Proceedings of the ASME 2020
International Design Engineering & Technical Conferences
and Computers and Information in Engineering Conference
IDETC/CIE 2020
August 16-19, 2020, St. Louis, MO, USA

DETC2020-22629

SCALABLE OBJECTIVE-DRIVEN BATCH SAMPLING IN SIMULATION-BASED DESIGN FOR MODELS WITH HETEROSCEDASTIC NOISE

Anton van Beek[1], Umar Farooq Ghumman[1], Joydeep Munshi[2], Siyu Tao[1], TeYu Chien[3], Ganesh Balasubramanian[2], Matthew Plumlee[4], Daniel Apley[4], Wei Chen[1*]

[1]Dept. of Mechanical Engineering, Northwestern University, Evanston, IL

[2]Dept. of Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, PA

[3]Dept. of Physics and Astronomy, University of Wyoming, Laramie, WY

[4]Dept. of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL

[*]Corresponding author, Email: weichen@northwestern.edu, URL: https://ideal.mech.northwestern.edu

ABSTRACT

Objective-driven adaptive sampling is a widely used tool for the optimization of deterministic black-box functions. However, the optimization of stochastic simulation models as found in the engineering, biological, and social sciences is still an elusive task. In this work, we propose a scalable adaptive batch sampling scheme for the optimization of stochastic simulation models with input-dependent noise. The developed algorithm has two primary advantages: (i) by recommending sampling batches, the designer can benefit from parallel computing capabilities, and (ii) by replicating of previously observed sampling locations the method can be scaled to higher-dimensional and more noisy functions. Replication improves numerical tractability as the computational cost of Bayesian optimization methods is known to grow cubicly with the number of unique sampling locations. Deciding when to replicate and when to explore depends on what alternative minimizes the posterior prediction accuracy at and around the spatial locations expected to contain the global optimum. The algorithm explores a new sampling location to reduce the interpolation uncertainty and replicates to improve the accuracy of the mean prediction at a single sampling location. Through the application of the proposed sampling scheme to two numerical test functions and one real engineering problem, we show that we can reliably and efficiently find the global optimum of stochastic simulation models with input-dependent noise.

1 INTRODUCTION

The use of sampling for the analysis of computer simulations is a well-studied subject with applications in the engineering, biological [1] and material sciences [2]. These research efforts have driven the optimization of increasingly sophisticated design objects and provided insights into complex physical phenomena. Many experimental design methods are tailored for deterministic simulation models, whereas the use of stochastic simulation models is becoming increasingly more commonplace in the engineering, biological, and social sciences. We are specifically motivated by the use of molecular dynamics (MD) simulations in the context of engineering design as they are known for having dramatically changing signal-to-noise ratios across experiments (sampling locations). An example of an MD simulation for the performance prediction of an organic photovoltaic cell (OPVC) is presented in Fig. 1. The left section of the figure shows the molecular structure used as an input to the MD simulation, while the right section shows the response surface approximation obtained from simulating many such structures at nine unique sampling locations. Observe from the purple bars that the noise of the simulation model differs for each of the nine sampling locations. In this work we propose a Gaussian process (GP)-based objective-driven adaptive sampling scheme for the optimization of stochastic functions with spatially varying noise.

When simulation models are computationally expensive,

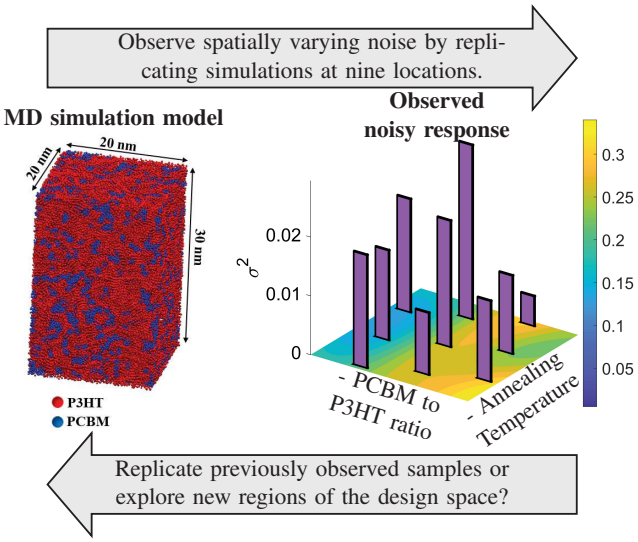


FIGURE 1: Example of an MD simulation for the efficiency prediction of an OPVC with spatially varying noise

training a response surface model to a data set obtained from an experimental design provides the designer with a surrogate that can predict the simulation response at an unobserved input location. For a general introduction into surrogate models and the validation of their fidelity, we refer the reader to [3]. One notable type of surrogate models are GPs, which have seen prolific use among many scientific communities [4]. The advantage of GP models over many 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 the largest. Such efforts are known as adaptive sampling for global surrogate modeling [5]. 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 and exploiting regions with a good mean response [6, 7]. The use of surrogate models has been extended to the optimization of multi-fidelity simulations [8, 9], non-myopic sampling [10–12], optimization of non-stationary functions [13], and robust design [14].

Despite the attention that objective-driven adaptive sampling has received, the challenge of finding the globally optimal mean of stochastic functions 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 (i.e., the noise is *homoscedastic*), and has been addressed in [15]. 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 only be simulated for a finite length and time scale. GP-based surrogate models have been developed to approximate stochastic functions, some popular examples of which are the variational heteroscedastic Gaussian process (VHGP) [16], practical Kriging (PK) [17, 18], and stochastic Kriging (SK) [19]. The VHGP model has been extended for application in global optimization in [20], but quickly becomes intractable because the covariance matrix of this model grows linearly with the number of samples. The SK and the PK models do not share this limitation as they allow samples 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 framework for the optimization of stochastic simulations. Preliminary efforts in this direction have been proposed in [21]. However, this work propose a rigid sampling scheme that requires the designer to define the ratio of replication to exploration a priori and greatly influences the efficiency and stability of the sampling process.

An additional challenge for adaptive sampling methods is the identification of sampling batches to facilitate parallel computing and subsequently improve computational efficiency. Examples of batch-based adaptive sampling methods for global surrogate modeling and optimization are given in [22, 23] and [24], respectively. Moreover, the desire for batch sampling is particularly prevalent in the MD community where it has become common practice to use supercomputers to run great numbers of 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, as acknowledged in [21, 25].

In this paper, we propose a tractable, objective-driven adaptive batch sampling approach for the optimization of stochastic simulation models with heteroscedastic noise. The main challenge of this work centers around an acquisition function that decides whether to explore a new sampling location or replicate a previously observed sample in an effort to improve the prediction accuracy only in the regions of interest. We show that by analyzing the surrogate models' posterior predictive variance we can separate the intrinsic model uncertainty from the surrogate model's interpolation uncertainty. Using this quantification, the algorithm decides whether to replicate (reduce intrinsic uncertainty) or explore (reduce interpolation uncertainty) by choosing the alternative 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. The performance of the proposed sampling scheme is demonstrated on two test functions and one engineering problem that involves the optimization of an organic photovoltaic cell (as presented in Fig. 1). 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 high signal-to-noise ratios.

