Efficient estimation of material property curves and surfaces via active learning

Yuan Tian ,[1] Dezhen Xue ,[1,*] Ruihao Yuan,[1] Yumei Zhou,[1,†] Xiangdong Ding,[1] Jun Sun,[1] and Turab Lookman[2,‡]

[1]*State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an 710049, China*
[2]*Los Alamos National Laboratory, Los Alamos, New Mexico 87545,USA*

The relationship between material properties and independent variables such as temperature, external field, or time is usually represented by a curve or surface in a multidimensional space. Determining such a curve or surface requires a series of experiments or calculations which are often time and cost consuming. A general strategy uses an appropriate utility function to sample the space to recommend the next optimal experiment or calculation within an active learning loop. However, knowing what optimal sampling strategy to use to minimize the number of experiments is an outstanding problem. We compare a number of strategies based on directed exploration on several materials problems of varying complexity using a Kriging-based model. These include one-dimensional curves such as the fatigue life curve for 304L stainless steel and the Liquidus line of the Fe-C phase diagram, surfaces such as the Hartmann 3 function in three-dimensional space and the fitted intermolecular potential for Ar-SH, and a four-dimensional data set of experimental measurements for $BaTiO_3$-based ceramics. We also consider the effects of experimental noise on the Hartmann 3 function. We find that directed exploration guided by maximum variance provides better performance overall, converging faster across several data sets. However, for certain problems, the tradeoff methods incorporating exploitation can perform at least as well, if not better than maximum variance. Thus, we discuss how the choice of the utility function depends on the distribution of the data, the model performance and uncertainties, additive noise, as well as the budget.

I. INTRODUCTION

The accurate prediction of the properties of materials as a function of independent variables is crucially important in exploiting their use in different applications. Such a functional relationship is usually described as a curve or surface between a property and the independent variable(s) in a multidimensional diagram [1]. These properties can be mechanical, thermal, electrical, magnetic, optical, and chemical, and the independent variables often include chemical composition, temperature, time, and heat treatment conditions [2]. The material property curves and surfaces determine the critical states and property optima, and consequently play a crucial role in the design of new materials, the assessment of hazards, and the optimization of processing parameters. Familiar examples include phase boundaries and surfaces in temperature versus composition space, fatigue life cycle curves describing the relationship between mechanical properties and loading cycles, and intermolecular potential energy surfaces for molecules.

Determining a property curve or surface is often time and cost consuming as a number of measurements or calculations are required depending on the accuracy needed. An adequate number of data needs to be accumulated as the independent variable is varied in given steps. The data requirements are sensitive to nonlinearities and sharp changes in the functional form as well as the presence and number of multiple extrema, including critical points. For example, establishing a phase diagram requires a series of experiments to determine the critical temperature for different compositions or pressures. Similarly, a number of parallel samples, each of which is used to obtain the ultimate stress for a given number of loading cycles, are required to obtain the fatigue curve of an alloy. Although regression algorithms have been employed to model the functional form between the property and the independent variables [3–6], the results of regression inevitably contain large uncertainties if the number of initial data points is relatively small, especially if the relationship between the property and independent variable is complex. Hence, there is a need for an approach that can predict general property curves and surfaces and successively refine them rapidly using as few new measurements or calculations as possible.

Active learning allows an algorithm to choose the data from which it learns so that it may learn more efficiently with less training data than otherwise [7–12]. This becomes particularly important in materials science where the size of a good quality labeled data set for supervised learning is often limited because of the expense associated with generating it [7,13–23]. Utility function in the active learning loop provides the criterion to sample the search space to recommend one or more unlabeled instances which would be most informative to be labeled by experiments or calculations [24–26]. Sampling the most important states is therefore key to avoiding excessive numbers of iterations or experiments, especially if it is unknown which states are most important.

————
*xuedezhen@xjtu.edu.cn
†zhouyumei@xjtu.edu.cn
‡turablookman@gmail.com

This requires exploring the allowed space efficiently, which has been studied in the context of reinforcement learning using exploration techniques, such as undirected and directed exploration [27]. Undirected exploration is uninformed and characterized by selecting actions randomly from a given distribution [28]. Directed exploration uses knowledge to guide the exploration search so that the exploration rule directly determines which action to take next. The goal is to select actions which maximize the improvement over time, which is impossible to determine as we do not know in advance how a given decision will improve the performance. Thus, all directed exploration techniques are heuristics designed to optimize knowledge gain. The exploration may be achieved by choosing states based on frequency of occurrence (counter-based exploration), or assumed to have a high prediction error (maximum variance), or those that include different degrees of exploitation functions based on using the best value of the model predictions at the time. Examples of the latter include tradeoff methods such as the efficient global optimization (EGO) and knowledge gradient (KG) schemes based on Bayesian optimization for finding maxima and minima of functions [24,29–34].

How much exploration needs to be performed depends on the costs of collecting new information and the value associated with that information. In the absence of analytical results for realistic problems and strategies, we need to compare different strategies on different sets of data of varying sizes and distributions to evaluate their relative performance as a function of the dimensionality of the problem as well as the influence of measurement noise. Here we compare the efficiencies of six utility functions to estimate material properties in terms of the number of new experiments required for each. Since selection via maximum variance (Max-v) is one of our utilities, we introduce the utility B.EGO, designed to search for the option with the maximum of the variability in the function over many bootstrap samples. This is in contrast to finding the maximum and minimum of the function that EGO and KG have so far been applied to. The uncertainties are given by a Kriging model and used in evaluating Max-v, B.EGO, EGO, KG, random exploration using a uniform distribution, and sequential Kriging optimization (SKO) to recommend the next candidate. The last utility considers the effects of experimental noise on the data.

We apply our approach to several problems with increasing complexity to determine which utilities are robust across all of them. As it is common to predict property curves from limited data, we examine two applications, the fatigue life curve for 304L stainless steel (SS304L) and estimating the Liquidus line of the Fe-C phase diagram. We show that two or three new experiments or calculations are all that is needed to complete the curve optimization. We study surfaces in the form of the three-dimensional (3D) Hartmann 3 function used in optimization tests, to which we also add experimental noise to study utility performance for noisy data, and the fitted surface for the intermolecular potential of Ar-SH. Finally, we apply our tests to a data set of experimental measurements for the Curie temperature of $BaTiO_3$ ferroelectrics modeled in four dimensions (4D).

Our principal conclusion is that for a range of materials data and problems with varying complexities, directed explo-

ration via maximum variance generally performs better than other utilities in mapping the property curve. The variability utility B.EGO based on bootstrap samples is also a good performer, following Max-v. However, for given problems, the tradeoff methods that add various degrees of exploitation can perform at least as well, if not better, than maximum variance Max-v. Thus, the choice of the utility function is sensitive to the distribution of the training and subsequently acquired data, the model performance, the noise as well as the budget, which determines the number of iterations allowed.

II. ACTIVE LEARNING STRATEGY

Figure 1 illustrates our active learning loop. We begin with a Kriging model that uses regression to estimate a property curve, i.e., $y = f(x)$, from the relatively small number of labeled data points available. The uncertainties associated with the estimation will be large due to limited data. The Kriging model predicts a value μ for each point in the curve as well as the variance of the prediction s^2 at that point. The utility functions are defined in terms of μ and s^2 and recommend the next unlabeled point for the curve for which the label is evaluated by the user via experiments or calculations. The variance serves as input to Max-v as well as B.EGO. The latter is defined in terms of the bootstrap mean error $\bar{s}$ and its standard deviation $se(s)$. The new point selected then augments the training data so that the regressor can be refined and provide an updated value for μ and s^2. The loop then continues so that the curve can be improved step by step until an adequate estimate of the curve is obtained.

A. Model: Gaussian process via Kriging

In Kriging [35–37], the spatial interpolated values are modeled by a Gaussian process governed by prior covariances. It is customary to consider noisy observations of the targeted property y, where $\tilde{y}_j = y(x_j) + \varepsilon_j$ and ε_j is a realization of a random variable so that ε_j follows an independent, identically distributed Gaussian distribution $\mathcal{N}(0, \tau_j^2)(1 \leqslant j \leqslant n)$ with homogeneous noise variance τ_j^2.

The property y is considered a realization of a Gaussian process Y following Kriging. That is,

$$Y = \mathbf{m} + Z = \sum \beta \mathbf{f} + Z, \qquad (1)$$

where $\mathbf{m}$ is a trend function, β is the coefficient, and the process Z is assumed Gaussian.

