			σ_{ND} [%]	.37(1)	$2(1)\cdot 10^{-2}$	$2(2)\cdot 10^{-2}$
	MOX pin	^{239}Pu	k_∞	.9124(2)	.9129(1)	.9158(2)
			σ_{ND} [%]	.49(2)	.34(1)	$.(2)\cdot 10^{-2}$
$E \geq 1\,\mathrm{MeV}$	Jezebel	^{239}Pu	k_{eff}	1.0008(4)	1.0010(2)	1.000(2)
			σ_{ND} [%]	.95(2)	.59(2)	.56(9)
	ELECTRA	^{239}Pu	k_{eff}	1.0010(3)	1.00123(9)	1.0011(7)
			σ_{ND} [%]	.74(2)	.206(7)	.23(6)
	FREIA	^{56}Fe	Dose [pSv/n]	1.324(5)	1.344(3)	1.308(5)
			σ_{ND} [%]	5.6(3)	5.7(3)	4.8(4)
	SFA	^{56}Fe	ϕ [10^9 cm^{-1}s^{-1}]	5.12(1)	5.12(1)	5.11(3)
			σ_{ND} [%]	2.5(2)	3.1(2)	3.1(3)

investigation shows that there are only a very small number of files that obtain significant weights which results in very peaked distributions, *i.e.*, the results are poorly converged. One such peaked distribution can be seen in Fig. 2, which views the distribution of k_∞ at EOL for the UO_2 pin varying ^{235}U data, with and without using the file weights. The observed standard deviation using weights ($0.04\,\% \pm 0.05\,\%$) is less than the random uncertaint. Therefore the uncertainty is reported as zero (directly applying Eq. (9) would give an imaginary uncertainty). Because of the small number of significant weights, the uncertainties of the ND uncertainties and the central values should be interpreted carefully, as mentioned in Sec. 2.2 – they are quite meaningless. The possible reasons for the few significant file weights observed in these cases are discussed in Sec. 5.2.

The change in ND uncertainty for FREIA can barely be judged as significant, but the central value is shifted significantly (see Sec. 5.6 for a discussion on statistical significance in this context). In the other cases, the weighted results in Table 3 do not differ significantly from the unweighted results, the reasons for which are discussed in Sec. 5.1. In almost all cases, including these, the uncertainty of the ND uncertainty and in the central value is increased, due to the smaller number of files being important for the distributions.

4.2. Sensitivity to EI parameters

4.2.1. Main findings

The main findings in the effect from the variation of the different EI parameters on the distributions for the results for the integral systems are described in the bullet list below. More detailed results from varying P_{tol} are found in Sec. 4.2.2.

- For most of the considered EI parameter values, the ND uncertainty for the thermal systems (UO_2 or MOX pin cell) varying ^{235}U or ^{239}Pu is estimated to be zero when the weights are applied, while the unweighted case yields a propagated ND uncertainties of more than 0.1 % for ^{235}U and around 0.4 % for ^{239}Pu. The central values may shift significantly when varying some EI parameters, though. Studying the average weight for the ^{235}U and ^{239}Pu random files reveals an explanation; it is less than 0.2 % and 0.4 % of $w_{\max}$, respectively, in almost all cases, and the file with the maximum weight alone contributes to $1/n \gtrsim 1/740 = 0.14\,\%$ of these (n is the number of files). Hence, a very small number of files contribute significantly to the weighted distribution. This was seen for the default weights in Sec. 4.1 and the corresponding narrow distribution was illustrated for ^{235}U in Fig. 2. The occasional shifts in the central values are simply due to that the actual file(s) with the greatest weights may vary with the EI parameter values, even though the number of significant files stays low. An example of this can be seen in Fig. 4, where the distribution of k_∞ is viewed, without weights and with weights using $\sigma_{\text{fully correlated}} = 1\,\%$ and $3\,\%$, respectively. Here, the shift in the central value is obtained from moving the majority of the weight from a few files in two bins to another single bin (in practice to one single random file), a process which clearly is highly sensitive to random fluctuations.
- The ^{239}Pu random files also get low average weights using experiments from $E \geq 1\,\mathrm{MeV}$. This does not propagate to give a zero uncertainty as for ^{235}U and ^{239}Pu-thermal; instead, neither the central values nor the ND uncertainty estimates are affected much by the weights. According to Fig. 3, which shows the unweighted and the default weighted distribution of k_{eff} for Jezebel, it seems that the weighting and the k_{eff} values are effectively uncorrelated which results

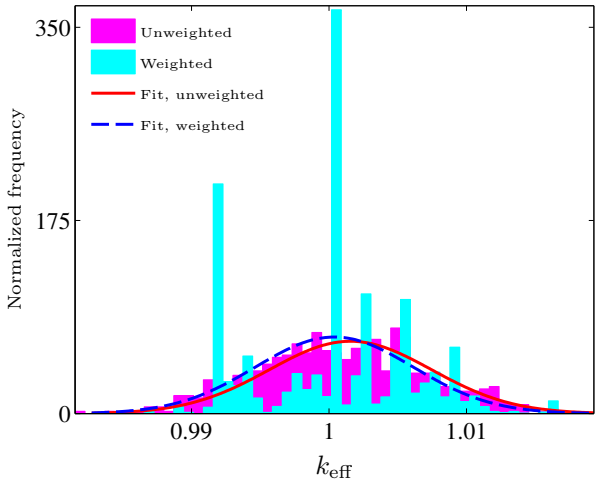

Figure 3: The unweighted and default weighted distribution of k_{eff} for Jezebel, varying ^{239}Pu data.

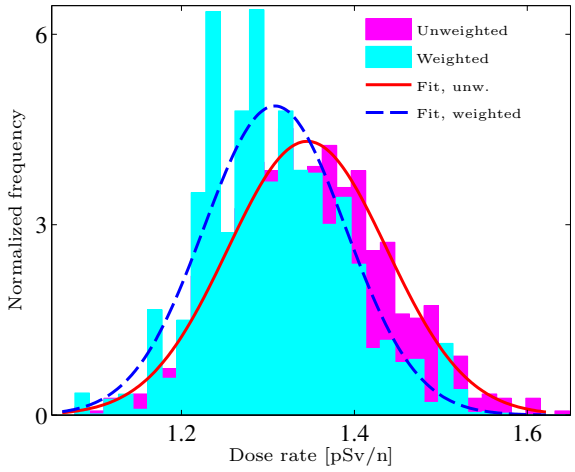

Figure 5: The unweighted and default weighted distribution of the FREIA dose rate, varying ^{56}Fe data.

in practically the same distribution, although more "noisy" because of the few files with a significant weight.

- Not much can be said about the weighted propagated uncertainty in the ^{238}U case since the unweighted distribution in the integral result is very narrow (using partial variations), as noted already in Sec. 4.1. The central value is very close to the unweighted value for all studied EI parameter settings, and the ND uncertainty estimate is very small. However, it can be noted that there are relatively many random ^{238}U files with significant weights, the average weight is on the order of 10 % for all EI parameter values.

- Neither the central value or the ND uncertainty estimate changes much with the EI parameters for the SFA, with one exception, namely in the case when $P_{\mathrm{tol}} = 0$, which is studied more in Sec. 4.2.2.

- For FREIA, the ND uncertainty is quite unaffected by the weights, but the central value is shifted towards lower values for most EI parameter choices. The distributions for the dose rate with and without using the default weights are seen in Fig. 5. The statistical significance of, *e.g.*, this shift is discussed in Sec. 5.6. In most cases, the shift decreases as the EI parameters increase (effectively giving larger experimental uncertainty).

- Increasing random experimental uncertainty typically gives more files with significant weights, which is easy to intuitively understand (the distributions become "wider"). For systematic experimental uncertainty, however, the picture is not as clear; sometimes an increased systematic uncertainty gives a more narrow distribution of the random files, and an effect of reaching an equilibrium is seen (referred to as "saturation" below). This saturation effect can be observed in Figs. 6 and 7. In Fig. 6, the central values and ND uncertainties for the quantities in the fast integral systems when varying $\sigma_{\text{fully correlated}}$ are viewed, and Fig. 7 shows the average weights for all cases relevant to the integral systems, also as a function of $\sigma_{\text{fully correlated}}$. One can see that the average weights can both increase and decrease as $\sigma_{\text{fully correlated}}$ increases, but when increasing $\sigma_{\text{fully correlated}}$ from the rather extreme 10^3 % to 10^4 %, the difference is small. In fact, the largest difference between the likelihood function in the two cases is less than 0.5 %. Following from this, the central values and the ND uncertainty estimates are also very similar in the two cases. This saturation as well as the fact that the average weights are observed to sometimes decrease when systematic uncertainties increase are discussed in Sec. 5.3.

- Many of the EI parameters do not impact neither of the applications much as long as they are within "reasonable ranges " (*i.e.* close to their default values) except for ^{239}Pu but this can be due to statistical fluctuations as mentioned above. The EI parameters $\sigma_{\text{fully correlated}}$, $\sigma_{\text{sys,extra rel}}$ and P_{tol}, however, have substantial effects on the *average file weights* even when they are varied close to their defaults.

- In practically all cases, the uncertainty of the central value and of the ND uncertainty estimate itself increases when the weights are applied, due to the effectively smaller number of random files.

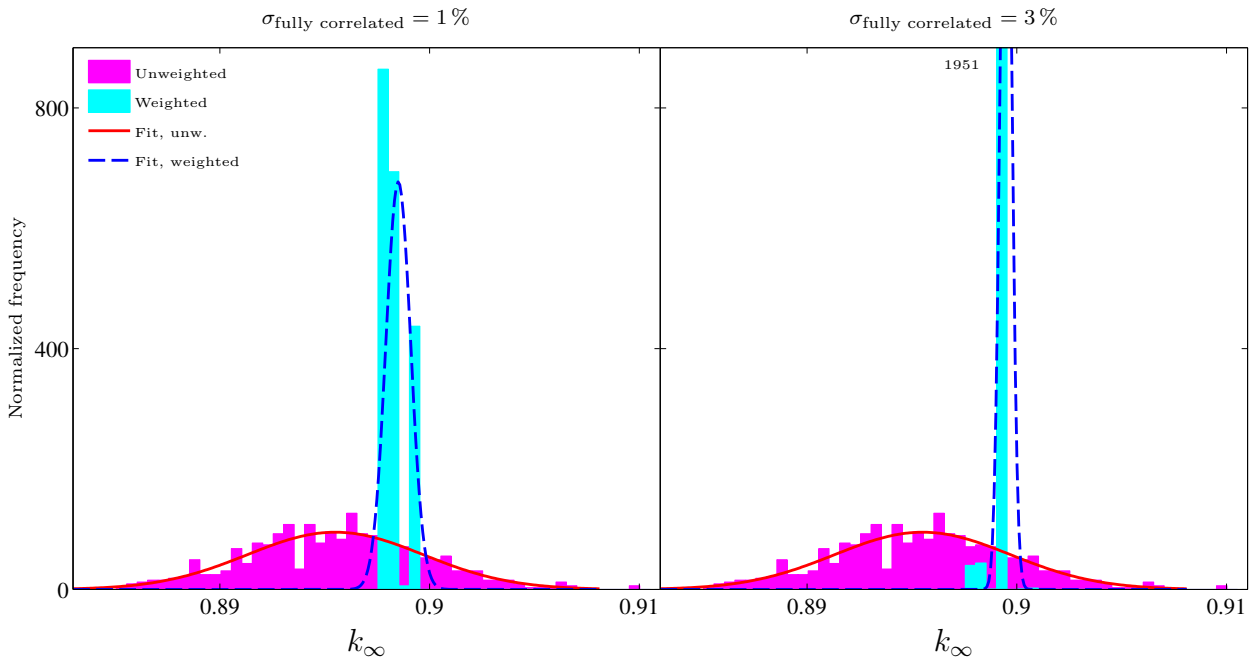

Figure 4: The k_∞ distribution of the UO_2 pin cell at EOL varying ^{239}Pu, with $\sigma_{\text{fully correlated}} = 1\,\%$ and $\sigma_{\text{fully correlated}} = 3\,\%$ (default). The peak of the Gaussian fit to the weighted case with $\sigma_{\text{fully correlated}} = 3\,\%$ is truncated in the figure.

