

Bayesian optimization-assisted approximate Bayesian computation and its application to identifying cyclic constitutive law of structural steels

Bach Do^{a,*}, Makoto Ohsaki^a

^a*Department of Architecture and Architectural Engineering, Graduate School of Engineering, Kyoto University, Kyoto 615-8540, Japan*

Abstract

A costly finite element model discourages Bayesian inference of the underlying parameters from noise-corrupted experimental datasets. This arises because the likelihood of such a model is often computationally intractable, leading to the impossibility of performing a large number of costly simulations. This study presents Bayesian optimization-assisted approximate Bayesian computation (BO-assisted ABC) and showcases its application to identifying the approximate posteriors of parameters for known statistical models and of cyclic elastoplastic parameters for structural steels. ABC bypasses likelihood evaluations by generating prior samples that are assigned as samples constituting the posterior if discrepancies between the experimental dataset and the corresponding simulated datasets do not exceed a small, positive threshold. With a modest number of costly simulations, BO facilitates ABC by intelligently constructing a Gaussian process model that approximates the discrepancy mean function. Identification results from illustrative examples show that the approximate posteriors by the proposed approach not only reproduce a given posterior with acceptable accuracy but also capture the true nominal parameters of a statistical model. Moreover, the approximate posteriors of material parameters can simulate the cyclic elastoplastic behavior of a steel specimen under different loading conditions. The dependence of identification results on definitions of BO acquisition function is also investigated.

Keywords: Cyclic constitutive law, Bayesian inference, Approximate Bayesian computation, Bayesian optimization, Parameter identification, Structural steels

*Corresponding author

Email addresses: `se.do@archi.kyoto-u.ac.jp` (Bach Do), `ohsaki@archi.kyoto-u.ac.jp` (Makoto Ohsaki)

1. Introduction

We pursue an approach that facilitates Bayesian inference of parameters for costly finite element (FE) models from noise-corrupted experimental datasets. In particular, we focus on solving the following inverse problem: given noise-corrupted experimental datasets obtained from cyclic tests of a steel specimen, identify the distributions of underlying material parameters for an FE model that simulates the elastoplastic behavior of the specimen.

The literature categorizes methods for solving an inverse problem into deterministic and probabilistic approaches [1]. The deterministic approach states the inverse problem as an optimization problem that minimizes the discrepancy between simulated and experimental datasets, where the simulated dataset is a function of the parameters. Solving this optimization problem produces a best-fit vector of parameters. Although such a parameter vector is acceptable for a well-posed inverse problem, it is not very useful for an ill-posed inverse problem because many choices of parameter vector can be consistent with the noise-corrupted experimental dataset [2, 3]. In other words, the best-fit parameter vector is incapable of capturing uncertainty in the experimental dataset caused by measurement noise, thereby lowering the reliability of the calibrated FE model. Nevertheless, the deterministic approach is still a cornerstone of solving inverse problems in the field of structural engineering because its mathematical foundation is easy to understand.

A main concern of the deterministic approach is to handle the ill-posedness of inverse problems using regularization techniques that bias the model for reducing the variance of identified parameters. The monograph by Mueller and Siltanen [4] provides insights into common regularization techniques for linear and nonlinear inverse problems. Benning and Burger [5] surveyed modern regularization techniques and their applications to image reconstruction. Another concern of the deterministic approach is the difficulty in selecting a proper algorithm for solving the optimization problem based on characteristics of the inverse problem, such as the problem dimensionality or the computational cost of the associated simulation. Although gradient-based algorithms [3, 6–9] are suitable for solving high-dimensional problems, population-based algorithms [10], hybrid algorithms [7], or Bayesian optimization (BO) approaches [11–13] show their robustness in some particular problems.

The probabilistic approach [14–16] overcomes the shortcoming of the deterministic approach by formulating the inverse problem over the space of uncertain parameters. It encapsulates the observational noise in random distributions of the identified parameters. The probabilistic approach is often performed through Bayesian inference. This inference method updates a prior probability density function (PDF) based on a likelihood function to provide a posterior PDF describing the distribution of parameters conditioned on the experimental dataset. Here the prior represents our belief about the parameters before the

experimental dataset is observed, while the likelihood describes the chance of observing the experimental dataset given the parameter values. From the obtained posterior, we can generate the samples of parameters using a posterior sampling technique, such as the Markov chain Monte-Carlo (MCMC) method [17], sequential Monte-Carlo simulation (MCS) sampler [18], ensemble MCMC algorithm [19], or transitional MCMC method [20], and then compute the mean, variance, and maximum a posteriori (MAP) estimates of each parameter for decision-making or for uncertainty quantification of the FE model. Note that the deterministic approach is a special case of Bayesian inference that assigns a constant and the MAP estimate as the prior and the best-fit parameter vector, respectively [12]. This gives a criterion to reasonably assess the quality of posterior samples if the best-fit parameter vector is available.

Although considered as an ideal candidate to handle noise-corrupted experimental datasets, Bayesian inference is difficult to implement on the basis of a costly FE model simulating the experimental dataset. This arises from the following main computational issues. First, the likelihood associated with the costly FE model is often not in an analytical form or is too expensive to evaluate. Second, it is impractical to perform a large number (say, many thousands or millions [21]) of costly simulations for the posterior sampling technique. In fact, parameter identification via Bayesian inference in the field of structural engineering still limits itself to problems with uni-axial constitutive laws and static loading conditions [22]. This inference method also seems unfamiliar to practicing structural engineers because it requires sophisticated probabilistic computations [12]. Recently, open-source codes have been released to facilitate solutions to various one-dimensional Bayesian inverse problems in computational mechanics [23].

To address the aforementioned computational issues, recent studies performed MCMC or MCS based upon an approximate likelihood [24–26]. This approach requires sufficient accuracy of the approximate likelihood for obtaining reliable posterior samples. However, a surrogate model, serving as the approximate likelihood, constructed from either the best goodness-of-fit coefficient of a test dataset [24] or a one-shot space-filling experimental design [25] may not be accurate enough to handle the trade-off between the number of costly training samples and the reliability of MCMC samples. This is due to the fact that such a surrogate model, which equally treats the parameter samples, does not incorporate the information from each sample into improving the accuracy of approximate likelihood, while it is totally free to improve the reliability of MCMC samples by ignoring the parameter region that is unlikely to be consistent with the experimental dataset. The question is, therefore, how to distinguish the parameter regions from which the approximate likelihood should be carefully constructed from the regions that should be ignored to reduce the number of costly simulations.

In this work, we present a posterior approximation approach to Bayesian inference of parameters for

costly FE models and showcase its application to identifying the approximate posteriors of cyclic elastoplastic material parameters for structural steels under observational noise. It complements our recent works [11, 12] of single- and multi-objective BO for deterministic identification of cyclic elastoplastic parameters. The abovementioned computational issues are addressed by combining the approximate Bayesian computation (ABC) [21, 27–29] with BO [30–32], leading to the BO-assisted ABC approach. In fact, ABC has been used in the field of model updating for estimating the parameters of structural dynamic models [33] and for calibrating uncertainty models from the NASA-Langley Uncertainty Quantification Challenge [34]. However, since these models are explicitly given as functions of the parameters, it is computationally inexpensive to generate a large set of parameter vectors and the corresponding simulated data for implementation of ABC. This large set in turn facilitates use of ABC as it can increase the number of parameter vectors to be accepted as posterior samples. In contrast, the focus of this paper is to infer the distributions of parameters for costly FE models that may have pronounced effect on the performance of ABC.

The present BO-assisted ABC approach is an extension of two recent works [35, 36] in the field of statistical modeling to the identification of parameters for costly FE models. In particular, ABC handles the computationally intractable likelihood using samples generated from the FE model. It draws parameter samples from the prior, evaluates the associated datasets simulated using the FE model, and accepts the samples to be included into the posterior if the discrepancies between the simulated and experimental datasets do not exceed a small, positive threshold. BO further facilitates ABC by intelligently constructing a Gaussian process (GP) model that approximates the discrepancy function. Guided by BO, the GP model improves its accuracy in the region where the parameter vectors are likely to be consistent with the experimental dataset or where the parameter vectors contribute to reducing the variance of ABC posterior approximation, thereby requiring a modest number of costly simulations.

The BO-assisted ABC algorithm extensively tests its performance against the benchmark analyses of two datasets generated from a Gaussian model and a g -and- k distribution, followed by the identification of approximate posteriors of material parameters that model the elastoplastic behavior of a steel specimen subjected to cyclic loading conditions. The approximate posteriors obtained from different definitions of BO acquisition function and different discrepancy threshold levels are also investigated.

The content of this paper is from Chapter 4 of the dissertation by the first author [37]. The remainder of this paper progresses as follows. Section 2 formulates the inverse identification problem of interest. Section 3 describes the ABC methodology, followed by Section 4 detailing the BO-assisted ABC approach. Section 5 illustrates the application of the BO-assisted ABC approach to identifying the underlying parameters for two benchmark data analyses and for a steel specimen subjected to cyclic loading. Finally, Section 6 concludes

this paper and provides possible research directions to extend the proposed approach to unresolved problems. Appendices A and B briefly describe the nonlinear combined isotropic/kinematic hardening model and the GP modeling, respectively.

In the following, we use $[\cdot]^\top$ and $[\cdot]^{-1}$ to denote the transpose and the inverse of matrix $[\cdot]$, respectively. All vectors are column vectors.

2. Identification problem

Consider an inverse problem to identify the vector of n_p parameters $\mathbf{x} \in \mathbb{R}^{n_p}$ incorporated into an FE model that simulates the cyclic elastoplastic behavior of a steel specimen. Let $\mathbf{d}_0 \in \mathbb{R}^{n_d}$ denote the vector of n_d response values measured from cyclic tests of the steel specimen, and $\mathbf{d}(\mathbf{x}) \in \mathbb{R}^{n_d}$ denote the corresponding vector simulated from the FE model that inputs $\mathbf{x}$ as material parameters. To solve the inverse problem, $\mathbf{d}_0$ and $\mathbf{d}(\mathbf{x})$ together formulate the discrepancy between them denoted as $\delta(\mathbf{x})$, which reads

$$\delta(\mathbf{x}) = \|\mathbf{d}(\mathbf{x}) - \mathbf{d}_0\| + e, \quad (1)$$

where $\|\cdot\| : \mathbb{R}^{n_d} \rightarrow \mathbb{R}_0^+$ denotes a distance measure (e.g., Euclidean or Mahalanobis distance), and e is the discrepancy noise describing random corruptions of the experimental dataset $\mathbf{d}_0$ and is assumed to be a Gaussian with zero mean and a specified standard deviation. If uncertainty due to idealization and discretization errors is not considered, the PDF of δ for a given $\mathbf{x}$ is

$$\pi(\delta|\mathbf{x}) = \mathcal{N}(\delta|y(\mathbf{x}), \tau^2), \quad (2)$$

where “ $\delta|\mathbf{x}$ ” is read as “ δ given (conditioned on) $\mathbf{x}$ ”, $\mathcal{N}$ denotes a Gaussian, $y(\mathbf{x}) = \|\mathbf{d}(\mathbf{x}) - \mathbf{d}_0\|$ is the mean function of δ , and τ is the standard deviation in the space of δ .

The deterministic approach minimizes $y(\mathbf{x})$ for finding a best-fit parameter vector. As previously mentioned, the best-fit parameter vector is not very useful in the presence of observational noise because many parameter vectors within the noise, especially for an ill-posed inverse problem, may be capable of capturing the experimental dataset [3]. In this circumstance, it is preferable to adopt the probabilistic approach for solving the inverse problem, by which observational noise is transformed into uncertainty in the identified parameters.

Through Bayesian inference, the probabilistic approach infers the distribution of $\mathbf{x}$, given the experimental dataset, by two steps. The first step expresses our prior belief about $\mathbf{x}$ through the prior PDF $\pi(\mathbf{x})$. The second step updates $\pi(\mathbf{x})$ based on the likelihood function $p(\mathbf{d}_0|\mathbf{x})$ of the FE model, resulting in a posterior PDF $\pi(\mathbf{x}|\mathbf{d}_0)$. In essence, if the likelihood function $p(\mathbf{d}_0|\mathbf{x})$ is available in a closed form, $\pi(\mathbf{x}|\mathbf{d}_0)$

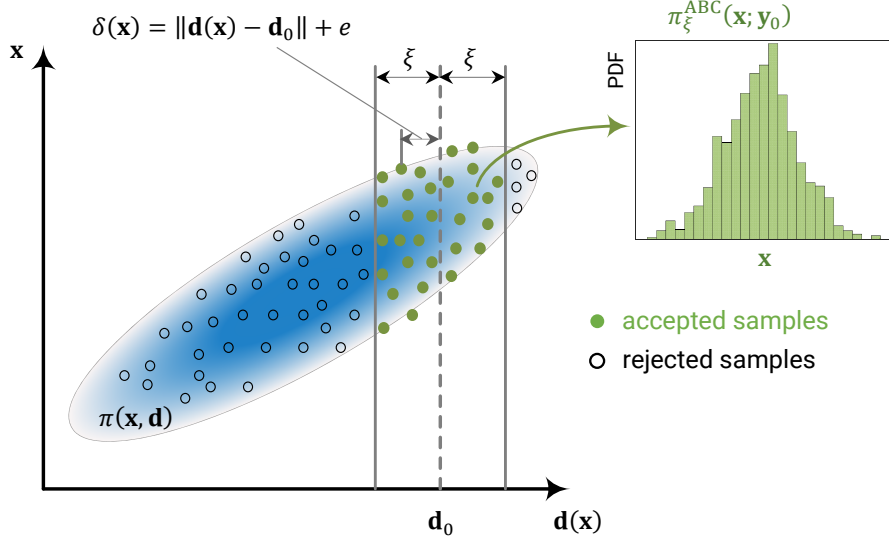


Figure 1: Schematic description of ABC (adapted from Beaumont [38]).

is straightforwardly computed from Bayes' theorem as

$$\pi(\mathbf{x}|\mathbf{d}_0) = \frac{\pi(\mathbf{x})p(\mathbf{d}_0|\mathbf{x})}{\pi(\mathbf{d}_0)} = \frac{\pi(\mathbf{x})p(\mathbf{d}_0|\mathbf{x})}{\int \pi(\mathbf{x})p(\mathbf{d}_0|\mathbf{x}) d\mathbf{x}} \propto \pi(\mathbf{x})p(\mathbf{d}_0|\mathbf{x}), \quad (3)$$

where the numerator describes the (unnormalized) joint PDF of $\mathbf{x}$ and $\mathbf{d}_0$, and the denominator $\pi(\mathbf{d}_0)$ (or evidence) is a normalization constant to ensure that $\pi(\mathbf{x}|\mathbf{d}_0)$ is a PDF. As a result, $\pi(\mathbf{x}|\mathbf{d}_0)$ is proportional to $\pi(\mathbf{x})p(\mathbf{d}_0|\mathbf{x})$.

Since it is difficult to compute $\pi(\mathbf{d}_0)$ using the integral in Eq. (3), Bayesian inference often approximates the posterior by generating a large number, say thousands or millions [21], of parameter samples using a posterior sampling technique, thereby requiring a large number of $p(\mathbf{d}_0|\mathbf{x})$ evaluations. In this study, however, it is computationally intractable to simply implement such a posterior sampling technique because the evaluation of likelihood function $p(\mathbf{d}_0|\mathbf{x})$ via modeling the cyclic elastoplastic behavior of structural steels demands a substantial computational cost. Consequently, it is impractical to use the posterior defined in Eq. (3) directly as the target distribution of a posterior sampling technique.

3. Approximate Bayesian computation

ABC appears as one of the posterior approximation methods [21, 27–29, 38, 39]. It bypasses the evaluations of likelihood $p(\mathbf{d}_0|\mathbf{x})$ using samples of simulated datasets generated by the FE model. Figure 1 shows a schematic description of ABC. It draws a parameter sample $\mathbf{x}$ from the prior $\pi(\mathbf{x})$ and evaluates $\mathbf{d}(\mathbf{x})$ using the FE model, which is equivalent to drawing sample $(\mathbf{x}, \mathbf{d})$ from the joint PDF $\pi(\mathbf{x}, \mathbf{d})$. Then, it computes $y(\mathbf{x}) = \|\mathbf{d}(\mathbf{x}) - \mathbf{d}_0\|$, generates δ from the distribution in Eq. (2), and accepts $\mathbf{x}$ as a sample constituting

the posterior PDF in Eq. (3) if δ is not greater than a small threshold value $\xi > 0$. Otherwise, reject $\mathbf{x}$. In fact, the mechanism of ABC is governed by the generation of $(\mathbf{x}, \delta)$ from the following joint PDF:

$$\pi_{\xi}^{\text{ABC}}(\mathbf{x}, \delta | \mathbf{d}_0) \propto \pi(\mathbf{x})\pi(\delta | \mathbf{x})K_{\xi}(\delta), \quad (4)$$

where $K_{\xi}(\delta)$ is a kernel function that indicates whether a parameter sample from the prior is accepted as a posterior sample or not. Most implementations of ABC algorithms use $K_{\xi}(\delta) = \mathbb{I}[\delta \leq \xi]$, where the binary indicator function $\mathbb{I}[\cdot]$ returns 1 if $[\cdot]$ is true and 0 otherwise. As a result, the ABC posterior is derived by integrating the right-hand side of Eq. (4) with respect to δ , as

$$\pi_{\xi}^{\text{ABC}}(\mathbf{x} | \mathbf{d}_0) \propto \pi(\mathbf{x}) \int \pi(\delta | \mathbf{x})K_{\xi}(\delta) d\delta. \quad (5)$$

Here $K_{\xi}(\delta)$ can be interpreted as a weight assigned to each parameter sample. A larger weight corresponds to the parameter sample with smaller δ , which is more obvious when a smooth kernel function (e.g., Epanechnikov or Gaussian) is used.