2 BACKGROUND

In this section we provide an overview of surrogate modeling for stochastic functions, and the concept of adaptive sampling for global optimization. More importantly, this delineation provides justification for why replicating samples is critical when optimizing high-dimensional stochastic functions.

2.1 Surrogate Modeling of Stochastic Functions

Given a data set of noisy outputs $\mathbf{Y} = \{y_1, \dots, y_N\}^T$ observed at a set of d-dimensional sampling locations $\boldsymbol{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}^T$ we want to train a surrogate on the scalar-valued function $f : \mathbb{R}^d \rightarrow \mathbb{R}$. Under the assumption that the response 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 $k : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$. Normalizing the observed response to have zero mean centers the surrogate modeling effort around the covariance/kernel structure (e.g., power exponential, or Matérn [4]). In this work 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 hyperparameters that characterize the correlation function $c(\cdot)$.

Using a GP we can model the observations as a function of their spatial location $y_i = f(\mathbf{x}_i) + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0, r(\mathbf{x}_i))$ 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, in many cases 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 training data set as

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

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 $\boldsymbol{\Sigma}_N = \mathrm{diag}\left(r(\mathbf{x}_1), \dots, r(\mathbf{x}_N)\right)$. Note that under this formulation the ε_i's are independently and identically distributed.

Conditioning the GP prior on a set of observations provides a 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

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

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

and $\mathbf{k}_N(\mathbf{x}) = \{k(\mathbf{x}, \mathbf{x}_1), \dots, k(\mathbf{x}, \mathbf{x}_N)\}^T$. Prediction of the posterior response requires the identification of the hyper-parameters $\boldsymbol{\omega}$ that characterize the kernel function. By reformulating the covariance structure of our GP predictor using the correlation function as $\mathbf{K}_N + \boldsymbol{\Sigma}_N = \sigma^2\left(\mathbf{C}_N + \boldsymbol{\Delta}_N\right)$, we obtain the maximum likelihood estimation of the hyperparameters as

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

where we have removed all the constant terms, $\Omega \subset \mathbb{R}^d$ is the admissible design space of the hyperparameters, and we use the first-order optimality conditions to identify $\hat{\sigma}^2 = N^{-1}\mathbf{Y}^T\left(\mathbf{C}_N + \boldsymbol{\Delta}_N\right)^{-1}\mathbf{Y}$ [22, 26]. Note that both terms in Eqn. 4 depend on the roughness parameters $\boldsymbol{\omega}$ through the correlation function.

One function evaluation during the optimization of Eqn. 4 requires the inversion and determinant computation of $\mathbf{C}_N + \boldsymbol{\Delta}_N$ that come at a cubic computational expense $O(N^3)$. This is admissible for most deterministic (i.e., $r(\mathbf{x}) = 0$) or homoscedastic cases (i.e., $r(\mathbf{x}) = \gamma$) [27]; however, in the heteroscedastic case it is reasonable to expect that more simulation model evaluations are necessary. The computational cost can be addressed by allowing replication at previously observed sampling locations. Consider that we have n unique sampling locations $\bar{\mathbf{x}}_i \quad (i = 1, \dots, n)$, where at the i^{th} sampling location we have observed a_i replicates $y_i^{(j)}, \quad (j = 1, \dots a_i)$, (i.e., $\sum_{i=1}^{n} a_i = N$). $\bar{\mathbf{Y}} = \{\bar{y}_1, \dots, \bar{y}_n\}^T$ taken to be the sampling average over replicates (i.e., $\bar{y}_i = \frac{1}{a_i}\sum_{j=1}^{a_i} y_i^{(j)}$), then the posterior mean and variance are obtained as

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

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

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

The formulation given in Eqn. 5 and Eqn. 6 is known as SK and holds two challenges. The first challenge is that $\boldsymbol{\Sigma}_n$ requires the designer to known the intrinsic noise at each sampled location $r(\bar{\mathbf{x}}_i) \quad (i = 1, \dots, n)$; however, in many practical cases a designer has no access to this information. As an alternative, [19]

proposes an estimate for $\hat{\boldsymbol{\Sigma}}_n = \text{diag}\left(\hat{r}(\bar{\boldsymbol{x}}_1), \ldots, \hat{r}(\bar{\boldsymbol{x}}_n)\right)$ 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 Eqn. 7 results in an unbiased approximation of $\mu_n(\mathbf{x})$ when $a_i \gg 1$ (it is recommended to have $a_i \geq \beta = 10$). The second limitation is the formulation of the posterior predictive variance σ_n^2 for which 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 $r(\mathbf{x})$ and be satisfied with a "denoised" predictive variance as $\tilde{\sigma}_n^2(\mathbf{x}) = \sigma_n^2(\mathbf{x}) - r(\mathbf{x})$ [17]. An alternative approach is to place a separate GP prior on r as proposed in [19].

2.2 Bayesian Optimization of Deterministic Functions

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 [28], probability of improvement [6], expected improvement (EI) [29], knowledge gradient [30], and entropy search [31]. Despite the attention that objective-driven adaptive sampling methods 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 access to this knowledge a priori.

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

$$EI(\mathbf{x}) = \left((y_{min} - \mu(\mathbf{x}))\,\Phi(u) + S^2(\mathbf{x})\varphi(u)\right), \tag{8}$$

where y_{min} is the current best function observation, $\Phi(\cdot)$ is the standard normal cumulative distribution function, $\varphi(\cdot)$ 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 $u = \frac{y_{min} - \mu(\mathbf{x})}{\sigma^2(\mathbf{x})}$. A new sample is selected by maximizing the EI of the objective function as

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

where $\chi \in \mathbb{R}^d$ is the admissible design space. The EI function has appeared to be a versatile acquisition function that has good convergence properties for a broad range of problems [28, 32]; however, its application to simulation models with intrinsic noise has been hindered by its inability to replicate at previously observed sampling locations.