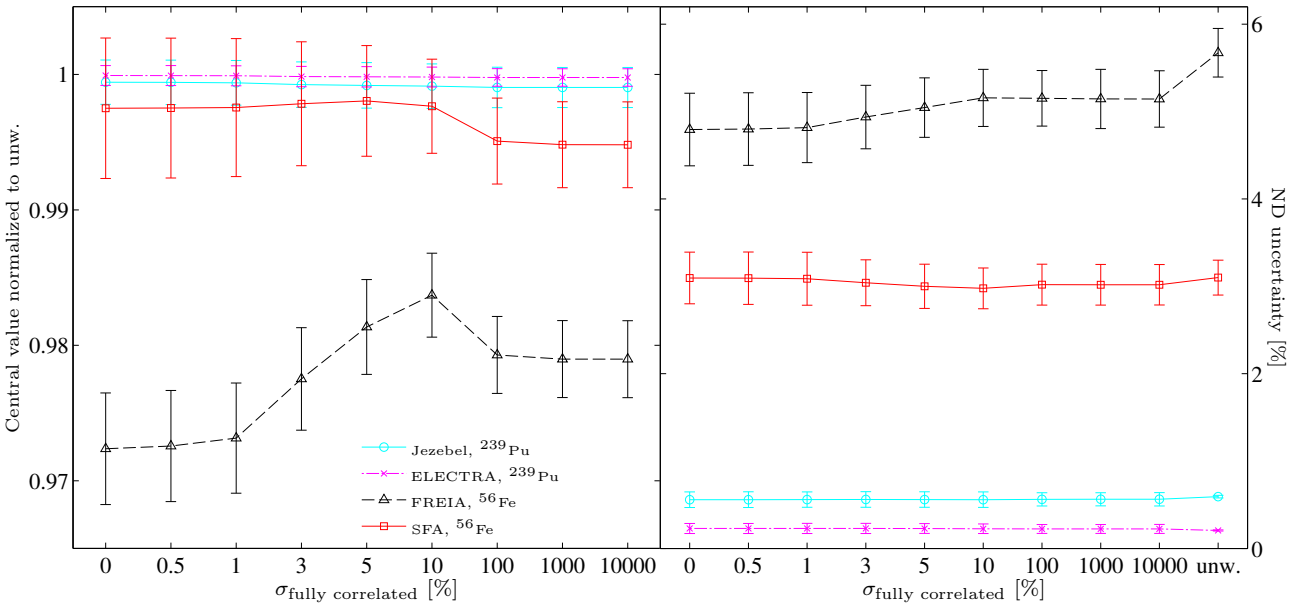

Figure 6: Central values and ND uncertainties for the fast integral systems as functions of $\sigma_{\text{fully correlated}}$ (default 1 %). The central values are normalized to the unweighted results and the ND uncertainties are relative to the central values. Note that the last value along the $\sigma_{\text{fully correlated}}$-axis corresponds to the unweighted case. The uncertainty bars may appear unreasonably large compared to the differences between the values for different parameter values but this is since the weights for different parameter values are strongly correlated.

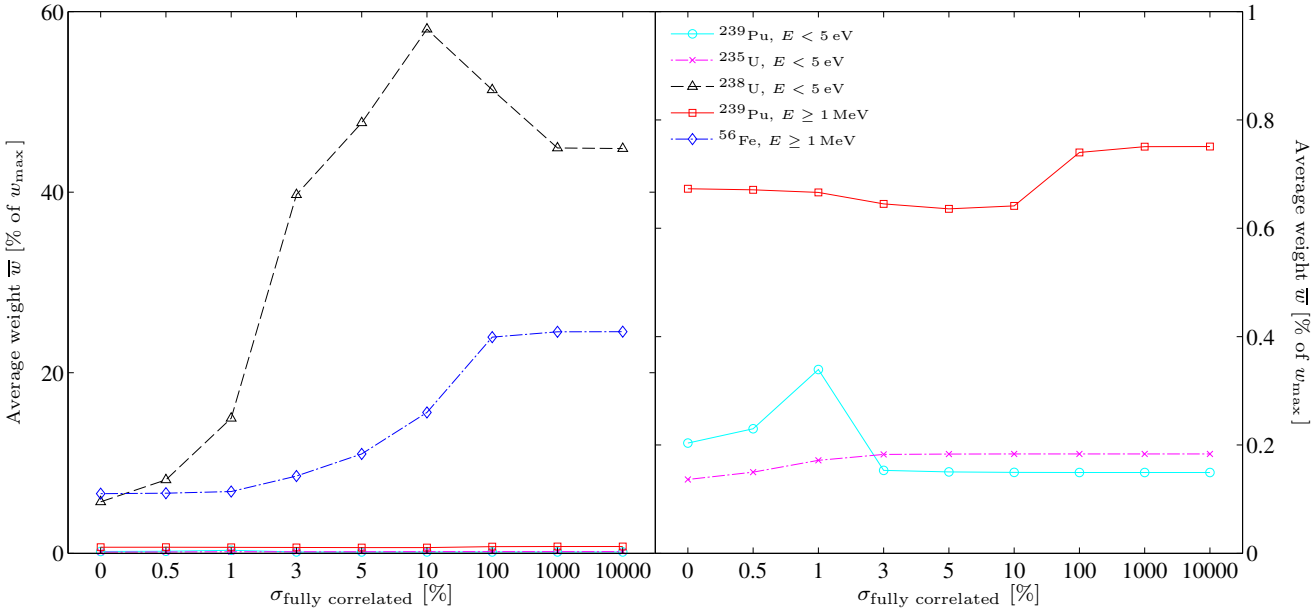

Figure 7: The average weights for the sets of random files relevant to the integral systems as functions of $\sigma_{\text{fully correlated}}$ (default 1 %). The right hand figure is a zoom-in on small average weights.

4.2.2. Varying P_{tol}

Because of the particular nature of P_{tol} (determining the "p-value" to be the limit for discarding experiments and "energy zones"), it is studied in some more detail here. The central values and ND uncertainties for the fast integral systems when P_{tol} is varied are seen in Fig. 8. In all cases shown in the figure, the ND uncertainty estimate vanishes when $P_{\text{tol}} = 0\,\%$, *i.e.*, when all experiments are accepted. The same holds in all other cases except when ^{238}U is varied. The average weights, see Fig. 9, expose that $P_{\text{tol}} = 0\,\%$ implies that almost only one file becomes significant in all these cases, which leads to this zero uncertainty estimate. These very narrow distributions are not very surprising since some of the experiments contained in EXFOR are likely to be erroneous and without *any* selection of experiments, experiments which differ very much from the random files can be included. They will have a large impact on the weights due to the exponential behavior of the likelihood function, see Eq. (1). ^{238}U is unaffected since almost no experiments are rejected anyway (except for $P_{\text{tol}} = 50\,\%$).

Tables 4 to 7 show which energy zones are rejected for each choice of P_{tol}. The tables also view the number of experiments which have points in each zone and the number of experimental points in each zone. $P_{\text{tol}} = 0\,\%$ is omitted from the tables since no zones are rejected with this choice. One can see that in many cases, either the zone gets rejected for all used non-zero choices of P_{tol} or not at all; many other zones get rejected for $P_{\text{tol}} = 50\,\%$ only.

For the default, $P_{\text{tol}} = 5\,\%$, one can note that the total cross section often is rejected; *e.g.*, for ^{239}Pu and ^{235}U this happens for $E \in [0.1, 5)$ eV. For ^{238}U, no zones with low energies are rejected using $P_{\text{tol}} = 5\,\%$, indicating a good agreement between experiments and the random files, which is also indicated by the relatively large weights.

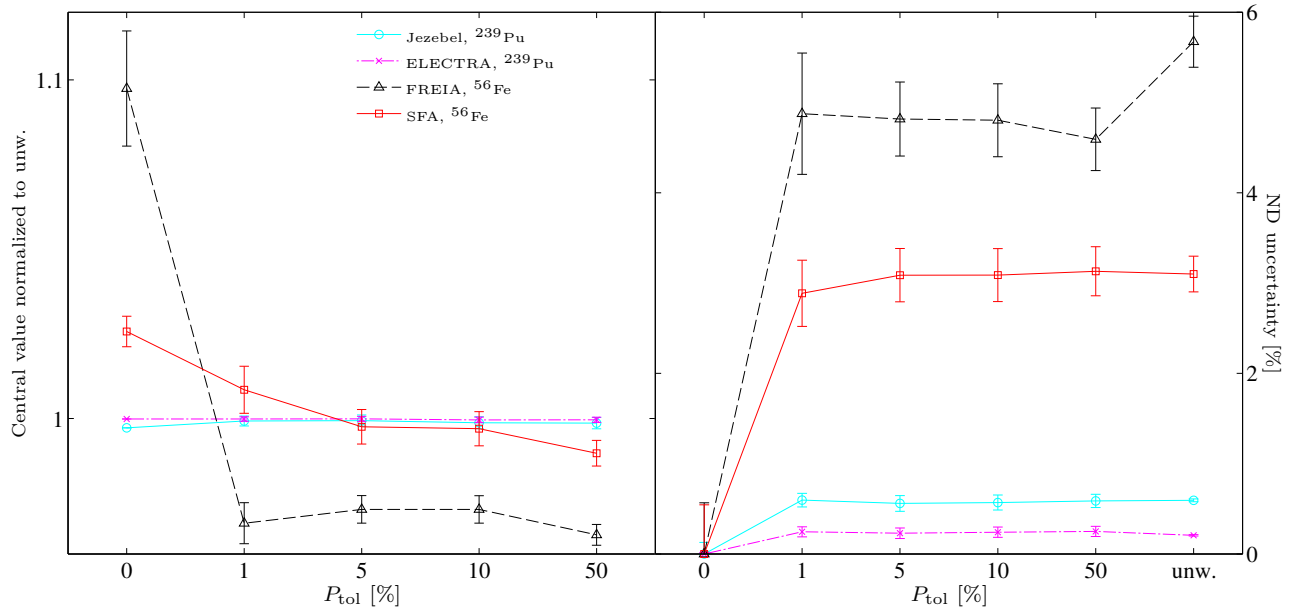

Figure 8: Central values and ND uncertainties for the fast integral systems as functions of P_{tol} (default: 5 %).

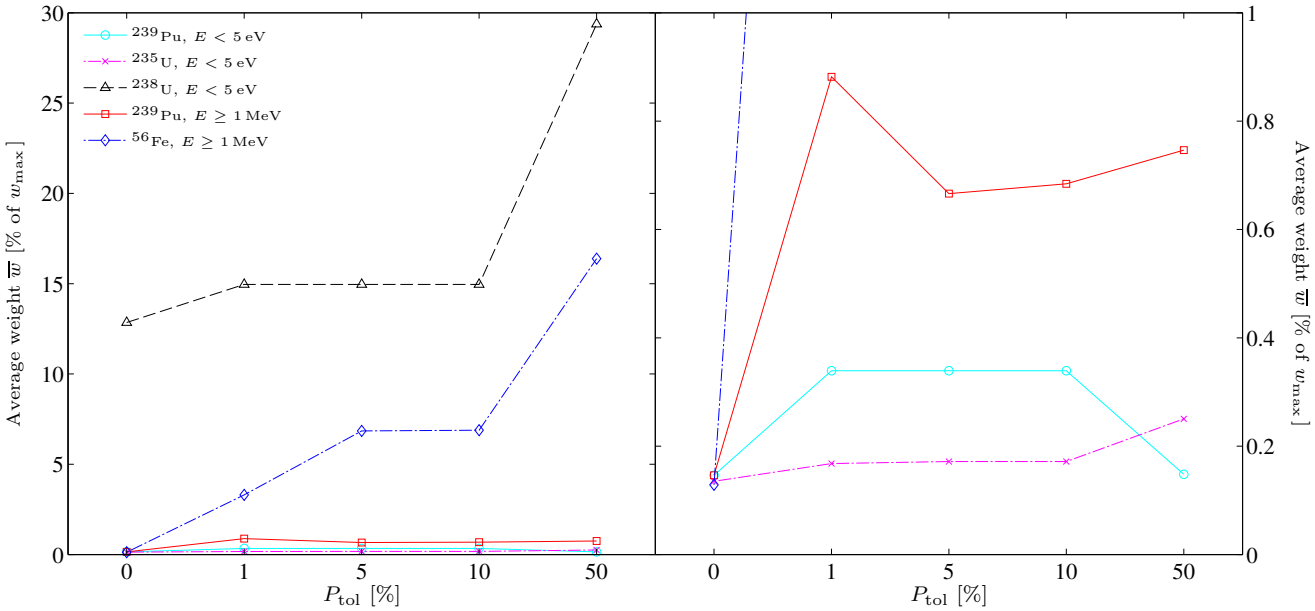

Figure 9: The average weights for the sets of random files relevant to the integral systems as functions of P_{tol} (default: 5 %). The right hand figure is a zoom-in on small average weights.

