

Adaptive active subspace-based efficient multifidelity materials design

Danial Khatamsaz^a, Abhilash Molkeri^b, Richard Couperthwaite^b, Jaylen James^b, Raymundo Arróyave^{a,b,c}, Ankit Srivastava^b, Douglas Allaire^{a,*}

^aJ. Mike Walker '66 Department of Mechanical Engineering, Texas A&M University, College Station, TX, USA

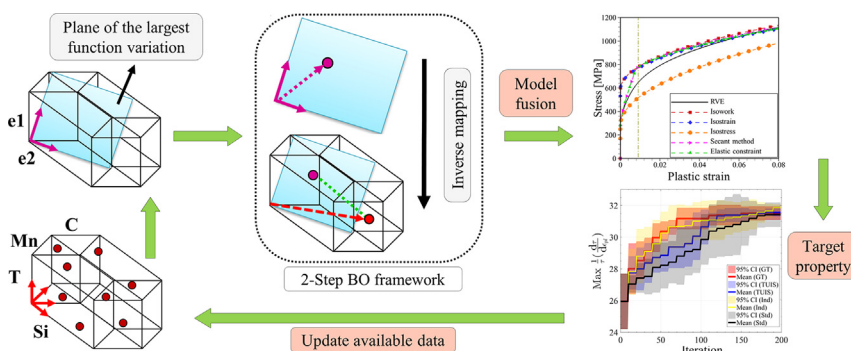
^bDepartment of Materials Science and Engineering, Texas A&M University, College Station, TX, USA

^cDepartment of Industrial and Systems Engineering, Texas A&M University, College Station, TX, USA

HIGHLIGHTS

- A technique using input variable's correlation to represent data in lower dimensions.
- The method uses the objective value to reduce the dimensionality of the design space.
- The active subspace is adaptively updated as new data becomes available.
- Optimizing over the informative design subspaces, increases the design efficiency.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 8 April 2021

Revised 5 July 2021

Accepted 21 July 2021

Available online 24 July 2021

Keywords:

Adaptive dimensionality reduction

Active subspace

Multifidelity design

Bayesian optimization

PSP relationships

Dual-phase materials

ABSTRACT

Materials design calls for an optimal exploration and exploitation of the process-structure-property (PSP) relationships to produce materials with targeted properties. Recently, we developed and deployed a closed-loop multi-information source fusion (multi-fidelity) Bayesian Optimization (BO) framework to optimize the mechanical performance of a dual-phase material by adjusting the material composition and processing parameters. While promising, BO frameworks tend to underperform as the dimensionality of the problem increases. Herein, we employ an adaptive active subspace method to efficiently handle the large dimensionality of the design space of a typical PSP-based material design problem within our multi-fidelity BO framework. Our adaptive active subspace method significantly accelerates the design process by prioritizing searches in the important regions of the high-dimensional design space. A detailed discussion of the various components and demonstration of three approaches to implementing the adaptive active subspace method within the multi-fidelity BO framework is presented.

© 2021 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Integrated Computational Materials Engineering (ICME) – based material design [1] relies on solving the inverse problem connecting target properties/performance metrics to material chemistry and processing. This connection is established through (forward)

process-structure-property (PSP) relationships [2,3], which are in turn established through physics- or machine learning-based models [4–6] and/or experimental data. The solution to this inverse problem entails the exploration and exploitation of PSP relationships to identify the required chemistry-processing combinations that yield desired properties [7]. Practical implementation of ICME frameworks requires addressing *three major challenges*: the need to explicitly connect the different models along the PSP chain; the considerable cost associated with the evaluation of each of the

* Corresponding author.

E-mail address: dallaire@tamu.edu (D. Allaire).

models/linkages; and the potentially large dimensionality of the design space.

A significant amount of work has been done to address the *first challenge*, at least in the context of microstructure sensitive materials design, which aims to uncover optimal microstructures that meet specific performance objectives by focusing exclusively on the microstructure-property space [8–13]. While promising, this approach assumes that the design space consists of a universe of microstructures that are all feasible, potentially, through suitable chemistry-processing combinations. This is an unwarranted assumption as there is no guarantee that an optimal microstructure is feasible, in the sense that it can be attained through an adequate processing protocol. To date, there has been some measure of success in the deployment of fully integrated PSP model chains for materials design [14,15]. However, this is not a trivial task largely due to the complex, highly coupled, multi-scale nature of the linkages along the PSP chain [16].

To address the *second challenge* associated with the considerable cost of querying the PSP relationships, the materials design community has focused on the development and deployment of closed-loop Bayesian Optimization (BO) frameworks to efficiently explore and exploit the material design space [15,17–21]. These frameworks seek a balance between exploration and exploitation in order to efficiently arrive at optimal materials solutions. These approaches are suitable and have been used successfully in both simulation-driven and experiment-centered materials design problems.

The *third challenge*, however, has largely remained unaddressed. This is despite the fact that, more often than not, the design space is large, [12,15] and BO frameworks tend to underperform as the dimensionality of the problem increases [22]. There are some works done toward scaling BO frameworks to higher dimensions to address the issue of employing a BO approach in high-dimensional design spaces. For example, a technique is to consider additive models [23,24] and treating a function as an additive function of mutually exclusive lower dimensional subspaces [25]. Also, in Ref. [26], a random embedding idea is used to lower the dimensionality of the design space. By embedding different dimensions and optimizing an acquisition function, they search for dimensions in the design space that are more important in optimizing an objective function. However, the results of an study comparing similar approaches in Ref. [27] have shown that none of these methods can outperform the other ones in all situations and there are trade-offs in employing each technique compromising the obtained results. It is certainly possible to carry out statistical tests to determine the most influential design variables in any optimization task and then to focus exclusively on those degrees of freedom during the design process. However, this requires sufficient data connecting design inputs to design outputs. Thus, there is a need for techniques that can effectively locate the most important (and/or informative) design regions to increase the efficiency of the materials design process. These efficiency gains can be more pronounced if these potentially productive design regions can be identified with limited data in an adaptive manner, while more information about the design space is gained in the course of exploring/exploiting it [28,29].

There are techniques to deal with the curse of dimensionality that stems from large design spaces, usually by defining a representative response surface in a lower-dimensional space while maintaining the relationship between the design variables as much as possible. For instance, global sensitivity analysis is used to measure the importance of different design variables in the variation of a quantity of interest [30–34]. This approach assigns a nominal value to the design variables that have little effect on the objective and perform the optimization over the remaining design variables. Approximating a subspace of the original large design space is also

among the most common approaches to dimensionality reduction and can be used to represent data in a lower-dimensional space to ease machine learning objectives, increase the efficiency of optimization tasks [35,36], aid in model reduction [37], or facilitate optimal control of dynamic systems [38]. Another technique in dimensionality reduction is the Principal Component Analysis (PCA) [39], which linearly projects a large dimensional dataset onto a lower-dimensional space [40–42] while trying to keep as much information as possible by determining the principal components that capture a majority of the variance in the data. PCA has been employed for microstructure sensitive design to build models to predict (mechanical) properties using a lower-dimensional representation of the complex material microstructure [43–50]. Note that PCA simply decreases the dimensionality of the design space by considering the correlations among design variables, without accounting for the connection between the design variables and the quantities of interest amenable to optimization. In a PSP-based materials design problem, where the focus is on optimizing the performance metrics by exploring material chemistry and processing options, directly decreasing the dimensionality of the design space without accounting for the design objectives may not be feasible.

Herein, we employ an adaptive active subspace method [51–53] to efficiently handle the large dimensionality of the design space of a typical PSP-based materials design problem within our recently developed closed-loop multi-information source fusion (multi-fidelity) BO framework [15]. Specifically, we demonstrate the efficacy of this framework by optimizing the stress, τ , normalized strain hardening rate, $d\tau/d\epsilon_{pl}$, at an arbitrary equivalent plastic strain, $\epsilon_{pl} = 0.9\%$, of a dual-phase material (ferrite-martensite steel) by adjusting the content of the alloying elements C, Mn and Si in the Fe-based alloy, and the processing condition, i.e., the intercritical annealing temperature, T . The normalized strain hardening rate, $(1/\tau)(d\tau/d\epsilon_{pl})$, is a useful mechanical performance metric, and a higher value of this parameter indicates better ductility and formability of the material. In the design framework, we utilize the thermodynamic results to predict the chemistry and composition of the constituent phases after the single-stage heat-treatment (intercritical annealing followed by quenching) [15,21]. This information is then used to predict the mechanical performance of the dual-phase material using a variety of (reduced-order) micromechanical models referred to as Isotress, Isostrain, Isowork, Secant method and Elastic constraint, and a high-throughput microstructure-based finite element model that utilizes a three-dimensional representative volume element (RVE) of the material microstructure [15,20]. All these models, low fidelity micromechanical models, as well as high fidelity microstructure-based finite element models (referred to as RVE and assumed to be the ‘ground truth’) are treated as information sources. We represent the response of each information source as Gaussian process surrogates and fuse them using standard approaches for the fusion of normally distributed data.