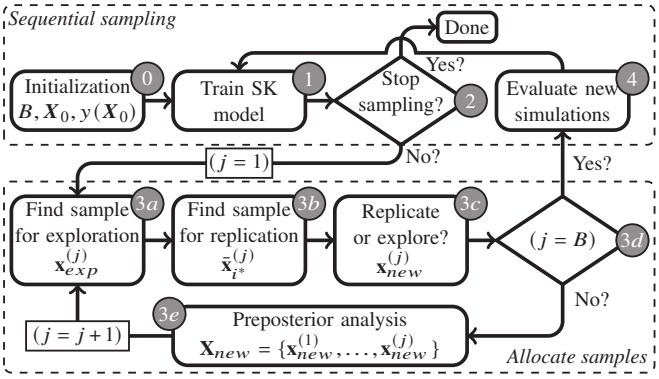


FIGURE 2: Flowchart for the proposed objective-driven adaptive sampling of stochastic simulation models

3 BAYESIAN OPTIMIZATION OF STOCHASTIC FUNCTIONS

In this section we introduce an SK-based objective-driven adaptive sampling algorithm that can balance the tradeoff between exploration and replication.

3.1 The Proposed Adaptive Sampling Scheme

The proposed adaptive sampling scheme as presented in Fig. 2 starts with an initialization phase (Step 0) where the designer has to evaluate the costly simulation model for a set of samples randomly dispersed over the design space (e.g., through a Latin hypercube design [33], or a Sobol sequence [34]). In addition, a 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 $2d - 5d$ [5], and the minimum number of recommended replications is $\beta \geq 10$ as suggested in [19]. In addition, the designer has to provide a preferred batch size B that will determine how many new samples will be recommended by the acquisition function during each sampling stage.

The subsequent two steps are similar to most adaptive sampling schemes, where in Step 1 of Fig. 2 we train our SK surrogate model to get a preliminary approximation of the mean response $f(\mathbf{x})$. Next, in Step 2 in Fig. 2, we determine if the stopping criterion is met, and if so, use the predicted optimal mean response from the current surrogate model. Two typical stopping criteria are: (i) the designer has exhausted all simulation resources, and (ii) the "denoised" posterior variance $\tilde{\sigma}_n^2(\mathbf{x})$ at the predicted optimal design is below a specific threshold γ.

The third step in the proposed sampling scheme is the identification of a batch of samples for which to evaluate the simulation model. This is where the proposed adaptive sampling scheme differs from other sampling methods. As shown in Fig. 2, Step 3 is iterated B times until a batch containing B samples is filled. In each iteration, Step 3a starts by identifying a candidate sampling location $\mathbf{x}_{exp}^{(1)}$ for exploration by maximizing a modified EI function as

$$EI_{\varepsilon}(\mathbf{x}) = ((\bar{y}_{min} - \mu_n(\mathbf{x}))\,\Phi(\bar{u}) + S_n^2(\mathbf{x})\varphi(\bar{u}), \qquad (10)$$

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 posterior variance if all unique sampling locations were to be sampled an infinite number of times (i.e., $\mathbf{A}^{-1} = diag(0,\ldots,0)$ in Eqn. 6) and $u = \frac{\bar{y}_{min} - \mu_n(\mathbf{x})}{S^2(\mathbf{x})}$. The intuition behind using $S_n(\mathbf{x})$ as the posterior variance in the EI function is that exploring a new sampling location will provide no information on the intrinsic noise of the objective function and is therefore not considered when deciding where to explore (a similar argument is presented in [21]). Note that, $\mathbf{x}_{exp}^{(1)}$ is only a candidate sample for exploration, rather we first need to find the best candidate sample for replication (Step 3b) before deciding whether to replicate or explore (Step 3c).

The best candidate sample for replication ($\bar{\mathbf{x}}_{i*}^{(1)}$) is found in Step 3b by choosing among all previously observed sampling locations $\bar{\mathbf{x}}_i \quad (i = 1,\ldots,n)$ the sample that provides the most information of the response $y(\cdot)$ at $\mathbf{x}_{exp}^{(1)}$. The purpose of replicating previous observed samples is to reduce the computational cost from $O(N^3)$ to $O(n^3)$ (see Section 2.1). We propose a heuristic that approximates the contribution of each previously observed sampling location to the variance of the predictive distribution $\hat{Y}(\mathbf{x}_{exp}^{(1)})|\mathbf{Y}$. The assumption behind this approach is that we assume to learn the most about the simulation response at $\mathbf{x}_{exp}^{(1)}$ by choosing to replicate the sampling location that minimizes the posterior predictive variance at that spatial location.

In Step 3c a decision is made to sample the costly simulation model either at $\mathbf{x}_{exp}^{(1)}$ or at $\bar{\mathbf{x}}_{i*}^{(1)}$. Adopting a similar consideration as for Step 3b, we want to choose the sample that minimizes the posterior variance at $\mathbf{x}_{exp}^{(1)}$ the most. For this purpose the proposed sampling scheme selects the best sample $\mathbf{x}_{new}^{(1)}$ from the set $\left\{\mathbf{x}_{exp}^{(1)}, \bar{\mathbf{x}}_{i*}^{(1)}\right\}$ to improve the surrogate model accuracy at the spatial location of interest.

It should be noted that going through Step 3a - Step 3c will only provide one sample, but prior to evaluating the costly simulation model (Step 4) we would like a batch containing B samples. Consequently, the algorithm will go through a preposterior analysis in Step 3e so that it can return to Step 3a and find a new sample $\mathbf{x}_{new}^{(2)}$. This process is repeated until a sampling batch $\mathbf{X}_{new} = \left\{\mathbf{x}_{new}^{(1)},\ldots,\mathbf{x}_{new}^{(B)}\right\}$ containing B samples has been identified. Next, we evaluate the new batch of samples with the simulation model in Step 4 and return to Step 1 to repeat the process until the stopping criterion in Step 2 is met.

The proposed sampling scheme pivots around the analysis of the posterior variance (Step 3a - Step 3c) and the preposterior analysis (Step 3e) that will be further discussed in Section 3.2 and Section 3.3, respectively.

3.2 Analysis of Posterior Predictive Variance

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

The first observation that needs to be made is the identification of two sources of uncertainty, (i) the interpolation uncertainty for not having explored the design space at $\mathbf{x}_{exp}$, and (ii) the uncertainty in the mean predictions of $\bar{y}_i, (i = 1,\ldots,n)$. Next, consider the contribution of the i^{th} sampling location to the variance of the posterior predictive distribution $Y(\mathbf{x}_{exp}^{(j)})|\mathbf{Y}$, the hypothetical maximum that can be learned from this location is to sample it an infinite number of times. By defining a matrix $\mathbf{A}_i^{-1} = \text{diag}\left(a_1^{-1},\ldots,a_{i-1}^{-1},0,a_{i+1}^{-1},\ldots,a_n^{-1}\right)$ we can approximate the posterior "denoised" variance after replicating at the i^{th} location an infinite number of times as

$$\tilde{\sigma}_{i,n}^2(\boldsymbol{x}) = k(\mathbf{x},\mathbf{x}) - \mathbf{k}_n^T(\mathbf{x})\left(\mathbf{K}_n + \mathbf{A}_i^{-1}\boldsymbol{\Sigma}_n\right)^{-1}\mathbf{k}_n(\mathbf{x}). \qquad (11)$$

