ORNL/TM-2022/2448

Bayesian Optimization Framework for *Imperfect* Data or Models

Jesse M. Brown
Goran Arbanas
Dorothea Wiarda
Andrew Holcomb

April 2022

Approved for public release.
Distribution is unlimited.

OAK RIDGE National Laboratory

ORNL IS MANAGED BY UT-BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY

ORNL/TM-2022/2448

Nuclear Energy and Fuel Cycle Division

BAYESIAN OPTIMIZATION FRAMEWORK FOR *IMPERFECT* DATA OR MODELS

Jesse M. Brown, Goran Arbanas, Dorothea Wiarda, Andrew Holcomb

April 2022

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831-6283
managed by
UT-Battelle LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

LIST OF FIGURES

ABBREVIATIONS

ENDF	Evaluated Nuclear Data File
GLLS	Generalized Linear Least Squares
MC	Monte Carlo
MHMC	Metropolis Hastings Monte Carlo
PDF	Probability Distribution Function
PPP	Peelle's Pertinent Puzzle

ABSTRACT

Conventional Bayesian optimization methods implicitly assume that the data and model being optimized are "perfect." This assumption leads to inaccurate posterior probability distribution functions (PDFs) when applied to "imperfect" data or models. The new Bayesian optimization framework presented in this report provides a way to parameterize the effect of imperfections usually encountered in a prior PDF of generalized data* or a model on the posterior PDF. The effects of imperfections are parameterized by a set of constraints imposed on the posterior expectation values of deviations between the data and the model and on their covariance matrix elements. A particular set of values for these constraints conveys an evaluator's best estimate of the effect of imperfections on the corresponding posterior expectation values. When a prior PDF of generalized data is assumed to be normal, an expression for a posterior PDF satisfying an arbitrary set of constraints is derived analytically for linear models. An analogous iterative algorithm is given for nonlinear models. The corresponding posterior PDF should be used to estimate any posterior expectation values in the presence of imperfections parameterized by that set of constraints. A posterior PDF of a conventional Bayesian optimization method is recovered analytically when all evaluator-specified constraints are set to zero (i.e., in the absence of any imperfections). The analytical expressions derived in this report for normal PDFs and linear models were verified numerically by a Metropolis–Hastings Monte Carlo method. The methods presented herein could be applied to any kind of data or models, including differential cross-section data or integral benchmark experiments.

1. INTRODUCTION

Although Bayes' theorem has been used for model optimization and the evaluation of scientific data since its publication in 1763, the common assumptions and approximations these assumptions entail are infrequently stated, leading to several distinct but synonymous formulations of Bayes' theorem in the scientific literature. To differentiate among the various forms of the Bayes' theorem, it useful to state the assumptions that lead to its most common and simplest formulation.

1. The prior† probability distribution function (PDF) of generalized data is specified perfectly.‡
2. The model (used for data evaluation) is perfect.
3. The posterior and prior PDFs are of the normal form.
4. The model is linear.

When all four listed assumptions are satisfied, then Bayes' theorem is equivalent to the minimization of the generalized linear least squares (GLLS) cost function. For nonlinear models, the minimization is conventionally performed using the Newton–Raphson method and a linear approximation of the model being optimized. These methods have been in continuous use in nuclear data evaluations for several decades; see, for example, Fröhner's report on nuclear data evaluation methods [1], *R*-matrix resonance parameter evaluation method described in the *SAMMY Updated Users' Guide* [2], or the integral

*_Generalized data_ are the union of model parameters and measured data.

†In the terminology of Bayes' theorem, a prior PDF is updated to yield a corresponding posterior PDF. The posterior PDF is conventionally referred to as *data evaluation* in the nuclear data community.

‡For a normal PDF, this implies that the mean values and the covariance matrix of generalized data are known perfectly. Conversely, the imperfections may be introduced via unspecified covariance matrix elements that are set to zero in their absence or by inconsistent datasets.

benchmark experiment evaluation code TSURFER in the SCALE code suite [3]. Nuclear data evaluations, including their uncertainties, are collected and disseminated via the US Evaluated Nuclear Data File (ENDF) [4] or one of the standard nuclear databases for neutron transport simulation codes, such as SCALE [3] and Monte Carlo (MC) N-Particle (MCNP) [5].

One undesirable but unfortunately common consequence of the first two assumptions is that they lead to underestimated posterior uncertainties. These uncertainties are most often caused by missing data covariance, which is set to zero in its absence. Because the underestimation is proportional to the size of the measured datasets, the posterior uncertainties become unreasonably small with increasingly large datasets. Nuclear data evaluators address this issue by inflating the posterior uncertainties by a number significantly larger than unity, which forces the uncertainties to fall into some intuitively acceptable range that is in harmony with the uncertainties of the measured data. However, this manual adjustment forsakes the formal consistency with Bayes' theorem that is a foundation of inferential probability theory. Identifying the first two assumptions is essential for resolving this and other problems in nuclear data evaluations. Recognizing the fact that the first two assumptions are rarely satisfied for nuclear data motivated the treatment of imperfections presented in this work.

The first aforementioned assumption is the most deeply ingrained and has been the most difficult to discern [6; 7]. This work seems to be the first to remove this assumption, both analytically and computationally, in complete harmony with Bayes' theorem. In addition to the absence of covariance information among distinct data, other data imperfections that may lead to the violation of the first assumption in practical applications include inconsistent datasets, imperfections in prior model parameters, and missing covariance information between prior parameter values and measured data. These imperfections are various aspects of imperfections stemming from the prior PDF of generalized[§] data.

In principle, the second assumption could be made to hold when an analytical model is used to create synthetic data for numerical studies. In natural sciences, because any model of the natural phenomena is expected to be at least slightly imperfect, the second assumption is not expected to hold exactly. The effects of data and model imperfection are intertwined in the absence of any information about their respective source of imperfections, so one welcome feature of the formalism presented in this work is that it simultaneously addresses both sources of imperfection. By contrast, previous attempts have addressed the second assumption—without discerning the first one—by introducing a model of the "model defect" (see Refs. [8; 9; 10; 11; 12]), which is added to the original model before proceeding with the data evaluation. However, an explicit treatment of a model defect is not necessary to present the formalism in this report.

The effects of the remaining two assumptions have been generally recognized and quantified by using the Metropolis–Hastings Monte Carlo (MHMC) method [13]. The MHMC method was used herein to quantify the effects of the first two assumptions for the first time.

Although the initial motivation for this work was to find a completely Bayesian solution to the problem of unrealistically small posterior uncertainties, another welcome outcome was a transparent derivation of Bayes' theorem with all four of the listed assumptions removed simultaneously from which the conventional Bayes' theorem can be seen to be a special case obtained in the limit of vanishing imperfections.

[§]Generalized data, a formal union of model parameter and measured data, simplify the derivations of the general form of Bayes' theorem presented in this report and make it conceptually transparent.

2. MATHEMATICAL FORMALISM

Section 2.1 outlines a general derivation of Bayes' theorem in the presence of imperfections by introducing evaluator-defined constraints on posterior expectation values of deviations between the model and the data and demonstrates that the conventional form of Bayes' theorem is seamlessly recovered in the limit of vanishing imperfections. Section 2.2 derives an analytically exact posterior PDF for linear models and normal PDFs and for the arbitrary values of the said evaluator-defined constraints on posterior expectation values. Section 2.3 describes an iterative algorithm that—by halving the covariance matrix of deviations between the model and data in each iteration—makes the posterior PDF of *generalized* data (parameters and data) converge to the posterior PDF of the parameters (only) in the form of the conventional Bayes' theorem, as expected in the limit of vanishing imperfections. This algorithm is useful for computing the conventional posterior PDF by a Bayesian MHMC method because a direct MC sampling of that PDF becomes unfeasible for sufficiently large data sets. Section 2.4 describes an iterative algorithm for minimizing the cost function of generalized data for normal PDFs and nonlinear models in the presence of imperfections; this algorithm is a generalization of the Newton–Raphson minimization algorithm for the conventional cost function. Section 2.5 describes an algorithm for finding a posterior PDF for non-normal PDFs and nonlinear models, and for arbitrary evaluator-defined constraints. Section 2.6 shows how the new formalism designed for imperfections could be used to find an infinite number of ways to bypass the Peelle's Pertinent Puzzle (PPP) (see Refs. [14; 15]) that has confounded the conventional applications of Bayes' theorem.

2.1 BAYES' THEOREM FOR IMPERFECT GENERALIZED DATA OR MODELS

The efficiency and clarity of this derivation is partly due to the concept of a generalized data vector, denoted by z, that is a concatenation of parameter vector, P, and data vector, D:

$$z \equiv \begin{pmatrix} P \\ D \end{pmatrix}. \tag{1}$$

A general form of Bayes' theorem expresses the posterior PDF, $p'(z|\beta\gamma)$, of generalized data, z, as a product of the prior PDF, $p(z|\gamma)$, and the likelihood function, $\mathcal{L}(\beta|z\gamma)$,

$$p'(z|\beta\gamma) = p'(z|\gamma\beta) = N'\mathcal{L}(\beta|z\gamma) \times p(z|\gamma), \tag{2}$$

where

$$\begin{aligned} \beta &\leftarrow \{\text{any constraints on } \textit{posteriors} \text{ imposed by an } \textit{evaluator}\}, \\ \gamma &\leftarrow \{\text{any parameters needed to define the prior PDF, } p(z|\gamma)\}. \end{aligned} \tag{3}$$

and where N' is a normalization constant [11; 6; 7]. From this vantage point, the conventional form of Bayes' theorem will be shown to correspond to the constraints that enforce the assumption that data and model are perfect. When applied to imperfect data or models, this assumption will lead to an inaccurate posterior PDF. In practice, it often leads to posterior uncertainties that are unreasonably small, forcing the evaluators to resort to a manual inflation of those uncertainties in a formal departure from Bayes' theorem. The data evaluation framework described herein is an attempt to remedy this situation by using the constraints imposed on posterior expectation values as a tool to modulate the effect of imperfections in harmony with Bayes' theorem.