Our approach to implementing the adaptive active subspace method within the multi-fidelity BO framework is schematically shown in Fig. 1. The active subspace method is a technique to look for the directions in the design space for which a function has the largest variability. By forming a subspace using those directions, an approximation of the function is obtained on a lower-dimensional space referred to as the *active subspace* [52]. Thus, increasing the efficiency of the design process by more effectively searching for the optimal solution within the high-dimensional design space [51,53]. In the context of materials design, the function is the PSP relationship that is being evaluated, and the basic idea is to find the directions in the design space (T , C , Mn , and Si) that give the largest variation in the objective value (normalized strain harden-

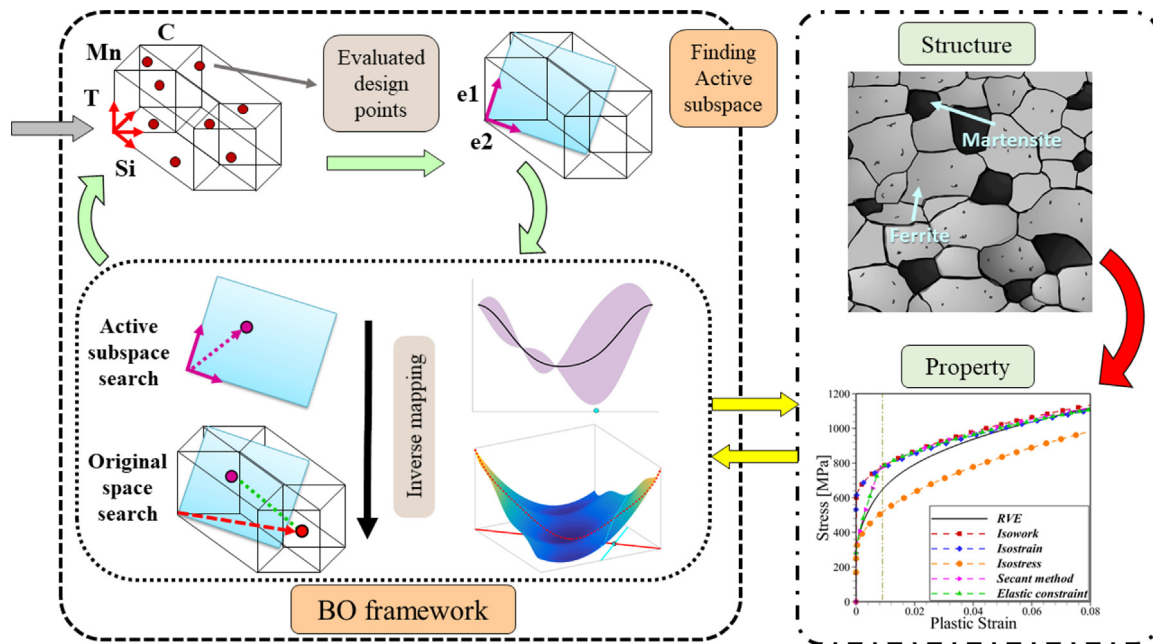


Fig. 1. Implementation of adaptive active subspace method within a multifidelity Bayesian Optimization (BO) framework. The basic idea is to find the active subspace, i.e., the directions in the material design space – intercritical annealing temperature (T), and alloying elements C , Mn , and Si – that give the largest variation in the mechanical property (normalized strain hardening rate) by using the available data at every stage of the optimization task. Next, the process – structure-property (PSP) relationship is mapped to the active subspace, and the first step of the BO framework is applied to find the ‘next best point’ to evaluate within the active subspace. The ‘next best point’ is then mapped back to the original design space by implementing a second BO step. Finally, the PSP relationship is evaluated at this best point using the thermodynamic-based model and the selected micromechanical model to estimate the objective value. This new data is added to the framework for the next iteration.

ing rate) by using the available data at every stage of the optimization task. The directions suggesting larger variation than a user-specified value then form the active subspace. Following this, we employ a knowledge gradient acquisition function to determine the ‘next best point’ to evaluate within the active subspace. In order to obtain the true input values for the ‘next best point,’ the chosen point in the active subspace must be inversely mapped to the true design space. Since there is no unique solution for this inverse mapping problem, a second BO step is performed to determine the ‘next best point’ in the true design space. At this stage, a decision about which information source (low fidelity micromechanical model) to query is also made by temporarily updating each information source and comparing their results. Finally, the PSP relationship is evaluated at this ‘best point’ using the thermodynamic-based model and the selected micromechanical model to estimate the objective value.

2. Methods

In this work, our objective is to maximize the stress, τ , normalized strain hardening rate, $d\tau/d\epsilon_{pl}$, i.e. $(1/\tau)(d\tau/d\epsilon_{pl})$ at an arbitrary equivalent plastic strain, $\epsilon_{pl} = 0.9\%$, of a dual-phase material. The dual-phase material system considered is a ferritic-martensitic steel which is produced by subjecting the material system composed of Fe , C , Mn and Si to a single-stage intercritical annealing heat treatment followed by rapid quenching. Therefore, our optimization problem aims to find the values of the intercritical annealing temperature, T , and C , Mn and Si content of the Fe -based alloy that correspond to the maximum value of the $(1/\tau)(d\tau/d\epsilon_{pl})$.

In previous work, we addressed the optimization problem discussed above with a multifidelity BO framework [15] to incorporate the response of different mechanical models, which enabled the collection of information about the optimum design in a less

costly manner in comparison to employing a finite element model alone. The multifidelity aspect of the approach was used to exploit the fact that in most materials design problems we have available several different models that can potentially be used to estimate a quantity of interest. These models are usually based on different physics-based and numerical assumptions, which leads to models with varying expense in terms of computational resources required for a query to the model and varying fidelity. The exploitation of each possible model, or information source, was achieved via an information fusion process described in Refs. [54,55]. While the overall multifidelity BO approach of our prior work was shown to be more efficient than traditional BO approaches, the process can still be computationally impractical when applied over large design spaces. Here, to address this challenge, we consider the application of adaptive dimensionality reduction in the context of our multifidelity BO framework using the active subspace method, which is described in detail below in Section 2.4.

Fig. 2 illustrates the steps in our proposed adaptive active subspace-based multifidelity BO framework. The framework starts by reducing the dimensionality of the design space and projecting all evaluated designs to a lower dimensional design space (the current active subspace), and then two steps of Bayesian optimization are executed. The first step of BO is applied over the active subspace to find a best design candidate to query in the lower dimensional space. Next, the obtained design is inverse mapped to the full dimensional design space, which results in a subspace of potential solutions of the inverse problem. The second step of BO is applied over this subspace to obtain the next design and information source to query.

In every step of the Bayesian optimization process, we use surrogate models to estimate the expected objective values of design points that have yet to be evaluated. In a multifidelity setting, we therefore have multiple surrogate models to construct (one for each information source) and for use in predicting design points not yet queried. Since every information source contains useful

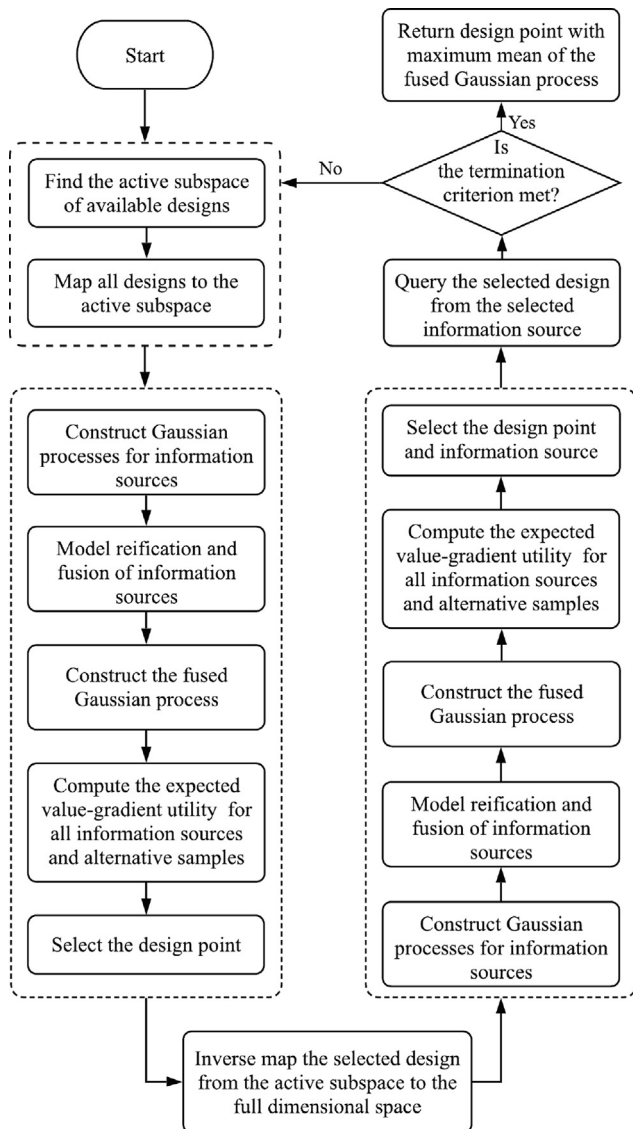


Fig. 2. Flowchart of the proposed approach. At the start of every iteration, the active subspace is found and all data are projected onto it. Then, the first step of BO is applied over this active subspace. The best design candidate is mapped back to the full dimensional design space, resulting in a solution subspace. The second step of BO then is applied over this subspace to select the best design and information source to query.

data regarding the expensive objective function to be optimized, we also employ a fusion technique, known as model reification [54–56] to fuse data from all information sources to obtain a fused predictive model to estimate the expensive objective function. This fused model encompasses our current state of knowledge during the design process. An update to any of the information sources results in a fused model update representing the system's new state of knowledge. During every iteration, we generate a set of potential design points using a space filling technique, for example, Latin hypercube sampling, evaluate them from each information source's surrogate model and temporarily update the fused model. We then use an acquisition function to quantify the expected change in the system's knowledge about the maximum objective value when evaluating that design point from the information source. The next design point to be evaluated is then selected by choosing the information source and design point that resulted in the largest expected change in the system's knowledge of the maximum objective value.