The maximum reduction in the posterior variance at $\mathbf{x}_{exp}^{(j)}$ that can be achieved by replicating $\bar{\mathbf{x}}_i$ is then given as

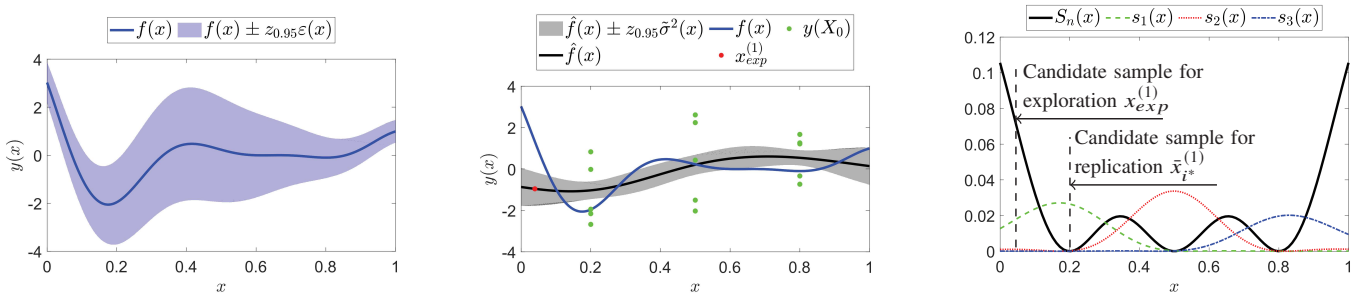
$$s_i(\mathbf{x}_{exp}) \approx \tilde{\sigma}_n^2(\mathbf{x}_{exp}^{(j)}) - \tilde{\sigma}_{i,n}^2(\mathbf{x}_{exp}^{(j)}). \qquad (12)$$

Using these approximations we find that the best sample for replication is $\bar{\mathbf{x}}_{i^*}^{(j)}$ where $i^* = \arg\max_i [s_i]_{1\le i\le n, i\in\mathbb{N}^+}$ (i.e., Step 3b in Fig. 2).

Similarly, we decide whether to replicate $\bar{\mathbf{x}}_{i^*}^{(j)}$ or explore $\mathbf{x}_{exp}^{(j)}$ as

$$\mathbf{x}_{new}^{(j)} = \begin{cases} \mathbf{x}_{exp}^{(j)} & \text{if } S_n(\mathbf{x}_{exp}^{(j)}) > s_{i^*} \\ \bar{\mathbf{x}}_{i^*}^{(j)} & \text{otherwise} \end{cases}, \qquad (13)$$

The intuition behind Eqn. 13 is that $S_n(\mathbf{x}_{exp}^{(j)})$ considers the interpolation uncertainty that will be reduced to zero once a sample is



(a) Example of a one-dimensional function with heteroscedastic intrinsic noise.

(b) SK approximation of a one-dimensional function with three unique sampling locations with five replications each.

(c) Analysis of posterior variance and the decision to explore or to replicate.

FIGURE 3: Visual example of approximating a one-dimensional stochastic function and the decision of whether to sample for exploration or for replication.

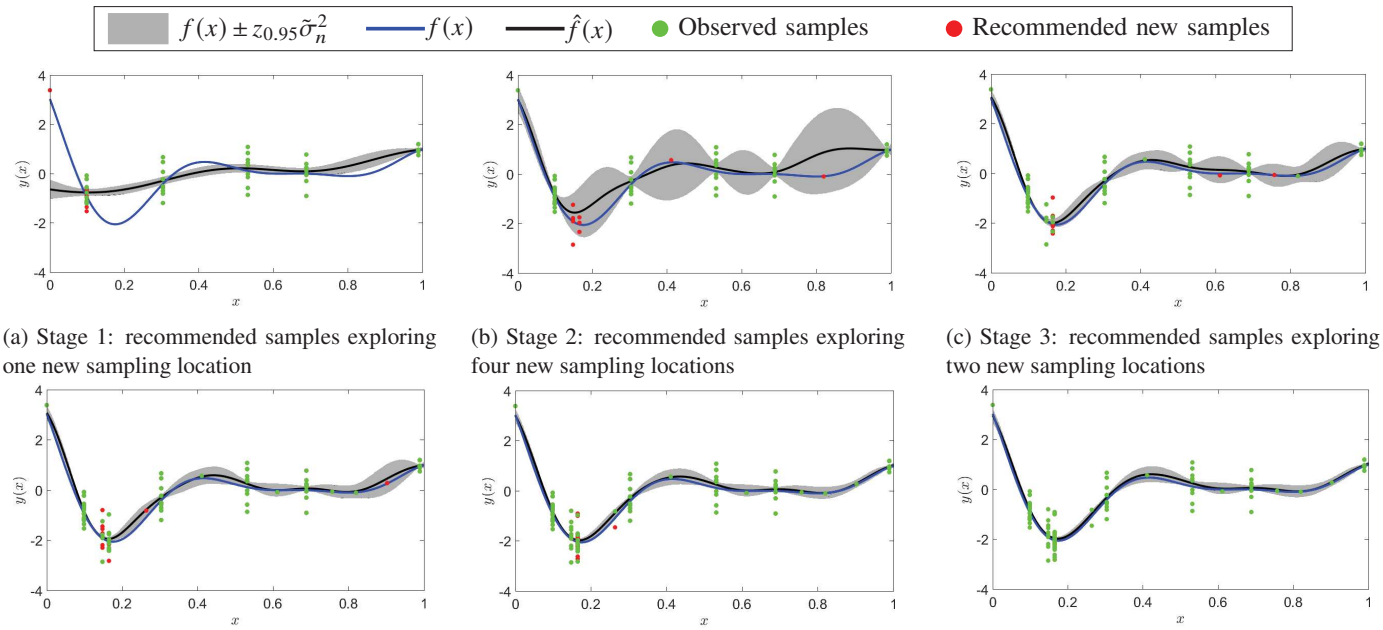
observed at that new spatial location $\mathbf{x}_{exp}^{(j)}$, while replicating the $(i^*)^{th}$ sample can only minimize the intrinsic modeling uncertainty that enter the surrogate model through the mean prediction at each associated spatial location.

