

Marginalization methods for the production of conservative covariance on nuclear data

Pierre TAMAGNO^{1,*}

¹CEA, DES, IRESNE, DER, SPRC, LEPH,
Cadarache, F-13108 Saint Paul Les Durance, France

Abstract. The production of evaluated nuclear data consists not only in the determination of best estimate values for the quantities of interest but also on the estimation of the related uncertainties and correlations. When nuclear data are evaluated with underlying nuclear reaction models, model parameters are expected to synthesize all the information that is extracted from the experimental data they are adjusted on. When dealing with models with a small number of parameters compared to the number of experimental data points – *e.g.* in resonant cross section analysis – one sometimes faces excessively small evaluated uncertainty compared for instance with model/experimental data agreement. To solve this issue, an attempt was to propagate the uncertainty coming from experimental parameters involved in the data reduction process on the nuclear physics model parameters. It pushed experimentalists to separately supply random (statistical) and systematic uncertainties. It also pushed evaluators to include or mimic the data reduction process in the evaluation. In this way experimental parameters – also called nuisance parameters – could be used to increase evaluated parameter uncertainty through marginalization techniques. Two of these methods: Matrix and Bayesian marginalizations – respectively called sometimes Analytical and Monte-Carlo Marginalizations – that are currently used for evaluation will be discussed here and some limitations highlighted. A third alternative method, also based on a Bayesian approach but using the spectral decomposition of the correlation matrix, is also presented on a toy model, and on a simple case of resonant cross section analysis.

1 Introduction

Since the discovery of the atomic nucleus, technological applications of nuclear energy have flourished. The design and modeling of such devices rely on technological properties (geometry, compositions, etc.) but also on fundamental constants describing the properties of spontaneous or induced nuclear reactions. These constants, called *nuclear data*, are however not known up to an arbitrarily high precision and special efforts must be brought to quantify related uncertainties, generally understood as *covariance data*. Nuclear data are obtained in a process called *evaluation* that combines experimental data and theoretical models. Once evaluated, these are usually further stored in files in ENDF format [1], these latter are then processed to create libraries specific to application codes, for instance to produce effective cross sections for medium at given temperature. The calculation scheme of the application codes can even internally imply some modifications of the nuclear data, for instance by the production of self-shielded energy-averaged cross sections. Each transformation of the nuclear data should imply a corresponding transformation of the related covariance data for a consistent propagation of nuclear data uncertainty to the quantities of interest for the applications. In practice, most of these transfers are handled in terms of best-estimate – or mean values –

and *covariance matrices*, therefore focus is made here on these first two moments of the probability density function (PDF) describing nuclear data. The uncertainty propagation thus usually assumes mono-modale Gaussian distributions.

Quantifying uncertainties are a natural aspect of experimental science where for instance raw count rates follow Poisson distributions and detector response functions correlate eventual results. On the theoretical side, uncertainties can consist in “truncation errors” or in uncertain determination of the value of model parameters. The model parameter values are usually estimated from comparison with experimental data using least-square fit or within a more general Bayesian approach.

The production of “safe” covariance data has been a concern leading to several attempts to be conservative in the estimation of model parameters. A long-standing idea [2–6] has been to consider that the uncertainty on experimental setup characteristics (*i.e.* detector efficiency, target composition, etc.) – called *nuisance parameters* or *active variables*, written $\vec{\theta}$, involved in the setup response function – should be “transferred” to theoretical model parameters – called *physical parameters* or *passive variables*, written \vec{x} .

The present paper first reviews two techniques used to transfer uncertainty from $\vec{\theta}$ to \vec{x} . The first approach (hereafter called the matrix marginalization method) is shown

*e-mail: pierre.tamagno@cea.fr

to be ill-defined and does not guarantee to be conservative with respect to Bayesian posterior. The second approach (the Bayesian marginalization method) is shown to be possibly biased, namely when nuisance parameters are inaccurate. These methods have been tested on a previous study on a simple but realistic analysis of Prompt Neutron Fission Spectrum [7]. These approaches are tested here on a specifically designed toy model test-case, where they are shown to fail in producing consistent best-estimates and reliable propagated uncertainties. This test-case has inconsistent data, and is specifically design to exhibit the possible limits of the methods. The failure highlights that the two presented techniques – focusing on the transfer of nuisance parameter uncertainty – neglect the verification of agreement between posterior theory and experiment. An *ad hoc* procedure [7] is then presented in the paper, it aims at artificially increasing the estimated uncertainties but in a conservative manner with respect to the Bayesian approach and with consideration of agreement between posterior theory and experimental. This new technique is further exemplified on resonant cross section analysis.

2 Nuclear data evaluation with Bayesian inference

Evaluated nuclear data are often obtained from the synthesis of experimental results and theoretical models. The necessity of theoretical models may come from:

1. missing experimental data (comprehensiveness)
2. targeted uncertainties not met (precision)
3. different mixed-up phenomena blurred in recorded data (selectivity)

Theoretical models are not always fully predictive and sometimes require adjustment of model parameters. Note that the notion of “model parameters” can be understood in a larger sens. When one only aggregates several experimental values, the obtained mean value can also be seen as a model parameter; and the set of provided values supplied in an ENDF file can be seen as a piece-wise linear model reproducing the evaluated data.