Table 4: The number of experiments and experimental points in each energy zone for ^{239}Pu, and which zones are rejected for different choices of $P_{\rm tol}$. "×" means that a zone is rejected and "✓" that a zone is kept. The number of experiments and experimental points refer to all those in EXFOR passing rule number 1 in Sec. 2.3.

				$P_{\rm tol}$ [%]			
		# exp's	# points	1	5	10	50
(n,tot)	[0, 100) meV	10	108	✓	✓	✓	✓
	[100, 1000) meV	11	360	×	×	×	×
	[1, 5) eV	10	1123	×	×	×	×
	[1, 5) MeV	8	1649	✓	✓	✓	✓
	[5, 10) MeV	5	383	✓	✓	×	×
	[10, 20) MeV	6	186	✓	✓	×	×
(n,el)	[0, 100) meV	0	0				
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	2	17	✓	✓	✓	×
	[5, 10) MeV	1	20	✓	✓	✓	✓
	[10, 20) MeV	1	18	✓	✓	✓	✓
(n,2n)	[0, 100) meV	0	0				
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	0	0				
	[5, 10) MeV	3	17	×	×	×	×
	[10, 20) MeV	4	18	✓	✓	×	×
(n,f)	[0, 100) meV	28	1542	✓	✓	✓	×
	[100, 1000) meV	12	1034	✓	✓	×	×
	[1, 5) eV	12	1083	✓	✓	✓	✓
	[1, 5) MeV	12	300	✓	✓	✓	✓
	[5, 10) MeV	7	120	✓	✓	✓	✓
	[10, 20) MeV	15	121	✓	✓	✓	✓
(n,γ)	[0, 100) meV	2	2	✓	✓	✓	✓
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	1	1	✓	✓	✓	✓
	[5, 10) MeV	0	0				
	[10, 20) MeV	0	0				

Table 5: The number of experiments and experimental points in each energy zone for ^{235}U, and which zones are rejected for different choices of $P_{\rm tol}$. "×" means that a zone is rejected and "✓" that a zone is kept. The number of experiments and experimental points refer to all those in EXFOR passing rule number 1 in Sec. 2.3.

				$P_{\rm tol}$ [%]			
		# exp's	# points	1	5	10	50
(n,tot)	[0, 100) meV	17	178	✓	✓	✓	×
	[100, 1000) meV	9	279	×	×	×	×
	[1, 5) eV	7	1910	×	×	×	×
	[1, 5) MeV	13	2531	✓	✓	✓	✓
	[5, 10) MeV	10	661	✓	✓	✓	×
	[10, 20) MeV	9	305	✓	✓	✓	×
(n,el)	[0, 100) meV	3	17	×	×	×	×
	[100, 1000) meV	1	17	✓	✓	✓	×
	[1, 5) eV	1	1	✓	✓	✓	✓
	[1, 5) MeV	2	2	✓	✓	✓	✓
	[5, 10) MeV	0	0				
	[10, 20) MeV	0	0				
(n,2n)	[0, 100) meV	0	0				
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	0	0				
	[5, 10) MeV	2	11	✓	✓	✓	×
	[10, 20) MeV	2	9	✓	✓	×	×
(n,f)	[0, 100) meV	27	2224	✓	✓	✓	✓
	[100, 1000) meV	18	1966	✓	✓	✓	✓
	[1, 5) eV	16	3040	✓	✓	✓	✓
	[1, 5) MeV	35	527	✓	✓	✓	✓
	[5, 10) MeV	15	183	✓	✓	✓	✓
	[10, 20) MeV	22	140	✓	✓	✓	✓
(n,γ)	[0, 100) meV	4	4	✓	✓	✓	✓
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	1	1	✓	✓	✓	✓
	[5, 10) MeV	0	0				
	[10, 20) MeV	0	0				

Table 6: The number of experiments and experimental points in each energy zone for ^{238}U, and which zones are rejected for different choices of $P_{\rm tol}$. "×" means that a zone is rejected and "✓" that a zone is kept. The number of experiments and experimental points refer to all those in EXFOR passing rule number 1 in Sec. 2.3.

				$P_{\rm tol}$ [%]			
		# exp's	# points	1	5	10	50
(n,tot)	[0, 100) meV	0	0				
	[100, 1000) meV	1	13	✓	✓	✓	✓
	[1, 5) eV	1	28	✓	✓	✓	✓
	[1, 5) MeV	13	1971	×	×	×	×
	[5, 10) MeV	7	564	✓	✓	✓	✓
	[10, 20) MeV	7	261	✓	✓	✓	✓
(n,el)	[0, 100) meV	0	0				
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	4	10	✓	✓	✓	×
	[5, 10) MeV	2	2	✓	✓	✓	×
	[10, 20) MeV	3	3	✓	✓	✓	✓
(n,2n)	[0, 100) meV	0	0				
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	0	0				
	[5, 10) MeV	5	26	✓	✓	✓	✓
	[10, 20) MeV	19	99	✓	✓	✓	✓
(n,f)	[0, 100) meV	1	1	✓	✓	✓	×
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	15	574	✓	✓	✓	✓
	[5, 10) MeV	8	112	✓	✓	✓	✓
	[10, 20) MeV	24	137	✓	×	×	×
(n,γ)	[0, 100) meV	12	155	✓	✓	✓	✓
	[100, 1000) meV	0	0				
	[1, 5) eV	2	10	✓	✓	✓	✓
	[1, 5) MeV	7	33	✓	✓	✓	✓
	[5, 10) MeV	2	8	✓	✓	✓	✓
	[10, 20) MeV	5	16	✓	✓	✓	✓

Table 7: The number of experiments and experimental points in each energy zone for ^{56}Fe, and which zones are rejected for different choices of $P_{\rm tol}$. "×" means that a zone is rejected and "✓" that a zone is kept. The number of experiments and experimental points refer to all those in EXFOR passing rule number 1 in Sec. 2.3.

				$P_{\rm tol}$ [%]			
		# exp's	# points	1	5	10	50
(n,tot)	[0, 100) meV	1	58	×	×	×	×
	[100, 1000) meV	1	15	✓	✓	✓	×
	[1, 5) eV	2	4	✓	✓	✓	✓
	[1, 5) MeV	2	4671	×	×	×	×
	[5, 10) MeV	3	1065	✓	✓	×	×
	[10, 20) MeV	3	625	✓	✓	✓	×
(n,el)	[0, 100) meV	1	1	✓	✓	✓	✓
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	4	7	✓	✓	✓	×
	[5, 10) MeV	2	8	✓	✓	✓	✓
	[10, 20) MeV	0	0				
(n,2n)	[0, 100) meV	0	0				
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	0	0				
	[5, 10) MeV	0	0				
	[10, 20) MeV	5	13	✓	✓	✓	✓
(n,γ)	[0, 100) meV	3	61	✓	✓	✓	✓
	[100, 1000) meV	1	28	✓	✓	✓	✓
	[1, 5) eV	1	19	✓	✓	×	×
	[1, 5) MeV	0	0				
	[5, 10) MeV	0	0				
	[10, 20) MeV	1	1	×	×	×	×
(n,n′)	[0, 100) meV	0	0				
	[100, 1000) meV	0	0				
	[1, 5) eV	0	0				
	[1, 5) MeV	2	7	✓	✓	✓	✓
	[5, 10) MeV	0	0				
	[10, 20) MeV	3	3	✓	✓	✓	✓

4.3. Weights for other libraries

Table 8 shows the values for the likelihood functions that are obtained with the methodology of this paper for a selection of ND libraries other than the considered random files. The values are normalized to the maximum likelihood function obtained for the random ND files. Thus, values greater than 1 indicate a better agreement to the experiments for the considered library compared to best random file and vice versa. Note that "better agreement to the experiments" means to the automatic selection of experiments made according to Secs. 2.3 and 2.5. Since outliers are identified based on the deviation to the random files, the random files may obtain unjustly large likelihood function values compared to the other libraries.

For $E \geq 1\,\text{MeV}$, ENDF/B-VII.1 succeeds quite well, with higher likelihood than any of the random files, except for ^{235}U, where no library seems to compete with the best random file. CENDL 3.1 also gets higher likelihood than the random files for ^{239}Pu and ^{238}U in this energy range, note the really large value for ^{239}Pu.

For $E < 5\,\text{eV}$, none of the libraries beats the best random file, even though the difference is pretty small in the cases of ^{238}U and ^{56}Fe, where the difference between the different libraries are really small, too.

JEFF 3.1, JENDL 4.0 and TENDL 2013 all have lower likelihood values than the best random files in all cases.

4.4. The shielding fuel assemblies (SFA)

Since there has been no previous studies of the ^{56}Fe data uncertainties in the SFA, these results (using the default weights) are studied with some more detail here.

In Fig. 10, the high energy flux at the reactor pressure vessel and the relative uncertainty in the same quantity due to ^{56}Fe data is seen. Where the weld crosses the planes $x = 0$ and $y = 0$, the sensitive points that should be protected by the SFA are located, and one can see that the flux is lower here than at greater z values for the same azimuthal angles. The flux pattern agrees with that found in Ref. [38] (where it is only reported along the weld).

The relative uncertainty due to ^{56}Fe data is (except for local variations at the top of the pressure vessel due to large random uncertainty), greatest in the vicinity of the sensitive points, where it is about $2 - 3\,\%$; this is not surprising since this is where the ^{56}Fe data is expected to be most important. With full variation of the random files, the unweighted uncertainty of the flux due to ^{56}Fe data is $2.5 \pm 0.2\,\%$ and $2.6 \pm 0.2\,\%$ at the sensitive point at $y = 0$ and at $x = 0$, respectively. The corresponding unweighted values using partial variations are $3.1 \pm 0.2\,\%$ and $2.8 \pm 0.2\,\%$. That the ND uncertainty estimates increase using partial variations is surprising, since less ND is varied; this is discussed in Sec. 5.7.

Applying the default weights (and using partial variation), ND uncertainties of $3.1 \pm 0.3\,\%$ and $2.6 \pm 0.3\,\%$ are obtained. Since the uncertainties are effectively unaffected by the weights from experimental data, there is a risk that the true ^{56}Fe data uncertainty is greater, as discussed in Sec. 5.1.

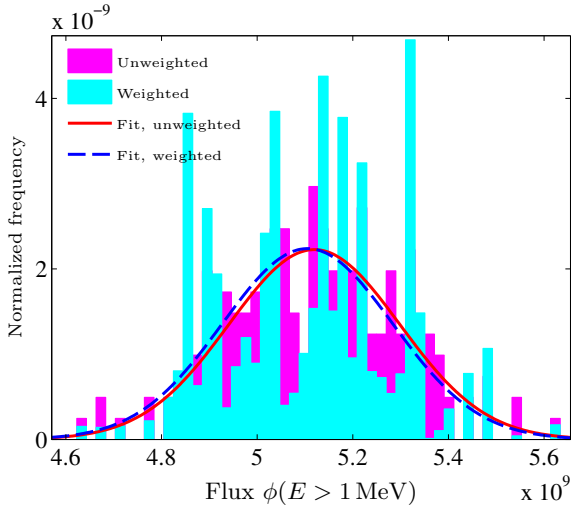

Figure 11: The distribution of the high energy flux ($E > 1\,\text{MeV}$) at a sensitive point of the SFA application (at the weld and $y = 0$) using partial variations of the ^{56}Fe data, with and without the default random file weights.