In the following two subsections, we describe the methods used in the dual-phase steel application for thermodynamic phase prediction and mechanical response prediction. The subsections that follow then detail the different ingredients of our design framework, including a discussion of Gaussian process regression, the active subspace for dimensionality reduction, information fusion via model reification, and acquisition via the knowledge gradient. We end this section with a discussion of different strategies and implementation options of the framework, which are then used in the presentation of results in Section 3.

2.1. Microstructure space prediction

One of the conventional methods for estimating a material's microstructure is to use thermodynamic predictions of the phases. This approach typically relies on CALPHAD based models such as those implemented in the Thermo-Calc™ software. Thermodynamic models are capable of predicting the equilibrium phase fraction and phase composition of a material. As such, they provide critical information on the microstructure of a material. In order to avoid needing to call the thermodynamic software explicitly and also to ensure that the calculations for the microstructure of the material can be achieved quickly, a Gaussian process surrogate model was built from the Thermo-Calc™ predictions. To achieve this, a uniform sampling of the design space was conducted and at each point, the volume fraction and composition of the austenite phase was predicted. The volume fraction of martensite was estimated using the Koistinen-Marburger relation and it was assumed that the martensite had the same composition as the austenite. A Gaussian process model was then fit to each of the volume fractions, and the Si, Mn, and C weight fractions of the martensite phase [57]. Since the material only consists of two phases, the volume fraction and composition of the ferrite phase is achieved by completing a mass balance. These predictions from the thermodynamic model are connected to the mechanical models (described in the following section) to establish a (chemistry) processing-structure-property linkage.

2.2. Mechanical response prediction

In the current work, the prediction of homogenized mechanical response is carried out using five reduced order mechanical models (which we refer to as information sources in the optimization framework) with varying degrees of complexity (computational cost) and fidelity and finally, a high fidelity microstructure-based finite element model (considered as the 'ground truth'). Here, the finite element model utilizes a three-dimensional representative volume element (RVE) to predict the overall stress-strain response of the microstructure [58,59]. The first three reduced order models, 'isostrain' [60], 'isostress' [61] and 'isowork' [62] are mechanical models and are based on simple assumptions of how the strain, stress and work is partitioned respectively among the constituent phases, (martensite and ferrite phases). The other two reduced order models 'secant method' and 'elastic constraint method' are two slightly more sophisticated micromechanical models based on the homogenization schemes proposed in Ref. [63]. The secant method model is based on Hill's weakening constraint power in a plastically-deforming matrix and the elastic constraint model is based on Kröner's treatment of the matrix-inclusion system under elastic constraints.

Here, in all the mechanical models, the constituent ferrite and martensite phases are discretely modeled as an isotropic elastic-plastic material. The Young's modulus and Poisson's ratio are assumed to be the same for both phases. Finally, we relied on several simplifying assumptions when linking the composition of the phases to the mechanical properties. For the ferrite phase there are

typically three strengthening mechanisms that are commonly considered, solid solution strengthening, grain size strengthening, and Precipitation strengthening [64]. In the current work we considered only a single grain size, and we assumed that we were not producing a precipitate phase and so the grain size strengthening effect is constant and the precipitate strengthening effect is ignored. Therefore, the only effect that varies with the composition is the solid solution strengthening. Following a further assumption that the heat treatment is long enough for the austenite and ferrite to reach equilibrium, we know that the concentration of carbon in ferrite will have a negligible effect on the solution strengthening and so we ignore that contribution to simplify the calculations. As a result, the only two components that contribute to the ferrite strength are the Mn and Si concentrations.

While martensite does have some effect from solid solution strengthening, most research has shown that it is predominantly the carbon concentration that controls the mechanical properties of the martensite phase [65]. Therefore, we made a simplifying assumption that only the carbon concentration of the martensite would determine the mechanical properties of the martensite phase. Additional descriptions and details of all the equations and parameters used for these mechanical models can be found in Ref. [15].

2.3. Gaussian process regression

In the context of Bayesian optimization, surrogate models are constructed to estimate the objective values before making a query directly from functions, so that the potential information gain toward the optimum design can be predicted beforehand. Gaussian process models are easy to update and cheap to evaluate, thus we use them to build the surrogates for the Bayesian framework using squared exponential kernel as in Eq. (1).

$$k_i(\mathbf{x}, \mathbf{x}') = \sigma_s^2 \exp \left(-\sum_{h=1}^d \frac{(x_h - x'_h)^2}{2l_h^2} \right), \quad (1)$$

Here, d is the dimensionality of the design space, σ_s^2 is the signal variance, and l_h is a d -dimensional vector with the characteristic length-scales defining the correlations between the points within each dimension. Signal variance and length-scales are estimated using maximum likelihood optimization or set based on the expert opinion about the objective functions. More information regarding Gaussian processes and kernels are provided in Ref. [66].

2.4. Active subspace

The active subspace method is a technique to look for the directions in the design space for which a function has the largest variability. By forming a subspace using those directions, an approximation of a function is obtained on a lower dimensional space called the active subspace. The advantage of constructing a subspace to approximate a function is that learning a subspace of the original high-dimensional design space is easier [51,53]. This advantage leads to significant efficiency gains, speeding up the optimization in design applications and reducing resource usage [67]. Briefly, the idea is to find the directions in the design space which contain the largest variation in the objective value. In other words, a new coordinate system is built based on eigen vectors of the space with eigenvalues defining how strong the variation of the objective value is when moving toward that direction. The matrix $\mathbf{U}$ has n eigen vectors corresponding to the first n largest eigen vectors and is called the transformation matrix. Other eigen vectors are stored in matrix $\mathbf{V}$ which defines an orthogonal space to the active subspace. Although n can be a fixed value for the entire process, a more flexible approach would be to normalize the eigenvalues with respect to the largest one and select the eigen vectors

with eigenvalues larger than a user-defined threshold. This approach helps to form active subspaces different in dimensionality in every iteration if considerable information is available toward any eigen vectors. Any design point in the original design space can be transformed to the active subspace using the transformation matrix:

$$\mathbf{z} = \mathbf{U}^T \mathbf{x} \quad (2)$$

The function g represents the original function f in the active subspace as

$$g(\mathbf{z}) = g(\mathbf{U}^T \mathbf{x}) \approx f(\mathbf{x}) \quad (3)$$

Now, we seek to learn the objective function g in the active subspace instead of the original objective function f on the design space $\mathcal{X}$. A detailed discussion on how to compute the active subspace associated with an objective function is presented in Ref. [53].

After performing the Bayesian optimization over the active subspace and once a candidate point, $\mathbf{z}^*$ which is the projection of $\mathbf{x}^*$ in the high-dimensional space, is selected, it needs to be mapped back to the original design space. This allows the second step of the optimization to identify the best point and information source to query. The challenge here is that there are an infinite number of high dimensional vectors that have the same projection to the point in the lower dimensional space. Therefore, we propose a method to overcome this problem.

Using the definition of orthogonality of eigen vectors of a symmetric matrix, which is the covariance matrix calculated to obtain the eigen vectors, any eigenvector in matrix $\mathbf{V}$ is orthogonal to any eigenvector in the transformation matrix $\mathbf{U}$ or in general, the active subspace. Consequently, any linear combination of eigen vectors in matrix $\mathbf{V}$ is orthogonal to the active subspace. Thus, by writing the equation of vectors created by the linear combination of orthogonal eigen vectors to the active subspace which pass from the design point selected in the first step of optimization, $\mathbf{x}^*$, we are able to generate an infinite number of design points in the higher dimensional space with the same projection to the active subspace.

Assuming the original design space has m dimensions and the active subspace has n dimensions, the matrix $\mathbf{V}$ will have $m - n$ eigen vectors. An orthogonal vector to the active subspace is given as

$$\vec{\mathbf{P}} = \sum_{k=1}^{m-n} a_k \vec{\mathbf{e}}_k, \quad (4)$$

where a_k is a random number that for simplicity, is generated from 0 to 1 and $\vec{\mathbf{e}}_k$ is an eigenvector in $\mathbf{V}$. Then using the orthogonal vector $\vec{\mathbf{P}}$ and $\mathbf{x}^*$, the corresponding design point in the high-dimensional space to $\mathbf{z}^*$, the equation of the linear subspace passing through $\mathbf{x}^*$ and orthogonal to the active subspace is given by

$$\frac{x(1) - x^*(1)}{P(1)} = \frac{x(2) - x^*(2)}{P(2)} = \dots = \frac{x(m) - x^*(m)}{P(m)} = t \quad (5)$$

where a design vector in m -dimensional design space, $\mathbf{x} = [x(1), x(2), \dots, x(m)]^T$, is found by solving m sub-equations for a given t . All design vectors obtained in this step satisfy the relation $\mathbf{z}^* = \mathbf{U}^T \mathbf{x}$. Although all generated design vectors satisfy the relations mathematically, the constraint here is to have all m design variables in the bounds defined by the designer. Therefore, before generating random t values, its range of variability should be specified. First, by replacing the lower bound for every design variable in sub-equations in Eq. (5), m different values are obtained for t

$$t_{low,i} = \frac{lb(i) - x^*(i)}{P(i)}, \quad 1 \leq i \leq m \quad (6)$$

and the same calculations are done for the design variable upper bounds

$$t_{up,i} = \frac{ub(i) - x^*(i)}{p(i)}, \quad 1 \leq i \leq m \quad (7)$$

Next, to have all design variables within the bounds, the lower bound for t from $2m$ values found is the closest value to zero between all negative t values and the upper bound is the closest value to zero between all positive t values. This way, it is guaranteed that all the design variables will remain in their bounds when being mapped back to the m -dimensional design space for any random t generated. Note that since a linear subspace can be expanded in any direction starting from $\mathbf{x}^*$, it is ensured that t can take both negative and positive values. Finally, a set of samples in the original design space is generated, $\mathcal{X}_f$, by mapping back the design point $\mathbf{z}^*$. The acquisition function is then employed for the second time to find the best design to be evaluated next.

