Why trust nuclear data evaluations with Bayesian networks

Georg Schnabel1,2*, Roberto Capote1, and Daniel L. Aldama1,2
1NAPC-Nuclear Data Section, International Atomic Energy Agency, A-1040 Vienna, Austria
2Centro de Aplicaciones Tecnológicas y Desarrollo Nuclear (CEADEN)

Abstract. Bayesian networks were recently suggested as a framework for nuclear data evaluation. Their theory was to some extent described in a recent preprint and some example evaluations were presented. However, due to their newness in the context of nuclear data evaluation and consequently the lack of experience with them within the community makes it difficult to develop trust in the underlying methodology and consequently also the results produced by it. In this contribution, we aim to make a case why evaluators can trust this methodology in principle but will also elaborate on the fact that Bayesian networks are not a silver bullet for evaluation work. On the contrary, evaluators must assess and quantify essential assumptions about nuclear models and experiments with the same diligence that is already necessary for the application of the well-established Generalized Least Squares (GLS) method. We also explain that the increased ease and flexibility to introduce assumptions regarding nuclear models, experiments and their relationships can help an evaluator to rigorously account for assumptions that are very often neglected in evaluations with the GLS method, such as the non-negativity of cross sections, relative experimental normalization uncertainties and the non-linearity in ratios of cross sections. We believe that adopting the Bayesian network paradigm can help both humans to produce evaluations with clearly traceable assumptions and machines to deal with nuclear data more efficiently in terms of execution speed and storage size requirements.

1 Introduction

The production of a good nuclear data evaluation that will be trusted and, therefore, broadly used and accepted depends on several ingredients. Experimental data must be collected, analyzed and their uncertainties assessed. A good model must be selected to fit the data. The model may be an optical model for fast incident neutrons, a Padé approximation or spline for purely data-driven evaluations, such as in the IRDFF-II library [1], or an empirical model extracted from systematic studies of experimental data for different isotopes, such as the Wahl model [2] for fission-product yields.

Several methods to quantify uncertainties have been explored and to some extent used in the past, such as [3-5]. However, for evaluations based on a rigorous and quantitative assessment of experimental error sources in form of a covariance matrix, the GLS method [6, 7] is the broadly accepted and employed method in the nuclear data field since a long time, e.g., [8].

To increase the trust in a new approach, it is pertinent to review the foundation of the GLS method and to highlight the methodological differences of the new approach. In this contribution, we want to perform this investigation for the Bayesian network approach in the form suggested for nuclear data evaluation in [9].

To this end, we first elaborate on the foundation of the GLS method, which in our opinion explains why the GLS method is broadly accepted. Then we explain how the linearity assumption of the GLS method can be abandoned and a refined and methodologically sound estimation procedure constructed. Finally, we discuss the Bayesian network perspective and how it facilitates the evaluation process.

2 Why trust the GLS method

Evaluation methods are specific realizations of statistical estimation methods that allow to estimate quantities of interest using measurements and in some situations model calculations. There are different approaches to derive statistical estimation methods and their relative merits to each other can be expressed in terms of bias and expected mean squared error. If an estimator is unbiased, it will produce in average over a large number of measurement campaigns the right result. The expected mean squared error is a measure how close the value of an estimation procedure can be expected to be to the true value. Both unbiasedness and an as small as possible expected mean squared error are two highly desirable properties of an estimator in general and for nuclear data evaluation in particular.

If measurements \( \tilde{y} \) and quantities of interest \( \tilde{x} \) are related by

\[
\tilde{y} = S \tilde{x} + \tilde{e}
\]

(1)

with the measurement error vector \( \tilde{e} \) being the realization of a random vector with its expectation given by a vector

*E-mail: g.schnabel@iaea.org

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
of zeros and covariance matrix $\Sigma$, the GLS equation
\[
\hat{x}_{\text{est}} = \left( S^T \Sigma^{-1} S \right)^{-1} S^T \Sigma^{-1} \hat{y}
\] (2)

yields an estimate $\hat{x}_{\text{est}}$ for the vector with the quantities of interest. It can be shown that this estimation formula represents the best linear unbiased estimator (BLUE) [7, 10]. Hence, there is no other estimation formula of the form $\hat{x}_{\text{est}} = X\hat{y} + \hat{a}$ with some choice for the matrix $X$ and vector $\hat{a}$ that yields an unbiased estimate of $\hat{x}$ with smaller expected mean squared error.

