

On the use of the BMC to resolve Bayesian inference with nuisance parameters

Edwin Privas, C. de Saint-Jean, Gilles Noguere

▶ **To cite this version:**

Edwin Privas, C. de Saint-Jean, Gilles Noguere. On the use of the BMC to resolve Bayesian inference with nuisance parameters. EPJ N - Nuclear Sciences & Technologies, 2018, 4, pp.36. 10.1051/epjn/2018042 . cea-02305641

HAL Id: cea-02305641

https://cea.hal.science/cea-02305641

Submitted on 4 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

EPJ Nuclear Sci. Technol. **4**, 36 (2018)
© E. Privas et al., published by EDP Sciences, 2018
https://doi.org/10.1051/epjn/2018042

EPJ Nuclear Sciences & Technologies

Available online at:
https://www.epj-n.org

REGULAR ARTICLE OPEN ACCESS

On the use of the BMC to resolve Bayesian inference with nuisance parameters

Edwin Privas*, Cyrille De Saint Jean, and Gilles Noguere

CEA, DEN, Cadarache, 13108 Saint Paul les Durance, France

Received: 31 October 2017 / Received in final form: 23 January 2018 / Accepted: 7 June 2018

Abstract. Nuclear data are widely used in many research fields. In particular, neutron-induced reaction cross sections play a major role in safety and criticality assessment of nuclear technology for existing power reactors and future nuclear systems as in Generation IV. Because both stochastic and deterministic codes are becoming very efficient and accurate with limited bias, nuclear data remain the main uncertainty sources. A worldwide effort is done to make improvement on nuclear data knowledge thanks to new experiments and new adjustment methods in the evaluation processes. This paper gives an overview of the evaluation processes used for nuclear data at CEA. After giving Bayesian inference and associated methods used in the CONRAD code [P. Archier et al., Nucl. Data Sheets **118**, 488 (2014)], a focus on systematic uncertainties will be given. This last can be deal by using marginalization methods during the analysis of differential measurements as well as integral experiments. They have to be taken into account properly in order to give well-estimated uncertainties on adjusted model parameters or multigroup cross sections. In order to give a reference method, a new stochastic approach is presented, enabling marginalization of nuisance parameters (background, normalization...). It can be seen as a validation tool, but also as a general framework that can be used with any given distribution. An analytic example based on a fictitious experiment is presented to show the good ad-equations between the stochastic and deterministic methods. Advantages of such stochastic method are meanwhile moderated by the time required, limiting it's application for large evaluation cases. Faster calculation can be foreseen with nuclear model implemented in the CONRAD code or using bias technique. The paper ends with perspectives about new problematic and time optimization.

1 Introduction

Nuclear data continue to play a key role, as well as numerical methods and the associated calculation schemes, in reactor design, fuel cycle management and safety calculations. Due to the intensive use of Monte-Carlo tools in order to reduce numerical biases, the final accuracy of neutronic calculations depends increasingly on the quality of nuclear data used. The knowledge of neutron induced cross section in the 0 eV and 20 MeV energy range is traduced by the uncertainty levels. This paper focuses on the neutron induced cross sections uncertainties evaluation. The latter is evaluated by using experimental data – either microscopic or integral, and associated uncertainties. It is very common to take into account the statistical part of the uncertainty using the Bayesian inference. However, systematic uncertainties are not often taken into account either because of the lack of information from the experiment or the lack of description by the evaluators.

A first part presents the ingredients needed in the evaluation of nuclear data: theoretical models, microscopic and integral measurements. A second part is devoted to the presentation of a general mathematical framework related to Bayesian parameters estimations. Two approaches are then studied: a deterministic and analytic resolution of the Bayesian inference and a method using Monte-Carlo sampling. The next part deals with systematic uncertainties. More precisely, a new method has been developed to solve the Bayesian inference using only Monte-Carlo integration. A final part gives a fictitious example on the ^{235}U total cross section and comparison between the different methods.

2 Nuclear data evaluation

2.1 Bayesian inference

Let $\boldsymbol{y} = \vec{y_i}(i = 1 \dots N_y)$ denote some experimentally measured variables, and let $\boldsymbol{x}$ denote the parameters defining the model used to simulate theoretically these variables and $\boldsymbol{t}$ is the associated calculated values to be compared with $\boldsymbol{y}$. Using Bayes' theorem [1] and especially its generalization to

* e-mail: edwin.privas@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

continuous variables, one can obtain the well known relation between conditional probability density functions (written $p(.)$) when the analysis of a new dataset $\boldsymbol{y}$ is performed:

$$p(x|y,U) = \frac{p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|\boldsymbol{x},U)}{\int d\boldsymbol{x}\cdot p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|\boldsymbol{x},U)}, \quad (1)$$

where U represents the "background" or "prior" information from which the prior knowledge of $\boldsymbol{x}$ is assumed. U is supposed independent of $\boldsymbol{y}$. In this framework, the denominator is just a normalization constant.

The formal rule [2] used to take into account information coming from new analyzed experiments is:

$$\text{posterior} \propto \text{prior} \cdot \text{likelihood}. \quad (2)$$

The idea behind fitting procedures is to find an estimation of at least the first two moments of the posterior density probability of a set of parameters $\boldsymbol{x}$, knowing an a priori information (or first guess) and a likelihood which gives the probability density function of observing a data set knowing $\boldsymbol{x}$.