5. Discussion

5.1. Possible explanations for constant likelihood functions

For many of the nuclides, applications and choices of EI parameters, the weights had almost no effect on the integral distributions, *e.g.* for the SFA, see Fig. 11. This effectively means that the distributions are entirely determined by the prior distribution, and that the likelihood function is practically constant with respect to the studied integral quantity in the part of the parameter space covered by the prior distribution.

One possible reason for this is that the distribution used as prior covers a much too small part of the parameter space, relative to the experimental data set and its uncertainties, such that the experimental data cannot limit the distribution anymore. There is an obvious risk for this, since the random files used as prior files are not intended to describe prior distributions. In Ref. [18], a very wide prior distribution of model parameters is taken, validated against coverage of all EXFOR data using global nuclear model calculations, after which zoom-in to nuclide-specific data takes place to obtain a reliable posterior. There, this is only applied to differential cross sections, but it can equally well be applied to entire nuclear data libraries and applied reactor calculations.

Another possibility is that the application is sensitive to resonances (which are in the considered energy region) and that the resonances are so many that weighting of full random files becomes an inefficient way to limit them.

It is also possible that the preliminary algorithms for automatic selection of data or generation of experimental

Table 8: The likelihood functions for a selection of nuclear data libraries for each nuclide and energy range, relative to the maximum likelihood obtained for the random files. For values less than 10^{-10}, the value is rounded to the nearest power of 10. The greatest weight for each case (if greater than the maximum likelihood for the random files) is highlighted with a boldface font.

	^{239}Pu		^{235}U		^{238}U		^{56}Fe	
Library \ E	≥ 1 MeV	< 5 eV	≥ 1 MeV	< 5 eV	≥ 1 MeV	< 5 eV	≥ 1 MeV	< 5 eV
ENDF/B-VII.1	12	10^{-153}	$3.0{\cdot}10^{-6}$	$6.5{\cdot}10^{-7}$	**3.5**	.53	**69**	.18
JEFF 3.1	.028	$< 10^{-300}$	$7.7{\cdot}10^{-5}$	$6.5{\cdot}10^{-7}$	.011	.49	10^{-16}	.16
JENDL 4.0	$2.2{\cdot}10^{-5}$	$< 10^{-300}$	10^{-16}	$6.5{\cdot}10^{-7}$	$8.4{\cdot}10^{-4}$	.52	10^{-35}	.18
CENDL 3.1	$\mathbf{8.0{\cdot}10^{4}}$	10^{-44}	$3.1{\cdot}10^{-5}$	$6.3{\cdot}10^{-7}$	4.7	.53	10^{-19}	.61
TENDL 2013	$4.5{\cdot}10^{-5}$	10^{-153}	$< 10^{-300}$	$6.6{\cdot}10^{-7}$	$8.3{\cdot}10^{-5}$	.53	10^{-22}	.18

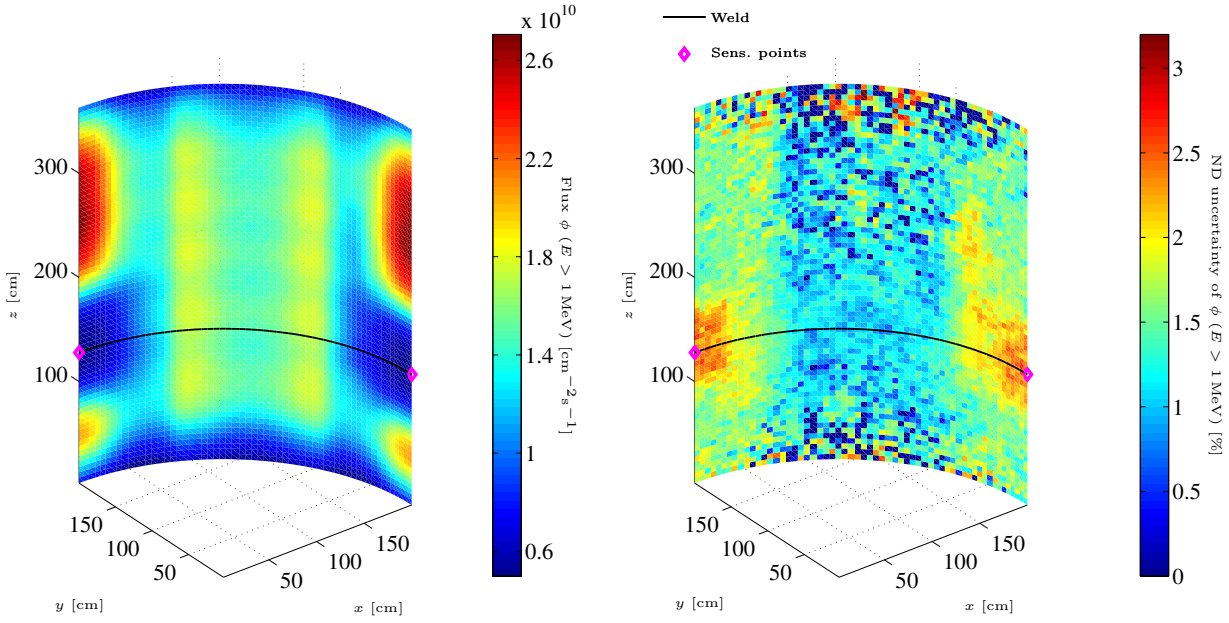

Figure 10: High energy neutron flux at the reactor pressure vessel in the SFA model, and the corresponding propagated ^{56}Fe data uncertainty.

covariance matrices, or any other limitation of the method such as disregarding model defects (*i.e.* the disagreement between reality and the nuclear models using the best possible parameters), can cause the practically constant likelihood functions by effectively overestimating the experimental uncertainties.

5.2. Cases with few significant files

In other cases than those discussed in Sec. 5.1, the file weights have a very large impact on the results; only a small number of files get significant weights. This may give the result that the ND uncertainty is estimated to be zero. Actually, the cases that seem to be most sensitive to variation of the EI parameters are such cases, and what appears to be strong sensitivities can simply be a result of sensitivity to random fluctuations – if a change in an EI parameter gives a change in the weight of one of the very few significantly "heavy" files, this may cause large changes in the estimates of both central value and ND uncertainty, see *e.g.* Fig. 4. Therefore, one cannot draw any conclusions on the actual sensitivity to the EI parameters of the estimates for these systems.

Now, there can be several reasons for the few significant files. Assuming that the nuclear reaction models are correct, there are two possibilities:

- The parameters may be sampled from a region of the parameter space where the likelihood is low, such that all files get large χ^2-values, which leads to large absolute differences between the different χ^2-values which in turn leads to very large relative differences between the likelihood function values, due to the exponential behavior of the likelihood function, see Eq. (1).
- The parameters are sampled from a "too" large region of the parameter space, giving a bad resolution in the region where the likelihood is large.

In other words, the results are poorly converged; this may be resolved by having a *much* larger set of random ND files, but this is a time consuming solution. Instead, the parameter distribution to sample from may have to be determined using feedback from the likelihood function values, possibly using a Markov Chain Monte Carlo algorithm such as the Metropolis-Hastings algorithm, as is done for UMC-G in [40].

However, the above assumed that the nuclear reaction models are perfect, and this cannot be said to be the case. Even with the best possible set of model parameters, there can still be a mismatch between theory and experiments which is so large that a problem similar the former of the two above applies. The observed cases with very few significant file weights are therefore another argument to prioritize the treatment of model defects in the future development of the methodology, possibly inspired by Refs. [41, 42].

The few significantly weighted files could be helped by a practical remedy used in Refs [15] and [18]: to normalize the χ^2_k used to compute the likelihood in 1 by the minimum χ^2_k. The statistical interpretation of the uncertainties resulting from this is unclear, however. Normalization by a number related to the *reduced* χ^2 could also be considered, since we would expect this to be close to 1 for experimental data consistent to the model predictions. This is, however, outside the scope of the present work.

5.3. Saturation effect when increasing correlated uncertainty

In Sec. 4.2.1 it was observed that the likelihood functions for each file seem to converge towards some limits as $\sigma_{\text{fully correlated}}$ increases. This may come as a surprise, since one normally expects the impact of the experiments to decrease as their uncertainties increase. This can be explained by the fact that even if a larger $\sigma_{\text{fully correlated}}$ allows for a larger offset between theory and experiments, it *does not allow* for a larger deviation in *shape* for each reaction channel. An example in Appendix B illustrates this.

Similar reasoning may also help to understand the narrowing of the weight distributions with increasing $\sigma_{\text{fully correlated}}$ which also is observed in Sec. 4.2.1 for some cases. Imagine a theoretical curve which has a shape which agrees really good with the experiments, but at the same time a large offset with respect to the experiments. This file will have a low weight as long as the correlated uncertainties are moderate. As $\sigma_{\text{fully correlated}}$ increases, all files get larger likelihood functions, but as $\sigma_{\text{fully correlated}}$ continues to increase, more and more files have their likelihood functions saturated since their shape only agrees up to a certain extent. However, the likelihood function for the file with the good shape and the large offset will continue to increase longer than the average file and possibly it will end up being the file with the by far greatest weight. Since the weights are studied relatively to the maximum weight, this can lead to a decreasing average weight, which also is observed in the example in Appendix B.

5.4. Outliers and rejection of energy zones

Outlying experiments are identified by the deviation to the data in the random files, and as pointed out in Sec. 2.5, this is not appropriate. Automatic rejection or down-weighting of experimental data is difficult and (if possible at all) it should be based on deviation between different experiments and not between experiments and theory, since erroneous theory never can be disproved using such a scheme. Possibly, a development towards a more sound automatic treatment of outliers could be inspired by Ref. [43]. If it is found necessary to retreat to some manual selection of experiments, this should be done using clear and transparent rules to reduce personal judgment to a minimum and enable reproducible results.

The rejection of energy zones used in this work suffers from the same problem as the identification and treatment

of outliers, but the motivation for it has a different nature; it is rather the defects in the model than unknown systematic experimental errors that cause the deviation in this case. This makes the rejection of energy zones even more problematic than the rejection of experiments, and it shall be abandoned in future studies. To reject an energy zone if *half* of the experimental subsets are judged as outliers also introduces some extra arbitrariness which is not investigated in this work. The rejection can partially be replaced by covering a greater part of the parameter space in the prior distribution of the random files, but it may prove necessary to also use experimental data more directly, for example along the lines of Ref. [42].

5.5. Sensitivity to the EXFOR interpretation parameters

It is seen in Sec. 4 that many of the different EI parameters do not have a very strong impact on the propagated ND uncertainties nor on the distribution of the weights, as long as they are within reasonable ranges. Many applications are generally insensitive to the weighting (as discussed in Sec. 5.1), why the effects of these parameters may not show, but in many cases the distributions of the weights seem not to be affected strongly either, judging from the average weights. The cases which appear as most sensitive to the EI parameters are primarily the thermal systems with ^{239}Pu or ^{235}U varied. However (as discussed in Sec. 5.2), there are very few significant weights for the random files in these cases, and therefore the results are very sensitive to random fluctuations and no conclusions can be drawn regarding the actual sensitivity to the EI parameters for the propagated results.

By comparing the results of this work with the results in Ref. [19] where several of the EI parameters had other default values and the outliers and energy zones were treated somewhat differently, one finds really substantial differences. For example, *many* more ^{235}U random files obtain a significant weight using the rules of Ref. [19]. Thus, even if the results seem not to be that sensitive to the choice of one single EI parameter at a time, they indeed have an impact as a group and they should be discussed and motivated more carefully in the future. Also, some of them should be replaced by more accurate estimates of corresponding uncertainties and correlations; for example, the uncertainty correlated over each reaction channel should at least partially be replaced by an analysis of the true correlations (identifying experiments using the same monitor cross section, *etc.*). Also, one could obtain a statistical distribution of the uncertainties of all measurements ever made for a particular reaction channel and use that to assign uncertainties to incomplete data. Such information is available in the EXFOR validation of Ref. [44].