Prior and posterior expectation values of any function of z are computed as integrals over z, weighted by a corresponding normalized PDF:

$$\langle f(z)\rangle = \int f(z)p(z|\gamma)dz \qquad \text{and} \tag{4}$$

$$\langle f(z)\rangle' = \int f(z)p'(z|\beta\gamma)dz, \tag{5}$$

respectively, where primes on expectation values indicate that posterior PDF, $p'(z|\beta\gamma)$, also denoted by a prime, has been used. For example, prior expectation values of generalized data are obtained for $f(z) = z$:

$$\langle z\rangle \equiv \begin{pmatrix}\langle P\rangle \\ \langle D\rangle\end{pmatrix}, \tag{6}$$

and the prior covariance matrix, $\mathbf{C}$ of generalized data is obtained for $f(z) = (z - \langle z\rangle)(z - \langle z\rangle)^\intercal$; that is,

$$\mathbf{C} \equiv \langle (z - \langle z\rangle)(z - \langle z\rangle)^\intercal\rangle, \tag{7}$$

which is a 2×2 block matrix,

$$\mathbf{C} = \begin{pmatrix}\mathbf{M} & \mathbf{W} \\ \mathbf{W}^\intercal & \mathbf{V}\end{pmatrix}. \tag{8}$$

Here, the matrices $\mathbf{M}$ and $\mathbf{V}$ correspond to the covariance for P and D, respectively, along with their cross covariance denoted by $\mathbf{W}$. When the only prior information given are the expectation values and the covariance data, which is usually the case for nuclear data evaluations, the prior PDF is then assumed to be a normal PDF; that is, $\gamma \leftarrow (\langle z\rangle, \mathbf{C})$, and

$$p(z|\gamma) \leftarrow p(z|\langle z\rangle, \mathbf{C}) = \mathcal{N}(z|\langle z\rangle, \mathbf{C}) = \frac{1}{\sqrt{(2\pi)^k|C|}} e^{-\frac{1}{2}(z-\langle z\rangle)^\intercal \mathbf{C}^{-1}(z-\langle z\rangle)}, \tag{9}$$

where k is the dimension of z, and $|\cdot|$ denotes the matrix determinant. However, the derivation in this section is general and applies to any arbitrary non-normal prior PDF, $p(z|\gamma)$.

A set of evaluator-defined constraints, β, are imposed on the posterior expectation values of some evaluator-defined function, $\delta(z, T(\cdot))$, where $T(\cdot)$ represents a model used for data evaluation, usually defined as $T(\cdot) \leftarrow T(P)$. In this work, evaluator-defined constraints on posteriors are limited to $\langle\delta(z, T(P))\rangle'$ and its posterior covariance matrix, $\mathbf{\Delta}'$. A particular form of a function $\delta(z, T(P))$ is chosen by the evaluator to reflect some property of the evaluated data. A generic form used in this work[¶] for illustration purposes is as follows:

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

whose prior and posterior expectation values—$\langle\delta\rangle$ and $\langle\delta\rangle'$, respectively—are computed by setting $f(z) = \delta$ and whose prior and posterior covariances—$\mathbf{\Delta}$ and $\mathbf{\Delta}'$, respectively—are computed accordingly. However, the constraints are imposed on the posterior expectation values and covariance matrix only (i.e., on the posterior PDF) so that $\langle\delta\rangle'$ and $\mathbf{\Delta}'$ computed using such a posterior PDF must be equal to the values imposed by an evaluator. The prior expectation values of δ, $\langle\delta\rangle$, and its prior covariance matrix, $\mathbf{\Delta}$, are uniquely determined by the prior PDF, $p(z|\gamma)$, and are neither constrained nor affected by the user-defined

[¶]When applied to line fitting, $\delta(T(P), D)$ could be alternatively defined as the shortest distance between each point in the dataset, D, and the line defined by $T(P)$.

constraints on posterior PDF. The posterior PDF is constrained by evaluator-defined posterior expectation values, $\langle\delta\rangle'$, and by constraining its posterior covariance matrix $\mathbf{\Delta}'$; that is,

$$\mathbf{\Delta}' \equiv \langle(\delta - \langle\delta\rangle')(\delta - \langle\delta\rangle')^\intercal\rangle', \tag{11}$$

where posterior expectation values are indicated by primes.

To constrain the posterior expectation values to the evaluator-provided values, the likelihood function takes the form of

$$\mathcal{L}(\beta|z,\gamma) \leftarrow \mathcal{L}(\langle\delta\rangle', \mathbf{\Delta}'|z, \langle z\rangle, \mathbf{C}) = e^{-\frac{1}{2}(\delta-\lambda)^\intercal \mathbf{\Lambda}^{-1}(\delta-\lambda)}, \tag{12}$$

where the vector λ and the matrix $\mathbf{\Lambda}$ consist of values determined by the constraints on the posterior expectation values, $\langle\delta\rangle'$, and the corresponding posterior covariance matrix, $\mathbf{\Delta}$' [11].

Combining the prior PDF and the likelihood function into Eq. (2) yields a posterior PDF of the form

$$p'(z|\langle\delta\rangle', \mathbf{\Delta}', \langle z\rangle, \mathbf{C}) \propto \mathcal{L}(\langle\delta\rangle', \mathbf{\Delta}'|z, \langle z\rangle, \mathbf{C}) \times \mathcal{N}(z|\langle z\rangle, \mathbf{C}), \tag{13}$$

which should be used to compute any posterior expectation values. In the most general case, the posterior PDF can be computed by an MHMC method. For normal PDFs, a generalized cost function (to be minimized) is inferred by collecting the terms in the exponents of the likelihood function and the prior PDF. An expression for generalized cost function is stated and minimized analytically for linear models in Section 2.1 and for nonlinear models in Section 2.5. The remainder of this section is used to make an elegant connection to the conventional form of Bayes' theorem.

All conventional evaluation methods implicitly set $\langle\delta\rangle' = 0$ and $\mathbf{\Delta}' = \mathbb{0}$, for which $\lambda = 0$ and $\mathbf{\Lambda} = \mathbb{0}$, so that $\mathcal{L}(\langle\delta\rangle', \mathbf{\Delta}'|z, \langle z\rangle, \mathbf{C}) \propto \delta_{\text{Dirac}}(T(P) - D)$, and the posterior PDF becomes[‖]

$$p'(P|\langle\delta\rangle' = 0, \mathbf{\Delta}' = \mathbb{0}, \langle z\rangle, \mathbf{C}) \propto \mathcal{N}((z|_{D=T(P)})|\langle z\rangle, \mathbf{C}) \propto e^{-\chi^2(P)/2}, \tag{14}$$

demonstrating that a conventional posterior PDF is recovered in the limit of vanishing imperfections and is naturally expressed in a generalized data notation. For linear and nonlinear models, this posterior PDF can be computed using the MHMC method. However, for linear models, this posterior PDF can be determined analytically by minimizing the conventional expression for $\chi^2(P)$ gleaned from the exponent of the normal PDF in Eq. (14):

$$\chi^2(P) = (z|_{D=T(P)} - \langle z\rangle)^\intercal \mathbf{C}^{-1}(z|_{D=T(P)} - \langle z\rangle), \tag{15}$$

as is done analytically in Section 2.2. For nonlinear models, the posterior PDF is approximated by a normal PDF whose posterior expectation values, $\langle P\rangle'$, are approximated by the values $\hat{P}$ that minimize $\chi^2(P)$ and whose posterior covariance matrix, $\hat{\mathbf{M}}'$, is approximated by the Hessian of $\chi^2(P)$ evaluated at $\hat{P}$. This approximation is used for conventional data evaluations reported as $T(\hat{P})$, and the corresponding evaluated covariance matrix is computed to a first order approximation from the posterior covariance matrix of model parameters and the first-order derivatives of the model with respect to parameters. This procedure is described in Refs. [1; 2].

The Bayesian evaluation framework for imperfect data enables evaluators to parameterize imperfections for each data point in each dataset, which is particularly useful for dealing with inconsistent datasets. A simple

[‖]The dependence on data, D, is eliminated by integrating the posterior PDF over all possible values of data, which is equivalent to replacing any occurrence of D by $T(P)$ because the likelihood function is now $\delta_{\text{Dirac}}(T(P) - D)$. The integration over all data values is conventionally referred to as *data marginalization*.

model—$T(P) = P$, where P is a single scalar parameter—could be used to illustrate this framework for the following prior expectation values: $\langle P\rangle = 2$, $\langle D\rangle = (1, 3)$. In this hypothetical scenario, an evaluator conventionally aligns the data to some value (e.g., $(2, 2)$) before initiating the evaluation; this action constitutes a formal departure from Bayes' theorem. In the new framework, the inconsistent dataset could be evaluated intact by setting values of $\langle\delta\rangle'$ equal to, for example, $(0, 0)$ without any need to adjust inconsistent data before evaluation, thus preserving the harmony with Bayes' theorem. The exact posterior expectation values depend on the exact value of the prior generalized data covariance matrix, $\mathbf{C}$, and the user-specified posterior imperfection covariance matrix, $\langle\mathbf{\Delta}\rangle'$, which is illustrated in Section 2.2.

2.2 LINEAR MODELS AND NORMAL PDFs

The general concepts introduced in Section 2.1 become transparent when the model, $T(P)$, is a linear function of parameters, P, and when all PDFs are assumed to be normal because of the simplifications this special case affords and because the transition to the conventional χ^2-minimization method is shown to be straightforward. A generalized cost function is defined by collecting the exponents of the likelihood function and the prior PDF on the right-hand side of Eq. (13) to obtain

$$p'(z|\beta\gamma) \leftarrow N' e^{-\frac{1}{2}X^2(z)}, \tag{16}$$

where

$$\begin{aligned} X^2(z) &\equiv (\delta-\lambda)^\intercal \mathbf{\Lambda}^{-1}(\delta-\lambda) + (z-\langle z\rangle)^\intercal \mathbf{C}^{-1}(z-\langle z\rangle) &(17)\\ &= X^2(\langle z\rangle') + (z-\langle z\rangle')^\intercal \mathbf{C}'^{-1}(z-\langle z\rangle'), &(18)\end{aligned}$$

and where

$$X^2(\langle z\rangle') = (\langle\delta\rangle - \langle\delta\rangle')^\intercal \frac{1}{\mathbf{\Delta}-\mathbf{\Delta}'}(\langle\delta\rangle - \langle\delta\rangle') \tag{19}$$

is the value of the cost function at its minimum.[**] A linear model, $T(P)$, is assumed to be perfect, λ and $\mathbf{\Lambda}$ play a role akin to Lagrange multipliers whose values are to be determined from the constraints on the posterior expectation values $\langle\delta\rangle'$ and $\mathbf{\Delta}' = \langle(\delta-\langle\delta\rangle')(\delta-\langle\delta\rangle')^\intercal\rangle'$, and $N' = 1/\sqrt{(2\pi)^{(k)}|C'|}$ is a normalization constant and k is the dimension of z.[††]

The posterior covariance matrix, $\mathbf{C}'$, is a Hessian of the cost function

$$\mathbf{C}'^{-1} = [\nabla_z \nabla_z^\intercal] X(z), \tag{20}$$

which, when applied to Eq. (17), becomes

$$\mathbf{C}'^{-1} = \mathcal{S}\mathbf{\Lambda}^{-1}\mathcal{S}^\intercal + \mathbf{C}^{-1}, \tag{21}$$

where $\mathcal{S}$ is a generalized data sensitivity matrix:

$$\mathcal{S} \equiv \nabla_z \delta^\intercal = \begin{pmatrix}\nabla_P \\ \nabla_D\end{pmatrix}(T(P)-D)^\intercal = \begin{pmatrix}\nabla_P T(P)^\intercal \\ -\nabla_D D^\intercal\end{pmatrix} = \begin{pmatrix}\mathbf{S} \\ -\mathbb{1}\end{pmatrix}, \tag{22}$$

[**]This expression may be seen as a generalization of the minimum value of the conventional cost function that is recovered by setting $\langle\delta\rangle' = 0$ and $\mathbf{\Delta}' = \mathbb{0}$; for example, see Eq. (6.6.36) of the TSURFER code documentation [3].

