EPJ Nuclear
Sciences
& Technologies

REGULAR ARTICLE

OPEN ◂ ACCESS

Bayesian optimization of generalized data

Goran Arbanas[1,*], Jinghua Feng[1,2], Zia J. Clifton[1,3], Andrew M. Holcomb[1], Marco T. Pigni[1], Dorothea Wiarda[1], Christopher W. Chapman[1,4], Vladimir Sobes[1], Li Emily Liu[2], and Yaron Danon[2]

[1] Nuclear Data and Criticality Safety Group, Reactor and Nuclear Systems Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6171, USA
[2] Department of Mechanical, Aerospace, and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, USA
[3] Department of Physics and Astronomy, University of Alabama, Huntsville, AL 35899, USA
[4] Nuclear & Radiological Engineering & Medical Physics, Georgia Institute of Technology, Atlanta, GA 30332-0745, USA

Abstract. Direct application of Bayes' theorem to generalized data yields a posterior probability distribution function (PDF) that is a product of a prior PDF of generalized data and a likelihood function, where generalized data consists of model parameters, measured data, and model defect data. The prior PDF of generalized data is defined by prior expectation values and a prior covariance matrix of generalized data that naturally includes covariance between any two components of generalized data. A set of constraints imposed on the posterior expectation values and covariances of generalized data via a given model is formally solved by the method of Lagrange multipliers. Posterior expectation values of the constraints and their covariance matrix are conventionally set to zero, leading to a likelihood function that is a Dirac delta function of the constraining equation. It is shown that setting constraints to values other than zero is analogous to introducing a model defect. Since posterior expectation values of any function of generalized data are integrals of that function over all generalized data weighted by the posterior PDF, all elements of generalized data may be viewed as nuisance parameters marginalized by this integration. One simple form of posterior PDF is obtained when the prior PDF and the likelihood function are normal PDFs. For linear models without a defect this PDF becomes equivalent to constrained least squares (CLS) method, that is, the χ^2 minimization method.

1 Introduction

Advancement of scientific understanding of natural phenomena is partly due to a complementary activity of conceiving better models while performing experiments that test those models or explore their limits of validity. There are many historical examples of constructive interplay between conceptual models and experiments that have resulted in improved understanding of some phenomena. Although experiments are often viewed as objective and independent from the model they were designed to test, their conception and design may be based upon some model in which measured data are to be interpreted and compared to the prediction of the model being tested.

As our understanding of nature increases, the harmony between models and experiments will likely increase to reveal a complementary relationship between the conceptual models and experimental data. One practical way of formalizing the complementary nature of model parame-

ters and experimental data is by defining their union that is conventionally called generalized data, although generalized parameters may be a more appropriate term because the aspect of fitting, which is conventionally restricted to parameters, is extended to experimental data by virtue of this generalization. This union is referred to as generalized data herein to maintain consistency with nomenclature used in literature. For defective models, this concept is extended to include the model defect data as the third component of generalized data.

In the process of applying Bayes' theorem [1] to generalized data, the farsightedness of Edwin T. James insight is apparent in this remark: "But every Bayesian problem is open ended; no matter how much analysis you have completed, this only suggests still other kinds of prior information that you might have had and therefore still more interesting calculations that need to be done to get still deeper insight into the problem" [2]. From this perspective, prior expectation values of the generalized data (model parameters, measured experimental data, and the model defect data) and their generalized covariance matrix that by definition contains all pair-wise covariances among the three components of generalized data, are

---

* e-mail: arbanasg@ornl.gov

formally introduced as priors in the context of Bayes' theorem.

Direct application of Bayes' theorem to this generalized data yields a posterior PDF of generalized data from which posterior expectation values of generalized data and their posterior generalized covariance matrix can be computed. Furthermore, the constraint relating posterior expectation values of the three components of the generalized data (model parameters, measured experimental data, and the model defect data) and their covariance is imposed. The formal framework used in this work has been conducive to recognizing a connection between model defect and constraints presented here.

Key expressions related to Bayesian treatment of model defect have already been derived using standard notation in [3], and much has already been learned from various implementations of various model defects and their effects on neutron cross sections in [3–7].

