Anton van Beek
Department of Mechanical Engineering,
Northwestern University,
Evanston, IL 60208
e-mail: antonvanbeek2022@u.northwestern.edu

Umar Farooq Ghumman
Department of Mechanical Engineering,
Northwestern University,
Evanston, IL 60208
e-mail: umarghumman2018@u.northwestern.edu

Joydeep Munshi
Department of Mechanical Engineering and Mechanics,
Lehigh University,
Bethlehem, PA 18015
e-mail: jom317@lehigh.edu

Siyu Tao
Department of Mechanical Engineering,
Northwestern University,
Evanston, IL 60208
e-mail: siyutao2020@u.northwestern.edu

TeYu Chien
Department of Physics and Astronomy,
University of Wyoming,
Laramie, WY 82071
e-mail: tchien@uwhy.edu

Ganesh Balasubramanian
Department of Mechanical Engineering and Mechanics,
Lehigh University,
Bethlehem, PA 18015
e-mail: bganesh@lehigh.edu

Matthew Plumlee
Department of Industrial Engineering and Management Sciences,
Northwestern University,
Evanston, IL 60208
e-mail: mplumlee@northwestern.edu

Daniel Apley
Department of Mechanical Engineering,
Northwestern University,
Evanston, IL 60208
e-mail: apley@northwestern.edu

Wei Chen[1]
Department of Mechanical Engineering,
Northwestern University,
Evanston, IL 60208
e-mail: weichen@northwestern.edu

Scalable Adaptive Batch Sampling in Simulation-Based Design With Heteroscedastic Noise

In this study, we propose a scalable batch sampling scheme for optimization of simulation models with spatially varying noise. The proposed scheme has two primary advantages: (i) reduced simulation cost by recommending batches of samples at carefully selected spatial locations and (ii) improved scalability by actively considering replicating at previously observed sampling locations. Replication improves the scalability of the proposed sampling scheme as the computational cost of adaptive sampling schemes grow cubicly with the number of unique sampling locations. Our main consideration for the allocation of computational resources is the minimization of the uncertainty in the optimal design. We analytically derive the relationship between the "exploration versus replication decision" and the posterior variance of the spatial random process used to approximate the simulation model's mean response. Leveraging this reformulation in a novel objective-driven adaptive sampling scheme, we show that we can identify batches of samples that minimize the prediction uncertainty only in the regions of the design space expected to contain the global optimum. Finally, the proposed sampling scheme adopts a modified preposterior analysis that uses a zeroth-order interpolation of the spatially varying simulation noise to identify sampling batches. Through the optimization of three numerical test functions and one engineering problem, we demonstrate (i) the efficacy and of the proposed sampling scheme to deal with a wide array of stochastic functions, (ii) the superior performance of the proposed method on all test functions compared to existing methods, (iii) the empirical validity of using a zeroth-order approximation for the allocation of sampling batches, and (iv) its applicability to molecular dynamics simulations by optimizing the performance of an organic photovoltaic cell as a function of its processing settings.
[DOI: 10.1115/1.4049134]

Keywords: uncertainty quantification, surrogate modeling, simulation-based design

1 Introduction

Sampling of computer simulations is a well-studied subject in the field of simulation-based design with applications in engineering [1], biological [2], and social sciences [3]. These research efforts

[1]Corresponding author.
Contributed by the Design Engineering Division of ASME for publication in the Journal of Mechanical Design. Manuscript received June 1, 2020; final manuscript received October 12, 2020; published online December 15, 2020. Special Editor Mian Li.

have driven the optimization of increasingly sophisticated designs and provided insight into complex social and physical phenomena. Available sampling methods are tailored for deterministic simulation models and are either unable or inefficient at optimizing stochastic simulations. What is more, the use of stochastic simulation models is becoming more commonplace in many scientific fields. We are specifically motivated by molecular dynamics (MD) simulations in the context of engineering design as MD simulations are known for having dramatically changing signal-to-noise ratios across sampling locations. An example of using MD simulations for the performance prediction and design of an organic photovoltaic cell (OPVC) is presented in Fig. 1. The left section of this figure shows the initial experimental design for two design variables that correspond to a set of nine unique combinations of design variables that are used as inputs to the MD simulation (central left figure). Running these simulations provides a training data set to which a surrogate model can be trained as illustrated in the central right image. The surrogate model provides insight into the response surface and helps decide what sample(s) to simulate next. This cycle is repeated until a convergence criterion has been met and an OPVC with globally optimal properties is successfully identified. The challenge in this effort is learning the spatially varying noise of MD simulations, as represented by the vertical purple bars in the central right image. In this study, we propose an optimization scheme for objective-driven adaptive sampling of costly stochastic functions with spatially varying noise.

When simulation models are computationally expensive, training a response surface model to a data set obtained from an experimental design provides the designer with a prediction of the response at unobserved input locations. For a general introduction into surrogate modeling and the validation of their fidelity, we refer the reader to Ref. [4]. A notable type of surrogate model are Gaussian processes (GPs), which have seen prolific use among many scientific communities [5]. One advantage of GP models over other types of surrogate models is that they provide a predictive distribution of the response at unobserved sampling locations. GP models enable a designer to improve the predictive capabilities of their surrogate model by running additional computer experiments in regions where the uncertainty is largest. Such efforts are known as adaptive sampling for global surrogate modeling [6]. Alternatively, we are interested in the allocation of simulation resources to only minimize the prediction uncertainty at and around sampling locations expected to contain the global optimum. This type of effort is known as adaptive sampling for global optimization and involves a delicate balance between exploring regions of the design space with large uncertainty versus exploiting regions with a good mean response [7,8]. The use of surrogate models has been extended to the optimization of multifidelity simulations [9,10], nonmyopic sampling [11–13], optimization of nonstationary functions [14], and robust design [15].

Despite the abundance of work on objective-driven adaptive batch sampling, the optimization of stochastic simulation models is still an elusive task. A simple version of the problem is the case where the noise intrinsic to the simulation model is constant over the design space. In other words, the noise is *homoscedastic*. Conventional adaptive sampling schemes are inefficient at optimizing stochastic functions because the existence of prediction uncertainty at observed spatial locations results in sampling decisions that greatly favor exploitation. In Ref. [16], a weighted acquisition function is proposed that overcomes this limitation for simulations models with homoscedastic noise. However, when the *intrinsic noise* varies over the design space (i.e., the noise is *heteroscedastic*), the problem becomes more complex. In the case of MD simulations, the intrinsic noise comes from random initial conditions (i.e., the momentum and coordinates of each particle) that can be only simulated for a finite length and time scale. GP-based surrogate models have been extended to approximate stochastic functions, some popular examples of which are the variational heteroscedastic Gaussian process (VHGP) [17], practical Kriging (PK) [18,19], and stochastic Kriging (SK) [20]. The VHGP model has been extended for application in global optimization in Ref. [21], but quickly becomes intractable as the computational cost increases cubicly with the number of samples. However, SK and PK do no share this limitation as they allow sampling locations to be *replicated* (i.e., resample at previously observed sampling locations). Replicating samples provides insight into the pure variance of the simulation at a single sampling location. However, PK has many parameters that are challenging to tune during the training process. For this reason, SK provides a promising surrogate for the optimization of stochastic simulations. Preliminary efforts in this direction have been proposed in Refs. [22–24]. However, the ratio of exploration to replication must be defined before initiating these sampling schemes. This is a decision that greatly influences the efficiency and the stability of the sampling process as the optimal ratio of exploration to replication depends on the magnitude and variability of the intrinsic modeling uncertainty and will on most cases be unknown to the designer a priori.

A desirable property for any sampling scheme is the recommendation of sampling batches as this facilitates parallel evaluation of the costly simulation model to mitigate the computational cost. Examples of batch-based adaptive sampling schemes for global surrogate modeling and optimization are given in Refs. [25–27]. Batch-based sampling for optimization is a more nuanced challenge compared to global surrogate modeling as it requires the observations of the costly objective function at yet to be sampled spatial locations. The desire for batch sampling is particularly prevalent in the MD community, where it has become a common practice to use supercomputers to run many simulations in parallel. Regardless of the application, batch sampling is a desirable feature in optimizing stochastic functions as an increased number of evaluations are required to distinguish the mean response from the intrinsic noise [22,28].

In this article, we propose a tractable, objective-driven adaptive batch sampling scheme for the optimization of stochastic simulation models with heteroscedastic noise. The main challenge of this study centers around the decision of whether to explore a new sampling

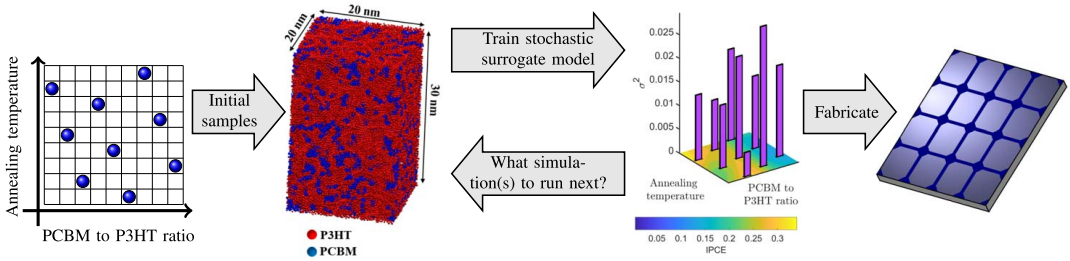

Fig. 1 Example of an MD simulation (with spatially varying noise) for the efficiency prediction and optimization of an OPVC (Color version online.)

location or replicate at a previously observed sampling location, a decision that is to be made under the consideration that we want to minimize the predictive variance only in the region of the design space expected to contain the global minimum. The contribution of this study centers around a new sampling scheme that includes the active consideration of deciding when and where to replicate and when and where to explore. We show that by using the Sherman–Morrison–Woodbury formula, we can reformulate the surrogate models' posterior predictive variance to separate the individual contribution of intrinsic modeling uncertainty and the interpolation uncertainty to the surrogate model's posterior predictive variance. Subsequently, the algorithm decides whether to replicate (reduce intrinsic uncertainty) or explore (reduce interpolation uncertainty) by choosing the candidate that minimizes the posterior variance at the region expected to contain the global optimum. Finally, to benefit from parallel computing capabilities, we propose a preposterior analysis that facilitates the allocation of batches of samples that is compatible with the SK and PK surrogate models. The performance of the proposed sampling scheme is demonstrated on three test functions and one engineering problem that involves the optimization of an OPVC (as presented in Fig. 1). Through the three test functions we (i) demonstrate the proposed sampling scheme's comparatively strong convergence properties, (ii) demonstrate its ability to scale to higher dimensional problems, and (iii) empirically prove its structural validity. The results exemplify that the developed sampling scheme provides a reliable and efficient approach for the optimization of stochastic functions with high-dimensional inputs and/or complex intrinsic noise functions.

2 Background

In this section, we first provide an introduction into surrogate modeling for stochastic functions and elucidate the importance of replication when optimizing high-dimensional or noisy functions. Next, we provide an introduction of two existing adaptive sampling schemes for optimization of stochastic functions and their limitations.

2.1 Surrogate Modeling of Simulations With Intrinsic Noise.

Given a data set of noisy responses $\mathbf{Y} = \{y_1, \ldots, y_N\}^T$ observed at a set of d-dimensional sampling locations $\boldsymbol{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}^T$, we want to train a surrogate on the scalar-valued function $f: \mathbb{R}^d \to \mathbb{R}$. Under the assumption that the observed responses at a set of inputs are jointly normally distributed, we can place a GP prior on the unknown function f so that it can be characterized by a mean trend and a covariance or kernel function $k: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. Normalizing the observed responses to have zero mean centers the surrogate modeling effort around the covariance/kernel structure (e.g., power exponential or Matérn [5]). In this study, we consider the common case of a stationary kernel $k(\mathbf{x}, \mathbf{x}') = \sigma^2\ c(\mathbf{x} - \mathbf{x}'|\boldsymbol{\omega})$, where σ^2 is known as the prior variance and $\boldsymbol{\omega}$ are the roughness parameters that characterize the correlation function $c(\cdot)$.

By using a GP, we can model the observations as a function of their spatial location $y_i = f(\mathbf{x}_i) + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0, r(\mathbf{x}_i))$ are normally independently distributed and accounts for the intrinsic model uncertainty. If the intrinsic uncertainty in the observations is constant (i.e., $r(\mathbf{x}) = \gamma$), then we are dealing with a model that has homoscedastic noise; however, it is often found that the noise varies as a function of the design variables, in which case we are dealing with a model that exhibits heteroscedastic noise. In both scenarios, we can model the observed training data set as follows:

$$\mathbf{Y} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{K}_N + \boldsymbol{\Sigma}_N) \tag{1}$$

where $\mathbf{K}_N$ is an $N \times N$ covariance matrix with the (i, j)th element being $k(\mathbf{x}_i, \mathbf{x}_j)$ and the intrinsic simulation uncertainty is captured by $\boldsymbol{\Sigma}_N = \text{diag}(r(\mathbf{x}_1), \ldots, r(\mathbf{x}_N))$.