2.5. Information fusion

When approximating a quantity of interest in a multi-information source setting, it is assumed that every information source provides some useful information about the ground truth. Therefore, we aim to gather as much information as can be supplied by these information sources to make the most reliable estimations regarding the ground truth quantity of interest. Several approaches exist for fusing multiple sources of information such as Bayesian modeling averaging [68–73], the use of adjustment factors [74–77], covariance intersection methods [78], and fusion under known correlation [79–81]. Some other approaches developed recently in Refs. [82–84] perform joint updates to share information between different fidelity levels considering a continuous fidelity space. Here, we follow Refs. [81,85] for fusion of information sources. In the presence of multiple information sources, the fused mean and variance at a particular design point $\mathbf{x}$ are defined as in Eqs. 8 and 9.

$$\mathbb{E}[\hat{f}(\mathbf{x})] = \frac{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \boldsymbol{\mu}(\mathbf{x})}{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \mathbf{e}}, \quad (8)$$

$$\text{Var}(\hat{f}(\mathbf{x})) = \frac{1}{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \mathbf{e}}, \quad (9)$$

where, $\mathbf{e} = [1, \dots, 1]^T$. Also, $\boldsymbol{\mu}(\mathbf{x}) = [\mu_1(\mathbf{x}), \dots, \mu_S(\mathbf{x})]^T$ and $\tilde{\Sigma}(\mathbf{x})$ are the mean values predicted by each information source and the covariance matrix defining the correlation between the information sources respectively. More detail on this fusion technique and examples of its applications are presented in Refs. [15,20,53,55,56,85–89].

2.6. Knowledge gradient

In the Bayesian optimization framework, an acquisition function is employed to determine the value of a potential query from an objective function looking for the best experiment to run next. We define a two-step look-ahead policy employing knowledge gradient (KG) as in Eqs. 10 and 11.

$$v^{KG}(\mathbf{x}) = \mathbb{E}[V^{N+1}(H^{N+1}(\mathbf{x})) - V^N(H^N)|H^N], \quad (10)$$

$$Q = \mu_{\text{fused}}^* + \alpha x_{\mathbf{x} \in \mathcal{X}} v^{KG}(\mathbf{x}), \quad (11)$$

In a two-step look-ahead strategy, the immediate improvement of a query and the expected improvement in the next iteration are taken into account. The term $v^{KG}(\mathbf{x})$ is the expected improvement estimated using knowledge gradient, while the term μ_{fused}^* repre-

sents the immediate improvement that, in the multi-information source setting, is the maximum mean function value of the fused Gaussian process. By computing Eq. (11) for a set of potential queries from information sources, the best pair of information source and design point maximizing this equation will be selected for the next experiment to run. More details are presented in Ref. [90].

2.7. Strategies and implementation

There are different strategies available for combining the concept of Bayesian optimization of multifidelity systems and the active subspace method. We consider three such strategies here.

The first approach is to build the active subspace upon the ground truth response surface (GT active subspace). The intention behind this decision is to focus on searching the subspace directly related to the design space of the highest fidelity model. At the beginning of every iteration, the ground truth active subspace is formed and all data from other information sources are projected to this subspace. This results in new models defined on a lower dimensional design space. The Bayesian optimization framework is then exploited to search this lower dimensional design space looking for the best potential design to be evaluated to provide the most information about the optimum design.

The point that maximizes the acquisition function value is then selected as the next-best point to evaluate. This point is then mapped back to the original high-dimensional space. Since there are an infinite number of possible solutions when mapping from a low-dimensional to high-dimensional design space, the Bayesian optimization approach is repeated for this solution set. From this optimization approach, the next-best design point and an information source are chosen to be queried.

The second approach is to transform all information sources to the active subspace of the temporary updated information source (TUIS active subspace). In this context, instead of transforming to the active subspace at the beginning of every iteration, transformations are done every time an information source is temporarily updated. Therefore, different active subspaces corresponding to each information source are taken into account and the system might find an information source suggesting larger variation in the objective value and by extension the fused model. Thus, it is likely that a more informative point may be identified. By transforming all active subspaces associated with each information source together in turn, the performance of other information sources in different active subspaces is investigated as well. This results in investigating all information sources and fused models in different active subspaces, resulting in a more informative decision making process.

The last approach is to let every information source operate in its own active subspace independently (Independent active subspaces). In this case, when a design point is to be evaluated, it should be transformed to the corresponding active subspace first. Again, all the processes explained earlier remain the same. This approach offers a cheaper framework in comparison to the second approach since the number of transformations and modeling time will be decreased.

3. Results and discussions

We implement an adaptive active subspace method to efficiently handle the large dimensionality of the design space of a typical PSP-based material design problem within our recently developed closed-loop multifidelity BO framework [15]. Here, we demonstrate the obtained results from utilizing three approaches discussed earlier to implement the adaptive active subspace

method within the multifidelity BO framework. For comparison purposes, the multifidelity BO framework developed in our recent work [15] that does not take advantage of the active subspace method (referred here as the standard approach or Std) is also considered. To show how implementing each of these strategies improves the performance of the optimization, which is crucial in higher dimensional design spaces, the results of Random sampling (Rand) is included as well. In random sampling, an information source and a design are randomly chosen to query at every iteration and no heuristic is employed to search for valuable queries from the information sources.

Fig. 3 shows how quickly each method attained progressively higher objective values as a function of the number of iterations. All models are initialized with 10 random points in the design space. The results are the average of 5 replicas with different initializations, and the filled region shows the 95% confidence intervals. The same 5 sets of 10 random points were used for each of the 5 replications of each method to enable valid comparison of the results. At the end of every iteration, the system chooses a point and an information source to query. Then, after every 10 iterations, the best estimation of the optimal solution suggested by the fused model is evaluated from the ground truth model.

There are some important conclusions that can be drawn from these results. First, the active subspace approach improves the performance of the optimization significantly as each of these methods leads to faster improvement in the objective and reaches the optimal design region more quickly than traditional multi-information source BO. This is because the active subspace approaches search more effectively over lower dimensional regions in the design space by adaptively locating the regions of largest variation (i.e., the most active regions) in the objective value. Therefore, greater jumps in the objective value are observed when the active subspace method is implemented. Next, using the TUIS active subspace method results in slower improvement rates as compared with the GT and Ind approaches. This is related to the need to construct more active subspaces in this approach and force possible deleterious connections between information sources in these temporary subspaces. Employing the independent active subspaces, or Ind approach for each information source has similar performance to using the

Ground Truth active subspace, or GT approach. In both approaches we see superior improvement in the objective as compared with the traditional approach and the TUIS approach. Finally, at iteration 200, all methods have arrived at the optimum design region and there is no advantage of using a particular strategy beyond this point. This is to be expected as each method has acquired enough information at this point in its respective approach to accurately approximate the ground truth objective. Overall, superior improvement rates of the active subspace approaches are associated with the ability of these methods to avoid costly exploration in less important regions of the high-dimensional design space in the early stages of the optimization process.

Although comparing the progression toward the optimum design region based on the number of function evaluations gives a sense of the effectiveness of queries made by employing the active subspace method, a more thoughtful comparison is to compare the time required for each approach to reach a target value. In this case, we consider the cost of modeling, including updating and evaluating the Gaussian processes, calculation of the active subspaces and knowledge gradient in addition to the function evaluations done during the optimization process. In Fig. 3(b), we have illustrated the objective value attained and time required for each approach. These results suggest that using the active subspace method, in particular the GT active subspace or Ind active subspace approaches, results in higher objective values in less computational time in comparison to the standard multifidelity or multi-information source optimization. On the other hand, using the TUIS active subspace approach shows little to no improvement, which is related to the number of active subspace computations and transformations made in a single iteration. While these results show that the conventional multifidelity approach performs as well as the active subspace approach after a certain amount of time, the active subspace approaches provide significant improvement in the results at early stages of the process. In addition to this, the active subspace approach also shows lower variability in the results—this last aspect is very important as low uncertainty is a desired attribute of any design framework.

We note here that the uncertainty (or variance) in the results stems from several sources. First of all, we have used different

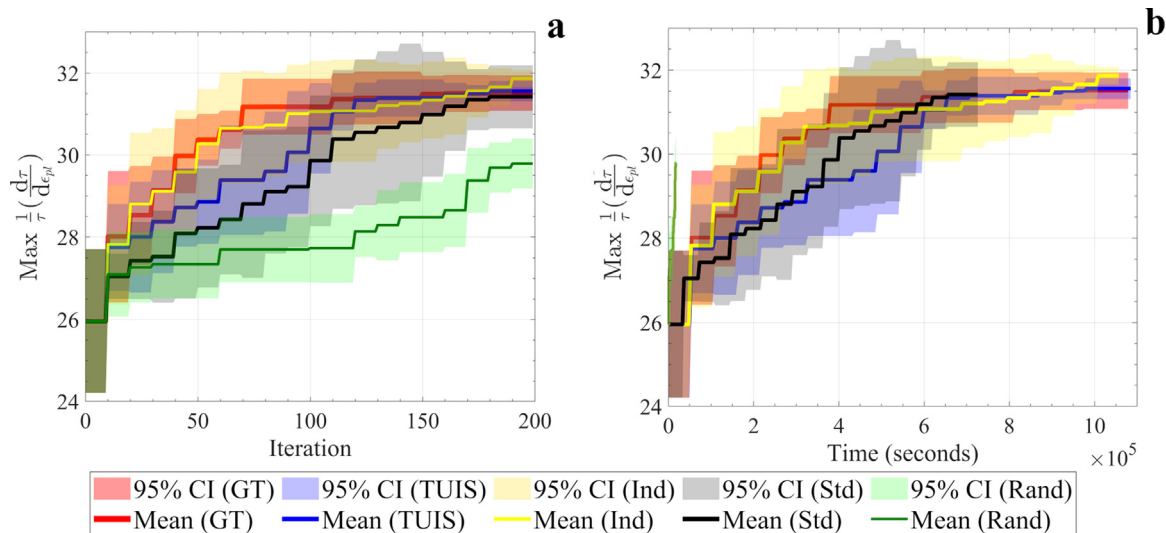


Fig. 3. Estimated optimum objective value as a function of (a) iteration and (b) time. (a) The active subspace methods using Ground Truth (GT), Temporary Updated Information Source (TUIS), and Independent (Ind) active subspaces have been shown to outperform the Standard (Std) approach without applying the active subspace method. Additionally, the result of using Random sampling (Rand) is included to show the superiority of other approaches. (b) In terms of computational cost, the active subspace approaches are again superior. The computational cost accounts for modeling, active subspace and knowledge gradient calculations in addition to the function evaluations. Using the TUIS active subspace is slightly more expensive due to the larger number of active subspace calculations and transformations required. The results are obtained from 5 different initializations and the mean and 95% confidence intervals are shown.