5.6. Statistical significance of differences

In a study like this, where one is looking for differences between several different pairs of observations, there is a risk of finding "statistically significant" differences even though they are due to normal statistical fluctuations. For example, comparing N independent pairs with a type 1 error rate of α (significance level of $1-\alpha$) in each comparison, can give a type 1 error rate of up to $1-(1-\alpha)^N$ (the Bonferroni inequality [45]). Simply phrased, some significant deviations will be observed by chance if you study enough cases. In classical Analysis of Variance (ANOVA), such pairwise comparisons are performed using *e.g.* Tukey's method [45] which handles the inflated error rate correctly. In this study, there are strong correlations between different pairs (*e.g.* for the propagated ND uncertainty compared for $\sigma_{\text{fully correlated}} = 0\,\%$ and $\sigma_{\text{fully correlated}} = 1\,\%$ versus $\sigma_{\text{fully correlated}} = 0\,\%$ and $\sigma_{\text{fully correlated}} = 3\,\%$) and there are attempts to identify trends, and therefore the inflated error rate is taken into account quite informally by stating that a difference is statistically significant if the values ± 2 standard deviations do not overlap (which would correspond to $\alpha \approx 0.5\,\%$ for each comparison assuming normal distributions and similar uncertainties).

5.7. Miscellaneous comments

- As briefly discussed in Ref. [19], the efficiency of the methodology can be improved using a Russian Roulette methodology, randomly discarding light-weight random files before simulations without introducing a bias, similar to the removal of light-weight particles in Monte Carlo transport codes. Such a methodology can also be used to obtain implicitly weighted random files which all have the same weight, which simplifies the understanding for an end user which would not have to bother about weights.

- No renormalization of EXFOR has been applied. That is, cross sections that are measured relatively to some reference cross section have been used as they stand in EXFOR, even though it is a good idea to obtain the original ratios and compute new values using the currently best estimate for the reference cross section. Such a renormalization would probably reduce the overall *error* in data but it would also introduce even more experimental correlations. However, if the renormalization is handled, the resulting correlations can be estimated simultaneously.

- As can be seen in Table 3 and in Sec. 4.4, the estimated ND uncertainty for the fast flux for the SFA is greater using partial variations than using variation of the full random files, *e.g.*, $3.1 \pm 0.2\,\%$ compared to $2.5 \pm 0.2\,\%$. In other words, the uncertainty is greater when less ND is varied, which can be surprising. It would rather be expected to get very similar results as with full variations since only flux with neutron energy $E \geq 1\,\text{MeV}$ is included, and the partial variations consider all cross sections with such energies.

  Considering the uncertainty of the uncertainty, one can argue that the difference is likely to be this large

by chance. However, the same ND random files are used in both cases, so even if the uncertainty of the estimates is "large enough", the values are strongly correlated. The uncertainty of the uncertainty which can cause differences between these values by chance should be the part of the uncertainty which is due the to Monte Carlo code statistics. Taking a closer look at the computations it turns out that the uncertainty of uncertainty from Monte Carlo code statistics is less than 1/3 of the total uncertainty of uncertainty, making it much less likely that the observed difference is only by chance.

A possible explanation is a compensating effect in the angular distributions which is excluded when not varying the angular distributions, *i.e.*, ^{56}Fe cross sections which give better shielding (typically higher cross sections) are correlated to angular distributions which give worse shielding (more directed in the forward direction), and vice versa. When this negative correlation is removed by fixing the angular distributions, the ND uncertainty increases.

- From comparing the results with partial variation and variation of full random files, it is clear that other nuclear data than cross section data below or above the resonance range (varied in the partial variations) can have a large impact in some cases, especially for ELECTRA and the pin cell varying ^{238}U. In particular, the treatment of resonance parameter distributions shall be revisited to ensure a statistically rigorous treatment, also using other methodologies than file weighting since this may be very inefficient due to the large number of resonance parameters.

6. Conclusions and outlook

The study presented in this paper further explores the incorporation of experimental data into TMC by assigning weights to the random files as suggested in Ref. [19], *i.e.*, by combining TMC and UMC-B [17]. Experimental covariance matrix estimates are automatically generated from EXFOR using a set of reviewable rules, which enables taking estimated experimental correlations into account in the computation of weights on a large scale. The rules include, *e.g.*, a coarse guess for typical cross-experimental correlations and a translation of the uncertainty due to the combination of energy resolution and energy uncertainty into cross section uncertainty. The sensitivities to certain different choices made in the interpretation of EXFOR are also studied.

A thorough explanation of the file weighting methodology is provided, as well as a proof that it is a consistent implementation of Bayes' theorem. Further, a general way to estimate the uncertainty of the uncertainty, taking file weights into account, is suggested and implemented.

The methodology is applied to assess propagated ND uncertainties in the integral results for several applications, covering both fast and thermal systems. One of the considered systems is the shielding fuel assemblies which are designed to protect the reactor pressure vessel at Ringhals 3 and 4 in Sweden from the fast neutron flux – an application which has not previously been studied with respect to ND uncertainty.

The results do depend on the choices made for the EXFOR interpretation and the effects that are observed are understood and explained, but the sensitivity seems not to be very strong from what can be seen in these results. This can partially be because some of the applications are generally insensitive to the weighting applied here, which in turn can be explained by that the used random files are not intended to describe prior distributions and cover a too small part of the parameter space. For other nuclides and energy ranges, only a very few weights are significantly large, which gives practical problems with resolution and random fluctuations (bad convergence). The bad convergence can also be due to the parameter distributions used as priors or due to model defects. It is either necessary to generate new prior distributions or (more likely) to use feedback from the likelihood directly to search for a parameter distribution, possibly using a Markov Chain Monte Carlo algorithm. Further, model defects need to be taken into account, possibly along the lines of Ref. [42]. Also, the probability distribution of the resonance parameters shall be studied.

The results are strongly affected by the automatic rejection of outliers, and this treatment needs to be further developed and investigated; not the least, it is important that experimental data is not rejected on the basis of deviation to (possibly erroneous) models. The related rejection of energy zones shall be abandoned in favor of more motivated parameter distributions and a sound treatment of model defects. This rejection may, however, help us to identify for which reaction channels and energy ranges the nuclear models and experiments agree the least, *i.e.*, which parts of the nuclear modeling that needs the most improvement, see Tables 4 to 7.

It is further observed that the effect of increasing a fully correlated uncertainty becomes saturated – all weights converge towards certain limits. This effect is explained by that the fully correlated uncertainty allows for larger offsets between theory and experiments but not for larger deviation in shape.

The uncertainty due to ^{56}Fe data in the flux at the sensitive points in the Shielding Fuel Assemblies from Ref. [38] is $2.5 \pm 0.2\,\%$, not using any weights but with the variation of the full random files. These values are not affected much by the weights, why there is reason to consider the possibility that a too small part of the parameter space is covered and that the true uncertainties are greater.

In summary, we conclude that to make the best use of the wealth of information in EXFOR database, a continuous research effort has to be pursued, both regarding better modeling and by establishing better motivated experimental covariance information. This will allow us to

approach transparent, highly automatized and statistically rigorous ND uncertainty estimates.

Acknowledgements

The authors would like to thank Klaes-Håkan Bejmer and John Loberg at Vattenfall Nuclear Fuel, Solna, Sweden, for providing the MCNP-6 model for the Shielding Fuel Assemblies (see Sec. 4.4), and for generously answering questions regarding this. Further, we would like to thank Denise Neudecker at Los Alamos National Lab, New Mexico, for interesting and fruitful discussions on the topics of the article.

Funding for this work was received from Uppsala University, Nuclear Research and Consultancy Group NRG, and the Swedish Center for Nuclear Technology SKC.

Appendix A. Eq. (7) gives a consistent estimate

This whole section adapts the notation of Sec. 2. It is important to motivate why an estimator, such as $\widehat{q^j}$ in Eq. (7), is "good", in some sense. For example, for an estimator θ_n^* of some quantity θ which depends on the sample size n, it can be considered important that the estimator is consistent, *i.e.*, that θ_n^* converges in probability towards θ as $n \to \infty$, meaning that for each $\varepsilon > 0$, $P(|\theta_n^* - \theta| > \varepsilon) \to 0$ as $n \to \infty$ [46].

To show consistency of an estimator θ_n^* of a quantity θ, depending on the sample size n, it suffices to show [46]

- that the estimate is asymptotically unbiased, *i.e.* that
$$\lim_{n\to\infty} \langle \theta_n^* \rangle = \theta$$
- and that the variance of the estimator approaches zero with increasing sample size, *i.e.*,
$$\lim_{n\to\infty} V(\theta_n^*) = 0.$$

In Sec. Appendix A.1, both these features of $\widehat{q^j}$ in Eq. (7) are motivated in an intuitive, but mathematically dubious, way. In Sec. Appendix A.3, the asymptotic unbiasedness is formally proven. Together with Sec. Appendix A.4, where it is shown that $\lim_{n\to\infty} V(\theta_n^*) = 0$, this shows consistency according to the above.

In the formal proof, it is assumed that $V\left(\left[q^j(\mathbf{p})\right]^2\right)$ is finite and that the expected values $\left\langle \left[q^j(\mathbf{p})\right]^2 \right\rangle$ and $\left\langle q^j(\mathbf{p}) \right\rangle_{f(\mathbf{p}|\mathbf{x})}$ exist, where the latter is the expected value of $q^j(\mathbf{p})$ given $\mathbf{x}$, *i.e.* with respect to the PDF $f(\mathbf{p}|\mathbf{x})$. One sufficient condition for all these assumptions is that $q(\mathbf{p})$ can be bounded below and above. In the example $q = k_{\text{eff}}$, this is easily seen to be the case from physics, since $k_{\text{eff}} \geq 0$ and limited above *at least* by the number of neutrons in a nuclide.

Appendix A.1. Informally motivating that the estimate is consistent

With $w_k = CL(\mathbf{p}^{(k)};\mathbf{x})$ and $C = 1/\sum_{\kappa=1}^{n} L(\mathbf{p}^{(\kappa)};\mathbf{x})$, Eq. (7) can be rewritten as

$$\widehat{q^j(\mathbf{p})} = \sum_{k=1}^{n} CL(\mathbf{p}^{(k)};\mathbf{x}) q^j\left(\mathbf{p}^{(k)}\right). \tag{A.1}$$

Assuming that n is large enough so that C can be assumed constant with respect to a particular $\mathbf{p}^{(k)}$ (this is the mathematically dubious step), and using that the $\mathbf{p}^{(k)}$ are sampled from $f_0(\mathbf{p})$, the expected value of $\widehat{q^j}$ becomes

$$\left\langle \widehat{q^j} \right\rangle = C \sum_{k=1}^{n} \int L(\mathbf{p}^{(k)};\mathbf{x}) q^j(\mathbf{p}^{(k)}) f_0(\mathbf{p}^{(k)})\, \mathrm{d}\mathbf{p}^{(k)} = Cn \int L(\mathbf{p};\mathbf{x}) q^j(\mathbf{p}) f_0(\mathbf{p})\, \mathrm{d}\mathbf{p}. \tag{A.2}$$

Since $Cn = n/\sum_{\kappa=1}^{n} L(\mathbf{p}^{(\kappa)};\mathbf{x}) \to 1/\langle L(\mathbf{p};\mathbf{x})\rangle$ as $n \to \infty$, this expression approaches unity for $j = 0$ since $f_0(\mathbf{p})$ is a PDF. In other words, $CnL(\mathbf{p};\mathbf{x})f_0(\mathbf{p})$ approaches a PDF which obviously is proportional to $L(\mathbf{p};\mathbf{x})f_0(\mathbf{p})$, *i.e.*

$$CnL(\mathbf{p};\mathbf{x})f_0(\mathbf{p}) \to f(\mathbf{p}|\mathbf{x}) \text{ as } n \to \infty, \tag{A.3}$$

giving $\left\langle \widehat{q^j} \right\rangle \to \left\langle q^j \right\rangle_{f(\mathbf{p}|\mathbf{x})}$, where the subscript $f(\mathbf{p}|\mathbf{x})$ denotes that this is the expected value of q^j with respect to the PDF $f(\mathbf{p}|\mathbf{x})$. In words, $\widehat{q^j}$ is an asymptotically unbiased estimate of the j'th moment of q given $\mathbf{x}$.