Conditioning the GP prior on a set of observations provides the posterior predictive distribution at an unobserved sampling location as $Y(\mathbf{x})|\mathbf{Y} \sim \mathcal{N}\left(\mu_N(\boldsymbol{x}), \sigma_N^2(\boldsymbol{x})\right)$, where

$$\mu_N(\boldsymbol{x}) = \mathbf{k}_N^T(\mathbf{x})(\mathbf{K}_N + \boldsymbol{\Sigma}_N)^{-1}\mathbf{Y} \tag{2}$$

$$\sigma_N^2(\boldsymbol{x}) = k(\mathbf{x}, \mathbf{x}) + r(\mathbf{x}) - \mathbf{k}_N^T(\mathbf{x})(\mathbf{K}_N + \boldsymbol{\Sigma}_N)^{-1}\mathbf{k}_N(\mathbf{x}) \tag{3}$$

and $\mathbf{k}_N(\mathbf{x}) = \{k(\mathbf{x}, \mathbf{x}_1), \ldots, k(\mathbf{x}, \mathbf{x}_N)\}^T$. Prediction of the posterior response requires the identification of the hyperparameters $\boldsymbol{\omega}$ that characterize the kernel function. By rewriting the covariance structure of our GP predictor as $\mathbf{K}_N + \boldsymbol{\Sigma}_N = \sigma^2(\mathbf{C}_N + \boldsymbol{\Delta}_N)$, we obtain the maximum likelihood estimation of the hyperparameters as follows:

$$\hat{\boldsymbol{\omega}} = \arg\max_{\boldsymbol{\omega} \in \Omega}\left(-N \log \hat{\sigma}^2 - \log|\mathbf{C}_N + \boldsymbol{\Delta}_N|\right) \tag{4}$$

where $\Omega \subset \mathbb{R}^d$ is the admissible space of the hyperparameters and we use the well-known closed-form expression to identify $\hat{\sigma}^2 = N^{-1}\mathbf{Y}^T(\mathbf{C}_N + \boldsymbol{\Delta}_N)^{-1}\mathbf{Y}$ for the MLE of σ^2 [25,29]. Note that both terms on the right side of Eq. (4) depend on the roughness parameters $\boldsymbol{\omega}$ through the correlation function.

One function evaluation during the optimization of Eq. (4) requires the inversion and determinant computation of $\mathbf{C}_N + \boldsymbol{\Delta}_N$ that come at a cubic computational expense $\mathcal{O}(N^3)$. This is admissible for most deterministic (i.e., $r(\mathbf{x}) = 0$) or homoscedastic cases (i.e., $r(\mathbf{x}) = \gamma$) [30]; however, in the heteroscedastic case, it is reasonable to expect that more simulation model evaluations are necessary. We address this issue by allowing replication at previously observed sampling locations. Consider that we have n unique sampling locations $\bar{\mathbf{x}}_i \quad (i = 1, \ldots, n)$, where at the ith sampling location, we have observed a_i replicates $y_i^{(j)}, \quad (j = 1, \ldots a_i)$, (i.e., $\sum_{i=1}^n a_i = N$). If we define $\bar{\mathbf{Y}} = \left\{\bar{y}_1, \ldots, \bar{y}_n\right\}^T$ to be the sampling average over replicates (i.e., $\bar{y}_i = \dfrac{1}{a_i}\sum_{j=1}^{a_i} y_i^{(j)}$), then the posterior mean and variance are obtained as follows:

$$\mu_n(\boldsymbol{x}) = \mathbf{k}_n^T(\mathbf{x})\left(\mathbf{K}_n + \mathbf{A}^{-1}\boldsymbol{\Sigma}_n\right)^{-1}\bar{\mathbf{Y}} \tag{5}$$

$$\sigma_n^2(\boldsymbol{x}) = k(\mathbf{x}, \mathbf{x}) + r(\mathbf{x}) - \mathbf{k}_n^T(\mathbf{x})\left(\mathbf{K}_n + \mathbf{A}^{-1}\boldsymbol{\Sigma}_n\right)^{-1}\mathbf{k}_n(\mathbf{x}) \tag{6}$$

where $\mathbf{K}_n = \left[k(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j)\right]_{1 \le i,j \le n}$, $\mathbf{A} = \text{diag}(a_1, \ldots, a_n)$, and the computational complexity has been reduced to $\mathcal{O}(n^3)$.

The formulation given in Eqs. (5) and (6) is known as SK and holds two challenges. The first challenge is that $\boldsymbol{\Sigma}_n$ requires the designer to know the intrinsic noise at each sampled location $r(\bar{\mathbf{x}}_i) \quad (i = 1, \ldots, n)$; however, in many practical cases, a designer has no access to this information. As an alternative, Ref. [20] proposes an estimate for $\hat{\boldsymbol{\Sigma}}_n = \text{diag}(\hat{r}(\bar{\boldsymbol{x}}_1), \ldots, \hat{r}(\bar{\boldsymbol{x}}_n))$ by taking

$$\hat{r}(\bar{\boldsymbol{x}}_i) = \frac{1}{a_i - 1}\sum_{j=1}^{a_i}\left(y_i^{(j)} - \bar{y}_i\right)^2 \tag{7}$$

Using the sampling variance of Eq. (7) results in an unbiased approximation of $\mu_n(\mathbf{x})$ when $a_i \gg 1$ (it is recommended to have $a_i \ge 10$). The second limitation in the formulation of the posterior predictive variance σ_n^2 is that the designer requires to know the intrinsic noise over the entire design space $r(\mathbf{x})$. One approach to address this issue is to omit $r(\mathbf{x})$ in Eq. (5) and be satisfied with a "denoised" predictive variance as $\tilde{\sigma}_n^2(\mathbf{x}) = \sigma_n^2(\mathbf{x}) - r(\mathbf{x})$ [18]. In fact, in our case, this is more meaningful than including it, as we want to quantify the uncertainty in the underlying response surface, as opposed to the noisy response observations. An alternative approach is to place a separate GP prior on $r(\mathbf{x})$ as proposed in Ref. [20], and for which the joint likelihood of both surrogates and a set of latent parameters associated with the intrinsic noise has been derived in Ref. [18] (i.e., the PK surrogate mentioned in Sec. 1). In this study, we will be using the SK model as it is easier to train; however, it should be noted that the proposed sampling scheme will also work with the PK model.

2.2 Objective-Driven Adaptive Sampling for Deterministic and Stochastic Simulations. Throughout the literature, a great number of acquisition functions have been proposed for objective-driven sampling of deterministic functions (i.e., functions with $\boldsymbol{\Sigma}_N = \mathbf{0}$). Some popular acquisition functions include statistical lower bound [31], probability of improvement [7], expected improvement (EI) [32], knowledge gradient [24], and entropy search [33]. Despite the attention that objective-driven adaptive sampling schemes have received, none of them outperforms the others on all optimization problems. Also, many of these functions have unique properties that make them suitable for specific types of problems, but the designer will frequently not have the necessary information to make this assessment a priori.

Objective-driven sampling balances the need for reducing the posterior predictive variance by exploring new sampling locations, versus exploitation of samples near the predicted global optimum. The EI for the minimization of an objective function is defined as follows:

$$EI(\mathbf{x}) = \big((y_{\min} - \mu(\mathbf{x})\big)\Phi(u) + S^2(\mathbf{x})\phi(u) \tag{8}$$

where $y_{\min}$ is the current best function observation, $\Phi(\cdot)$ is the standard normal cumulative distribution function, $\phi(\cdot)$ is the standard normal probability density function, $\mu(\cdot)$ is the posterior mean prediction of a deterministic GP model, $S(\cdot)$ is the posterior variance of a deterministic GP model, and $u = (y_{\min} - \mu(\mathbf{x}))/\sigma^2(\mathbf{x})$. A new sample $\mathbf{x}_{\text{new}}$ is selected by maximizing the EI of the objective function, i.e.,

$$\mathbf{x}_{\text{new}} = \arg\max_{\mathbf{x}\in\chi} EI(\mathbf{x}) \tag{9}$$

where $\chi \in \mathbb{R}^d$ is the admissible design space. The EI function presents a versatile acquisition function that has good convergence properties for a broad range of problems [31,34]; however, its application to simulation models with intrinsic noise has been hindered by its inability to replicate at previously observed sampling locations.

Example of sampling schemes that have been proposed for the optimization of simulation models with heteroscedastic noise include minimum quantile (MQ) [23], correlated knowledge gradient (CKG) [35], expected quantile improvement (EQI) [36], and two-stage sequential optimization (TSSO) [22,37]. The MQ scheme is relatively straightforward in that it selects the point with the minimum Kriging quantile defined as follows:

$$q(\mathbf{x}) = \mu_n(\mathbf{x}) + \Phi^{-1}(\theta)\sigma_n(x) \tag{10}$$

where $\theta \in (0, 0.5]$. The advantage of MQ over the CKG and EQI is that it does not require information on the intrinsic modeling uncertainty $r(\mathbf{x})$. Moreover, the scheme allocates a fixed number of replicates B to each identified additional sample, a number that must be set by the designer and is referred to as the batch size. Theoretically the acquisition function in MQ is never zero and thus allow replication at previously observed sampling locations; however, this is unlikely to happen in practice as we have a real-valued search space.

One drawback of the MQ scheme is that has a fixed number of replicates at each sampling location, a limitation that is overcome by the TSSO scheme [22]. The TSSO scheme adopts a search and replication step during each micro-trial (i.e., one cycle of training the surrogate model, optimizing the acquisition function, and simulating new design(s)). During the search step, $n_{\text{rep}} < B$ replicates are allocated to a new sampling location found through a modified EI acquisition function given as follows:

$$EI_\epsilon(\mathbf{x}) = \big((\bar{y}_{\min} - \mu_n(\mathbf{x})\big)\Phi(\bar{u}) + S_n^2(\mathbf{x})\phi(\bar{u}) \tag{11}$$

where $\bar{y}_{\min}$ is the lowest predicted mean response at any of the previously observed sampling locations, $S_n^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}_n^T(\mathbf{x})\mathbf{K}_n^{-1}\mathbf{k}_n(\mathbf{x})$ is the deterministic posterior variance (i.e., $\mathbf{A}^{-1} = \text{diag}(0, \ldots, 0)$ and $r(\mathbf{x}) = 0$ in Eq. (6)), and $\bar{u} = \dfrac{\bar{y}_{\min} - \mu_n(\mathbf{x})}{S^2(\mathbf{x})}$. The intuition behind using $S_n(\mathbf{x})$ as the posterior variance is that exploring a new sampling location will provide no information on the intrinsic noise of the objective function and should therefore not be considered when deciding where to explore. Finally, during the replication step, the remaining $B - n_{\text{rep}}$ samples are allocated to the previously simulated sampling locations according to an optimal computing budget allocation heuristic. This approach ensures that more replications are allocated at and around the global optimum, but TSSO still has a fixed ratio of exploration to replication, and therefore, its performance greatly depends on the parameters selected by the designer (the parameters here are B and n_{rep}).

3 Proposed Scheme: When to Replicate and When to Explore

In this section, we introduce the proposed adaptive sampling scheme that actively decides where and when to explore a new sampling location and where and when to replicate at a previously observed sampling location.

3.1 A New Scheme for the Optimization of Functions With Heteroscedastic Noise. To initialize the sampling scheme as depicted by step 0 in Fig. 2, the designer has to provide a set of space-filling samples $\mathbf{X}_0$ uniformly dispersed over the design space (e.g., through a Latin hypercube design [38] or a Sobol sequence [39]). This implies that the designer has to determine the total number of unique sampling locations and the number of replications. A typical number of initial sampling locations for the purpose of adaptive sampling is in between $2d$ to $5d$ [6], and the minimum number of recommended replications is $\beta \geq 10$ as suggested in Ref. [20]. A designer can choose to add a higher

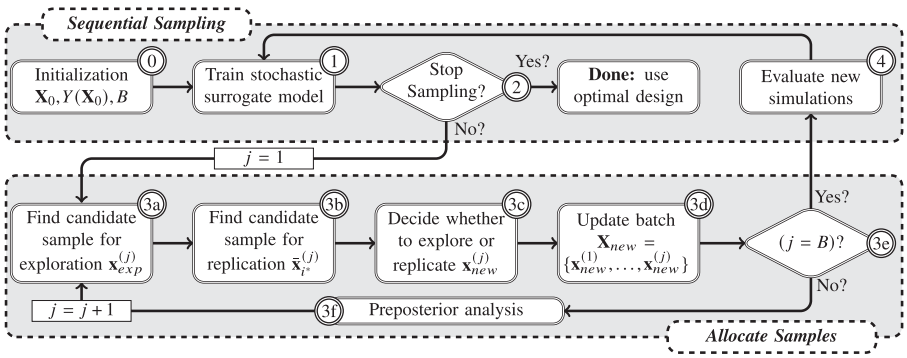

Fig. 2 Proposed sequential adaptive batch sampling scheme for the optimization of simulation models with heteroscedastic noise

number of replicates when dealing with a simulation model that has a high magnitude of intrinsic noise although this is not necessary as shown in Sec. 4.1.2. In addition to the initial experimental design, the designer has to indicate the desired batch size B, and this informs how many samples the proposed sampling scheme will recommend during each micro-trial. A micro-trial is defined as going through steps 1–4 in Fig. 2.

In step 1 of Fig. 2, we train the SK model to provide a predictive distribution of the response $f(\mathbf{x})$ at unobserved spatial locations. Subsequently, in step 2, we determine if a stopping criterion is met, and if so, use the predicted optimal mean response from the current surrogate model as the optimal design. The common stopping criteria are as follows: (i) the designer has exhausted all simulation resources or (ii) the "denoised" posterior variance $\tilde{\sigma}_n^2(\mathbf{x})$ at the predicted optimal design is below a specific threshold η. Note that steps 0–2 are common in most adaptive sampling schemes.