Algorithms described in this section are summarized and detailed description can be found in paper linked to the CONRAD code in which they are implemented [3]. A general overview of the evaluation process and the conrad code is given in Figures 1 and 2.

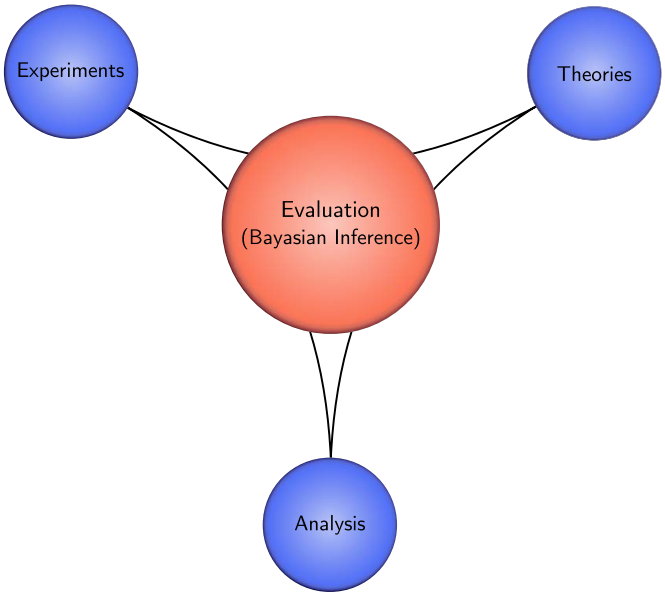


Fig. 1. General overview of the evaluation process.

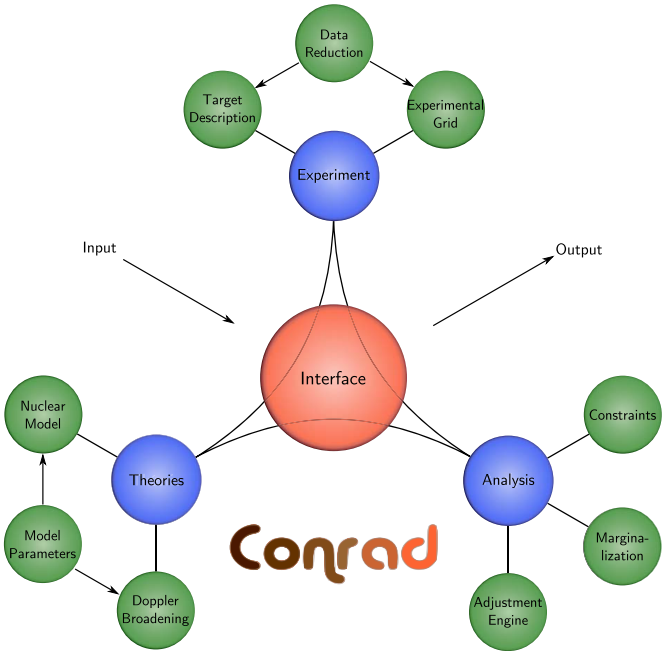


Fig. 2. General overview of the CONRAD code.

2.2 Deterministic theory

To obtain an equation to be solved, one has to make some assumptions on the prior probability distribution involved.

Given a covariance matrix and mean values, the choice of a multivariate joint normal for the probability density $p(\boldsymbol{x}|U)$ and for the likelihood is maximizing the entropy [4]. Adding Bayes' theorem, equation (1) can be written as follow:

$$p(\boldsymbol{x}|\boldsymbol{y},U) \propto e^{-\frac{1}{2}[(\boldsymbol{x}-\boldsymbol{x}_m)^T M_x^{-1}(\boldsymbol{x}-\boldsymbol{x}_m)+(\boldsymbol{y}-\boldsymbol{t})^T M_y^{-1}(\boldsymbol{y}-\boldsymbol{t})]}, \quad (3)$$

where $\boldsymbol{x}_m$ (expectation) and M_x (covariance matrix) are prior information on $\boldsymbol{x}$, $\boldsymbol{y}$ an experimental set and $\mathrm{M}_{\boldsymbol{y}}$ the associated covariance matrix. $\boldsymbol{t}$ represents the theoretical model predictions.

The Laplace approximation is also made. It enables to approximate the posterior distribution by a multivariate normal distribution with the same maximum and curvature of the right side of equation (3). Then, it can be demonstrated that both the posterior expectation and the covariances can be calculated by finding the minimum of the following cost function (Generalized Least Square):

$$\chi^2_{GLS} = (\boldsymbol{x}-\boldsymbol{x}_m)^T M_x^{-1}(\boldsymbol{x}-\boldsymbol{x}_m) + (\boldsymbol{y}-\boldsymbol{t})^T M_y^{-1}(\boldsymbol{y}-\boldsymbol{t}). \quad (4)$$

To take into account non-linear effects and ensure a proper convergence of the algorithm, a Gauss–Newton iterative solution can be used [5].

Thus, from a mathematical point of view, the evaluation of parameters through a GLS procedure suffers from the Gaussian choice as guessed distribution for the prior and the likelihood, the use of Laplace approximation, the linearization around the prior for Gauss-Newton algorithm and at last a 2nd order terms neglected in the Gauss–Newton iterative procedure.

2.3 Bayesian Monte-Carlo

Bayesian Monte-Carlo (BMC) methods are natural solutions for Bayesian inference problems. They avoid approximations and propose alternatives in probability

density distribution choice for priors and likelihoods. This paper exposes the use of BMC in the whole energy range from thermal, resonance to continuum range.