The nature of ABC implies that a high dimensional dataset $\mathbf{d}_0$ and a small threshold value ξ could result in a small probability of accepting parameter samples as posterior samples at which the associated vectors of simulated responses are close to $\mathbf{d}_0$ [21]. This circumstance may require a large value of ξ to increase the acceptance probability. A relatively large ξ , ironically, leads to a poor approximation of the posterior because the simulated response vectors corresponding to the accepted parameter samples are not actually close to $\mathbf{d}_0$. To deal with this issue, let $\mathbf{z}(\mathbf{x}) = \eta(\mathbf{d}(\mathbf{x})) : \mathbb{R}^{n_d} \rightarrow \mathbb{R}^{n_f}$ be a mapping (or summary statistics [21]) to reduce the dimension of $\mathbf{d}$, where $n_f \ll n_d$. Here $\mathbf{z}(\mathbf{x})$ can be referred to as a feature vector of the dataset $\mathbf{d}(\mathbf{x})$. For example, a vector $\mathbf{z}(\mathbf{x})$ including the mean and standard deviation values can represent a huge number of samples generated from a Gaussian. By choosing an appropriate mapping $\eta(\cdot)$, the ABC posterior in Eq. (5) can be approximated by [21]

$$\pi_{\xi}^{\text{ABC}}(\mathbf{x} | \mathbf{d}_0) \approx \pi_{\zeta}^{\text{ABC}}(\mathbf{x} | \mathbf{z}_0) \propto \pi(\mathbf{x}) \int \pi(\Delta | \mathbf{x})K_{\zeta}(\Delta) d\Delta, \quad (6)$$

where $\mathbf{z}_0 = \eta(\mathbf{d}_0)$, $\Delta \sim \pi(\Delta | \mathbf{x}) = \mathcal{N}(\Delta | f(\mathbf{x}), \omega^2)$ is the discrepancy between $\mathbf{z}(\mathbf{x})$ and $\mathbf{z}_0$ with $f(\mathbf{x}) = \|\mathbf{z}(\mathbf{x}) - \mathbf{z}_0\|$ and the standard deviation ω in the space of Δ , and $K_{\zeta}(\Delta) = \mathbb{I}[\Delta \leq \zeta]$ with a small threshold value $\zeta > 0$.

By further generating T samples of Δ , namely $\{\Delta^1, \dots, \Delta^T\}$, from the PDF $\pi(\Delta | \mathbf{x}) = \mathcal{N}(\Delta | f(\mathbf{x}), \omega^2)$, the ABC posterior can be estimated by replacing the integral in Eq. (6) with the following Monte-Carlo sum:

$$\pi_{\zeta}^{\text{ABC}}(\mathbf{x} | \mathbf{z}_0) \propto \frac{\pi(\mathbf{x})}{T} \sum_{t=1}^T \mathbb{I}[\Delta^t \leq \zeta] \approx \pi(\mathbf{x}) \mathbb{P}[\Delta(\mathbf{x}) \leq \zeta], \quad (7)$$

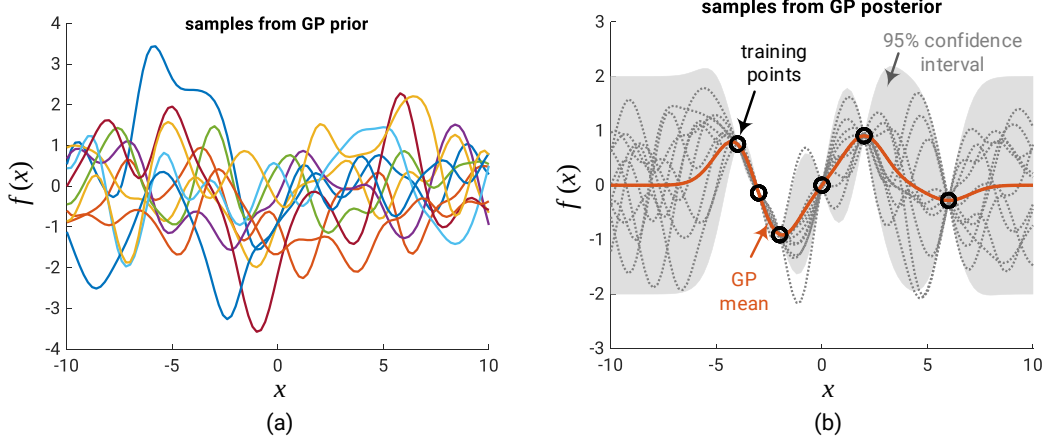


Figure 2: Examples of functions sampled from (a) the GP prior and (b) the GP posterior constructed from six training points.

where $\mathbb{P}[\cdot]$ denotes the probability of occurring $[\cdot]$, and $\mathbb{P}[\Delta(\mathbf{x}) \leq \zeta]$ is similar to the approximate likelihood. Applying the Markov’s inequality to $\mathbb{P}[\Delta(\mathbf{x}) \leq \zeta] = 1 - \mathbb{P}[\Delta(\mathbf{x}) \geq \zeta]$ gives [35]

$$\pi_{\zeta}^{\text{ABC}}(\mathbf{x}|\mathbf{z}_0) \propto \pi(\mathbf{x})\mathbb{P}[\Delta(\mathbf{x}) \leq \zeta] \geq \pi(\mathbf{x}) \left[1 - \frac{\mathbb{E}[\Delta(\mathbf{x})]}{\zeta} \right], \quad (8)$$

where “ $\propto$ ” and $\mathbb{E}[\cdot]$ denote “approximately proportional” and the mean of $[\cdot]$, respectively; and $\mathbb{E}[\Delta(\mathbf{x})] = f(\mathbf{x})$.

Since $\pi_{\zeta}^{\text{ABC}}(\mathbf{x}|\mathbf{z}_0) \propto \pi(\mathbf{x})\mathbb{P}[\Delta(\mathbf{x}) \leq \zeta]$, a large error in estimation of $\mathbb{P}[\Delta(\mathbf{x}) \leq \zeta]$ can lead to a substantial error in the posterior estimate. The statistical modeling community reduces this error by adopting two main approaches. The first approach explores an important region of the parameter space where the probability of accepting a parameter sample $\mathbf{x}$ is high [35]. This is equivalent to maximizing the lower bound of $\mathbb{P}[\Delta(\mathbf{x}) \leq \zeta]$ or minimizing $\mathbb{E}[\Delta(\mathbf{x})] = f(\mathbf{x})$ in Eq. (8). The second approach sequentially reduces the variance of $\pi_{\zeta}^{\text{ABC}}(\mathbf{x}|\mathbf{z}_0)$ [36]. Each of these approaches can be further guided by solving an optimization problem with a costly objective function and therefore can be performed by BO. We detail these approaches in the next section.

4. Bayesian optimization-assisted approximate Bayesian computation

4.1. Gaussian process as surrogate for the discrepancy mean function

To develop the BO-assisted ABC approach, we first approximate the mean $f(\mathbf{x})$ of the discrepancy $\Delta(\mathbf{x})$ using a GP model. We build this GP model from a training dataset $\mathcal{D} = \{\mathbf{x}^n, f^n\}_{n=1}^N$, where $\mathbf{x}^n$ is a parameter sample, $f^n = \|\mathbf{z}(\mathbf{x}^n) - \mathbf{z}_0\|$, and N is the number of initial training samples, which is problem-dependent [11]. The sample $\mathbf{x}^n$ is randomly generated from the prior $\pi(\mathbf{x})$. It may also be better to generate

$\mathbf{x}^n$ using a space-filling Latin-hypercube sampling method [40] that can provide a good training dataset with a uniform spread of samples. The value f^n is calculated through $\mathbf{d}(\mathbf{x}^n)$, or equivalently $\mathbf{z}(\mathbf{x}^n)$, by performing the costly FE analysis for the input vector $\mathbf{x}^n$.

The GP model describes the relationship between f and $\mathbf{x}$ using a conditional Gaussian $\hat{f}(\mathbf{x})$. The mathematical foundation of $\hat{f}(\mathbf{x})$ is detailed in Appendix B. Following Eq. (B.6), the GP prediction of $f(\mathbf{x})$ reads

$$\hat{f}(\mathbf{x}) \sim \mathcal{N}(f|\mu_f(\mathbf{x}), \tau_f^2(\mathbf{x})), \quad (9)$$

where $\mu_f(\mathbf{x})$ and $\tau_f(\mathbf{x})$ are the mean and standard deviation of $\hat{f}(\mathbf{x})$, respectively.

Figure 2 illustrates the underlying foundation of GP modeling as a Bayesian regression method for prediction of a univariate function $f(x)$. The GP prior defines a spectrum of functions and then converts this spectrum into a conditional spectrum called the GP posterior when some data points are observed. Figure 2(a) shows ten functions sampled from the GP prior. Figure 2(b) shows ten functions sampled from the GP posterior constructed from six training points, the associated GP posterior mean function μ_f , and the 95% confidence interval derived from the GP posterior mean and standard deviation in Eq. (9). The GP posterior is utilized as the predictive model for $f(x)$.

4.2. BO-assisted ABC

The costly simulation of cyclic elastoplastic behavior of structural steels motivates use of BO to intelligently guide the improvement of ABC posterior approximation. BO is a sequential global optimization method well suited for solving an optimization problem with an expensive-to-evaluate objective function [30–32]. Suppose that BO has completed the s th iteration at which a solution to the optimization problem has been found among the samples of the training dataset. By maximizing an acquisition function formulated from the current GP model for the objective function, BO strategically selects a new, good design vector (i.e., parameter vector) for updating the current solution in the $(s + 1)$ th iteration. Since the formulation and maximization of the acquisition function do not require any costly simulation, the number of costly simulations can be reduced considerably. In particular, BO determines the new parameter vector $\mathbf{x}^{s+1}$ by solving

$$\mathbf{x}^{s+1} = \operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} \alpha(\mathbf{x}), \quad (10)$$

where $\alpha(\mathbf{x})$ and $\mathcal{X}$ are the acquisition function and the feasible space of parameters, respectively.

After $\mathbf{x}^{s+1}$ is determined, we evaluate the corresponding objective function by carrying out the FE analysis for the input vector $\mathbf{x}^{s+1}$. Then, we add $\mathbf{x}^{s+1}$ and the corresponding objective function to the current training dataset for updating the solution as well as the current GP model.

The standard BO formulates $\alpha(\mathbf{x})$ based on the following two main criteria: (1) improving the current best-found solution (i.e., exploitation) and (2) reducing the uncertainty in GP predictions (i.e., exploration). The first criterion exploits the mean function of GP prediction, while the second criterion explores the GP standard deviation. In addition to these criteria, this study formulates $\alpha(\mathbf{x})$ based on how the error in approximation of the ABC posterior can be reduced.

The first approach to reducing the error in ABC posterior approximation refines $\hat{f}(\mathbf{x})$ in the parameter region that minimizes the discrepancy mean function $f(\mathbf{x})$ for an increase in the probability of accepting parameter samples generated from the prior [35]. This approach is similar to the exploitation of BO when the objective function is conditioned on the associated GP model. The acquisition functions for minimizing $f(\mathbf{x})$ using the standard BO have been well developed [30–32].

Let $f_{\min}$ denote the current best value of f associated with the best-found solution, and $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function (CDF) and the PDF of the standard Gaussian, respectively. The first acquisition function for minimizing $f(\mathbf{x})$ is the probability of improvement (PI), which is [41]

$$\text{(PI)} \quad \alpha(\mathbf{x}) = \mathbb{P} \left[f(\mathbf{x}) \leq f_{\min} | \hat{f}(\mathbf{x}) \right] = \Phi \left(\frac{f_{\min} - \mu_f(\mathbf{x})}{\tau_f(\mathbf{x})} \right). \quad (11)$$

By maximizing PI, we maximize the chance that the discrepancy mean function at $\mathbf{x}^{s+1}$ is smaller (better) than the best value $f_{\min}$ observed so far.

The second acquisition function for minimizing $f(\mathbf{x})$ is the expected improvement (EI), which is [30]

$$\text{(EI)} \quad \alpha(\mathbf{x}) = \mathbb{E} \left[\max(0, f_{\min} - f(\mathbf{x})) | \hat{f}(\mathbf{x}) \right] = (f_{\min} - \mu_f(\mathbf{x})) \Phi \left(\frac{f_{\min} - \mu_f(\mathbf{x})}{\tau_f(\mathbf{x})} \right) + \tau_f(\mathbf{x}) \phi \left(\frac{f_{\min} - \mu_f(\mathbf{x})}{\tau_f(\mathbf{x})} \right). \quad (12)$$

New parameter vector $\mathbf{x}^{s+1}$ obtained by maximizing EI is to maximize the expected improvement over the best-found solution $f_{\min}$. Here EI perfectly balances exploitation (i.e., the first term) and exploration (i.e., the second term).

The third acquisition function for minimizing $f(\mathbf{x})$ is the negative lower confidence bound (LCB), which reads [42]

$$\text{(LCB)} \quad \alpha(\mathbf{x}) = -[\mu_f(\mathbf{x}) - \beta \tau_f(\mathbf{x})], \quad (13)$$

where β is a tuning parameter assigned by the user to control the trade-off between exploitation (i.e., term $-\mu_f(\mathbf{x})$) and exploration (i.e., term $\beta \tau_f(\mathbf{x})$). $\beta = 2$ is used in this study. By maximizing LCB, we find $\mathbf{x}^{s+1}$ so that it improves the lower bound of $f(\mathbf{x})$ derived from the current GP model as much as possible.

The second approach to reducing the error in ABC posterior approximation refines $\hat{f}(\mathbf{x})$ in the parameter region to reduce the variance of $\pi_{\zeta}^{\text{ABC}}(\mathbf{x}|\mathbf{z}_0)$ estimate [36]. To formulate an acquisition function for this approach, we evaluate the mean and variance of $\pi_{\zeta}^{\text{ABC}}(\mathbf{x}|\mathbf{z}_0)$ stated in Eq. (7) when conditioned on the GP

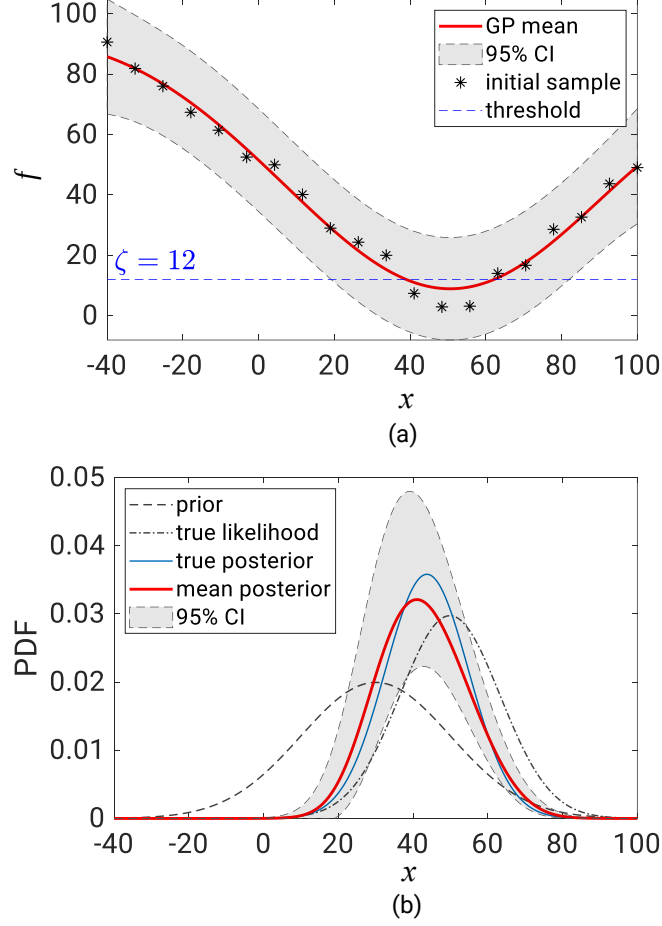


Figure 3: Schematic illustration of calculating the approximate posterior. (a) GP approximation for the discrepancy mean, and (b) the mean posterior calculated from Eq. (14) and uncertainty in the posterior estimate (i.e., shaded area of 95% confidence interval) incorporated in Eq. (15).

model $\hat{f}(\mathbf{x})$, such that

$$\mathbb{E} \left[\pi_{\zeta}^{\text{ABC}}(\mathbf{x}|\mathbf{z}_0) | \hat{f}(\mathbf{x}) \right] \propto \pi(\mathbf{x}) \Phi \left(\frac{\zeta - \mu_f(\mathbf{x})}{\sqrt{\omega^2 + \tau_f^2(\mathbf{x})}} \right), \quad (14)$$

$$\mathbb{V} \left[\pi_{\zeta}^{\text{ABC}}(\mathbf{x}|\mathbf{z}_0) | \hat{f}(\mathbf{x}) \right] \propto \pi^2(\mathbf{x}) \left[\Phi \left(\frac{\zeta - \mu_f(\mathbf{x})}{\sqrt{\omega^2 + \tau_f^2(\mathbf{x})}} \right) \Phi \left(\frac{\mu_f(\mathbf{x}) - \zeta}{\sqrt{\omega^2 + \tau_f^2(\mathbf{x})}} \right) - 2\mathcal{T} \left(\frac{\zeta - \mu_f(\mathbf{x})}{\sqrt{\omega^2 + \tau_f^2(\mathbf{x})}}, \frac{\omega}{\sqrt{\omega^2 + 2\tau_f^2(\mathbf{x})}} \right) \right], \quad (15)$$

where $\mathbb{V}[\cdot]$ denotes the variance of $[\cdot]$ and $\mathcal{T}(\cdot, \cdot)$ is the Owen's T-function, which reads

$$\mathcal{T}(t_1, t_2) = \frac{1}{2\pi} \int_0^{t_2} \frac{e^{-0.5t_1^2(1+t^2)}}{1+t^2} dt. \quad (16)$$

Note that Eq. (14) is derived from the assumption that $\pi(\Delta|\mathbf{x}) = \mathcal{N}(\Delta|f(\mathbf{x}), \omega^2)$, while the derivation of Eq. (15) can be found in the supplementary material of the work by Järvenpää et al. [36]. The Owen's T-function in Eq. (16) is numerically calculated [43].