For the purpose of visualizing the analysis of the posterior prediction variance, consider the one-dimensional simulation model with intrinsic noise presented in Fig. 3a. Moreover, assume that an initial batch of samples containing three unique sampling locations $\{0.2, 0.5, 0.8\}$ with five replications each has been simulated as given in Fig. 3b. Note that five replications are too few to have an unbiased predictor, but this number is used for illustration purposes. Next, assume that a candidate sample for exploration has been found at $x_{exp}^{(1)} = 0.04$ as presented by the red dot in Fig. 3b, it would then make sense that the corresponding candidate for replication is the nearest sample as given by $\bar{x}_{i^*}^{(1)} = 0.2$. The sampling locations for exploration and replication are also presented in Fig. 3c by the two vertical lines, the approximation of the interpolation uncertainty has been plotted by the black line $S_n(x)$, and the approximation of the reduction in the posterior variance for replicating the three sampling locations have been plotted by the dashed colored lines (i.e., the green line $s_1(x)$ captures the intrinsic uncertainty that enters the surrogate model through sampling location $x = 0.2$, the red line $s_2(x)$ is for $x = 0.5$, and the blue line $s_3(x)$ is for $x = 0.8$). What is more, from this figure we can observe that the uncertainty at the candidate sample for exploration is mostly driven by the interpolation uncertainty (i.e., $S_n(0.04) > s_i(0.04)$, for $i = 1, 2, 3$), and thus the proposed method 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 $s_2(0.45) > S_n(0.45) > s_1(0.45) > s_3(0.45)$ and thus the algorithm would choose to replicate $x = 0.45$.

3.3 Batch Sampling Through a Preposterior Analysis

Concerning the preposterior analysis (Step 3e Fig. 2) [35], this is a sampling method used to facilitate parallel computing by the allocation of batches of samples [36], and is also known as the "Kriging believer" method. For the deterministic case, the initial GP model $Y(\mathbf{x})|\mathbf{Y}$ is used to identify a new sampling location $\mathbf{x}_{new}$. The new sampling location is then used to obtain the posterior mean prediction $\mu(\mathbf{x}_{new})$ and is considered to be the true response value. Subsequently, the new observation $\{\mathbf{x}_{new}, \mu(\mathbf{x}_{new})\}$ is added to the training data set and is used for updating the new GP model (i.e., we do not need to retrain the GP model).

Using the preposterior analysis for an SK surrogate model differs from the GP model in that the posterior prediction depends not only on $\{\mathbf{x}_{new}, \mu(\mathbf{x}_{new})\}$, but also requires the intrinsic noise of the simulation model $r(\mathbf{x}_{new})$ and the number of replicates at $\mathbf{x}_{new}$. This is relatively straightforward when the algorithm decides to replicate the $i^{th}(i = 1, \ldots, n)$ sampling location as we can incrementally increase the number of replicates a_i by one. However, when exploring a new sampling location, the designer needs an approximation for the intrinsic modeling uncertainty $r(\mathbf{x})$. One approximation is to train an individual GP model to $\log(\hat{r}(\mathbf{x}))$. Nevertheless, 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 i^{th} sampling locations where $a_i < \beta$ will be assigned the sampling variance of its highest correlated neighbour that does have enough replicates. 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) when multiple samples are in proximity of each other the GP interpo-



(a) Stage 1: recommended samples exploring one new sampling location

(b) Stage 2: recommended samples exploring four new sampling locations

(c) Stage 3: recommended samples exploring two new sampling locations

(d) Stage 4: recommended samples exploring two new sampling locations

(e) Stage 5: recommended samples exploring no new sampling locations

(f) Final surrogate model approximation after observing a total of 100 simulations

FIGURE 4: Surrogate model approximations and recommended sampling batches for five stages of a one-dimensional test function

lation approximation can overestimate the intrinsic uncertainty at an unobserved sampling location, potentially resulting in an unnecessary large number of replicates.

4 RESULTS

In this section we present the results of the proposed sampling scheme tested on three examples: (i) a one-dimensional expository test function to visualize the proposed method's sampling characteristics, (ii) a six-dimensional test function to demonstrate the proposed method's performance on a higher dimensional problem, and (iii) a real two-dimensional engineering example for the performance optimization of an OPVC.

4.1 One-Dimensional Test Function

We modify the one dimensional test function used in [32,37] to exhibit heteroscedastic intrinsic noise as

$$y(x) = (3x-2)^2 \sin(12x-4) + \varepsilon(x), \quad 0 \le x \le 1, \tag{14}$$

$$\varepsilon(x) \sim \mathcal{N}\left(0, \frac{1+\exp(1.2-3x)^2}{\sqrt{6}}\right). \tag{15}$$

The optimization is started by defining an initial set of samples containing five unique locations allocated through a Latin hypercube design each with ten replicates (i.e., $\mathbf{A} = \text{diag}(10,\ldots,10)$ for a total of 50 samples). In addition, we want the sampling scheme to allocate batches containing ten samples, so we set the batch size $B = 10$. Next, we evaluate these samples with Eqn. 14 and train an SK surrogate model as visualized in Fig. 4a.

During the first call of the sampling scheme as presented in Fig. 4a the SK model predicts that the underlying mean function is relatively linear. Consequently, the sampling scheme decides to explore one new sampling location $x = 0$ and suggests nine replications at the current best observed sampling location as presented by the red dots. In the next stage Fig. 4b, we find that the underlying mean function is more nonlinear than expected, and therefore all new samples are allocated to explore four new sampling locations, with more replicates allocated to the samples that are expected to contain the global optimum. As the algorithm progresses through three additional sampling stages as given by Fig. 4c through Fig. 4e, we observe that fewer samples are used to explore new sampling locations, and more samples are used to reduce the "denoised" posterior variance around the global optimum as presented by the gray shaded regions. After observing 50 new samples, most of which have been allocated in the region $[0.1, 0.3]$ as can be seen from Fig. 4f, we successfully identified the global optimum at $x^* = 1.76$.

4.2 Six-Dimensional Test Function

To show that the proposed sampling scheme can deal with relatively high-dimensional optimization problems, we use it to minimize a modified version of the six-dimensional Hartmann function presented in [38] and given as

$$y(\mathbf{x}) = 5 - \sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} Q_{ij}(x_j - P_{ij})^2\right) + \varepsilon(x), \tag{16}$$

$$\varepsilon(x) \sim \mathcal{N}(0, 0.1y(\mathbf{x})), \quad \mathbf{x} \in [0,1]^6, \tag{17}$$

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 sampling problem with twenty unique sampling locations using a Latin hypercube design with twenty replications each. We then go through twenty stages, during each of which a batch containing 50 samples is identified (i.e., $B = 50$). Moreover, this process is repeated ten times, and the resulting convergence history for these simulations is presented by the black line in Fig. 5. From this figure we observe that the proposed sampling scheme quickly narrows in on 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 consider an alternative sampling scheme where the EI algorithm of Eqn. 8 is used to identify a new sampling location to which a fixed number of replicates is added. The result for this approach considering 5, 25 and 50 replicates for each of these new sampling locations is presented by the green line with square markers, blue line with pentagram markers and red line with circle markers, respectively. Note that neither of these methods outperforms the proposed sampling scheme, this is because some spatial locations only require a small number of samples for the algorithm to be confident that it does not contain the global optimum, while spatial locations with a better mean response closer to the global optimum require more replicates.