BMC can be seen as a reference calculation tool to validate the GLS calculations and approximations. In addition, it allows to test probability density distributions effects and to find higher distribution moments with no approximations.

2.3.1 Classical Bayesian Monte-Carlo

The main idea of "classical" BMC is the use of Monte-Carlo techniques to calculate integrals. For any function f of random variable set $\boldsymbol{x}$, any integral can be calculated by a Monte-Carlo sampling:

$$\int p(\boldsymbol{x})\cdot f(\boldsymbol{x})d\boldsymbol{x} = \underset{n\overrightarrow{\infty}}{lim}\left(\frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{x}_k)\right), \tag{5}$$

where $p(\boldsymbol{x})$ is a probability density function. One can thus estimate any moments of the probability function $p(\boldsymbol{x})$. By definition, the mean value is given by:

$$\langle\, \boldsymbol{x}\,\rangle = \int \boldsymbol{x}\cdot p(\boldsymbol{x})d\boldsymbol{x}. \tag{6}$$

Application of these simple features to Bayesian inference analysis gives for posterior distribution's expectation values:

$$\langle\, x\,\rangle = \int \boldsymbol{x}\cdot \frac{p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|x,U)}{\int p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|\boldsymbol{x},U)d\boldsymbol{x}}d\boldsymbol{x}. \tag{7}$$

The proposed algorithm to find out the first two moments of the posterior distribution is to sample the prior probability distribution function $p(\boldsymbol{x}|U)\mathrm{N}_x$ times and for each $\boldsymbol{x}_k$ realization evaluate the likelihood:

$$\mathcal{L}_k = p(\boldsymbol{y}_k|\boldsymbol{x}_k,U). \tag{8}$$

Finally, the posterior expectation and covariance is given by the following equation:

$$\langle\, x_i\,\rangle_{N_x} = \frac{\sum_{k=1}^{N_x} x_{i,k}\mathcal{L}_k}{\sum_{k=1}^{N_x}\mathcal{L}_k}, \tag{9}$$

$$\langle\, x_i, x_j\,\rangle_{N_x} = \frac{\sum_{k=1}^{N_x} x_{i,k}\cdot x_{j,k}\cdot\mathcal{L}_k}{\sum_{k=1}^{N_x}\mathcal{L}_k} - \langle\, x_i\,\rangle_{N_x}\langle\, x_j\,\rangle_{N_x}. \tag{10}$$

The choice of the prior distribution depends on what kind of analysis is done. If no prior information is given, a non-informative prior could be chosen (uniform distribution). On the contrary, for an update of a given parameters set, the prior is related to a previous analysis with a known probability distribution function.

The use of $\mathcal{L}_k$ weights indicates clearly the major drawbacks of this classical BMC method: if the prior is far from high likelihood values and/or by nature $\mathcal{L}_k$ values are small (because of the number of experimental points for example), then the algorithm will have difficulties to converge.

Thus, the main issue is that the covered phase space of sampling is not favorable to convergence. In practice a trial function close to the posterior distribution should be chosen for sampling.

More details can be found in paper [6].

2.3.2 Importance sampling

As previously exposed, the estimation of the integral of $f(\boldsymbol{x})$ times a probability density function $p(\boldsymbol{x})$ is not straightforward. Especially in this Bayesian inference case, sampling the prior $p(\boldsymbol{x}|U)$ distribution, when it is far away from the posterior distribution or when the likelihood weights are difficult to evaluate properly, could be very expensive and time consuming without any valuable estimation of posterior distributions. The idea is then to sample in a different phase space region by respecting statistics.

This trial probability density function $p_{trial}(\boldsymbol{x})$ can be introduced by a trick as follow:

$$p(\boldsymbol{x}) = \frac{p(\boldsymbol{x})}{p_{trial}(\boldsymbol{x})}\cdot p_{trial}(\boldsymbol{x}). \tag{11}$$

Thus, putting this expression in equation (5), one obtains the following equation:

$$\int p(\boldsymbol{x})\cdot f(\boldsymbol{x})d\boldsymbol{x} = \underset{n\overrightarrow{\infty}}{lim}\left(\frac{1}{n}\sum_{k=1}^{n}\frac{p(\boldsymbol{x}_k)}{p_{trial}(\boldsymbol{x}_k)}\cdot f(\boldsymbol{x}_k)\right). \tag{12}$$

Then, sampling is done on the trial probability density function $p_{trial}(\boldsymbol{x})$ getting a new set of $\{\boldsymbol{x}_k\}$. For each realization $\boldsymbol{x}_k$, an evaluation of additional terms $\left(\frac{p(\boldsymbol{x})}{p_{bias}(\boldsymbol{x})}\right)$ is necessary.

As a result, expectation and covariances are defined by:

$$\langle\, x_i\,\rangle_{N_x} = \frac{\sum_{k=1}^{N_x} x_{i,k}\cdot h(\boldsymbol{x}_k)\cdot\mathcal{L}_k}{\sum_{k=1}^{N_x} h(\boldsymbol{x}_k)\cdot\mathcal{L}_k}, \tag{13}$$

and

$$\langle\, x_i, x_j\,\rangle_{N_x} = \frac{\sum_{k=1}^{N_x} x_{i,k}\cdot x_{j,k}\cdot h(\boldsymbol{x}_k)\cdot\mathcal{L}_k}{\sum_{k=1}^{N_x} h(\boldsymbol{x}_k)\cdot\mathcal{L}_k} - \langle\, x_i\,\rangle_{N_x}\langle\, x_j\,\rangle_{N_x}, \tag{14}$$

with $h(\boldsymbol{x}_k) = \frac{p(\boldsymbol{x}_k|U)}{p_{trial}(\boldsymbol{x}_k)}$.