In nuclear data evaluation, it is therefore general to consider parameter value estimation, and the Bayesian inference is an interesting mathematical framework for this task. The main idea is to treat parameter values \vec{p} as random quantities described with related probability density functions (PDF). Starting with a possible initial *prior* PDF $P(\vec{p})$, information provided by \vec{y} (a direct or indirect measurement of \vec{p}) yields an updated *posterior* PDF $P(\vec{p}|\vec{y})$. This relation is given by the Bayes’ theorem [8]

$$P(\vec{p}|\vec{y}) = \frac{P(\vec{p})P(\vec{y}|\vec{p})}{\int d\vec{p}' P(\vec{p}')P(\vec{y}|\vec{p}')} \quad (1)$$

where $P(\vec{y}|\vec{p})$ is the likelihood describing the probability to observe \vec{y} given a value of \vec{p} . The posterior distribution can further be used to obtain best-estimate values of a parameter p_i and related covariance matrix as the first and

second moments of the posterior PDF:

$$\begin{aligned} \langle p_i \rangle &= \int d\vec{p} p_i P(\vec{p}|\vec{y}) \\ M_{ij} &= \text{Cov}(p_i, p_j) = \int d\vec{p} (p_i - \langle p_i \rangle)(p_j - \langle p_j \rangle) P(\vec{p}|\vec{y}) \end{aligned} \quad (2)$$

Assuming no prior knowledge for \vec{p} and if the likelihood is taken as

$$P(\vec{y}|\vec{p}) = \exp\left(-\frac{1}{2}(\vec{y} - \vec{t}(\vec{p}))^t M_y^{-1} (\vec{y} - \vec{t}(\vec{p}))\right) \quad (3)$$

the Bayesian approach reduces to a least-square fit but provides also a posterior covariance matrix. Here the likelihood is related to the usual χ^2 , with M_y being the covariance matrix on the observation \vec{y} and $\vec{t}(\vec{p})$ being the modeled values to be compared to \vec{y} .

To first order, meaning for small variations of \vec{p} about $\langle \vec{p} \rangle$, the uncertainty on \vec{p} can be propagated to modeled values $\vec{\sigma}(\vec{p})$ – which can be different from $\vec{t}(\vec{p})$ – using the sensitivity matrix G_{ij} defined as

$$G_{ij} = \frac{\partial \sigma_i}{\partial p_j} \quad (4)$$

The eventual covariance matrix on the $\vec{\sigma}$ quantities are obtained as

$$\begin{aligned} N &= GMG^t \\ \text{where } N_{ij} &= \text{Cov}(\sigma_i, \sigma_j) \end{aligned} \quad (5)$$

The model – providing theoretical values \vec{t}_1 to be compared with experimental values \vec{y} – can be a composite model. It can depends on parameters \vec{p}_1 and on a sub-model \vec{t}_2 depending on its own parameters \vec{p}_2 . This Russian dolls aspect can also apply for \vec{t}_2 . The reason for this nested structure is that some experimental corrections can be handled during the Bayesian inference (described at the \vec{t} level) by the evaluator. This can be for instance normalization or sample temperature for resonant cross section analysis. The parameters can also be frozen and fixed in the M_y matrix when treated by the experimentalist, for instance detector efficiency or dead time correction. However this latter approach may lead to so-called Peel’s Pertinent Puzzles that are situation where the eventual χ^2 is minimized for values \vec{t} that obviously mismatch \vec{y} . In general one can separate the “experimental” or “nuisance” parameters $\vec{\theta}$ to the theoretical parameters \vec{x} that describe the fundamental nuclear data. The nested model will be written in short as $\vec{t}(\vec{\theta}, \vec{x})$ in the rest of the paper.

3 Uncertainty estimate – Traditional marginalization methods

The Bayesian inference described above produces best-estimates for \vec{x} and $\vec{\theta}$ but also a covariance matrix between those parameters:

$$M_p = \begin{bmatrix} M_x & M_{x,\theta} \\ M_{\theta,x} & M_\theta \end{bmatrix} \quad (6)$$

where only the sub-matrix M_x will be used to propagate uncertainties on evaluated data $\sigma(\vec{x})$. A long-standing idea

to produce conservative covariance data has been to modify M_x so that uncertainties propagated on \vec{r} could be considered to originate from \vec{x} alone, not from \vec{x} and $\vec{\theta}$. Two methods will be reviewed in this section and exemplified on a toy model that will be presented first.

3.1 Toy model analysis

The toy model considered here consists in a linear model

$$\vec{r}(E, x, \theta) = \theta E + x \quad (7)$$

depending on a nuisance parameter θ (slope), a theoretical parameter x (value at origin) and is applied on an experimental condition parameter E . The parameters x and θ are inferred from “experimental” data points $y_i(E_i)$ generated as the following:

- E_i are sampled uniformly on $[0, 1]$
- $y_i = 0.55 \pm 0.1$ (with even probability of sign) are altered by a noise uniformly sampled in $[-0.01, 0.01]$
- A statistical uncertainty of 0.01 is associated to all y_i .

It is clear that expected final values should be

$$\theta_{\text{fin}} = 0 \quad \text{and} \quad x_{\text{fin}} = 0.55 \quad (8)$$

It is also clear that the model (7) is designed not to be able to reproduce the sampled data that follows a bi-modal distribution. This is done to emulate either model defects or experimental artifacts that could bias model parameter estimation. It is yet useful to have assimilation methods robust enough to yield realistic (mono-modal) mean values and standard deviations even in such cases. The sampled data are first analyzed with model (7) by Bayesian inference without consideration of prior knowledge for x and θ , which resumes to a least-square fit. The initial values used for the parameters are

$$\theta_{\text{ini}} = 0.055 \pm 0.055 \quad \text{and} \quad x_{\text{ini}} = 0.3 \quad (9)$$

The initial uncertainty on θ is provided as it will be used latter for marginalization techniques. One can either adjust θ along with x – these cases are called “informative” hereafter – or keeping θ constant, which is called here “non-informative”. The data set is shown in Fig. 1 with only 2 sampled points analyzed. Of course with this small number of parameters, adjusting θ yields a perfect match of experimental data but terribly wrong values of x_{fin} and θ_{fin} . The simple model (7) allows to easily read the value of x : the model value for $E = 0$. The value of θ can also be estimated as the slope of the curves. The analysis for 20 sampled points is shown on Figure 2. It shows that increasing the amount of experimental information (increasing the number of data points) improves the estimation of final values in both cases (the final values are indicated in the figures). However either x or θ has a final uncertainty that is inconsistent with the expected values (8). It is striking however that the propagated uncertainty band (in green) does not cover the spread of the experimental data points. Even in the most favorable case (non-informative),