Formal expressions in generalized data notation are derived in Section 2. In Section 2.1, Bayes' theorem [1] is applied directly to generalized data, yielding general expressions for generalized data optimization, that is, for simultaneous optimization of model parameters, measured experimental data, and model defect data. From the perspective of Bayes' theorem, generalized data have prior and posterior PDF, meaning that each of its three components will also have posterior values just like model parameters would.

In Section 2.2, normal PDFs are used to derive a special case of the posterior PDF, and in Section 2.3, an alternative derivation focusing on posterior PDF of model parameters is presented.

In Section 4, the connection between our Bayes' generalized data optimization method and the constrained generalized least squares (CGLS) method [8] is discussed. A linear form of CGLS has been implemented in the module TSURFER [9] of the SCALE system [10]. The CGLS method defines an objective function, often called χ^2, along with a constraint that is enforced by Lagrange multipliers in those codes. The relationship to the conventional χ^2 minimization is also discussed.

2 Derivation

This application of Bayes' theorem is relatively simple, illustrating the words of Bayesian advocate late Edwin T. Jaynes: "The difficulties are never mathematical," but they could be described more accurately as "conceptual difficulties" [2].

The posterior PDF of generalized data is derived by application of the Bayes' theorem in Section 2.1. Compact generalized data notation is used whenever possible, except when particular aspects of model parameters and experimental data must be distinguished. A distinction is maintained in this derivation between probability $p(z|\langle z \rangle, \ldots)$ that a variable attains a particular value z and the corresponding expectation values denoted by $\langle z \rangle$.

Application of Bayes' theorem to generalized data implies that both prior parameters and the new experi-

mental data together constitute prior generalized data. That is, since a given experimental data has not yet been used in an evaluation, it should be viewed as a prior from the perspective of the Bayes' theorem. Since experimental data are to be treated as a prior, it may be expected that Bayes' theorem would naturally yield an optimal posterior PDF of experimental data, just as it does conventionally for model parameter values. Prior model parameters, measured data, and model defect data, simultaneously inform each other in a way that yields their optimal posterior joint PDF.

An expression for posterior PDF of generalized data is derived and then used to compute posterior expectation values and covariances. In Section 2.3, conventional methods are used to derive posterior probability distribution of model parameters, where the total probability theorem is used to derive an expression equivalent to that derived in Section 2.1. The derivation based on generalized data notation may be more compact, but the conventional derivation may be more intuitive to those accustomed to PDFs of model parameters. Mathematical guidelines and nomenclature used in derivations are listed here for convenience:

– Generalized data, $z \equiv (P, D, \delta)$, is a union of model parameters (P), measured experimental data (D), and model defect data (δ).
– The covariance matrix of generalized data, **C**, by definition contains covariance among all three components of generalized data.
– Bayes' theorem is applied directly to generalized data.
– A set of constraints, f, on posterior expectation values and covariances of generalized data is imposed on the posterior PDF, and formally solved by the Lagrange multiplier method.
– A model, i.e., theory $T(\cdot)$, relates model parameters, measured data, and model defect data in definition of constraints.
– Distinction between expectation values, i.e., $\langle z \rangle$, and their instance value, i.e., z, is maintained in PDFs, and
– Posterior expectation values are indicated by a prime, $\langle z \rangle'$ and **C**$'$, while unprimed ones represent prior expectation values, $\langle z \rangle$ and **C**.

The following expressions are referenced within the paper by assigning a context-dependent meaning to generic variables α, β, and γ used below. A generic Bayes' theorem could be stated as

$$p(\alpha|\beta\gamma) = p(\alpha|\gamma\beta) \propto p(\alpha|\gamma)p(\beta|\alpha\gamma), \qquad (1)$$

while a generic product rule of probability theory is

$$p(\alpha\beta|\gamma) = p(\alpha|\beta\gamma)p(\beta|\gamma). \qquad (2)$$

Integrating equation (2) over β yields the law of total probability,

$$p(\alpha|\gamma) = \int d\beta \quad p(\alpha|\beta\gamma) \times p(\beta|\gamma), \qquad (3)$$

that is equivalent to marginalization of nuisance parameter β by integrating over its all possible values.