Assuming p training data x^*, with unknown data points x, the universal Kriging (UK) equations are given by [38,39]

$$\mu = \mathbf{m}(x) + K(x, x^*)(K + \Delta)^{-1}[\tilde{y} - \mathbf{m}(x^*)], \qquad (2)$$

$$s^2 = s_{SK}^2 + [\mathbf{f}(x)^\mathsf{T} - K(x, x^*)^\mathsf{T}(K + \Delta)^{-1}\mathbf{f}(x^*)]^\mathsf{T}[\mathbf{f}(x^*)^\mathsf{T}$$
$$\times (K + \Delta)^{-1}\mathbf{f}(x^*)]^{-1}[\mathbf{f}(x)^\mathsf{T}$$
$$- K(x, x^*)^\mathsf{T}(K + \Delta)^{-1}\mathbf{f}(x^*)], \qquad (3)$$

where $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_p)^T$, K is covariance between training data points, Δ is a diagonal matrix with diagonal terms $\tau_1^2, \ldots, \tau_p^2$. The simple Kriging (SK) variance s_{SK}^2 is given by

$$s_{SK}^2 = K(x, x) - K(x, x^*)(K + \Delta)^{-1}K(x^*, x). \qquad (4)$$
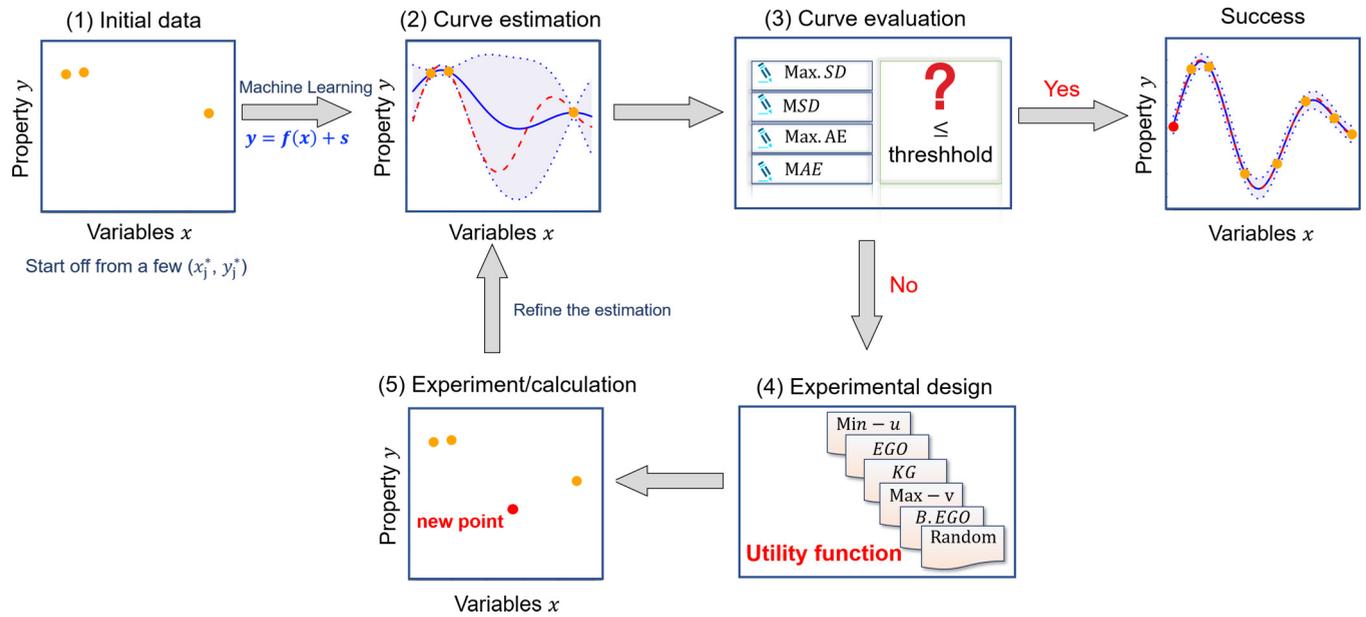
FIG. 1. Flow chart of active learning strategy for efficient estimation of material property curves. (1) A small, labeled training data serves as the starting point for the targeted curve. (2) A surrogate model using machine learning methods provides an estimate of the fit with uncertainties. (3) Several criteria are used to check if the estimated fit is adequate and fulfills conditions for success. If not, (4) utility functions are evaluated and ranked to recommend the next data point to use from the pool of possible points for measurement or calculation to determine the label. (5) The recommended experiment or calculation is performed and the new labeled point augments the training data to obtain a revised estimate of the curve. The loop continues until the criteria for success are met.

We use the covariance kernel $g(h) = \exp[-\frac{1}{2}(\frac{h}{\theta})^2]$, where h and θ are hyper-parameters of the model and set characteristic length scales associated with the data. Note that the variance value s^2 at x depends on the distance from known point x^*. If x is close to known point x^*, it is influenced by x^* and the variance at x will be small. If x is separated from known points, the variance at x will be large.

B. Evaluating goodness of fit for the targeted curve

We determine the quality of the model by tracking the deviation of the regressed curve from the true curve. In our testing case, as the true curve is known, we can use the mean absolute error (MAE) and maximum absolute error (Max.AE) defined by

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}(|y_j - \mu_j|), \qquad (5)$$

$$\text{Max.AE} = \max(|y_j - \mu_j|), \qquad (6)$$

where n is the total number of possible points in the function, y_j are the true values, and μ_j are the estimated values from Kriging model. The error MAE is the average deviation of the estimate value from the true value whereas Max.AE is the largest error over the range of data points.

As the true curve is usually not known, we utilize the uncertainty associated with the regressor prediction as an estimate of the model quality. We thus use instead the mean standard deviation (MSD) and the maximum standard deviation

(Max.SD) defined as follows:

$$\text{MSD} = \frac{1}{n}\sum_{j=1}^{n}(s_j), \qquad (7)$$

$$\text{Max.SD} = \max(s_j), \qquad (8)$$

where s_j is the standard deviation associated with each prediction (μ_j) in the curve. We will monitor the evolution of MAE, Max.AE, MSD, and Max.SD as we iterate the active learning loop until the accuracy threshold is reached.

C. Utility functions

Small data sets, characteristics of many materials science problems, typically give rise to large uncertainties in prediction and therefore additional statistical design criteria need to be invoked. A utility function allows us to choose by maximizing an expected utility, where the utilities are defined with respect to information-theoretic considerations. By ranking the expected value of the information for possible alternatives for observation or calculation, the utility function provides the means to prioritize the decision-making process based on the information gained or reduced by observing a potential new data point [13,40]. The utilities we compare in this work are defined below; in the noise case we employ SKO instead of EGO.

Min-u. Min-u is a greedy choice in which the candidate with the lowest predicted mean value from the model is chosen. That is,

$$\nu_{\text{Min-u}} = \mu. \qquad (9)$$

Max-v. The variability in the predictions can be characterized by the variance at that point obtained from the Kriging covariance equation (3). Thus, Max-v is a risk-averse utility function selecting the next candidate point based on the magnitude of the variance in the property at a given point. That is,

$$\nu_{\text{Max-v}} = s^2. \tag{10}$$

EGO. Efficient global optimization (EGO) balances exploration and exploitation by evaluating the "expected improvement" (EI) and choosing the candidate with the largest (EI). If $\tilde{y}^*_{min}$ is the minimum value in the training data, the improvement at a point x_j is $I = \max(\tilde{y}^*_{min} - Y_j, 0)$, where Y_j is distributed normally, $\mathcal{N}(\mu_j, s_j^2)$. As the tail of the density function at point x_j extends into $\tilde{y}^*_{min}$, improvement is then possible. Different amounts of improvement or distances from $\tilde{y}^*_{min}$ are associated with different density values. The EI is obtained by weighting all these improvement values by the associated density values. The EI of each potential measurement is the expectation of I at that point given by [41]

$$\nu_{\text{EGO}} = E[\max(\tilde{y}^*_{min} - Y_j, 0)] = s\mathcal{G}\left(\frac{\tilde{y}^*_{min} - \mu}{s}\right)$$
$$= (\tilde{y}^*_{min} - \mu)\Phi\left(\frac{\tilde{y}^*_{min} - \mu}{s}\right) + s\phi\left(\frac{\tilde{y}^*_{min} - \mu}{s}\right), \tag{11}$$

where $\mathcal{G}(z_0) = z_0\Phi(z_0) + \phi(z_0)$, $z_0 = \frac{\tilde{y}^*_{min} - \mu}{s}$, s is the standard deviation associated with the mean value μ of the model prediction, $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, respectively. If the measurements are noise free, ν_{EGO} is zero at the sampled points (points that are already measured) and is positive elsewhere. In our discretized version of the problem here, EGO simply evaluates EI at each unexplored point and recommends a point with the largest EI to be measured next.

SKO. In the noise case, the current best estimate $\tilde{y}^*_{min}$ also suffers from noise, and the actual minimum is indeed unknown. We therefore utilize an extension of EGO, sequential Kriging optimization (SKO) [42], in which $\tilde{y}^*_{min}$ in EGO is modified through the model predictions μ^{**}. A prefactor $1 - \frac{\tau}{\sqrt{\tau^2 + s^2}}$ is introduced to enhance the exploration

$$\nu_{\text{SKO}} = \left(1 - \frac{\tau}{\sqrt{\tau^2 + s^2}}\right)s\mathcal{G}\left(\frac{\mu^{**} - \mu}{s}\right), \tag{12}$$

where $\mu^{**} = \mu(\text{argmin}[\mu + \lambda s])$, λ is the "risk-avoidance" parameter.