The contribution of the proposed sampling scheme lies in the decision between replication versus exploration to identify batches of samples that minimize the uncertainty in the optimal design. A schematic representation of the proposed sampling scheme to make this decision has been visualized in the bottom half of Fig. 2. More specifically, step 3 consists of a cycle of steps that are repeated B times where during each cycle one new sample is identified. Starting with step 3a, we identify a candidate sampling location $\mathbf{x}_{\text{exp}}^{(j)}$, $(j = 1, \ldots, B)$ for exploration by maximizing the modified EI function as presented in Eq. (11) [22]. It should be noted that the proposed sampling scheme is not limited to this specific choice of acquisition function as any of the existing acquisition functions for simulation models with heteroscedastic noise can be used (e.g., MQ [23], SKO [16], CKG [24], and EQI [36]). Note that step 3a only provides a candidate sample for exploration $\mathbf{x}_{\text{exp}}^{(j)}$ that will not necessarily be simulated. Rather, we first need to find the best candidate sample for replication (step 3b) before deciding which one of the two to simulate using the costly simulation model (step 3c).

The location of the candidate sample for replication $\bar{\mathbf{x}}_{i*}^{(j)}$, $(j = 1, \ldots, B)$ is found in step 3b by choosing the sample location among all previously observed sampling locations $\bar{\mathbf{x}}_i$, $(i = 1, \ldots, n)$ that provides the most "information" on the response $y(\cdot)$ at $\mathbf{x}_{\text{exp}}^{(j)}$. The purpose of replicating at previously observed sampling locations is to reduce the sampling scheme's computational cost from $\mathcal{O}(N^3)$ to $\mathcal{O}(n^3)$ (see Sec. 2.1). In previous study [40], we proposed a heuristic that approximates the contribution of each previously observed sampling location to the posterior variance of the predictive distribution $\hat{Y}(\mathbf{x}_{\text{exp}}^{(j)})|\mathbf{Y}$. However, in this work, we use the Sherman–Morrison–Woodbury formula to reformulate the posterior predictive variance into its individual constituents associated with the interpolation uncertainty and the intrinsic modeling uncertainty. The assumption behind this approach is that we expect to learn the most about the simulation response at $\mathbf{x}_{\text{exp}}^{(j)}$ by choosing to replicate at the sampling location that minimizes the posterior predictive variance at that spatial location.

With the candidate samples for replication $\bar{\mathbf{x}}_{i*}^{(j)}$ and exploration $\mathbf{x}_{\text{exp}}^{(j)}$ available, we use step 3c to decide which of the two should be added to the batch of samples in step 3d. Adopting a similar consideration as for step 3b, we want to select the sample from the set $\left\{\mathbf{x}_{\text{exp}}^{(j)}, \bar{\mathbf{x}}_{i*}^{(j)}\right\}$ that minimizes the posterior variance at $\mathbf{x}_{\text{exp}}^{(j)}$.

Going through step 3a to step 3d will only add one additional sample to the batch of samples $\mathbf{X}_{\text{new}}$, but prior to evaluating the costly simulation model (step 4), we would like a batch containing B samples. The batch size is determined by a designer based on the available computational resource. In the context of parallel computing, the size often depends on the number of available parallel processors. The algorithm will go through a preposterior analysis in step 3f so that it can return to step 3a so that it can identify the next sample. This process is repeated until a sampling batch $\mathbf{X}_{\text{new}} = \left\{\mathbf{x}_{\text{new}}^{(1)}, \ldots, \mathbf{x}_{\text{new}}^{(B)}\right\}$ containing B samples has been identified. Subsequently, the sampling scheme will advance to step 4, where the new batch of samples $\mathbf{X}_{\text{new}}$ will be evaluated by the costly simulation model before returning to step 1. This overarching process is repeated until a stopping criterion has been met, and we have successfully identified the optimal design.

Crucial to the performance of the proposed sampling scheme is the reformulation of the posterior variance (step 3a–3c) and the preposterior analysis (step 3f) for which further details are provided in Secs. 3.2 and 3.3, respectively.

3.2 Reformulation of the Posterior Predictive Variance.

The objective behind the decision to replicate or explore (step 3a–3c in Fig. 2) is the minimization of the posterior variance at the location that is likely to contain the global optimum (i.e., $\mathbf{x}_{\text{exp}}^{(j)}$). This raises the question: "What source of uncertainty has the largest contribution to the variance of the posterior predictive distribution $Y(\mathbf{x}_{\text{exp}}^{(j)})|\mathbf{Y}$?"

Taking a closer look at the formulation of the posterior variance given in Eq. (6), we observe that the contribution of the interpolation uncertainty is captured by the covariance matrix $\mathbf{K}_n$ and the intrinsic modeling uncertainties are captured by the diagonal elements of $\mathbf{A}^{-1}\boldsymbol{\Sigma}_n$. Note that the intrinsic modeling uncertainty comes from not knowing the population mean at the observed spatial locations. Subsequently, we are interested in quantifying the contribution of the interpolation uncertainty and intrinsic modeling uncertainty to the posterior prediction variance $\sigma_n^2(\mathbf{x})$.

Introducing an n-dimensional column vector of zeros $\mathbf{0}_{n,i}$ with its ith element equal to one and a matrix $\mathbf{A}_i^{-1} = \text{diag}\left(0, \ldots, 0, a_{i+1}^{-1}, a_{i+2}^{-1}, \ldots, a_n^{-1}\right)$ (i.e., the matrix $\mathbf{A}^{-1}$ with its first i diagonal elements set equal to 0). By setting $i = 1$, we can then rewrite the matrix inversion $(\mathbf{K}_n - \mathbf{A}^{-1}\boldsymbol{\Sigma}_n)^{-1}$ inside the posterior prediction variance as follows:

$$(\mathbf{K}_n - \mathbf{A}^{-1}\boldsymbol{\Sigma}_n)^{-1} = \left(\mathbf{K}_n + \mathbf{A}_1^{-1}\boldsymbol{\Sigma}_n + \mathbf{0}_{n,1}\left(\frac{\hat{r}(\bar{\mathbf{x}}_1)}{a_1}\right)\mathbf{0}_{n,1}^T\right)^{-1} \tag{12}$$

We then find that a rank-1 correction $\mathbf{0}_{n,1}\left(\frac{\hat{r}(\bar{\mathbf{x}}_1)}{a_1}\right)\mathbf{0}_{n,1}^T$ is applied to the matrix $\boldsymbol{\Lambda}_1 = \mathbf{K}_n + \mathbf{A}_1^{-1}\boldsymbol{\Sigma}_n$. Consequently, we can adopt the Sherman–Morrison–Woodbury formula [41] to rewrite the right-hand side of Eq. (12) as follows:

$$\begin{aligned}(\mathbf{K}_n - \mathbf{A}^{-1}\boldsymbol{\Sigma}_n)^{-1} &= \boldsymbol{\Lambda}_1^{-1} - \boldsymbol{\Lambda}_1^{-1}\mathbf{0}_{n,1}\left(\frac{\hat{r}(\bar{\mathbf{x}}_1)}{a_1} + \mathbf{0}_{n,1}^T\boldsymbol{\Lambda}_1^{-1}\mathbf{0}_{n,1}\right)\mathbf{0}_{n,1}^T\boldsymbol{\Lambda}_1^{-1} \\ &= \boldsymbol{\Lambda}_1^{-1} - \boldsymbol{\lambda}_1(a_1)\end{aligned} \tag{13}$$

where we have taken $\boldsymbol{\lambda}_1(a_1) = \boldsymbol{\Lambda}_1^{-1}\mathbf{0}_{n,1}\left(\frac{\hat{r}(\bar{\mathbf{x}}_1)}{a_1} + \mathbf{0}_{n,1}^T\boldsymbol{\Lambda}_1^{-1}\mathbf{0}_{n,1}\right)\mathbf{0}_{n,1}^T\boldsymbol{\Lambda}_1^{-1}$. Note that the first element on the right-hand side of Eq. (13) is once again a matrix inversion that is analogous to Eq. (12) and can be reformulated to have a rank-1 correction. Subsequently, we can again rewrite this matrix inversion using the Sherman–Morrison–Woodbury formula. In fact, by realizing that $\boldsymbol{\Lambda}_{i-1}^{-1} = \boldsymbol{\Lambda}_i^{-1} - \boldsymbol{\lambda}_i$, we can repeat this recursive procedure n times to find

$$(\mathbf{K}_n - \mathbf{A}^{-1}\boldsymbol{\Sigma}_n)^{-1} = \mathbf{K}_n^{-1} - \sum_{i=1}^{n}\boldsymbol{\lambda}_i(a_i) \tag{14}$$

where we have recovered $\mathbf{K}_n^{-1}$ because $\mathbf{A}_n^{-1}$ is an $n \times n$ dimensional vector of zeros so that $\boldsymbol{\Lambda}_n^{-1} = \left(\mathbf{K}_n + \mathbf{A}_n^{-1}\boldsymbol{\Sigma}_n\right)^{-1} = \mathbf{K}_n^{-1}$.

By substitution of Eq. (14) into Eq. (6), we obtain the following formulation of the posterior variance

$$\sigma_n^2(\boldsymbol{x}) = k(\mathbf{x}, \mathbf{x}) + r(\mathbf{x}) - \mathbf{k}_n^T(\mathbf{x})\mathbf{K}_n^{-1}\mathbf{k}_n(\mathbf{x}) + \sum_{i=1}^{n}\mathbf{k}_n^T(\mathbf{x})\boldsymbol{\lambda}_i(a_i)\mathbf{k}_n(\mathbf{x}) \tag{15}$$

Observation of the right-hand side of Eq. (15) reveals that we can capture the contribution of the interpolation uncertainty to the posterior variance as $k(\mathbf{x}, \mathbf{x}) - \mathbf{k}_n^T(\mathbf{x})\mathbf{K}_n^{-1}\mathbf{k}_n(\mathbf{x})$. Note that this is the

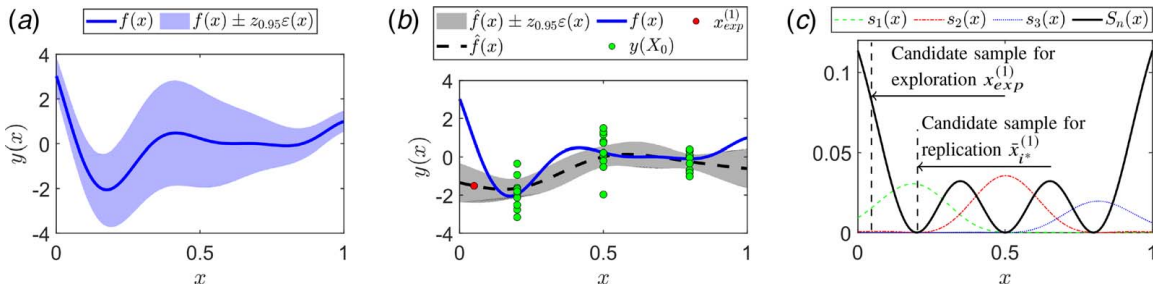

Fig. 3 Visualization of a one-dimensional test function with heteroscedastic noise, its approximation through SK, and the analysis of the posterior variance to decide whether to replicate or explore: (*a*) example of a one-dimensional test function with heteroscedastic noise, (*b*) approximation of a one-dimensional test function with three unique sampling locations with ten replicates each, and (c) analysis of the posterior prediction variance and its constituents (Color version online.)

exact same expression for the posterior variance of a GP trained on the data set obtained from a deterministic simulation model. In addition, the contribution of the intrinsic modeling uncertainty that enters the SK surrogate by not knowing the population mean at the ith unique sampling location is captured by $\mathbf{k}_n^T(\mathbf{x})\boldsymbol{\lambda}_i(a_i)\mathbf{k}_n(\mathbf{x})$. Finally, the intrinsic modeling uncertainty $r(\mathbf{x})$ in Eq. (15) has a fixed and typically unknown contribution to the posterior variance, but as will be shown in the following paragraph, we do not need this information to identify what sample(s) to simulate next.

Concerning the identification of the best candidate sample for replication (step 3b), from the formulation of Eq. (15), we note that the contribution of not knowing the population mean at the ith sampling location is given by $\mathbf{k}_n^T(\mathbf{x})\boldsymbol{\lambda}_i(a_i)\mathbf{k}_n(\mathbf{x})$, where $\boldsymbol{\lambda}_i(a_i)$ depends on the sampled data through the number of replicates a_i at x_i. When deciding to replicate a design, its number of replicates will increase by one, in which case we can approximate the reduction in the posterior variance $s_i(\cdot)$ at the sampling location for exploration $\mathbf{x}_{\text{exp}}^{(j)}$ by replicating at the ith sampling location as follows:

$$s_i(\mathbf{x}_{\text{exp}}^{(j)}) = \mathbf{k}_n^T(\mathbf{x}_{\text{exp}}^{(j)})(\boldsymbol{\lambda}_i(a_i) - \boldsymbol{\lambda}_i(a_i+1))\mathbf{k}_n(\mathbf{x}_{\text{exp}}^{(j)}) \tag{16}$$

Observe that in Eq. (16), we approximate the difference in the posterior variance by subtracting the hypothetical future value from its current value. We can then (with slight abuse of notation) find the optimal sample for replication by maximizing Eq. (16) with respect to all unique observed sampling locations $\bar{\mathbf{x}}_i$ $(i = 1, \ldots, n)$ to find the candidate sample for replication $\mathbf{x}_{i^*}^{(j)}$.