[††]As discussed in Section 2.1, for $\lambda = 0$, and in the limit $\mathbf{\Lambda} \to \mathbb{0}$, the likelihood function becomes a Dirac delta function in $(T(P) - D)$, which enables data, D, to be straightforwardly integrated out (i.e., marginalized). Consequently, D is replaced by $T(P)$ in the prior PDF, thus yielding the conventional form of the cost function, known as χ^2.

where

$$\mathbf{S} \equiv \nabla_P T(P)^\intercal, \tag{23}$$

is a conventional sensitivity matrix, which is a constant matrix (i.e., independent of P) for linear models. The dimensions of $\mathbf{S}$ are $(\#P \times \#D)$, where $\#D$ is the number of points in dataset D, and $\#P$ is the number of parameters in parameter set P, and where

$$\nabla_D D^\intercal = \mathbb{1} \tag{24}$$

is a $(\#D \times \#D)$ identity matrix. Analogously,

$$\mathcal{S}^\intercal \equiv (\mathbf{S}^\intercal, -\mathbb{1}). \tag{25}$$

Prior expectation values of δ are

$$\langle \delta \rangle \equiv \langle T(P) \rangle - \langle D \rangle, \tag{26}$$

which, for a linear model $T(P)$, is equal to

$$\langle \delta \rangle \equiv T(\langle P \rangle) - \langle D \rangle. \tag{27}$$

Then, for linear models,

$$T(P) = T(\langle P \rangle) + \mathbf{S}^\intercal (P - \langle P \rangle), \tag{28}$$

which yields

$$\delta - \langle \delta \rangle = \mathcal{S}^\intercal (z - \langle z \rangle). \tag{29}$$

Inserting Eq. (29) into the expression for the prior covariance matrix of δ, namely

$$\mathbf{\Delta} \equiv \langle (\delta - \langle \delta \rangle)(\delta - \langle \delta \rangle)^\intercal \rangle, \tag{30}$$

yields

$$\mathbf{\Delta} = \mathcal{S}^\intercal \mathbf{C} \mathcal{S}, \tag{31}$$

expressed in terms of the prior generalized data covariance matrix $\mathbf{C}$. A completely analogous expression for the corresponding posterior matrix is:

$$\mathbf{\Delta}' = \mathcal{S}^\intercal \mathbf{C}' \mathcal{S}, \tag{32}$$

Inserting the expression for $\mathbf{C}'$ from Eq. (21) into Eq. (32) yields

$$\begin{aligned}
\mathbf{\Delta}' &= \mathcal{S}^\intercal (\mathbf{C}^{-1} + \mathcal{S}\mathbf{\Delta}^{-1}\mathcal{S}^\intercal)^{-1}\mathcal{S} \\
&= \mathcal{S}^\intercal [(\mathbb{1} + \mathcal{S}\mathbf{\Delta}^{-1}\mathcal{S}^\intercal \mathbf{C})\mathbf{C}^{-1}]^{-1}\mathcal{S} \\
&= \mathcal{S}^\intercal \mathbf{C} (\mathbb{1} + \mathcal{S}\mathbf{\Delta}^{-1}\mathcal{S}^\intercal \mathbf{C})^{-1}\mathcal{S} \\
&= \mathcal{S}^\intercal \mathbf{C}\mathcal{S} (\mathbb{1} + \mathbf{\Delta}^{-1}\mathcal{S}^\intercal \mathbf{C}\mathcal{S})^{-1} \\
&= \mathbf{\Delta}(\mathbb{1} + \mathbf{\Delta}^{-1}\mathbf{\Delta})^{-1} \\
&= [(\mathbb{1} + \mathbf{\Delta}^{-1}\mathbf{\Delta})\mathbf{\Delta}^{-1}]^{-1} \\
&= (\mathbf{\Delta}^{-1} + \mathbf{\Delta}^{-1})^{-1},
\end{aligned} \tag{33}$$

where the following matrix identities have been used:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}, \tag{34}$$

$$(\mathbb{1} + \mathbf{AB})^{-1}\mathbf{A} = \mathbf{A}(\mathbb{1} + \mathbf{BA})^{-1}. \tag{35}$$

Equation (33) yields an expression for $\mathbf{\Lambda}$ in terms of $\mathbf{\Delta}'$ and $\mathbf{\Delta}$,

$$\mathbf{\Lambda}^{-1} = \mathbf{\Delta}'^{-1} - \mathbf{\Delta}^{-1}, \tag{36}$$

so that the posterior covariance matrix $\mathbf{C}'$ can be expressed in terms of $\mathbf{\Delta}'$ and the prior covariance matrix $\mathbf{C}$ via Eq. (21):

$$\mathbf{C}'^{-1} = \mathcal{S}[\mathbf{\Delta}'^{-1} - \mathbf{\Delta}^{-1}]\mathcal{S}^{\intercal} + \mathbf{C}^{-1}. \tag{37}$$

The posterior expectation value of generalized data, $\langle z\rangle'$, can found explicitly by minimizing $X(z)$:

$$\nabla_z X(z)|_{z=\langle z\rangle'} = 0, \tag{38}$$

which yields

$$\mathcal{S}\mathbf{\Lambda}^{-1}(\langle\delta\rangle' - \lambda) + \mathbf{C}^{-1}(\langle z\rangle' - \langle z\rangle) = 0, \tag{39}$$

because for linear models, the cost function is minimized by the posterior expectation values of generalized data, $\langle z\rangle'$. This is in accordance with the constraint set at the outset: $\langle z\rangle'$ must be consistent with the evaluator-specified value of $\langle\delta\rangle'$. Multiplying both sides of Equation (39) by $\mathbf{C}$ yields

$$\langle z\rangle' - \langle z\rangle = -\mathbf{C}\mathcal{S}\mathbf{\Lambda}^{-1}(\langle\delta\rangle' - \lambda). \tag{40}$$

To find an explicit expression for λ, Eq. (40) is multiplied by $\mathcal{S}^{\intercal}$ to yield, via Eqs. (29) and (31),

$$\langle\delta\rangle' - \langle\delta\rangle = -\mathbf{\Delta}\mathbf{\Lambda}^{-1}(\langle\delta\rangle' - \lambda), \tag{41}$$

which yields

$$\langle\delta\rangle' - \lambda = -\mathbf{\Lambda}\mathbf{\Delta}^{-1}(\langle\delta\rangle' - \langle\delta\rangle), \tag{42}$$

and

$$\begin{aligned} \lambda &= \langle\delta\rangle' + \mathbf{\Lambda}\mathbf{\Delta}^{-1}(\langle\delta\rangle' - \langle\delta\rangle), & (43)\\ &= (\mathbb{1} + \mathbf{\Lambda}\mathbf{\Delta}^{-1})\langle\delta\rangle' - \mathbf{\Lambda}\mathbf{\Delta}^{-1}\langle\delta\rangle. & (44) \end{aligned}$$

Multiplying both sides of Eq. (44) by $\mathbf{\Lambda}^{-1}$ yields

$$\mathbf{\Lambda}^{-1}\lambda = (\mathbf{\Lambda}^{-1} + \mathbf{\Delta}^{-1})\langle\delta\rangle' - \mathbf{\Delta}^{-1}\langle\delta\rangle. \tag{45}$$

Expression for $\mathbf{\Delta}'^{-1}$ in Eq. (33) can be used to rewrite it as

$$\mathbf{\Lambda}^{-1}\lambda = \mathbf{\Delta}'^{-1}\langle\delta\rangle' - \mathbf{\Delta}^{-1}\langle\delta\rangle. \tag{46}$$

Eq. (41) could be used to express $\langle\delta\rangle'$ in terms of λ and $\langle\delta\rangle$:

$$\langle\delta\rangle' = (1 + \mathbf{\Delta}\mathbf{\Lambda}^{-1})^{-1}(\langle\delta\rangle + \mathbf{\Delta}\mathbf{\Lambda}^{-1}\lambda) \tag{47}$$

Finally, inserting Eq. (42) into Eq. (40) yields an expression for $\langle z\rangle'$ in terms of $\langle\delta\rangle'$ and the prior covariance matrix $\mathbf{C}$:

$$\langle z\rangle' - \langle z\rangle = \mathbf{C}\mathcal{S}\mathbf{\Delta}^{-1}(\langle\delta\rangle' - \langle\delta\rangle), \tag{48}$$

or, equivalently,

$$\langle z\rangle' = \langle z\rangle + \mathbf{C}\mathcal{S}\boldsymbol{\Delta}^{-1}(\langle\delta\rangle' - \langle\delta\rangle). \tag{49}$$

The posterior PDF has thus been completely determined by the explicit solutions for λ and $\boldsymbol{\Lambda}$ expressed in terms of user-specified posterior expectation values of δ, $\langle\delta\rangle'$; its posterior covariance matrix, $\boldsymbol{\Delta}'$; the prior expectation values of generalized data, $\langle z\rangle$; and its prior covariance matrix, $\mathbf{C}$. Equivalent expressions are also provided for corresponding posterior expectation values and covariance of generalized data, $\langle z\rangle'$ and $\mathbf{C}'$, respectively. Note that the priors $\langle\delta\rangle$ and $\boldsymbol{\Delta}$ are defined by Eqs. (29) and (31). Either form of the posterior PDF, whether defined using Eq. (17) or Eq. (18), provides an equivalent form of the posterior PDF for calculating posterior expectation values consistently with Bayes' theorem.