The choice of trial functions is crucial and the closer to the true solution $p_{trial}(\boldsymbol{x})$ is, the quicker the algorithm will be. In this paper, a trial function used by default

comes from the result of the generalized least square (with additional standard deviation enhancement). Many other solutions can be used depending on the problem.

3 Systematic uncertainties treatment

3.1 Theory

Let recall some definitions and principles. First, it is possible to link model parameters $\boldsymbol{x}$ and nuisance parameters $\boldsymbol{\theta}$ with conditional probability:

$$p(\boldsymbol{x},\boldsymbol{\theta}|y,U) = p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y},U)\cdot p(\boldsymbol{\theta}|\boldsymbol{y},U), \tag{15}$$

with U the prior information on both model and nuisance parameters. The latter is supposed independent to the measurement. It means nuisance parameters are considered acting on experimental model. It induces $p(\boldsymbol{\theta}|\boldsymbol{y},U) = p(\boldsymbol{\theta}|U)$, giving the following equation:

$$p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y},U) = p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y},U)\cdot p(\boldsymbol{\theta}|U). \tag{16}$$

Moreover, evaluators are looking at the marginal probability density $p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y},U)$, also written $p_\theta(\boldsymbol{x}|\boldsymbol{y},U)$. It is given by the integration of the probability density function over marginal variables as follow:

$$p_\theta(\boldsymbol{x}|\boldsymbol{y},U) = \int p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y},U)d\boldsymbol{\theta}. \tag{17}$$

According to (16), the follow equation is obtained:

$$p_\theta(\boldsymbol{x}|\boldsymbol{y},U) = \int p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y},U)\cdot p(\boldsymbol{\theta}|U)d\boldsymbol{\theta}. \tag{18}$$

Then the Bayes theorem is used to calculate the first integral term of (18):

$$p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y},U) = \frac{p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x},U)}{\int p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x},U)dx}. \tag{19}$$

This expression is right if both model and nuisance parameters are supposed independent. According to (18) and (19), the marginal probability density function of the a posteriori model parameters is given by:

$$p_\theta(\boldsymbol{x}|\boldsymbol{y},U) = \int p(\boldsymbol{\theta}|U)d\boldsymbol{\theta}\frac{p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x},U)}{\int p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x},U)dx}. \tag{20}$$

3.2 Deterministic resolution

The deterministic resolution is well described in Habert thesis [7]. Several works have been first performed in 1972 by H. Mitani and H. Kuroi [8,9] and later by Gandini [10] giving a formalism to the multigroup adjustment and a way to take into account the systematic uncertainties. These were the first attempts to consider the possible presence of systematic errors in a data adjustment process. Equations are not detailed here, only the idea and the final equation is provided.

Let M_x^{stat} be the a posteriori covariance matrix obtained after an adjustment. The a posteriori covariance after marginalization M_x^{marg} can be found as follow:

$$M_x^{marg} = M_x^{stat} + (G_x^T G_x)^{-1}\cdot G_x^T\cdot G_\theta M_\theta G_\theta^T\cdot G_x\cdot(G_x^T G_x)^{-1} \tag{21}$$

with G_x sensitivities vector of the calculated model values to the model parameters and G_θ sensitivities vector of the calculated model values to the nuisance parameters. Similar expressions have been given in reference [8,9] where two terms appear: one for classical resolution and the second for some added systematic uncertainties. $(G_x^T G_x)$ is a square matrix supposed reversal. It is often the case when there are more experimental points than fitted parameters. If numeric issues appeared, it is mandatory to find another way, giving by a stochastic approach. Further study should be undertaken to compare the deterministic method proposed here and the one identified in Mitani's papers in order to provide a more robust approach.

3.3 Semi-stochastic resolution

This method (written MC_Margi) is easy to understand starting from the equation (20): nuisance parameters are sampled according to a Gaussian distribution and for each history, a deterministic resolution is done (GLS). At the end of every simulation, parameters and covariances are stored. When all the histories have been simulated, the covariance total theorem gives the final model parameters covariance. The methods is not developed here but more detailed can be found in papers [7,11].

3.4 BMC with systematic treatment

BMC method can deal with marginal parameters without deterministic approach. This work has been successfully implemented in the CONRAD code. One wants to find the posterior marginal probability function defined in equation (20). It is similar to the case with no nuisance parameters but with two integrals. Same weighting principle can be applied by replacing the likelihood term by a new weight $w_\theta(\boldsymbol{x}|\boldsymbol{y})$ defined by:

$$w_\theta(\boldsymbol{x}|\boldsymbol{y}) = \int \frac{p(\boldsymbol{\theta}|U)\cdot p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x},U)}{\int p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x},U)d\boldsymbol{x}}d\boldsymbol{\theta}. \tag{22}$$

The very close similarities between the case with no marginal parameter enabled a quick implementation and understanding. Finally, the previous equation gives:

$$p_\theta(\boldsymbol{x}|\boldsymbol{y},U) = p(\boldsymbol{x}|U)\cdot w_\theta(\boldsymbol{x}|\boldsymbol{y}). \tag{23}$$

Table 1. ^{238}U spin configuration considered for resonance waves s and p.