training sets for the initialization of the models, so each iteration starts from different initial conditions. Second, we have modeled the information sources using a stochastic process, namely Gaussian process models. These probabilistic models predict the objective value with normally distributed uncertainty in the prediction. In addition, at every stage of the optimization, we generate random samples using Latin hypercube sampling. Therefore, for each different run, there will be different samples to evaluate as candidate design points. Finally, the ground truth function in this particular design application, RVE, is noisy and can provide different objective values for the same design input [15]. The confidence intervals in Fig. 3 show the total uncertainty since differentiating between each source of uncertainty was not practical.

We are also interested in knowing which are the active subspaces that are preferred at every stage of the design process and what design variables are contributing most to the active subspace formation. The different algorithms used in the current work lead to diverse active subspace configurations and subsequently selections of information sources. In Fig. 4, the cumulative sum of the times every active subspace is chosen has been plotted. The labels show the primary design variables that the active subspace is composed of. As such, the labels indicate the dimensionality of an active subspace and which design variables will be searched preferentially. In other words, all design variables participate and are searched in a particular active subspace, but to different degrees. In Fig. 4 only the main participants of each active subspace are shown and are included in the labels. Note that, the ground truth

is queried every 10^{th} iteration and the corresponding active subspace will be updated accordingly. Fig. 4(a) shows that when using the GT active subspace, the system initially starts searching the carbon space, then it searches the manganese space, and finally, the temperature space.

The same results when using the TUIS and Independent active subspace approaches are shown in Fig. 4(b) and (c). The point here is that the one-dimensional active subspaces are preferred over the higher dimensional active subspaces. This shows that the system finds more value in searching active subspaces mainly composed of one important design variable at a time, and, once the information from the single design variables is exhausted, the system starts searching subspaces with main contributions from a combination of design variables. However, these active subspaces are still smaller than the original design space.

While Fig. 4 provides useful information about important design variables at every iteration in different case studies, it is beneficial to look at how much other variables are participating in the active subspace formation. As mentioned earlier, active subspaces are built upon the most effective directions in the design space that considers the change in all design variables but in different degrees. These directions are the eigen vectors of the covariance matrix defined as

$$\mathbf{C} \approx \frac{1}{M} \sum_{i=1}^M \nabla_{\mathbf{x}} f(\mathbf{x}_i) \nabla_{\mathbf{x}} f(\mathbf{x}_i)^T \quad (12)$$

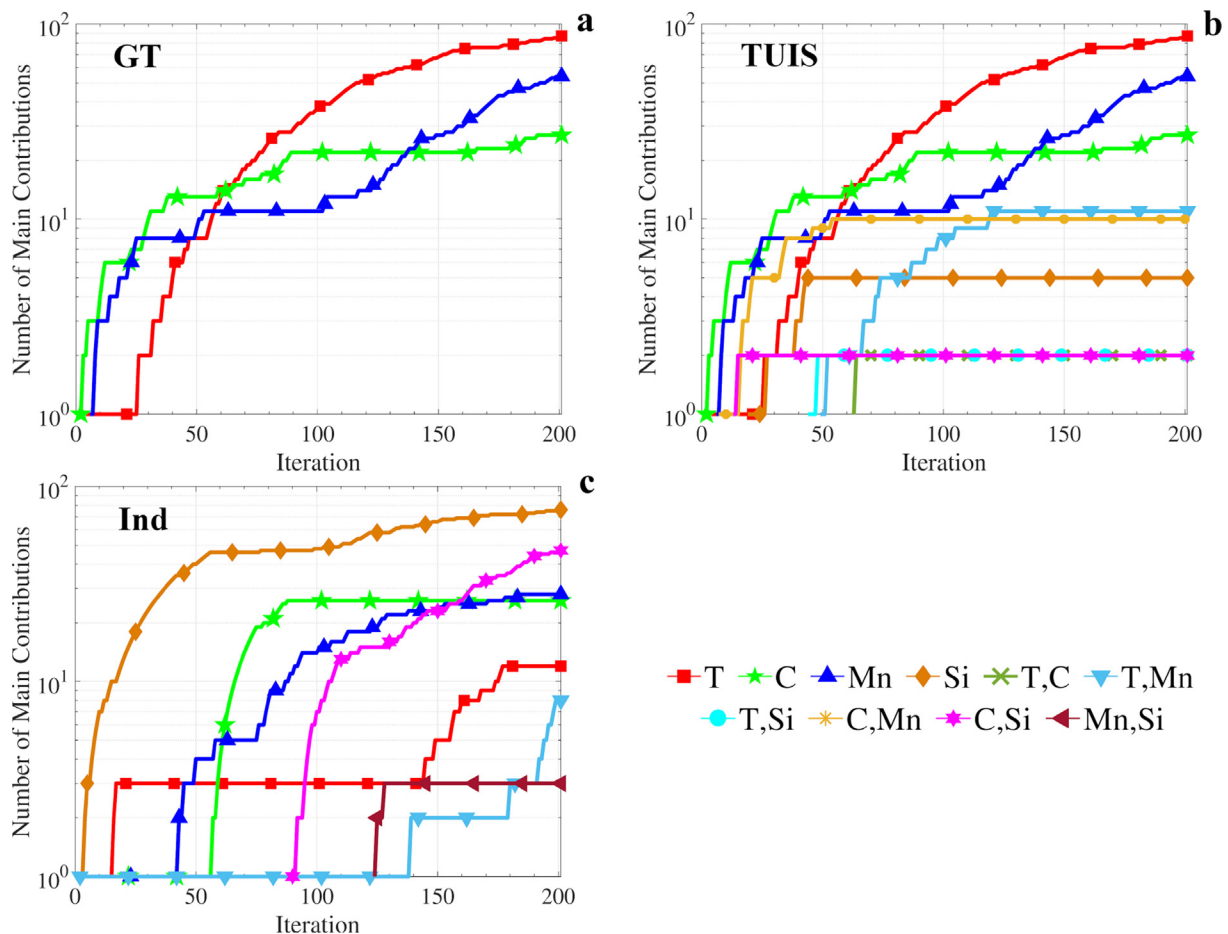


Fig. 4. Number of times a design variable has contributed to form the active subspace. (a) using Ground Truth (GT) active subspace (b) using Temporary Updated Information Source (TUIS) active subspace (c) using Independent (Ind) active subspaces. While Si is not showing any contribution in forming the ground truth active subspace, in the other cases, a variety of design variable combinations are participating to build the active subspace.

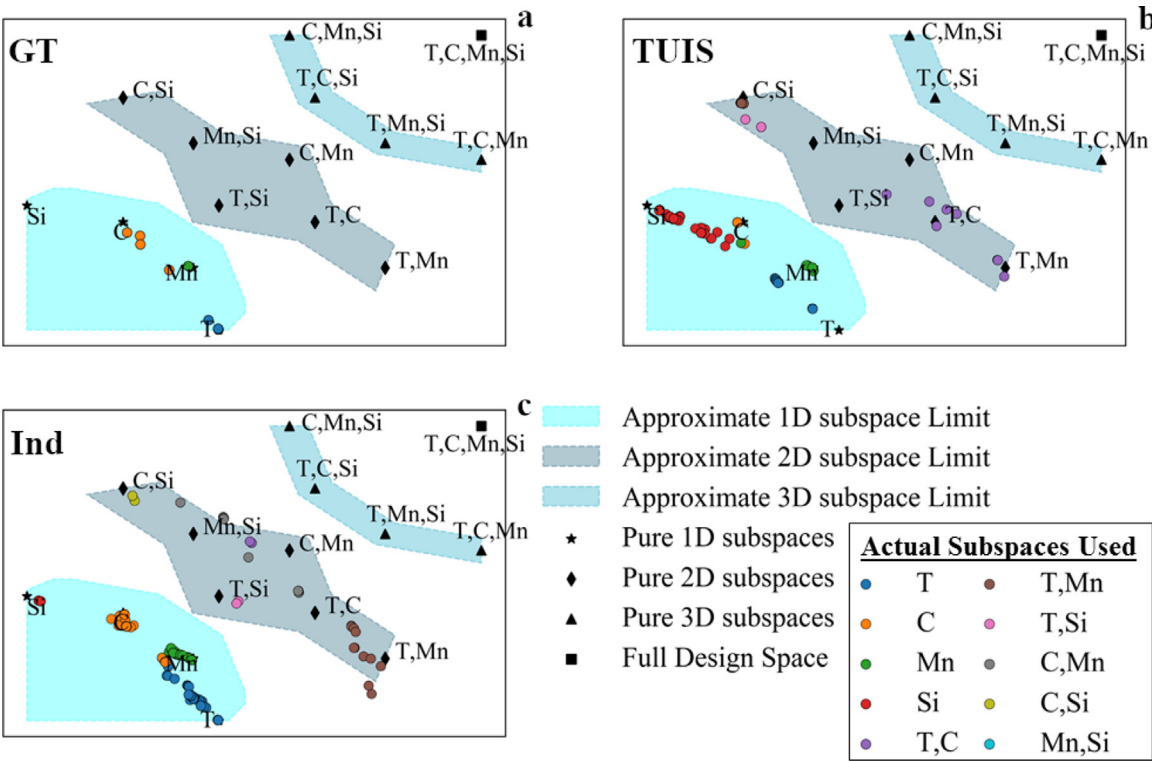


Fig. 5. Distribution of active subspaces within the design space. A 2-D projection over the 4-D design space (a) using Ground Truth (GT) active subspace (b) using Temporary Updated Information Source (TUIS) active subspace (c) using Independent (Ind) active subspaces.