KG. Knowledge gradient (KG) aims at maximizing the current reward [43,44]. KG can be calculated using

$$\nu_{\text{KG}} = S\mathcal{G}\left(-\left|\frac{\mu' - \mu}{S}\right|\right), \tag{13}$$

where μ' is the minimum of the predicted values in the virtual unexplored space (μ_j) without the next recommended sample (μ_i), i.e., $\mu' = \min\mu_j$, for $j \neq i$. If the measurements are noise free, S is the standard deviation s provided by the Kriging model directly. KG is also available for the noise case and the modified standard deviation S is given by $S = s^2/\sqrt{\tau^2 + s^2}$, where s^2 is the variance of the prediction and τ^2 is the variance associated with noise [40,43,44].

B.EGO. In order to find the point with the maximum variability in the property using EGO or KG, we introduce a utility that complements Max-v. Whereas the variability in Max-v is directly obtained from the Kriging model for a given data set, the *variability* in B.EGO is obtained by considering many bootstrap samples. Let s^*_{max} be the largest so-far value of the standard error of the bootstrap uncertainties in the training data, so that $I = \max(\text{ERROR}_j - s^*_{\text{max}}, 0)$, where ERROR_j is distributed normally $\mathcal{N}(\bar{s}_j, \text{se}(s)_j^2)$. The mean $\bar{s}$ is given by $\frac{\sum_{b=1}^{B} s_b}{B}$, where B are the bootstrap replicates or samples, and se(s) is the standard error of the bootstrap uncertainties corresponding to $\bar{s}$. We use a value of B of 50. Then, the expected improvement EI of each potential measurement is the expectation of I at that point given by [41]

$$\nu_{\text{B.EGO}} = E[\max(\text{ERROR}_j - s^*_{\text{max}}, 0)]$$
$$= \text{se}(s)\mathcal{G}\left(\frac{\bar{s} - s^*_{\text{max}}}{\text{se}(s)}\right)$$
$$= (\bar{s} - s^*_{\text{max}})\Phi\left(\frac{\bar{s} - s^*_{\text{max}}}{\text{se}(s)}\right) + \text{se}(s)\phi\left(\frac{\bar{s} - s^*_{\text{max}}}{s}\right), \tag{14}$$

where $\mathcal{G}(z_0) = \int_{-\infty}^{z_0} \Phi(z)dz = z_0\Phi(z_0) + \phi(z_0)$, $z_0 = \frac{\bar{s} - s^*_{\text{max}}}{\text{se}(s)}$. B.EGO is designed to aim at searching for the point with the maximum expected improvement, however, the improvement is not for the current minimum function value but for the current maximum standard deviation of all labeled observations.

Random. This involves a random choice of the unmeasured candidate, such that if there are a total of N choices, x_i is chosen with probability $1/N$.

III. RESULTS

We present results for three classes of problems with varying complexity. We consider 1D and multidimensional cases with a finite number of measurements of relevance in materials science. These include the following. Case I: (a) the fatigue life cycle curve for 304L stainless steel (SS304L) and (b) the Liquidus line in the Fe-C phase diagram. Case II: (a) a standard optimization test function, the Hartmann 3 function in 3D, and (b) the fitted intermolecular potential surface for Ar-SH. Case III: measurements of the Curie temperature for ferroelectric samples with four variables or features. In case II(a) we also vary the data set size in the presence of "experimental" noise. In each case a small subset $\tilde{y}^*_j$ and x^*_j of the data set is randomly chosen as the initial training data and the remaining data $\tilde{y}_j$ and x_j comprise the unexplored search space. We implemented the feedback loop of Fig. 1, monitoring the departures from the true result (the optimizing function) given by MAE, Max.AE, MSD, and Max.SD defined previously. We employed Kriging to perform regression and once the new measurement x_i is selected by the utility function, the new observed value $\tilde{y}_i$ augments the training data and the loop repeats itself, refining successive estimates. We monitor the number of iterations (N) of the loop, i.e., number of new measurements made with a given utility that minimize (N). To garner adequate statistics, the design process was repeated 100 times with different initial training data randomly selected from all the discretized points in the data set.

A. Case study I: 1D materials cases

1. (a) Fatigue life curve for 304L stainless steel

Fatigue properties of materials are often described using the fatigue curve, which describes the relationship between cyclic stress and number of cycles to failure. It is critical in assessing material failure and obtaining it experimentally requires a series of tests to find the ultimate stress for a given number of loading cycles, which is quite time and cost consuming.

We choose a monotonic fatigue life curve for 304L stainless steel from the simulation work of Mozafari *et al.* [45] as a typical example to validate our design loop in Fig. 1. We consider the number of cycles to failure N_f as the independent variable and the stress amplitude σ_a as the output or property and discretize the curve into 201 data points. We randomly choose five data points as the initial training data with known x_j^* and $\tilde{y}_j^*$ to employ in the loop of Fig. 1 to optimize the curve. The next measurement is recommended from the remaining 196 data points using different utility functions. Typical examples of the optimization process for Max-v, Random compared to EGO and B.EGO from one initial data set are shown in Fig. S1 [46]. The function Max-v converges to the true function in only three new measurements, outperforming the other functions which need more measurements.

For a more robust comparison, we used an initial random training data set of $n = 5$ training points and tracked the values of MAE, Max.AE, MSD, and Max.SD as a function of number of new measurements for the different policies. By repeating over 100 trials, Figs. 2(a) and 2(b) show the average values and 95% confidence intervals for MAE and Max.AE for the different utility functions. The results of MSD and Max.SD are shown in Fig. S2 [46]. Both Max-v and our function B.EGO perform very well, converging in relatively few iterations followed by Random which also converges but with more iterations. The tradeoff methods EGO and KG decrease the error quickly in the first three iterations, but then relax slowly but nevertheless also converge. The greedy, pure exploitation Min-u shows very little relaxation after a few iterations.

2. (b) Liquidus line in the iron-carbon phase diagram

The iron-carbon (Fe-C) phase diagram displays the phases, compositions, and transformations in iron-carbon alloys as a result of heating and cooling, and therefore serves as the basis for composition design and optimizing heat treatment of steels. The liquidus line is the phase boundary in the phase diagram limiting the bottom of the liquid field, and the liquidus line exhibits a eutectic point at C composition of 4.3% between γ and Fe_3C. The temperature curve in the phase diagram is irregular and the challenge is to obtain it with as few measurements as possible. We discretize the liquidus line into 118 data points, i.e., 118 composition-temperature data points and randomly choose 5 initial points. The estimated curves initially deviate significantly from the true curve, which gives rise to large values of MAE, Max.AE, MSD, and Max.SD. We show in Fig. S3 [46] that the function Max-v only requires two new measurements to match the true function and outperforms all the other functions which need more measurements. The function B.EGO also does well in the optimization as it works
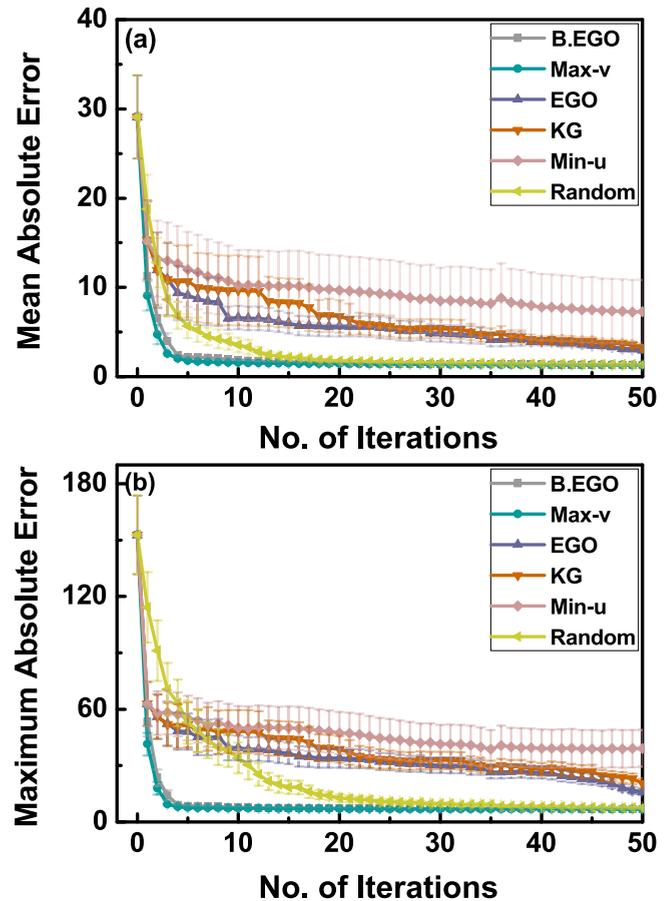
FIG. 2. Comparison of the performance of utility functions in optimizing the fatigue life curve for SS304L steel. The initial data size contains $n = 5$ training points and the mean values and error bars showing the 95% confidence levels of the points are evaluated over 100 trials. Shown is the behavior of (a) mean absolute error and (b) maximum absolute error.

directly on the prediction variability. Both EGO and Random predict a curve close to the true function in the second iteration, but then get worse as the iteration number increases.