	l	s	J^{π}		g_J		onde
^{238}U (0^+)	0	1/2	$1/2^+$		1		s
	1	1/2	$1/2^-$	$3/2^-$	1	2	p

Both integrals can be solved using a Monte-Carlo approach. The integral calculation of the equation (22) is done as follow:

$$w_\theta(\boldsymbol{x}|\boldsymbol{y}) = \lim_{n_\theta\to\infty}\left(\frac{1}{n_\theta}\sum_{l=0}^{n_\theta}\frac{p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x},U)}{\int p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x},U)d\boldsymbol{x}}\right). \quad (24)$$

The n_θ histories are calculated by sampling according to $p(\boldsymbol{\theta}|U)$. The denominator's integral of equation (24) is then computed:

$$\forall l\in[\![1,n_\theta]\!],\quad \int p(\boldsymbol{x}|U)\cdot p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x},U)d\boldsymbol{x} = \lim_{n_x\to\infty}\left(\frac{1}{n_x}\sum_{m=0}^{n_x}p(\boldsymbol{y}|\boldsymbol{\theta}_l,x_m,U)\right). \quad (25)$$

The n_x histories are evaluated by sampling according to $p(\boldsymbol{x}|U)$. Mixing equations (24) and (25), the new weight $w_\theta(\boldsymbol{x}|\boldsymbol{y})$ is given by:

$$\lim_{n_\theta\to\infty}\left(\frac{1}{n_\theta}\sum_{l=0}^{n_\theta}\frac{p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x},U)}{\lim\limits_{n_x\to\infty}\left(\frac{1}{n_x}\sum\limits_{m=0}^{n_x}p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x}_m,U)\right)}\right). \quad (26)$$

Let us consider N_x is the number of history sampled according the prior parameters and N_θ is the number of history sampled according the marginal parameters. The larger those number are, the more converge the results will be. The previous equation can be numerically written with no limits as follow:

$$w_\theta(\boldsymbol{x}|\boldsymbol{y}) = \frac{N_x}{N_\theta}\sum_{l=0}^{N_\theta}\frac{p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x},U)}{\sum\limits_{m=0}^{N_x}p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x}_m,U)}. \quad (27)$$

In order to simplify the algorithm, N_θ and N_x are considered equal (introducing N as $N=N_\theta=N_x$). First, N samples are drawn according to $\boldsymbol{\theta}_k$ and $\boldsymbol{x}_k$. Equation (26) can be simplified as follow:

$$w_\theta(\boldsymbol{x}|\boldsymbol{y}) = \sum_{l=0}^{N}\frac{p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x},U)}{\sum\limits_{m=0}^{N}p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x}_m,U)}. \quad (28)$$

Table 2. Initial URR parameters with no correlation.

Parameters	Values
S_0	$1.290\times10^{-4}\pm10\%$
S_1	$2.170\times10^{-4}\pm10\%$

In CONRAD, the N^*N values of the likelihood are stored (i.e. $\forall(i,j)\in[\![1,N]\!]^2, p(\boldsymbol{y}|\theta_j,\boldsymbol{x}_i,U)$. Those values are required to perform statistical analysis at the end of the simulation. The weight for an history k is then calculated:

$$w_\theta(\boldsymbol{x}_k|\boldsymbol{y}_k) = \sum_{l=0}^{N}\frac{p(\boldsymbol{y}_k|\boldsymbol{\theta}_l,\boldsymbol{x}_k,U)}{\sum\limits_{m=0}^{N}p(\boldsymbol{y}_k|\boldsymbol{\theta}_l,\boldsymbol{x}_m,U)}. \quad (29)$$

To get the posterior mean values and the posterior correlation, one should apply the statistical definition and get the two next equations:

$$\langle x\rangle_N = \frac{1}{N}\sum_{k=0}^{N}\boldsymbol{x}_k\cdot\sum_{l=0}^{N}\frac{p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x}_k,U)}{\sum\limits_{m=0}^{N}p(\boldsymbol{y}|\boldsymbol{\theta}_l,\boldsymbol{x}_m,U)}, \quad (30)$$

$$\begin{aligned}\mathrm{Cov}(x_i,x_j)_N &= (M_{x_{post}})_{ij}^{N_x}\\ &= \langle x_ix_j\rangle_N - \langle x_i\rangle_N\langle x_j\rangle_N,\end{aligned} \quad (31)$$

with $\langle x_ix_j\rangle_N$ defined as the weighting mean of the two parameters product:

$$\langle x_ix_j\rangle_N = \frac{1}{N}\sum_{k=1}^{N}x_{i,k}\cdot x_{j,k}\cdot w_\theta(\boldsymbol{x}_k|\boldsymbol{y}_k). \quad (32)$$

4 Illustrative analysis on ^{238}U total cross section

4.1 Study case

The selected study case is just an illustrative example giving a very first step towards the validation of the method, its applicability and potential limitations. The ^{238}U total cross section is chosen and fitted on the unresolved resonance range, between 25 and 175 keV. The theoretical cross section is calculated using the R mean matrix model. The main sensitives parameters in this energy range is the two first strength functions $S_{l=0}$ and $S_{l=1}$. Tables 1 and 2 give respectively the spin configurations and the prior parameters governing the total cross section. An initial relative uncertainties of 15% is taken into account with no correlations. The experimental dataset used comes from the EXFOR database [12]. A 1% arbitrary statistical uncertainty is chosen.