It should be noted here that the specifications 'linear' and 'unbiased' are essential. For example, removing the restriction that the estimator needs to be unbiased opens up the class of James-Stein estimators [11] that have lower or equal expected mean squared error. Equally important, the unbiasedness and minimum mean squared error property only hold if the covariance matrix $\Sigma$ appearing in the estimation formula coincides with the covariance matrix of $\hat{x}$ in the true data generating process specified in eq. (1). Likewise, these two properties are also lost if the matrix $S$ in the estimation formula does not correspond to the true one in the data generating process. Therefore, in order to take advantage of the desirable properties of the GLS method, due diligence is required in the assessment of experimental data and their covariance matrices and the selection of a suitable model (manifested in the choice of $S$).

The proof that the GLS method is BLUE does not depend on the assumption of a specific distribution but only linearity and the knowledge of the first and second moments are required. However, if we impose the additional assumption that $\hat{x}$ comes from a multivariate normal distribution with a zero mean vector and covariance matrix $\Sigma$, another way to derive the GLS formula is maximum likelihood estimation. The likelihood to obtain a specific measurement vector $\hat{y}$ is given by
\[
\rho(\hat{y} | \hat{x}) = N(\hat{y} | S\hat{x}, \Sigma)
\] (3)

with $S\hat{x}$ being the mean vector and $\Sigma$ the covariance matrix of the multivariate normal distribution. As we don’t know $\hat{x}_{\text{true}}$ in the distribution above, the maximum likelihood principle prescribes that we should select as estimate $\hat{x}_{\text{est}}$ the vector that maximizes $\rho(\hat{y} | \hat{x})$ based on the known measurement vector $\hat{y}$ and covariance matrix $\Sigma$. The solution of this optimization problem is again given by the GLS method in eq. (2) [7].

Another approach to derive the Bayesian version of the GLS method is to impose a multivariate normal prior with mean vector $\bar{x}_0$ and covariance matrix $Q$ and solve the Bayesian update equation,
\[
\rho(x | \hat{y}) \propto \rho(\hat{y} | x) \rho(x)
\] (4)

with $\rho(\hat{y} | x)$ given by eq. (3) and $\rho(x) = N(x | \bar{x}_0, Q)$. It can be shown that the posterior distribution $\rho(x | \hat{y})$ is multivariate normal and its mean vector given by
\[
\hat{x}_{\text{est}} = \left( S^T \Sigma^{-1} S + Q^{-1} \right)^{-1} \left( Q^{-1} \bar{x}_0 + S^T \Sigma^{-1} \hat{y} \right). 
\] (5)

Choosing an uninformative prior covariance matrix $Q = \eta Q$ with $Q$ being an arbitrary full rank matrix and taking the limit $\eta \to \infty$, yields again the frequentist GLS estimate in eq. (2). It can also be shown that with an increasing number of data points, the Bayesian GLS estimate converges to the frequentist one also for a finite prior covariance matrix $Q$.

In summary, the authors believe that the various ways to derive the GLS formula from distinct assumptions and reasonable principles are reasons for the broad acceptance of the GLS method in the nuclear data field. Other contributing factor may be the simplicity of the estimation formula and the long and successful history over decades of the GLS method to perform high-quality evaluations. However, we also emphasized that the GLS method loses its desirable properties if some of the underlying assumptions are violated, such as wrongly specified experimental covariance matrices, models or in the presence of outliers.

3 Going beyond GLS

A serious limitation of the GLS method is the assumption that quantities of interest are linearly related to the measurements in eq. (1). Typical relationships in nuclear data evaluation are non-linear:

- In the neutron data standards project [13], quantities of interest are cross sections on an energy mesh and some measurements are ratios of cross sections.
- Padé approximations, R-matrix fits and optical models all exhibit non-linear relationships between parameters and cross sections.

One important improvement is therefore to remove the linearity assumption. To this end, we abandon the linear model $M_{\text{lin}}(\hat{x}) = S\hat{x}$ in eq. (3) and replace it by a non-linear one $M(\hat{x})$ for which we don’t assume any particular functional shape,
\[
\rho(\hat{y} | \hat{x}) = N(\hat{y} | M(\hat{x}), \Sigma).
\] (6)

The maximum likelihood principle and the Bayesian approach can be employed irrespective of whether $M(\hat{x})$ is linear or non-linear. In the case of a non-linear model, however, we can’t identify the vector $\hat{x}_{\text{map}}$ associated with the posterior maximum by an analytic estimation formula anymore and depend on numerical optimization routines.

In summary, the authors believe that the various ways to derive the GLS formula from distinct assumptions and reasonable principles are reasons for the broad acceptance of the GLS method in the nuclear data field. Other contributing factor may be the simplicity of the estimation formula and the long and successful history over decades of the GLS method to perform high-quality evaluations. However, we also emphasized that the GLS method loses its desirable properties if some of the underlying assumptions are violated, such as wrongly specified experimental covariance matrices, models or in the presence of outliers.