In step 3c, a decision has to be made whether to replicate $\mathbf{x}_{i^*}^{(j)}$ or explore $\mathbf{x}_{\text{exp}}^{(j)}$. This decision is made based on which alternative minimizes the posterior variance. The candidate sample for replication is identified by its predicted reduction in the posterior variance $s_{i^*}(\mathbf{x}_{\text{exp}}^{(j)})$ at $\mathbf{x}_{\text{exp}}^{(j)}$, whereas the reduction in the posterior variance when deciding the explore is given as as follows:

$$S_n(\mathbf{x}_{\text{exp}}^{(j)}) = k(\mathbf{x}_{\text{exp}}^{(j)}, \mathbf{x}_{\text{exp}}^{(j)}) - \mathbf{k}_n^T(\mathbf{x}_{\text{exp}}^{(j)})\mathbf{K}_n^{-1}\mathbf{k}_n(\mathbf{x}_{\text{exp}}^{(j)}) \tag{17}$$

Subsequently, we then decide whether to explore a new sampling location or replicate at an existing one according to the following equation:

$$\mathbf{x}_{\text{new}}^{(j)} = \begin{cases} \mathbf{x}_{\text{exp}}^{(j)} & \text{if } S_n(\mathbf{x}_{\text{exp}}^{(j)}) > s_{i^*}(\mathbf{x}_{\text{exp}}^{(j)}) \\ \bar{\mathbf{x}}_{i^*}^{(j)} & \text{otherwise} \end{cases} \tag{18}$$

The intuition behind Eq. (18) is that $S_n(\mathbf{x}_{\text{exp}}^{(j)})$ considers the interpolation uncertainty that will be reduced to zero once a sample is observed at the new spatial location $\mathbf{x}_{\text{exp}}^{(j)}$, while replicating at $\bar{\mathbf{x}}_{i^*}^{(j)}$ sample can only reduce the intrinsic modeling uncertainty that enters the surrogate model through the sample mean at each unique sampling location.

For visualization purposes, consider that we want to optimize a one-dimensional simulation model with heteroscedastic noise as shown in Fig. 3(*a*). Next, assume that we start with an initial batch of samples containing three unique sampling locations {0.2, 0.5, 0.8}, each with ten replicates as shown by the green dots in Fig. 3(*b*). Given that a candidate sample for exploration has been found at $x_{\text{exp}}^{(1)} = 0.05$ as presented by the left most red dot in Fig. 3(*b*), it would make sense that the corresponding candidate for replication is the nearest sample $\bar{x}_{i^*}^{(1)} = 0.2$. The sampling locations for exploration and replication are also presented in Fig. 3(*c*) by the two vertical dashed lines, the approximation of the interpolation uncertainty has been plotted by the solid black line $S_n(x)$, and the approximation of the reduction in the posterior variance for replicating at the three sampling locations has been plotted by the colored lines (i.e., the long dashed green line $s_1(x)$ captures the intrinsic uncertainty that enters the surrogate model through sampling location $x = 0.2$, the long-short dashed red line $s_2(x)$ is for $x = 0.5$, and the short dashed blue line $s_3(x)$ is for $x = 0.8$). From this figure, we observe that the uncertainty at the candidate sample for exploration is mostly driven by the interpolation uncertainty (i.e., $S_n(0.05) > s_i(0.05)$, for $i = 1, 2, 3$), and thus, the proposed scheme will decide to explore the new sampling location. However, if for example the candidate for exploration was found at $x = 0.45$, we observe that $s_2(0.45) > S_n(0.45) > s_1(0.45) > s_3(0.45)$, in which case the proposed sampling scheme would choose to replicate at $\bar{x}_2 = 0.5$.

3.3 Preposterior Analysis Through a Zeroth-Order Interpolation of the Intrinsic Noise. In this section, we present the proposed preposterior analysis used for the allocation of sampling batches (step 3f in Fig. 2). The preposterior analysis allows the sampling scheme to return to step 3a without having to run the costly simulation model and without getting a new maximum likelihood approximate of the hyperparameters (Eq. (4)). For the deterministic case, the initial GP model $Y(\mathbf{x})|\mathbf{Y}$ is used to identify a new sampling location $\mathbf{x}_{\text{new}}$ [42,43]. Assuming that $\mu(\mathbf{x}_{\text{new}})$ is an accurate approximation of the response at that location, we can temporarily add $\{\mathbf{x}_{\text{new}}, \mu(\mathbf{x}_{\text{new}})\}$ to the training data set. The updated training data set is then used to condition the posterior predictive distribution of Eq. (6), allowing the designer to identify the next sampling location.

Using the preposterior analysis for an SK surrogate model not only deviates from the deterministic GP model in that the posterior prediction not only depends on $\{\mathbf{x}_{\text{new}}^{(j)}, \mu_n(\mathbf{x}_{\text{new}}^{(j)})\}$ but also requires the intrinsic noise of the simulation model $r(\mathbf{x}_{\text{new}}^{(j)})$ and the number of replicates at $\mathbf{x}_{\text{new}}^{(j)}$. This is relatively straightforward when the algorithm decides to replicate the ith$(i = 1, \ldots, n)$ sampling location as we can incrementally increase the number of replicates a_i by one. For example, assume that the original data set on which we trained our model and conditioned the posterior predictive distribution is given as follows:

$$D^{(j)} = \left\{ \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \vdots \\ \bar{\mathbf{x}}_n \end{bmatrix}, \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{bmatrix}, \begin{bmatrix} \hat{r}(\mathbf{y}_1) \\ \vdots \\ \hat{r}(\mathbf{y}_n) \end{bmatrix}, \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \right\} \tag{19}$$

In the case that the algorithm decides to replicate at the ith sampling location, then the data on which we condition the posterior response can be approximated as follows:

$$D^{(j+1)} = \left\{ \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \vdots \\ \bar{\mathbf{x}}_{i-1} \\ \bar{\mathbf{x}}_i \\ \bar{\mathbf{x}}_{i+1} \\ \vdots \\ \bar{\mathbf{x}}_n \end{bmatrix}, \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_{i-1} \\ \bar{y}_i \\ \bar{y}_{i+1} \\ \vdots \\ \bar{y}_n \end{bmatrix}, \begin{bmatrix} \hat{r}(\mathbf{y}_1) \\ \vdots \\ \hat{r}(\mathbf{y}_{i-1}) \\ \hat{r}(\mathbf{y}_i) \\ \hat{r}(\mathbf{y}_{i+1}) \\ \vdots \\ \hat{r}(\mathbf{y}_n) \end{bmatrix}, \begin{bmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_i + 1 \\ a_{i+1} \\ \vdots \\ a_n \end{bmatrix} \right\} \tag{20}$$

The resulting training data set of Eq. (20) is equivalent to the scenario where the response of simulating the new replicate is equivalent to the sample mean response at the ith sampling location $\bar{y}_i$.

When exploring a new sampling location the designer needs an approximation of the intrinsic modeling uncertainty $r(\mathbf{x}_{\text{exp}}^{(j)})$. One approximation is to train an individual GP model to the sampling variances $\hat{r}(\mathbf{x}_i)$ that have more than $a_i > \beta$ samples. However, as we are expecting that in the final experimental design sufficient sampling locations and replications will be allocated around the global optimum of the design space, we propose to use a zeroth-order interpolation. This implies that the ith sampling locations where $a_i < \beta$ will be assigned the sampling variance $\hat{r}_{ngb}(\bar{\mathbf{x}}_{\text{exp}}^{(j)})$ of its highest correlated neighbor that does have sufficient replicates. The resulting data set to which we condition the posterior predictive response is then given as follows:

$$D^{(j+1)} = \left\{ \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \vdots \\ \bar{\mathbf{x}}_n \\ \bar{\mathbf{x}}_{\text{exp}}^{(j)} \end{bmatrix}, \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \\ \mu_n(\bar{\mathbf{x}}_{\text{exp}}^{(j)}) \end{bmatrix}, \begin{bmatrix} \hat{r}(\mathbf{y}_1) \\ \vdots \\ \hat{r}(\mathbf{y}_n) \\ \hat{r}_{ngb}(\bar{\mathbf{x}}_{\text{exp}}^{(j)}) \end{bmatrix}, \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ 1 \end{bmatrix} \right\} \tag{21}$$

The advantage of using a zeroth-order interpolation over GP interpolation for the approximation of the intrinsic noise is twofold: (i) the algorithm is less likely to run into numerical issues when multiple samples in proximity of each other have different sampling variances and (ii) the zeroth-order interpolation of the intrinsic noise is less likely to waste resources, whereas GP interpolation of the intrinsic noise can result in poor sampling decisions resulting from greatly underestimating or overestimating the intrinsic uncertainty at unobserved spatial locations (this will be further explained in Sec. 4.1.2).

4 Results and Observations

In this section, we present the results of the proposed sampling scheme when applied to the optimization of three test functions with different input dimensionalities and degrees of intrinsic model uncertainty. In addition, the proposed scheme is applied to a real engineering problem that involves the optimization of an OPVC as a function of its processing settings.

4.1 Two-Dimensional Test Functions. To highlight the strength of the proposed sampling scheme, we apply it to two two-dimensional test functions that exhibit various types of intrinsic heteroscedastic noise. Moreover, the results are compared to the performance of the MQ and TSSO schemes as introduced in Sec. 2.2 and to our previously introduced scheme that adopts a heuristic to approximate $s_i(\cdot)$ [25].

4.1.1 Problem Formulation and Implementation. We consider the six-hump Camel function and the rescaled Branin function as presented in Table 1 [44]. The six-hump Camel function has a relatively rough response surface with six local minimum spread throughout the design space. Subsequently, it will be interesting to observe at what rate the proposed sampling scheme can identify the global minimum from the local minima. The rescaled Branin function is a relatively smooth function with three local minima. It will be interesting to observe how the proposed sampling scheme will utilize its computational resources to distinguish between each local minima.

In this work, we consider four different noise scenarios that are analogous to the study presented in Ref. [45]. More specifically, we assume that the variance of the noise varies linearly with respect to the response of the function as follows:

$$\varepsilon(x_1, x_2) \sim \mathcal{N}\left(0, a\left(\mathbb{E}\left(f(x_1, x_2)\right) + b\right)\right) \tag{22}$$

To determine the parameters a and b, we consider that the magnitude of the intrinsic noise can take on one of two levels: (i) light noise where the standard deviation of the noise varies between 15% and 60% of the range of the response (i.e., $r(\mathbf{x}) \in \left(0.15\sqrt{R_f}, 0.6\sqrt{R_f}\right)$) and (ii) heavy noise where the standard deviation of the noise varies between 150% and 600% of the range of the response (i.e., $r(\mathbf{x}) \in \left(1.5\sqrt{R_f}, 6\sqrt{R_f}\right)$). In addition, we consider two levels for the functional form of the intrinsic simulation noise: (i) best-case noise where the intrinsic uncertainty is smallest at the global optimum of the function (i.e., min ($r(\mathbf{x})$) at argmin$f(\mathbf{x})$) and (ii) worst-case noise where the intrinsic uncertainty is largest at the global optimum (i.e., max ($r(\mathbf{x})$) at argmin$f(\mathbf{x})$). This means that we have two test functions, each with four test cases, for which the total number of micro-trials N_t, batch size B, and the parameters a, b are presented in Table 2. Note that the two noise parameters are associated with the formulation of the test functions and do not need to be defined by a designer.

We optimize the eight cases summarized in Table 2 using the proposed adaptive sampling scheme, our previously published heuristics-based sampling scheme [25], the MQ scheme (Eq. (10) in Sec. 2.2) and the TSSO scheme (Eq. (11) in Sec. 2.2). We then create an initial experimental design $\mathbf{X}_0$ through a Latin hypercube containing nine unique sampling locations with 50 replicates each. We have chosen a relatively large number of initial replicates so that our results can be compared with those presented in Ref. [45]. This process is repeated 20 times, each time starting from a different initial design. Each repetition will be referred to as a macro-trial. Finally, in terms of parameter selection, for the MQ scheme,

Table 1 Two-dimensional test functions for comparing the proposed sampling scheme to existing sampling schemes

Six-hump Camel	Rescaled Branin
$d = 2$	$d = 2$
$f(x_1, x_2) = 4x_1^2 - 2.1x_1^4 + \frac{x_1^6}{3} + x_1x_2 - 4x_2^2 + 4x_2^4 + \varepsilon(x_1, x_2)$	$f(x_1, x_2) = \frac{1}{51.95}\left(15x_2 - \frac{5.1(15x_1 - 5)^2}{4\pi^2} + \frac{5(15x_1 - 5)}{\pi} - 6\right)^2 + \frac{1}{51.95}\left[\left(10 - \frac{10}{8\pi}\right)\cos(15x_1 - 5) - 44.81\right] + \varepsilon(x_1, x_2)$
$-2 \le x_1 \le 2, -1 \le x_2 \le 1$	$0 \le x_1 \le 1, 0 \le x_2 \le 1$
Range: $R_f = 7.3$	Range: $R_f = 6$

Table 2 Summary of the intrinsic noise of eight test problems, number of macro-trials, and selected batch size for each the test functions presented in Table 1

		Noise magnitude	
		Light noise $r(\mathbf{x}) \in (0.15\sqrt{R_f}, 0.6\sqrt{R_f})$	Heavy noise $r(\mathbf{x}) \in (1.5\sqrt{R_f}, 6\sqrt{R_f})$
Best-case noise	Six-hump Camel	$N_t = 20 \quad B = 50$	$N_t = 40 \quad B = 50$
min $(r(\mathbf{x}))$ at argmin $f(\mathbf{x})$		$a = 0.45 \quad b = 3.46$	$a = 4.5 \quad b = 3.46$
	Rescaled Branin	$N_t = 20 \quad B = 50$ $a = 0.45 \quad b = 3.05$	$N_t = 40 \quad B = 50$ $a = 4.5 \quad b = 3.05$
Worst-case noise max $(r(\mathbf{x}))$ at argmin $f(\mathbf{x})$	Six-hump Camel	$N_t = 20 \quad B = 50$ $a = -0.45 \quad b = -8.704$	$N_t = 40 \quad B = 50$ $a = -4.5 \quad b = -8.704$
	Rescaled Branin	$N_t = 20 \quad B = 50$ $a = -0.45 \quad b = -6.95$	$N_t = 40 \quad B = 50$ $a = -4.5 \quad b = -6.95$

we have chosen $\theta = 0.25$ and set the number of replicates per sampling location equal to the batch size, while in the TSSO scheme, we set the number of replicates per new sampling location $r_{rep} = 10$.