2.1 Derivation in generalized data notation

A definition of generalized data vector z is extended to include model defect data δ in addition to parameters P and measured data D, namely:

$$z \equiv (P, D, \delta), \tag{4}$$

where prior values of generalized data are

$$\langle z \rangle \equiv (\langle P \rangle, \langle D \rangle, \langle \delta \rangle), \tag{5}$$

and where the prior covariance matrix of generalized data is represented by a 3×3 block diagonal matrix C

$$C \equiv \langle (z - \langle z \rangle)(z - \langle z \rangle)^{\mathsf{T}} \rangle \tag{6}$$

$$\equiv \begin{pmatrix} \mathbf{M} & \mathbf{W} & \mathbf{X} \\ \mathbf{W}^{\mathsf{T}} & \mathbf{V} & \mathbf{Y} \\ \mathbf{X}^{\mathsf{T}} & \mathbf{Y}^{\mathsf{T}} & \mathbf{\Delta} \end{pmatrix}, \tag{7}$$

where square matrices $\mathbf{M}$, $\mathbf{V}$, and $\mathbf{\Delta}$ along the diagonal represent covariance matrix of parameters, measured data, and the model defect, respectively, while $\mathbf{W}$, $\mathbf{X}$, and $\mathbf{Y}$ are their respective pair-wise covariances. Prior expectation value of model defect $\langle \delta \rangle$ is a vector of the same size as measured data $\langle D \rangle$, and it is expectation value of deviations between model predictions $T(P)$ and the measured data caused by the model defect alone. The Bayes' theorem is used to write a posterior PDF for $z \equiv (P, D, \delta)$ by making the following substitution in equation (1),

$$\begin{aligned} \alpha &\rightarrow z \\ \beta &\rightarrow f \\ \gamma &\rightarrow \langle z \rangle, C \end{aligned} \tag{8}$$

to obtain

$$p(z | \langle z \rangle, \mathbf{C}, f) \propto p(z | \langle z \rangle, \mathbf{C}) \times p(f | z, \langle z \rangle, \mathbf{C}) \tag{9}$$

where $p(z | \langle z \rangle, \mathbf{C})$ is the prior PDF, f is a set of constraints imposed on the posterior expectation values, and where $p(f | z, \langle z \rangle, \mathbf{C})$ is the likelihood function. Constraints in f are defined by an auxiliary quantity that relates components of generalized data as

$$\omega \equiv \omega(T(\cdot), z) \equiv T(P) - D - \delta, \tag{10}$$

as constraints on their posterior expectation values and their posterior covariance matrix elements,

$$\begin{aligned} \langle \omega \rangle' &= \omega'_f \\ \mathbf{\Omega}' \equiv \langle (\omega - \langle \omega \rangle')(\omega - \langle \omega \rangle')^{\mathsf{T}} \rangle' &= \mathbf{\Omega}'_f, \end{aligned} \tag{11}$$

where ω'_f and $\mathbf{\Omega}'_f$ are given, and where posterior expectation values are indicated by primes.

Constraints on posterior expectation values and covariances yield a likelihood function formally expressed via Lagrange multipliers, so that the posterior PDF becomes

$$p(f | z, \langle z \rangle, \mathbf{C}) = e^{-\sum_i \lambda_i \omega_i - \sum_{ij} \Lambda_{ij} (\omega - \langle \omega \rangle')_i (\omega - \langle \omega \rangle')_j} \tag{12}$$

where $\{\lambda_i\}_f$ and $\{\Lambda_{ij}\}_f$ constitute a set of Lagrange multipliers to be determined from the constraint set f. This posterior PDF of generalized data implicitly contains a combined posterior PDF of parameters, measured data, and model defect data, that has been informed by all prior information available, namely, by $\langle z \rangle$, $\mathbf{C}$, and the constraint f enforced on the posterior expectation values and covariances.