Choosing an uninformative prior covariance matrix $Q = \eta Q$ with $Q$ being an arbitrary full rank matrix and taking the limit $\eta \to \infty$, yields again the frequentist GLS estimate in eq. (2). It can also be shown that with an increasing number of data points, the Bayesian GLS estimate converges to the frequentist one also for a finite prior covariance matrix $Q$.

In summary, the authors believe that the various ways to derive the GLS formula from distinct assumptions and reasonable principles are reasons for the broad acceptance of the GLS method in the nuclear data field. Other contributing factor may be the simplicity of the estimation formula and the long and successful history over decades of the GLS method to perform high-quality evaluations. However, we also emphasized that the GLS method loses its desirable properties if some of the underlying assumptions are violated, such as wrongly specified experimental covariance matrices, models or in the presence of outliers.
the integral whose solution is then given by the GLS estimation formula eq. (5). Posterior expectation \( \hat{x}_{\text{est}} \) and the vector \( \hat{x}_{\text{map}} \) associated with the posterior maximum will be very close to each other. In this case, the same properties as for the GLS method approximately hold even though we deal with a non-linear model. Of course, if in the data generating process,

\[
\hat{y} = M(\hat{x}) + \hat{e},
\]

the non-linear model or covariance matrix associated with \( \hat{e} \) is misspecified, the Bayesian approach is prone to the same distortions as the GLS method.

If the likelihood is not peaked sharply enough, the non-linearity of the model will impact the result. The posterior expectation of \( \hat{x}_{\text{est}} \) will differ from the vector \( \hat{x}_{\text{map}} \) associated with the maximum of the posterior distribution. In this situation it is not entirely clear whether \( \hat{x}_{\text{est}} \) or \( \hat{x}_{\text{map}} \) is a better summary statistic of the posterior distribution. The approach in the Bayesian network paper [9] is to use \( \hat{x}_{\text{map}} \) as it can be obtained by optimization, which is usually more tractable than efficient sampling to target the posterior distribution for time consuming models. The posterior covariance matrix \( \Sigma_{\text{post}} \) obtained by using the sandwich formula with the Jacobian matrix of the non-linear model computed at the posterior maximum will then also differ from the covariance matrix obtained by

\[
\Sigma_{\text{post}} = \int (\hat{x} - \hat{x}_{\text{est}})(\hat{x} - \hat{x}_{\text{est}})^T \rho(\hat{y}|\hat{x}) \rho(\hat{x}).
\]

In order to obtain the correct covariance matrix of the posterior distribution, numerical integration routines may be used if the dimension of \( \hat{x} \) is not too large or Monte Carlo integration otherwise.

Despite these potential shortcomings, the estimate \( \hat{x}_{\text{map}} \) will still be a better choice than the result of the GLS method with an expansion point different from the posterior maximum. Likewise the approximation to the posterior covariance matrix \( \Sigma_p \) is more defensible than the GLS solution based on a linear approximation of the non-linear model at some location in a parameter domain without significant posterior weight.

Finally, we remark that the Levenberg-Marquardt algorithm [14, 15] is an efficient algorithm to reliably locate the maximum of the posterior distribution in eq. (4) as long as the Jacobian matrix of the non-linear model is available. In practice, the Jacobian matrix can be computed numerically by a finite difference approximation and sometimes more efficiently by an adjoint approach or automatic differentiation.

We also want to stress that the approach sometimes referred to as iterative GLS is inferior to the Levenberg-Marquardt algorithm as it does not control for the step size nor ensures that the updated vector for the next iteration \( \hat{x}_{\text{est}} \) is really an improvement according to some metric, such as the value of the posterior density function in eq. (4).

The outlined assumptions, which are a multivariate normal likelihood, a multivariate normal prior distribution and a potentially non-linear model are exactly the assumptions taken in [9] elaborating on Bayesian networks in the nuclear data context in more technical detail.

### 4 Why adopting the Bayesian network paradigm

Equation (6) establishes the probabilistic link from quantities of interest \( \hat{x} \) to the observed quantities \( \hat{y} \). In other words, it specifies that the model \( M(\hat{x}) \) with the right choice of \( \hat{x} \) predicts the data up to a difference characterized by the covariance matrix \( \Sigma \). Regarding both the covariance matrix and the model as black-box monoliths is detrimental to understanding the assumptions taken and prevents numerical algorithms to take advantage of their structure. In the Bayesian network paradigm, such structures are preserved and exploited during computation.