Table 3. Results comparison between the different methods implemented in CONRAD for ^{238}U total cross section.

Physical quantities	Prior	GLS	BMC	Importance
S_0 ($\times 10^{-4}$)	1.290	1.073	1.072	1.073
δS_0 ($\times 10^{-6}$)	19.35	9.013	9.122	9.020
S_1 ($\times 10^{-4}$)	2.170	1.192	1.193	1.192
δS_1 ($\times 10^{-6}$)	32.55	6.089	6.135	6.095
Correlation	0.000	−0.446	−0.425	−0.447
χ^2	381.6	8.78	8.79	8.78
$\langle \delta_y \rangle$ (%)	2.78	0.64	0.65	0.64

Table 4. Results comparison when a normalisation marginal parameter is taken into account. AN_Margi represents the deterministic method, MC_Margi represents the semi-stochastic resolution, BMC is the classical method and the last column called Importance is the BMC where an importance function is used for the sampling.

Physical quantities	AN	MC_Margi	BMC	Importance
S_0 ($\times 10^{-4}$)	1.073	1.073	1.063	1.074
δS_0 ($\times 10^{-6}$)	9.634	9.469	8.939	9.490
S_1 ($\times 10^{-4}$)	1.192	1.194	1.215	1.193
δS_1 ($\times 10^{-6}$)	11.60	11.52	8.945	11.54
Correlation	0.081	0.035	−0.061	0.044
$\langle \delta_y \rangle$ (%)	1.19	1.17	0.93	1.18

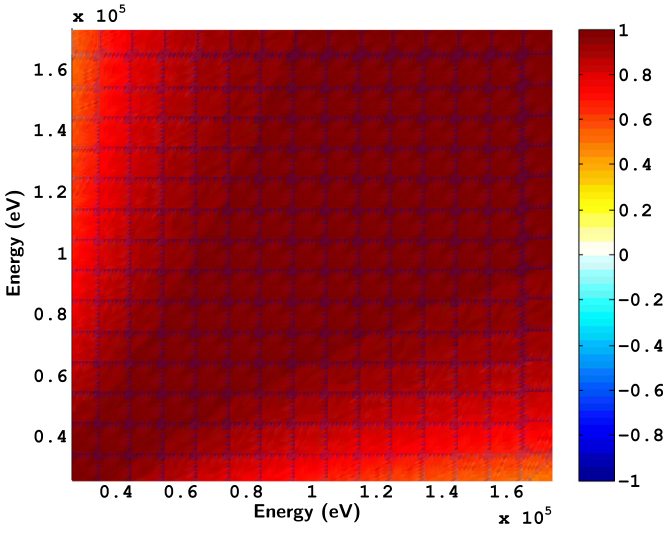


Fig. 3. A priori covariance matrix on ^{238}U total cross section.

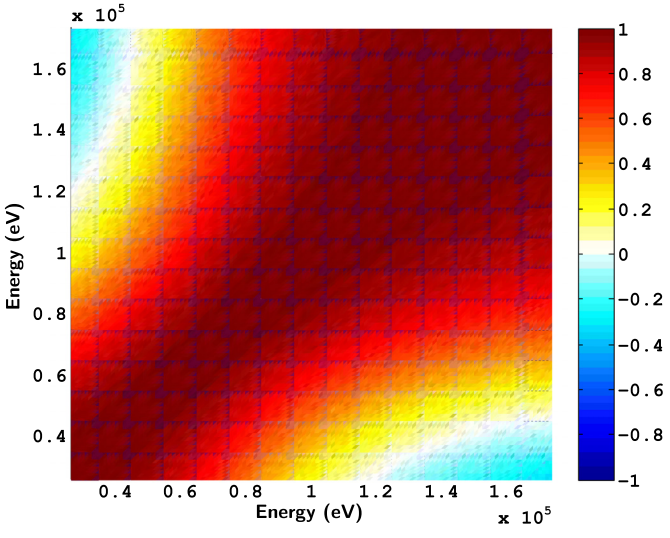


Fig. 4. A posteriori covariance matrix on ^{238}U total cross section.

4.2 Classical resolution with no marginal parameters

All the methods have been used to perform the adjustment. One million histories haven been simulated in order to get the statistical values converged below 0.1% for the mean values and 0.4% for the posterior correlation. The convergence is driven by how far the solution is from the prior values. Table 3 shows the overall good results coherence. Very small discrepancies can be seen for the classical methods caused by the convergence issues. For a similar number of history, the importance methods converged better than the classical BMC. The prior covariance on the total cross section is given on Figure 3. The anti-correlation created between S_0 and S_1 directly give correlations between the low energy and the high energy (see Fig. 4). The prior and posterior distribution engaged in the BMC methods are given in Figure 5. One can notice the Gaussian distributions for all the parameters (both prior and posterior).