2.3 CONVENTIONAL BAYESIAN EVALUATION BY MHMC

This section describes two variants of an iterative MHMC algorithm for the efficient stochastic computation of posterior PDFs corresponding to the conventional χ^2-minimization problem (i.e., for $z \leftarrow (P, T(P))$).

Iteration Scheme A: For a given prior PDF defined by $\langle z\rangle$ and $\mathbf{C}$, initiate the iterations by setting:

$$\lambda \leftarrow 0, \tag{50}$$

$$\boldsymbol{\Lambda} \leftarrow \boldsymbol{\Delta}. \tag{51}$$

Then:

1. For the current values of $\langle z\rangle$, $\mathbf{C}$, λ and $\boldsymbol{\Lambda}$, use the MHMC method to compute the posterior PDF, then calculate the posteriors: $\langle z\rangle'$, $\mathbf{C}'$, $\langle\delta\rangle'$, and $\boldsymbol{\Delta}'$. (For this particular choice—namely, $\boldsymbol{\Lambda} = \boldsymbol{\Delta}$ and $\lambda = 0$—and for a linear model, the analytical expressions could be used to show that $\boldsymbol{\Delta}' = \boldsymbol{\Delta}/2$ and $\langle\delta\rangle' = \langle\delta\rangle/2$ for comparison with the corresponding posteriors calculated numerically by the MHMC method in this step.)

2. Define a new prior PDF by using the posterior expectation values calculated by the MHMC method in the previous step:

$$\langle z\rangle \leftarrow \langle z\rangle' \tag{52}$$

$$\mathbf{C} \leftarrow \mathbf{C}'. \tag{53}$$

   (This definition also implies that $\boldsymbol{\Delta} \leftarrow \boldsymbol{\Delta}'$ and $\langle\delta\rangle \leftarrow \langle\delta\rangle'$.)

3. Update $\boldsymbol{\Lambda}$ by the new prior matrix $\boldsymbol{\Delta}$, ($\lambda = 0$ is kept fixed):

$$\boldsymbol{\Lambda} \leftarrow \boldsymbol{\Delta}. \tag{54}$$

4. Return to the Step 1, until convergence (e.g.,, when $\boldsymbol{\Delta}'$ and $\langle\delta\rangle'$ start to fluctuate around zero as a result of the MHMC method's finite accuracy).

Because the matrix $\boldsymbol{\Lambda}$ is halved in each iteration, this scheme should converge to $\boldsymbol{\Delta}' \to 0$ and $\langle\delta\rangle' \to 0$, corresponding to the solution of the conventional χ^2-minimization (Bayesian evaluation) method. (In principle, a factor other than 2 could be chosen as the divisor in each iteration.)

Iterative Scheme B: For a given prior PDF defined by $\langle z\rangle$ and $\mathbf{C}$, initiate the iteration scheme by setting

$$\lambda \leftarrow -\langle\delta\rangle, \tag{55}$$
$$\mathbf{\Lambda} \leftarrow \mathbf{\Delta}. \tag{56}$$

For a linear model, this choice leads to posteriors $\langle\delta\rangle' = 0$ and $\mathbf{\Delta}' = \mathbf{\Delta}/2$. This posterior PDF can be used as a new prior PDF to be fed directly into the initiation step of Iteration Scheme A because $\lambda \leftarrow \langle\delta\rangle \leftarrow \langle\delta\rangle' = 0$ will cause λ to remain 0 in all subsequent iterations, like in Iteration Scheme A. Note that $\langle\delta\rangle' = 0$ will also remain zero in all subsequent iterations for this choice, which is not necessarily the case for Iteration Scheme A alone.

The posterior expectation values of generalized data in Iteration Scheme A can be shown to converge as

$$\langle z\rangle_{n+1} \leftarrow \langle z\rangle'_n = \langle z\rangle - \mathbf{C}\mathcal{S}\mathbf{\Delta}^{-1}\langle\delta\rangle(1-2^{-n}), \tag{57}$$
$$\xrightarrow[n\to\infty]{} \langle z\rangle_{\chi^2} = \langle z\rangle - \mathbf{C}\mathcal{S}\mathbf{\Delta}^{-1}\langle\delta\rangle, \tag{58}$$

whereas Iteration Scheme B achieves the same limit above in its first iteration due to $\langle\delta\rangle' = 0$. An equivalent expression for $\langle z\rangle_{\chi^2}$ in the conventional χ^2-minimization formalism is provided in Eq. (67). The posterior covariance matrix of generalized data of both iteration schemes can be shown to converge as

$$\mathbf{C}^{-1}_{n+1} \leftarrow \mathbf{C}'^{-1}_n = \mathbf{C}^{-1}_n + \mathcal{S}\left(\frac{\mathbf{\Delta}}{2^{n-1}}\right)^{-1}\mathcal{S}^\intercal \quad (n>0), \tag{59}$$
$$= \mathbf{C}^{-1} + \mathcal{S}\left(\frac{\mathbf{\Delta}}{2^{n}-1}\right)^{-1}\mathcal{S}^\intercal, \tag{60}$$
$$\xrightarrow[n\to\infty]{} \mathbf{C}^{-1}_{\chi^2}, \tag{61}$$

where $\langle z\rangle$ and $\mathbf{C}$ define the original prior PDF, all expectation values ($\langle\delta\rangle$ and $\mathbf{\Delta}$) are computed using the original prior PDF, $\mathbf{C}_{\chi^2}$ is the covariance matrix calculated using a conventional χ^2-minimization method, and n is the ordinal number of an iteration. This is confirmed by a converged posterior PDF computed by either iteration scheme and by the analytical solution to the conventional χ^2-minimization problem, stated below for a linear model. These equations were derived by observing that in this iterative scheme, the original prior, defined by $\langle z\rangle$ and $\mathbf{C}$, could be considered as the formal prior at any iteration—not just for the first iteration—because of the recursive expressions for posteriors $\langle z\rangle'$ and $\mathbf{C}'$ that make regrouping terms within it possible at each iteration. Thus, the original prior PDF can be considered a prior PDF of each iteration, including of the converged posterior PDF, that is the conventional posterior PDF, also defined by the same priors $\langle z\rangle$ and $\mathbf{C}$. Each time a posterior PDF is computed using MHMC, it counts as one iteration, and the iteration number, n, is incremented by one. Although $\mathbf{C}^{-1}_{n+1} \to \infty$ as $n \to \infty$, the iteration scheme should make this limit approachable to the extent allowed by the accuracy of the MHMC method. The sensitivity matrix defined by Fröhner in JEFF Report 18 [1] for the conventional χ^2-minimization method is a transpose of the sensitivity matrix defined in Eq. (22):

$$\mathbf{S} = \mathbf{S}^\intercal_{\chi^2}, \tag{62}$$

where $\mathbf{S}$ is a $(\#P \times \#D)$ sensitivity matrix. However, this difference in notation is simply due to an inconsequential arbitrary choice in defining a sensitivity matrix as its transpose. For comparison,

$$\mathcal{S}_{\chi^2} = \nabla^\intercal_P \begin{pmatrix} P \\ T(P) \end{pmatrix} = \begin{pmatrix} \mathbb{1}_P \\ \mathbf{S}_{\chi^2} \end{pmatrix} \quad cf. \quad \mathcal{S} = \nabla_z (T(P)-D)^\intercal = \begin{pmatrix} \mathbf{S} \\ -\mathbb{1}_D \end{pmatrix}, \tag{63}$$

where $\mathcal{S}_{\chi^2}$ and $\mathcal{S}$ are $((\#P + \#D) \times \#P)$ and $((\#P + \#D) \times \#D)$ generalized sensitivity matrices, respectively, which illuminates the origin of this otherwise inconsequential difference.) For comparison, a solution to the conventional χ^2-minimization problem for a linear model is

$$\mathbf{M}'^{-1}_{\chi^2} = \mathcal{S}^{\mathsf{T}}_{\chi^2}\mathbf{C}^{-1}\mathcal{S}_{\chi^2}, \tag{64}$$

$$\langle P\rangle'_{\chi^2} = \langle P\rangle + [\mathcal{S}^{\mathsf{T}}_{\chi^2}\mathbf{C}^{-1}\mathcal{S}_{\chi^2}]^{-1}\mathcal{S}^{\mathsf{T}}_{\chi^2}\mathbf{C}^{-1}\begin{pmatrix}\langle P\rangle - \langle P\rangle \\ \langle D\rangle - T(\langle P\rangle)\end{pmatrix} \tag{65}$$

$$= \langle P\rangle - \mathbf{M}'_{\chi^2}\mathcal{S}^{\mathsf{T}}_{\chi^2}\mathbf{C}^{-1}\begin{pmatrix}0 \\ \langle\delta\rangle\end{pmatrix}. \tag{66}$$

With the posterior expectation values and covariance of parameters, P, for the conventional χ^2 method thus in hand, corresponding posterior expectation values and covariances for generalized data become

$$\langle z\rangle'_{\chi^2} = (\langle P\rangle'_{\chi^2}, T(\langle P\rangle'_{\chi^2})) \quad \text{and} \tag{67}$$

$$\mathbf{C}'_{\chi^2} = \begin{pmatrix}\mathbf{M}'_{\chi^2} & \mathbf{M}'_{\chi^2}\mathbf{S}^{\mathsf{T}}_{\chi^2} \\ \mathbf{S}_{\chi^2}\mathbf{M}'_{\chi^2} & \mathbf{S}_{\chi^2}\mathbf{M}'_{\chi^2}\mathbf{S}^{\mathsf{T}}_{\chi^2}\end{pmatrix}, \tag{68}$$

respectively, which are equivalent to the $n \to \infty$ limit solution of the iterative scheme, as derived in Eqs. (58) and (60), respectively.