To reduce the error in $\pi_{\zeta}^{\text{ABC}}(\mathbf{x}|\mathbf{z}_0)$ estimate, we wish to generate new parameter samples in the region that has a large value of $\mathbb{V}[\pi_{\zeta}^{\text{ABC}}(\mathbf{x}|\mathbf{z}_0)]$. This is similar to the exploration of BO when the objective function is conditioned on the associated GP model. As a result, we use the following acquisition function, called VAR, to determine the new parameter vector $\mathbf{x}^{s+1}$ for the second approach:

$$\text{(VAR)} \quad \alpha(\mathbf{x}) = \pi^2(\mathbf{x}) \left[\Phi \left(\frac{\zeta - \mu_f(\mathbf{x})}{\sqrt{\omega^2 + \tau_f^2(\mathbf{x})}} \right) \Phi \left(\frac{\mu_f(\mathbf{x}) - \zeta}{\sqrt{\omega^2 + \tau_f^2(\mathbf{x})}} \right) - 2\mathcal{T} \left(\frac{\zeta - \mu_f(\mathbf{x})}{\sqrt{\omega^2 + \tau_f^2(\mathbf{x})}}, \frac{\omega}{\sqrt{\omega^2 + 2\tau_f^2(\mathbf{x})}} \right) \right]. \quad (17)$$

Figure 3 schematically shows how to calculate the approximate posterior as well as how to improve the approximate posterior using the aforementioned two refinement approaches. Since the discrepancy mean function $f(\mathbf{x})$ is represented by the GP model $\hat{f}(\mathbf{x})$ depicted in Fig. 3(a), the estimation of ABC posterior in Fig. 3(b) has two sorts of information, namely the mean posterior (i.e., the red solid curve) and uncertainty in the posterior estimate (i.e., the shaded area). By selecting the new parameter vector that minimizes $f(\mathbf{x})$ in the first approach, we are likely to increase the approximation accuracy in the parameter region that corresponds to large posterior values. By sequentially selecting the new parameter vector that maximizes VAR in the second approach, we incrementally reduce the shaded area in Fig. 3(b).

To this end, Algorithm 1 summarizes the identification procedure utilizing the BO-assisted ABC approach. In step 1, $s_{\max}$ and N_s are the pre-specified limit of the number of BO iterations and the number of approximate posterior samples to be generated, respectively. Specifying ω is optional because a reasonable value of ω can be determined as a hyperparameter in each BO iteration by maximizing the approximate likelihood of $\pi(\Delta|\mathbf{x})$ derived from the GP model $\hat{f}(\mathbf{x})$; see Appendix B. The BO nested in the algorithm (steps 11-22) terminates if either the number of its iterations reaches $s_{\max}$ or $\mathbf{x}^{s+1}$ is identical to any member of $\mathcal{D}$, i.e., the condition in step 14 is not satisfied. Step 13 requires an optimization algorithm for solving problem (10). For this task, any conventional optimization algorithm can be used because the analytical forms of different definitions of $\alpha(\mathbf{x})$ are given in Eqs. (11), (12), (13), and (17).

After the refinement process, the GP from step 12 and the prior $\pi(\mathbf{x})$ are incorporated into Eq. (14) and then assigned to step 23 as input to an adaptive MCMC for generating a total of N_s parameter samples, which are the representatives of posterior samples. Details of the adaptive MCMC can be found in Refs. [44, 45]. The MATLAB toolbox for the adaptive MCMC is available at <https://github.com/mjlaine/mcmcstat.git> (accessed on May 24, 2023). Note that the initial parameter vector for starting the MCMC in step 23 takes the value of new parameter vector obtained from the very last BO iteration of step 13.

It is also desirable to compare the quality of ABC samples from different posterior sampling techniques such as the adaptive MCMC [44], sequential MCS [18], ensemble MCMC [19], and transitional MCMC [20]

Algorithm 1 BO-assisted ABC

```
1: Specify  $\mathbf{d}_0, \pi(\cdot), \eta(\cdot), \zeta, s_{\max}, N, N_s, \omega$  (optional);
2:  $\mathbf{z}_0 \leftarrow \eta(\mathbf{d}_0)$ ;
3:  $\mathcal{D} \leftarrow \emptyset$ ;
4: Generate  $N$  samples of parameter vector  $\mathbf{x}^n \sim \pi(\mathbf{x})$ ;
5: for  $n = 1 : N$  do
6:   Evaluate  $\mathbf{d}(\mathbf{x}^n)$  using the FE model; ▷ Costly step
7:    $\mathbf{z}(\mathbf{x}^n) \leftarrow \eta(\mathbf{d}(\mathbf{x}^n))$ ;
8:    $f^n \leftarrow \|\mathbf{z}(\mathbf{x}^n) - \mathbf{z}_0\|$ ;
9:    $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{x}^n, f^n\}$ ;
10: end for
11: for  $s = 0 : s_{\max}$  do
12:   Construct  $\hat{f}(\mathbf{x})$  in Eq. (9) from  $\mathcal{D}$ ;
13:   Find  $\mathbf{x}^{s+1}$  by solving problem (10);
14:   if  $\mathbf{x}^{s+1} \notin \mathcal{D}$  then
15:     Evaluate  $\mathbf{d}(\mathbf{x}^{s+1})$  using the FE model; ▷ Costly step
16:      $\mathbf{z}(\mathbf{x}^{s+1}) \leftarrow \eta(\mathbf{d}(\mathbf{x}^{s+1}))$ ;
17:      $f^{s+1} \leftarrow \|\mathbf{z}(\mathbf{x}^{s+1}) - \mathbf{z}_0\|$ ;
18:      $\mathcal{D} \leftarrow \mathcal{D} \cup \{\mathbf{x}^{s+1}, f^{s+1}\}$ ;
19:   else
20:     exit for
21:   end if
22: end for
23: Generate  $N_s$  parameter samples from Eq. (14) using the adaptive MCMC [44, 45];
24: return  $N_s$  samples of parameters from the approximate posterior.
```

after the refinement process is done. Such a comparison, however, belongs to another research topic that is beyond the scope of this study. Readers may consult with Sisson and Fan [46] for a survey of various sampling techniques for ABC. A useful review of advanced MCS methods for solving Bayesian model updating problems of engineering applications can be found in Ref. [47].

Table 1: Parameters for genetic algorithm.

Parameter	Value
Population size	4000
Maximum number of generations	50
Crossover fraction	65%
Elite transfer	2
Fitness function tolerance	10^{-6}

5. Illustrative examples

In this section, we extensively test the performance of Algorithm 1 against the benchmark analyses of two datasets generated from a bivariate Gaussian model [36] (Section 5.1) and from a so-called g -and- k distribution [48] (Section 5.2). We then perform the algorithm to identify the approximate posteriors of material parameters modeling the cyclic elastoplastic behavior of a steel specimen tested under different static cyclic loading conditions (Section 5.3).

We use a genetic algorithm to maximize the acquisition function in step 13 of Algorithm 1. Parameters for the genetic algorithm are listed in Table 1. Parameters for Algorithm 1 solving each identification example are provided thereof.

5.1. Example 1: Analysis of Gaussian model data

We find the approximate posterior of parameter vector $\mathbf{x} = [x_1, x_2]^\top \in \mathcal{X} = [0, 8]^2$ when the experimental dataset $\mathbf{d}_0 = \sum_{i=1}^5 \mathbf{d}_i/5$ is observed, where $\mathbf{d}_i$ ($i = 1, \dots, 5$) is distributed according to $p(\mathbf{d}|\mathbf{x}) = \mathcal{N}(\mathbf{d}|\mathbf{x}, \mathbf{\Sigma})$ and the covariance matrix $\mathbf{\Sigma}$ is known. If the prior $\pi(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mathbf{a}, \mathbf{B})$ is truncated to $\mathcal{X}$ for given $\mathbf{a}$ and $\mathbf{B}$, the true posterior can be derived as $\pi(\mathbf{x}|\mathbf{d}_0) = \mathcal{N}(\mathbf{x}|\mathbf{a}^*, \mathbf{B}^*)$, where $\mathbf{a}^* = \mathbf{B}^*(\mathbf{B}^{-1}\mathbf{a} + 5\mathbf{\Sigma}^{-1}\mathbf{d}_0)$ and $\mathbf{B}^* = (\mathbf{B}^{-1} + 5\mathbf{\Sigma}^{-1})^{-1}$ (see Ref. [36]). The true posterior is the baseline to assess the accuracy of approximate posteriors obtained from Algorithm 1.

To generate the experimental dataset $\mathbf{d}_0$ and specify the prior $\pi(\mathbf{x})$, we set $\mathbf{x} = [2, 2]^\top$, $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \Sigma_{21} = 0.5$, $\mathbf{a} = [5, 5]^\top$, and $\mathbf{B} = \mathbf{I}$ with the identity mapping $\mathbf{I}$. The feature vector and the discrepancy are defined by $\mathbf{z}(\mathbf{x}) = \mathbf{d}(\mathbf{x})$ (i.e., $\eta(\cdot)$ is an identity mapping) and a Mahalanobis distance $f(\mathbf{x}) = \sqrt{(\mathbf{z}(\mathbf{x}) - \mathbf{z}_0)^\top \mathbf{\Sigma}^{-1}(\mathbf{z}(\mathbf{x}) - \mathbf{z}_0)}$, respectively. The parameters for Algorithm 1 include the discrepancy threshold $\zeta \in \{1, 0.25\}$, the standard deviation of discrepancy noise $\omega = 0.5$, the number of initial samples $N = 10$, the number of BO iterations $s_{\max} = 100$, and the number of MCMC samples $N_s = 5 \times 10^4$.

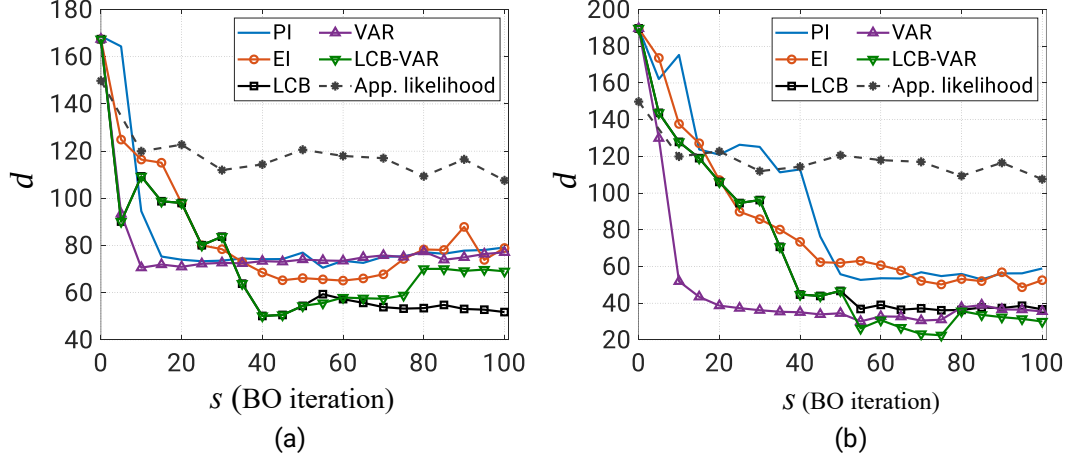


Figure 4: Example 1: Comparison of distance criterion values for different acquisition functions and two levels of discrepancy threshold. (a) $\zeta = 1$; (b) $\zeta = 0.25$.

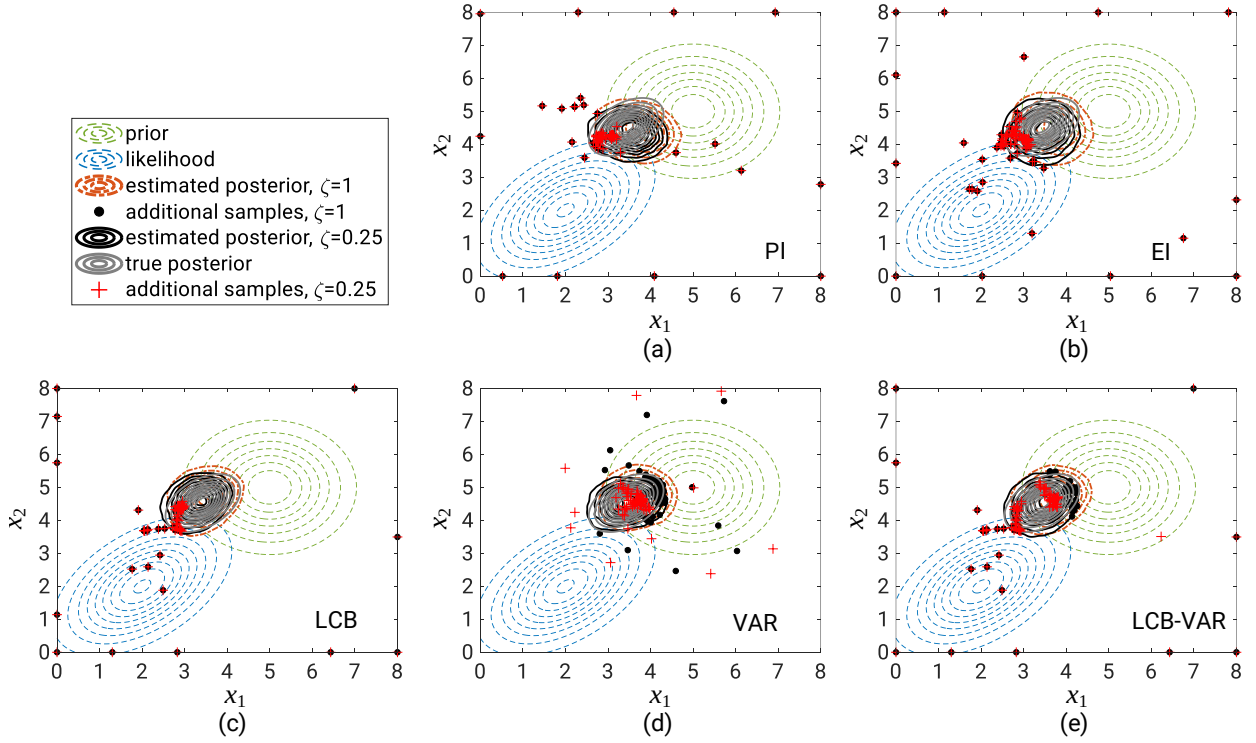


Figure 5: Example 1: Comparison of approximate posteriors for different acquisition functions and two levels of discrepancy threshold. (a) PI; (b) EI; (c) LCB; (d) VAR; (e) LCB-VAR.

In addition to utilizing only one of four acquisition functions (i.e., PI, EI, LCB, and VAR) stated in Section 4.2 for each identification attempt, we employ another strategy, namely LCB-VAR that switches the acquisition function from LCB to VAR during the identification process. Accordingly, LCB and VAR guide

BO to determine new parameter vectors for the first half and the second half of $s_{\max}$ iterations, respectively. To further assess the accuracy of approximate posterior, we define the following distance criterion:

$$d = \sum_{i=1}^{n_n} |\pi_{\zeta}^{\text{ABC}}(\mathbf{x}_i|\mathbf{z}_0) - \pi(\mathbf{x}_i|\mathbf{z}_0)|, \quad (18)$$

where $\mathbf{x}_i$ are coordinate vectors of the grid points of grid $[0, 0.1, \dots, 8]^2$ and $n_n = 81^2$ is the number of grid points. The estimation of $\pi_{\zeta}^{\text{ABC}}(\mathbf{x}_i|\mathbf{z}_0)$ from the ABC posterior samples requires a kernel density estimator.

For comparison purpose, we also compute d for the approximate posteriors found from carrying out the adaptive MCMC using a set of approximate likelihood functions defined as the GP models constructed from 11 different datasets corresponding to $N \in \{10, 20, \dots, 110\}$ samples generated by Latin-hypercube sampling [40]. This is called the approximate-likelihood approach hereafter. The approximate-likelihood approach is similar to the one-shot space-filling method by Fernandez-Zelaia et al. [25]. As a result, the number of training samples for $N \in \{10, 20, \dots, 110\}$ of the approximate-likelihood approach is the same as that for $s_{\max} \in \{0, 10, \dots, 100\}$ of the BO-assisted ABC, respectively.

