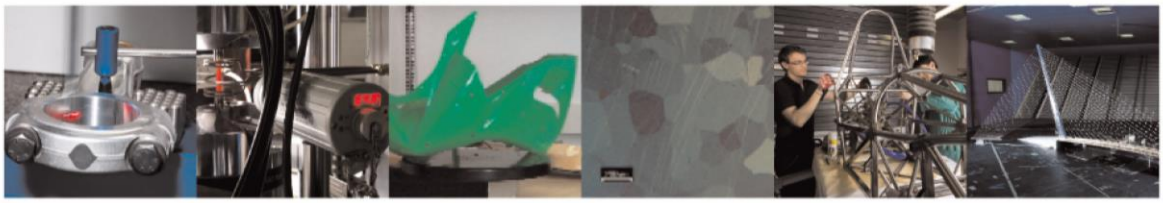


An Adaptive Fused Sampling Approach of High-Accuracy Data in the Presence of Low-Accuracy Data

Reisi Gahrooei, Mostafa; Paynabar, Kamran; Pacella, Massimo; Colosimo, BIANCA MARIA

This is an Accepted Manuscript of an article published by Taylor & Francis in IISE TRANSACTIONS on 20 Dec 2018, available online:
http://dx.doi.org/10.1080/24725854.2018.1540901

Mostafa Reisi Gahrooei, Kamran Paynabar, Massimo Pacella,Bianca M Colosimo

An Adaptive Fused Sampling Approach of High-Accuracy Data in the Presence of Low-Accuracy Data

Abstract

In several applications, a large amount of low-accuracy (LA) data can be acquired at a small cost. However, in many situations, such LA data is not sufficient for generating a high-fidelity model of a system. To adjust and improve the model constructed by LA data, a small sample of high-accuracy (HA) data, which is expensive to obtain, is usually fused with the LA data. Unfortunately, current techniques assume that the HA data is already collected and concentrate on fusion strategies, without providing guidelines on how to sample the HA data. This work addresses the problem of collecting HA data adaptively and sequentially so when it is integrated with the LA data a more accurate surrogate model is achieved. For this purpose, we propose an approach that takes advantage of the information provided by LA data as well as the previously selected HA data points and computes an improvement criterion over a design space to choose the next HA data point. The performance of the proposed method is evaluated, using both simulation and case studies. The results show the benefits of the proposed method in generating an accurate surrogate model when compared to three other benchmarks.

Keywords: Data fusion, adaptive sampling, expected improvement criterion

1 Introduction

Accurate modeling of a complex system requires exploration of a large design space. However, this exploration often requires a large number of high-accuracy (HA) simulations or experiments that are too costly or time-consuming to conduct. Alternatively, the model can be built based upon less accurate simulations or experiments that are faster and less costly to undertake and can provide a large number of data points. These approximate simulations, however, compromise the model accuracy and may introduce bias in the model. A practical approach is to fuse the LA data obtained

from crude but fast experiments with a few HA data points to remove the bias of the LA model and construct a reasonably accurate and cost-effective model (Kennedy & O'Hagan, 2000; Qian *et al.*, 2006; Qian & Wu, 2008; Xia *et al.*, 2011; Colosimo *et al.*, 2015). Nevertheless, the proposed data fusion strategies assume the availability of LA and HA data and do not provide any strategy for sampling expensive HA data using the information that the abundant LA data can provide. The goal of this paper is to propose an LA-data-led approach for collecting HA data to be fused with LA data for constructing a more accurate model. Therefore, this paper lies at the intersection of design of experiments and data fusion literature. Such LA-data-led sampling approach of HA data can be used in many applications. For example, in geometric inspection and metrology (Colosimo *et al.*, 2010; Xia *et al.*, 2011), measurements from less accurate metrology devices (e.g., structured light scanner) can lead the sampling trajectory of the HA data obtained from more accurate tools such as traditional Coordinate Measuring Machines (CMMs) that use contact touch probes.