Simultaneously, the variance of $\widehat{q^j}$ approaches 0 analogously to the variance of a simple mean value. Together with the asymptotic unbiasedness, this yields that $\widehat{q^j}$ is a consistent estimate of $\left\langle q^j \right\rangle$ [46]. □

Appendix A.2. Some prerequisites for the formal proof

The following results will be used several times in the proofs.

- The limit of a product of sequences equals the product of the corresponding limits, provided that the limits exist [47]. That is, if a_n and b_n are some sequences, and if there exist numbers A and B such that $\lim_{n\to\infty} a_n = A$ and $\lim_{n\to\infty} b_n = B$, then
$$\lim_{n\to\infty} a_n b_n = AB. \tag{A.4}$$
- Slutsky's theorem [48]: If an estimator $\hat{\theta}_n$ converges in probability to θ as $n \to \infty$, and the function g is continuous at θ and independent of n, then $g(\hat{\theta}_n)$ converges in probability to $g(\theta)$, *i.e.*,
$$\hat{\theta}_n \xrightarrow{\text{P}} \theta \Rightarrow g(\hat{\theta}_n) \xrightarrow{\text{P}} g(\theta). \tag{A.5}$$
- The squeeze theorem for sequences [49]: If, for N large enough, the sequences a_n, b_n and c_n fulfill $a_n \leq b_n \leq c_n$ for $n \geq N$, and $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L$, then
$$\lim_{n\to\infty} b_n = L. \tag{A.6}$$

Appendix A.3. Proving that the estimate is asymptotically unbiased

From the definition of the likelihood function, it can be seen that

$$0 < L\left(\mathbf{p}^{(k)};\mathbf{x}\right) \leq \frac{1}{(2\pi)^{m/2}\sqrt{\det\left(\mathbf{C}_{\mathrm{E}}\right)}} =: \alpha, \tag{A.7}$$

where "=:" indicates that α is defined by this equation. Thus,

$$\frac{1}{\alpha + \sum_{\kappa\neq k} L(\mathbf{p}^{(\kappa)};\mathbf{x})} \leq \frac{1}{\sum_{\kappa=1}^{n} L(\mathbf{p}^{(\kappa)};\mathbf{x})} \leq \frac{1}{\sum_{\kappa\neq k} L(\mathbf{p}^{(\kappa)};\mathbf{x})} \tag{A.8}$$

where $\kappa \neq k$ is short notation for $\kappa \in \{1,2,...,n\}\setminus k$. From the definition of $\widehat{q^j(\mathbf{p})}$ in Eq. (7), one thus obtains

$$\sum_{k=1}^{n} \frac{L(\mathbf{p}^{(k)};\mathbf{x})q^j(\mathbf{p}^{(k)})}{\alpha + \sum_{\kappa\neq k} L(\mathbf{p}^{(\kappa)};\mathbf{x})} \leq \widehat{q^j(\mathbf{p})} \leq \sum_{k=1}^{n} \frac{L(\mathbf{p}^{(k)};\mathbf{x})q^j(\mathbf{p}^{(k)})}{\sum_{\kappa\neq k} L(\mathbf{p}^{(\kappa)};\mathbf{x})}. \tag{A.9}$$

Taking expectation values on all sides,

$$\begin{aligned} &\sum_{k=1}^{n} \left\langle \frac{1}{\alpha + \sum_{\kappa\neq k} L(\mathbf{p}^{(\kappa)};\mathbf{x})} \right\rangle \\ &\int L(\mathbf{p}^{(k)};\mathbf{x})q^j(\mathbf{p}^{(k)})f_0(\mathbf{p}^{(k)})\,\mathrm{d}\mathbf{p}^{(k)} \leq \\ &\left\langle \widehat{q^j(\mathbf{p})} \right\rangle \leq \sum_{k=1}^{n} \left\langle \frac{1}{\sum_{\kappa\neq k} L(\mathbf{p}^{(\kappa)};\mathbf{x})} \right\rangle \\ &\int L(\mathbf{p}^{(k)};\mathbf{x})q^j(\mathbf{p}^{(k)})f_0(\mathbf{p}^{(k)})\,\mathrm{d}\mathbf{p}^{(k)}. \end{aligned} \tag{A.10}$$

The separation of the numerators and the denominators in each summand is possible since the denominators are constant with respect to $\mathbf{p}^{(k)}$ and the numerators are constant with respect to $\mathbf{p}^{(\kappa)}$ for $\kappa \neq k$. Since all terms in the summands are the same (the $\mathbf{p}^{(k)}$ are sampled from the same distribution, namely $f_0(\mathbf{p})$), one may omit the superscript (k), yielding

$$\begin{aligned} &\left\langle \frac{n}{\sum_{\kappa=1}^{n-1}\left(\frac{\alpha}{n-1} + L(\mathbf{p}^{(\kappa)};\mathbf{x})\right)} \right\rangle \int L(\mathbf{p};\mathbf{x})q^j(\mathbf{p})f_0(\mathbf{p})\,\mathrm{d}\mathbf{p} \leq \\ &\left\langle \widehat{q^j(\mathbf{p})} \right\rangle \leq \left\langle \frac{n}{\sum_{\kappa=1}^{n-1} L(\mathbf{p}^{(\kappa)};\mathbf{x})} \right\rangle \int L(\mathbf{p};\mathbf{x})q^j(\mathbf{p})f_0(\mathbf{p})\,\mathrm{d}\mathbf{p}. \end{aligned} \tag{A.11}$$

Since, by definition, $f(\mathbf{p}|\mathbf{x}) \propto L(\mathbf{p};\mathbf{x})f_0(\mathbf{p})$, the expectation value of q^j given $\mathbf{x}$ (it has been assumed to exist) is

$$\left\langle q^j \right\rangle_{f(\mathbf{p}|\mathbf{x})} = \frac{\int q^j(\mathbf{p})L(\mathbf{p};\mathbf{x})f_0(\mathbf{p})\,\mathrm{d}\mathbf{p}}{\int L(\mathbf{p};\mathbf{x})f_0(\mathbf{p})\,\mathrm{d}\mathbf{p}} = \frac{\int q^j(\mathbf{p})L(\mathbf{p};\mathbf{x})f_0(\mathbf{p})\,\mathrm{d}\mathbf{p}}{\langle L(\mathbf{p};\mathbf{x})\rangle}, \tag{A.12}$$

where the denominator is necessary to normalize $L(\mathbf{p};\mathbf{x})f_0(\mathbf{p})$ into a PDF. Solving for the integral in the numerator, inserting the result into Eq. (A.11) and rewriting n as $n(n-1)/(n-1)$, one obtains

$$\begin{aligned} &\frac{n}{n-1}\left\langle \frac{n-1}{\sum_{\kappa=1}^{n-1}\left(\frac{\alpha}{n-1} + L(\mathbf{p}^{(\kappa)};\mathbf{x})\right)} \right\rangle \langle L(\mathbf{p};\mathbf{x})\rangle \left\langle q^j \right\rangle_{f(\mathbf{p}|\mathbf{x})} \leq \\ &\left\langle \widehat{q^j(\mathbf{p})} \right\rangle \leq \frac{n}{n-1}\left\langle \frac{n-1}{\sum_{\kappa=1}^{n-1} L(\mathbf{p}^{(\kappa)};\mathbf{x})} \right\rangle \langle L(\mathbf{p};\mathbf{x})\rangle \left\langle q^j \right\rangle_{f(\mathbf{p}|\mathbf{x})}, \end{aligned} \tag{A.13}$$

Now, since the mean value of a random sample is a consistent estimate of the expectation value of the underlying random variable [46], Slutsky's theorem (Eq. (A.5)) implies that

$$\lim_{n\to\infty}\left\langle \frac{1}{\bar{Y}_n} \right\rangle = \frac{1}{\langle Y\rangle}, \tag{A.14}$$

where $\bar{Y}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$ and the Y_k are independent and distributed as Y. Consequently, by identifying the expressions as expectation values of reciprocals of mean values, it can be seen that

$$\begin{aligned} &\lim_{n\to\infty}\left\langle \frac{n-1}{\sum_{\kappa=1}^{n-1}\left(\frac{\alpha}{n-1} + L(\mathbf{p}^{(\kappa)};\mathbf{x})\right)} \right\rangle = \\ &\lim_{n\to\infty} \frac{1}{\left\langle \frac{\alpha}{n-1} + L(\mathbf{p};\mathbf{x}) \right\rangle} = \frac{1}{\langle L(\mathbf{p};\mathbf{x})\rangle}, \end{aligned} \tag{A.15}$$

and

$$\lim_{n\to\infty}\left\langle \frac{n-1}{\sum_{\kappa=1}^{n-1}\left(L(\mathbf{p}^{(\kappa)};\mathbf{x})\right)} \right\rangle = \frac{1}{\langle L(\mathbf{p};\mathbf{x})\rangle}. \tag{A.16}$$

Applying these equalities to the inequalities of Eq. (A.13), the squeeze theorem for sequences (Eq. (A.6)) gives

$$\lim_{n\to\infty}\left\langle \widehat{q^j(\mathbf{p})} \right\rangle = \left\langle q^j \right\rangle_{f(\mathbf{p}|\mathbf{x})}, \tag{A.17}$$

also using that $\lim_{n\to\infty}\frac{n}{n-1} = 1$ and the result in Eq. (A.4). Thus, the estimate in Eq. (7) is asymptotically unbiased. □

Appendix A.4. Proving that the variance approaches zero

It is well-known that [46]

$$C(Y^2, Z^2) = \left\langle Y^2Z^2 \right\rangle - \left\langle Y^2 \right\rangle \left\langle Z^2 \right\rangle, \tag{A.18}$$

where $C(Y^2, Z^2)$ denotes the covariance of the random variables Y^2 and Z^2. Further [46],

$$V(YZ) = \left\langle Y^2Z^2 \right\rangle - \left\langle YZ \right\rangle^2 \leq \left\langle Y^2Z^2 \right\rangle. \tag{A.19}$$

Solving Eq. (A.18) for $\left\langle Y^2Z^2 \right\rangle$ it is thus seen that the variance of the product of two possibly correlated random variables Y and Z can be bounded as

$$\begin{aligned} V(YZ) &\leq C(Y^2, Z^2) + \left\langle Y^2 \right\rangle \left\langle Z^2 \right\rangle \leq = \\ &\sqrt{V(Y^2)V(Z^2)} + \left\langle Y^2 \right\rangle \left\langle Z^2 \right\rangle, \end{aligned} \tag{A.20}$$

noting that $C(Y^2, Z^2) \leq \sqrt{V(Y^2)V(Z^2)}$ [46]. Substituting Y to w_k and Z to $q^j\left(\mathbf{p}^{(k)}\right)$ (from the summand of Eq. (7)) yields

$$V\left(w_k q^j\left(\mathbf{p}^{(k)}\right)\right) \leq \sqrt{V(w_k^2)V\left(\left[q^j\left(\mathbf{p}^{(k)}\right)\right]^2\right)} + \left\langle w_k^2\right\rangle \left\langle \left[q^j\left(\mathbf{p}^{(k)}\right)\right]^2\right\rangle. \tag{A.21}$$