Figure 4 shows the histories of the distance criterion d for different acquisition functions and two levels of discrepancy threshold $\zeta = 1$ and 0.25. The history of d from the approximate-likelihood approach is also plotted (i.e., dashed lines). There are some sudden jumps of d because the initial MCMC samples in step 23 of Algorithm 1 assigned as the new parameter vectors found by BO for different s values are generally different, showing that the improvement of approximate posterior depends on the initial sample of MCMC. In overall, however, the refinement of GP model for the discrepancy mean function considerably improves the accuracy of ABC posterior approximations for both threshold levels. The approximate ABC posterior is better under the smaller discrepancy threshold as the final values of d associated with $\zeta = 0.25$ in Fig. 4(b) are much smaller than those associated with $\zeta = 1$ in Fig. 4(a). For $\zeta = 0.25$, similar accuracy levels of the approximate posteriors are obtained from LCB and VAR. Switching the acquisition function from LCB to VAR during the identification attempt can slightly improve the approximate posterior when using a small discrepancy threshold; see Fig. 4(b). In addition, the speed of posterior improvement by the approximate-likelihood approach is much slower than that by the BO-assisted ABC approach, even though these approaches use the same number of training samples.

Figure 5 compares the approximate and true posteriors for different combinations of five acquisition function definitions and two discrepancy threshold levels. It is confirmed that the obtained approximate posteriors can reproduce the true posterior with acceptable accuracy. For the small discrepancy threshold level, the approximate posteriors brought by LCB, VAR, and LCB-VAR are more accurate than those by PI and EI, which is also indicated in Fig. 4(b).

Figure 5 also describes how new parameter samples are added during the refinement process for each

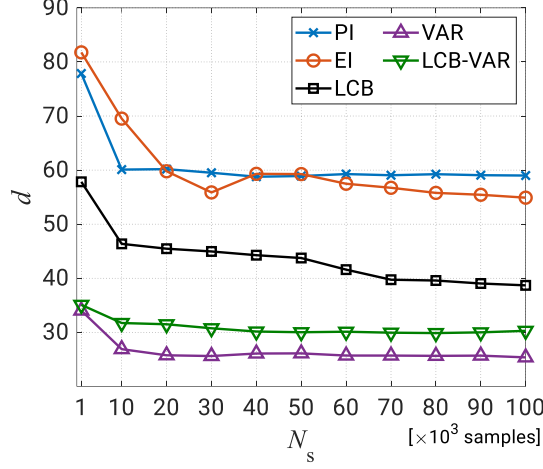


Figure 6: Example 1: Approximation errors versus numbers of MCMC samples for different acquisition functions and $\zeta = 0.25$.

acquisition function definition and a specific discrepancy threshold level. Accordingly, EI and LCB tend to draw new samples both in the likelihood area where the discrepancy is small, and on the boundary where the uncertainty in GP predictions is large due to the lack of training samples. VAR draws new samples in the prior area only since it incorporates the prior information as stated in Eq. (17). An interesting observation is that the new samples associated with two levels of ζ are identical when using PI, EI, and LCB, but are different when using VAR. This is because PI, EI, and LCB are independent of ζ , while VAR is dependent on ζ . Thus, it is convenient to investigate how different levels of discrepancy threshold affect the quality of approximate posterior when using PI, EI, or LCB as different levels of discrepancy threshold share the same refinement process. This also facilitates the selection of a proper discrepancy threshold ζ .

Figure 6 shows the effect of the number of MCMC samples on the quality of approximate posteriors for another initial training dataset and $\zeta = 0.25$. An increase in the number of MCMC samples when it is still small can improve the accuracy of approximate posteriors as the number of accepted samples in this situation is not enough to capture the true posterior accurately. However, the approximation quality almost remains unchanged when the number of MCMC samples reaches a large value, which is likely to depend on the type of acquisition function.

5.2. Example 2: Analysis of g -and- k data

This example performs Algorithm 1 to capture the true parameters for a univariate g -and- k distribution that is a unimodal PDF describing data with large amounts of skewness and kurtosis, which are controlled by parameters g and k , respectively. Although the g -and- k distribution has no closed form, it can be defined

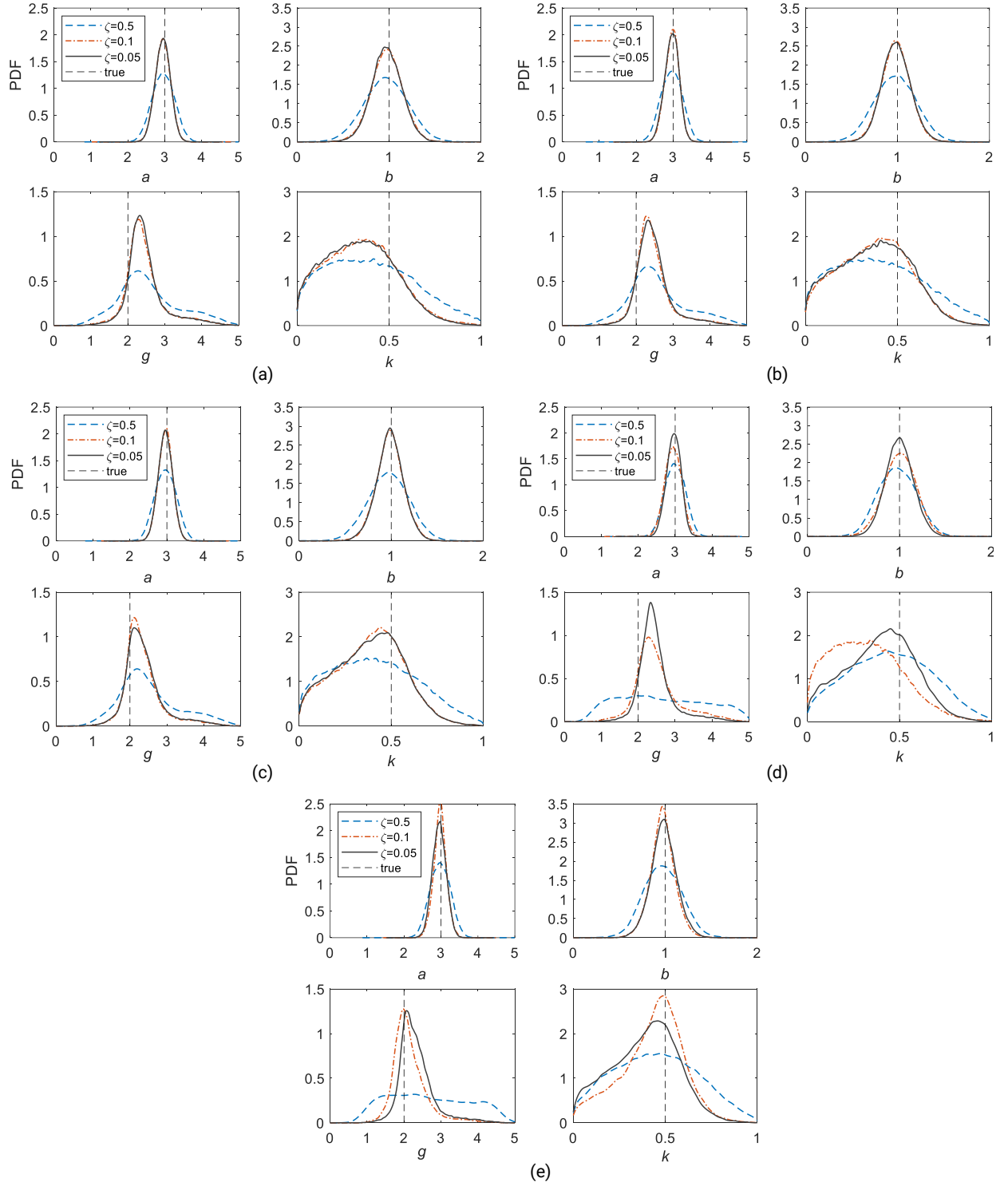


Figure 7: Example 2: Approximate posteriors of each parameter for different acquisition functions and three levels of discrepancy threshold. (a) PI; (b) EI; (c) LCB; (d) VAR; (e) LCB-VAR.

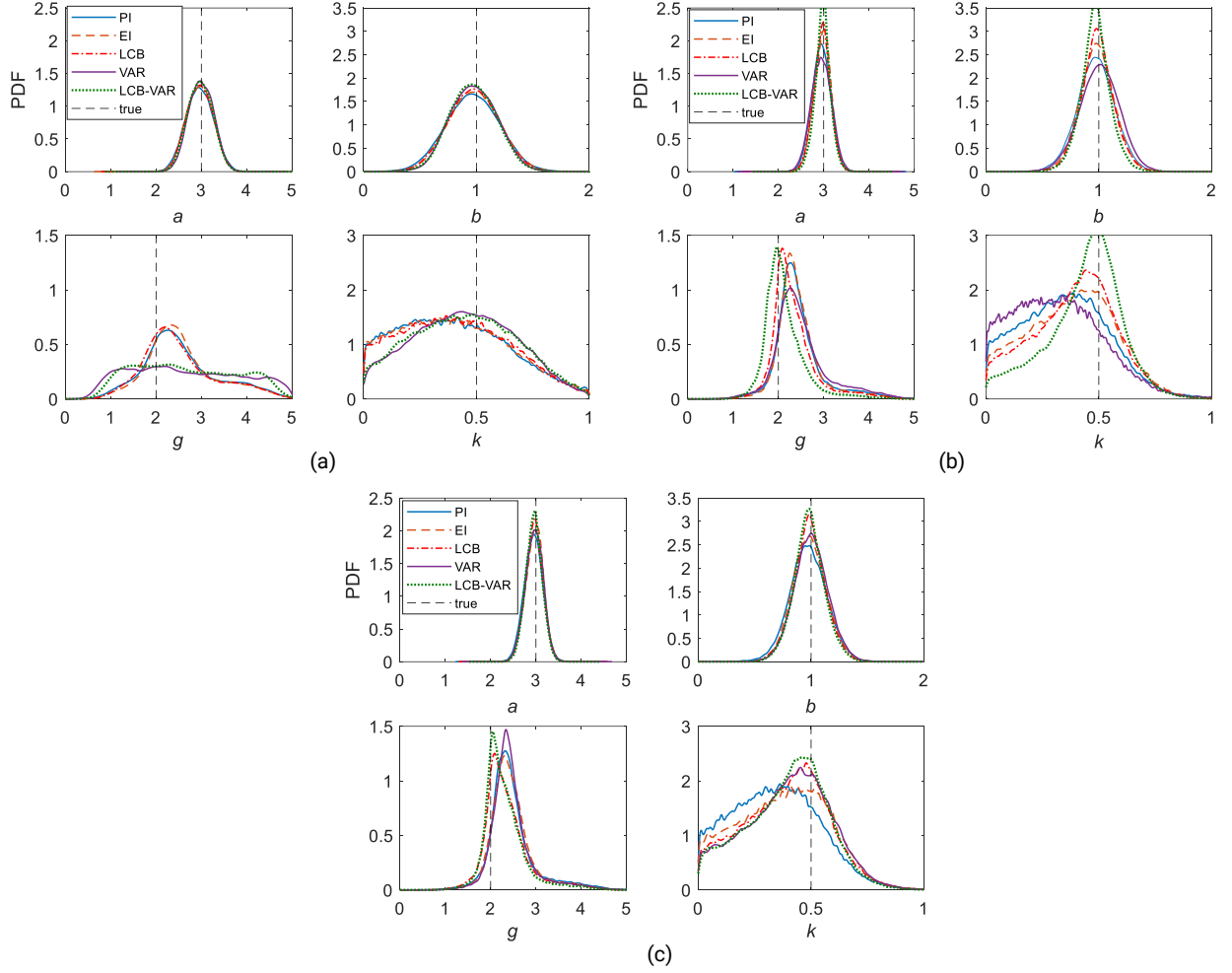


Figure 8: Example 2: Comparison of approximate posteriors obtained from different acquisition functions for each level of discrepancy threshold. (a) $\zeta = 0.5$; (b) $\zeta = 0.1$; (c) $\zeta = 0.05$.

through the following quantile function (i.e., inverse CDF) [48]:

$$Q_{\mathbf{gk}}(\lambda(q)|\mathbf{x}) = a + b \left(1 + c \frac{1 - \exp(-g\lambda(q))}{1 + \exp(-g\lambda(q))} \right) (1 + \lambda^2(q))^k \lambda(q) \quad (19)$$

where $\lambda(q) = \Phi^{-1}(q)$ is the q th quantile of the standard Gaussian; $\mathbf{x} = [a, b, g, k]^\top$ is the vector of four parameters characterizing the distribution location, scale, skewness, and kurtosis, respectively; and c is fixed at 0.8. Our goal is to use Algorithm 1 for identifying the parameter vector $\mathbf{x} = [a, b, g, k]^\top$ that best fits the g -and- k distribution to an observed dataset $\mathbf{d}_0$.

A synthetic dataset $\mathbf{d}_0$ of 10^4 independent samples is generated from Eq. (19) with $\mathbf{x} = [3, 1, 2, 0.5]^\top$. More specifically, 10^4 independent samples of $\lambda(q)$ are drawn from the standard Gaussian, and each sample is then substituted into Eq. (19) with $\mathbf{x} = [3, 1, 2, 0.5]^\top$ to compute each element of $\mathbf{d}_0$. By employing Algorithm 1,

Table 2: Example 2: Comparison of MAP estimates of each parameter for different acquisition functions with $\zeta = 0.5$ and 0.05 .

ζ	Parameter	PI	EI	LCB	VAR	LCB-VAR	true
0.5	a	2.939	2.981	2.967	2.963	2.972	3.00
	b	0.956	1.004	0.969	0.952	0.959	1.00
	g	2.893	2.896	2.759	2.873	2.838	2.00
	k	0.415	0.340	0.366	0.436	0.474	0.50
0.05	a	2.962	2.985	2.958	2.973	2.974	3.00
	b	0.951	0.990	0.987	1.002	0.986	1.00
	g	2.743	2.852	2.735	2.863	2.710	2.00
	k	0.361	0.409	0.479	0.451	0.458	0.50

Table 3: Example 2: Comparison of standard deviation values of each parameter for different acquisition functions with $\zeta = 0.5$ and 0.05 .

ζ	Parameter	PI	EI	LCB	VAR	LCB-VAR
0.5	a	0.297	0.284	0.284	0.270	0.273
	b	0.236	0.226	0.222	0.206	0.205
	g	0.838	0.790	0.816	1.188	1.098
	k	0.237	0.238	0.234	0.223	0.229
0.05	a	0.203	0.191	0.190	0.192	0.181
	b	0.165	0.154	0.144	0.152	0.137
	g	0.524	0.480	0.514	0.469	0.460
	k	0.188	0.195	0.186	0.186	0.178

we wish to find the approximate posterior of $\mathbf{x}$ so that the associated MAP estimate is close to the true parameter vector $\mathbf{x} = [3, 1, 2, 0.5]^\top$. For this purpose, let $\pi(\mathbf{x}) = \pi(a)\pi(b)\pi(g)\pi(k) = U(0, 5)U(0, 2)U(0, 5)U(0, 1)$, where $U(\cdot)$ denotes the uniform PDF over the interval $(\cdot)$, and $\mathbf{z}(\mathbf{x}) = [z_1, z_2, z_3, z_4]^\top$ with

$$z_1 = l_2; z_2 = l_3 - l_1; z_3 = (l_3 + l_1 - 2l_2)/z_2; z_4 = (o_7 - o_5 + o_3 - o_1)/z_2. \quad (20)$$

Here l_i and o_j are the i th quartile and the j th octile of the elements of $\mathbf{d}(\mathbf{x}) \in \mathbb{R}^{10^4}$, respectively. The feature vector $\mathbf{z}(\mathbf{x})$, following Ref. [48], is called robust summary statistics with z_1, z_2, z_3 , and z_4 representing the median, interquartile range, skewness, and kurtosis of $\mathbf{d}(\mathbf{x})$, respectively. Since the generation of $\mathbf{d}(\mathbf{x})$ from Eq. (19) is similar to that of $\mathbf{d}_0$, each sample of $\mathbf{x}$ drawn from $\pi(\mathbf{x})$ yields a unique vector $\mathbf{z}(\mathbf{x})$ when a

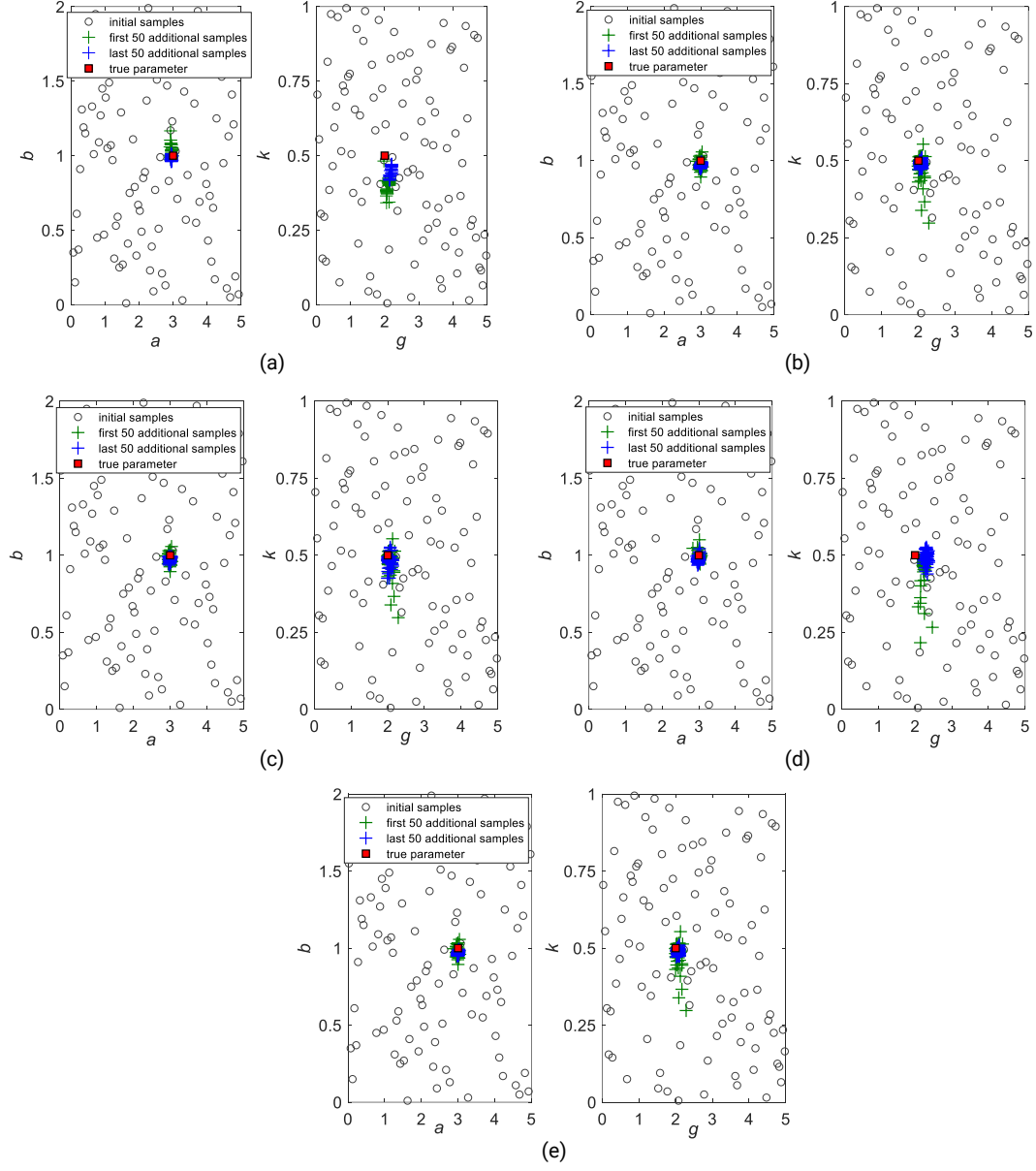


Figure 9: Example 2: Additional samples for different acquisition functions with $\zeta = 0.05$. (a) PI; (b) EI; (c) LCB; (d) VAR; (e) LCB-VAR.

random seed is used throughout the identification process.