For example, an experimental covariance matrix for an experimental dataset with statistical uncertainties and a normalization uncertainty can be written as \( \Sigma = D + \eta I \) with a diagonal matrix \( D \) capturing the statistical uncertainties and a scalar value \( \eta \) for the normalization uncertainty. The notation \( I \) designates a matrix of ones. The resulting matrix is dense and a naive inversion of such a matrix of dimension 50000 × 50000 would be infeasible on desktop computers but is trivial if taking advantage of its structure as sum of two components by applying the Sherman-Morrison formula, \( \Sigma^{-1} = D^{-1} - \eta D^{-1}IT(I + \eta IT)^{-1} \). Furthermore, apart from the ability to scale problem sizes up, keeping uncertainty components separated helps humans to understand the assumptions taken as decomposing an experimental covariance matrix into the individual uncertainty contributions is in general not possible.

Not only the covariance matrix possesses in general structure but also the model \( M \) does. For example, the predictions of a reaction model \( M_{\text{reaction}} \) must be convoluted to compare them with an experiment with finite energy resolution. Convolution can be regarded as its own model \( M_{\text{conv}} \) and the resulting probabilistic link would take the form

\[
\hat{y} = M_{\text{conv}}(M_{\text{reaction}}(\hat{x})) + \hat{e}.
\]

If the reaction model is a microscopic one, the ability to fit it closely to available experimental data is sacrificed for global predictive power. In such a case, we can introduce a model defect \( \hat{g} \) with a Gaussian process prior to obtain the specification:

\[
\hat{y} = M_{\text{conv}}(M_{\text{reaction}}(\hat{x}) + \hat{g}) + \hat{e}.
\]

This model specification is even more nested and has structure that can be exploited. This structure can also be visualized as a graph, a Bayesian network,
with \( \tilde{z} = M_{\text{rec}}(\tilde{x}) + \tilde{g} \).

The modular perspective coming along with the Bayesian networks paradigm also helps in the construction of statistical models for the evaluation process. Once an essential set of models (in our understanding) is implemented, such as convolution, linear interpolation, exponentiation (to implement log-normal distributions), ReLU transformations (to implement truncated normal distributions), energy shifts and scalings (for TOF experiments) they can be reused and stacked together according to the specifics of an evaluation.

In practice, an evaluator builds up a database—for simplicity one can think of a table with columns—that contains the quantities of interest and the measurements. Each row is associated with an index. Probabilistic links between variables can then be established by constructing JSON-like objects with fields specific to the defined mapping. Listing 1 shows an example how it would be done with the nucdataBaynet R package [16]. Database query operations are used to extract the information that is required for the specification of the mapping, such as the indices src idx of the rows with the source quantities and for which target indices tar idx the interpolation should be calculated for. The fields src_x and tar_x contain the energies of the source and target mesh, respectively. More details can be found in [9] from where this example is taken.

### 5 Summary and outlook

We reviewed the GLS method and sketched several ways to derive it using different principles and assumptions. This exercise also made the strengths and limitations of the GLS method more tangible. We went on and dropped the linearity assumption and explained how a refined iteration can be constructed in this case. We explained that the refined procedure approximately shares the same mathematical properties if the likelihood is sharply peaked and argued that the refined scheme is still superior to the GLS method otherwise. One reason being that in each iteration of the refined scheme, it is ensured that the proposed vector for the next iteration is indeed an improvement.

Finally, we explained that the Bayesian network perspective is about preserving and exploiting the structure present in the covariance matrix and the models. We emphasized that for the purpose of this paper, models were understood in a broad sense and encompassed apart from reaction models also mathematical operations, such as convolution and linear interpolation.

Finally, we discussed a simple example of an evaluation situation involving a reaction model, experimental data with limited energy-resolution and a model defect. We showed how such models can be depicted as a Bayesian network and gave a glimpse on the construction of such Bayesian networks from the perspective of the practitioner.

We believe that the Bayesian network approach is very well suited to link together information of different reaction systems and models in ways that go beyond the status quo. We are planning to couple the Bayesian network approach with codes for reading and writing essential nuclear data formats, such as the ENDF format, to facilitate the production of high-quality evaluations within the Bayesian network paradigm in the future.

### References


```r
list(
  mapname = "truexs_to_experiment",
  maptype = "linearinterpol_map",
  src_idx = dt[NODE=="truexs", IDX],
  tar_idx = dt[NODE=="exp", IDX],
  src_x = dt[ NODE=="truexs", ENERGY],
  tar_x = dt[ NODE=="exp", ENERGY],
)
```

Listing 1. Example R code to create a list with the definition of a linear interpolation mapping