Upon normalizing the posterior PDF of generalized data to unity, posterior expectation values of any function $g(z)$ of posterior generalized data z could be computed as an integral over generalized data

$$\langle g(z) \rangle' = \int dz g(z) p(z | \langle z \rangle, \mathbf{C}, f) \tag{13}$$

that are also used to compute posterior expectation values, $\langle z \rangle'$, and their posterior covariance matrix $\mathbf{C}'$. This posterior PDF yields $\langle \omega \rangle' = \omega'_f$ and $\mathbf{\Omega}' = \mathbf{\Omega}'_f$.

One consequence of the derivation of the posterior PDF of generalized data is that any computation of expectation values computed with this PDF entails integration over all generalized data. Therefore, generalized data (model parameters, measured data, and model defect data) could be viewed as nuisance parameters marginalized by integration.

Expectation values of the constraint parameters are generally set to

$$\omega'_f = 0 \tag{14}$$

$$\text{and} \quad \mathbf{\Omega}'_f = 0, \tag{15}$$

where this choice defines a particular set of constraints labeled f_0. Since the diagonal elements of $\mathbf{\Omega}'$ are the posterior expectation values of ω_i^2, and since PDF's are positive functions, this constraint could be satisfied by enforcing $\omega = 0$ for all values of z inside the integral. This suggests an effective likelihood function

$$p(f_0 | z, \langle z \rangle, \mathbf{C}) = \delta_{\text{Dirac}}(\omega). \tag{16}$$

With this likelihood function, the expectation values computed by the posterior PDF become

$$\langle g(z) \rangle'_{f_0} = \int dz g(z) p(z | \langle z \rangle, \mathbf{C}) \delta_{\text{Dirac}}(\omega), \tag{17}$$

where $p(z | \langle z \rangle, \mathbf{C})$ is the prior PDF of generalized data.

A $\delta_{\text{Dirac}}(\omega)$ likelihood function of a defective model effectively reduces integration over $z = (P, D, \delta)$ to (P, D), and the model defect variable δ is replaced by $T(P) - D$ in the prior PDF. This component of the prior PDF has similar features as the likelihood function obtained by setting constraints $\omega'_f \leftarrow \langle \delta \rangle'$ and $\mathbf{\Omega}'_f \leftarrow \mathbf{\Delta}'$ for a perfect model. Conversely, non-zero values of ω'_f and $\mathbf{\Omega}'_f$ for a perfect model with $\langle \delta \rangle = \langle \delta \rangle' = 0$ and $\mathbf{\Delta} = \mathbf{\Delta}' = \mathbf{0}$ yield a PDF analogous to that obtained by setting constraints to zero and introducing a model defect $\langle \delta \rangle' \leftarrow \omega'_f$ and $\mathbf{\Delta}' \leftarrow \mathbf{\Omega}'_f$. This could be phrased as

$$(\omega'_f, \mathbf{\Omega}'_f, (\langle \delta \rangle, \mathbf{\Delta}) = 0) \leftrightarrow ((\omega'_f, \mathbf{\Omega}'_f) = 0, \langle \delta \rangle', \mathbf{\Delta}') \tag{18}$$

or in shorthand

$$(\omega'_f, \mathbf{\Omega}'_f) \leftrightarrow (\delta', \mathbf{\Delta}'). \tag{19}$$

This point will be elaborated upon in Sections 2.2 and 3. The exact connection between the two approaches may be intricate because constraints are defined on posterior expectation values and covariances, while the model defect is defined by prior model defect data expectation values and covariances.

The posterior generalized data PDF enables computation of covariance matrix $\mathbf{\Theta}$ of posterior model values $\langle T(P) \rangle'$, corresponding to experimental data D, to all orders

$$\mathbf{\Theta} \equiv \langle (T(P) - \langle T(P) \rangle')(T(P) - \langle T(P) \rangle')^{\mathsf{T}} \rangle', \tag{20}$$

in contrast to the first-order approximation expression,

$$\left.\frac{\partial T(P)}{\partial P}\right|_{P = \langle P \rangle'} \langle (P - \langle \mathrm{P} \rangle')(P - \langle P \rangle')^{\mathsf{T}} \rangle' \left.\frac{\partial T(P)}{\partial P}\right|^{\mathsf{T}}_{P = \langle P \rangle'}, \tag{21}$$

reported in evaluated nuclear data files like the ENDF [11].