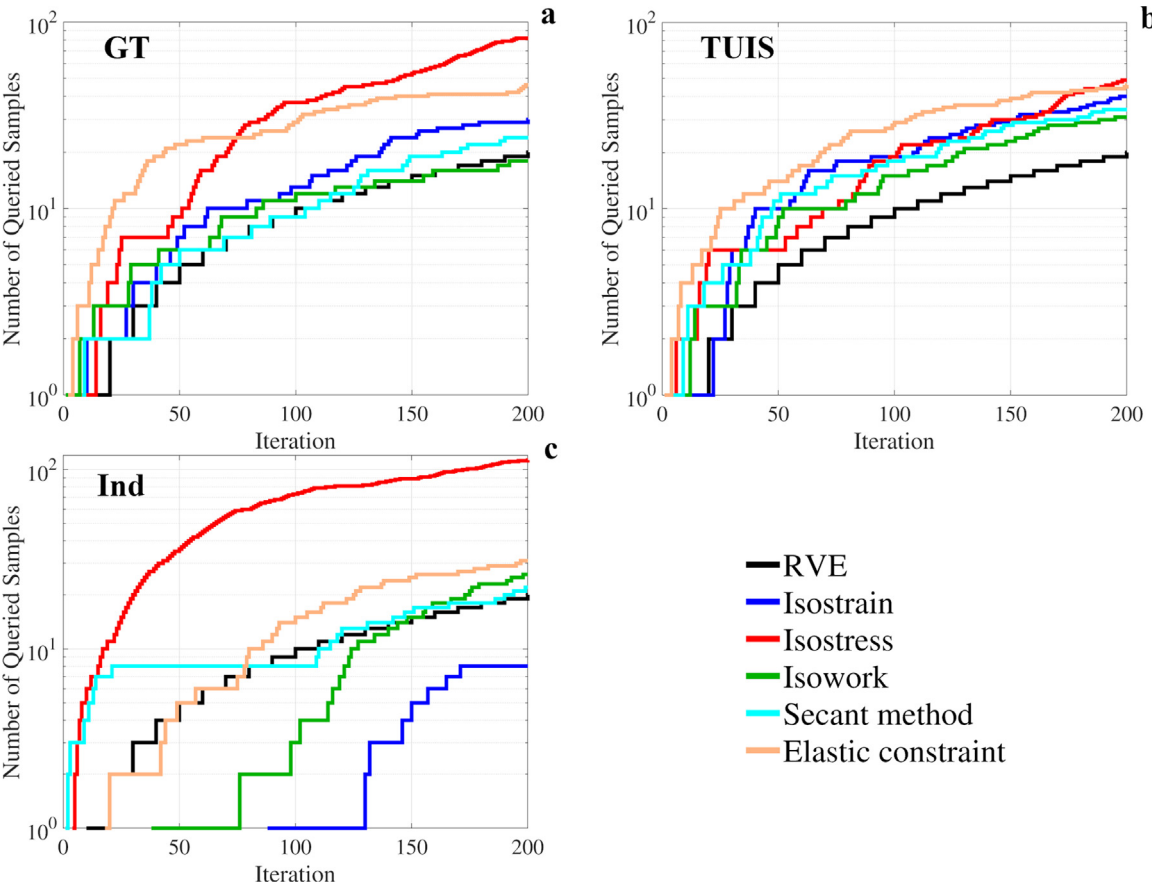


Fig. 6. Number of queries from different information sources and the ground truth (RVE). (a) using Ground Truth (GT) active subspace (b) using Temporary Updated Information Source (TUIS) active subspace (c) using Independent (Ind) active subspaces. the isostress and elastic constraint models highly contributing to provide valuable information regarding the optimum design in all approaches.

assuming M samples are evaluated from the function f previously and the gradient is calculated numerically using a finite difference method. This is done since the function is a 'black-box' and there is no closed-form expression for the gradient. Once the eigen vectors and the associated eigenvalues are found, a single vector is formed using the linear combination of all eigen vectors, each multiplied by their eigenvalue to emphasize the strength of each variable toward a particular direction. At every iteration, we will have a 4-dimensional vector showing the effective participation of every variable based on a scalar value. Since we are not able to show a 4-dimensional space, a simple mathematical projection of the 4-dimensional space to 2-dimensional space was used. This projection allows us to graphically show the distribution of active subspaces within the design space. These results are shown in Fig. 5.

Every point on these projection plots corresponds to a point in the 4-dimensional space. The projections allow us to show the locations of each of the pure 1D, 2D and 3D subspaces (labeled on the figure). The actual subspaces that are used in the calculations for the three approaches are plotted in relation to these. The labels of each of these subspaces is determined by the magnitude of the eigenvalues, where an eigenvalue greater than 0.5 assigns that input dimension to the label. As illustrated in Fig. 5, this approach to labeling allows the points to deviate quite significantly from the pure subspaces, however, the clusters are still visible. We can also quite easily observe that all approaches mostly use 1D and 2D subspaces in the calculations. This is a promising result since it shows that the active subspace approach is operating as expected. We also observe that very few unique subspaces are used in the GT-subspace approach which is likely a result of only using the active subspace of the ground truth model. Both the Individual and TUIS active subspace approaches show a much broader selection of active-subspaces, with many 2D subspaces also being utilized. As a final note, these differences in the active-subspaces used in the optimization do not appear to significantly affect the optimization process, as shown in the results above. The importance of this result is that the use of an active subspace approach is not dependent on the active-subspaces that are used in the optimization.

The contribution of the information sources in the optimization process and the number of queries made from each information source can show which lower fidelity models are providing more valuable knowledge about the optimum design in different case studies. Additionally, since every information source can have a different active subspace, that changes over time as more data is added to the model, the selection of the information source to query is directly affecting results in Fig. 5. It is thus interesting to know about the participation of the information sources as well. Fig. 6 shows the cumulative number of times an information source is queried based on the iterations. At every iteration, only a single information source is queried.

Fig. 6 suggests that for all cases, the isostress and elastic constraint information sources have been selected more than any other information source. It shows that these two models have a smaller discrepancy with the ground truth model around the optimum design point in comparison to the other information sources. In the present work, the focus is on how using active subspace approach results in a more efficient and faster search for the optimal designs in the initial iterations. Therefore, at every iteration, we let the system query the best information source providing the highest information value to the system about the optimal design regardless of how much such queries costs.

4. Conclusions

In this work, we suggested strategies to equip a multifidelity BO framework with a gradient-based active subspace approach to

address the issue of underperforming BO frameworks when the dimensionality of the design space increases. We have demonstrated this framework on a microstructure-sensitive design problem. Although employing a multifidelity BO approach alone results in less costly design procedures, exploring a high-dimensional design space could still be costly and thus, efficiency gains are desired in this regard. The results suggest that by taking advantage of incorporating the active subspace method in the multifidelity BO frameworks, with fewer function evaluations, it is possible to obtain better estimation of the optimal design faster. This is due to the active subspace method prioritizing searches in the important regions in the high-dimensional design space and representing the data on the lower dimensional active subspace that ease the curse of dimensionality problem. By investigating different strategies to use the active subspace method in a multifidelity BO framework, it has been shown that the Ground Truth and Individual Active Subspace approaches performed better than the conventional multifidelity BO approach. Therefore, these two methods can be beneficial to reach a target objective value faster with larger initial steps toward the optimum design. Finally, while we have applied this approach to a very specific class of materials design problems, the framework has wider applicability, as it is often the case that in materials design problems only a small fraction of the degrees of freedom are active at any one time.

Data Availability

The raw data and code required to reproduce the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the financial support provided by the U.S. National Science Foundation through Grant No.NSF-CMMI-1663130, *DEMS: Multi-Information Source Value of Information Based Design of Multiphase Structural Materials*. D.Allaire and R. Arroyave also acknowledge the Grant No. NSF-DGE-1545403. R. Arroyave also acknowledges the Grant Nos. NSF-IIS-1849085 and NSF-IIS-1835690.