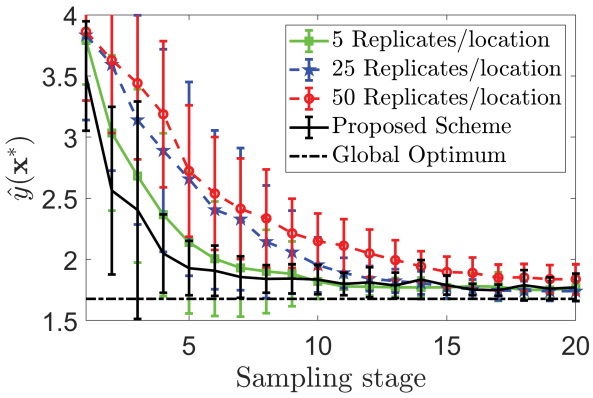
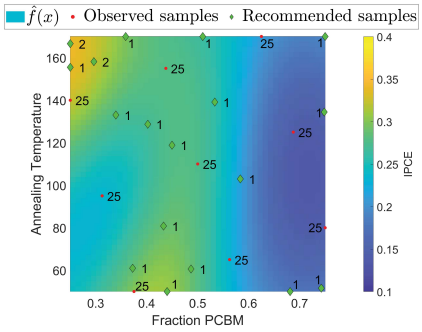


FIGURE 5: Convergence history of the proposed adaptive sampling scheme for the optimization of the six-dimensional Hartmann function, adding 50 samples per stage.

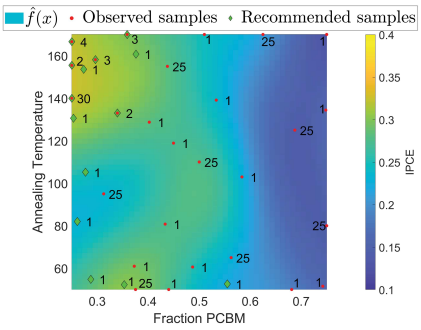
4.3 Optimization of an Organic Photovoltaic Cell

In this section we use the proposed adaptive sampling scheme for the optimization of an OPVC [39, 40]. The preferred choice of architecture for an OPVC is bulk heterojunction [41] 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. All the four processes are directly affected by the microstructure of the OPVC, which in turn is a product of its processing conditions.

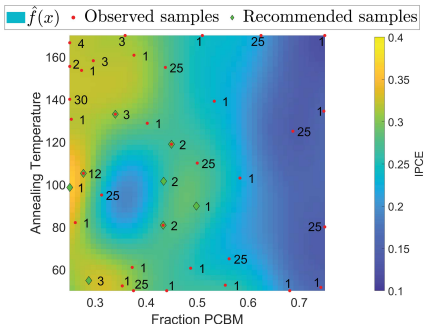
The production process used for the fabrication of these thin film OPVCs is known as spin coating, where a small amount of coating material is deposited onto a substrate and rotated. The centrifugal forces spread the coating evenly over the substrate, and rotation is continued until the desired film 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. Though there are other important processing conditions, only the two predominant ones were considered: ratio of PCBM:P3HT, and annealing temperature. In a previous work [42], we developed a structure-performance simulation incorporating the four aforementioned physical phenomena to predict the IPCE value for any digital OPVC microstructure. The same simulation, with slight modifi-



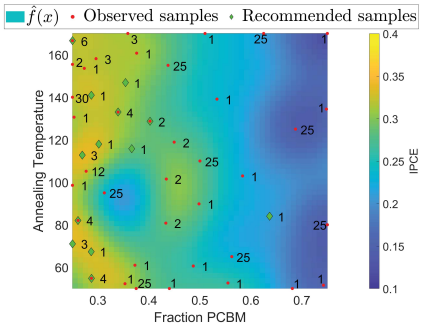
(a) Stage 1: recommended samples exploring eighteen new sampling locations



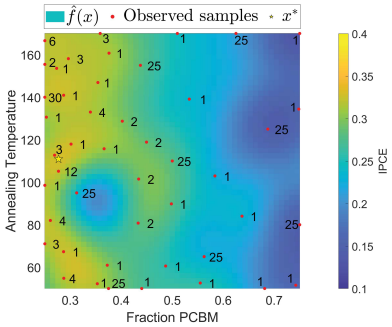
(b) Stage 2: recommended samples exploring ten new sampling locations



(c) Stage 3: recommended samples exploring four new sampling locations



(d) Stage 4: recommended samples exploring eight new sampling locations



(e) Final surrogate model after observing 305 simulations

FIGURE 6: Surrogate model approximation and recommended sampling batches for five stages of a one dimensional test function

cations, is employed in this work to evaluate the IPCE value for all the data points. The total simulation time for the processing-structure-performance linkage is about 36-38 hours.

Because CGMD simulations can only be evaluated at a finite length and time scale, and have random initial conditions, we find that the prediction of the IPCE exhibits spatially varying intrinsic noise (see purple bars in Fig. 1). We start the optimization scheme with an optimal Latin hypercube design containing nine unique sampling locations, 25 replications at each sampling location, and batch size set as $B = 20$. Next, we run 225 CGMD simulations and train an SK model on the obtained data set. The mean prediction of the surrogate model has been presented in Fig. 6a, where the red dots indicate the observed sampling locations, the green diamonds are the recommended next sampling locations, and the numbers next to each sampling location indicate the number of replicates. During the first stage we observe that most recommended samples are used to explore the two sampling locations predicted to contain the global optimum.

After evaluating the twenty samples as recommended by the proposed sampling scheme, we observe from Fig. 6b that the region around $\mathbf{x} = \{0.3, 160\}$ is expected to contain the global optimum. Consequently, half of the samples of the new batch are placed in this region, while the other half are 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. 6c. However, after exploring this new region with the next batch of samples (shown in Fig. 6d), we find that this was an overly optimistic prediction and now observe that the global optimum is found at $\mathbf{x}^* = \{0.2704, 113.7\}$. However, an additional batch of samples is required, because the spatial location of the global optimum has shifted significantly and the posterior predictive distributing 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 global optimum response as $\hat{y}(\mathbf{x}^*) = \mathcal{N}(0.3299, 0.0067)$ at $\mathbf{x}^* = \{0.2769, 110.89\}$ as shown by the yellow star in Fig. 6e. After this stage the sampling procedure is stopped as no significant change in the spatial location of the optimal response is observed, and the posterior predictive variance is satisfactorily low.

The optimal volume fraction of PCBM in literature is around 0.4 [43], but the experiment samples are larger in thickness (generally more than a 100 nm) compared to our simulations ($20nm$). At this simulation length scale, it is expected that the most promi-

nent physical phenomenon, from the four mentioned earlier, is the exciton creation, that is a function of the material's absorption coefficient. PCBM has a lower absorption coefficient than P3HT [44], and thus more P3HT is favored than PCBM. This leads to a general trend of high performance with low PCBM fraction. However, even with this small thickness, a prominent feature is observed at ≈ 0.45 PCBM fraction. The capture of this feature is intriguing as it substantiates that charge diffusion and collection, charge separation, and light absorption are valuable physics to include in our CGMD model.