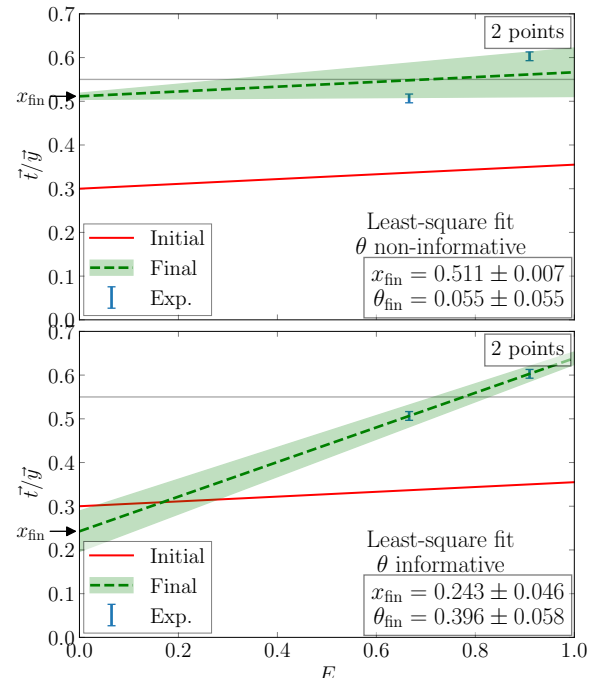


Figure 1. Analysis of the 2 points data set with the toy model data by Bayesian inference with and without fit of θ .

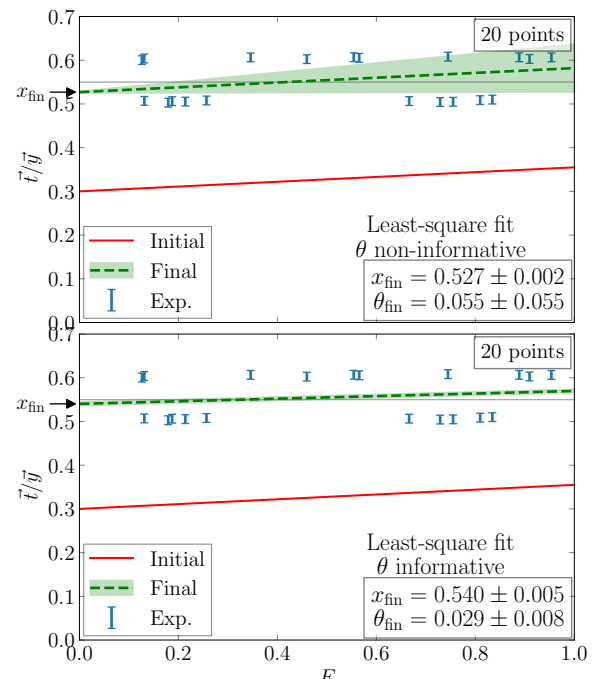


Figure 2. Analysis of the 20 points data set with the toy model data by Bayesian inference with and without fit of θ .

the large band is mostly due to the propagation of the initial uncertainty on θ . This can be seen as the uncertainty band becomes larger as E increases. Since only x is used for the applications $\sigma(\vec{x})$, the propagated uncertainty will be extremely small compared to the experimental knowledge of x . It is therefore misleading to propagate uncertainty on θ on such figures, for the rest of the paper only the uncertainty on x will be shown in the final uncertainty band. The initial relative uncertainty on θ (Eq. (9)) can

be questioning since it stands for a calibration parameter that is in principle well known. It should be understood as θ is expected to be smaller than a given value – as for a background for instance – with an average propagated uncertainty on \vec{t} of 5%. This can be high for a calibration parameter but for readability we restrict the toy model with a limited number of data points, which pushes for exaggerating the uncertainty on θ . This prior uncertainty of 0.055 for θ is rather consistent (for E between 0.5 and 1) with the bi-modal spread of experimental data points. However θ is not related to this bi-modal feature and thus cannot be expected to account for it. It will be shown latter that the uncertainty on θ is not necessarily the key for realistic uncertainty estimation.

3.2 “Matrix marginalization”

Several techniques have been designed to transfer uncertainty from θ on x . The first method presented here that is called matrix marginalization but is also called analytical marginalization [4, 6, 7] or zero-variance penalty marginalization [5] in other works. This method attempts to find a matrix M_x^m on x alone so that the propagated covariance matrix on \vec{t} is the same than the one obtained with M_p of Eq. (6), namely

$$\underbrace{G_x M_x^m G_x^t}_{\text{expected matrix}} = \underbrace{G_p M_p G_p^t}_{\text{(propagated) posterior matrix}} \quad (10)$$

where

$$[G_x]_{ij} = \frac{\partial t_i}{\partial x_j}, \quad [G_\theta]_{ij} = \frac{\partial t_i}{\partial \theta_j}, \quad G_p = [G_x G_\theta] \quad (11)$$

Multiplying Eq. (10) from the left by G_x^t and from the right by G_x one has

$$G_x^t G_x M_x^m G_x^t G_x = G_x^t G_p M_p G_p^t G_x \quad (12)$$

After multiplication on both sides by $(G_x^t G_x)^{-1}$ one obtains

$$M_x^m = (G_x^t G_x)^{-1} G_x^t G_p M_p G_p^t G_x (G_x^t G_x)^{-1} \quad (13)$$

However inserting Eq. (13) in Eq. (10) one does not generally obtain the propagated posterior matrix $G_p M_p G_p^t$, this could occur but in special cases where $G_x (G_x^t G_x)^{-1} G_x^t = \mathbb{1}$. This method does not yield the claimed properties, but it is still often used in practice.