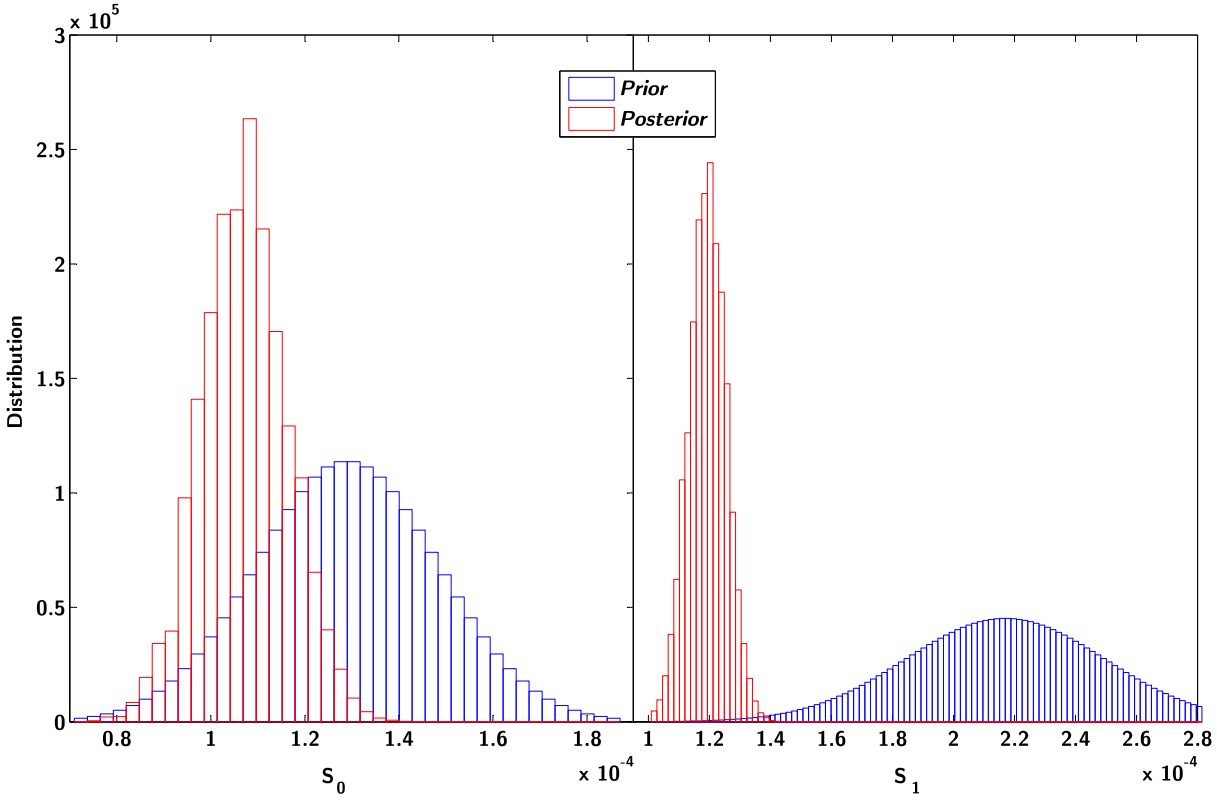


Fig. 5. S_0 and S_1 distributions obtained with the classical BMC method.

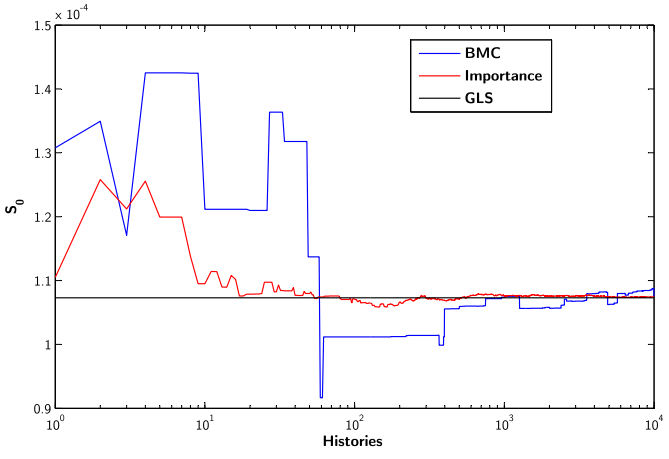


Fig. 6. S_0 mean value convergence.

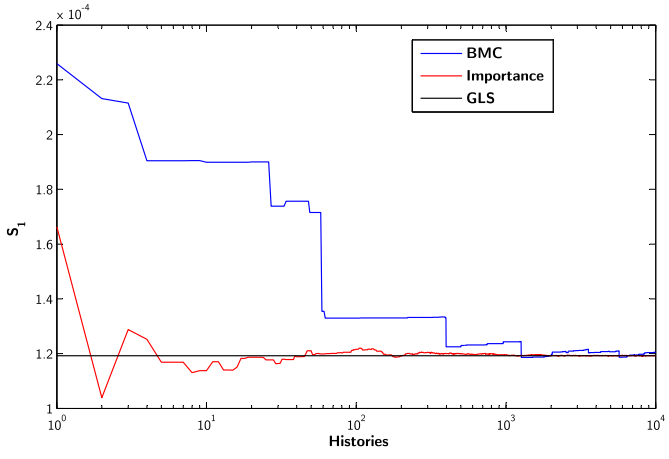


Fig. 7. S_1 mean value convergence.

4.3 Using marginalization methods

This case is very closed to the previous paragraph. But this time, a nuisance parameter is taken into account. More precisely, a normalization is considered with a systematic uncertainties of 1%. 10 000 histories are sampled for the MC_Margi case (semi-stochastic resolution) and $10\,000 \times 10\,000$ for BMC methods. For the importance resolution, the biasing function is the posterior solution coming from the deterministic resolution. All the results seem to be equivalent, as shown in Table 4. However, the classical BMC is not fully converged because slight differences are found between the means values. Figures 6 and 7 show the mean values convergence using the stochastic resolutions, showing one more time not converged results with the classical BMC method. Calculation time are longer with marginal parameters. This is explained by the method which the idea is to perform a double Monte-Carlo integration. The good coherence on the mean values and correlation between parameters give identical posterior correlation on the total cross section. Figure 8 shows the a posteriori covariance, whatever methods chosen.