4.4 Discussion on the Proposed Sampling Scheme

The proposed adaptive sampling scheme holds several assumptions that warrant additional discussion.

1. Though the proposed 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 intrinsic uncertainty introduced through 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 at sampling locations around the global optimum. The number of required replications depends on the signal-to-noise ratio of the simulation model, but the proposed method has shown that in many cases, 20 to 100 replications at carefully selected sampling locations will suffice.
2. In the introduced adaptive sampling scheme we have proposed an approximation for the individual contribution of each unique sampling location to the posterior predictive variance. This approach has two limitations, (i) it is only an approximation and might therefore lead to incorrect sampling decisions, and (ii) it uses considerable computational resources to invert $\mathbf{K}_n + \mathbf{A}^{-1}\boldsymbol{\Sigma}_n$ a total of n times. This implies that the proposed sampling scheme is at least n times more costly than conventional optimization schemes (i.e., i.e., objective-driven batch sampling for deterministic simulation models). Future work will include an investigation into the use of the Sherman-Morisson formulation as presented in [45] to provide fast and accurate prediction of $s_i(\cdot) \quad (i = 1, \ldots, n)$.
3. The proposed sampling scheme has to make an approximation of the sampling variance at sampling locations with fewer than β replicates. We proposed a zeroth-order interpolation for this purpose, but the validity of such an approach has not yet been tested. The intrinsic noise of the presented test problems varies only gradually; however, the assumption of a zeroth-order interpolation might become more consequential when the intrinsic simulation noise varies more wildly over the design space. Future work will include a detailed study to validate the efficacy of using a zeroth-order interpolation for the approximation of the heteroscedastic intrinsic simulation noise.
4. The computational cost of the proposed sampling scheme depends predominantly on Step 3a and Step 3b as presented in Fig. 2. More specifically, in Step 3a the algorithm has to invert the covariance matrix one time, While in Step 3b the covariance matrix has to be inverted p times. Where p is the number of previously observed sampling location for which to evaluate Eqn. 12 (i.e., $1 \leq p \leq n$). By evaluating Eqn. 12 for the previously observed sampling location in the order of their highest correlation to $\mathbf{x}_{exp}$ we typically find that after $p \ll n$ no better sample for replication exists. Consequently, we can proceed to Step 3c without evaluating Eqn. 12 for all n previously observed sampling locations. The computational cost for allocating a new sample equals $O((p+1)n^3)$ and should be taken into consideration when deciding to use the proposed adaptive sampling scheme.

5 Concluding Remarks

We propose a batch-based objective-driven adaptive sampling scheme for the optimization of simulation models with intrinsic heteroscedastic noise. Simulation models in various engineering domains increasingly exhibit such "noisy" behaviour and are typically associated with considerable computational cost. The advantage of the proposed sampling scheme is twofold, (i) the computational cost of running simulation models is minimized 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 methods by reducing the size of the covariance matrix through replication at previously simulated sampling locations. The functionality behind the proposed method comes from analyzing the variance of the posterior prediction distribution and approximating the contribution of the interpolation uncertainty and the intrinsic modeling noise associated with each previously simulated sampling location. Consequently, the proposed sampling scheme recommends batches of samples to minimize the variance of the posterior distribution at and around the spatial locations expected to contain the globally optimal mean.

The proposed adaptive sampling scheme has been tested on three design problems to illustrate its sampling characteristics, its ability to deal with relatively high-dimensional problems, and its effectiveness in finding the optimal processing settings using coarse-grained molecular dynamic simulations for the fabrication of an organic photovoltaic cell. The investigation into the functionality of the proposed method shows promising results and opens the door for future research. One avenue of future work includes the derivation and implementation of a closed-form solution for the contribution of the intrinsic uncertainty of each previously observed sampling location to the posterior vari-

ance at a new sampling location. A second avenue of future work includes the investigation into the influence of selected batch size on 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-based adaptive sampling schemes. In conclusion, we proposed a potent sampling scheme for the optimization of costly stochastic functions, which are becoming increasingly more commonplace in a multitude of scientific fields.

ACKNOWLEDGMENT

Support from AFOSR FA9550-18-1-0381, 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.

REFERENCES

[1] Ludkovski, M., and Niemi, J., 2011. "Optimal disease outbreak decisions using stochastic simulation". In IEEE, Winter Simulation Conference, pp. 3844–3853.

[2] 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), July, pp. 7946–7958.

[3] Mehmani, A., Chowdhury, S., and Achille, M., 2015. "Predictive quantification of surrogate model fidelity based on modal variations with sample density". *Structural and Multidisciplinary Optimization,* ***52***(2), August, pp. 353–373.

[4] Rasmussen, C. E., and Williams, C. K. I., 2006. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA.

[5] Chen, R. J. W., and Sudjianto, A., 2002. "On sequential sampling for global metamodeling in engineering design". In Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. **463**, pp. 1–10.

[6] Jones, D. R., 2001. "A taxonomy of global optimization methods based on response surfaces". *Journal ofGlobal Optimization,* ***21***(1), June, pp. 345–383.

[7] I.J.Forrester, A., and Keane, A. J., 2009. "Recent advances in surrogate-based optimization". *Progress in Aerospace Sciences,* ***45***(1-3), January - April, pp. 50–79.

[8] Chen, S., Jiang, Z., Yang, S., Apley, D. W., and Chen, W., 2016. "Nonhierarchical multi-model fusion using spatial random processes". *AIAA,* ***106***(7), August, pp. 503–526.

[9] Chen, S., Jiang, Z., Yang, S., and Chen, W., 2017. "Multimodel fusion based sequential optimization". *AIAA,* ***55***(1), August, pp. 241–254.

[10] Lam, R., Willcox, K., and Wolpert, D. H., 2016. "Bayesian optimization with a finite budget: An approximate dynamic programming approach". In Conference on Neural Information Processing Systems, Vol. **29**, pp. 1–11.

[11] Lam, R., and Willcox, K., 2017. "Lookahead bayesian optimization with inequality constraints". In Conference on Neural Information Processing Systems, Vol. **30**, pp. 1–11.

[12] González, J., Osborne, M., and Lawrence, N. D., 2016. "Glasses: Relieving the myopia of bayesian optimisation". In International Conference on Artificial Intelligence and Statistics, Vol. **19**, pp. 790–799.

[13] Assael, J.-A. M., Wang, Z., Shahriari, B., and de Freitas, N., 2015. Heteroscedastic treed bayesian optimisation. arXiv, March. URL `arxiv.org /pdf /1410.7172.pdf`.