Depending on how $\vec{\theta}$ are treated, three different cases can be considered with this method

1. *informative* case: mean value and standard deviation of $\vec{\theta}$ are updated by Bayesian inference
2. *non-informative* case: mean value and standard deviation of $\vec{\theta}$ are *not* updated by Bayesian inference
3. *conservative-informative* case: mean value of $\vec{\theta}$ is updated by Bayesian inference but its standard deviation is restored to the original value. One can replace M_θ in Eq. (6) by M_θ^{ini} and use the altered M'_p in Eq. (13):

$$M'_p = \begin{bmatrix} M_x & M_{x,\theta} \\ M_{\theta,x} & M_\theta^{\text{ini}} \end{bmatrix} \quad (14)$$

where it is not clear if cross terms should also be restored to their original values. In the two latter cases a initial uncertainty on θ must be used. The results for the three cases are shown in Fig. 3.

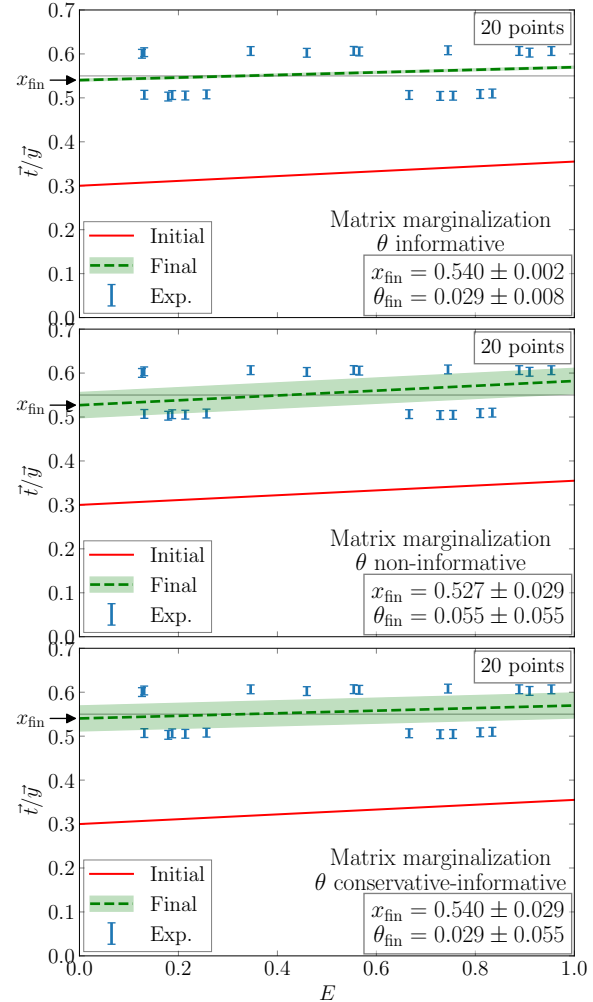


Figure 3. Uncertainty propagation in the three “matrix marginalization” methods: informative (top), non-informative (middle), conservative-informative (bottom).

The informative case does not yield much larger uncertainty band than in the case of strict Bayesian inference. This is due to the large number of data points that narrows the likelihood. For the non-informative case, the uncertainty band is indeed increased but the posterior value of x is less accurate than in the original Bayesian case (x should tend to 0.55). Finally the conservative-informative case exhibits the best behavior, x and θ tend both to good values and a larger uncertainty band is obtained. Two comments can yet be made:

1. M_x^m as given in Eq. (13) does not guarantee to obtain the desired propagated covariance matrix of Eq. (10).
2. The larger uncertainty band obtained here is only due to the prior θ uncertainty. In the sampled data points θ does not play any role and thus a much smaller initial uncertainty on θ would have been

plausible. As mentioned above, θ is not related to the bi-modal behavior of the sampled data. For instance in Fig. 4 a smaller initial value of θ is used (and a smaller prior uncertainty) on the same set. It yields this time a narrow uncertainty band that does not render the effect of the bi-modal distribution. This time the final parameter uncertainties do not overlap the true values.

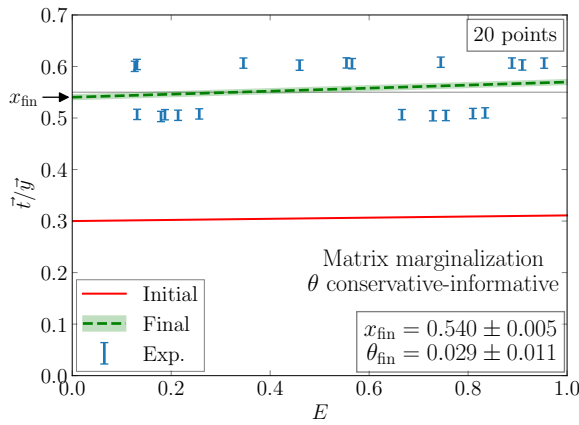


Figure 4. Uncertainty propagation with the conservative-informative “matrix marginalization” technique using a reduced initial uncertainty on θ : $\theta_{mi} = 0.011 \pm 0.011$.