By repeating over 100 trials, Figs. 3(a) and 3(b) show the average values and 95% confidence intervals for MAE and Max.AE for the different utilities. The results for MSD and Max.SD are shown in Fig. S4 [46]. Figure 3 essentially bears out our previous findings for the fatigue curve seen in Fig. 2, and the general features are very similar to those discussed previously for the fatigue curve. The uncertainty-based Max-v and B. EGO perform well and converge readily compared to Random and the trade-off methods, all of which do converge although require more iterations. Max-v relaxes more quickly than B.EGO if compared to fatigue, however, other than pure exploitation Min-u, all the exploratory utilities (including Random) converge in the 1D materials data sets.

B. Case study II: Higher-dimensional surfaces

1. (a) Hartmann 3 function

We utilize a well-known optimization test function, the Hartmann 3 function, to generate data for a three-dimensional
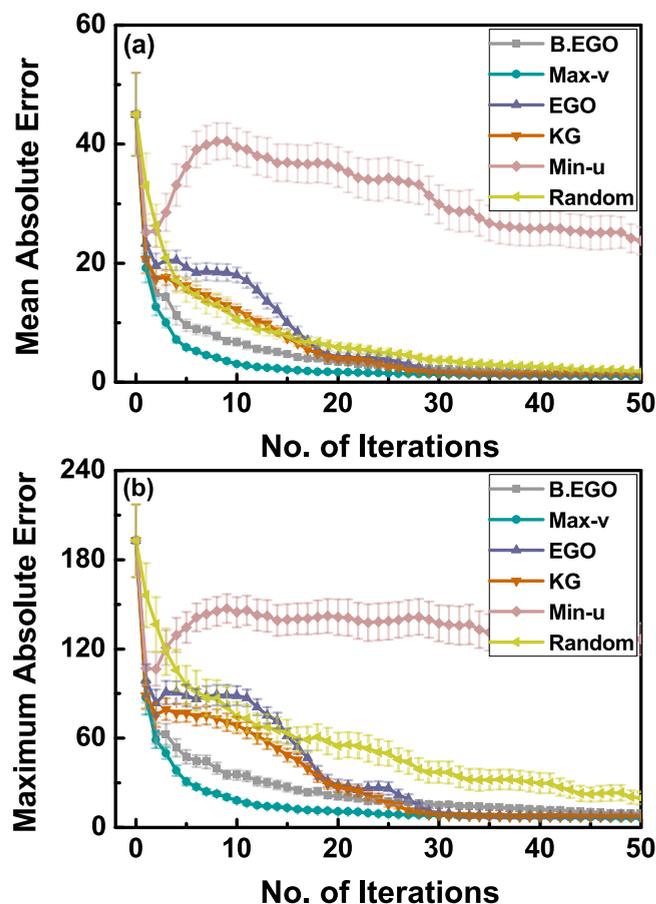
FIG. 3. Comparison of the performance of utility functions in optimizing the Liquidus line of the Fe-C phase diagram. The initial data size contains $n = 5$ training points and the mean values and error bars showing the 95% confidence intervals of the points are evaluated over 100 trials. Shown is the behavior of (a) mean absolute error and (b) maximum absolute error.

FIG. 4. Comparison of the performance of utility functions in optimizing the targeted Hartmann 3 function. The initial data size contains $n = 80$ training points and the mean values and error bars showing 95% confidence intervals of the points are evaluated over 100 trials. Shown is the behavior of (a) mean absolute error and (b) maximum absolute error.

mathematical case with multiple local minima and a global minimum. The function is defined by

$$y = -\sum_{n=1}^{4} \alpha_n \exp\left(-\sum_{m=1}^{3} A_{nm}(x_m - P_{nm})^2\right),$$

where $\alpha = (1.0, 1.2, 3.0, 3.2)^\top$,

$$A = \begin{pmatrix} 3.0 & 10 & 30 \\ 0.1 & 10 & 35 \\ 3.0 & 10 & 30 \\ 0.1 & 10 & 35 \end{pmatrix},$$

$$P = 10^{-4} \begin{pmatrix} 3689 & 1170 & 2673 \\ 4699 & 4387 & 7470 \\ 1091 & 8732 & 5547 \\ 381 & 5743 & 8828 \end{pmatrix}. \quad (15)$$

The whole space is discretized into 400 points (x_{mj}, $m = 1, 2, 3$; $j = 1, 2 \ldots 400$) using Latin hypercube sampling. Their corresponding y_j is the value evaluated by the function. We randomly select 80 data points (20% of the total search space) as the initial training data, with known x_{mj}^* and y_j^*. The surface as well as the optimization results after running 10

steps of different utility functions is shown in Fig. S5 [46]. From the distribution of the new 10 points, we can see that the points chosen by Max-v and B.EGO are widely distributed on the whole surface and initially points on the edge of the contours are even sampled. With EGO very few points are distributed away from the local or the global minimum.

To compare the efficiency of the utility function introduced above in the optimization process, we used an initial randomly training data set of $n = 80$ training points and tracked the values of MAE, Max.AE, MSD, and Max.SD as a function of number of new measurements for the different policies. Figures 4(a) and 4(b) show the average values and 95% confidence intervals for MAE and Max.AE for 100 trials. The results for MSD and Max.SD are shown in Fig. S6 [46]. Max-v and the tradeoff policies perform better than the rest, including B.EGO and Random. This example also suggests that the actual variance lends itself better to exploration of the space than the variability across bootstrap samples. The performance of EGO and KG is slightly better than Max-v for 30 or less iterations, which can be explained through the optimization sequence shown in Fig. S5 [46]. We observe that the training data are further away from the global minimum in Fig. S5 [46], suggesting that the points near the minimum
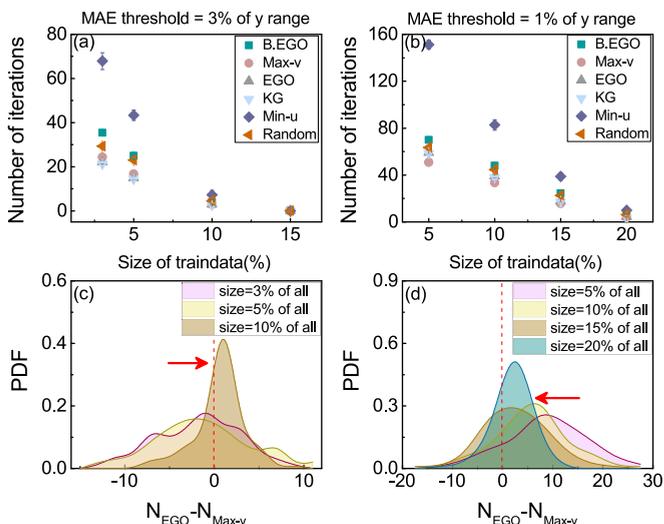
FIG. 5. The number of new measurements needed to achieve a given accuracy (3% and 1% of the y range) for the curve as a function of the training data size [(a) and (b)]. Each optimization process is repeated 100 times. The points with the error bar are the mean values associated with 95% confidence levels over 100 trials. The probability density of the difference in the number of iterations using EGO and Max-v. Plotted along the x axis is $N_{EGO} - N_{Max\text{-}v}$.

will not be well predicted. EGO samples some points near the global minimum in the first few iterations (shown by dark purple stars), whereas Max-v, B.EGO, and Random do not sample points in this area. Thus, it is not surprising that EGO does better initially.