2.4 NEWTON–RAPHSON ITERATIVE COST MINIMIZATION FOR NONLINEAR MODELS

The analytical solutions for λ and $\mathbf{\Lambda}$ that satisfy constraints on posterior expectation values ($\langle\delta\rangle'$ and $\mathbf{\Delta}'$) are much more difficult to find for nonlinear models than for linear models. Therefore, an iterative solution for cost function minimization is provided in Section 2.4 for arbitrary λ and $\mathbf{\Lambda}$ without imposing constraints on the posterior expectation values $\langle\delta\rangle'$ and $\mathbf{\Delta}'$. (The constraints are imposed in Section 2.5.)

For nonlinear models, Eq. (29) can be stated as a first order approximation,

$$\delta_{n+1} - \delta_n \approx \mathcal{S}^{\mathsf{T}}_n(z_{n+1} - z_n), \tag{69}$$

where

$$\mathcal{S}_n \equiv \mathcal{S}|_{z=z_n}, \tag{70}$$

$$\delta_n \equiv \delta(z_n) = T(P_n) - D_n. \tag{71}$$

The Newton–Raphson iterative scheme can be constructed by replacing $\langle\delta\rangle'$ and $\langle z\rangle'$ by δ_{n+1} and z_{n+1}, respectively, in Eq. (39) and then using Eq. (69) to eliminate δ_{n+1}. Finally, solving for z_{n+1} yields

$$z_{n+1} = [\mathbb{1} + \mathbf{C}\mathcal{S}_n\mathbf{\Lambda}^{-1}\mathcal{S}^{\mathsf{T}}_n]^{-1}[\langle z\rangle - \mathbf{C}\mathcal{S}_n\mathbf{\Lambda}^{-1}(\delta_n - \mathcal{S}^{\mathsf{T}}_n z_n - \lambda)]. \tag{72}$$

This iterative scheme converges to the value of z that minimizes the cost function, called $z_{\min}$, for convenience. Conventionally, the posterior expectation value, $\langle z\rangle'$, is approximated by $z_{\min}$. The accuracy of this approximation could be quantified by comparing $z_{\min}$ with the $\langle z\rangle'$ computed by the MHMC method.

The analytical solutions for linear models are a special case of this iterative scheme—as shown by setting $z_n = \langle z \rangle$, $z_{n+1} = \langle z \rangle'$, $\delta_n = \langle \delta \rangle$, and $\mathcal{S}_n = \mathcal{S}$—because sensitivity $\mathcal{S}$ is constant for linear models.

The expression for the posterior covariance matrix for nonlinear models is obtained in the same way as for linear models, that is, by applying the Hessian operator to the cost function. For nonlinear models, the posterior covariance matrix attains two new terms in addition to the terms already seen in Eq. (21):

$$\mathbf{C}'^{-1} = \mathbf{C}^{-1} + \mathcal{S}\mathbf{\Lambda}^{-1}\mathcal{S}^{\intercal} + \frac{1}{2}\{[(\nabla_z \nabla_z^{\intercal})\delta^{\intercal}]\mathbf{\Lambda}^{-1}(\delta - \lambda) + (\delta - \lambda)^{\intercal}\mathbf{\Lambda}^{-1}[(\nabla_z \nabla_z^{\intercal})\delta]\}, \tag{73}$$

where all the quantities (e.g. δ, the sensitivities, and the Hessian) are evaluated at the iterative solution, $z_{\min}$, and the sum of the two new terms is a symmetric matrix because the two new terms are transposes of each other. The functional form defining δ is not limited to $\delta = T(P) - D$ but could take an alternative form that may be more appropriate in some other context, which will affect the evaluation of the Hessian operator. For $\delta = T(P) - D$ and a nonlinear model, $T(P)$, the application of the Hessian operator yields a $(2 \times 2$ block) matrix that is zero everywhere except in the block corresponding to the parameter subset of generalized data. However, when δ is defined as the shortest distance from a line, as in the application to line fitting, the Hessian must be evaluated accordingly. The algorithm presented in Section 2.4 will be verified numerically in a future work.

2.5 DETERMINING $\mathbf{\Lambda}^{-1}$ AND λ FOR NONLINEAR MODELS

An iterative solution for $\mathbf{\Lambda}^{-1}$ and λ, constrained by user-defined posterior expectation values $\mathbf{\Delta}'$ and δ', proceeds by initializing $\mathbf{\Lambda}^{-1}$ and λ and iterating until convergence is reached.

For initiation, Eq. (36) is used to initiate the $\mathbf{\Lambda}^{-1}$ matrix as

$$\mathbf{\Lambda}_1^{-1} = \mathbf{\Delta}'^{-1} - \mathbf{\Delta}^{-1}. \tag{74}$$

Then, this initial value, $\mathbf{\Lambda}_1^{-1}$, is inserted into Eq. (46) to yield

$$\mathbf{\Lambda}_1^{-1}\lambda_1 = \mathbf{\Delta}'^{-1}\langle \delta \rangle' - \mathbf{\Delta}^{-1}\langle \delta \rangle, \tag{75}$$

which is multiplied by $\mathbf{\Lambda}_1$ to yield the initial value λ_1.

The inverse of matrix $\mathbf{\Lambda}_{n+1}$ is found by correcting the current matrix $\mathbf{\Lambda}_n^{-1}$ by the amount in which the corresponding posterior expectation matrix $\mathbf{\Delta}_n'^{-1}$ is found to deviate from the user-defined posterior expectation matrix $\mathbf{\Delta}'^{-1}$, namely,

$$\mathbf{\Lambda}_{n+1}^{-1} = \mathbf{\Lambda}_n^{-1} + (\mathbf{\Delta}'^{-1} - \mathbf{\Delta}_n'^{-1}), \tag{76}$$

where the term in parentheses is the amount by which $\mathbf{\Lambda}_n^{-1}$ is corrected, and $\mathbf{\Delta}_n'^{-1}$ is a posterior expectation matrix calculated for the PDF defined by $\mathbf{\Lambda}_n^{-1}$ and λ_n (as is $\langle \delta \rangle_n'$ used below). For example, when $\mathbf{\Delta}_n'^{-1}$ is smaller than $\mathbf{\Delta}'^{-1}$, then $\mathbf{\Lambda}_n^{-1}$ is increased by that amount in an attempt to force $\mathbf{\Delta}_{n+1}'^{-1}$ closer to $\mathbf{\Delta}'^{-1}$.

Similar logic is applied to Eq. (75) to set up an iterative scheme for $\mathbf{\Lambda}_{n+1}^{-1}\lambda_{n+1}$:

$$\mathbf{\Lambda}_{n+1}^{-1}\lambda_{n+1} = \mathbf{\Lambda}_n^{-1}\lambda_n + (\mathbf{\Delta}'^{-1}\langle \delta \rangle' - \mathbf{\Delta}_n^{-1}\langle \delta \rangle_n'), \tag{77}$$

where the term in the parentheses is the amount by which $\mathbf{\Lambda}_n^{-1}\lambda_n$ is corrected. A straightforward multiplication of the previous expression by $\mathbf{\Lambda}_{n+1}$, already computed via Eq. (76), yields an expression for λ_{n+1}.

Posterior expectation values in each iteration could be computed via the MHMC method or via the Newton–Raphson method in Section 2.4 in which the former would be expected to be more accurate than the latter because Newton–Raphson assumes normal PDFs throughout. When using the Newton–Raphson method to compute posterior expectation values in each iteration, that is, for $\mathbf{\Lambda}_n^{-1}$ and λ_n, the sensitivities needed for computing the posterior expectation matrix $\mathbf{\Delta}'_n$ are to be evaluated at the posterior expectation values, $\langle z\rangle'_n$. Convergence toward the true posterior computed by MHMC could be monitored in each iteration, starting with the first one (i.e., for $\mathbf{\Lambda}_1$ and λ_1). The solution calculated by the linear approximation could be used as a reference calculation to quantify the full spectrum of approximations and the corresponding deviations from the presumably most accurate MHMC results. The $\mathbf{\Lambda}^{-1}$ and λ thus determined are used to define the posterior PDF for the computation of any desired posterior expectation values. The algorithm presented in Section 2.5 will be verified numerically in a future work.

2.6 PEELLE'S PERTINENT PUZZLE REVISITED

One way to resolve PPP* by this formalism is to set $\mathbf{\Delta}' = \mathbb{0}$ while retaining $\langle\delta\rangle'$ as a user-imposed constraint on the posterior expectation values of $\delta = T(P) - D$; Ref. [15] provides further discussion on PPP. The example used to demonstrate PPP is specified by $T(P) = P$, where P is a single scalar parameter, and prior data, $\langle D\rangle$, is a vector of two scalar data points with a covariance matrix defined so that the posterior expectation value of P, $\langle P\rangle'$, is outside the range of prior data.

For $\mathbf{\Delta}' = \mathbb{0}$, the analytical expressions yield $\mathbf{\Lambda} = \mathbb{0}$ and $\lambda = \langle\delta\rangle'$. Therefore,

$$\mathcal{L}(\langle\delta\rangle', (\mathbf{\Delta}' = \mathbb{0})|z, \langle z\rangle, \mathbf{C}) = \delta_{\text{Dirac}}(T(P) - D - \langle\delta\rangle'), \tag{78}$$

and the posterior PDF becomes,†

$$p'(P|\beta\gamma) \leftarrow N'\mathcal{N}((z|_{D=T(P)-\langle\delta\rangle'})|\langle z\rangle, \mathbf{C}). \tag{79}$$

Then, $\chi^2(P|\langle\delta\rangle')$ appears in the exponent of the normal PDF above, $e^{-\chi^2(P|\langle\delta\rangle')/2}$, parameterized by the user-imposed constraint on posterior expectation values, $\langle\delta\rangle'$:

$$\chi^2(P|\langle\delta\rangle') = (z|_{D=(T(P)-\langle\delta\rangle')} - \langle z\rangle)^\intercal \mathbf{C}^{-1}(z|_{D=(T(P)-\langle\delta\rangle')} - \langle z\rangle), \tag{80}$$

which can be minimized by using conventional methods to yield the posterior expectation value of parameter P, $\langle P\rangle'$, in a way that is completely determined by the evaluator-imposed constraint on the posterior, $\langle\delta\rangle'$. For $\mathbf{\Delta}' = \mathbb{0}$, the posterior covariance matrix of parameters is equal to that calculated by the conventional minimization method, as seen in Ref. [1]. Also, note that setting $D = (T(P) - \langle\delta\rangle')$ is equivalent to setting $D = T(P)$ and adding $\langle\delta\rangle'$ to the prior expectation values of data, $\langle D\rangle$ (i.e., $\langle D\rangle \leftarrow \langle D\rangle + \langle\delta\rangle'$). This provides a direct way to modify prior expectation values of data only—without making any changes to the covariances—while maintaining harmony with Bayes' theorem. However, for nonlinear models, the posterior sensitivity matrix is evaluated at the $\langle z\rangle' = (\langle P\rangle', T(\langle P\rangle'))$, where, for normal PDFs, $\langle P\rangle'$ are approximated by the parameter values that minimize $\chi^2(P)$ computed by the aforementioned Newton–Raphson iterative method. In particular, the sensitivity matrix is evaluated at $\langle z\rangle'$ when computing the posterior covariance matrix of model predictions (i.e., the "evaluated" covariance

*Thanks to Klaus Guber for this suggestion made on February 18, 2021.