2.2 Posterior PDF for normal PDFs

Although this formalism applies to arbitrary PDFs, a particularly simple form is attained when normal form is assumed for all PDFs. In that case, the prior PDF becomes

$$p(z|\langle z\rangle, \mathbf{C}) \propto e^{-\frac{1}{2}(z-\langle z\rangle)^{\mathsf{T}} C^{-1}(z-\langle z\rangle)}, \tag{22}$$

$$\equiv \mathcal{N}(z|\langle z \rangle, \mathbf{C}) \tag{23}$$

where N stands for a normal PDF, and the likelihood function could be stated as

$$p(f|z, \langle z \rangle, \mathbf{C}) \propto e^{-\frac{1}{2}(\omega-\omega_f)^{\mathsf{T}} \mathbf{\Omega}_f^{-1}(\omega-\omega_f)}, \tag{24}$$

$$\equiv \mathcal{N}(\omega|\omega_f, \mathbf{\Omega}_f) \tag{25}$$

where ω is defined in equation (10), ω_f is an effective parameter vector and $\mathbf{\Omega}_f$ is an effective covariance matrix such that this posterior PDF obeys the constraint f on posterior expectation value $\langle \omega \rangle' = \omega'_f$ and $\mathbf{\Omega} = \mathbf{\Omega}_f$. Unknown parameters ω_f and $\mathbf{\Omega}_f$ play a role equivalent to Lagrange multipliers $\{\lambda_i\}_f$ and $\{\Lambda_{ij}\}_f$ in equation (12).

Combining the normal prior PDF and the normal likelihood functions yields a posterior PDF

$$p(z|\langle z \rangle, \mathbf{C}, f) \propto \mathcal{N}(z|\langle z \rangle, \mathbf{C}) \times \mathcal{N}(\omega|\omega_f, \mathbf{\Omega}_f), \tag{26}$$

subject to aforementioned constraints in f.

Constraint set, f_0, namely $\langle \omega \rangle' = 0$ and $\mathbf{\Omega}' \equiv \langle \omega\omega^{\mathsf{T}} \rangle' = 0$, are satisfied for $\omega_f = 0$ and $\mathbf{\Omega}_F = \mathbf{0}$, for which the normal likelihood function in equation (24) becomes a Dirac delta function, so that the posterior PDF

becomes

$$p(z|\langle z \rangle, \mathbf{C}, F) \propto \mathcal{N}(z|\langle z \rangle, \mathbf{C}) \times \delta_{\mathrm{Dirac}}(\omega). \tag{27}$$

Furthermore, for models without a defect, that is, $\langle \delta \rangle = 0$ and $\mathbf{\Delta} = \mathbf{0}$, this posterior PDF becomes

$$p(\hat{z}|\langle \hat{z} \rangle, \mathbf{C}, F) \propto \mathcal{N}(\hat{z}|\langle \hat{z} \rangle, \hat{\mathbf{C}}) \times \delta_{\mathrm{Dirac}}(T(P) - D) \tag{28}$$

where

$$\hat{z} \equiv (P, D) \tag{29}$$

and $\hat{\mathbf{C}}$ is the covariance matrix corresponding to $\hat{z}$. In Section 4, it will be shown that the expression for the posterior PDF in this limit is equivalent to the CGLS method implemented in the APLCON code, or to its linear approximation implemented in the TSURFER module of the SCALE code system.

2.3 Conventional derivation of posterior parameter PDF

Making the following substitutions into a generic Bayes' theorem in equation (1):

$$\begin{array}{rcl} \alpha & \to & P \\ \beta & \to & f \\ \gamma & \to & (\langle P \rangle, \langle D \rangle, \langle \delta \rangle, \mathbf{C}) = (\langle z \rangle, \mathbf{C}), \end{array} \tag{30}$$

one obtains

$$p(P|\langle z \rangle, \mathbf{C}, f) \propto p(P|\langle z \rangle, \mathbf{C}) \times p(f|P, \langle z \rangle, \mathbf{C}), \tag{31}$$

where the second factor on the right hand side can be expressed as a nested integral over all possible values of measured data D and model defect data δ, given their expectation values $\langle D \rangle$ and $\langle \delta \rangle$, respectively, and their covariance matrix $\mathbf{C}$, by using the total probability theorem in equation (3):

$$p(f|P, \langle z \rangle, \mathbf{C}) = \int dD \int (d\delta) p(f|D, \delta, P, \langle z \rangle, \mathbf{C}) \\ \times p(D, \delta|P, \langle z \rangle, \mathbf{C}). \tag{32}$$