The performance of each competing alternative is measured through a set of three criteria that are equivalent to the criteria presented in Ref. [45]:

(1) The GAP is the difference between the true global optimum $f(\mathbf{x}^*)$ and the response at the predicted global optimum $f(\hat{\mathbf{x}}^*_{m,s})$ $(m = 1, \ldots, 20, s = 1, \ldots, N_t)$, where $\hat{\mathbf{x}}^*_{m,s}$ is the predicted globally optimal design for the sth micro-trial of the mth macro-trial.
(2) The $NV_{\mathcal{X}}$ is the fraction of macro-trials where the sampling scheme explores a new sampling location that is within $(1 - \mathcal{X})100\%$ of the range $f(\mathbf{x})$ to the globally optimal response.
(3) The $NR_{\mathcal{X}}$ is the fraction of macro-trials where the returned solution is within $(1 - \mathcal{X})100\%$ of the range $f(\mathbf{x})$ to the globally optimal response. Note that this implies that $NR_{\mathcal{X}} \leq NV_{\mathcal{X}}$.

In this work, we have chosen $\mathcal{X} = 0.975$, and this means that $NV_{\mathcal{X}}$ and $NR_{\mathcal{X}}$ only count designs that differ from the global optimal design $f(\mathbf{x}^*)$ no more than 2.5% of the range of $f(\mathbf{x})$.

4.1.2 Empirical Validation of the Proposed Sampling Scheme. The results for the six-hump Camel function and the rescaled Branin function have been plotted in Figs. 4 and 5, respectively. First, when looking at the GAP for the six-hump Camel function, we find that the proposed sampling scheme as depicted by the continuous blue lines in Figs. 4(*a*)–4(*d*) is highly competitive to the alternative sampling schemes in terms of the GAP criterion, particularly for the low-noise problems. What is more, we find that our previous heuristic scheme as depicted by the dashed red lines performs relatively well in the low best-case noise scenario, but quickly degrades in performance when the noise becomes larger. Under closer observation, we found that in the high-noise scenario the heuristics-based scheme underestimates the intrinsic noise when many samples are in proximity of one another. Therefore, the heuristics-based scheme places more emphasis on exploring new sampling locations and is unlikely to successfully identify the global optimum as it has no accurate representation of the intrinsic uncertainty. In terms of the MQ (dotted gray line) and TSSO scheme (dashed dotted green line), we find that the TSSO algorithm is more likely to explore regions in the design space compared to MQ scheme. Consequently, the MQ scheme converges quicker to the global

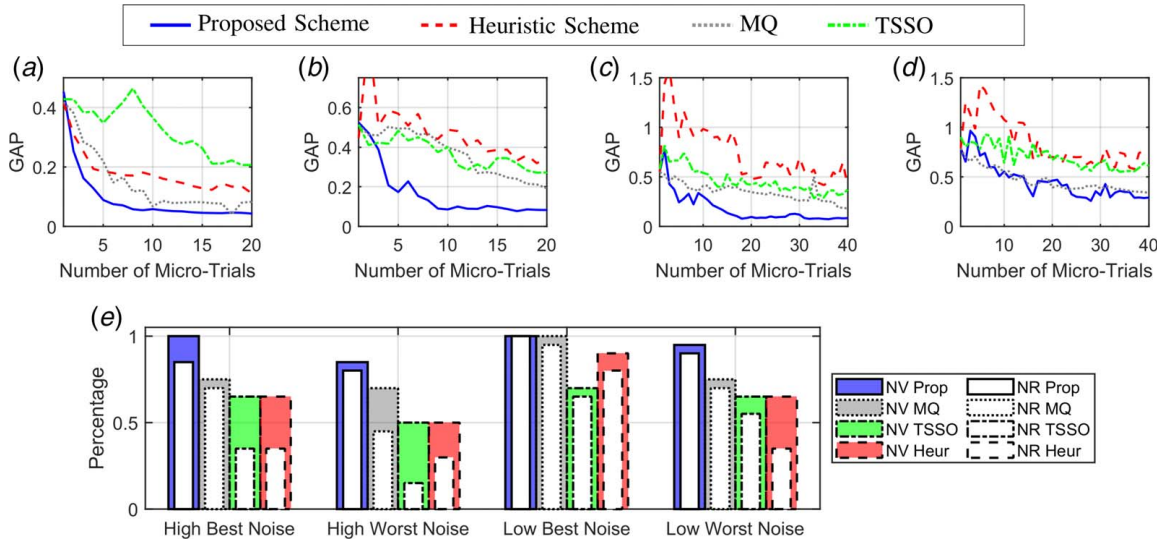

Fig. 4 Comparison of the proposed sampling scheme to the TSSO, MQ, and previously published heuristics based schemes on the six-hump Camel function: (*a*) low best-case noise average GAP for the six-hump Camel function, (*b*) low worst-case noise average GAP for the six-hump Camel function, (*c*) high best-case noise average GAP for the six-hump Camel function, (*d*) high worst-case noise average GAP for the six-hump Camel function, and (e) fractions of times a micro-trial explored the global minimum (Nv) and times it converged to the global optimum (Nr) (Color version online.)

optimum when starting from a good initial design but was occasionally unable to identify the global optimum when starting from a poor initial design.

From the results depicted in Figs. 4(*a*)–4(*d*), we observe that the proposed sampling scheme has the best performance in terms of the GAP criterion in nearly all cases, except from the high worst-case noise scenario where it is on par with the MQ scheme. However, when looking at the $NV_{\mathcal{X}}$ and $NR_{\mathcal{X}}$ of the sampling schemes on each of the four optimization problems as shown in Fig. 4(*e*), we find that the proposed sampling scheme outperforms all three competing schemes in terms of the fraction of times it was able to return the global optimum and the number of times it explored the region containing the global optimum. Moreover, we observe that the proposed scheme is comparatively better for the noisier functions. In fact, in the high worst noise and high best noise scenario, the $NR_{\mathcal{X}}$ of the proposed sampling scheme trumps the $NV_{\mathcal{X}}$ of all other schemes.

When the four sampling schemes under consideration are used for the optimization of the rescaled Branin function as presented in Fig. 5, we find that all schemes perform relatively well. This is because the rescaled Branin function is relatively smooth, and the region of the design space that is within 2.5% of the range of the global optimum is relatively large. For this reason, the results in Figs. 5(*a*)–5(*d*) have been plotted on a logarithmic scale. We observe similar results here as we did for the six-hump Camel function where the proposed scheme shows very strong performance for both the low-noise scenarios and is highly competitive with the MQ scheme for the worst-case high-noise scenario. However, in terms of the $NR_{\mathcal{X}}$ and $NV_{\mathcal{X}}$, the performance of the proposed scheme is more distinct from the competing schemes when the intrinsic noise increases and maximizes at the minimum of the objective function. The results confirm that the active decision when to replicate and when to explore based on which alternative minimizes the posterior predictive variance is a potent and valuable consideration when optimizing stochastic functions. This observation is bolstered by realizing that any acquisition function can be used to identify the candidate samples for exploration.

4.1.3 Empirical Validation of the Zeroth-Order Interpolation Assumption. One key construct of the proposed sampling scheme is the zeroth-order interpolation to approximate the posterior variance used in the preposterior analysis. There are two alternatives to this approach: (i) the designer adopts an additional GP model to approximate the posterior variance or (ii) the designer assumes to know the actual function of the intrinsic simulation noise. In this section, we will perform the same simulations as presented in Tables 1 and 2 using the proposed sampling scheme with the zeroth-order interpolation, GP interpolation, and the actual function for the intrinsic noise. For the GP interpolation of the intrinsic noise, we trained a surrogate to the natural log of the sample variance for all unique sampling locations that have ten or more replicates $\beta > 10$. Taking the natural logarithm ensures that we never predict a negative variance for the intrinsic noise as this could potentially lead to an ill-conditioned covariance matrix.

The GAP for all eight test scenarios has been plotted in Fig. 6, where the continuous blue lines correspond to the zeroth-order interpolation, the dashed black lines corresponds to the GP model interpolation, and the dotted black lines correspond to actually knowing the intrinsic noise. We observe that knowing the intrinsic noise and using the zeroth-order interpolation have very comparable performance in all eight scenarios. Moreover, the GP model interpolation performs similarly well in the low-noise cases, while in the high-noise cases, its performance quickly degrades. It is found that using a GP model sometimes greatly underestimates the intrinsic noise, in which case many of the new samples will be allocated for exploration. The use of more samples for exploration has two negative consequences: (i) the computational cost increases significantly as the size of the covariance matrix balloons and (ii) because insufficient replicates are allocated to new sampling location we can no longer have an accurate prediction of the response surface and its noise. This phenomenon may happen during every micro-trial and is more likely to occur in later micro-trials. Consequently, we observe that initially the average GAP of all 20 micro-trials decreases, but as we go through more trials, the gap increases.

The presented observations provide three valuable insights. First, the use of GP interpolation for the intrinsic noise can have a detrimental effect to the performance of the algorithm. Second, the use of the zeroth-order interpolation is robust regardless of the functional form of the simulation model's intrinsic noise. Third, the performance of the zeroth-order interpolation is comparable to the case is which we know the intrinsic noise. Consequently, we postulate that we can adopt the zeroth-order interpolation in the preposterior analysis with negligible negative consequences to the performance of the proposed sampling scheme.

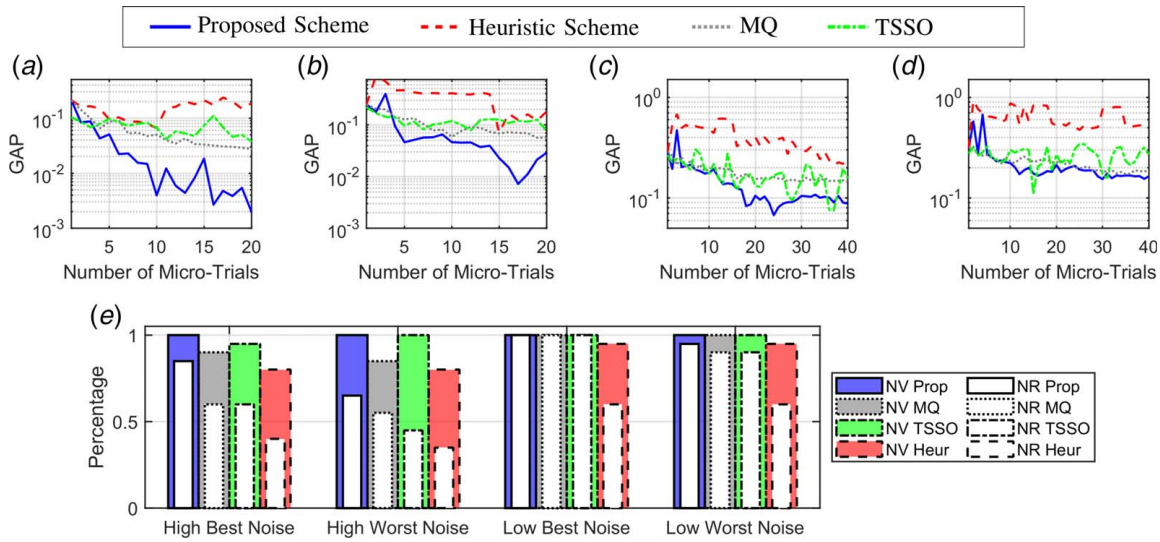

Fig. 5 Comparison of the proposed sampling scheme to the TSSO, MQ, and previously published heuristics based schemes to the rescaled Branin function: (*a*) low best-case noise average GAP for the rescaled Branin function, (*b*) low worst-case noise average GAP for the rescaled Branin function, (*c*) high best-case noise average GAP for the rescaled Branin function, (*d*) high worst-case noise average GAP for the rescaled Branin function, and (*e*) fractions of times a micro-trial explored the global minimum (Nv) and times it converged to the global optimum (Nx)

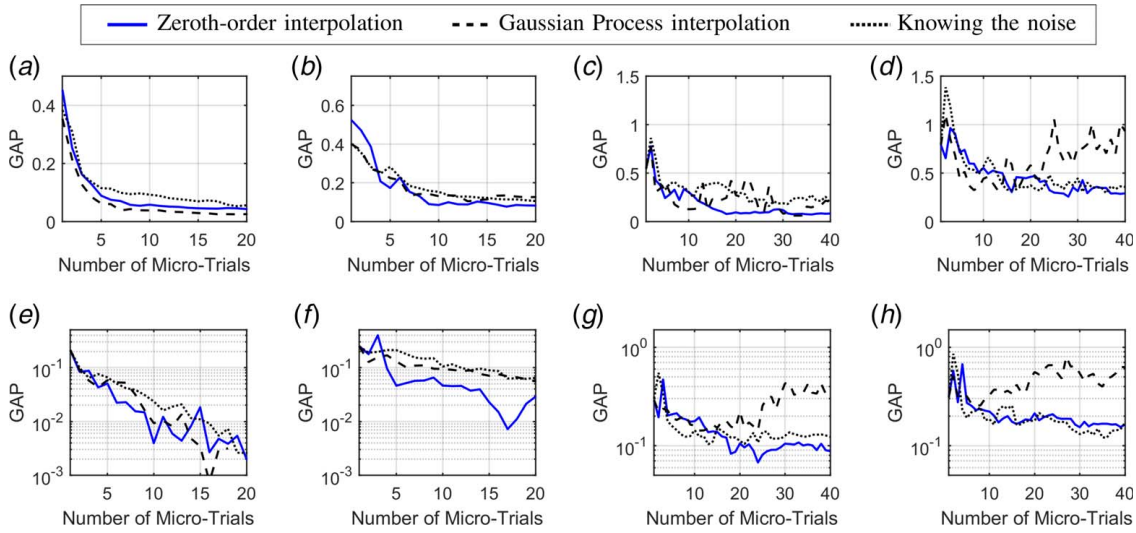

Fig. 6 Comparison of the zeroth-order interpolation assumption for the approximation of the heteroscedastic noise with the interpolation of the intrinsic noise through an additional GP, and knowing the noise: (*a*) low best-case noise GAP for the six-hump Camel function, (*b*) low worst-case noise GAP for the six-hump Camel function, (*c*) high best-case noise GAP for the six-hump Camel function, (*d*) high worst-case noise GAP for the six-hump Camel function, (*e*) low best-case noise GAP for the rescaled Branin function, (*f*) low worst-case noise GAP for the rescaled Branin function, (*g*) high best-case noise GAP for the rescaled Branin function, and (*h*) high worst-case noise GAP for the rescaled Branin function (Color version online.)