†The PDF changes by integrating over all data D, which is equivalent to replacing any occurrence of D by $T(P) - \langle\delta\rangle'$ due to the Dirac delta function. This eliminates any dependence on data D.

matrix.) Furthermore, this formalism also provides a way to affect the posterior covariance matrices by letting an evaluator simultaneously impose constraint on $\mathbf{\Lambda}'(\neq 0)$ and $\langle\delta\rangle'$, revealing a completely new space of possible solutions to the PPP. Application of the Bayesian framework to evaluations in which PPP is present will be discussed in a future publication.

3. NUMERICAL IMPLEMENTATION

3.1 METROPOLIS-HASTINGS MONTE CARLO

To solve Bayes' equation, an MC method was designed to calculate posteriors. The most apt types of MC methods for this application fall under the category of Markov Chain MC, which all generate an autocorrelated[‡] ensemble of samples from a probability distribution. A Metropolis–Hastings criterion for accepting posterior samples was chosen for this work; the algorithm is fully described by Refs. [16; 17; 18]. A general outline of the MHMC method is given in Algorithm 1.

Algorithm 1 Metropolis–Hastings Algorithm

$N \leftarrow$ iterations ▷ necessary number of iterations for convergence
$i \leftarrow 0$
$z_0 \leftarrow$ arbitrary values ▷ z_i can be of arbitrary dimension
while $i < (N + 1)$ **do**
    Generate random candidate sample z' from $g(z'|z_i)$
    $A = \min\left(1, \frac{p(z')}{p(z_i)} \frac{g(z_i|z')}{g(z'|z_i)}\right)$
    Generate random value of u from uniform distribution between $(0, 1)$
    if $u < A$ **then**
        $z_{i+1} \leftarrow z'$
    else
        $z_{i+1} \leftarrow z_i$
    $i \leftarrow i + 1$

Here, $g(z'|z_i)$ is defined as the "proposal distribution" or "sampling distribution," and $p(z)$ is a PDF that is proportional to the desired posterior PDF $p(z)'$. The MHMC algorithm is designed so that it should explore the entire probability space where $p(z) > 0$ and randomly exit local minima. This prevents the algorithm from getting "stuck". The algorithm itself is simple, but the shape of the sampling distribution can be difficult to optimize for a given problem. The optimum acceptance rate was estimated to be approximately 23% [16] and is directly influenced by the sampling distribution.

This implementation of the MHMC algorithm generates samples of z from a sampling distribution $g(z|z_i)$ by sampling the covariance matrix from the prior distribution of z. The sampling distribution and the prior distribution do not necessarily need to be related. In theory, any sampling distribution will eventually produce a true posterior, but in practice, the posterior solution may never be reached in an acceptable amount of time. Therefore, it is important to select a sampling distribution so that the posterior distributions of z can be efficiently generated; the more similar to the posterior PDF, the more efficient the algorithm will typically be. For the linear analytical posterior solution shown herein, the true posterior solution could be generated by sampling the analytically exact posterior solution.

Because many prior distributions have been documented in a covariance matrix, the present implementation samples priors by Cholesky decomposition, where the prior covariance matrix

$$\mathbf{C} = \mathbf{L}\mathbf{L}^{\intercal}, \tag{81}$$

[‡]The correlation of samples in a set must be accounted for in this case by only taking a subset of the posterior ensemble.

and $\mathbf{L}$ is the lower triangular matrix found by the Cholesky decomposition. Then, each new random candidate sample z' is generated from the prior distribution by setting

$$z' = z_i + \mathbf{L}\Re K, \tag{82}$$

where K is a vector of multiplicative factors that scale the sampling distribution for each element of z', and $\Re$ is a vector of random numbers from the normal (Gaussian) distribution (mean = 0 and variance = 1). The application of K a crucial step to control the width of the sampling distribution for every element of z so that the random walk taken by the MHMC algorithm can be made to move about the domain (z) at different rates with respect to the iteration number i. At the conclusion of the MHMC algorithm, a posterior set (i.e. distribution) of z_i has been generated from which the various moments of that distribution (e.g., $\langle z\rangle'$ and $\mathbf{C}'$) can be determined and, in the present case, compared with other methods. This comparison is shown in the next section.

3.2 VALIDATION WITH LINEAR MODELS

To validate the current numerical implementation of MHMC, it is first compared with the analytical solutions of Bayes' equation for linear models. This exercise follows and validates the derivations found in Section 2.2. These methods are demonstrated with different examples to highlight certain features and make the use of the method more clear. As the basis of the linear examples, $\mathbf{C}$ and $\mathbf{\Delta}$ are defined by creating a prior distribution of P and D by the Cholesky decomposition of a starting matrix $\mathbf{C}_{st} = L_{st}L_{st}^{\mathsf{T}}$. Values of $\mathbf{C}_{st}$ are chosen arbitrarily, and subscript "*st*" indicates that this is the starting covariance.

$$\mathbf{C}_{st} = \begin{bmatrix} \Delta P_1^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \Delta P_2^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \Delta D_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Delta D_2^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Delta D_3^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Delta D_4^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Delta D_5^2 \end{bmatrix}, \tag{83}$$

where $\Delta P_1^2 = 1.1$, $\Delta P_2^2 = 1.2$. The variance on D is given by $\Delta D^2 = 0.15\langle D\rangle$, where

$$\begin{aligned} \langle P\rangle &= [5, 8], \\ \langle D\rangle &= [13.1, 17.8, 23.3, 27.2, 33.5], \\ x &= [1, 2, 3, 4, 5]. \end{aligned} \tag{84}$$

It was assumed—for convenience, not because the formalism demands it—that no uncertainty is in the abscissa, x, of the data, D. The prior ensemble of $z = [P, D]$ is calculated by

$$z = \langle z\rangle + L_{st}\Re, \tag{85}$$

where $\mathfrak{R}$ is a vector of random values drawn from the normal distribution (mean = 0 and variance = 1)[§]. Finally, with this linear model defined as

$$T(P) = P_1 x + P_2, \tag{86}$$

the prior covariance matrices $\mathbf{C}$ and $\mathbf{\Lambda}$ can be determined (using Eqs. (7) and (11) along with expectation values $\langle z \rangle$ and $\langle \delta \rangle$. The prior distributions of the elements of z are shown in Fig. 1.

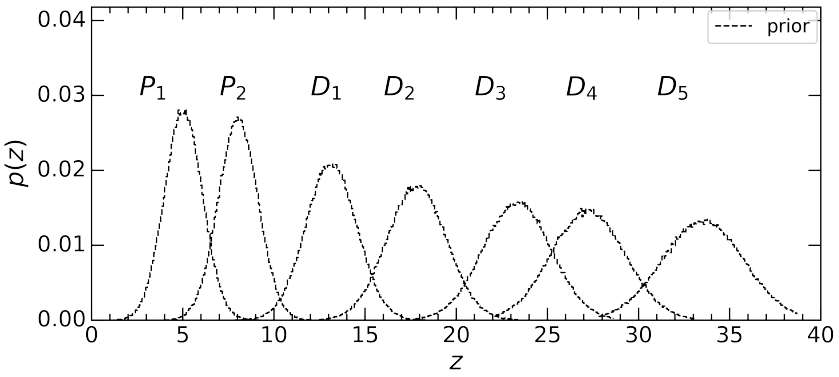


Figure 1. Prior distributions of the elements of z as generated by a random sampling of the starting matrix $\mathbf{C}_{st}$ by Cholesky decomposition (Eq. (85)). Here, the PDF of each component of the generalized data, z, is labeled by the element corresponding to that PDF.

3.2.1 Posterior Δ' Reduced by Two

Using the priors defined previously and following Iteration Scheme A from Section 2.3, consider the case in which the posterior covariance of $\delta = T(P) - D$, $\mathbf{\Delta}'$, is equal to $\mathbf{\Delta}/2$. Starting with Eq. (33) and then replacing $\mathbf{\Delta}'$ with $\mathbf{\Delta}/2$ yields the following:

$$\begin{aligned} \mathbf{\Delta}'^{-1} &= \mathbf{\Lambda}^{-1} + \mathbf{\Delta}^{-1} \\ (\mathbf{\Delta}/2)^{-1} &= \mathbf{\Lambda}^{-1} + \mathbf{\Delta}^{-1} \\ \mathbf{\Lambda} &= \left(2\mathbf{\Delta}^{-1} - \mathbf{\Delta}^{-1}\right)^{-1} \\ \mathbf{\Lambda} &= \mathbf{\Delta}. \end{aligned} \tag{87}$$

The values of $\mathbf{\Lambda}$ can be set to the prior $\mathbf{\Delta}$ for input to the MHMC algorithm. If, at the same time, $\langle \delta \rangle' \leftarrow \langle \delta \rangle / 2$, then starting with Eq. (44), it produces the following:

[§]This Cholesky decomposition sampling to create a prior is entirely separate from the MHMC evaluation process in which the same decomposition method is used to generate new candidates z'.

$$\lambda = (\mathbb{1} + \mathbf{\Lambda}\mathbf{\Delta}^{-1})\langle\delta\rangle' - \mathbf{\Lambda}\mathbf{\Delta}^{-1}\langle\delta\rangle \quad (88)$$

$$= (\mathbb{1} + \mathbf{\Delta}\mathbf{\Delta}^{-1})\langle\delta\rangle/2 - \mathbf{\Delta}\mathbf{\Delta}^{-1}\langle\delta\rangle \quad (89)$$

$$= 0, \quad (90)$$

and λ is simply set equal to 0. With $\mathbf{\Lambda}$ and λ now defined, all input to the MHMC algorithm are defined, and the algorithm can be tested to discover whether it reproduces the expected posterior values of $\langle\delta\rangle'$ and $\mathbf{\Delta}'$. The MHMC solutions are compared with the analytical solutions in Fig. 2.