3.3 “Bayesian marginalization”

Even if the matrix marginalization shows somehow reasonable behavior, in practice it was observed to be unstable in cases with larger number of parameters. This is possibly due to the improper connection from Eq. (13) to Eq. (10). Another traditional method has a stronger connection with the Bayesian framework [3]. This method is derived in few steps from the Bayesian posterior given in Eq. (1). First \vec{p} is split into \vec{x} and $\vec{\theta}$, then the conditional probability theorem is applied:

$$P(\vec{p}|\vec{y}) = P(\vec{x}, \vec{\theta}|\vec{y}) = P(\vec{x}|\vec{\theta}, \vec{y})P(\vec{\theta}|\vec{y}) \quad (15)$$

It is then assumed that $\vec{\theta}$ do not benefit from information from \vec{y} , so that

$$P(\vec{\theta}|\vec{y}) = P(\vec{\theta}) \quad (16)$$

The marginalized PDF of \vec{x} is then defined as

$$P^m(\vec{x}) = \int d\vec{\theta} P(\vec{x}, \vec{\theta}|\vec{y}) = \int d\vec{\theta} P(\vec{x}|\vec{\theta}, \vec{y})P(\vec{\theta}) \quad (17)$$

The Bayesian inference is then applied on $P(\vec{x}|\vec{\theta}, \vec{y})$ yielding

$$P(\vec{x}|\vec{\theta}, \vec{y}) = \frac{P(\vec{x}|\vec{\theta})P(\vec{y}|\vec{\theta}, \vec{x})}{\int d\vec{x}' P(\vec{x}'|\vec{\theta})P(\vec{y}|\vec{\theta}, \vec{x}')} \quad (18)$$

Finally one assumes that there is no prior correlation between \vec{x} and $\vec{\theta}$, namely

$$P(\vec{x}|\vec{\theta}) = P(\vec{x}) \quad (19)$$

so that the marginalized PDF of \vec{x} is

$$P^m(\vec{x}) = \int d\vec{\theta} P(\vec{\theta}) \frac{P(\vec{x})P(\vec{y}|\vec{\theta}, \vec{x})}{\int d\vec{x}' P(\vec{x}'|\vec{\theta})P(\vec{y}|\vec{\theta}, \vec{x}')} \quad (20)$$

From this distribution, one obtains the mean value and covariance matrix on \vec{x} . The corresponding results are shown in Fig. 5.

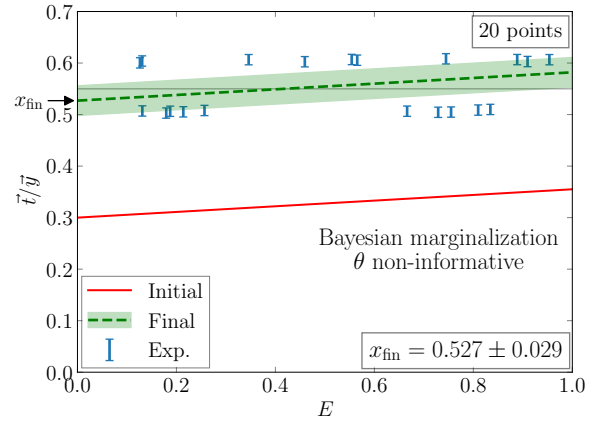


Figure 5. Uncertainty propagation in the Bayesian marginalization method.

One can see that the results of this method has the same behavior than the non-informative matrix marginalization (middle plot of Fig. 3). Namely since $\vec{\theta}$ is assumed not to benefit from information from \vec{y} (Eq. (16)), the constant value of $\theta = \theta_{mi}$ bias the posterior estimate of x which is not as close to 0.55 as in the “informative” cases.

4 Spectral decomposition method

A third method [7] combines a good mathematical definition and the possibility to let the $\vec{\theta}$ parameters evolve if necessary. This method is based on the preservation of conservative uncertainty propagation for *any* quantity varying (linearly) with parameters \vec{x} . To demonstrate the method it is first recalled that covariance matrices must be positive (semi-)definite, namely all its eigenvalues must be positive or zero. This is true for the posterior covariance matrix on M_x , which can be decomposed as

$$M_x = Q\Lambda Q^t \quad (21)$$

where Q stores M_x eigenvectors and Λ is a diagonal matrix of its eigenvalues. The fact that $\forall i, \Lambda_i \geq 0$ implies that any quantity computed from \vec{x} that varies linearly with \vec{x} (possibly only locally), with sensitivity vector \vec{S} , has a positive propagated variance given by

$$\vec{S}^t M_x \vec{S} = (Q^t \vec{S})^t \Lambda Q^t \vec{S} = \sum_i \Lambda_i [Q^t \vec{S}]_i^2 \geq 0 \quad (22)$$

A conservative covariance matrix can thus be searched in the form $M_x^m = Q\Lambda^m Q^t$ where the Λ_i^m verify

$$\forall i, \Lambda_i^m \geq \Lambda_i \quad (23)$$

This parameterization, where the Λ_i^m parameters are adjusted to satisfy some criteria that will be defined below, ensures that the propagated variance will be larger than the posterior one. Note that in practice one can keep standard deviation outside the covariance matrix and work on the eigenvalue decomposition of the correlation matrix instead. This can improve numerical stability of the method.

It seems reasonable to search for a covariance matrix that would be conservative *just* as necessary. This requirement with condition Eq. (23) can be mathematically formulated as a minimization under constraints problem, for instance with the cost function

$$S = \frac{1}{2} \sum_i \left(\frac{\Lambda_i^m}{\Lambda_i} - 1 \right)^2 \quad (24)$$

to be minimized and with constraints

$$\forall i, g_i = \frac{\Lambda_i^m}{\Lambda_i} - 1 \geq 0 \quad (25)$$

Stated as such, the problem has a trivial solution: $\forall i, \Lambda_i^m = \Lambda_i$. To be more conservative, additional constraints can be considered. In the same spirit as in the matrix marginalization, one can impose the variance propagated from M_x^m on the t_n to be larger than the one obtained by propagation of M_p (or by M'_p given by Eq. (14)). This gives the constraints

$$\forall n, g_n = \frac{v_n^m}{v_n} - 1 \geq 0, \quad \text{or} \quad g_n = \frac{v_n^m}{v_n'} - 1 \geq 0 \quad (26)$$

with

$$\begin{aligned} v_n^m &= [G_x M_x^m G_x^t]_{nn} \\ v_n &= [G_p M_p G_p^t]_{nn} \\ v_n' &= [G_p M'_p G_p^t]_{nn} \end{aligned} \quad (27)$$

Figure 6 shows the propagated results obtained with the informative approach (use of M_p and v_n) and in the conservative-informative approach (use of M'_p and v_n').