The first term in equation (31) and the second term in equation (32) could be combined by making the following substitutions:

$$\begin{array}{rcl} \alpha & \to & P \\ \beta & \to & D, \delta \\ \gamma & \to & \langle z \rangle, \mathbf{C} \end{array} \tag{33}$$

into the product rule in equation (2) to obtain

$$p(z|\langle z \rangle, \mathbf{C}) = p(P, D, \delta|\langle z \rangle, \mathbf{C}) \\ \propto p(P|\langle z \rangle, \mathbf{C}) \times p(D, \delta|P, \langle z \rangle, \mathbf{C}). \tag{34}$$

Combining all terms yields

$$p(P|\langle z \rangle, \mathbf{C}, f) \propto \int dD \int (d\delta) p(z|\langle z \rangle, \mathbf{C}) p(f|z, \langle z \rangle, \mathbf{C}) \\ \propto \int dD \int (d\delta) p(z|\langle z \rangle, \mathbf{C}, f), \tag{35}$$

where Bayes' theorem stated by equation (9) was used to introduce $p(z|\langle z \rangle, \mathbf{C}, f)$ in the integrand on the last line above. This shows that a partial posterior PDF of parameters, P, is simply an integral of the posterior PDF over all measured data, D, and model defect δ.

3 Simple example

The correspondence between a constraint set f and a model defect suggested above is illustrated on a simple analytically solvable example. A simple model without a defect is defined as $T(P) = P$ with a single scalar parameter P, and a single data point D with constraints f: $\langle \omega \rangle' = 0$ and $\mathbf{\Omega}' \equiv \langle \omega^2 \rangle' = 2/3$. A corresponding model with a defect $\langle \delta \rangle = 0$ and $\mathbf{\Delta} = 1$ has constraint set to zero, namely f_0: $\langle \omega \rangle' = 0$ and $\mathbf{\Omega}' = 0$, while its covariance matrix has a component corresponding to model defect, namely $C_{3,3} = \mathbf{\Delta} = 1$. For simplicity, the prior expectation values of generalized data are set to zero, $\langle z \rangle = 0$, and the prior covariance matrix $\mathbf{C}$ for both models is set to the identity matrix, so that posterior expectation values remain unchanged for both models, that is, $\langle z \rangle = \langle z \rangle' = 0$.

First we consider a model without a defect, that is, $\hat{z} = (P, D)$ whose covariance matrix is defined as

$$\hat{\mathbf{C}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{36}$$

and whose posterior PDF is expressed as a product of its prior PDF and the likelihood function expressed in terms of Lagrange multipliers,

$$p(\hat{z}|\langle \hat{z} \rangle, \hat{\mathbf{C}}, f) = \mathcal{N}(\hat{z}|\langle \hat{z} \rangle, \hat{\mathbf{C}}) e^{-\lambda(x-y) - \Lambda(x-y)^2}, \tag{37}$$

where $\lambda = 0$ and $\Lambda = 1/2$ yield a posterior PDF that satisfies the imposed constraints, that is, $\langle \omega \rangle' = 0$ and $\mathbf{\Omega}' = 2/3$.

For a model with a defect, $z = (P, D, \delta)$, with a prior covariance matrix

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{38}$$

yields a posterior PDF that satisfies f_0: $\langle \omega \rangle' = 0$ and $\mathbf{\Omega}' = 0$, as

$$p(z|\langle z \rangle, \mathbf{C}, f) = N(z|\langle z \rangle, \mathbf{C})\,\delta_{\text{Dirac}}(\omega). \tag{39}$$

Upon integration over δ, the remaining PDF becomes equivalent to the posterior PDF in equation (37). In this simple illustration, numerical constants were chosen to make the correspondence between the two approaches exact.