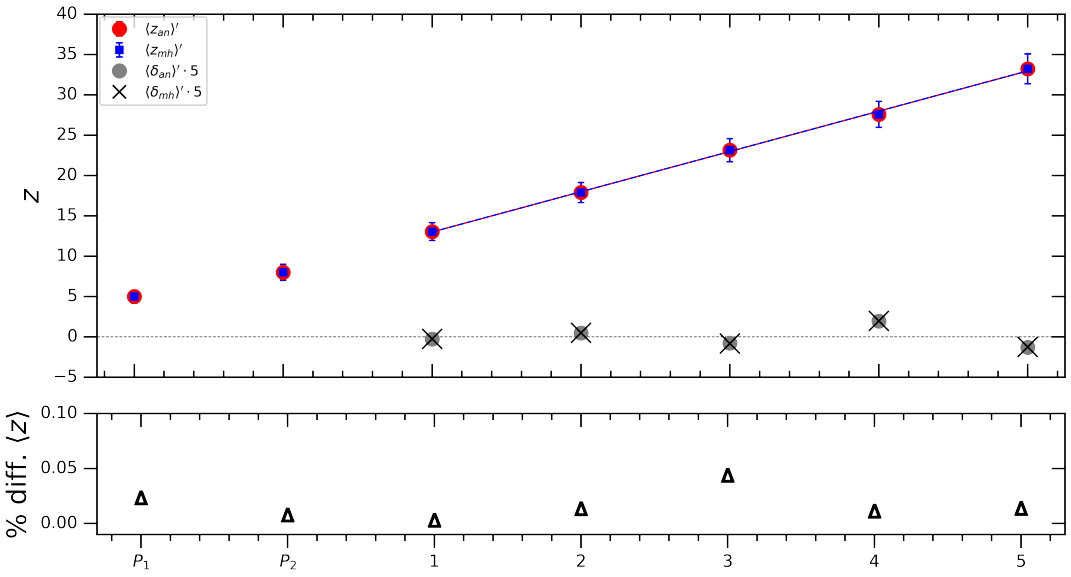


Figure 2. The upper plot shows the perfect (within statistical uncertainty) agreement between the analytical and MHMC values for $\langle z\rangle'$ and $\langle\delta\rangle'$. The two different solutions for $\langle\delta\rangle'$ are multiplied by five to better see the deviation from zero. The lower plot shows that the difference between the analytical and MHMC $\langle z\rangle'$ is not >0.05% for any element of z.

For the posterior values shown, the MHMC algorithm was run for 10^6 iterations to reduce the statistical uncertainty by a reasonable amount. The analytical and MHMC posterior expectation values of z and δ match each other perfectly within the statistical uncertainty of the MC algorithm. The lower plot shows that the difference between the analytical and MHMC $\langle z\rangle'$ is not greater than 0.05% for any element of z. The actual posterior PDFs of z from the MHMC algorithm are shown alongside the prior PDFs in Fig. 3.

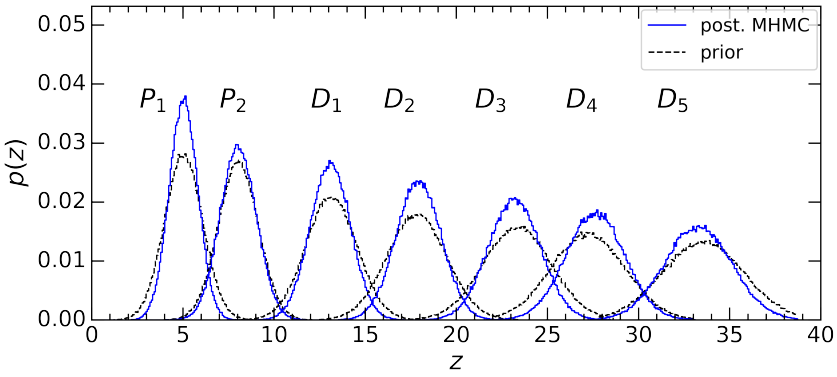
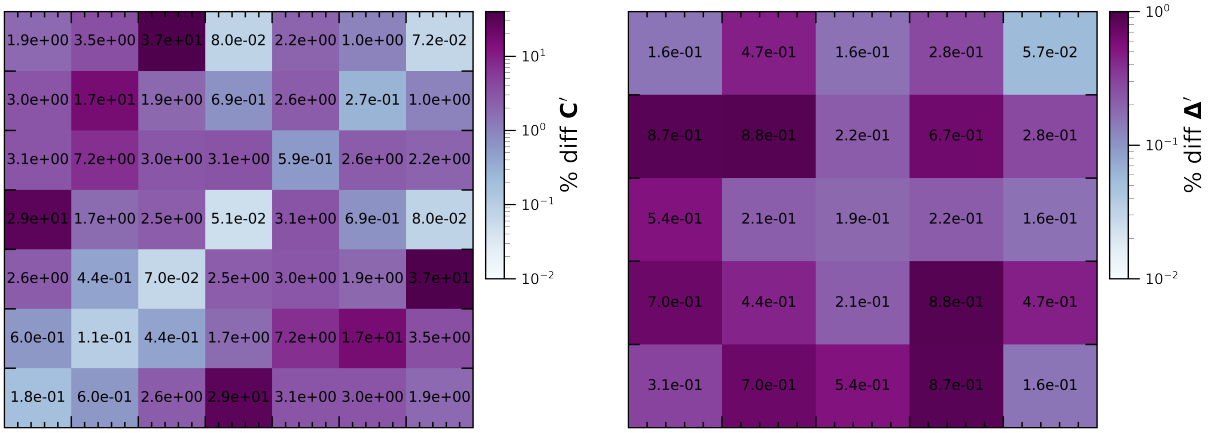


Figure 3. The MHMC algorithm is given the prior distributions as input along with a likelihood function and produces a Bayesian posterior of narrower width for all dimensions of z.

Here, choosing $\mathbf{\Lambda}/\lambda$ to produce posteriors $\mathbf{\Delta}'$ and $\langle\delta\rangle'$ results in the posterior PDFs of z being narrower than the prior PDFs. This posterior PDF width reduction is not as dramatic as when using conventional methods because defects are nonzero, as shown in the following sections.

The exact match of $\langle z\rangle'$ and $\langle\delta\rangle'$ for MHMC and analytical solutions is demonstrated in Fig. 2, but the question remains as to whether MHMC also closely approximates the analytical solutions for $\mathbf{C}'$ and $\mathbf{\Delta}'$. This comparison is shown in Fig. 4 in which the percent difference between the analytical and MHMC solutions is plotted.



(a) (b)

Figure 4. In 4a the difference in variance between analytical and MHMC $\mathbf{C}'$ is shown as never >0.6%; however, the off-diagonal covariances are significantly larger because they require many more iterations than 10^6 to converge to a higher accuracy. The differences in $\mathbf{\Delta}'$ shown in 4b are better behaved by comparison. Each plot gives confidence that the MHMC algorithm is approaching the analytical solution for a linear model.

The MHMC solution for $\mathbf{C}'$ can very closely (<0.6%) match the variances for elements of z, but predictions become worse for off-diagonal covariances, indicating that the algorithm will require many iterations (>10^6) to perfectly match off-diagonal covariances. The posterior $\mathbf{\Delta}'$ is easier to predict, and all elements are within 1% after 10^6 iterations.

To highlight the differences between this new method and the conventional one, the MHMC solution should also be compared with that of the conventional evaluation method described in the beginning of Section 2.. In that method, the assumption is essentially made that in the present notation, $\delta' = \lambda = 0$ and $\mathbf{\Delta}' = \mathbf{\Lambda}' = 0$, resulting in a likelihood function $\mathcal{L} = \delta_{Dirac}(T(P) - D)$. Figure 5 shows that the GLLS posteriors for all elements of z are much narrower than those of the MHMC.

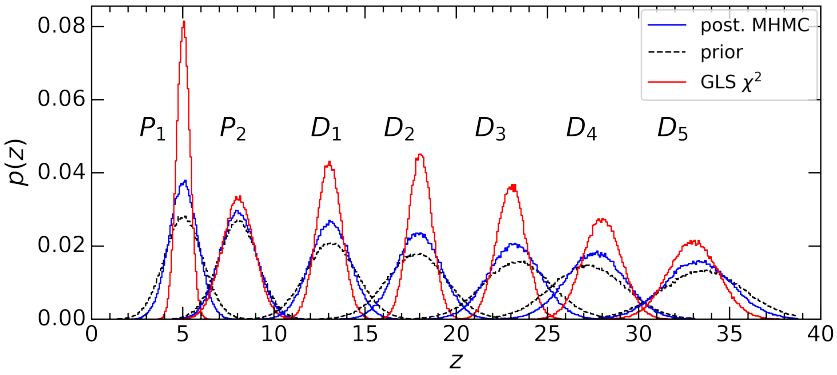


Figure 5. The GLLS predicted posterior values of z are compared with those of the MHMC and the prior distributions to show the effects of assumptions made in the GLLS method. These assumptions result in a smaller uncertainties in predicted values of z.

This result is expected because of the different likelihood functions used by the two methods. In this example, the MHMC framework was given a relatively conservative estimate of posterior $\mathbf{\Delta}' = \mathbf{\Delta}/2$ and $\delta' = \delta/2$, whereas the conventional GLLS method assumes that the model will perfectly match the data ($\mathbf{\Delta}' = \mathbb{0}$ and $\delta' = 0$). Solving for the linear GLLS posteriors uses the framework discussed by Fröhner in Ref. [1] and explicitly in Eqs. (62)–(68). Section 2.3 discusses how the MHMC evaluation process can be used to mimic this conventional evaluation.