The discrepancy mean function is defined by the Euclidean norm $f(\mathbf{x}) = \|\mathbf{z}(\mathbf{x}) - \mathbf{z}_0\|$. The parameters for Algorithm 1 include $\zeta \in \{0.5, 0.1, 0.05\}$, $N = 100$ samples, $s_{\max} = 100$ iterations, and $N_s = 10^5$ samples. Since this example focuses on capturing the best-fit parameter vector, ω is automatically optimized in each BO iteration by maximizing the approximate likelihood of $\pi(\Delta|\mathbf{x})$ given the GP model $\hat{f}(\mathbf{x})$.

Figure 7 shows the approximate posteriors of each parameter for different acquisition function definitions and three levels of discrepancy threshold. The MAP estimates of each parameter by Algorithm 1 can be found in the neighborhood of the true parameter value. The PDF values at the MAP estimates of all parameters associated with $\zeta = 0.5$ are much smaller than those associated with $\zeta = 0.1$ and 0.05 as there is a large number of accepted samples found from the prior when using the larger ζ , which in turn flattens the approximate posteriors.

Figure 8 compares the approximate posteriors of each parameter obtained from different acquisition functions for a specific discrepancy threshold level. For parameters a and b , there is no major difference between the approximate posteriors obtained from different acquisition functions at the same discrepancy threshold level. However, the approximate posteriors of parameters g and k are strongly affected by the acquisition function used for the refinement process.

Tables 2 and 3 respectively compare the MAP estimates and standard deviation values associated with the approximate posteriors of each parameter obtained from different acquisition functions and two levels of discrepancy threshold $\zeta = 0.5$ and 0.05 . As our expectation, the MAP estimates of each parameter are close to the true parameter. The smaller discrepancy threshold value, i.e., $\zeta = 0.05$, can provide more accurate MAP estimates as well as smaller standard deviation values of the approximate posteriors than the larger one, i.e., $\zeta = 0.5$. Among the five acquisition function definitions, LCB-VAR with $\zeta = 0.05$ offers the most accurate posterior approximation.

The ability of MAP estimates to capture the true parameters for the g -and- k distribution is due to the special characteristic of additional samples found by BO. As depicted in Fig. 9, the additional samples found from maximizing all the acquisition functions tend to concentrate in the neighborhood of the true parameter vector, where samples drawn from the prior have a high chance to be accepted as posterior samples.

5.3. Example 3: A steel specimen

Algorithm 1 is performed to find the approximate posteriors of cyclic elastoplastic parameters for a steel specimen shown in Fig. 10. The specimen was tested under two different static cyclic loading histories SS1 and SS2 [49]. The experimental results consist of two datasets of axial true stress σ and axial true strain ϵ associated with the loading histories. These experimental results were used in our previous works [11, 12, 50] to solve different deterministic inverse problems with single and multiple objective functions. The reader may refer to Yamada and Jiao [49] for the detail of experimental program as well as the calculation of σ and ϵ from the axial force and the corresponding axial deformation of the specimen measured at each time instant of interest.

The material behavior of the specimen is simulated using the nonlinear combined isotropic/kinematic

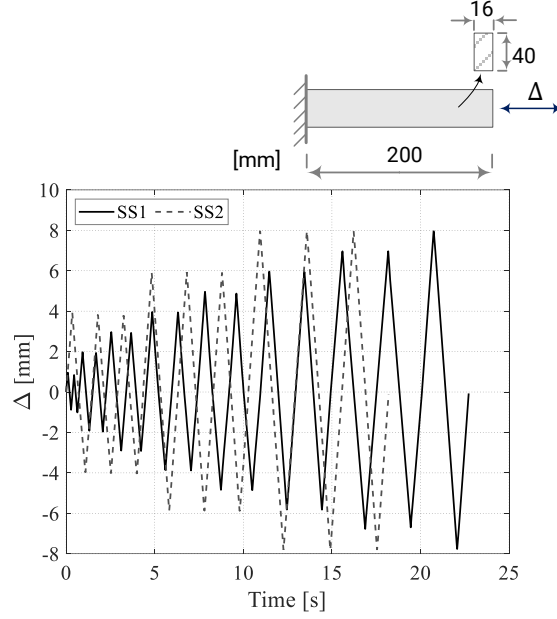


Figure 10: Example 3: Steel specimen and two loading histories for cyclic tests [11, 49].

hardening model, which is detailed in Appendix A and implemented by Abaqus 2020 [51]. Such a hardening model, when using only one back-stress component, is characterized by seven parameters, namely Poisson's ratio ν , Young's modulus E , initial yield stress $\sigma_{y,0}$, two nonlinear kinematic parameters C_1 and γ_1 , and two isotropic hardening parameters Q_∞ and b . In this study, however, Poisson's ratio and Young's modulus are kept constant as $\nu = 0.3$ and $E = 205.94$ GPa [11], respectively. Thus, the approximate posteriors of five material parameters are identified for the specimen, i.e., $\mathbf{x} = [\sigma_{y,0}, C_1, \gamma_1, Q_\infty, b]^\top$.

Since the specimen was axially loaded during the cyclic tests, we model it using one element of type C3D8R of Abaqus 2020 [51]. For every time instant of interest, the FE program directly extracts the numerical (simulated) values of σ and ϵ in the direction of loading at the integration point of the element.

Let $\sigma_{i,t}^s(\mathbf{x})$ denote the simulated value of σ extracted from Abaqus at the t th time step of the i th cyclic loading history with N_i discrete steps, and $\sigma_{i,t}^m$ denote the corresponding measured value. Thus, the experimental dataset $\mathbf{d}_0$ for the i th cyclic loading history is a vector containing N_i measured values of σ . If we further define the feature vector $\mathbf{z}(\mathbf{x}) = \mathbf{d}(\mathbf{x})$ (i.e., $\eta(\cdot)$ is an identity mapping), the discrepancy mean associated with the i th cyclic loading history for the identification problem of material parameters for the specimen can read the following root-mean-square deviation [11, 50]:

$$f_i(\mathbf{x}) = \sqrt{\frac{1}{N_i} \sum_{t=1}^{N_i} (\sigma_{i,t}^s(\mathbf{x}) - \sigma_{i,t}^m)^2}. \quad (21)$$

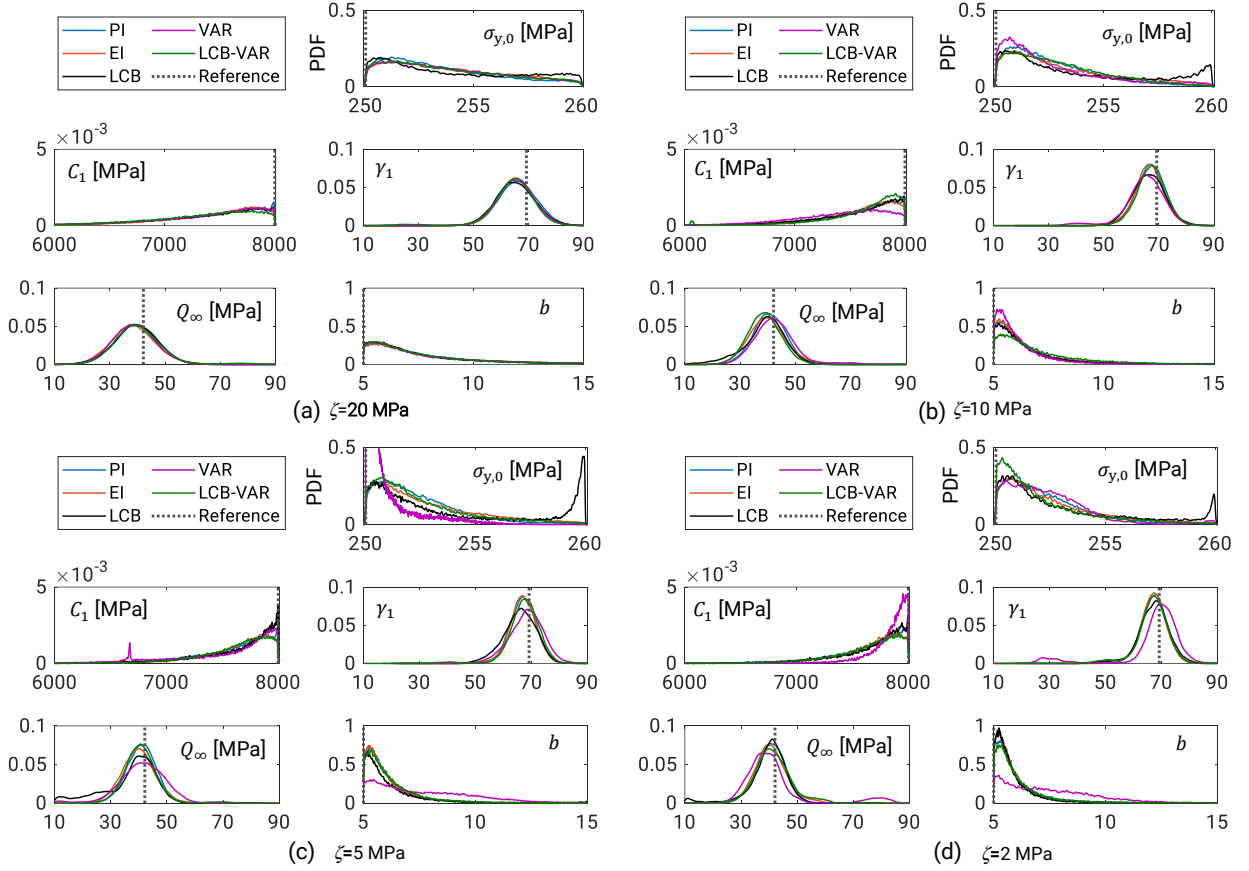


Figure 11: Example 3: Approximate posteriors of each parameter obtained from five acquisition functions, four levels of discrepancy threshold, and dataset SS1. Reference values are from the deterministic approach [11].

Using the interval for each parameter provided in Ref. [11], we define the prior of independent material parameters as $\pi(\mathbf{x}) = \pi(\sigma_{y,0})\pi(C_1)\pi(\gamma_1)\pi(Q_\infty)\pi(b)$, where $\pi(\sigma_{y,0}) = U(250, 260)$ MPa, $\pi(C_1) = U(2000, 8000)$ MPa, $\pi(\gamma_1) = U(10, 100)$, $\pi(Q_\infty) = U(10, 100)$ MPa, and $\pi(b) = U(5, 25)$. For identification, we set the parameters for Algorithm 1 as $\zeta \in \{20, 10, 5, 2\}$ MPa, $\omega = 3$ MPa, $N = 100$ samples, $s_{\max} = 100$ iterations, and $N_s = 10^5$ samples. Thus, each identification attempt requires 200 simulations in total.

Although we assign the same ω value for all cyclic tests during the identification process, it is also possible to estimate ω for each test by performing the following four steps. First, observational noise involved in each stress measurement can be calibrated by acquiring data without use of the steel specimen. Second, the calibrated noise is assigned to each measured stress value (i.e., each element of $\mathbf{d}_0$), leading to the generation of a large number of $\mathbf{d}_0$ samples, say 10^4 . Third, a small number (e.g., 3) of parameter vectors $\mathbf{x}$ is randomly generated and ω associated with each parameter vector is evaluated from $\mathbf{d}(\mathbf{x})$ and the samples of $\mathbf{d}_0$ generated in the second step. Finally, ω for Algorithm 1 is defined as the average of the ω values

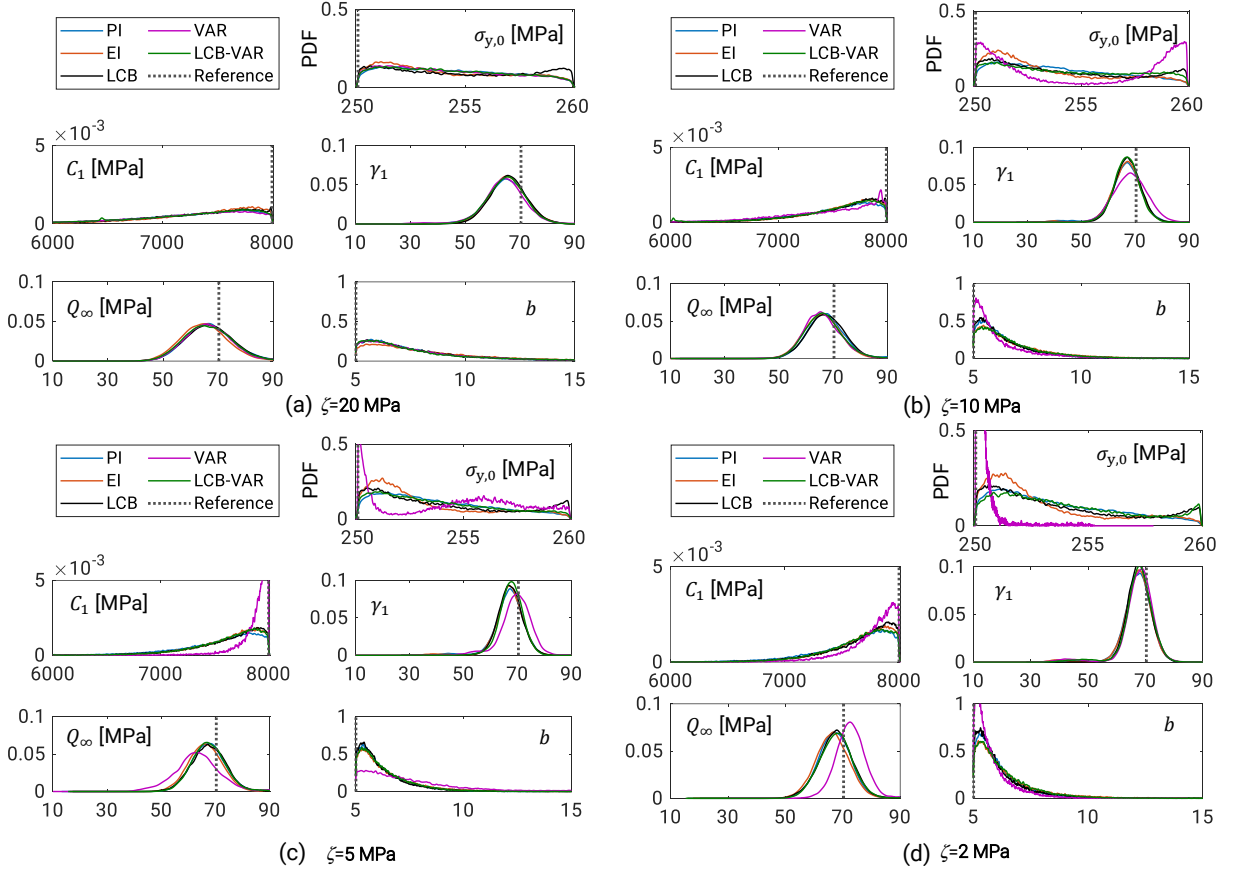


Figure 12: Example 3: Approximate posteriors of each parameter obtained from five acquisition functions, four levels of discrepancy threshold, and dataset SS2. Reference values are from the deterministic approach [11].

obtained from the third step. The reader may consult with Do and Ohsaki [11] for the detail of how to estimate ω given noise involved in each stress measurement.

Algorithm 1 inputs the experimental dataset from SS1 or SS2 for finding the approximate posterior of each parameter. Then, the MAP estimate associated with the resulting approximate posterior is compared with the corresponding optimal deterministic value (i.e., reference value) provided in Ref. [11].