4 Connections to other methods

To establish connections to other methods all PDFs are assumed to be normal, and consequently, the exponents in equations (22) and (24) could be combined to define a generalized cost function:

$$Q(z) \equiv (\tilde{z} - \langle z \rangle)^{\mathsf{T}} \mathbf{C}^{-1} (\tilde{z} - \langle z \rangle) \tag{40}$$

where the constraint $T(P) - D = \delta$ is enforced by defining $\tilde{z} \equiv (P, D, T(P) - D)$. This cost function can be minimized by using Laplace transform and Newton-Raphson method to yield approximate posterior expectation values of generalized data $\langle z \rangle' \approx z_{\min}$ and of its covariance matrix $\mathbf{C}' \approx \mathbf{C}_{\min}$ [12]. For a perfect model, one may set $\langle \delta \rangle \to 0$ and $\mathbf{\Delta} \to \mathbf{0}$ to obtain

$$Q(z) \approx (\hat{z} - \langle \hat{z} \rangle)^{\mathsf{T}} \hat{\mathbf{C}}^{-1} (\hat{z} - \langle \hat{z} \rangle), \tag{41}$$

where $\hat{z} \equiv (P, D)$ and $\hat{\mathbf{C}}$ is the corresponding covariance matrix, with the constraint $T(P) = D$. A constrained minimization of this cost function performed by the TSURFER code [9] and the APLCON code [13], where the constraint is enforced by the Lagrange multiplier method. The values of $z_{\min}$ that minimize χ^2 are then approximate expectation values of posterior generalized data $\langle z \rangle' \approx z_{\min}$. Since TSURFER makes a linear approximation of the model, its method is referred to as generalized linear least squares (GLLS). In conventional GLS, which is also known as the χ^2 minimization method, the constraint is applied to the generalized data $\hat{z} \equiv (P, D) \to (P, T(P))$, and the difference $(\hat{z} - \langle \hat{z} \rangle)$ in equation (41) is replaced by

$$\begin{aligned} (\hat{z} - \langle \hat{z} \rangle) &\equiv (P - \langle P \rangle, D - \langle D \rangle) \\ &\to (P - \langle P \rangle, T(P) - \langle D \rangle), \end{aligned} \tag{42}$$

with no constraint enforced.

A common approximation to the χ^2-function is obtained for a block-diagonal generalized data covariance matrix $\mathbf{C}$, with parameter covariance matrix $\mathbf{M}$ and experimental data covariance matrix $\mathbf{V}$ along the diagonal:

$$\begin{aligned} \chi^2 &= (P - \langle P \rangle)^{\mathsf{T}} \mathbf{M}^{-1}(P - \langle P \rangle) \\ &+ (T(P) - \langle D \rangle)^{\mathsf{T}} \mathbf{V}^{-1}(T(P) - \langle D \rangle). \end{aligned} \tag{43}$$

Minimization of χ^2 with respect to P yields a solution vector $P_{\min}$, that is approximately equal to the posterior expectation value of parameters $\langle P \rangle'$ [14]. This definition of χ^2 has been used in nuclear data evaluations and is also the quantity that is minimized in generic optimization codes like MINUIT [15]. That these methods could lead to incorrect posterior values and covariances has been recognized by Capote [16].

5 Conclusions and outlook

A new, general expression for the posterior PDF of generalized data has been derived, where generalized data refers to a union of model parameters, measured data, and model defect data, starting from the Bayes' theorem. An analogy between the constraints on posterior expectation values and model defect data has been suggested in this work, and further study may be needed to better understand their connections and potential applications. Key ingredients used in this derivation are:
– use of generalized data that is a union of parameters, measured data, and model defect;

– formal recognition constraints on posterior expectation values and covariances when applying Bayes' theorem;
– formal and consistent separation between expectation and instance values of generalized data.

A direct consequence of application of Bayes' theorem to generalized data, of which experimental data are a subset, is that the posterior PDF of generalized data yields posterior expectation values and covariances for experimental data and model defect data. This is in contrast to the prevalent nuclear data evaluation practice where posterior PDF of experimental data and its covariances are identical to those of model predictions.