3.2.2 Iterative Solution Converging to Conventional χ^2 Approach

Following the discussion in Section 2.3, an attempt can be made to sample a PDF with likelihood $\mathcal{L} = \delta_{Dirac}(T(P) - D)$ and the prior defined as before, limiting the distribution of data D to $D = T(P)$. This sampling is accomplished by iteratively solving Bayes' equation using MHMC in which the prior PDFs for

an iteration are defined by the posteriors of the previous one. More explicitly, such that for each iteration k,

$$\begin{aligned} \mathbf{C}_{k+1} &= \mathbf{C}'_k, \\ z_{k+1} &= z'_k, \\ \mathbf{\Lambda}_{k+1} &= \mathbf{\Delta}'_k. \end{aligned} \tag{91}$$

Using this scheme and the analytical solution described in Section 2.3, the condition $\mathbf{\Lambda} \to 0$ is approximated more closely as $\mathbf{\Delta}'$ is reduced by half upon each iteration of i. To match each iteration of the MHMC solution closely with the analytical solution, a relatively large number of Markov chain samples must be used (i.e., 10^6). The MHMC algorithm solves Bayes' equation iteratively until $\mathbf{C}'_{\chi^2}$ and $\mathbf{C}'_k$ converge. Convergence is defined as where the average percent difference of each respective element of the posterior covariance matrices is less than 1%. For this example case, convergence is achieved after around 12 iterations of MHMC solutions. Figure 6 shows a comparison of the final MHMC posterior PDFs for z with the original prior PDFs before the iterative solving scheme began.

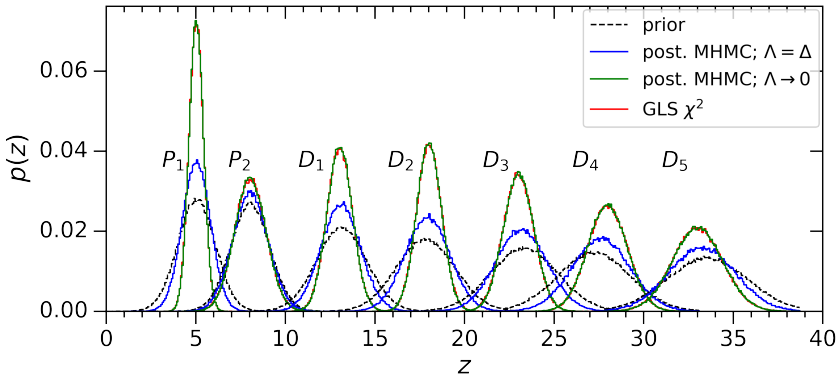


Figure 6. The general solution to Bayes' equation given by the normal MHMC solution (blue) more conservatively estimates posterior variance in z than the iterative MHMC solution (green), which approximates the conventional χ^2 minimization (red). This analysis demonstrates the effect of the conventional assumption that the likelihood function $\mathcal{L} = \delta_{Dirac}(T(P) - D)$.

The PDFs shown in Fig. 6 are the summation of all linear example cases discussed so that they can easily be compared simultaneously. The conventional solution often predicts much smaller variance in parameters when compared with the MHMC solution in which nonzero values of $\mathbf{\Delta}'$—in this case, $\mathbf{\Delta}' = \mathbf{\Delta}/2$—are accounted for. This phenomenon is expected because a less restrictive likelihood function is applied to the prior. As expected, the MHMC algorithm also (1) matches the analytical solutions for mean values $\langle z \rangle'$, $\langle \delta \rangle'$, $\mathbf{\Delta}'$, and $\mathbf{C}'$, and (2) can be used to estimate the conventional method (χ^2) by iteratively approaching a likelihood function that resembles a Dirac delta distribution. These linear solutions validate the algorithm's methodology and provide confidence that it can be applied to more realistic problems.

4. CONCLUSIONS

This report describes a framework for the Bayesian optimization of generalized data in the presence of imperfect data or models (i.e., prior PDF of generalized data). The effects of imperfections are parameterized by evaluator-defined constraints on the posterior expectation values of deviations between the data and model ($\langle\delta\rangle'$) and the corresponding posterior covariance matrix ($\mathbf{\Delta}'$). These constraints provide a method that is in complete harmony with Bayes' theorem, enabling evaluators to address issues such as inconsistent data or a missing/inaccurate prior generalized data covariance matrix. Skillfully applying these constraints should obviate the need to adjust (1) the inconsistent data before the evaluation and (2) the posterior (unreasonably small) covariances after the evaluation because neither one of these adjustments is in harmony with Bayes' theorem. These constraints uniquely define the likelihood function and consequently the posterior PDF, and setting $\langle\delta\rangle' = 0$ and $\mathbf{\Delta}' \to \mathbb{0}$ recovers the conventional Bayes' theorem. Furthermore, for linear models and normal PDFs, the formalism seamlessly recovers the extant methods used for conventional nuclear data evaluations (e.g., SAMMY or the TSURFER module of the SCALE code system).

In the case of linear models and normal PDFs, analytical solutions are derived for a posterior PDF of generalized data for arbitrary values of evaluator-specified constraints. These analytical solutions are appealing because their computational burden is comparable to that of the extant methods; therefore, it will facilitate computationally efficient revisiting of extant evaluations in the context of imperfections introduced by this formalism. The analytical convergence of these analytical solutions to the conventional methods in the limit of vanishing imperfections was confirmed numerically by a Bayesian MHMC method. Application of the Bayesian MHMC to the evaluation of resolved resonance cross-section data by a (nonlinear) *R*-matrix model will be presented in a forthcoming publication.

The Bayesian MHMC is a general and mathematically rigorous method for data evaluations of nonlinear models and non-normal PDFs that was extended in this work to address imperfections of data or models used for optimization. Bayesian MHMC would be particularly useful for testing the linear approximation and the assumption of normal PDFs, both of which are assumed for evaluations of differential cross-section data and integral benchmark experiments. For other applications, the Bayesian MHMC may be the only alternative when the sensitivity of a given response with respect to the model parameter may not be generally known, such as with thermal neutron scattering evaluations at the time of writing this report. Other applications of the Bayesian MHMC may span more than one energy region (e.g., thermal, resolved resonance, unresolved resonance) in a way that would yield a cross-covariance among the regions, which are ignored by the extant piecemeal evaluations. Various forms of the presented Bayesian optimization framework should enable improved nuclear data evaluations and advance the nuclear data evaluation methods of both differential cross sections and integral benchmark experiments.

5. REFERENCES

References

[1] F. Fröhner, "Evaluation and analysis of nuclear resonance data," OECD NEA, Paris, France, Tech. Rep. JEFF Report 18, 2000.

[2] N. Larson, "Updated users' guide for SAMMY: Multilevel r-matrix fits to neutron data using Bayes' equations," ORNL, Oak Ridge, TN, USA, Tech. Rep. ORNL/TM-9179/R8, 2008.

[3] W. A. Wieselquist and R. A. Lefebvre, "SCALE 6.3.0 User Manual," ORNL, Oak Ridge, TN, USA, Tech. Rep. ORNL/TM-SCALE-6.3.0, Dec. 2021.

[4] D. Brown *et al.*, "ENDF/B-VIII.0: The 8th major release of the nuclear reaction data library with CIELO-project cross sections, new standards and thermal scattering data," *Nuc. Dat. Sheets*, vol. 148, pp. 1 – 142, Feb. 2018. Accessed: Jul. 2, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0090375218300206

[5] T. Goorley, et. al., "Initial MCNP6 release overview," *Nuc. Technology*, vol. 180, no. 3, pp. 298–315, Dec 2012.

[6] G. Arbanas, J. Feng, Z. J. Clifton, A. M. Holcomb, M. T. Pigni, D. Wiarda, C. W. Chapman, V. Sobes, L. E. Liu, and Y. Danon, "Bayesian optimization of generalized data," *EPJ Nuclear Sciences and Technology*, vol. 4, p. 30, 2018. [Online]. Available: https://doi.org/10.1051/epjn/2018038

[7] J. M. Brown, G. Arbanas, D. Wiarda, and A. Holcomb, "Bayesian monte carlo evaluation framework for imperfect nuclear data," in *Transactions of the American Nuclear Society, Vol. 125, 2021 ANS Virtual Winter Meeting, December, 2021*, Washington, D.C., 2021.

[8] M. Pigni and H. Leeb, "Uncertainty estimates of evaluated ^{56}Fe cross sections based on extensive modelling at energies beyond 20 MeV," in *Proc. Int. Workshop on Nuclear Data for the Transmutation of Nuclear Waste. GSI-Darmstadt, Germany*. Citeseer, 2003.

[9] H. Leeb, D. Neudecker, and T. Srdinko, "Consistent procedure for nuclear data evaluation based on modeling," *Nuclear Data Sheets*, vol. 109, no. 12, pp. 2762–2767, 2008.

[10] D. Neudecker, R. Capote, and H. Leeb, "Impact of model defect and experimental uncertainties on evaluated output," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 723, pp. 163–172, 2013.

[11] S. Davis, "Exponential family models from bayes' theorem under expectation constraints," 2016. [Online]. Available: https://arxiv.org/pdf/1503.03451.pdf

[12] G. Schnabel and H. Leeb, "Differential Cross Sections and the Impact of Model Defects in Nuclear Data Evaluation," *EPJ Web of Conferences*, vol. 111, p. 09001, 2016. [Online]. Available: https://doi.org/10.1051/epjconf/201611109001

[13] G. King, A. Lovell, L. Neufcourt, and F. Nunes, "Direct comparison between bayesian and frequentist uncertainty quantification for nuclear reactions," *Physical Review Letters*, vol. 122, no. 23, p. 232502, 2019.

[14] Z. Zhao and F. Perey, "The covariance matrix of derived qauntities and their combination," ORNL, Oak Ridge, TN, USA, Tech. Rep. ORNL/TM-12106, 1992.

[15] T. Burr, T. Kawano, P. Talou, F. Pan, and N. Hengartner, "Defense of the Least Squares Solution to Peelle's Pertinent Puzzle," *Algorithms*, vol. 4, no. 1, pp. 28–39, 2011.

[16] S. D. Hill and J. C. Spall, "Stationarity and convergence of the metropolis-hastings algorithm: Insights into theoretical aspects," *IEEE Control Systems Magazine*, vol. 39, no. 1, pp. 56–67, 2019.

[17] J. C. Spall, "Estimation via markov chain monte carlo," in *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, vol. 4, 2002, pp. 2559–2564 vol.4.

[18] S. Chib and E. Greenberg, "Understanding the metropolis-hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.

6. ACKNOWLEDGEMENTS

This work was supported by the Nuclear Criticality Safety Program, funded and managed by the National Nuclear Security Administration for the Department of Energy.