Now,

$$0 \leq \left\langle w_k^2\right\rangle = \left\langle \left(\frac{L\left(\mathbf{p}^{(k)}\right)}{\sum_{\kappa=1}^{n} L\left(\mathbf{p}^{(\kappa)}\right)}\right)^2\right\rangle \leq \left\langle \left(\frac{L\left(\mathbf{p}^{(k)}\right)}{\sum_{\kappa\neq k} L\left(\mathbf{p}^{(\kappa)}\right)}\right)^2\right\rangle = \left\langle \frac{1}{\left(\sum_{\kappa\neq k} L\left(\mathbf{p}^{(\kappa)}\right)\right)^2}\right\rangle \left\langle \left[L\left(\mathbf{p}^{(k)}\right)\right]^2\right\rangle, \tag{A.22}$$

again using that the numerator depends only on $\mathbf{p}^{(k)}$ and the denominator on $\mathbf{p}^{(\kappa)}$ with $\kappa \neq k$. Multiplying with n yields

$$0 \leq n\left\langle w_k^2\right\rangle \leq \left\langle \left(\frac{n-1}{\sum_{\kappa\neq k} L\left(\mathbf{p}^{(\kappa)}\right)}\right)^2\right\rangle \frac{n\left\langle \left[L\left(\mathbf{p}^{(k)}\right)\right]^2\right\rangle}{(n-1)^2}, \tag{A.23}$$

also multiplying both numerator and denominator with $(n-1)^2$ to obtain the expected value of the reciprocal of a squared mean value of the $L(\mathbf{p}^{(\kappa)}; \mathbf{x})$. With a similar argument as in connection with Eqs. (A.14) and (A.15), Slutsky's theorem (Eq. (A.5)) gives

$$\lim_{n\to\infty} \left\langle \left(\frac{n-1}{\sum_{\kappa\neq k} L\left(\mathbf{p}^{(\kappa)}\right)}\right)^2\right\rangle = \frac{1}{\left\langle L(\mathbf{p};\mathbf{x})\right\rangle^2}. \tag{A.24}$$

The expected values $\langle L(\mathbf{p})\rangle$ and $\left\langle [L(\mathbf{p})]^2\right\rangle$ exist since $L(\mathbf{p})$ is positive and bounded above (easily seen from the definition of expected value). Thus, inserting Eq. (A.24) into Eq. (A.23), using Eq. (A.4) and applying the squeeze theorem for sequences (Eq. (A.6)) gives

$$\lim_{n\to\infty} n\left\langle w_k^2\right\rangle = 0. \tag{A.25}$$

Analogously,

$$0 \leq n^2\left\langle w_k^4\right\rangle = n^2\left\langle \left(\frac{L\left(\mathbf{p}^{(k)}\right)}{\sum_{\kappa=1}^{n} L\left(\mathbf{p}^{(\kappa)}\right)}\right)^4\right\rangle \leq n^2\left\langle \left(\frac{L\left(\mathbf{p}^{(k)}\right)}{\sum_{\kappa\neq k} L\left(\mathbf{p}^{(\kappa)}\right)}\right)^4\right\rangle = \left\langle \left(\frac{n-1}{\sum_{\kappa\neq k} L\left(\mathbf{p}^{(\kappa)}\right)}\right)^4\right\rangle \frac{n^2\left\langle \left[L\left(\mathbf{p}^{(k)}\right)\right]^4\right\rangle}{(n-1)^4}, \tag{A.26}$$

such that Slutsky's theorem (Eq. (A.5)), the squeeze theorem (Eq. (A.6)) and Eq. (A.4) give

$$\lim_{n\to\infty} n^2\left\langle w_k^4\right\rangle = 0. \tag{A.27}$$

Observing that

$$0 \leq n^2 V(w_k^2) = n^2\left(\left\langle w_k^4\right\rangle - \left\langle w_k^2\right\rangle^2\right) \leq n^2\left\langle w_k^4\right\rangle, \tag{A.28}$$

one can conclude from the squeeze theorem (Eq. (A.6)) that

$$\lim_{n\to\infty} n^2 V(w_k^2) = 0. \tag{A.29}$$

Multiplying Eq. (A.21) with n gives

$$nV\left(w_k q^j\left(\mathbf{p}^{(k)}\right)\right) \leq \sqrt{n^2 V(w_k^2)V\left(\left[q^j\left(\mathbf{p}^{(k)}\right)\right]^2\right)} + n\left\langle w_k^2\right\rangle \left\langle \left[q^j\left(\mathbf{p}^{(k)}\right)\right]^2\right\rangle. \tag{A.30}$$

Since it is assumed that $V\left(\left[q^j\left(\mathbf{p}^{(k)}\right)\right]^2\right)$ is finite and that $\left\langle \left[q^j\left(\mathbf{p}^{(k)}\right)\right]^2\right\rangle$ exists, the results in Eqs. (A.25) and (A.29), together with the squeeze theorem (Eq. (A.6)) and Eq. (A.4), yield

$$\lim_{n\to\infty} nV\left(w_k q^j\left(\mathbf{p}^{(k)}\right)\right) = 0, \tag{A.31}$$

for any $k \in \{1, 2, ..., n\}$. Thus, taking the variance of Eq. (7) gives

$$\lim_{n\to\infty} V\left(\widehat{q^j(\mathbf{p})}\right) = \lim_{n\to\infty} \sum_{k=1}^{n} V\left(w_k q^j\left(\mathbf{p}^{(k)}\right)\right) = \lim_{n\to\infty} nV\left(w_k q^j\left(\mathbf{p}^{(k)}\right)\right)\Big|_{k\in\{1,2,...,n\}} = 0. \tag{A.32}$$

□

Appendix B. An example illustrating the saturation effect

In this section, the likelihood function is here computed analytically in the case of two experimental points, *i.e.* using a 2×2 experimental covariance matrix, with $\sigma_{\text{fully correlated}}$ seen as a variable, here written as $\sigma_{\text{f.c.}}$ to save space. Further, the solution is applied to a simple special case, where three linear "theoretical curves" are compared.

We begin by describing the simple special case: Consider a situation with two experimental points, X_1 and X_2, for energies $E = 1\,\text{eV}$ and $E = 2\,\text{eV}$, centered about $\mu_1 = 10\,\text{mb}$ and $\mu_2 = 20\,\text{mb}$, respectively, and with random uncertainties of $\sigma_1 = \sigma_2 = 10\,\text{mb}$ each. Further, there is a fully correlated uncertainty $\sigma_{\text{f.c.}}$ relative to the central experimental values (which is variable here). We have three candidates to theoretical curves which we want to assign weights to, making use of X_1 and X_2, namely

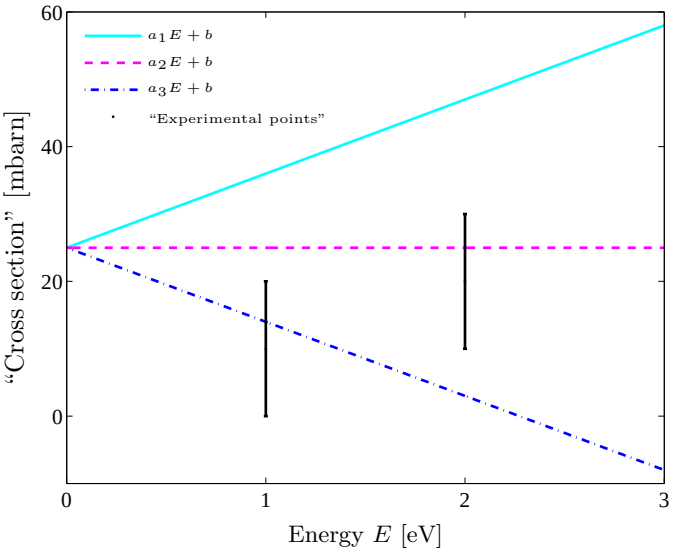

Figure B.12: Fictive experimental points and theoretical cross section curves as functions of neutron energy. The error bars denote the random uncertainty only.

$\varsigma_k(E) = a_kE + b$, where $a_1 = 11\,\text{mb/eV}, a_2 = 0, a_3 = -11\,\text{mb/eV}$ and $b = 25\,\text{mb}$. The three curves and the two experimental points are seen in Fig. B.12. As one can see, the curve $\varsigma_1(E)$ is furthest away from the experimental points but the shape of this curve agrees better with how the experiments are related to each other.

Denoting the reference values for the relative uncertainty c_1 and c_2 (in the code, these are the average theoretical values τ_1 and τ_2), respectively, the experimental covariance matrix becomes

$$\mathbf{C}_\text{E} = \begin{pmatrix} \sigma_1^2 + c_1^2\sigma_\text{f.c.}^2 & c_1c_2\sigma_\text{f.c.}^2 \\ c_1c_2\sigma_\text{f.c.}^2 & \sigma_2^2 + c_2^2\sigma_\text{f.c.}^2 \end{pmatrix}. \tag{B.1}$$

If the inverse of $\mathbf{C}_\text{E}$ exists, it is simple to find since it is a 2×2 matrix. We start by finding the determinant, *i.e.* [50],

$$\det(\mathbf{C}_\text{E}) = (\sigma_1^2 + c_1^2\sigma_\text{f.c.}^2)(\sigma_2^2 + c_2^2\sigma_\text{f.c.}^2) - (c_1c_2\sigma_\text{f.c.}^2)^2 = \sigma_1^2\sigma_2^2 + \sigma_\text{f.c.}^2(c_1^2\sigma_2^2 + c_2^2\sigma_1^2). \tag{B.2}$$

Since the determinant is non-zero if (and only if) $\sigma_1 \neq 0$ or $\sigma_2 \neq 0$, $\mathbf{C}_\text{E}^{-1}$ exists under this condition [50], in which case [50]

$$\mathbf{C}_\text{E}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 + \sigma_\text{f.c.}^2(c_1^2\sigma_2^2 + c_2^2\sigma_1^2)} \begin{pmatrix} \sigma_2^2 + c_2^2\sigma_\text{f.c.}^2 & -c_1c_2\sigma_\text{f.c.}^2 \\ -c_1c_2\sigma_\text{f.c.}^2 & \sigma_1^2 + c_1^2\sigma_\text{f.c.}^2 \end{pmatrix}. \tag{B.3}$$

With an arbitrary difference vector $\mathbf{x}-\boldsymbol{\tau} = \mathbf{d} = (d_1, d_2)^\text{T} = (x_1 - \tau_1, x_2 - \tau_2)^\text{T}$, we get that

$$\chi^2 = \mathbf{d}^\text{T}\mathbf{C}_\text{E}^{-1}\mathbf{d} = \frac{1}{\sigma_1^2\sigma_2^2 + \sigma_\text{f.c.}^2(c_1^2\sigma_2^2 + c_2^2\sigma_1^2)} \begin{pmatrix} d_1 & d_2 \end{pmatrix} \begin{pmatrix} \sigma_2^2 + c_2^2\sigma_\text{f.c.}^2 & -c_1c_2\sigma_\text{f.c.}^2 \\ -c_1c_2\sigma_\text{f.c.}^2 & \sigma_1^2 + c_1^2\sigma_\text{f.c.}^2 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \frac{\sigma_2^2d_1^2 + \sigma_1^2d_2^2 + \sigma_\text{f.c.}^2(c_2d_1 - c_1d_2)^2}{\sigma_1^2\sigma_2^2 + \sigma_\text{f.c.}^2(c_1^2\sigma_2^2 + c_2^2\sigma_1^2)}. \tag{B.4}$$

Now, consider the two extremes $\sigma_\text{f.c.} = 0$ and $\sigma_\text{f.c.} \to \infty$. For $\sigma_\text{f.c.} = 0$,

$$\chi^2 = \frac{\sigma_2^2d_1^2 + \sigma_1^2d_2^2}{\sigma_1^2\sigma_2^2} = \sum_{i=1}^{2}\frac{d_i^2}{\sigma_i^2},$$

i.e. one retains the expression for "the independent χ^2", *cf.*, Eq. (3).

For $\sigma_\text{f.c.} \to \infty$,

$$\lim_{\sigma_\text{f.c.}\to\infty}\chi^2(\sigma_\text{f.c.}) = \frac{(c_2d_1 - c_1d_2)^2}{c_1^2\sigma_2^2 + c_2^2\sigma_1^2},$$

i.e., a limit is indeed approached.