Approximate posteriors and MAP estimates

Figures 11 and 12 show the approximate posteriors of each parameter obtained from the datasets SS1 and SS2, respectively, with four discrepancy threshold levels and five acquisition function definitions. The approximate posteriors of each parameter brought by PI, EI, LCB, and LCB-VAR are similar for each discrepancy threshold level. The difference between the posterior by VAR and those by other acquisition functions is larger when the discrepancy threshold level becomes smaller.

Tables 4 and 5 respectively show the MAP estimates and standard deviation values of each parameter

Table 4: Example 3: Comparison of MAP estimates of each parameter for different combinations of the acquisition function and experimental dataset with $\zeta = 20$ and 2 MPa.

Dataset	Parameter	PI	EI	LCB	VAR	LCB-VAR	Ref. [11]
SS1 ($\zeta = 20$ MPa)	$\sigma_{y,0}$ [MPa]	251.24	251.21	250.65	251.45	251.19	250.00
	C_1 [MPa]	8000.00	7804.99	8000.00	7848.57	7790.70	8000.00
	γ_1	65.50	65.22	64.64	66.51	65.50	69.19
	Q_∞ [MPa]	38.64	38.41	39.64	37.66	38.68	42.11
SS1 ($\zeta = 2$ MPa)	$\sigma_{y,0}$ [MPa]	250.71	250.64	250.56	250.49	250.28	250.00
	C_1 [MPa]	7967.22	7880.43	8000.00	8000.00	7915.22	8000.00
	γ_1	68.03	67.42	68.19	69.61	67.25	69.19
	Q_∞ [MPa]	40.27	39.55	41.15	37.53	39.91	42.11
	b	5.25	5.20	5.22	5.00	5.28	5.00
SS2 ($\zeta = 20$ MPa)	$\sigma_{y,0}$ [MPa]	251.81	251.10	250.75	251.30	251.25	250.01
	C_1 [MPa]	7714.71	7807.73	8000.00	7671.86	7741.45	7999.77
	γ_1	64.83	64.79	65.50	64.11	65.60	70.31
	Q_∞ [MPa]	66.58	64.81	65.92	65.73	64.68	70.28
	b	5.70	5.77	5.65	5.57	5.56	5.01
SS2 ($\zeta = 2$ MPa)	$\sigma_{y,0}$ [MPa]	251.12	251.23	250.67	250.05	251.33	250.01
	C_1 [MPa]	7868.71	7881.34	7906.98	7948.73	7847.83	7999.77
	γ_1	68.08	67.68	67.36	68.80	67.94	70.31
	Q_∞ [MPa]	67.15	66.73	67.92	72.52	68.46	70.28
	b	5.30	5.31	5.29	5.18	5.35	5.01

for different combinations of the acquisition function and experimental dataset with $\zeta = 20$ and 2 MPa. The reference value in the last column of Table 4 obtained from the deterministic approach [11] is the baseline to assess the accuracy of the corresponding MAP estimate. This assessment is reasonable because the use of uniform prior for each parameter leads to the fact that the MAP estimate can be found in the neighborhood of the maximum likelihood estimation, which is the solution by the deterministic approach [11]. Results in Table 4 show that the MAP estimates of all parameters from different acquisition function definitions are close to the reference values in both cases of $\zeta = 20$ and 2 MPa. Results in Table 5 confirm that the standard deviation of each parameter is small when a small discrepancy threshold is used together with PI or EI, which indicates an increase in the reliability of posterior approximation. This, however, is only apparent to

Table 5: Example 3: Comparison of the standard deviation values of each parameter for different combinations of the acquisition function and experimental dataset with $\zeta = 20$ and 2 MPa.

Dataset	Parameter	PI	EI	LCB	VAR	LCB-VAR
SS1 ($\zeta = 20$ MPa)	$\sigma_{y,0}$ [MPa]	2.62	2.72	3.10	2.69	2.75
	C_1 [MPa]	539.87	539.50	558.73	534.93	990.65
	γ_1	7.73	8.31	7.93	7.84	10.67
	Q_∞ [MPa]	3.67	3.94	3.85	3.61	3.56
SS1 ($\zeta = 2$ MPa)	$\sigma_{y,0}$ [MPa]	1.88	2.22	3.92	2.30	1.98
	C_1 [MPa]	341.99	464.31	322.63	957.01	542.55
	γ_1	4.50	4.95	8.65	9.62	5.27
	Q_∞ [MPa]	5.17	5.91	12.60	9.96	5.99
	b	1.37	1.06	5.16	2.04	8.84
SS2 ($\zeta = 20$ MPa)	$\sigma_{y,0}$ [MPa]	2.84	2.96	3.21	2.88	2.86
	C_1 [MPa]	692.79	601.52	593.90	1254.92	600.37
	γ_1	7.53	7.11	6.77	10.20	6.94
	Q_∞ [MPa]	9.12	8.87	9.18	11.16	9.01
	b	3.24	3.79	3.96	3.76	3.49
SS2 ($\zeta = 2$ MPa)	$\sigma_{y,0}$ [MPa]	2.60	2.68	3.06	0.47	2.94
	C_1 [MPa]	504.51	543.85	315.44	525.99	713.21
	γ_1	5.10	5.56	3.97	5.44	6.07
	Q_∞ [MPa]	6.38	7.00	6.02	6.02	8.55
	b	1.12	1.33	1.32	0.75	1.15

some parameters when using LCB, VAR, or LCB-VAR.

Effect of discrepancy threshold

Figures 13 and 14 compare the approximate posteriors of each parameter obtained from datasets SS1 and SS2, respectively, with use of PI, LCB-VAR, and four levels of discrepancy threshold. We see that a large value of the discrepancy threshold flattens the approximate posterior of each parameter. An increasing order of PDF values at the MAP estimate of each parameter from both PI and LCB-VAR corresponds to $\zeta = 20$, 10, 5, and 2 MPa. The difference between the approximate posteriors of each parameter by $\zeta = 5$ and 2 MPa is not pronounced, indicating that the approximate posterior could cease to improve when the discrepancy threshold reaches a small-enough value, which is also observed in Figs. 7(a), (b), and (c).

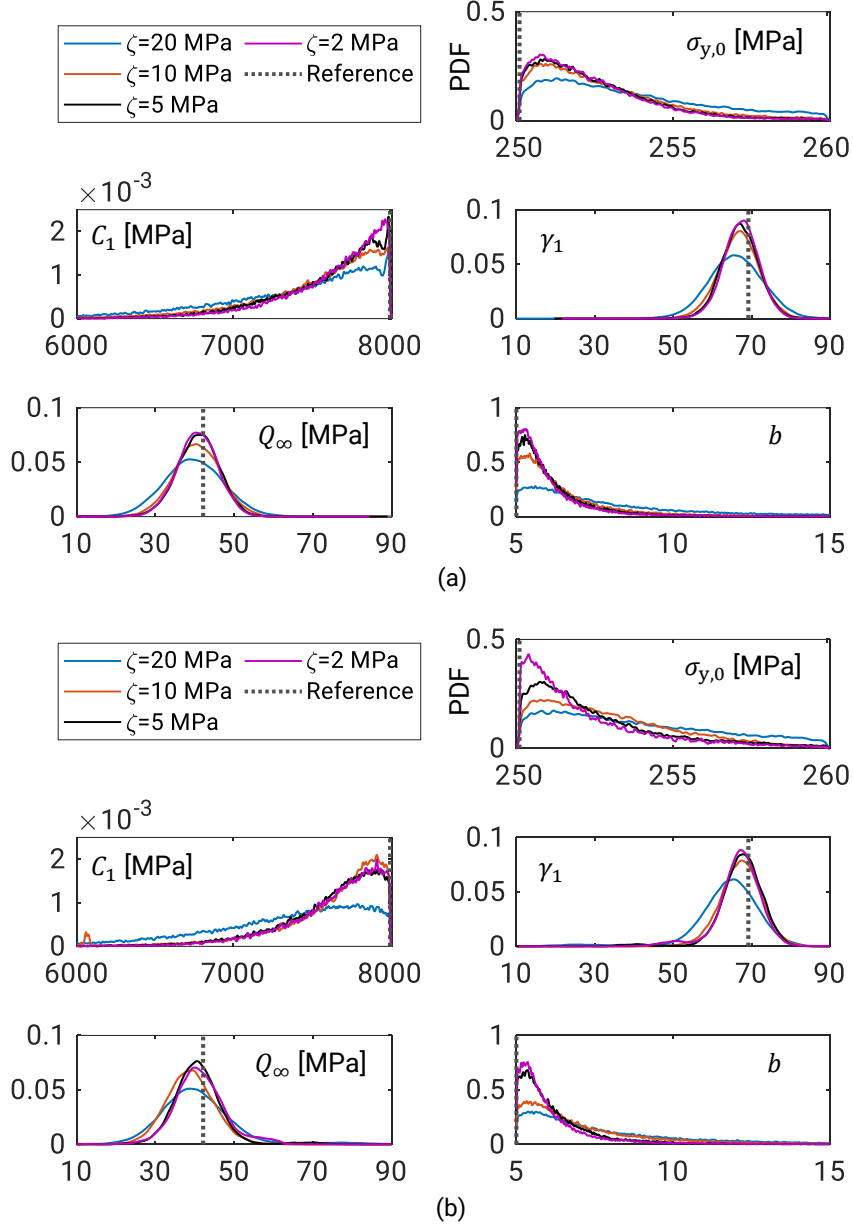


Figure 13: Example 3: Comparison of approximate posteriors of each parameter from four levels of discrepancy threshold and dataset SS1. (a) result from PI; (b) result from LCB-VAR. Reference values are from the deterministic approach [11].

Effect of initial training dataset

Since the GP model for the discrepancy mean is constructed from the training dataset $\mathcal{D}$, it is desirable to investigate how use of different initial training datasets affects the resulting approximate posteriors. For this purpose, we use Latin-hypercube sampling [40] to randomly generate three different initial training datasets, namely Dataset1, Dataset2, and Dataset3. Each dataset consists of 100 different sets of the

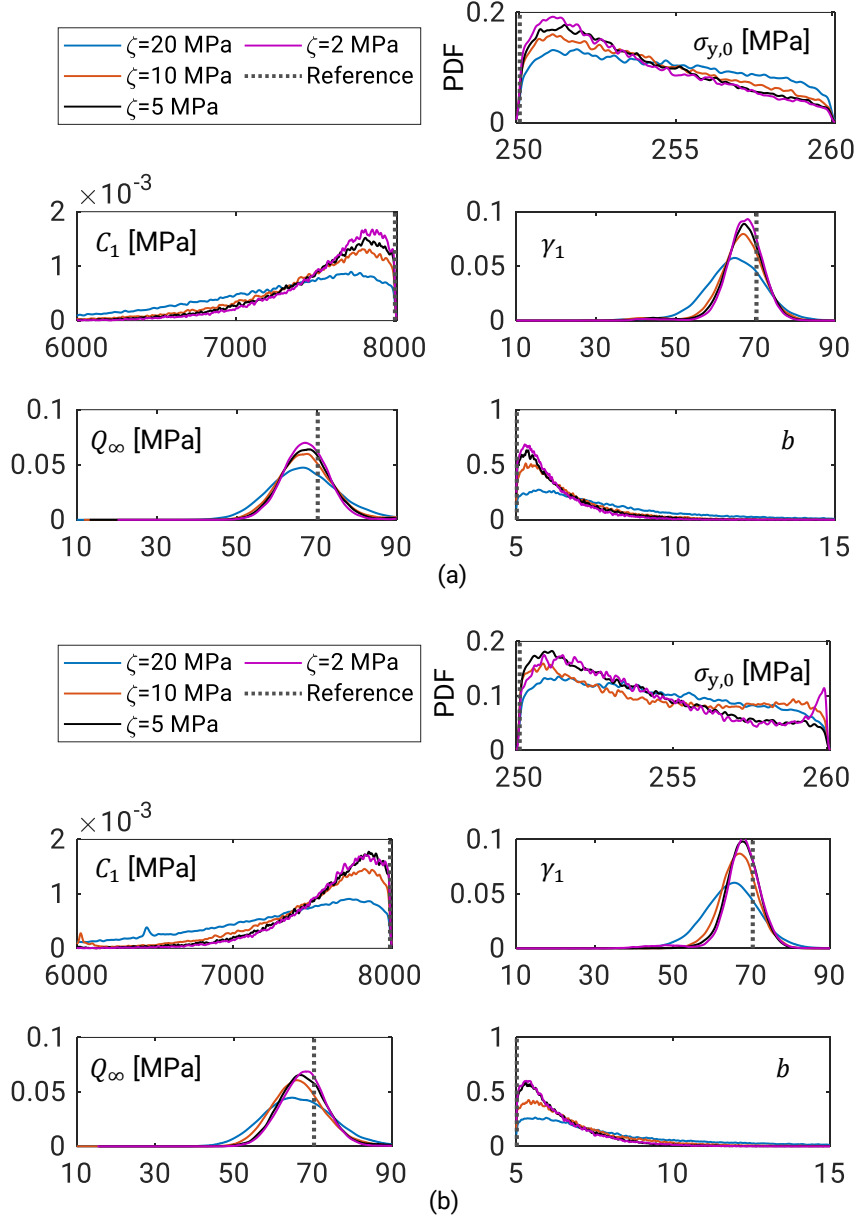


Figure 14: Example 3: Comparison of approximate posteriors of each parameter from four levels of discrepancy threshold and dataset SS2. (a) result from PI; (b) result from LCB-VAR. Reference values are from the deterministic approach [11].

parameter vectors $\mathbf{x}$ and the corresponding values of f computed from the numerical and measured values of σ as stated in Eq. (21). We then perform Algorithm 1 starting from each of these datasets.

Figures 15 and 16 compare the approximate posteriors of each parameter associated with PI, LCB-VAR, and three initial training datasets when ABC uses the experimental results from SS1 and SS2, respectively. For this identification problem, there is no major difference in the approximate posteriors obtained from the

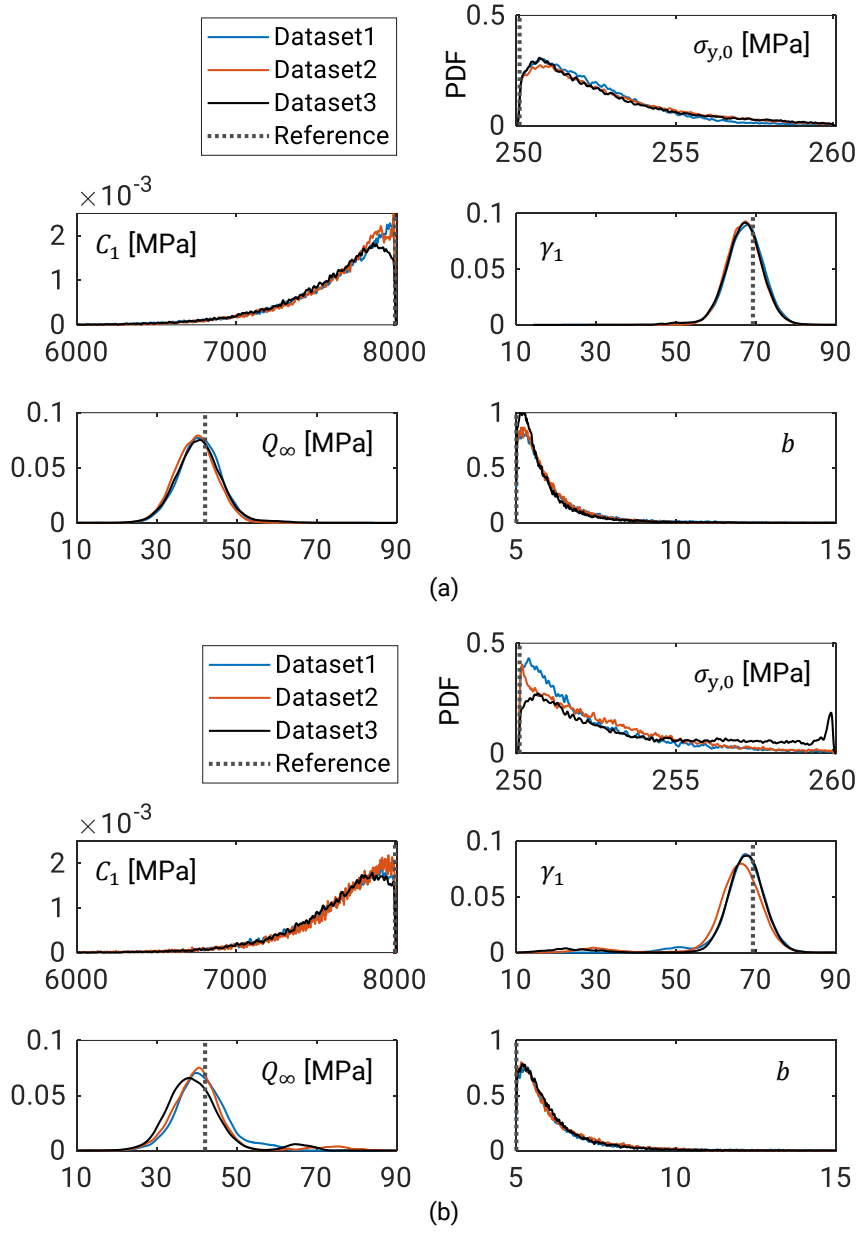


Figure 15: Example 3: Comparison of approximate posteriors of each parameter from dataset SS1, $\zeta = 2$ MPa, and three ABC attempts with different initial training datasets. (a) result from PI; (b) result from LCB-VAR. Reference values are from the deterministic approach [11].

three initial training datasets when using PI; see Figs. 15(a) and 16(a). However, different initial training datasets can lead to different approximate posteriors when using LCB-VAR; see Figs. 15(b) and 16(b).