[14] Arendt, P. D., Apley, D. W., and Chen, W., 2013. "Objective-oriented sequential sampling for simulation based robust design considering multiple sources of uncertainty". *Journal of Mechanical Design,* ***135***(5), May, pp. 1–10.

[15] Huang, D., Allen, T. T., Notz, W. I., and Zeng, N., 2006. "Global optimization of stochastic black-box systems via sequential kriging meta-models". *Journal of Global Optimization,* ***34***(3), March, pp. 441–466.

[16] Léazaro-Gredilla, M., and Titsias, M. K., 2011. "Variational heteroscedastic gaussian process regression". In International Conference on Machine Learning, Vol. **28**, pp. 1–8.

[17] Binois, M., Huang, J., Gramacy, R. B., and Ludkovski, M., 2018. "Replication or exploration? sequential design for stochastic simulation experiments". *Technometrics,* ***61***(1), May, pp. 7–23.

[18] Binois, M., Gramacy, R. B., and Ludkovski, M., 2018. "Practical heteroscedastic gaussian process modeling for large simulation experiments". *Journal of Computational and Graphical Statistics,* ***27***(4), April, pp. 808–821.

[19] Ankenman, B., Nelson, B. L., and Staum, J., 2010. "Stochastic kriging for simulation metamodeling". *Institute for Operations Research and the Management Sciences,* ***58***(2), March, pp. 371–382.

[20] Ariizumi, R., Tesch, M., Kato, K., Choset, H., and Matsuno, F., 2016. "Multiobjective optimization based on expensive robotic experiments under heteroscedastic noise". *IEEE Transactions on Robotics,* ***33***(2), December, pp. 468–483.

[21] 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 Transactions,* ***45***(7), April, pp. 763–780.

[22] 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". *Journal of Mechanical Design,* ***142***(3), January, pp. 1–11.

[23] L.Loeppky, J., J., L. M. B., and Williams, 2010. "Batch sequential designs for computer experiments". *Journal of Statistical Planning and Inference,* ***140***(6), June, pp. 1452–1464.

[24] Zhu, P., Zhang, S., and Chen, W., 2013. "Multi-point objective-oriented sequential sampling strategy for constrained robust design". *Engineering Optimization,* ***47***(3), March, pp. 287–307.

[25] Torossian, L., Picheny, V., and Durrande, N., 2020. Bayesian quantile and expectile optimisation. arXiv, January. URL `arxiv.org /pdf /2001.04833.pdf`.

[26] Martin, J. D., and Simpson, T. W., 2005. "On the use of kriging models to approximate deterministic computer models". *AIAA Journal,* ***43***(4), April, pp. 853–863.

[27] Plumlee, M., 2014. "Fast prediction of deterministic functions using sparse grid experimental designs". *Journal of the American Statistical Association,* ***109***(508), March, pp. 1581–1591.

[28] Xiong, Y., Chen, W., and Tsui, K.-L., 2008. "A new variable-fidelity optimization framework based on model fusion and objective-oriented sequential sampling". *Journal of Global Optimization,* ***130***(11), October, pp. 1–9.

[29] Jones, D. R., Schonlau, M., and Welch, W. J., 1998. "Efficient global optimization of expensive black-box functions". *Journal of Global Optimization,* ***13***(1), June, pp. 455–492.

[30] Hennig, P., and Schuler, C. J., 2011. "The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression". *SIAM Journal on Optimization,* ***21***(3), June, pp. 996–1026.

[31] Binois, M., Gramacy, R. B., and Ludkovski, M., 2012. "Entropy search for information-efficient global optimization". *Journal of Machine Learning Research,* ***13***(1), June, pp. 1809–1837.

[32] Forrester, A. I., Sóbester, A., and Keane, A. J., 2007. "Multi-fidelity optimization via surrogate modelling". In Proceedings of the Royal Society A, Vol. **463**, pp. 3251–3269.

[33] Olsson, A., Sandberg, G., and Dahlblom, O., 2003. "On latin hypercube sampling for structural reliability analysis". *Structural Safety,* ***25***(1), January, pp. 47–68.

[34] Sobol, I., 1976. "Uniformly distributed sequences with an additional uniform property". *USSR Computational Mathematics and Mathematical Physics,* ***16***(5), May, pp. 236–242.

[35] Jiang, Z., Apley, D. W., and Chen, W., 2015. "Surrogate preposterior analyses for predicting and enhancing identifiability in model calibration". *International Journal for Uncertainty Quantification,* ***5***(4), January, pp. 341–359.

[36] Ginsbourger, D., Riche, R. L., and Carraro, L., 2008. Simulation optimization via kriging: a sequential search using expected improvement with computing budget constraints. hal-00260579, March. URL `hal.archives-ouvertes.fr/hal-00260579`.

[37] van Beek, A., Tao, S., and Chen, W., 2019. "Global emulation through normative decision making and thrifty adaptive batch sampling". In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. **463**, pp. 1–13.

[38] Wu, J., Poloczek, M., Wilson, A. G., and Frazier, P. I., 2017. "Bayesian optimization with gradients". In Conference on Neural Information Processing Systems, Vol. **31**, pp. 1–12.

[39] Heeger, A. J., 2001. "Nobel lecture: Semiconducting and metallic polymers: The fourth generation of polymeric materials". *Reviews of Modern Physics,* ***73***(3), p. 681.

[40] Berger, P., and Kim, M., 2018. "Polymer solar cells: P3ht: Pcbm and beyond". *Journal of Renewable and Sustainable Energy,* ***10***(1), p. 013508.

[41] Grancini, G., Polli, D., Fazzi, D., Cabanillas-Gonzalez, J., Cerullo, G., and Lanzani, G., 2011. "Transient absorption imaging of p3ht: Pcbm photovoltaic blend: Evidence for interfacial charge transfer state". *The Journal of Physical Chemistry Letters,* ***2***(9), pp. 1099–1105.

[42] Farooq Ghumman, U., Iyer, A., Dulal, R., Munshi, J., Wang, A., Chien, T., Balasubramanian, G., and Chen, W., 2018. "A spectral density function approach for active layer design of organic photovoltaic cells". *Journal of Mechanical Design,* ***140***(11).

[43] Reyes-Reyes, M., Kim, 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". *Applied Physics Letters,* ***87***(8), p. 083506.

[44] Mihailetchi, V. D., Xie, H., de Boer, B., Koster, L. A., and Blom, P. W., 2006. "Charge transport and photocurrent generation in poly (3-hexylthiophene): methanofullerene bulk-heterojunction solar cells". *Advanced Functional Materials,* ***16***(5), pp. 699–708.

[45] Miller, K. S., 1981. "On the inverse of the sum of matrices". *Mathematics Magazine,* ***54***(2), March, pp. 67–72.