4.2 Six-Dimensional Test Function. In this section, we demonstrate the proposed sampling scheme's ability to be scaled to higher dimensional problems. We consider the six-dimensional Hartmann function presented in Ref. [46] and given as follows:

$$y(\mathbf{x}) = 5 - \sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} Q_{ij}(x_j - P_{ij})^2\right) + \varepsilon(\mathbf{x}) \tag{23}$$

$$\varepsilon(\mathbf{x}) \sim \mathcal{N}\big(0, 0.1y(\mathbf{x})\big), \quad \mathbf{x} \in [0, 1]^6 \tag{24}$$

where $\boldsymbol{\alpha} = \{1, 1.2, 3, 3.2\}$ and

$$\mathbf{Q} = \begin{Bmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{Bmatrix}$$

$$\mathbf{P} = \begin{Bmatrix} 1312 & 1696 & 5569 & 124 & 8283 & 5886 \\ 2329 & 4135 & 8307 & 3736 & 1004 & 9991 \\ 2348 & 1451 & 3522 & 2883 & 3047 & 6650 \\ 4047 & 8828 & 8732 & 5743 & 1091 & 381 \end{Bmatrix}$$

We initiate the optimization with an experimental design containing 20 unique sampling locations each replicated 20 times and obtained through a Latin hypercube design. We then go through 20 micro-trials, during each of which a batch containing 50 samples is identified (i.e., $B = 50$). Moreover, this process is repeated for ten macro-trials, and the GAP for the proposed sampling scheme is presented by the continuous blue line in Fig. 7. This figure shows that the proposed sampling scheme quickly narrows in the region containing the global optimum of the simulation model. Once the optimal region has been identified, samples are added to reduce the variance of the posterior distribution at this location. However, because it is a relatively high-dimensional problem, many replications need to be added to a multitude of sampling locations to accurately identify the global optimum of the function. Despite this apparent limitation, with a reasonable number of samples, the proposed adaptive sampling scheme can reliably and prudently find a design that is close to the global optimum.

In addition to the proposed sampling scheme, we also implemented the MQ and TSSO schemes using the same settings as in the previous section (i.e., $n_{\text{rep}} = 10$ and $\theta = 0.25$) to optimize the six-dimensional Hartmann function. The average GAP of each micro-trial for the MQ scheme is presented by the dashed gray line, and the result of the TSSO scheme is presented by the dashed green line. We observe that the proposed sampling scheme provides a competent alternative to the existing schemes as it rapidly identifies the region containing the global optimum, while the other schemes take longer as they only explore one new sampling location during each micro-trial. In addition, it could be argued that we selected poor values for the parameters B, n_{rep}, or θ. However, this is an additional advantage that the proposed method holds of the existing methods as it does not require the designer to have knowledge of the underlying function and its intrinsic noise.

4.3 Optimization of an OPVC. To highlight the applicability of the proposed adaptive sampling scheme, we applied it to the optimization of an OPVC considering its performance is a function of two processing settings [47,48]. The preferred choice of architecture for an OPVC is bulk heterojunction [49] and the "best seller" donor/acceptor combination is phenyl-C61-butyric-acid-methyl ester (PCBM) interspersed with poly(3-hexylthiophene-2,5-diyl) (P3HT). The efficiency of a solar cell is measured by the incident photon-to-converted-electron (IPCE) ratio and broadly depends on four physical phenomena: (i) light absorption, (ii) exciton creation, (iii) charge separation, and (iv) charge diffusion and collection. These four physical processes directly influence the OPVC's microstructure and in turn are influenced by the processing conditions.

The thin film OPVCs are manufactured through a process known as spin coating, which involves depositing a small amount of coating material on a rotating substrate. The centrifugal forces spread the coating evenly over the substrate, and rotation is continued until the desired film thickness has been achieved. Not only are these experiments expensive but also the thickness of the samples is hard to control. As an alternative, coarse-grained molecular dynamic (CGMD) simulations were carried out to replace the physical experiments. Although there are other important processing conditions, only the two predominant ones were considered: ratio of PCBM:P3HT and annealing temperature. In previous work

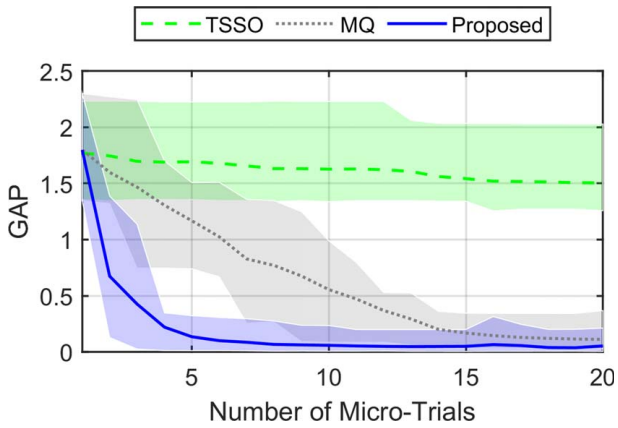

Fig. 7 Convergence history of the proposed adaptive sampling scheme for the optimization of the six-dimensional Hartmann function, adding 50 samples per micro-trial. The lines represent the mean of the GAP for each micro-trial, and the boundaries of their shaded region represent their best- and worst-case performance. (Color version online.)

[50], we presented a processing-structure-performance simulation incorporating the four aforementioned physical phenomena to predict the IPCE value for any digital OPVC microstructure. The same simulation, with slight modifications, is employed in this work to evaluate the IPCE value for all the data points. A single simulation of the processing-structure-performance linkage requires about 36–38 h, it is therefore desirable that computational resources are spend prudently.

We observe that the IPCE predicted from the CGMD model exhibits heteroscedastic noise (see purple bars in Fig. 1) that are a result from the random initial conditions and the limitation that we can only evaluate a finite number of molecules at a finite time scale. We start the optimization scheme with an optimal Latin hypercube design containing nine unique sampling locations, 25 replications at each sampling location, and set the batch size equal to $B = 20$. Next, we run the 225 CGMD simulations of the initial set of samples and train an SK model on the obtained data. The mean prediction of the surrogate model has been presented in Fig. 8(*a*), where the red circular dots indicate the observed sampling locations, the gray diamonds are the recommended next sampling locations, and the numbers next to each sampling location indicates its number of replicates. Over the first micro-trial, we find that the majority of the recommended samples are used to explore the predicted promising regions at $\mathbf{x} = \{0.25, 165\}$ and $\mathbf{x} = \{0.45, 55\}$.

After evaluating the first batch of 20 samples as recommended by the proposed sampling scheme, we observe that the updated response surface as presented in Fig. 8(*b*) predicts the global optimum to be around $\mathbf{x} = \{0.3, 160\}$. Consequently, half of the samples of the new batch are placed in this region, while the other half is used to explore the remainder of the design space. From allocating the samples for exploration, a more promising region is found around $\mathbf{x} = \{0.25, 100\}$ as shown in Fig. 8(*c*). However, after exploring this new region with the next batch of samples (shown in Fig. 8(*d*)), we find that this was an overly optimistic prediction and now observe that the global optimum is predicted to be $\mathbf{x}^* = \{0.2704, 113.7\}$. Consequently, an additional batch of samples is warranted because the spatial location of the global optimum has shifted significantly and the posterior predictive distribution has a relative large standard deviation (0.025). The new sampling batch adds function evaluations in the regions with a good mean response and we now predict the globally optimal mean response to be $\hat{y}(\mathbf{x}^*) = \mathcal{N}(0.3299, 0.0067)$ and is located at $\mathbf{x}^* = \{0.2769, 110.89\}$ as shown by the yellow star in Fig. 8(*e*). At this micro-trial, the spatial location of the global optimum did not change much with respect to the previous micro-trial, and the uncertainty in the posterior prediction is satisfactorily low. Consequently, we stop the sampling scheme and have confidence that we identified a design that is sufficiently close to the global optimum.

The optimal volume fraction of PCBM according to the literature is around 0.4 [51], but the samples in physical experiments are thicker (generally more than a 100 nm) compared to our simulations (20 nm). Since the simulation thickness is comparable or smaller than the mean free path of the excitons and charges, it is expected that the most prominent physical phenomenon, from the four mentioned earlier, is the exciton creation, which is directly associated with the material's absorption coefficient. PCBM has a lower absorption coefficient than P3HT in the solar spectrum range [52], and thus, more P3HT in the material results in higher IPCE. This leads to a general trend of high performance for designs with lower PCBM fraction. Nevertheless, even with these thinner microstructures, a prominent feature is observed at $\mathbf{x} \approx 0.45$ PCBM fraction and $T \approx 100\,^{\circ}\mathrm{C}$ (Fig. 8(*e*)). The capture of this feature is intriguing as it substantiates that the nanostructures have significant influences on the other physical phenomena such as the charge separation, the charge diffusion, and collection.

5 Discussion on the Proposed Scheme and Its Results

Following the validation square [53], we address the theoretical validity of the proposed sampling scheme by discussing its implementation details, its sampling characteristics, and limitations.

(1) The proposed sampling scheme has additional computational cost compared to the existing MQ and TSSO scheme. More specifically, for the allocation of each individual sample, the proposed sampling scheme must identify the best candidate sample for exploration. This step requires the inversion of the $n \times n$ covariance matrix needed to approximate the EI in Eq. (11). Moreover, the identification of the best candidate for replication requires the inversion of an $n \times n$ matrix a total of n times, as given by the recursive formulation of Eq. (13) used to obtain $\lambda_i(\cdot)$, $(i = 1, \ldots, n)$. Ordering the previously observed sampling locations from highest correlation to lowest with respect to $\mathbf{x}_{\text{exp}}^{(j)}$, we can prioritize the evaluation of Eq. (16) for the most promising samples. This enables the algorithm to reduce the necessary number of matrix inversions to typically be no more than $p \leq 5$. Consequently, as the computational cost of inverting an $n \times n$ matrix is $\mathcal{O}(n^3)$, we find that the computational cost per allocation of a batch containing B samples can be approximated as $\mathcal{O}((1+p)Bn^3)$. In contrast, the MQ and TSSO schemes only require one matrix inversion to allocate B samples, and thus, they have an approximated computational cost per allocated Batch as $\mathcal{O}(n^3)$. The proposed sampling scheme has a computational complexity that is roughly $(1+p)B$ more costly than conventional sampling schemes, and this increase in computational cost needs to be considered by a designer when selecting what optimization scheme to use. However, to put this in context, for the results presented in Sec. 4.1, the average time necessary to identify a new sampling location was approximately 2.1 s, an execution time that is negligible when juxtaposed to the simulation time of a single MD simulation (hours to days). We therefore believe that the proposed sampling scheme has great practical utility in many optimization problems involving stochastic objective functions.