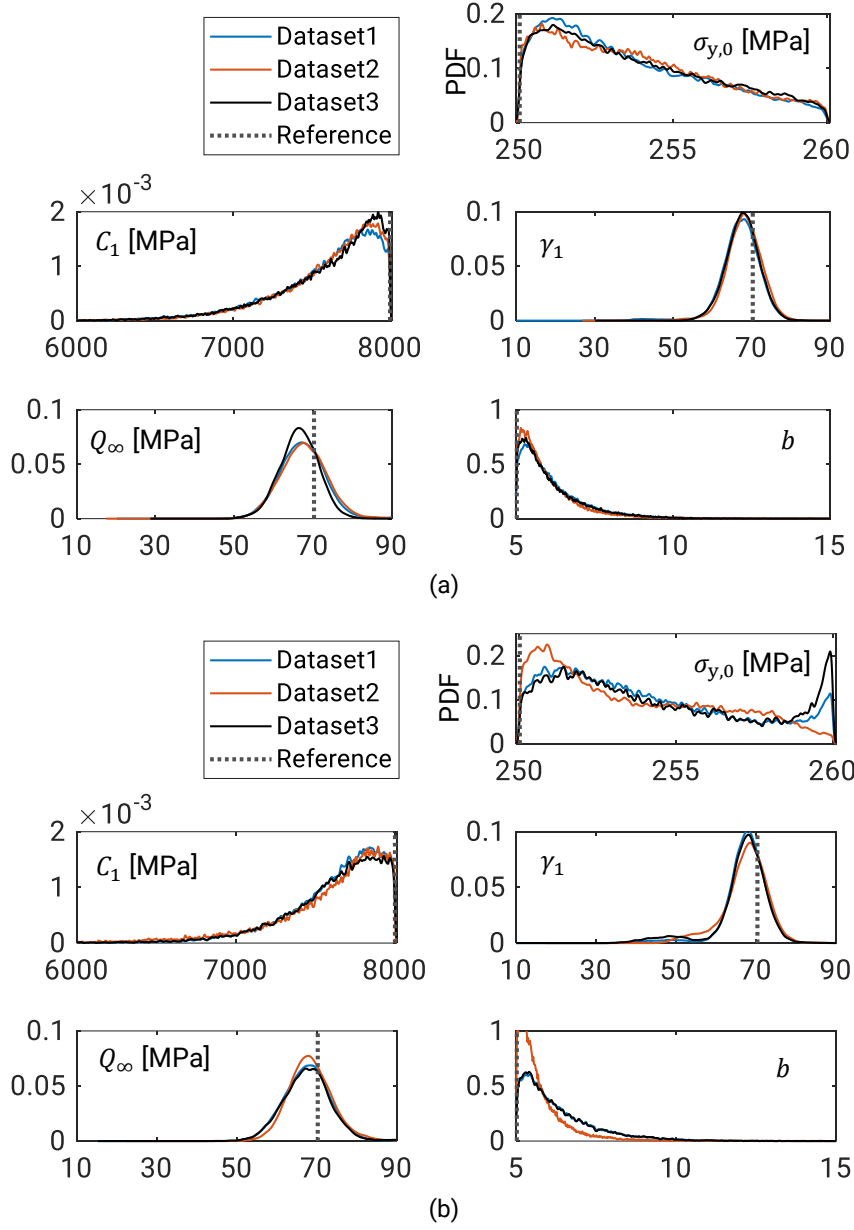


Figure 16: Example 3: Comparison of approximate posteriors of each parameter from dataset SS2, $\zeta = 2$ MPa, and three ABC attempts with different initial training datasets. (a) result from PI; (b) result from LCB-VAR. Reference values are from the deterministic approach [11].

Posterior update

Since the experimental datasets SS1 and SS2 are obtained from different loading histories, we wish to update the approximate posterior of each parameter in an online fashion when each dataset is sequentially observed. For this purpose, we carry out two updating schemes using PI and $\zeta = 2$ MPa, namely SS1-SS2 and SS2-

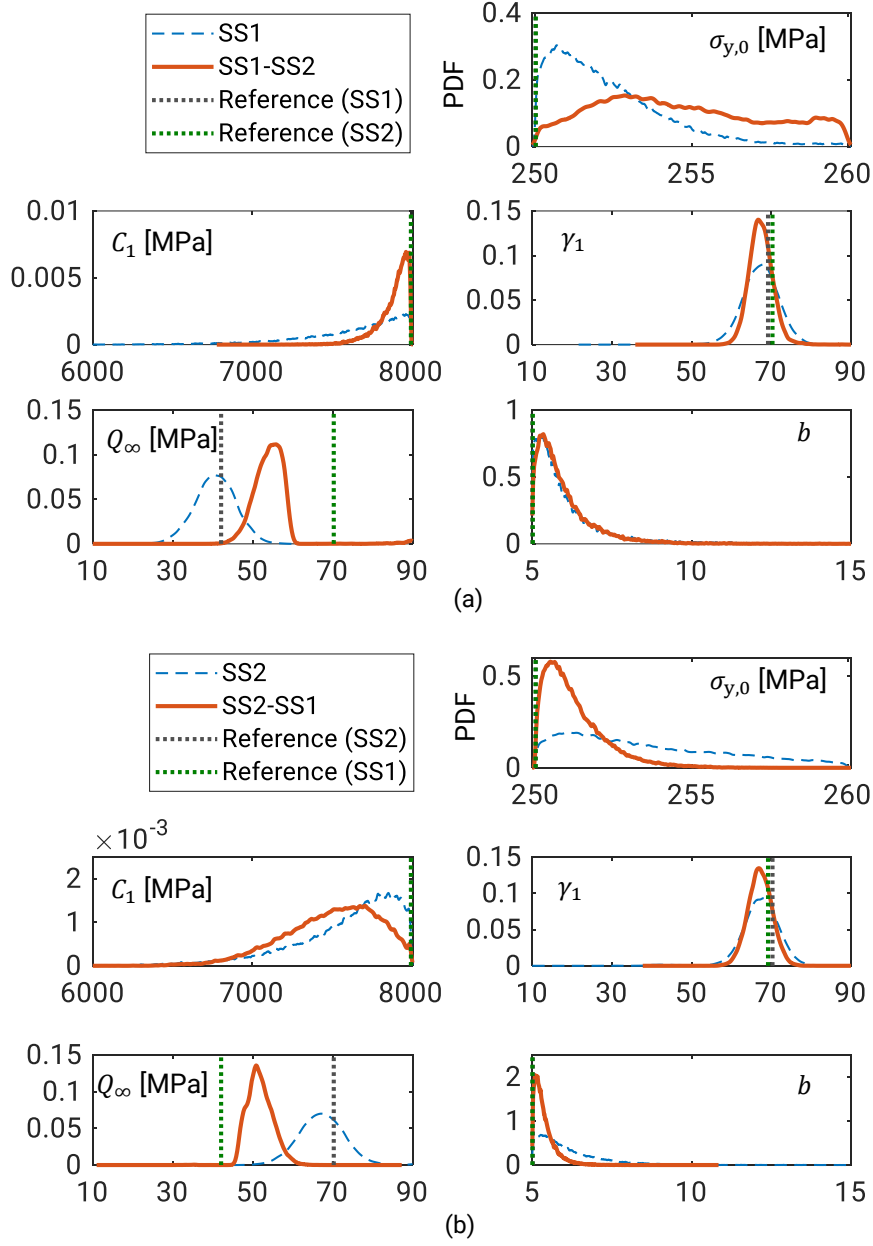


Figure 17: Example 3: Approximate posteriors obtained from two updating schemes with PI and $\zeta = 2$ MPa. (a) SS1-SS2 scheme; (b) SS2-SS1 scheme. Reference values are from the deterministic approach [11].

SS1, and the evolution of approximate posteriors during each scheme is shown in Fig. 17. In fact, the SS1-SS2 scheme uses the approximate posterior obtained from SS1 as prior to infer the posterior when the dataset of SS2 is observed. A similar interpretation can be made for the SS2-SS1 scheme. In most cases of Fig. 17, adding more experimental datasets can reduce the variance of each parameter estimated from the

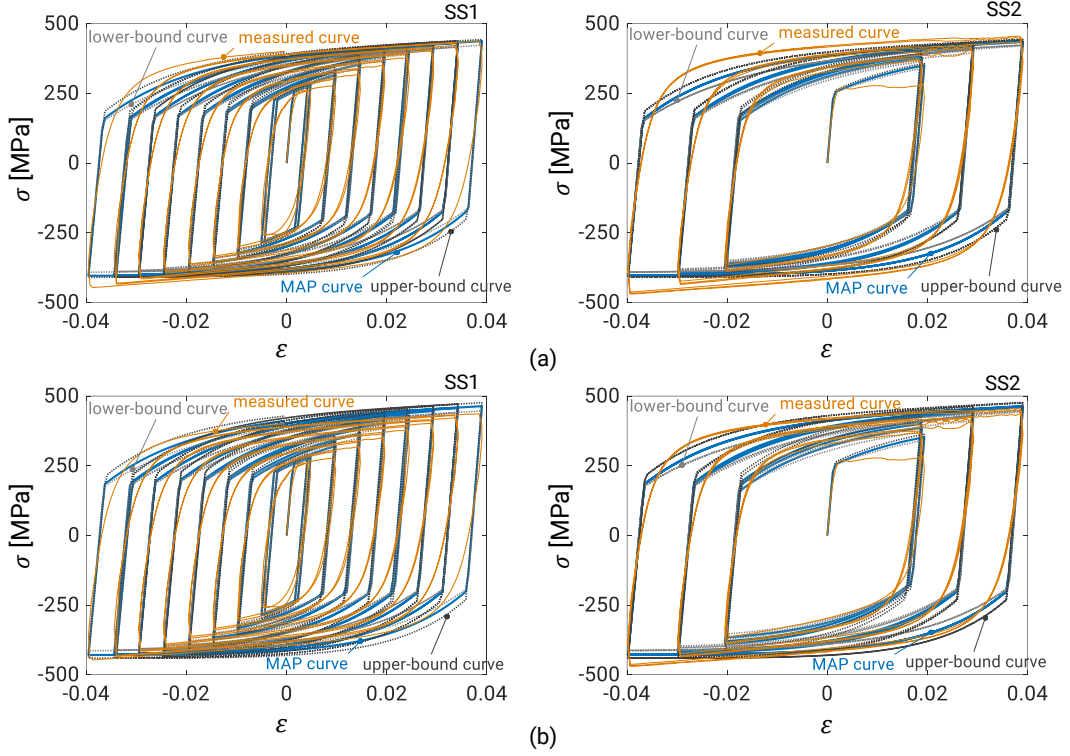


Figure 18: Example 3: Comparison of test data and model predictions for cyclic tests of the specimen with the posteriors obtained from PI. (a) parameters from dataset SS1; (b) parameters from dataset SS2.

corresponding posterior samples. An interesting observation is that the approximate posteriors of Q_∞ , which depends on the dataset used for deterministic identification [11], obtained from the two updating schemes are capable of representing the trade-off between the reference values found from individual experimental datasets.

Prediction

To illustrate one of their applications to predicting the cyclic elastoplastic behavior of the steel specimen, the approximate posteriors obtained from PI with $\zeta = 2$ MPa for SS1 and SS2 are used to reproduce the $\sigma - \epsilon$ curves of SS1 and SS2. We consider three sets of parameters found from the posterior of each parameter, namely the MAP set as listed in Table 4, and the upper and lower sets of the 95% confidence interval of each parameter, which is estimated from samples generated from the corresponding approximate posterior. Accordingly, the prediction of σ shown in Fig. 18 for the $\sigma - \epsilon$ curves of SS1 and SS2 at a specific time instant includes MAP, lower- and upper-bound values. We see that the MAP, lower- and upper-bound curves, especially in Fig. 18(b), are capable of capturing the measured curve from each loading condition.

6. Conclusions and opportunities for future research

We have presented the BO-assisted ABC approach and its application to identifying approximate posteriors of the parameters for known statistical models as well as the cyclic elastoplastic parameters for structural steels from noise-corrupted experimental datasets. The approach shows its potential for the structural engineering community as it can facilitate Bayesian inference of parameters characterizing a costly FE model to which the associated likelihood is often analytically or computationally intractable. At the heart of the approach, ABC addresses the computationally intractable likelihood, while BO embedded in ABC handles the costly simulation for intelligently reducing uncertainty in the approximation of ABC posterior. The main conclusions are summarized as follows:

- Results from the illustrative examples show that the BO-assisted ABC approach successfully offers approximate posteriors that can not only reproduce a given posterior with acceptable accuracy but also capture the true nominal parameters of a known statistical model. They also confirm the prediction ability of approximate posteriors of material parameters in modeling the cyclic elastoplastic behavior of a steel specimen tested under different static cyclic loading conditions. These results are brought by the performance of BO in refinement of the GP model approximating the discrepancy mean function in important regions of parameter space that contribute to the improvement of ABC posterior approximation.
- The ABC posterior can be more accurate for a smaller discrepancy threshold because a small discrepancy threshold, with a sufficiently large number of MCMC samples, can lead to the acceptance of good samples from the prior as samples constituting the approximate posterior. However, the approximate posterior may cease to improve when the discrepancy threshold is reduced and reaches a small-enough value.
- The accuracy of approximate posteriors is affected by the acquisition function used in each BO-assisted ABC attempt. However, it is not clear to indicate which acquisition function provides the best approximate posterior from the presented identification results. Switching the acquisition function from LCB to VAR during an identification attempt may improve the accuracy of approximate posteriors when a proper value of discrepancy threshold is used.
- The new samples associated with different levels of discrepancy threshold are identical when BO is guided by PI, EI, or LCB from the same initial training dataset. Hence, refinement schemes are the same for different levels of discrepancy threshold when PI, EI, or LCB is the acquisition function of interest. This facilitates both the investigation of how different levels of discrepancy threshold affect

the accuracy of the approximate posteriors and the selection of a proper discrepancy threshold.

Since this study is still in the first stage of developing a robust posterior approximation method to facilitate Bayesian inference of costly FE models, its scope is outside the realm of some important aspects that require further research.

- It is desirable to develop a scalable method for BO over high-dimensional parameter spaces because BO often limits itself to problems with less than 20 parameters [32].
- Since the quality of approximate posterior samples depends on the robustness of posterior sampling techniques used after the refinement process, it is needed to investigate the performance of different sampling methods such as the adaptive MCMC [44], sequential MCS [18], ensemble MCMC [19], and transitional MCMC [20].
- It is important to characterize robust priors to ensure the consistency of approximate posteriors. For this, hierarchical priors may be a viable choice. Robust Bayesian analysis [52] over a set of different priors can also be adopted to handle the sensitivity of posterior to the prior selection.
- Post-processing of parameters characterizing the approximate posterior may be needed to further consider the discrepancy between simulated and measured summary statistics. This process adjusts the obtained approximate posteriors based on a regression model, such as a local linear or nonlinear model [28].
- Modeling errors can be incorporated by defining the standard deviation τ in Eq. (1) as a function of material parameters $\mathbf{x}$, observational noise, and other parameters representing uncertainty due to idealization and discretization.
- The problem of selecting an appropriate constitutive law may be an opportunity for future research because the ABC approach has been widely used for selecting the best model among a finite number of candidate models [27], while many cyclic elastoplastic constitutive laws are available.

Acknowledgments

This work was supported by the JICA AUN/SEED-Net project. We gratefully thank Prof. Satoshi Yamada at The University of Tokyo for providing the experimental results used in Section 5.3.

Appendix A: Nonlinear combined isotropic/kinematic hardening model

Let $\boldsymbol{\sigma}$, $\boldsymbol{\alpha}$, and σ_y denote the stress tensor at a point of the steel specimen, the back-stress tensor, and the yield stress, respectively. Assume that our focus is on a small deformation problem without large rigid body

rotations. The relationship between strain and stress states of structural steel is described by its current status that is either elastic or plastic, and can be detected using the following von Mises yield criterion:

$$F = \sqrt{\frac{3}{2}}(\boldsymbol{\xi} : \boldsymbol{\xi})^{1/2} - \sigma_y = 0, \quad (\text{A.1})$$

where $\boldsymbol{\xi} = \text{dev}[\boldsymbol{\sigma}] - \text{dev}[\boldsymbol{\alpha}]$ is the shifted-stress tensor, $\text{dev}[\cdot]$ denotes the deviatoric part of $[\cdot]$, and the symbol “:” is the contraction of two second-order tensors.

We commonly use nonlinear isotropic and kinematic hardening models to describe the strain hardening process of structural steels subjected to cyclic loading. The isotropic hardening model only increases the size of the yield surface $F = 0$ during the evolution of plastic deformation. As a result, $\boldsymbol{\alpha}$ does not appear in Eq. (A.1), leading the yield surface to be an isotropic function of the stress that cannot capture the Bauschinger effect [53]. As structural steels exhibit a saturation point of the stress at large deformation, the isotropic hardening model can describe the increment of the yield surface size using the following Voce hardening law [54]:

$$\sigma_y = \sigma_{y,0} + Q_\infty[1 - \exp(-b\epsilon_{\text{eq}}^p)], \quad (\text{A.2})$$

where $\sigma_{y,0}$ denotes the initial yield stress, Q_∞ is the difference of the stress saturation and $\sigma_{y,0}$, b is the isotropic saturation rate, and ϵ_{eq}^p is the current equivalent plastic strain determined based on its previous state and the rate $\dot{\epsilon}_{\text{eq}}^p$.