References

- [1] J. Allison, Integrated computational materials engineering: A perspective on progress and future steps, *JOM* 63 (4) (2011) 15–18.
- [2] M.F. Horstemeyer, Integrated Computational Materials Engineering (ICME) for metals: using multiscale modeling to invigorate engineering design with science, John Wiley & Sons, 2012.
- [3] L. Johnson, R. Arróyave, An inverse design framework for prescribing precipitation heat treatments from a target microstructure, *Mater. Des.* 107 (2016) 7–17, <https://doi.org/10.1016/j.matdes.2016.06.009>, <https://www.sciencedirect.com/science/article/pii/S0264127516307596>.
- [4] W.Y. Wang, B. Tang, D. Lin, C. Zou, Y. Zhang, S.-L. Shang, Q. Guan, J. Gao, L. Fan, H. Kou, et al., A brief review of data-driven icme for intelligently discovering advanced structural metal materials: Insight into atomic and electronic building blocks, *J. Mater. Res.* 35 (8) (2020) 872–889.
- [5] B. Gautham, R. Kumar, S. Bothra, G. Mohapatra, N. Kulkarni, K. Padmanabhan, More efficient icme through materials informatics and process modeling, in: *Proceedings of the 1st World Congress on Integrated Computational Materials Engineering (ICME)*, Wiley Online Library, 2011, p. 35.
- [6] Q. Zhao, H. Yang, J. Liu, H. Zhou, H. Wang, W. Yang, Machine learning-assisted discovery of strong and conductive Cu alloys: Data mining from discarded experiments and physical features, *Mater. Des.* 197 (2021) 109248, <https://doi.org/10.1016/j.matdes.2020.109248>, <https://www.sciencedirect.com/science/article/pii/S0264127520307838>.

- [7] R. Arróyave, D.L. McDowell, Systems approaches to materials design: past, present, and future, *Annu. Rev. Mater. Res.* 49 (2019) 103–126.
- [8] B. Adams, A. Henrie, B. Henrie, M. Lyon, S. Kalidindi, H. Garmestani, Microstructure-sensitive design of a compliant beam, *J. Mech. Phys. Solids* 49 (8) (2001) 1639–1663, [https://doi.org/10.1016/S0022-5096\(01\)00016-3](https://doi.org/10.1016/S0022-5096(01)00016-3). <http://www.sciencedirect.com/science/article/pii/S0022509601000163>.
- [9] S.R. Kalidindi, J.R. Houskamp, M. Lyons, B.L. Adams, Microstructure sensitive design of an orthotropic plate subjected to tensile load, *Int. J. Plast.* 20 (8) (2004) 1561–1575, <https://doi.org/10.1016/j.ijplas.2003.11.007>. <http://www.sciencedirect.com/science/article/pii/S0749641903001633>.
- [10] G. Saheli, H. Garmestani, B.L. Adams, two phase composite using two-point correlation functions, *J. Comput. Aided Mater. Des.* 11 (2) (2004) 103–115, <https://doi.org/10.1007/s10820-005-3164-3>.
- [11] T. Fast, M. Knezevic, S.R. Kalidindi, Application of microstructure sensitive design to structural components produced from hexagonal polycrystalline metals, *Comput. Mater. Sci.* 43 (2) (2008) 374–383, <https://doi.org/10.1016/j.commatsci.2007.12.002>. <http://www.sciencedirect.com/science/article/pii/S0927025607003369>.
- [12] D.T. Fullwood, S.R. Niezgoda, B.L. Adams, S.R. Kalidindi, Microstructure sensitive design for performance optimization, *Prog. Mater. Sci.* 55 (6) (2010) 477–562, <https://doi.org/10.1016/j.pmatsci.2009.08.002>. <http://www.sciencedirect.com/science/article/pii/S0079642509000760>.
- [13] S. Allain, O. Bouaziz, I. Pushkareva, C. Scott, Towards the microstructure design of DP steels: A generic size-sensitive mean-field mechanical model, *Mater. Sci. Eng.: A* 637 (2015) 222–234, <https://doi.org/10.1016/j.msea.2015.04.017>. <http://www.sciencedirect.com/science/article/pii/S0921509315004141>.
- [14] L. Lin, W. Ren, An implementation of icme in materials information exchanging interfaces, *Mater. Discov.* 12 (2018) 9–19.
- [15] D. Khatamsaz, A. Molkeri, R. Couperthwaite, J. James, R. Arróyave, D. Allaire, A. Srivastava, Efficiently exploiting process-structure-property relationships in material design by multi-information source fusion, *Acta Mater.* 116619 (2021).
- [16] P. Voorhees, G. Spanos, et al., Modeling across scales: a roadmap study for connecting materials models and simulations across length and time scales, TMS, Warrendale, PA 14.
- [17] A. Talapatra, S. Boluki, T. Duong, X. Qian, E. Dougherty, R. Arróyave, Autonomous efficient experiment design for materials discovery with bayesian model averaging, *Phys. Rev. Mater.* 2 (11) (2018) 113803.
- [18] A. Solomou, G. Zhao, S. Boluki, J.K. Joy, X. Qian, I. Karaman, R. Arróyave, D.C. Lagoudas, Multi-objective bayesian materials discovery: Application on the discovery of precipitation strengthened niti shape memory alloys through micromechanical modeling, *Mater. Des.* 160 (2018) 810–827.
- [19] S.F. Ghoreishi, A. Molkeri, A. Srivastava, R. Arroyave, D. Allaire, Multi-information source fusion and optimization to realize icme: Application to dual-phase materials, *J. Mech. Des.* 140 (11) (2018).
- [20] S.F. Ghoreishi, A. Molkeri, R. Arróyave, D. Allaire, A. Srivastava, Efficient use of multiple information sources in material design, *Acta Mater.* 180 (2019) 260–271.
- [21] R. Couperthwaite, A. Molkeri, D. Khatamsaz, A. Srivastava, D. Allaire, R. Arróyave, Materials design through batch bayesian optimization with multisource information fusion, *JOM* 72 (12) (2020) 4431–4443.
- [22] E. Raponi, H. Wang, M. Bujny, S. Boria, C. Doerr, High dimensional bayesian optimization assisted by principal component analysis, in: *International Conference on Parallel Problem Solving from Nature*, Springer, 2020, pp. 169–183.
- [23] T. Hastie, R. Tibshirani, *Generalized additive models* London Chapman and Hall Inc.
- [24] D. Duvenaud, H. Nickisch, C.E. Rasmussen, Additive gaussian processes, *arXiv preprint arXiv:1112.4394*.
- [25] K. Kandasamy, J. Schneider, B. Póczos, High dimensional bayesian optimisation and bandits via additive models, in: *International conference on machine learning*, PMLR, 2015, pp. 295–304.
- [26] Z. Wang, M. Zoghi, F. Hutter, D. Matheson, N. De Freitas, et al., Bayesian optimization in high dimensions via random embeddings, in: *IJCAI*, 2013, pp. 1778–1784.
- [27] B. Choffin, N. Ueda, Scaling bayesian optimization up to higher dimensions: a review and comparison of recent algorithms, in: *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, 2018, pp. 1–6.
- [28] W. Jiang, W. Liao, T. Liu, X. Shi, C. Wang, J. Qi, Y. Chen, Z. Wang, C. Zhang, A voxel-based method of multiscale mechanical property optimization for the design of graded tpms structures, *Mater. Des.* (2021) 109655, <https://doi.org/10.1016/j.matdes.2021.109655>. <https://www.sciencedirect.com/science/article/pii/S0264127521002082>.
- [29] X. Chen, H. Zhou, Y. Li, Effective design space exploration of gradient nanostructured materials using active learning based surrogate models, *Mater. Des.* 183 (2019) 108085, <https://doi.org/10.1016/j.matdes.2019.108085>. <https://www.sciencedirect.com/science/article/pii/S0264127519305234>.
- [30] A. Saltelli, Sensitivity analysis for importance assessment, *Risk Anal.* 22 (3) (2002) 579–590.
- [31] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, *Sensitivity analysis in practice: a guide to assessing scientific models*, vol. 1, Wiley Online Library, 2004.
- [32] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, *Global sensitivity analysis: the primer*, John Wiley & Sons, 2008.
- [33] H. Christopher Frey, S.R. Patil, Identification and review of sensitivity analysis methods, *Risk Anal.* 22 (3) (2002) 553–578.
- [34] D.L. Allaire, K.E. Willcox, A variance-based sensitivity index function for factor prioritization, *Reliab. Eng. Syst. Saf.* 107 (2012) 107–114.
- [35] P.E. Gill, W. Murray, M.H. Wright, *Practical optimization*, SIAM, 2019.
- [36] J. Zhang, S. Bi, G. Zhang, A directional gaussian smoothing optimization method for computational inverse design in nanophotonics, *Mater. Des.* 197 (2021) 109213, <https://doi.org/10.1016/j.matdes.2020.109213>. <https://www.sciencedirect.com/science/article/pii/S0264127520307486>.
- [37] A.C. Antoulas, *Approximation of large-scale dynamical systems*, SIAM, 2005.
- [38] K. Zhou, J.C. Doyle, K. Glover, et al., *Robust and optimal control*, vol. 40, Prentice hall, New Jersey, 1996.
- [39] G.H. Duntman, *Principal components analysis*, no. 69, Sage, 1989.
- [40] H. Abdi, L.J. Williams, *Principal component analysis*, Wiley Interdiscipl. Rev. Comput. Stat. 2 (4) (2010) 433–459.
- [41] J. Shlens, A tutorial on principal component analysis, *arXiv preprint arXiv:1404.1100*.
- [42] S. Wold, K. Esbensen, P. Geladi, *Principal component analysis*, *Chemometrics Intell. Lab. Syst. 2* (1–3) (1987) 37–52.
- [43] C. Kunselman, S. Sheikh, M. Mikkelsen, V. Attari, R. Arróyave, Microstructure classification in the unsupervised context, *Acta Mater.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3683591.
- [44] S.R. Kalidindi, S.R. Niezgoda, A.A. Salem, Microstructure informatics using higher-order statistics and efficient data-mining protocols, *JOM* 63 (4) (2011) 34–41.
- [45] A. Cecen, H. Dai, Y.C. Yabansu, S.R. Kalidindi, L. Song, Material structure-property linkages using three-dimensional convolutional neural networks, *Acta Mater.* 146 (2018) 76–84, <https://doi.org/10.1016/j.actamat.2017.11.053>. <http://www.sciencedirect.com/science/article/pii/S1359645417310443>.
- [46] J.H. Panchal, S.R. Kalidindi, D.L. McDowell, Key computational modeling issues in integrated computational materials engineering, *Comput.-Aided Des.* 45 (1) (2013) 4–25, computer-aided multi-scale materials and product design. doi:10.1016/j.cad.2012.06.006. <http://www.sciencedirect.com/science/article/pii/S0010448512001352>.
- [47] A. Paul, P. Acar, W. keng Liao, A. Choudhary, V. Sundararaghavan, A. Agrawal, Microstructure optimization with constrained design objectives using machine learning-based feedback-aware data-generation, *Comput. Mater. Sci.* 160 (2019) 334–351, <https://doi.org/10.1016/j.commatsci.2019.01.015>. <http://www.sciencedirect.com/science/article/pii/S0927025619300151>.
- [48] J. Jung, J.I. Yoon, H.K. Park, J.Y. Kim, H.S. Kim, Bayesian approach in predicting mechanical properties of materials: Application to dual phase steels, *Mater. Sci. Eng. A* 743 (2019) 382–390, <https://doi.org/10.1016/j.msea.2018.11.106>. <http://www.sciencedirect.com/science/article/pii/S0921509318316320>.
- [49] Z. Li, B. Wen, N. Zabarar, Computing mechanical response variability of polycrystalline microstructures through dimensionality reduction techniques, *Comput. Mater. Sci.* 49 (3) (2010) 568–581, <https://doi.org/10.1016/j.commatsci.2010.05.051>. <http://www.sciencedirect.com/science/article/pii/S0927025610003411>.
- [50] X. Hu, J. Li, Z. Wang, J. Wang, A microstructure-informatic strategy for Vickers hardness forecast of austenitic steels from experimental data, *Mater. Des.* 201 (2021) 109497, <https://doi.org/10.1016/j.matdes.2021.109497>. <https://www.sciencedirect.com/science/article/pii/S0264127521000502>.
- [51] T.M. Russi, *Uncertainty quantification with experimental data and complex system models*, Ph.D. thesis, UC Berkeley (2010).
- [52] P.G. Constantine, E. Dow, Q. Wang, *Active subspace methods in theory and practice: applications to kriging surfaces*, *SIAM J. Sci. Comput.* 36 (4) (2014) A1500–A1524.
- [53] S.F. Ghoreishi, S. Friedman, D.L. Allaire, Adaptive dimensionality reduction for fast sequential optimization with gaussian processes, *Journal of Mechanical Design* 141 (7).
- [54] W.D. Thomason, D.L. Allaire, A model reification approach to fusing information from multifidelity information sources, in: *19th AIAA non-deterministic approaches conference*, 2017, p. 1949.
- [55] D. Allaire, K. Willcox, Fusing information from multifidelity computer models of physical systems, in: *2012 15th international conference on information fusion*, IEEE, 2012, pp. 2458–2465.
- [56] S.F. Ghoreishi, D.L. Allaire, A fusion-based multi-information source optimization approach using knowledge gradient policies, in: *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2018, p. 1159.
- [57] R. Couperthwaite, D. Allaire, R. Arróyave, Utilizing gaussian processes to fit high dimension thermodynamic data that includes estimated variability, *Comput. Mater. Sci.* (2020) 110133, <https://doi.org/10.1016/j.commatsci.2020.110133>. <http://www.sciencedirect.com/science/article/pii/S0927025620306248>.
- [58] A. Srivastava, A. Bower, L. Hector Jr, J. Carsley, L. Zhang, F. Abu-Farha, A multiscale approach to modeling formability of dual-phase steels, *Modell. Simul. Mater. Sci. Eng.* 24 (2) (2016) 025011.
- [59] D. Gerbig, A. Srivastava, S. Osovski, L.G. Hector, A. Bower, Analysis and design of dual-phase steel microstructure for enhanced ductile fracture resistance, *Int. J. Fract.* 209 (1–2) (2018) 3–26.
- [60] W. Voigt, On the relation between the elasticity constants of isotropic bodies, *Ann. Phys. Chem* 274 (1889) 573–587.
- [61] A. Reuss, Berechnung der fließgrenze von mischkristallen auf grund der plastizitätsbedingung für einkristalle, *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angew. Math. Mech.* 9 (1) (1929) 49–58.