To study the number of iterations required for different policies or utility functions to match a targeted objective function or curve, we set a threshold on MAE to stop our iteration loop. The threshold is set to (y_{max}-y_{min}) × 3% and 1%, respectively, to show how these utility functions perform to meet different demands of accuracy [shown in Figs. 5(a) and 5(c), 5(b) and 5(d)]. The initial training data with sizes from 3% to 15% times the number of total data for the first threshold, and 5% to 20% times the number of total data for the second threshold were selected randomly. A total of 200 iterations is set to stop the optimization loop. If after 200 iterations the loop does not reach the threshold, the number of new measurements needed is counted as 200. Figures 5(a) and 5(b) show the number of iterations required to meet the threshold as a function of initial training data size. Each point with the error bar represents the average value associated with 95% confidence level over 100 random trials. As the size of training data increases, all the utility functions perform better. Pure exploitation Min-u performs much worse than the others, almost two or three times slower, followed by B.EGO and Random. However, we notice the differences between the tradeoff methods and Max-v when the MAE threshold differs. If the MAE threshold equals 3% of y range, the tradeoff methods EGO and KG are a little better than Max-v. If the MAE threshold equals 1% of y range, a higher requirement on the accuracy, then Max-v is the best choice. Figures 5(c) and 5(d) show the probability density difference (δ) in iterations needed to reach the threshold using EGO and Max-v in 100 trials. The peak moves from negative to zero and becomes
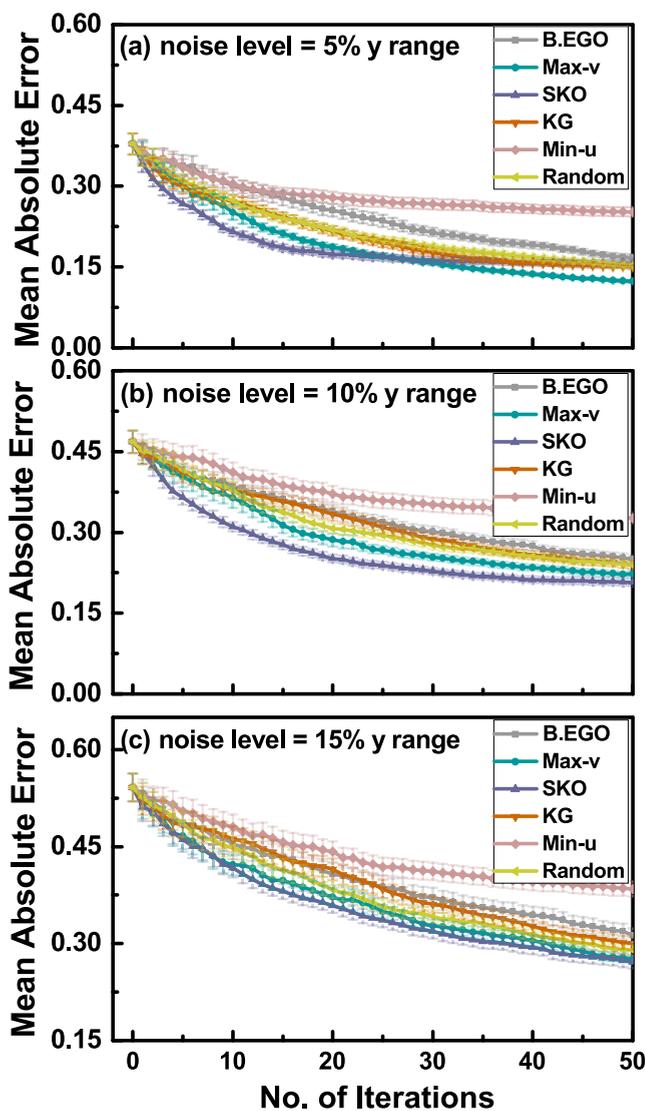
FIG. 6. Comparison of the performance of utility functions in optimizing the targeted Hartmann 3 function subject to noise levels of 5%, 10%, and 15% corresponding to (a)–(c). The initial data size contains $n = 80$ training points and the mean values and error bars showing the 95% confidence intervals of the points are evaluated over 100 trials. Shown is the behavior of mean absolute error (MAE).

narrower as the size of training data increases in Fig. 5(c), indicating that the advantages of EGO decrease with the increasing size until they finally perform similarly. For MAE threshold equal to 1% of y range, the opposite of Fig. 5(c) applies [shown in Fig. 5(d)].

Effects of noise. Using the function above, we introduce random errors to generate noisy data to simulate noisy measurements in experiments. We assume noise ε_j follows a normal distribution $\mathcal{N}(0, \tau^2)$, where τ is set to 5%, 10%, and 15% of y range, respectively. The observation values then can be calculated via $\tilde{y}_j = y(x_j) + \varepsilon_j$. A second measurement of the same candidate j, which has already been measured, is allowed. The results for the different utility functions after 100 trials are presented in Fig. 6. Each row corresponds to one level of noise, the noise level increases from top to

bottom. Compared to Fig. 4, if the number of iterations equals to zero, MAE and Max.AE (shown in Fig. S7 [46]) increase. The increase in these values in the beginning indicates that noise makes the prediction of the model deviate much more from the real curve. That is, the prediction suffers from both model uncertainties and measurement noise. SKO, the modified version of EGO with noise, performs very well, especially at noise levels of 5% and 10%. Max-v does surprisingly well and does converge to SKO with more iterations.

2. (b) Intermolecular potential energy surface for Ar-SH

The 3D intermolecular potential energy surface for Ar-SH has been determined by a combination of spectroscopic measurements and solutions to the Schrödinger equation [47]. The fitted surface and potential well reproduces all the known experimental data and we utilize this example to test the utility functions. The database includes in total 1050 points with the calculated potential energy in cm^{-1} and 3 variables, namely, the distance between Ar and the center of mass of SH in Å, the angle theta, and the SH bond length in Å. We randomly selected 52 data points (5% of the total search space) as the initial training data. The surface as well as the sequence of optimizations on the two of these variables planes are shown in Fig. S8 [46] for the 1050 points. The stars with the numbers refer to the sequence obtained. The distribution of the newly acquired 10 starred points by the utility functions is very similar to that for the Hartmann 3 function. That is, those chosen by Max-v and B.EGO are widely separated, whereas for EGO few points are dispersed away from the local or global maxima.

To compare the efficiency of the utility function introduced in the optimization process for the Ar-SH potential energy surface, we used an initial random training data set of $n = 52$ training points and monitored the values of MAE, Max.AE, MSD, and Max.SD as a function of number of new measurements for the different policies. Figures 7(a) and 7(b) show the average values and 95% confidence intervals for MAE and Max.AE for 100 trials. The results of MSD and Max.SD are shown in Fig. S9 [46]. The performance of EGO, KG, and Max-u is considerably better than Max-v in the first 50 iterations shown, and the optimization sequence in Fig. S8 [46] shows the evolution. As the training data are further away from the global maximum in Fig. S8 [46], we expect the predictions to have large uncertainties. EGO samples points near the global maximum in the first few iterations (shown by dark purple stars), whereas Max-v, B.EGO, and Random are sampling points further away. It is not surprising that the tradeoff methods, such as EGO, as well as the greedy Max-u perform well. Thus, as expected, the distribution of the data is a factor in the relaxation and performance.

C. Case study III: Multidimensional Curie transition temperature for BaTiO₃-based ceramics

Ferroelectric ceramics such as $BaTiO_3$ undergo a transition from a nonpolar paraelectric state to a ferroelectric polar state at a critical temperature known as the Curie temperature. This critical temperature is the upper limit of the temperature window for piezoelectric applications including sensors, transducers, etc. [32], and therefore it is of impor-
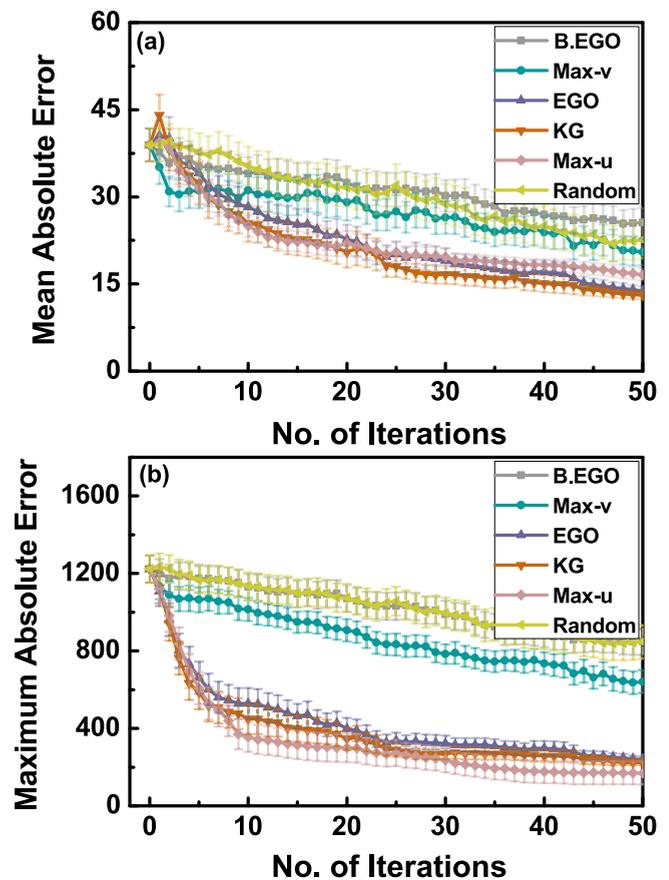
FIG. 7. Comparison of the performance of utility functions in optimizing the Ar-SH potential energy surface. Plotted is the deviation of the model prediction from that obtained using measurements fitted to the Schrödinger equation. The initial data size contains $n = 52$ training points and the mean values and error bars showing 95% confidence intervals of the points are evaluated over 100 trials. Shown is the behavior of (a) mean absolute error and (b) maximum absolute error.

tance to construct the Curie transition temperature surface in as few measurements as possible.

The Curie transition temperature in $BaTiO_3$-based ceramics is affected by some features or descriptors, which can be calculated for each composition by a weighted fraction of elemental properties including radius, electronegativity, valence electron numbers of the perovskite structure containing A and B site cations [48]. It has been found that four descriptors capture the salient physics of the system [49] and here we test how the Curie transition temperature behaves in terms of these descriptors in a discrete 4D space without a knowledge functional form for the surface.