It can be seen that final expectation values of x and θ are as good as the best cases in the other methods because x and θ are adjusted during the Bayesian inference step and that the spectral method does not alter best-estimates. Even if a large uncertainty band is obtained, this method still suffers from the inconsistent definition of M'_p (Eq. (14)). Additionally if the initial uncertainty in θ does not account for the model defect, the uncertainty band will not be reliable (just like in Fig 4). To avoid these two latter points (to keep sound maths and robustness when initial uncertainties on θ are small), additional constraints can be included.

The first additional set of constraints are called *local* constraints and enforce that the random variables $(t_n - y_n) / \sqrt{v_n^m + [M_y]_{nn}}$ – considered as following a centered standard distribution – fall within an interval $[-\alpha, +\alpha]$. The value of α corresponding to the a one- σ standard deviation is given by $\alpha = \sqrt{2} \operatorname{erf}^{-1}(68.27\%)$, where erf is the error function. With this choice $\alpha = 1$ but could be changed depending of the targeted confidence interval (here 68.27%). The corresponding constraints are thus

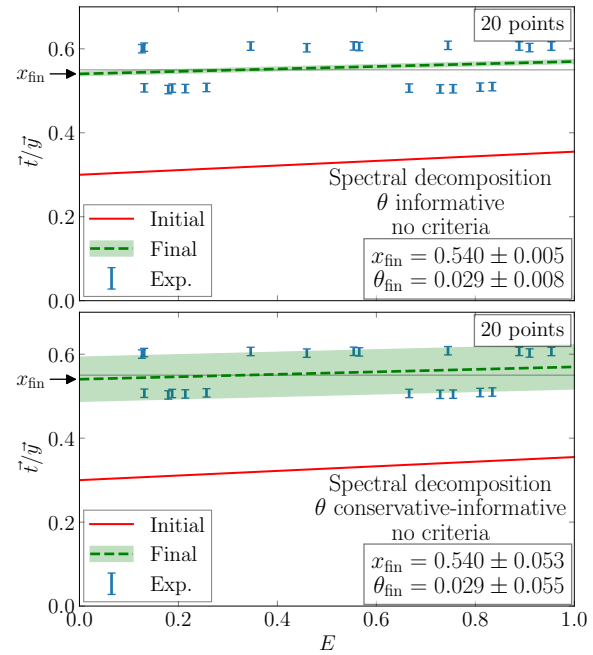


Figure 6. Uncertainty propagation in the informative (top) and conservative-informative (bottom) spectral methods.

given by

$$\forall n, g_n = \alpha^2 \frac{[M_y]_{nn} + v_n^m}{(t_n - y_n)^2} \geq 0 \quad (28)$$

The results obtained with this approach is shown in Fig. 7.

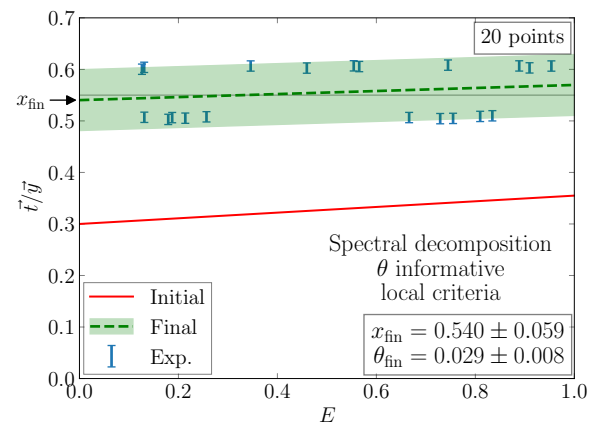


Figure 7. Uncertainty propagation in the spectral method with a local criteria constraints.

The local criteria constraint enforce that *all* points should fall within the experimental+propagated uncertainty band. Since the related constraints (Eq. (28)) are define for each point specifically, often a specific point will be responsible for the critical constraint ($g_n = 0$) limiting the minimization under constraint. This uncertainty propagation method can thus be used to identify possible outliers.

One may consider the local criteria as excessive. Indeed even with proper normally distributed random samples, it is not rare to have few points falling outside the

one- σ interval. A global criteria [7] constraint can thus be given by

$$g = 1 - \frac{X^2}{\alpha^2} \geq 0 \quad (29)$$

with

$$X = \frac{1}{\sqrt{N_y}} \left(\sum_n \frac{(t_n - y_n)^2}{[M_y]_{nn} + v_n^m} - N_y \right) \quad (30)$$

where N_y is the number of experimental data points. This latter criteria is valid if N_y is large because it assumes that an asymptotic behavior of a chi-square distribution with N_y degrees of freedom. The results obtained with this approach is shown in Fig. 8.

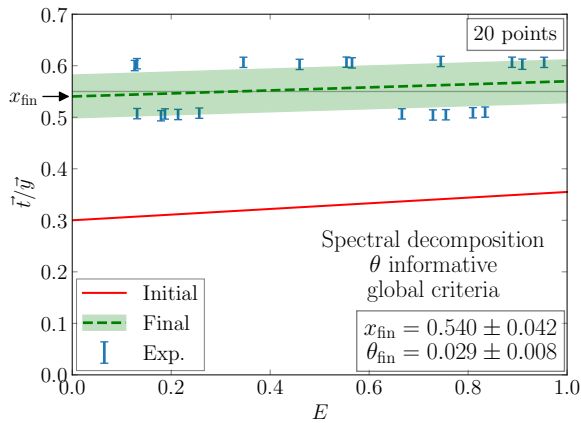


Figure 8. Uncertainty propagation in the spectral method with the global criteria constraint.