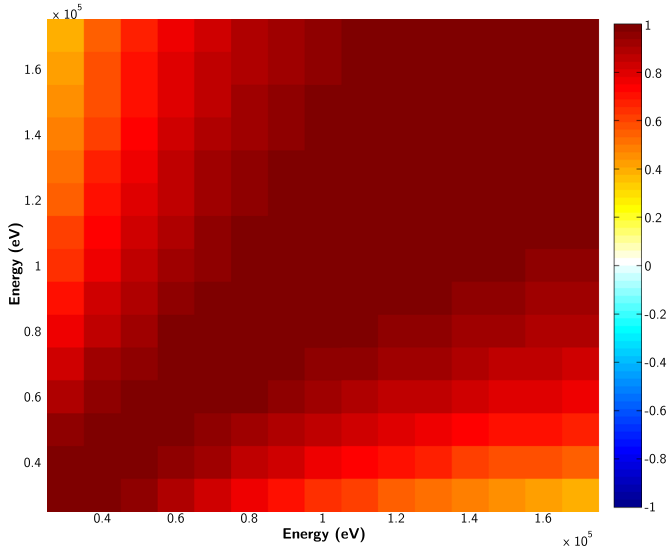


Fig. 8. A posteriori covariance of the ^{238}U total cross section.

5 Conclusions

The use of BMC methods was exposed in this paper and an illustrative analysis was detailed. One important point is that these methods could be used for resonance range analysis (both resolved and unresolved resonance) as well as higher energy models. In addition, both microscopic and integral data assimilation could be achieved. Nevertheless, the major issue is related to the convergence estimator: depending on which parameters are investigated (central values or correlation between them), the number of histories (sampling) could be very different. Indeed, special care should be taken for the correlations calculation because an additional order of magnitude of sampling histories could be necessary. Furthermore, it was shown that sampling priors are not a problem. It is more efficient to properly sample in the phase space region where the likelihood is large. In this aspect, Importance and Metropolis algorithms are working better than brute force ("classical") Monte-Carlo. It also highlights the fact that pre-sampling prior with a limited number of realizations could be inadequate for further inference analysis. The integral data assimilation with feedback directly on the model parameter is too much time consuming. However, a simplified model can be adopted by using a simple model with a linear approach for instance (to predict integral parameters response to input parameters). Such approximation would be less consuming but will erase non-linearity effect that may be observed in the posterior distribution. Such study should be performed with extensive cases to improve the Monte-Carlo methods.

Concerning BMC inference methods, in the future, other Markov chain algorithms will be developed in CONRAD code and efficient convergence estimators will be proposed as well. The choice of Gaussian probability functions for both prior and likelihood will be challenged and discussed.

More generally, an open range of scientific activities will be investigated. In particular, one major issue is related to a change in paradigm: to go beyond covariance matrices and deal with parameters knowledge taking into account full probability density distributions. In addition, for end-up users, it will be necessary to investigate the feasibility of a full Monte-Carlo approach, from nuclear reaction models to nuclear reactors or integral experiments (or any other applications) without format/files/processing issues which are most of the time bottlenecks.

The use of Monte-Carlo could solve a generic issue in nuclear data evaluation related to difference of information given in evaluated files: in the resonance range where cross section uncertainties and/or nuclear model parameters uncertainties are compiled and in the higher energy range where only cross section uncertainties are formatted. This could simplify the evaluation of full covariance over the whole energy range.

Author contribution statement

The main author (E. Privas) has done the development and the analysis of the new stochastic method. The other authors participated in the verification process and in the overall development of CONRAD.

References

1. T. Bayes, Philos. Trans. R. Soc. London **53**, 370 (1763)
2. F. Frohner, JEFF Report 18, 2000
3. P. Archier, C. De Saint Jean, O. Litaize, G. Noguère, L. Berge, E. Privas, P. Tamagno, Nucl. Data Sheets **118**, 488 (2014)
4. T. Cover, J. Thomas, *Elements of information theory* (Wiley-Interscience, New York, 2006)
5. R. Fletcher, *Practical methods of optimization* (John Wiley & Sons, New York, 1987)
6. P.A.C. De Saint Jean, E. Privas, G. Noguère, in *Proceeding of the International Conference on Nuclear Data, 2016*
7. B. Habert, Ph.D. thesis, Institut Polytechnique de Grenoble, 2009
8. H. Mitani, H. Kuroi, J. Nucl. Sci. Technol. **9**, 383 (1972)
9. H. Mitani, H. Kuroi, J. Nucl. Sci. Technol. **9**, 642 (1972)
10. A. Gandini, Y. Ronen Ed. *Uncertainty analysis* (CRC Press, Boca Raton, 1988)
11. C. De Saint Jean, P. Archier, E. Privas, G. Noguère, O. Litaize, P. Leconte, Nucl. Data Sheets **123**, 178 (2015)
12. C. Uttley, C. Newstead, K. Diment, in *Nuclear Data For Reactors, in Conference Proceedings* (Paris, 1966), Vol. 1, p. 165

Cite this article as: Edwin Privas, Cyrille De Saint Jean, Gilles Noguere, On the use of the BMC to resolve Bayesian inference with nuisance parameters, EPJ Nuclear Sci. Technol. **4**, 36 (2018)