Our data set consists of 182 samples of data obtained in our laboratory [46] and previously published [19,50]. This data set consists of the property, the Curie temperature, and the four descriptors or variables including density of B-site element (BD), the ratio of electronegativity of the A site and B site (EN), the ratio of the ionic radii of A site and B site (TA.B), and the product of atomic volume of A site and B site (av). The descriptors were obtained and studied previously from a large superset of descriptors using methods such as gradient
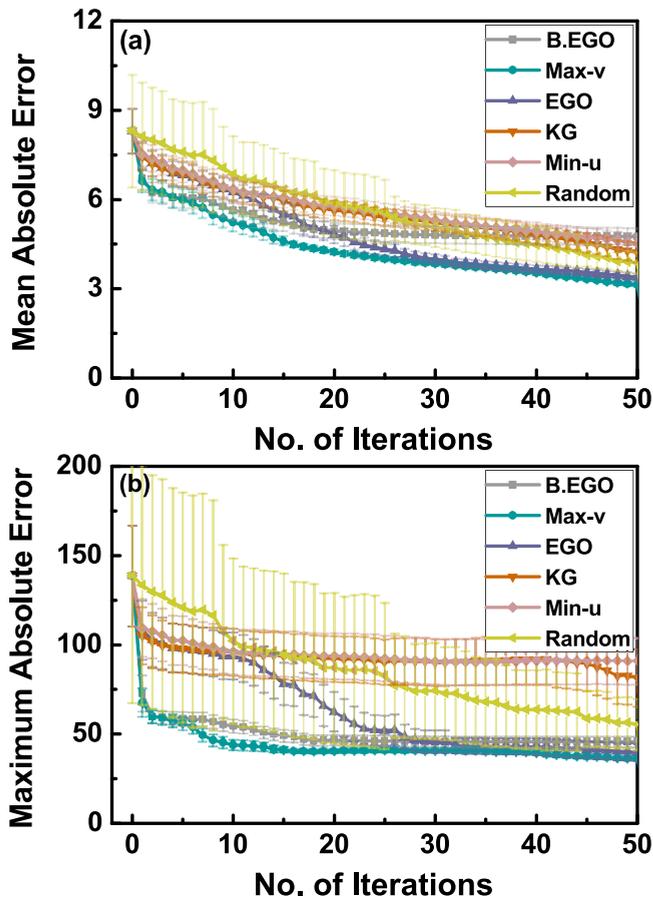
FIG. 8. Comparison of the performance of utility functions in optimizing the targeted Curie temperature for $BaTiO_3$-based ceramics. The initial data size contains $n = 55$ training points and the mean values and error bars showing the 95% confidence intervals of the points are evaluated over 100 trials. Shown is the behavior of (a) mean absolute error and (b) maximum absolute error.

boosting and 10-fold cross validation with a Kriging-based model [49]. Our target here is to minimize the difference between the predicted values from our Kriging model and the measurement values. For the initial model training, we randomly selected 30% of the data.

Figure 8 shows the results from 100 trials. There is a significant drop initially in the maximum error for Max-v and B. EGO, suggesting that the bulk of the uncertainty is reduced within a few iterations. Again, Max-v relaxes the most but the others are not far from converging to it. Unlike the other cases, Max.AE does not relax to zero and is indicative of the complexity of the problem. We have essentially used only four features to model these data from a sampling of 182 measurement points. There are uncertainties in the model itself, hence, the Kriging needs to be accurate, and we are assuming that the sampling of points is representative of the data in the whole space.

IV. DISCUSSION

Our objective has been to compare the influence of utility functions for curves, such as phase boundaries, fatigue lines,

and other multidimensional cases important in materials science. This is essential as in the absence of analytical results, it is difficult to predict *a priori* which utilities will be superior in reducing the costs of acquiring new information when learning from data.

Except for the random case, all the utilities we compare are based on directed exploration, which can incorporate different degrees of exploitation. Our key finding is that maximum variance (Max-v) performs well across a range of data sets with varying complexities, including the addition of experimental noise. The function B.EGO, which tracks the variability over bootstrap samples, and uses EGO to minimize the variability across the whole function, also shows relatively good performance, although it is not as robust as Max-v. Moreover, we also find that for some type of data set, there exists a utility which performs as well, if not better than, or at least competes with, Max-v. The distributions of the property values in the data set *y* can also influence the behavior. If the distributions depart from the uniform distribution, and typically there are relatively few training data points located near global minima and maxima, then the minima and maxima can be associated with large deviations from the true result. We find this to be the case for the Hartmann 3 function and intermolecular potential data sets, where the tradeoff methods EGO and KG perform as well, if not better than Max-v. Also, for a given problem, several utilities can converge but at varying iteration numbers. For the intermolecular potential, the convergence is far superior to Max-v even after 50 iterations. In cases where Random selection does converge, it requires a lot more iterations as the exploration is unguided. In the presence of experimental noise, SKO, which is essentially EGO with noise incorporated, is the superior performer at noise levels of 5% and 10%, although Max-v also does well, converging with more iterations.

Our results emphasize the importance of making the appropriate choice in ranking and selecting the next candidate for measurements or calculations.

To gain an understanding of the behavior of these functions, we plot the probability density functions (PDF) for the uncertainties from the Kriging model estimates and for the deviation of the estimate from the true curve, as a function of iterations. Figures 9 and 10 show the results for the Liquidus line of Fe-C phase diagram and the fatigue life curve of 304L steel, respectively. The Fe-C curve is more complex and its Kriging estimate would not be expected to be as good. Thus, for all the four utilities being studied, we see wide distributions in the uncertainty profile for the estimates and the deviation from true curve (panel No. = 0) of Fig. 9. With successive iterations as the next point is added, the distributions of the uncertainties begin to narrow and the mean value tends towards zero. All strategies are efficient in this sense with Max-v leading to a desired narrower sharply peaked distribution (No. = 5) of Fig. 9. For Random selection and EGO, points with very large uncertainties always occur because of the long tails in the distribution. The other utilities target such points with large uncertainties to reduce the tail in the uncertainty distribution.

The Kriging estimate for the fatigue line behavior shown in Fig. 10 is better than for Fe-C as the function by comparison is quite monotonic and so the initial distributions of "deviation
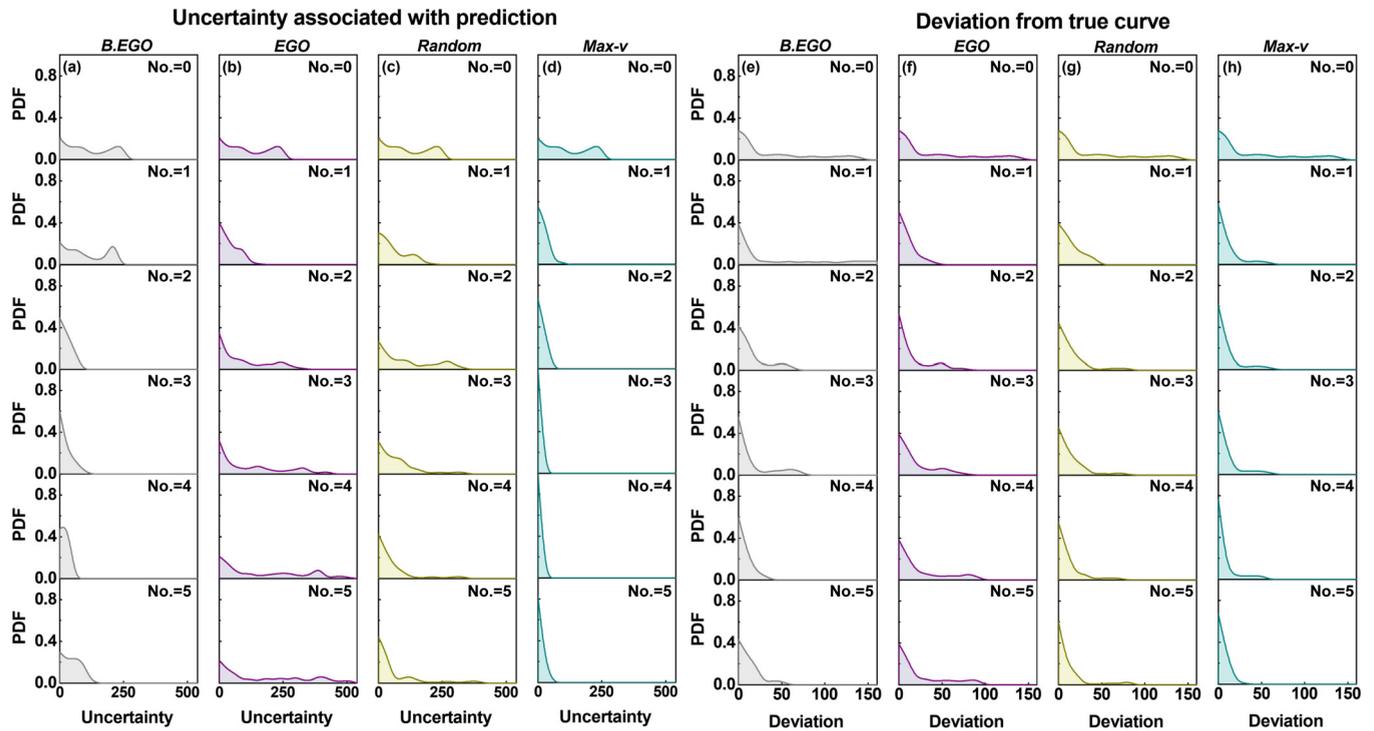
FIG. 9. For the Liquidus line of the Fe-C phase diagram, the probability density functions (PDF) of the uncertainties and deviation from true result associated with the Kriging model prediction for Random, Max-v, B.EGO, and EGO are shown with successive iterations.