In another example, reducing the number of tests performed on a combustion engine is one of the most important requirements to improve cost efficiency of the engine control units (ECU) modeling. The ECU includes models of different systems in a vehicle that are constructed based on a large number of physical tests performed at different levels of engine torque and speed. Often, these physical tests have already been performed on other engines with similar specifications when collecting data from a new engine, and can be considered as historical LA data. This LA data can guide the design values at which performing a test on the new engine can significantly improve the model. This intelligent sampling can significantly reduce the number of required experiments on the engine, benefiting the manufacturer both economically and environmentally.

In general, data fusion refers to the process of combining data obtained from different sources with the goal of achieving improved results over what could have been obtained from each source, separately (Colosimo *et al.*, 2015). A data fusion approach may consider one of the following settings: a) Integrating data obtained from computer codes with different levels of accuracy (Kennedy & O'Hagan, 2000), b) Integrating physical experiment data with the data obtained from a computer experiment (Kennedy & O'Hagan, 2001), and c) Integrating data collected from physical experiments with different levels of accuracy (Xia *et al.*, 2011). In this paper, we concentrate on a situation in which measurements contain noise and are collected from two sources, where one source is more accurate but more expensive than the other. We refer to the data from these two sources as low and

high-accuracy data. As a result, this paper is more concerned with the third setting, i.e., the fusion of physical experiments data (or stochastic computer experiment data). Nevertheless, our proposed approach can be modified to be used in the deterministic settings. This paper will not be concerned with the second setting, which is usually cast as a calibration of a computer model. In order to fuse data from LA and HA experiments, several approaches have been introduced in the literature. In the simplest approach, the two sets of data are directly merged together and are used to construct a surrogate model. For example, in metrology applications, Bradley & Chan (2001) used the CMM machine at the boundaries, where the laser scanner is less accurate, and combined all the data to construct a model for a surface. Co-kriging models have extensively been used to fuse LA and HA experiment data (Kennedy & O'Hagan, 2000; Forrester *et al.*, 2007). The co-kriging approach usually generates very accurate predictive models but are computationally expensive (Le Gratiet & Garnier, 2014). Another popular framework used for data fusion is the hierarchical approach. Reese *et al.* (2004) proposed a hierarchical method with linear models to integrate physical and computer experiments. Qian & Wu (2008) introduced Bayesian hierarchical Gaussian process (BHGP) models to fuse multiple sources of data. Their approach can be viewed as the Bayesian formulation of a co-kriging model. Xia *et al.* (2011) used a hierarchical Bayesian model to align and fuse CMM and laser scanner measurements. An alternative method that is tightly related to the Bayesian approaches uses a link function between the LA and HA observations. In fact, both of the co-kriging and hierarchical Bayesian employ a form of link function. Qian *et al.* (2006) proposed a two-step approach in which a surrogate model, referred to as a base model, is first fitted to the LA data. Next, the base model is corrected according to the HA data, using a link function that scales the base model linearly and captures the bias by a Gaussian process model. Xia *et al.* (2011) used a kernel regression link model and employed a hierarchical Bayesian approach for estimating model parameters.

The aforementioned methods focus on data fusion, assuming that the LA and HA experiments have already been conducted and the data is available. They, however, do not provide a strategy on how to collect few expensive HA points given LA data to achieve a more accurate model. Consequently, they may under or over sample from the HA design space that may result in a less accurate model or unnecessary sampling costs. Therefore, it would be essential to devise an adaptive approach that uses LA data to systematically explore the design space and guide the

sampling strategy for HA experiments. A large group of works in the field of computer design of experiments provides methods for sampling a set of points from a design space, either in one-step or sequentially, to construct an accurate model of a system. Space-filling designs such as Latin hypercube design (McKay *et al.*, 1979; Stein, 2012; Morris & Mitchell, 1995; Santner *et al.*, 2013), distance-based designs (Johnson *et al.*, 1990), uniform designs (Fang, 1980), and sequential versions of these designs (e.g., Sobol sequences) are effective approaches for initial exploration of the surface/model when no information is available. They, however, are constructed based on the assumption that the features of the true model are uniformly distributed across the design space. Criterion-based designs are obtained by minimizing/maximizing a statistical criterion such as mean square prediction error (MSPE) or entropy (Box & Draper, 1959; Shewry & Wynn, 1987; Sacks *et al.*, 1989a,b). These approaches can be converted into sequential designs by selecting the point that minimizes/maximizes the criterion at each step. The issue with these approaches is that they only capture the global behavior of the model using a correlation function but fail to concentrate on the areas of the design space where the model has high local variations. To capture both the global and local features, Lam (2008) introduced an expected improvement criterion for a global fit (EIGF). The EIGF at each point of a design space represents the amount of improvement (in the fitted model) that the corresponding point can introduce if added to the design. Unfortunately, none of the foregoing sequential design schemes take the fusion of HA and LA data into account. They only focus on constructing a design when no prior information (e.g., LA data) is available.