It is visible that the uncertainty band covers a large part of the data points but some points can lay outside since the global criteria is less strict than the local criteria. At first glance this latter method seems equivalent to the conservative-informative matrix marginalization results, here however:

1. the problem is mathematically well defined and is expected to scale better with increasing the number of parameters, namely because there is no consistency issues anymore between Eq. (13) and Eq. (10).
2. the large uncertainty band is longer driven by the prior uncertainty on θ but by the posterior agreement or disagreement between model values and experimental data.

5 Application to resonant cross section analysis and model defect

The spectral decomposition methods presented above can be used *even if there is no nuisance parameter considered in the analysis*. This application case of this marginalization technique is used to show its effects in case of realistic model defect. The spectral decomposition method is exemplified here in the case of an analysis of neutron total cross section measurement of sodium [9]. This measurement is reported to have a statistical uncertainty between 0.7% to 3.4% in the considered energy range, and 3% of

systematic uncertainty, usually assumed to correspond to a normalization factor – not considered here. In Fig. 9 the experimental values are shown with the evaluated data provided in the JEFF-3.2 international evaluated nuclear data library [10] that is used here as a prior.

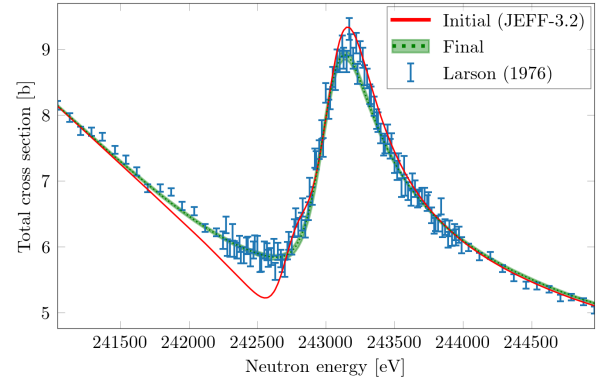


Figure 9. Resonance analysis using the JEFF-3.2 parameters.

The experimental data in the range [241 keV, 245 keV] of incident neutron kinetic energy have been used to adjust parameters provided in JEFF-3.2. These parameters are provided to an R-Matrix model [11] (with Reich-Moore approximation) used to reproduce cross sections in a large energy range. Here the setup response function is not considered, no resolution function and no Doppler broadening were included in the analysis, it was however tested that including them does not modify the presented conclusions. In the prior value of Fig. 9, an experimented eye can see that the structure is composed of two resonances, corresponding to two energies given in JEFF-3.2: $E_{\lambda} = 242.7913$ keV and $E_{\lambda} = 243.0485$ keV. For each of these two resonances, two other parameters (resonance partial widths) are given for the radiative capture and the elastic scattering channels. To perform consistent comparisons, these six parameters were adjusted on the experimental data set and yielded the green line in Fig. 9 where it is no longer possible to distinguish the contribution of the two resonances. To make sure of the necessity of having two resonances in this energy range, the three parameters related to each resonance (energy, gamma and neutron widths), have been, in turn, removed while those for the remaining resonance have been re-adjusted. The resulting χ^2 are reported in Tab. 1 along the χ^2 obtained while keeping both resonances.

Table 1. Summary of adjustments on $^{23}\text{Na}(n, \text{tot})$ measurement

nb. of parameters	χ^2_{fin}	nb. iterations
6 (two res.)	1.63	15
3 (first res. kept)	2.96	7
3 (second res. kept)	2.82	3
4	1.06	4

The significant reduction of χ^2 while keeping two sets of resonance parameters can be seen as a justification of the inclusion of a second resonance. However it was tested to keep only the second resonance but using a general-

ized version of the R-matrix model where several neutron widths can be used. Two partial widths were thus used corresponding to orbital angular momentum $\ell = 0$ and $\ell = 2$ that are coupled to the $J^\pi = 0^+$ (second) resonance. In this case only the second resonance is kept but an extra parameter is included. The corresponding χ^2 is also reported in the last line of Tab. 1. The large reduction of χ^2 with a reduced number of adjusted parameters with respect to JEFF-3.2 allows to conclude that *there is a model defect* in the precise energy range in JEFF-3.2. Note that the number of iterations required to converge as reported in Tab. 1 also argues in the same direction.

We now consider only the two models

- the 6-parameters model, *i.e.* JEFF-3.2 refitted – the model with a model defect
- the 4-parameters model, *i.e.* the extend reference R-matrix model

And we will see how they behave with the spectral marginalization methods. The results obtained with the local criteria are shown in Fig. 10.

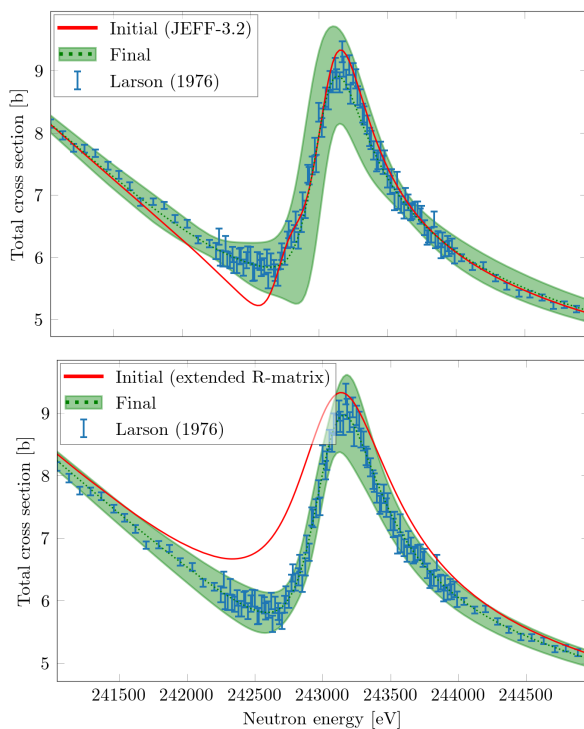


Figure 10. Uncertainty band obtained with the local criteria with the JEFF-3.2 prior (top) and with the extended model (bottom).