from true curve" shown in panels (No. = 0) of Fig. 10 are narrower. The evolution of the distributions for utility functions (except for Random) is similar to that in Fig. 9. Because the objective function is simpler, all utility functions show very similar performance in the first two iterations. Thereafter, the PDF of "deviation from true curve" employing Max-v quickly converges to zero compared to B.EGO and EGO.
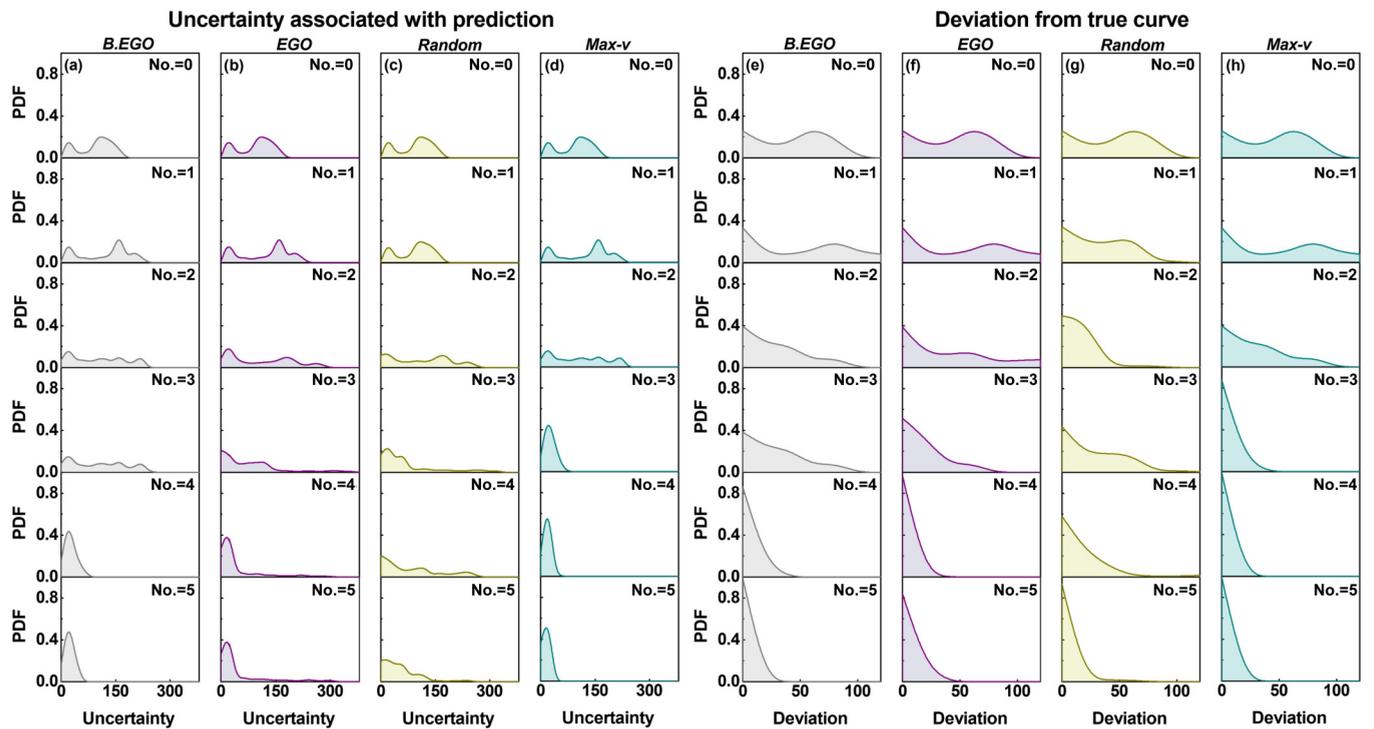
FIG. 10. For the fatigue life curve for SS304L steel, the probability density functions (PDF) of the uncertainties and deviation from true result associated with the Kriging model prediction for Random, Max-v, B.EGO, and EGO are shown with successive iterations.
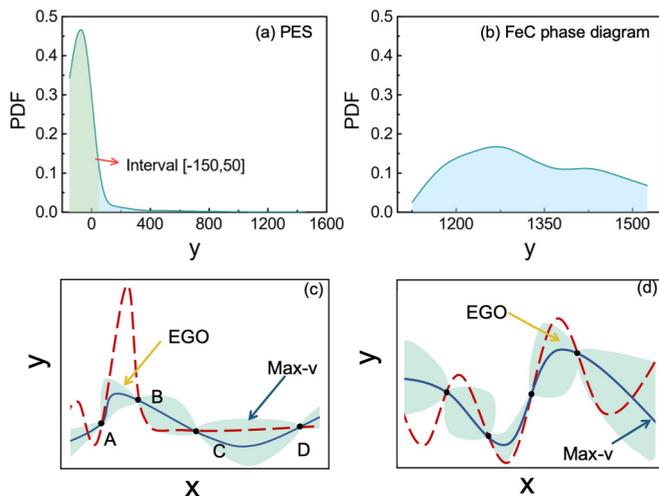
FIG. 11. Typical solutions and the probability density functions (PDF) of the property values y for the intermolecular potential energy surface (PES) and the FeC phase boundary data sets. The dark blue line represents the model prediction and the red line the true solution.

Thus, irrespective of how good initially the predictive model is, Max-v shows superior performance in these two 1D cases.

We conclude with some general remarks on circumstances that favor Max-v and tradeoff methods such as EGO. If the variance is large and the deviation from the true result also large, then selection by Max-v will have a significant effect in decreasing the deviation further. However, if the deviation is small (i.e., the model is good), then the reduction in deviation will not be significant. This is likely the case for both of the 1D curve examples in case 1 where Max-v is especially good. Similarly, in situations where the uncertainty is not too large, and the deviation from the true result is significant, then tradeoff methods such as EGO will have a substantial effect in locating the max and min of the curve, that is, decrease the deviation further. We suggest this is the case for the Hartmann

3 function and PES intermolecular potential. Tradeoff methods depend on balancing the uncertainty (exploration) with exploitation (model performance), and in the limits where the uncertainties are either very large or very small, for a given deviation, then from expression (4), EGO behaves either as Max-v or chooses the model prediction, respectively.

To illustrate graphically, we have plotted in Figs. 11(a) and 11(b) the distribution of the y-data values for the intermolecular potential energy surface (PES) and FeC phase boundary examples. We note that the former deviates strongly from a uniform distribution of values, whereas the latter is closer to uniform. Schematics of the solutions corresponding to these two cases are shown in Figs. 11(c) and 11(d), where the red line is the actual solution and the dark blue line is the model prediction. The distribution of y values in Fig. 11(a) typically gives rise to the curves of Fig. 11(c) with a maximum and there are relatively few training data points close to the maximum in the distribution. As expected, EGO works well near the maximum, whereas Max-v is better where the uncertainties are large. For a more uniform distribution of y-data values as in Fig. 11(b) for the 1D FeC example, the solution profile is more complex as in Fig. 11(d), with Max-v having a greater impact on driving the optimization towards the solution. We hope our work will motivate further studies on a variety of materials data to confirm our findings, as well as provide a deeper understanding of why Max-v works so well across data sets with varying complexities. All the data used to perform this work are given in [46].

[1] D. Askeland and W. Wright, *Essentials of Materials Science and Engineering* (Cengage Learning, Stamford, CT, 2013).

[2] R. Newnham, *Properties of Materials: Anisotropy, Symmetry, Structure, Properties of Materials: Anisotropy, Symmetry, Structure* (Oxford University Press, Oxford, 2005).

[3] C. Shen, C. Wang, X. Wei, Y. Li, S. Zwaag, and W. Xu, Physical metallurgy-guided machine learning and artificial intelligent design of ultrahigh-strength stainless steel, Acta Mater. **179**, 201 (2019).

[4] J. Im, S. Lee, T. W. Ko, H. W. Kim, Y. K. Hyon, and H. Chang, Identifying Pb-free perovskites for solar cells by machine learning, npj Comput. Mater. **5**, 37 (2019).

[5] V. Stanev, C. Oses, A. G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, and I. Takeuchi, Machine learning modeling of superconducting critical temperature, npj Comput. Mater. **4**, 29 (2018).

[6] D. Shin, Y. Yamamoto, M. P. Brady, S. Lee, and J. A. Haynes, Modern data analytics approach to predict creep of high-temperature alloys, Acta Mater. **168**, 321 (2019).

[7] T. Lookman, P. V. Balachandran, D. Xue, and R. Yuan, Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design, npj Comput. Mater. **5**, 21 (2019).

[8] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and Weinan E, Active learning of uniformly accurate interatomic potentials for materials simulation, Phys. Rev. Mater. **3**, 023804 (2019).