With the numerical values of the special case,

$$\lim_{\sigma_\text{f.c.}\to\infty}\chi^2(\sigma_\text{f.c.}) = \frac{(d_1 - d_2)^2}{200}.$$

and with $d_i^{(k)}$ being the difference between the i'th experimental point and the k'th "theoretical curve", $d_1^{(1)} = -26, d_2^{(1)} = -27, d_1^{(2)} = -15, d_2^{(2)} = -5, d_1^{(2)} = -4$ and $d_2^{(2)} = 17$, giving

$$\lim_{\sigma_\text{f.c.}\to\infty}\chi_1^2(\sigma_\text{f.c.}) = \frac{1}{200},$$

$$\lim_{\sigma_\text{f.c.}\to\infty}\chi_2^2(\sigma_\text{f.c.}) = \frac{100}{200},$$

$$\lim_{\sigma_\text{f.c.}\to\infty}\chi_3^2(\sigma_\text{f.c.}) = \frac{441}{200}.$$

This yields that the weights approach $w_1(\infty) = 1, w_2(\infty) = e^{-99/400} \approx 0.7808$ and $w_3(\infty) = e^{-440/400} \approx 0.3329$, for curve 1, 2 and 3, respectively.

In Fig. B.13 the weights are plotted as functions of $\sigma_\text{f.c.}$ when $\sigma_\text{f.c.}$ ranges from 10^{-4} to 10^3. The χ^2-values are computed directly from the definition of χ^2 in Eq. (B.4), why the results can be seen as an error check; the values for $w(\sigma_\text{f.c.} \to \infty)$ are indeed correct. Also, one can see that $\sigma_\text{f.c.}$ starts having a substantial impact on the weights when it reaches a few %, and that the effect is saturated when $\sigma_\text{f.c.}$ reaches a few hundred %.

[1] R. Forrest, Advances in nuclear data, Oral presentation, PHYSOR (2014).

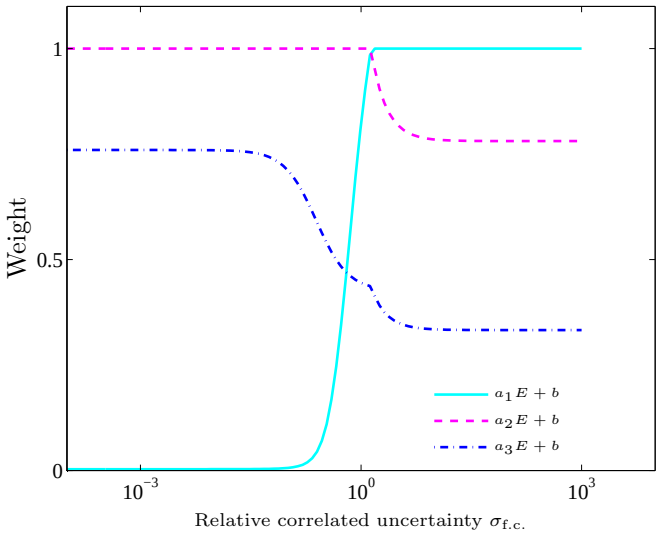

Figure B.13: The weights for the three fictive cross section curves as functions of the relative fully correlated uncertainty $\sigma_{\text{f.c.}}$. For low $\sigma_{\text{f.c.}}$, the shape of the curve is unimportant, and only the "closeness" to the experimental points matter.

[2] A. Koning, D. Rochman, Towards sustainable nuclear energy: Putting nuclear physics to work, Annals of Nuclear Energy 35 (2008) 2024–2030.

[3] A. Koning, D. Rochman, Modern nuclear data evaluation with the TALYS code system, Nuclear Data Sheets 113 (2012) 2841–2944.

[4] A. Koning, et al., TALYS-1.6, User Manual, Nuclear Research and Consultancy Group NRG (December 2013).

[5] A. Trkov, M. Herman, D. Brown, et al., ENDF-6 formats manual, Tech. Rep. BNL-990365-2009, Brookhaven National Laboratory (2011).

[6] H. Sjöstrand, et al., Total Monte Carlo evaluation for dose calculations, Radiation Protection Dosimetry 131 (2014) 312–315.

[7] P. Helgesson, D. Rochman, H. Sjöstrand, E. Alhassan, A. Koning, UO_2 vs MOX: Propagated nuclear data uncertainty for k_{eff}, with burnup, Nuclear Science and Engineering 177 (3) (2014) 321–336.

[8] D. Rochman, A. Koning, S. van der Marck, Uncertainties for criticality-safety benchmarks and k-eff distributions, Annals of Nuclear Energy 36 (2009) 810–831.

[9] S. van der Marck, D. Rochman, Nuclear data uncertainties for local power densities in the Martin-Hoogenboom benchmark, Joint International Conference on Supercomputing in Nuclear Applications + Monte Carlo 177. doi:http://dx.doi.org/10.1051/snamc/201402410.

[10] O. Cabellos, et al., Propagation of nuclear data uncertainties for PWR core analysis, Nuclear Engineering and Technology 46 (3) (2013) 299–312.

[11] D. da Cruz, et al., Propagation of nuclear data uncertainty for a control rod ejection accident using the Total Monte Carlo method, in: PHYSOR 2014 – The Role of Reactor Physics toward a Sustainable Future, 2014.

[12] E. Alhassan, H. Sjöstrand, P. Helgesson, A. Koning, M. Österlund, S. Pomp, D. Rochman, Uncertainty and correlation analysis of Lead nuclear data on reactor parameters for the european Lead cooled training reactor, Annals of Nuclear Energy 75 (2015) 26–37.

[13] R. Capote, D. Smith, A. Trkov, Nuclear data evaluation methodology including estimates of covariances, EPJ Web of Conferences 8, 04001 (2010). doi:10.1051/epjconf/20100804001.

[14] M. Rising, Quantification and propagation of nuclear data uncertainties, Ph.D. thesis, University of New Mexico (2012).

[15] E. Bauge, S. Hilaire, P. Dossantos-Uzarralde, Evaluation of the covariance matrix of neutronic cross sections with the Backward-Forward Monte Carlo method, Proc. Int. Conf. Nuclear Data for Science and Technology, April 22-27, 2007, Nice, France (2008) 259–264doi:10.1051/ndata:07339.

[16] D. Smith, A Unified Monte Carlo approach to fast neutron cross section data evaluation, Proceedings of the Eighth International Topical Meeting on Nuclear Applications and Utilization of Accelerators (AccApp'07), Pocatello, Idaho, July 29 - August 2, 2007, American Nuclear Society, LaGrange Park, IL (2007) 736 See also: Argonne National Laboratory Report ANL/NDM-166, 2008.

[17] R. Capote, D. Smith, A. Trkov, M. Meghzifene, A new formulation of the Unified Monte Carlo approach (UMC-B) and cross-section evaluation for the dosimetry reaction ^{55}Mn(n,γ)^{56}Mn, Fourteenth International Symposium on Reactor Dosimetry (ISRD-14), Bretton Woods, ASTM STP-1550, 179
See also: Journal of the ASTM International 9(4), JAI 104119 (2012), doi: 10.1520/JAI104115.

[18] A. Koning, Bayesian Monte Carlo for nuclear data evaluation, European Physics Journal A 51 (2015) 184–199.

[19] P. Helgesson, H. Sjöstrand, A. Koning, D. Rochman, E. Alhassan, S. Pomp, Incorporating experimental information in the TMC methodology using file weights, Nuclear Data Sheets 123 (2015) 214–219.

[20] O. Schwerer, EXFOR Formats Description for Users (EXFOR basics), Nuclear Data Section, IAEA (June 2008).

[21] D. Rochman, W. Zwermann, S. van der Marck, A. Koning, H. Sjöstrand, P. Helgesson, B. Krzykacz-Hausmann, Efficient use of Monte Carlo: Uncertainty propagation, Nuclear Science and Engineering 177 (3) (2014) 337–349.

[22] A. Koning, et al., TALYS-1.8, User Manual, Nuclear Research and Consultancy Group NRG (December 2015).

[23] J. Briesmeister, MCNP - a general Monte Carlo N-particle transport code, version 4C, Tech. Rep. LA-13709-M, Los Alamos National Laboratory (April 2000).

[24] A. Gut, An Intermediate Course in Probability, Springer-Verlag, New York, 1995.

[25] C. Shannon, A mathematical theory of communication, The Bell System Technical Journal (reprinted with corrections) 27 (1948) 379–523,626–656.

[26] E. Jaynes, Information theory and statistical mechanics, Physical Review 106 (1957) 620–630.

[27] F. Brown, A review of the best practices for Monte Carlo criticality calculations, in: ANS NSCD-2009, Richland, WA, 2009.

[28] N. Otuka, et al., Experimental uncertainty and covariance information in EXFOR library, EPJ Web of Conferences 27.

[29] M. Taboga, Lectures on Probability Theory and Mathematical Statistics, 2nd Edition, Amazon CreateSpace, 2012.

[30] P. Helgesson, H. Sjöstrand, A. Koning, J. Rydén, D. Rochman, E. Alhassan, S. Pomp, Sampling of systematic errors to estimate likelihood weights in nuclear data uncertainty propagation, Nuclear Instruments and Methods in Physics Research A 807 (2016) 137–149.

[31] R. MacFarlane, A. Kahler, Methods for processing ENDF/B-VII with NJOY, Nuclear Data Sheets 111 (2010) 2739–2890.

[32] J. Leppänen, Serpent – a Continuous-energy Monte Carlo Reactor Physics Burnup Calculation Code, User's Manual, VTT Technical research centre of Finland (March 2013).

[33] H. Sjöstrand, et al., Propagation of nuclear data uncertainties for ELECTRA burn-up calculations, in: Proc. of ND2013, 2013.

[34] J. Wallenius, et al., ELECTRA: European Lead cooled training reactor, Nuclear Technology 177 (2012) 303–313.

[35] D. Pelowitz, et al., MCNP6 User's Manual 2.7.0, Los Alamos National Lab. (December 2012).

[36] R. O'Dell, Bare sphere of plutonium-239 metal (4.5 at.% pu-240, 1.02 wt.% ga), Tech. rep., NEA/NSC (2002).

[37] OECD-NEA, International Handbook of Evaluated Criticality Safety Benchmark Experiments (September 2011).

[38] K.-H. Bejmer, J. Loberg, U. Sandberg, Second generation shielding assemblies – neutron flux impact on reactor pressure vessel and core design, in: Presented at PHYSOR 2014, 2014.
[39] H. Sjöstrand, et al., Propagation of nuclear data uncertainties for ELECTRA burn-up calculations, Nuclear Data Sheets 118 (2014) 527–530.
[40] R. Capote, D. Smith, An investigation of the performance of the Unified Monte Carlo method of neutron cross section data evaluation, Nuclear Data Sheets 109 (2008) 2768–2773.
[41] H. Leeb, et al., Consistent procedure for nuclear data evaluation based on modeling, Nuclear Data Sheets 109 (2008) 2762–2767.
[42] D. Rochman, A. Koning, E. Bauge, A. Plompen, From flatness to steepness: Updating TALYS covariances with experimental information, Annals of Nuclear Energy 73 (2014) 7–16.
[43] D. Muir, Treatment of discrepant data in the ZOTT99 Generalized Least Squares program, Proc. Nuclear Data Covariance Workshop, April 22-23, 1999, Brookhaven National Laboratory (2000) 11–15.
[44] A. Koning, Statistical verification and validation of the EXFOR database: (n,n'), (n,2n), (n,p), (n,a) and other neutron-induced threshold reaction cross-sections, Tech. Rep. DOC(2014)3, OECD/NEA Data Bank (2014).
[45] D. Montgomery, Design and Analysis of Experiments, 8th Edition, Wiley, 2012.
[46] G. Casella, R. Berger, Statistical Inference, 2nd Edition, Thomson Learning, 2001.
[47] R. Adams, Calculus - A Complete Course, 6th Edition, Prentice Hall, 2006.
[48] H. White, Asymptotic Theory For Econometricians, Academic Press, 2001.
[49] R. Larson, B. Edwards, Calculus of a Single Variable: Early Trancendental Functions, 5th Edition, Brooks/Cole, 2010.
[50] D. Harville, Matrix Algebra From a Statistician's Perspective, Springer, 1997.