The kinematic hardening model does not change the size and shape of the yield surface. Instead, it updates the center of the yield surface using a rigid translation in the evolution direction of plastic strain. This enables the kinematic hardening model to capture the Bauschinger effect. The back-stress tensor $\boldsymbol{\alpha}$ can be further decomposed into n_k back-stress components for a better approximation as [55]

$$\boldsymbol{\alpha} = \sum_{k=1}^{n_k} \boldsymbol{\alpha}_k, \quad (\text{A.3})$$

where the rate of $\boldsymbol{\alpha}_k$ can be described by a the following nonlinear kinematic hardening rule [56]:

$$\dot{\boldsymbol{\alpha}}_k = \sqrt{\frac{2}{3}}C_k\dot{\epsilon}_{\text{eq}}^p\mathbf{n} - \gamma_k\dot{\epsilon}_{\text{eq}}^p\boldsymbol{\alpha}_k, \quad (\text{A.4})$$

where $\mathbf{n} = \boldsymbol{\xi}/\|\boldsymbol{\xi}\|$ denotes the unit normal vector of the yield surface; and C_k and γ_k are the translation and relaxation rates of $\boldsymbol{\alpha}_k$, respectively.

The nonlinear combined isotropic/kinematic hardening model was developed using both Eqs. (A.2) and (A.4). Let $\mathbf{x} = [x_1, \dots, x_{n_p}]^\top \in \mathbb{R}^{n_p}$ denote the vector of n_p material parameters for this combined hardening model. Thus, $\mathbf{x} = [E, \nu, Q_\infty, b, \sigma_{y,0}, C_1, \gamma_1]^\top$ if we use one back-stress component in Eq. (A.3), where E and ν are Young’s modulus and Poisson’s ratio of the material, respectively.

Appendix B: Gaussian process modeling

Consider the training dataset $\mathcal{D} = \{\mathbf{X}, \mathbf{f}\} = \{\mathbf{x}^n, f^n\}_{n=1}^N$, where $\mathbf{x}^n$ are the vectors of parameters and f^n are the corresponding values of the function to be estimated. Based on $\mathcal{D}$, we wish to establish the mapping from $\mathbf{x}$ to the function value Δ , i.e., $\Delta = \hat{f}(\mathbf{x}) + \omega_f$, where $\hat{f}(\mathbf{x})$ is a Gaussian conditioned on $\mathcal{D}$ and $\omega_f \sim \mathcal{N}(0, \omega^2)$.

GP modeling assumes that any finite subset of an infinite set of f values has a joint Gaussian distribution [57]. Thus, for the set $\{\mathbf{x}^1, \dots, \mathbf{x}^N\}$, the corresponding values $\{f^1, \dots, f^N\}$ are distributed according to

$$\begin{bmatrix} f^1 \\ \vdots \\ f^N \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(\mathbf{x}^1) \\ \vdots \\ m(\mathbf{x}^N) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}^1, \mathbf{x}^1) & \cdots & k(\mathbf{x}^1, \mathbf{x}^N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}^N, \mathbf{x}^1) & \cdots & k(\mathbf{x}^N, \mathbf{x}^N) \end{bmatrix} \right), \quad (\text{B.1})$$

where $m(\mathbf{x}) = \mathbb{E}[\hat{f}(\mathbf{x})]$ and $k(\mathbf{x}, \mathbf{x}')$ are the mean and covariance kernel functions, respectively. $k(\mathbf{x}, \mathbf{x}')$ is defined for any pair of the parameter vectors $\mathbf{x}$ and $\mathbf{x}'$ to measure the similarity between $f = \hat{f}(\mathbf{x})$ and $f = \hat{f}(\mathbf{x}')$, such that

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(\hat{f}(\mathbf{x}) - m(\mathbf{x}))(\hat{f}(\mathbf{x}') - m(\mathbf{x}'))]. \quad (\text{B.2})$$

In this study, we use the squared exponential kernel as

$$k(\mathbf{x}, \mathbf{x}') = \exp \left(-\frac{(\mathbf{x} - \mathbf{x}')^\top (\mathbf{x} - \mathbf{x}')}{2l^2} \right), \quad (\text{B.3})$$

where l denotes the characteristic length-scale parameter that is determined by using the maximum likelihood estimation of $\mathcal{D}$ [57]. Note that an optimal value of ω can be found in the same way if the distribution of $\{\Delta^1, \dots, \Delta^N\}$ is derived from Eq. (B.1) and $\Delta = \hat{f}(\mathbf{x}) + \mathcal{N}(0, \omega^2)$.

After l is determined, we use the information in Eq. (B.1) for predicting the function value f^* at a new parameter vector $\mathbf{x}^*$, i.e., $f^*|\mathbf{f} = \hat{f}(\mathbf{x}^*)$. As the GP nature, the joint PDF of f^* and $\mathbf{f}$ is also a Gaussian as

$$\begin{bmatrix} f^* \\ \mathbf{f} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(\mathbf{x}^*) \\ \mathbf{m}(\mathbf{X}) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}^*, \mathbf{x}^*) & \mathbf{K}(\mathbf{x}^*, \mathbf{X}) \\ \mathbf{K}(\mathbf{x}^*, \mathbf{X})^\top & \mathbf{K}(\mathbf{X}, \mathbf{X}) \end{bmatrix} \right), \quad (\text{B.4})$$

where $\mathbf{m}(\mathbf{X}) = [m(\mathbf{x}_1), \dots, m(\mathbf{x}_N)]^\top$, and

$$\begin{aligned} \mathbf{K}(\mathbf{x}^*, \mathbf{X}) &= [k(\mathbf{x}^*, \mathbf{x}^1), \dots, k(\mathbf{x}^*, \mathbf{x}^N)], \\ \mathbf{K}(\mathbf{X}, \mathbf{X}) &= \begin{bmatrix} k(\mathbf{x}^1, \mathbf{x}^1) & \cdots & k(\mathbf{x}^1, \mathbf{x}^N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}^N, \mathbf{x}^1) & \cdots & k(\mathbf{x}^N, \mathbf{x}^N) \end{bmatrix}. \end{aligned} \quad (\text{B.5})$$

The conditional Gaussian variable $f^*|\mathbf{f} = \hat{f}(\mathbf{x}^*)$ can be derived from Eq. (B.4) using the standard conditioning rule [57], such that

$$f^*|\mathbf{f} = \hat{f}(\mathbf{x}^*) \sim \mathcal{N}(f|\mu_{f^*}(\mathbf{x}^*), \tau_{f^*}^2(\mathbf{x}^*)), \quad (\text{B.6})$$

where

$$\begin{aligned}\mu_{f^*}(\mathbf{x}^*) &= m(\mathbf{x}^*) + \mathbf{K}(\mathbf{x}^*, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}(\mathbf{f} - \mathbf{m}(\mathbf{X})), \\ \tau_{f^*}^2(\mathbf{x}^*) &= k(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{K}(\mathbf{x}^*, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{K}(\mathbf{x}^*, \mathbf{X})^\top.\end{aligned}\tag{B.7}$$

References

- [1] A. Tarantola, Inverse problem theory and methods for model parameter estimation, SIAM, 2005.
- [2] S. Arridge, P. Maass, O. Öktem, C.-B. Schönlieb, Solving inverse problems using data-driven models, *Acta Numerica* 28 (2019) 1–174.
- [3] O. Ghattas, K. Willcox, Learning physics-based models from data: perspectives from inverse problems and model reduction, *Acta Numerica* 30 (2021) 445–554.
- [4] J. L. Mueller, S. Siltanen, Linear and nonlinear inverse problems with practical applications, SIAM, 2012.
- [5] M. Benning, M. Burger, Modern regularization methods for inverse problems, *Acta Numerica* 27 (2018) 1–111.
- [6] G. Johansson, J. Ahlström, M. Ekh, Parameter identification and modeling of large ratcheting strains in carbon steel, *Computers & Structures* 84 (15) (2006) 1002–1011.
- [7] B. M. Chaparro, S. Thuillier, L. F. Menezes, P. Y. Manach, J. V. Fernandes, Material parameters identification: Gradient-based, genetic and hybrid optimization algorithms, *Computational Materials Science* 44 (2) (2008) 339–346.
- [8] R. de Carvalho, R. A. F. Valente, A. Andrade-Campos, Optimization strategies for non-linear material parameters identification in metal forming problems, *Computers & Structures* 89 (1) (2011) 246–255.
- [9] R. Hartloper Alexander, de Castro e Sousa Albano, G. Lignos Dimitrios, Constitutive modeling of structural steels: nonlinear isotropic/kinematic hardening material model and its calibration, *Journal of Structural Engineering* 147 (4) (2021) 04021031.
- [10] A. H. Mahmoudi, S. M. Pezeshki-Najafabadi, H. Badnava, Parameter determination of Chaboche kinematic hardening model using a multi objective genetic algorithm, *Computational Materials Science* 50 (3) (2011) 1114–1122.
- [11] B. Do, M. Ohsaki, Bayesian optimization for inverse identification of cyclic constitutive law of structural steels from cyclic structural tests, *Structures* 38 (2022a) 1079–1097.
- [12] B. Do, M. Ohsaki, Proximal-exploration multi-objective Bayesian optimization for inverse identification of cyclic constitutive law of structural steels, *Structural and Multidisciplinary Optimization* 65 (7) (2022b) 199.
- [13] M. Ohsaki, B. Do, J. Fujiwara, T. Kimura, T. Yamashita, Two-step parameter identification of multi-axial cyclic constitutive law of structural steels from cyclic structural responses, *Structures* 46 (2022) 2014–2030.
- [14] J. L. Beck, L. S. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, *Journal of Engineering Mechanics* 124 (4) (1998) 455–461.
- [15] Y. M. Marzouk, H. N. Najm, L. A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, *Journal of Computational Physics* 224 (2) (2007) 560–586.
- [16] B. V. Rosić, A. Kučerová, J. Sýkora, O. Pajonk, A. Litvinenko, H. G. Matthies, Parameter identification in a probabilistic setting, *Engineering Structures* 50 (2013) 179–196.
- [17] D. Gamerman, H. F. Lopes, Markov chain Monte Carlo: stochastic simulation for Bayesian inference, CRC press, 2006.
- [18] J. S. Liu, R. Chen, Sequential Monte Carlo methods for dynamic systems, *Journal of the American Statistical Association* 93 (443) (1998) 1032–1044.
- [19] J. Goodman, J. Weare, Ensemble samplers with affine invariance, *Communications in applied mathematics and computational science* 5 (1) (2010) 65–80.

- [20] C. Jianye, C. Yi-Chu, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, *Journal of Engineering Mechanics* 133 (7) (2007) 816–832.
- [21] S. A. Sisson, Y. Fan, M. Beaumont, *Handbook of approximate Bayesian computation*, CRC Press, 2018.
- [22] H. Rappel, L. A. A. Beex, J. S. Hale, L. Noels, S. P. A. Bordas, A tutorial on Bayesian inference to identify material parameters in solid mechanics, *Archives of Computational Methods in Engineering* 27 (2) (2020) 361–385.
- [23] N. Noii, A. Khodadadian, J. Ulloa, F. Aldakheel, T. Wick, S. François, P. Wriggers, Bayesian inversion with open-source codes for various one-dimensional model problems in computational mechanics, *Archives of Computational Methods in Engineering* (2022).
- [24] H. Wang, Y. Zeng, X. Yu, G. Li, E. Li, Surrogate-assisted Bayesian inference inverse material identification method and application to advanced high strength steel, *Inverse Problems in Science and Engineering* 24 (7) (2016) 1133–1161.
- [25] P. Fernandez-Zelaia, V. Roshan Joseph, S. R. Kalidindi, S. N. Melkote, Estimating mechanical properties from spherical indentation using Bayesian approaches, *Materials & Design* 147 (2018) 92–105.
- [26] E. Janouchová, A. Kučerová, J. Sýkora, J. Vorel, R. Wan-Wendner, Robust probabilistic calibration of a stochastic lattice discrete particle model for concrete, *Engineering Structures* 236 (2021) 112000.
- [27] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, M. P. H. Stumpf, Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, *Journal of The Royal Society Interface* 6 (31) (2009) 187–202.
- [28] J.-M. Marin, P. Pudlo, C. P. Robert, R. J. Ryder, Approximate Bayesian computational methods, *Statistics and Computing* 22 (6) (2012) 1167–1180.
- [29] B. M. Turner, T. Van Zandt, A tutorial on approximate Bayesian computation, *Journal of Mathematical Psychology* 56 (2) (2012) 69–85.
- [30] D. R. Jones, M. Schonlau, W. J. Welch, Efficient global optimization of expensive black-box functions, *Journal of Global Optimization* 13 (4) (1998) 455–492.
- [31] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, N. de Freitas, Taking the human out of the loop: A review of Bayesian optimization, *Proceedings of the IEEE* 104 (1) (2016) 148–175.
- [32] P. I. Frazier, A tutorial on Bayesian optimization, *arXiv preprint* (2018). [arXiv:1807.02811](https://arxiv.org/abs/1807.02811).
- [33] A. Ben Abdesslem, N. Dervilis, D. Wagg, K. Worden, Model selection and parameter estimation in structural dynamics using approximate Bayesian computation, *Mechanical Systems and Signal Processing* 99 (2018) 306–325.
- [34] A. Lye, M. Kitahara, M. Broggi, E. Patelli, Robust optimization of a dynamic Black-box system under severe uncertainty: A distribution-free framework, *Mechanical Systems and Signal Processing* 167 (2022) 108522.
- [35] M. U. Gutmann, J. Corander, Bayesian optimization for likelihood-free inference of simulator-based statistical models, *Journal of Machine Learning Research* 17 (125) (2016) 1–47.
- [36] M. Järvenpää, M. U. Gutmann, A. Pleska, A. Vehtari, P. Marttinen, Efficient acquisition rules for model-based approximate Bayesian computation, *Bayesian Analysis* 14 (2) (2019) 595–622.
- [37] B. Do, Probabilistic Approaches to Optimization of Steel Structures Considering Uncertainty, Ph.D. dissertation, Kyoto University (2023).
- [38] M. A. Beaumont, Approximate Bayesian Computation, *Annual Review of Statistics and Its Application* 6 (1) (2019) 379–403.
- [39] M. Sunnåker, A. G. Busetto, E. Numminen, J. Corander, M. Foll, C. Dessimoz, Approximate Bayesian Computation, *PLOS Computational Biology* 9 (1) (2013) e1002803.
- [40] T. J. Santner, B. J. Williams, W. I. Notz, *The design and analysis of computer experiments*, 2nd Edition, Springer, New York, NY, 2018.

- [41] H. J. Kushner, A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, *Journal Basic Engineering* 86 (1) (1964) 97–106.
- [42] N. Srinivas, A. Krause, S. M. Kakade, M. Seeger, Gaussian process optimization in the bandit setting: No regret and experimental design, in: *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 1015–1022.
- [43] M. Patefield, Fast and accurate calculation of Owen’s T function, *Journal of Statistical Software* 5 (5) (2000) 1 – 25.
- [44] H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis algorithm, *Bernoulli* 7 (2) (2001) 223–242.
- [45] H. Haario, M. Laine, A. Mira, E. Saksman, DRAM: Efficient adaptive MCMC, *Statistics and Computing* 16 (4) (2006) 339–354.
- [46] S. A. Sisson, Y. Fan, ABC samplers, in: *Handbook of approximate Bayesian computation*, Chapman and Hall/CRC, 2018, pp. 87–123.
- [47] A. Lye, A. Cicirello, E. Patelli, Sampling methods for solving Bayesian model updating problems: A tutorial, *Mechanical Systems and Signal Processing* 159 (2021) 107760.
- [48] C. C. Drovandi, A. N. Pettitt, Likelihood-free Bayesian estimation of multivariate quantile distributions, *Computational Statistics & Data Analysis* 55 (9) (2011) 2541–2556.
- [49] S. Yamada, Y. Jiao, A concise hysteretic model of structural steel considering the Bauschinger effect, *International Journal of Steel Structures* 16 (3) (2016) 671–683.
- [50] M. Ohsaki, T. Miyamura, J. Y. Zhang, A piecewise linear isotropic-kinematic hardening model with semi-implicit rules for cyclic loading and its parameter identification, *Computer Modeling in Engineering & Sciences* 111 (4) (2016) 303–333.
- [51] Dassault Systèmes, Abaqus User’s Manual Ver. 2020 (2020).
- [52] I. Raices Cruz, J. Lindström, M. C. M. Troffaes, U. Sahlin, Iterative importance sampling with Markov chain Monte Carlo sampling in robust Bayesian analysis, *Computational Statistics & Data Analysis* 176 (2022) 107558.
- [53] J. Lemaitre, J.-L. Chaboche, *Mechanics of solid materials*, Cambridge University Press, 1994.
- [54] E. Voce, The relationship between stress and strain for homogeneous deformation, *Journal of the Institute of Metals* 74 (1948) 537–562.
- [55] J. L. Chaboche, G. Rousselier, On the plastic and viscoplastic constitutive equations—Part I: Rules developed with internal variable concept, *Journal of Pressure Vessel Technology* 105 (2) (1983) 153–158.
- [56] P. J. Armstrong, C. O. Frederick, A mathematical representation of the multiaxial Bauschinger effect, Report RD/B/N731, Berkeley, UK (1966).
- [57] C. E. Rasmussen, C. K. I. Williams, *Gaussian processes for machine learning*, The MIT Press, Cambridge, Massachusetts, 2006.