A normal form of posterior PDF of generalized data, obtained by setting all constituent PDFs to be normal and assuming a perfect model, was found useful to establish a connection to extant optimization methods, namely CGLS implemented in APLCON and its linear approximation implemented in the TSURFER module of SCALE.

The appealing features of the posterior PDF of generalized data listed above make it a candidate for simultaneous and consistent optimization of model parameters and experimental data of differential cross section and integral benchmarks that would yield more accurate and complete evaluations and covariances. The presented method could simultaneously sample R-matrix resonance parameters, optical model potential parameters, and integral benchmark parameters such as spatial dimensions and material composition, to compute presently unknown covariances among integral benchmark experiments and cross section data. Although sensitivities of integral benchmark responses with respect to cross sections are not needed for calculations using the Bayesian Monte Carlo method, they could nevertheless be computed [17]. It is hoped that the derived method could complement conventional nuclear data adjustment methods [18].

Author contribution statement

All the authors were involved in the preparation of the manuscript. All the authors have read and approved the final manuscript.

References

1. T. Bayes, Phil. Trans. Roy. Soc. **53**, 370 (1763) [reprinted in E.S. Pearson and M.G. Kendall, *Studies in the History of Statistics and Probability*, (Hafner, Darien, Conn., 1970)]
2. E.T. Jaynes, Straight Line Fitting − a Bayesian Solution, http://bayes.wustl.edu/etj/articles/leapz.pdf (1991)
3. G. Schnabel, Ph.D. Thesis, Technischen Universität Wien, 2015
4. M.T. Pigni, H. Leeb, in *Proceedings of the International Workshop on Nuclear Data for the Transmutation of Nuclear Waste, GSI-Darmstadt*, Germany, 2003
5. H. Leeb, D. Neudecker, T. Srdinko, Consistent procedure for nuclear data evaluation based on modeling, Nucl. Data Sheets **109**, 2762 (2008)
6. D. Neudecker, R. Capote, H. Leeb, Impact of model defect and experimental uncertainties on evaluated output, Nucl. Instrum Meth. Phys. Res. A **723**, 163 (2013)
7. G. Schnabel, H. Leeb, Differential cross sections and the impact of model defects in nuclear data evaluation, EPJ Web Conf. **111**, 9001 (2016)
8. V. Blobel, Constrained Least Squares Methods with Correlated Data and Systematic Uncertainties (2010), http://www.desy.de/blobel/apltalk.pdf
9. M.L. Williams, B.L. Broadhead, M.A. Jessee, J.J. Wagschal, TSURFER: An Adjustment Code To Determine Biases and Uncertainties in Nuclear System Responses by Consolidating Differential Data and Benchmark Integral Experiments, Version 6.2.1, Vol. III, Sect. M21, ORNL/TM-2005/39 (2016)
10. B.T. Rearden, M.A. Jessee, Eds., *SCALE Code System, ORNL/ TM-2005/39, Version 6.2.1* (Oak Ridge National Laboratory, Oak Ridge, Tennessee, 2016) Available from Radiation Safety Information Computational Center as CCC-834
11. National Nuclear Data Center, Brookhaven National Laboratory, http://nndc.bnl.gov
12. F. Fröhner, Evaluation and Analysis of Nuclear Resonance Data, JEFF Report 18, 2000
13. V. Blobel (DESY), APLCON downloadable from http://www.desy.de/blobel/wwwcondl.html
14. N.M. Larson, *Updated Users' Guide for SAMMY: Multilevel R-matrix Fits to Neutron Data Using Bayes' Equations*, ORNL/TM-9179/R8 (2008)
15. F. James, M. Roos, Comput. Phys. Commun. **10**, 343 (1975)
16. R. Capote, D.L. Smith, An investigation of the performance of the unified Monte Carlo method of neutron cross section data evaluation, Nucl. Data Sheets **109**, 2768 (2008)
17. L. Fiorito et al., Nuclear data uncertainty propagation to integral responses using SANDY, Ann. Nucl. Energy **101**, 359 (2017)
18. V. Sobes, L. Leal, G. Arbanas, B. Forget, Resonance parameter adjustment based on integral experiments, Nucl. Sci. Eng. **183**, 347 (2016)