[9] D. Fan, G. Jodin, T. R. Consi, L. Bonfiglio, Y. Ma, L. R. Keyes, G. E. Karniadakis, and M. S. Triantafyllou, A robotic intelligent towing tank for learning complex fluid-structure dynamics, Sci. Robotics **4**, eaay5063 (2019).

[10] K. A. Brown, S. Brittman, N. Maccaferri, D. Jariwala, and U. Celano, Machine learning in nanoscience: Big data at small scales, Nano Lett. **20**, 2 (2020).

[11] K. Tran and Z. W. Ulissi, Active learning across intermetallics to guide discovery of electrocatalysts for CO2 reduction and H-2 evolution, Nat. Catal. **1**, 696 (2018).

[12] J. E. Gubernatis and T. Lookman, Machine learning in materials design and discovery: Examples from the present and

suggestions for the future, Phys. Rev. Mater. **2**, 120301 (2018).

[13] M. J. Casciato, S. Kim, J. C. Lu, D. W. Hess, and M. A. Grover, Optimization of a carbon dioxide-assisted nanoparticle deposition process using sequential experimental design with adaptive design space, Ind. Eng. Chem. Res. **51**, 4363 (2012).

[14] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, Less is more: Sampling chemical space with active learning, J. Chem. Phys. **148**, 241733 (2018).

[15] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, Machine learning in materials informatics: Recent applications and prospects, npj Comput. Mater. **3**, 54 (2017).

[16] J. Ling, M. Hutchinson, E. Antono, S. Paradiso, and B. Meredig, High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates, Integrat. Mater. Manufact. Innovat. **6**, 207 (2017).

[17] L. Bassman, P. Rajak, R. K. Kalia, A. Nakano, F. Sha, J. Sun, D. J. Singh, M. Aykol, P. Huck, K. Persson, and P. Vashishta, Active learning for accelerated design of layered materials, npj Comput. Mater. **4**, 74 (2018).

[18] C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman, and Y. Su, Machine learning assisted design of high entropy alloys with desired property, Acta Mater. **170**, 109 (2019).

[19] R. Yuan, Y. Tian, D. Xue, D. Xue, Y. Zhou, X. Ding, J. Sun, and T. Lookman, Accelerated search for $BaTiO_3$ based ceramics with large energy storage at low fields using machine learning and experimental design, Adv. Sci. **6**, 1901395 (2019).

[20] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, and T. Lookman, Accelerated search for materials with targeted properties by adaptive design, Nat. Commun. **7**, 11241 (2016).

[21] D. Xue, P. V. Balachandran, R. Yuan, T. Hu, X. Qian, E. R. Dougherty, and T. Lookman, Accelerated search for $BaTiO_3$-based piezoelectrics with vertical morphotropic phase boundary using bayesian learning, Proc. Natl. Acad. Sci. USA **113**, 13301 (2016).

[22] K. Terayama, R. Tamura, Y. Nose, H. Hiramatsu, H. Hosono, Y. Okuno, and K. Tsuda, Efficient construction method for phase diagrams using uncertainty sampling, Phys. Rev. Mater. **3**, 033802 (2019).

[23] T. Mazaheri, B. Sun, J. Scher-Zagier, A. S. Thind, D. Magee, P. Ronhovde, T. Lookman, R. Mishra, and Z. Nussinov, Stochastic replica voting machine prediction of stable cubic and double perovskite materials and binary alloys, Phys. Rev. Mater. **3**, 063802 (2019).

[24] T. Fukazawa, Y. Harashima, Z. Hou, and T. Miyake, Bayesian optimization of chemical composition: A comprehensive framework and its application to RFe12-type magnet compounds, Phys. Rev. Mater. **3**, 053807 (2019).

[25] A. Talapatra, S. Boluki, P. Honarmandi, A. Solomou, G. Zhao, S. F. Ghoreishi, A. Molkeri, D. Allaire, A. Srivastava, X. Qian, E. R. Dougherty, D. C. Lagoudas, and R. Arróyave, Experiment design frameworks for accelerated discovery of targeted materials across scales, Front. Mater. **6**, 82 (2019).

[26] R. K. Vasudevan, K. Choudhary, A. Mehta, R. Smith, G. Kusne, F. Tavazza, L. Vlcek, M. Ziatdinov, S. V. Kalinin, J. H. Simpers

et al., Materials science in the artificial intelligence age: high-throughput library generation, machine learning, and a pathway from correlations to the underpinning physics, MRS Commun. **9**, 821 (2019).

[27] S. B. Thrun, Efficient exploration in reinforcement learning, Technical Report No. CMU-CS-92-102, Carnegie Mellon University, USA, 1992.

[28] S. Whitehead, A study of cooperative mechanisms for faster reinforcement learning, Technical Report No. TR365, University of Rochester, 1991.

[29] M. K. Bisbo and B. Hammer, Efficient Global Structure Optimization with a Machine-Learned Surrogate Model, Phys. Rev. Lett. **124**, 086102 (2020).

[30] J. Theiler and B. G. Zimmer, Selecting the selector: Comparison of update rules for discrete global optimization, Stat. Anal. Data Mining **10**, 211 (2017), Conference on Data Analysis (CoDA), Santa Fe, NM, 2016.

[31] Y. Wang, K. G. Reyes, K. A. Brown, C. A. Mirkin, and W. B. Powell, Nested-batch-mode learning and stochastic optimization with an application to sequential multistage testing in materials science, SIAM J. Sci. Comput. **37**, B361 (2015).

[32] R. Yuan, Z. Liu, P. V. Balachandran, D. Xue, Y. Zhou, X. Ding, J. Sun, D. Xue, and T. Lookman, Accelerated discovery of large electrostrains in $BaTiO_3$-based piezoelectrics using active learning, Adv. Mater. **30**, 1702884 (2018).

[33] P. V. Balachandran, D. Xue, J. Theiler, J. Hogden, and T. Lookman, Adaptive Strategies for Materials Design using Uncertainties, Sci. Rep. **6**, 19660 (2016).

[34] D. Xue, D. Xue, R. Yuan, Y. Zhou, P. V. Balachandran, X. Ding, J. Sun, and T. Lookman, An informatics approach to transformation temperatures of NiTi-based shape memory alloys, Acta Mater. **125**, 532 (2017).

[35] N.Cressie, The origins of kriging, Math. Geol. **22**, 239 (1990).

[36] A. G. Journel, Kriging in terms of projections, J. Int. Assoc. Math. Geol. **9**, 563 (1977).

[37] D. Ginsbourger, R. Le Riche, and L. Carraro, *Kriging Is Well-Suited to Parallelize Optimization* (Springer, Berlin, 2010), pp. 131–162.

[38] C. E. Rasmussen, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).

[39] O. Roustant, D. Ginsbourger, and Y. Deville, DiceKriging, Diceoptim: Two R packages for the analysis of computer experiments by Kriging-based metamodeling and optimization, J. Stat. Software **51**, 1 (2012).

[40] W. Powell, *The Knowledge Gradient for Optimal Learning* (Princeton University Press, Princeton, NJ, 2011).

[41] D. Jones, M. Schonlau, and W. Welch, Efficient global optimization of expensive black-box functions, J. Global Optim. **13**, 455 (1998).

[42] D. Huang, T. Allen, W. Notz, and N. Zeng, Global optimization of stochastic black-box systems via sequential kriging metamodels, J. Global Optim. **34**, 441 (2006).

[43] P. I. Frazier, W. B. Powell, and S. Dayanik, A knowledge-gradient policy for sequential information collection, SIAM J. Control Optim. **47**, 2410 (2008).

[44] P. Frazier, W. Powell, and S. Dayanik, The Knowledge-Gradient Policy for Correlated Normal Beliefs, Informs J. Comput. **21**, 599 (2009).

[45] F. Mozafari, P. Thamburaja, A. R. Srinivasa, and N. Moslemi, A rate independent inelasticity model with smooth transition for

unifying low-cycle to high-cycle fatigue life prediction, Int. J. Mech. Sci. **159**, 325 (2019).

[46] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.5.013802 for the optimization process and MSD/Max.SD results of each case.

[47] Y. Sumiyoshi and Y. Endo, Spectroscopy of Ar-SH and Ar-SD. II. determination of the three-dimensional intermolecular potential-energy surface, J. Chem. Phys. **123**, 054325 (2005).

[48] A. Seko, H. Hayashi, K. Nakayama, A. Takahashi, and I. Tanaka, Representation of compounds for machine-learning prediction of physical properties, Phys. Rev. B **95**, 144110 (2017).

[49] Y. Tian, R. Yuan, D. Xue, Y. Zhou, Y. Wang, X. Ding, J. Sun, and T. Lookman, Determining multi-component phase diagrams with desired characteristics using active learning, Adv. Sci. 2003165 (2020), doi: 10.1002/advs.202003165.

[50] Y. Tian, R. Yuan, D. Xue, Y. Zhou, X. Ding, J. Sun, and T. Lookmanw, Role of uncertainty estimation in accelerating materials development via active learning, J. Appl. Phys. **128**, 014103 (2020).