While for the extended R-matrix case the local criteria yields a smooth uncertainty band following nicely experimental “error bars”, results for the uncertainty band obtained from the JEFF-3.2 model extends much farther, namely at the top and bottom of the resonance, highlighting the model defect penalty. This could indicate that even if a reasonable χ^2 is obtained with the addition of a second resonance (χ^2 divided by about a factor of two, *cf.* Tab. 1), the model has trouble overlapping the spread of experimental points within their reported uncertainties. The uncertainty band thus gives an idea of the model flexibility

allowing to reproduce or not the observed shape. This also indicates that a second measurement with consistent – but different – experimental data points could be difficult to assimilated in the Generalized Least-Square approach. Indeed in the 6-parameters models, parameters must vary a lot to accommodate experimental data spread and uncertainties, this would correspond to a strong penalty in successive data assimilation.

If the *global* criteria is used instead of the *local* ones, the width of the uncertainty band width reduces but remains larger than the band obtained with the reference model. This latter case is not shown here, instead focus is made on the 6-parameters model. Indeed in most cases we are not aware if there is a model defect or not. The tested situation presented here consists in comparing the uncertainty band obtained with the Bayesian inference only and the one obtained after applying the global criteria. The result is shown in Fig. 11, where the uncertainty band obtained with the global criteria is 20% to 50% larger than one given by the Bayesian posterior. Both methods share the same expectation values since the spectral decomposition method only deals with second order moments.

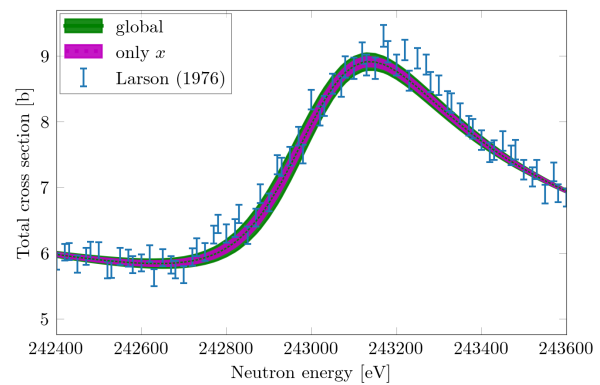


Figure 11. Uncertainty band obtained with the global criteria compared to the one obtained after the Bayesian analysis.

6 Conclusion

The necessity to estimate realistic uncertainties led to the development of marginalization techniques used to transfer uncertainties from nuisance (or experimental) parameters to evaluated parameters that are used to produce nuclear data. Two main methods have been reviewed here but shown to suffer from mathematical weakness or from lack of flexibility while keeping nuisance parameter values fixed. An alternative method based on the spectral decomposition of the covariance (or correlation) matrix has been presented. Here not only nuisances parameter values are free to be updated by data assimilation, but the produced uncertainties are guaranteed to be larger than the one propagated after Bayesian assimilation. This method allows to transfer uncertainty from nuisance parameters to model parameters of interest but can also be applied without nuisance parameters. In this latter cases, it is the compatibility of post-fit theoretical values and provided experimental

data that is used to condition estimated uncertainties. This method has been applied here on a toy model experiment analysis but also on a real-case resonant cross section analysis. It has been shown that the spectral method (using the local criteria) could be used to indicate the presence of a not-obvious model defect. The same method (with global criteria) is shown to increase the propagated uncertainty by 50% with respect to original Bayesian assimilation. In all cases, the spectral method is applied *after* the Bayesian (or least-square) fit, it is therefore very fast as no extra model computations or derivatives are required. It is also rather easy to implement in existing codes.

References

- [1] A. Trkov, D.A. Brown, Tech. Rep. BNL-203218-2018-INRE, Brookhaven National Lab.(BNL), Upton, NY (United States) (2018), <https://www.osti.gov/biblio/1425114>
- [2] F. Fröhner, Nuclear Science and Engineering **126**, 1 (1997)
- [3] C. De Saint Jean, G. Noguère, B. Habert, B. Iooss, Nuclear Science and Engineering **161**, 363 (2009)
- [4] B. Habert, C. De Saint Jean, G. Noguère, L.C. Leal, Y. Rugama, Nuclear Science and Engineering **166**, 276 (2010)
- [5] G. Noguère, P. Archier, C. De Saint Jean, B. Habert, Nuclear Science and Engineering **172**, 164 (2012)
- [6] C. De Saint Jean, P. Archier, E. Privas, G. Noguère, B. Habert, P. Tamagno, Nuclear Data Sheets **148**, 383 (2018), special Issue on Nuclear Reaction Data
- [7] P. Tamagno, European Physical Journal A **57**, 61 (2021)
- [8] T. Bayes, Philosophical Transactions **53**, 370 (1763)
- [9] D.C. Larson, J.A. Harvey, N.W. Hill, Tech. Rep. ORNL/TM-5614, Oak Ridge National Lab., Tenn. (USA) (1976), <https://www.osti.gov/scitech/servlets/purl/7342187>
- [10] A.J.M. Plompen, O. Cabellos, C. De Saint Jean, M. Fleming, A. Algora, M. Angelone, P. Archier, E. Bauge, O. Bersillon, A. Blokhin et al., The European Physical Journal A **56** (2020)
- [11] A.M. Lane, R.G. Thomas, Rev. Mod. Phys. **30**, 257 (1958)