(2) The number of costly simulation model evaluations allocated by the proposed sampling scheme depends implicitly on the selected batch size. Choosing a batch size that is smaller than the total number of simulations that can be run in parallel (n_{par}) has the disadvantage that computational resources are left idle. It would therefore be better to choose a batch size that is at least as large as n_{par}. However, choosing a much larger batch size is not recommended as the proposed batch

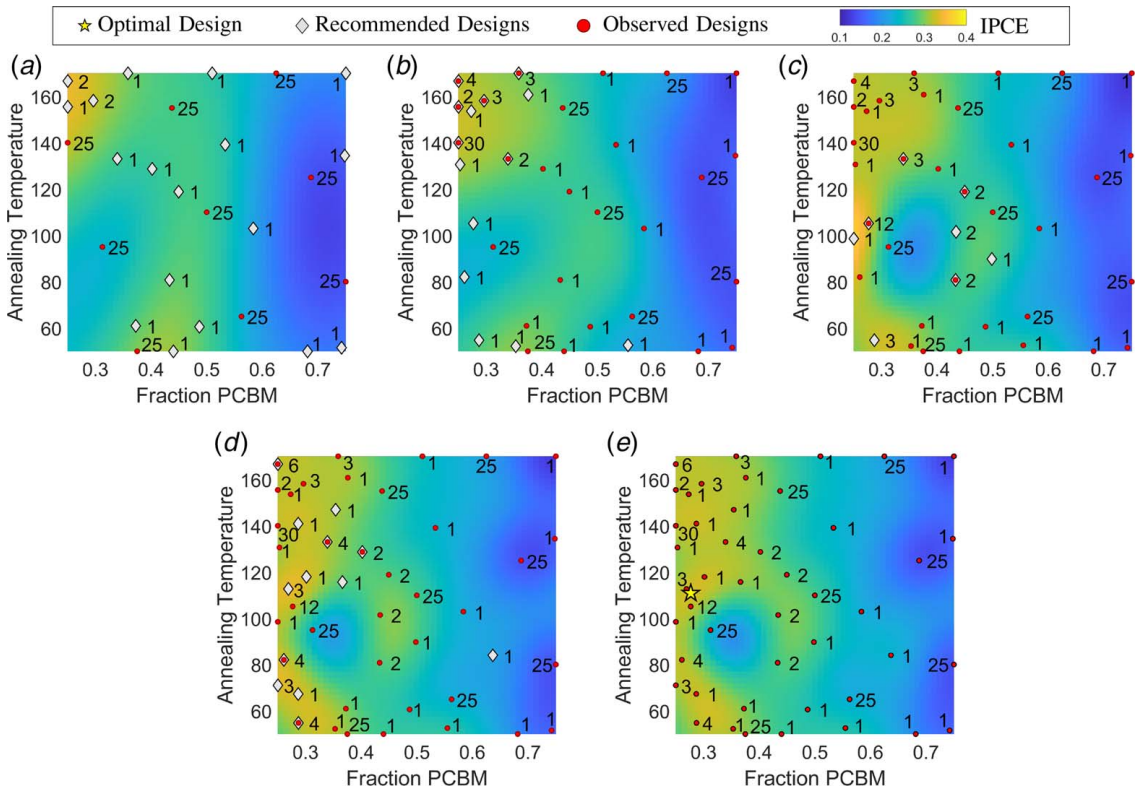

Fig. 8 Visualization of the initial experimental design and four subsequent micro-trials for the optimization of an OPVc as predicted from two processing conditions through a CGMD simulation. Numbers next to each sample indicates its number of replications. Overlapping samples indicate that the sampling scheme recommends replication at sampling locations that have been simulated in a previous micro-trial. (*a*) Micro-trial 1: recommended samples exploring 18 new sampling locations, (*b*) micro-trial 2: recommended samples exploring ten new sampling locations, (c) micro-trial 3: recommended samples exploring four new sampling locations, (*d*) micro-trial 4: recommended samples exploring eight new sampling locations, and (*e*) final surrogate model after observing 305 simulations. (Color version online.)

sampling scheme only uses the information of simulated samples. More specifically, in Eq. (18), for $B>1$, the jth, ($j=1, \ldots, B$) sample of a batch is allocated without knowing the simulation response of $j-1$ samples. Therefore, it is reasonable to expect that batch sampling schemes require more total number of simulations to identify the global optimum of a function in comparison to one sample at a time schemes. Nevertheless, by selecting a batch size $B \in (n_{\text{par}}, 3n_{\text{par}})$, a designer greatly reduces computational cost, while the proposed sampling scheme still makes good sampling decisions as the number of observed samples will be much larger than the batch size (i.e., $N \gg B$).

(3) Many of the existing sampling schemes proposed in the literature emphasize the choice of the acquisition function. However, the contribution of the presented work is the introduction of a sampling scheme that actively decides where and when to replicate and where and when to explore. The proposed work uses the acquisition function proposed in Ref. [22] as it neglects the contribution of the intrinsic uncertainty, thus focusing samples for exploration on minimizing the interpolation uncertainty. However, if designers wish for the sampling scheme to exploit more near currently promising design alternatives, then they could choose to exchange the modified EI acquisition function given in Eq. (11) used in step 3a of Fig. 2 with the MQ acquisition function given in Eq. (10) and select a θ that has the desired properties.

(4) Although the proposed sequential sampling scheme adds samples in a prudent manner, it still places a stringent demand on computational resources. This is because learning the true mean response of the samples around the global optimum requires many replications. In fact, the uncertainty introduced into the surrogate model by not knowing the population mean at a single unique sampling location will only reduce to zero once it has been sampled an infinite number of times. However, we have shown that a reasonably accurate approximation of the global optimum can be obtained by sufficiently replicating samples near the global optimum. The number of required replications depends on the variability and the magnitude of a simulation model's intrinsic noise; nevertheless, the proposed scheme has shown that in many cases, 20–100 replications at carefully selected sampling locations suffices.

6 Concluding Remarks

We proposed an objective-driven batch sampling scheme for the optimization of simulation models with intrinsic heteroscedastic noise. The use of costly simulations models that exhibit such "noisy" behavior is increasingly more commonplace in many engineering and scientific domains. The advantage of the proposed sampling scheme is twofold, (i) the computational cost of running simulation models is reduced by the recommendation of sampling batches (i.e., simulations can be run in parallel) and (ii) it can be scaled to higher dimensional and noisier simulation models than available schemes by reducing the size of the covariance matrix through replicating at previously simulated sampling locations. The functionality behind the proposed scheme comes from the reformulation of the posterior predictive variance to analytically derive the individual contribution of the interpolation uncertainty and the intrinsic modeling uncertainty through the

Sherman–Morrison–Woodbury formula. In addition, we introduce a preposterior analysis for the allocation of sampling batches that uses a zeroth-order interpolation for the intrinsic noise and is compatible with stochastic Kriging and practical Kriging surrogate models. Consequently, the proposed sampling scheme recommends batches of samples that minimize the variance of the posterior distribution only at and around the spatial locations expected to contain the globally optimal mean.

To illustrate its performance and sampling characteristics, we have applied the proposed adaptive sampling scheme to three test functions and one engineering design problem. This exercise demonstrated the scheme's ability to deal with relatively high-dimensional problems, various forms of intrinsic model uncertainty, and its superior performance with respect to existing schemes. Moreover, the usefulness of the proposed scheme is validated by demonstrating its effectiveness in identifying the optimal processing settings of an organic photovoltaic cell by optimizing a coarse-grained molecular dynamic (MD) simulation.

The presented investigation into the functionality of the proposed scheme shows promising results and opens the door for multiple future research directions. One avenue of future work includes the extension of the proposed scheme to integrate the decision at what length and time scale to run a new MD simulation. This would provide great insight into the intrinsic uncertainty that manifests in MD simulations, and other fields of science that use stochastic simulation models where the intrinsic uncertainty is negatively related to its controllable computational cost. A second avenue of future work includes the investigation into the influence of selected batch size to the efficiency of the sampling scheme. Such a study would not just benefit the optimization of simulation models with spatially varying noise, but all batch sampling schemes. In addition, the proposed sampling scheme can be extended to robust design optimization by integrating the acquisition function presented in Ref. [15]. This extension would provide designers with a tool to adaptively find an optimal design that is robust with respect to parametric, interpolation, and intrinsic uncertainty. In conclusion, the proposed sampling scheme provides a flexible and efficacious framework for the optimization of simulation models with heteroscedastic noise.

Acknowledgment

Support from AFOSRFA9550-18-1-0381, U.S. Department of Commerce under award No. 70NANB19H005 and National Institute of Standards and Technology as part of the Center for Hierarchical Materials Design (CHiMaD), and grant support from National Science Foundation (NSF) CMMI-1662435, 1662509, and 1753770 under the Design of Engineering Material Systems (DEMS) program are greatly appreciated.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request. The authors attest that all data for this study are included in the paper. No data, models, or code were generated or used for this paper.

Nomenclature

Symbols

n = total number of unique sampling locations
B = desired size of the sampling batches
N = total number of samples
$\mathbf{Y}$ = an N-dimensional vector of noisy simulation responses $\{y_1, \ldots, y_N\}^T$
$\bar{\mathbf{Y}}$ = an n-dimensional vector of sample means of observed noisy simulations $\{\bar{y}_1, \ldots, \bar{y}_n\}^T$ at n unique sampling locations
$\hat{r}_{ngb}(\bar{\mathbf{x}})$ = the sampling variance of the highest correlated unique sampling location to $\bar{\mathbf{x}}$ that has more than $\beta - 1$ replicates
n_{rep} = minimum number of replicates allocated to a new sampling location in the two-stage sequential optimization scheme
$y_{\min}$ = current best observed response
$\bar{y}_{\min}$ = current best observed sample mean of the response
$\mathbf{A}_n$ = an $n \times n$ diagonal matrix where the diagonal elements are the number of replicates $a_i \quad (i = 1, \ldots, n$ at each sampling location (i.e., $\mathbf{A} = \text{diag}(a_1, \ldots, a_n)$)
$\mathbf{K}_n$ = covariance matrix of a Gaussian process that facilitates replication
$\mathbf{K}_N$ = covariance matrix of a Gaussian process that does not facilitate replication
R_f = a function's range over design space χ
$\mathbf{X}_n$ = an n-dimensional tuple of d-dimensional unique sampling locations $\{\bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_n\}^T$
$\mathbf{X}_N$ = an N-dimensional tuple of d-dimensional sampling locations $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}^T$
$\mathbf{X}_{\text{new}}$ = new batch of samples $\{\mathbf{x}_{\text{new}}^{(1)}, \ldots, \mathbf{x}_{\text{new}}^{(B)}\}$, containing B new sampling locations
$\mathbf{X}_0$ = initial batch of samples
$NV_{\mathcal{X}}$ = fraction of macro-trials where the sampling scheme explores a new sampling location that is within $(1 - \mathcal{X})100\%$ of the range $f(\mathbf{x})$ to the globally optimal response
$NR_{\mathcal{X}}$ = fraction of macro-trials where the returned solution is within $(1 - \mathcal{X})100\%$ of the range $f(\mathbf{x})$ to the globally optimal response
$D^{(j)}$ = preposterior data set used for conditioning the posterior predictive distribution
$s_i(\mathbf{x}_{\text{exp}}^{(j)})$ = predicted reduction in the posterior variance at $\mathbf{x}_{\text{exp}}^{(j)}$ when replicating the ith sample
$\hat{\mathbf{x}}_{m,s}^*$ = predicted globally optimal design for the sth micro-trial of the mth macro-trial
$\bar{\mathbf{x}}_{i^*}^{(j)}$ = candidate sample for replication
$\mathbf{x}_{\text{exp}}^{(j)}$ = candidate sample for exploration
$\mathbf{x}_{\text{new}}^{(j)}$ = the jth sample added to the new sampling batch
$\mathbf{A}_i^{-1}$ = an $n \times n$ diagonal matrix $\text{diag}\left(0, \ldots, 0, a_{i+1}^{-1}, a_{i+2}^{-1}, \ldots, a_n^{-1}\right)$, i.e., the matrix $\mathbf{A}^{-1}$ with its first i diagonal elements equal to 0)
$c(\cdot, \cdot)$ = correlation function
$f(\cdot)$ = mean function of the costly simulation model
$k(\cdot, \cdot)$ = covariance function
$r(\cdot)$ = variance of the intrinsic model uncertainty
$q(\cdot)$ = minimum Kriging quantile
$\hat{r}(\bar{\mathbf{x}}_i)$ = sampling variance at the ith unique sampling location
$\mathbf{C}_n + \mathbf{\Delta}_n$ = correlation matrix of a Gaussian process that facilitates replication
$\mathbf{C}_N + \mathbf{\Delta}_N$ = correlation matrix of a Gaussian process that does not facilitate replication
$EI(\cdot)$ = expected improvement for a deterministic function
$EI_\varepsilon(\cdot)$ = expected improvement for a stochastic function
$\mathcal{N}_N(\cdot, \cdot)$ = Gaussian process that does not facilitate replication
$\mathcal{N}_n(\cdot, \cdot)$ = Gaussian process that facilitates replication
$S_n(\mathbf{x}_{\text{exp}}^{(j)})$ = predicted reduction in the posterior variance when exploring $\mathbf{x}_{\text{exp}}^{(j)}$
$\mathbf{0}_{n,i}$ = an n-dimensional vector of zeros, with its ith element equal to 1
$\mathcal{N}(\cdot, \cdot)$ = normal distribution
$\mathcal{O}(\cdot)$ = big O notation for computational complexity
$\mathcal{X}$ = tuning parameter used in $NV_{\mathcal{X}}$ and $NR_{\mathcal{X}}$