- [62] O. Bouaziz, P. Buessler, Mechanical behaviour of multiphase materials: an intermediate mixture law without fitting parameter, *Revue de Métallurgie-Int. J. Metall.* 99 (1) (2002) 71–77.
- [63] G. Weng, The overall elastoplastic stress-strain relations of dual-phase metals, *J. Mech. Phys. Solids* 38 (3) (1990) 419–441.
- [64] J. Orr, Strengthening mechanisms in high strength structural steels*, *J. Microscopy* 94 (3) (1971) 205–214, <https://doi.org/10.1111/j.1365-2818.1971.tb02370.x>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2818.1971.tb02370.x>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2818.1971.tb02370.x>.
- [65] G. Krauss, Martensite in steel: strength and structure, *Mater. Sci. Eng.: A* 273–275 (1999) 40–57, [https://doi.org/10.1016/S0921-5093\(99\)00288-9](https://doi.org/10.1016/S0921-5093(99)00288-9), <https://www.sciencedirect.com/science/article/pii/S0921509399002889>.
- [66] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 2005.
- [67] P. Liu, Y. Yan, X. Zhang, Y. Luo, Z. Kang, Topological design of microstructures using periodic material-field series-expansion and gradient-free optimization algorithm, *Mater. Des.* 199 (2021) 109437, <https://doi.org/10.1016/j.matdes.2020.109437>, <https://www.sciencedirect.com/science/article/pii/S0264127520309734>.
- [68] M. Clyde, Model averaging, In *Subjective and Objective Bayesian Statistics*, second ed., Wiley-Interscience, 2003 (Chapter 13).
- [69] M. Clyde, E. George, Model uncertainty, *Stat. Sci.* 19 (2004) 81–94.
- [70] D. Draper, Assessment and propagation of model uncertainty, *J. Roy. Stat. Soc. Ser. B* 57 (1) (1995) 45–97.
- [71] J. Hoeting, D. Madigan, A. Raftery, C. Volinsky, Bayesian model averaging: A tutorial, *Stat. Sci.* 14 (4) (1999) 382–417.
- [72] E. Leamer, *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, John Wiley & Sons, New York, NY, 1978.
- [73] D. Madigan, A. Raftery, Model selection and accounting for model uncertainty in graphical models using occam's window, *Am. Stat. Assoc.* 89 (428) (1994) 1535–1546.
- [74] A. Mosleh, G. Apostolakis, The assessment of probability distributions from expert opinions with an application to seismic fragility curves, *Risk Anal.* 6 (4) (1986) 447–461.
- [75] J. Reinert, G. Apostolakis, Including model uncertainty in risk-informed decision making, *Ann. Nucl. Energy* 33 (4) (2006) 354–369.
- [76] M. Riley, R. Grandhi, Quantification of modeling uncertainty in aeroelastic analyses, *J. Aircraft* 48 (3) (2011) 866–873.
- [77] E. Zio, G. Apostolakis, Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive waste repositories, *Reliab. Eng. Syst. Saf.* 54 (2–3) (1996) 225–241.
- [78] S. Julier, J. Uhlmann, A non-divergent estimation algorithm in the presence of unknown correlations, in: *proceedings of the American Control Conference*, 1997, pp. 2369–2373.
- [79] S. Geisser, A Bayes approach for combining correlated estimates, *J. Am. Stat. Assoc.* 60 (1965) 602–607.
- [80] P. Morris, Combining expert judgments: A bayesian approach, *Manage. Sci.* 23 (1977) 679–693.
- [81] R. Winkler, Combining probability distributions from dependent information sources, *Manage. Sci.* 27 (4) (1981) 479–488.
- [82] K. Kandasamy, G. Dasarathy, J. Schneider, B. Póczos, Multi-fidelity bayesian optimisation with continuous approximations, in: *International Conference on Machine Learning*, PMLR, 2017, pp. 1799–1808.
- [83] K. Kandasamy, G. Dasarathy, J. Oliva, J. Schneider, B. Poczos, Multi-fidelity gaussian process bandit optimisation, *J. Artif. Intell. Res.* 66 (2019) 151–196.
- [84] R. Sen, K. Kandasamy, S. Shakkottai, Multi-fidelity black-box optimization with hierarchical partitions, in: *International conference on machine learning*, PMLR, 2018, pp. 4538–4547.
- [85] R.L. Winkler, Combining probability distributions from dependent information sources, *Manage. Sci.* 27 (4) (1981) 479–488.
- [86] D. Khatamsaz, L. Peddareddygar, S. Friedman, D.L. Allaire, Efficient multi-information source multiobjective bayesian optimization, in: *AIAA Scitech 2020 Forum*, 2020, p. 2127.
- [87] D. Khatamsaz, D.L. Allaire, A comparison of reification and cokriging for sequential multi-information source fusion, in: *AIAA Scitech 2021 Forum*, 2021, p. 1477.
- [88] S.F. Ghoreishi, D. Allaire, Multi-information source constrained bayesian optimization, *Struct. Multidiscipl. Optimiz.* 59 (3) (2019) 977–991.
- [89] D. Khatamsaz, L. Peddareddygar, S. Friedman, D. Allaire, Bayesian optimization of multiobjective functions using multiple information sources, *AIAA J.* (2021) 1–11.
- [90] W.B. Powell, I.O. Ryzhov, *Optimal learning*, vol. 841, John Wiley & Sons, 2012.