β = minimum number of replications considered sufficient to use a sampling locations sampling variance
θ = tuning parameter of the minimum Kriging quantile
$\boldsymbol{\lambda}_i$ = the ith $n \times n$ rank-1 matrix used for reformulation of a stochastic Kriging surrogate model's covariance matrix
$\boldsymbol{\Lambda}_i$ = the ith $n \times n$ matrix used as an intermediate matrix for reformulating a stochastic Kriging surrogate model's covariance matrix
$\mu(\cdot)$ = posterior mean prediction of a Gaussian process trained to a noiseless data set
$\mu_n(\cdot)$ = posterior mean prediction of a Gaussian process that facilitates replication
$\mu_N(\cdot)$ = posterior mean prediction of a Gaussian process that does not facilitate replication
$\sigma^2(\cdot)$ = posterior predictive variance of a Gaussian process trained to a noiseless data set
$\sigma_n^2(\cdot)$ = posterior predictive variance of a Gaussian process that facilitates replication
$\tilde{\sigma}_n^2(\cdot)$ = denoised posterior predictive variance of a Gaussian process that facilitates replication
$\sigma_N^2(\cdot)$ = posterior predictive variance of a Gaussian process that does not facilitate replication
$\boldsymbol{\Sigma}_n$ = an $n \times n$ Diagonal matrix of i.i.d. Gaussian process modeling noise that facilitates replication
$\boldsymbol{\Sigma}_n$ = an $n \times n$ Diagonal matrix of i.i.d. Gaussian process noise approximated through the sampling variance of the observed responses at each unique sampling location
$\boldsymbol{\Sigma}_N$ = an $N \times N$ Diagonal matrix of i.i.d. Gaussian process modeling noise that does not facilitate replication
$\phi(\cdot)$ = standard normal probability density function
$\Phi(\cdot)$ = standard normal cumulative distribution function
χ = $d-$dimensional space of afdmissible designs
ω, σ^2 = Gaussian process model hyperparameters
$\hat{\omega}$, $\hat{\sigma}^2$ = approximated Gaussian process model hyperparameters
Ω = d-dimensional admissible real design space for the Gaussian process hyperparameters

References

[1] Hansoge, N. K., Huang, T., Sinko, R., Xia, W., Chen, W., and Keten, S., 2018, "Materials by Design for Stiff and Tough Hairy Nanoparticle Assemblies," ACS. Nano., **12**(8), pp. 7946–7958.
[2] Ludkovski, M., and Niemi, J., 2011, "Optimal Disease Outbreak Decisions Using Stochastic Simulation," IEEE, Winter Simulation Conference, Phoenix, AZ, Dec. 11–14, pp. 3844–3853.
[3] Cioffi-Revilla, C., 2014, *Introduction to Computational Social Science*, Springer, Berlin/New York.
[4] Mehmani, A., Chowdhury, S., and Achille, M., 2015, "Predictive Quantification of Surrogate Model Fidelity Based on Modal Variations With Sample Density," Structural Multidiscipl. Optim., **52**(2), pp. 353–373.
[5] Rasmussen, C. E., and Williams, C. K. I., 2006, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA.
[6] Chen, R. J. W., and Sudjianto, A., 2002, "On Sequential Sampling for Global Metamodeling in Engineering Design," Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Montreal, QC, Canada, Sept. 29–Oct. 2, Vol. 463, pp. 1–10.
[7] Jones, D. R., 2001, "A Taxonomy of Global Optimization Methods Based on Response Surfaces," J. Global Optim., **21**(1), pp. 345–383.
[8] Forrester, A.I.J., and Keane, A. J., 2009, "Recent Advances in Surrogate-Based Optimization," Progress Aeros. Sci., **45**(1–3), pp. 50–79.
[9] Chen, S., Jiang, Z., Yang, S., Apley, D. W., and Chen, W., 2016, "Nonhierarchical Multi-Mmodel Fusion Using Spatial Random Processes," AIAA, **106**(7), pp. 503–526.
[10] Chen, S., Jiang, Z., Yang, S., and Chen, W., 2017, "Multimodel Fusion Based Sequential Optimization," AIAA, **55**(1), pp. 241–254.
[11] Lam, R., Willcox, K., and Wolpert, D. H., 2016, "Bayesian Optimization With a Finite Budget: An Approximate Dynamic Programming Approach," Conference on Neural Information Processing Systems, Barcelona, Catalonia, Spain, Dec. 5–11, Vol. 29, pp. 1–11.
[12] Lam, R., and Willcox, K., 2017, "Lookahead Bayesian Optimization With Inequality Constraints," Conference on Neural Information Processing Systems, Long Beach, CA, Dec. 4–9, Vol. 30, pp. 1–11.
[13] González, J., Osborne, M., and Lawrence, N. D., 2016, "Glasses: Relieving the Myopia of Bayesian Optimisation," International Conference on Artificial Intelligence and Statistics, Cadiz, Spain, May 9–11, Vol. 19, pp. 790–799.
[14] Assael, J.-A. M., Wang, Z., Shahriari, B., and de Freitas, N., 2015, "Heteroscedastic Treed Bayesian Optimisation," arXiv, https://arxiv.org/abs/1410.7172
[15] Arendt, P. D., Apley, D. W., and Chen, W., 2013, "Objective-Oriented Sequential Sampling for Simulation Based Robust Design Considering Multiple Sources of Uncertainty," ASME J. Mech. Des., **135**(5), p. 051005.
[16] Huang, D., Allen, T. T., Notz, W. I., and Zeng, N., 2006, "Global Optimization of Stochastic Black-Box Systems Via Sequential Kriging Meta-Models," J. Global Optim., **34**(3), pp. 441–466.
[17] Léazaro-Gredilla, M., and Titsias, M. K., 2011, "Variational Heteroscedastic Gaussian Process Regression," International Conference on Machine Learning, Bellevue, WA, June 28–July 2, Vol. 28, pp. 1–8.
[18] Binois, M., Huang, J., Gramacy, R. B., and Ludkovski, M., 2018, "Replication or Exploration? Sequential Design for Stochastic Simulation Experiments," Technometrics, **61**(1), pp. 7–23.
[19] Binois, M., Gramacy, R. B., and Ludkovski, M., 2018, "Practical Heteroscedastic Gaussian Process Modeling for Large Simulation Experiments," J. Comput. Graphical Stat., **27**(4), pp. 808–821.
[20] Ankenman, B., Nelson, B. L., and Staum, J., 2010, "Stochastic Kriging for Simulation Metamodeling," Inst. Operat. Res. Management Sci., **58**(2), pp. 371–382.
[21] Ariizumi, R., Tesch, M., Kato, K., Choset, H., and Matsuno, F., 2016, "Multiobjective Optimization Based on Expensive Robotic Experiments Under Heteroscedastic Noise," IEEE Trans. Rob., **33**(2), pp. 468–483.
[22] Quan, N., Yin, J., Ng, S. H., and Lee, L. H., 2013, "Simulation Optimization Via Kriging: A Sequential Search Using Expected Improvement With Computing Budget Constraints," IIE Trans., **45**(7), pp. 763–780.
[23] Picheny, V., Wagner, T., and Ginsbourger, D., 2013, "A Benchmark of Kriging-Based Infill Criteria for Noisy Optimization," Structural Multidiscipl. Optim., **48**(3), pp. 607–626.
[24] Hennig, P., and Schuler, C. J., 2011, "The Correlated Knowledge Gradient for Simulation Optimization of Continuous Parameters Using Gaussian Process Regression," SIAM J. Optim., **21**(3), pp. 996–1026.
[25] van Beek, A., Tao, S., Plumlee, M., Apley, D. W., and Chen, W., 2020, "Integration of Normative Decision-Making and Batch Sampling for Global Metamodeling," ASME J. Mech. Des., **142**(3), p. 031114.
[26] Loeppky, J. L., Morore, L. M., and Williams, B. J., 2010, "Batch Sequential Designs for Computer Experiments," J. Stat. Planning Inference, **140**(6), pp. 1452–1464.
[27] Zhu, P., Zhang, S., and Chen, W., 2013, "Multi-Point Objective-Oriented Sequential Sampling Strategy for Constrained Robust Design," Eng. Optim., **47**(3), pp. 287–307.
[28] Torossian, L., Picheny, V., and Durrande, N., 2020, "Bayesian Quantile and Expectile Optimisation," arXiv, https://arxiv.org/abs/2001.04833
[29] Martin, J. D., and Simpson, T. W., 2005, "On the Use of Kriging Models to Approximate Deterministic Computer Models," AIAA J., **43**(4), pp. 853–863.
[30] Plumlee, M., 2014, "Fast Prediction of Deterministic Functions Using Sparse Grid Experimental Designs," J. Am. Stat. Assoc., **109**(508), pp. 1581–1591.
[31] Xiong, Y., Chen, W., and Tsui, K.-L., 2008, "A New Variable-Fidelity Optimization Framework Based on Model Fusion and Objective-Oriented Sequential Sampling," J. Global Optim., **130**(11), pp. 1–9.
[32] Jones, D. R., Schonlau, M., and Welch, W. J., 1998, "Efficient Global Optimization of Expensive Black-Box Functions," J. Global Optim., **13**(1), pp. 455–492.
[33] Binois, M., Gramacy, R. B., and Ludkovski, M., 2012, "Entropy Search for Information-Efficient Global Optimization," J. Mach. Learn. Res., **13**(1), pp. 1809–1837.
[34] Forrester, A. I., Sóbester, A., and Keane, A. J., 2007, "Multi-Fidelity Optimization Via Surrogate Modelling," Proceedings of the Royal Society A, York, UK, July 17–20, Vol. 463, pp. 3251–3269.
[35] Frazier, P., Powell, W., and Dayanik, S., 2009, "The Knowledge-Gradient Policy for Correlated Normal Beliefs," INFORMS J. Comput., **21**(4), pp. 599–613.
[36] Picheny, V., Ginsbourger, D., Richet, Y., and Caplin, G., 2013, "Quantile-Based Optimization of Noisy Computer Experiments With Tunable Precision," Technometrics, **55**(1), pp. 2–13.
[37] Pedrielli, G., Wang, S., and Ng, S. H., 2020, "An extended Two-Stage Sequential Optimization approach: Properties and performance," Eur. J. Oper. Res., **45**, pp. 929–945.
[38] Olsson, A., Sandberg, G., and Dahlblom, O., 2003, "On Latin Hypercube Sampling for Structural Reliability Analysis," Structural Safety, **25**(1), pp. 47–68.
[39] Sobol, I., 1976, "Uniformly Distributed Sequences With an Additional Uniform Property," USSR Comput. Math. Math. Phys., **16**(5), pp. 236–242.
[40] van Beek, A., Ghumman, U. F., Munshi, J., Tao, S., Chien, T., Balasubramanian, G., Plumlee, M., Apley, D., and Chen, W., 2020, "Scalable Objective-Driven Batch Sampling in Simulation-Based Design for Models With Heteroscedastic Noise," Proceedings of the ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Virtual, online, Aug. 17–19.
[41] Sherman, J., and Morrison, W. J., 1950, "Adjustment of An Inverse Matrix Corresponding to a Change in One Element of a Given Matrix," Ann. Math. Stat., **21**(1), pp. 124–127.

[42] Jiang, Z., Apley, D. W., and Chen, W., 2015, "Surrogate Preposterior Analyses for Predicting and Enhancing Identifiability in Model Calibration," Int. J. Uncertainty Quantification, **5**(4), pp. 341–359.
[43] Ginsbourger, D., Riche, R. L., and Carraro, L., 2008, "A Multi-Points Criterion for Deterministic Parallel Global Optimization Based on Gaussian Processes." https://hal.archivesouvertes.fr/hal-00260579
[44] Dixon, L. C. W., and Szego, G. P., 1978, *The Global Optimization Problem: An Introduction*, North-Holland Pub. Co, Amsterdam, North-Holland.
[45] Jalali, H., van Nieuwenhuyse, I., and Picheny, V., 2016, "Comparison of Kriging-Based Methods for Simulation Optimization With Heterogeneous Noise," Eur. J. Oper. Res., **261**(1), pp. 279–301.
[46] Wu, J., Poloczek, M., Wilson, A. G., and Frazier, P. I., 2017, "Bayesian Optimization With Gradients," Conference on Neural Information Processing Systems, Long Beach, CA, Dec. 4–9, Vol. 31, pp. 1–12.
[47] Heeger, A. J., 2001, "Nobel Lecture: Semiconducting and Metallic Polymers: The Fourth Generation of Polymeric Materials," Rev. Mod. Phys., **73**(3), p. 681.
[48] Berger, P. R., and Kim, M., 2018, "Polymer Solar Cells: P3ht: Pcbm and Beyond," J. Renew. Sustainable Energy, **10**(1), p. 013508.
[49] Giulia, G., Juan, C.-G., Giulio, C., and Lanzani, G., 2011, "Transient Absorption Imaging of P3ht: Pcbm Photovoltaic Blend: Evidence for Interfacial Charge Transfer State," J. Phys. Chem. Lett., **2**(9), pp. 1099–1105.
[50] Umar, F. G., Akshay, I., Rabindra, D., Joydeep, M., Aaron, W., TeYu, C., Ganesh, B., and Wei, C., 2018, "A Spectral Density Function Approach for Active Layer Design of Organic Photovoltaic Cells," ASME J. Mech. Des., **140**(11), p. 111408.
[51] Marisol, R.-R., Kyungkon, K., and Carroll, D.L, 2005, "High-Efficiency Photovoltaic Devices Based on Annealed Poly (3-hexylthiophene) and 1-(3-methoxycarbonyl)-propyl-1-phenyl-(6, 6) C 61 Blends," Appl. Phys. Lett., **87**(8), p. 083506.
[52] Mihailetchi, V. D., Xie, H. X., de Boer, B., Koster, L. J. A, and Blom, P. W. M., 2006, "Charge Transport and Photocurrent Generation in Poly (3-hexylthiophene): Methanofullerene Bulk-Heterojunction Solar Cells," Adv. Funct. Mater., **16**(5), pp. 699–708.
[53] Pederson, K., Emblemsvåg, J., Bailey, R., Allen, J. K., and Mistree, F., 2000, "Validating Design Methods & Research: The Validation Square," Design Engineering Technical Conferences, Baltimore, MD, Sept. 10–14, pp. 1–13.