A relevant approach that considers the fusion of LA and HA computer experiment data is the sequential nested Latin hypercube design (LHD) proposed by Xiong *et al.* (2013). In this approach the HA computer experiments are conducted at locations sampled by small LHDs and the LA computer experiments are conducted at points selected by the large LHDs obtained by enlarging the small LHDs. The issue with this method is that when sampling the HA data, the information from the LA data is not used. Ezzat *et al.* (2017) proposed sampling the LA and HA data using separate designs to calibrate a computer model based on a physical experiment. Although this approach provides flexibility in terms of choice of designs, the LA information remains unused when sampling the HA data. Please note that, although these two works are different in scope (i.e., one focuses on the fusion of data obtained from computer experiments and the other focuses on a calibration problem), both ignore the information from the LA data when sampling the HA data.

This limitation is the main focus of this paper.

The main goal of this paper is then to introduce a sequential sampling method for HA data that utilizes LA data to explore the design space and adaptively identify the points where HA data can have the highest impact on improving the accuracy of the surrogate model constructed by fusion of LA data and the sampled HA data. Additionally, as the proposed method samples the design space sequentially, it can choose just enough HA data samples for a pre-specified model accuracy and hence reduce the sampling costs. Our method combines the EIGF experiment design approach with a data fusion method that employs a link function to adaptively select the HA data. Specifically, we first fit a Gaussian process (GP) to LA data to obtain a base surrogate model of the system. This model captures the global behavior of the system but lacks accuracy. To improve this model, we then select an initial set of HA points using a space filling approach. Next, we modify the EIGF criterion based upon both HA and LA data, and employ it to select the next HA points. Because the information is limited early on in the sampling procedure, we propose a simulated annealing procedure to explore more when data is not reliable, and exploit the sampled data as the sampling moves forward and more information becomes available.

The rest of the paper is organized as follows: In the next section, we review the fusion of the LA and HA data using a Gaussian process model. Section 3 describes the adaptive sampling approach and how it can be used for selecting HA data. Next, in Section 4, we evaluate the performance of the proposed method using two simulation studies. We also compare the proposed method to three benchmarks based on the model prediction error. Section 5 describes two case studies used to assess the proposed method in real-world applications. Finally, we summarize the paper in Section 7.

2 Review of the one-step data fusion model using a Gaussian process and link function

In this section, we outline the data fusion method for the situation that both LA and HA data are available. Next, we modify this framework to adaptively select and fuse the HA data. When data from both experiments is already available, the common practice is to follow a two-step approach for data integration that involves fitting a statistical model (usually a GP model) to the LA data to generate a base surrogate model. Next, this model is adjusted using the HA data to obtain the

final surrogate model. Before we describe these two steps, we provide a set of notations used in this paper.

In this paper, we consider two sources of data: one obtained from an LA but inexpensive experiment, and one that is acquired (or to be acquired) from HA, expensive experiments. We assume that a design value in both the HA and LA experiments consists of same p factors denoted by $\boldsymbol{v} = (v_1, \cdots, v_p)$. We also denote an experiment data measured at a design value $\boldsymbol{v}$ by $z(\boldsymbol{v})$. For instance, in the metrology example, $\boldsymbol{v} \in \mathbb{R}^2$ is the location of a point in a $x-y$ plane and $z(\boldsymbol{v})$ is the height of the product at that point. In the engine example, $\boldsymbol{v} \in \mathbb{R}^2$ is a vector consisting of torque and rpm, and $z(\boldsymbol{v})$ is the engine performance (e.g., air mass) at a particular design value. We denote an LA point with an index L and an HA point with an index H, i.e., $(\boldsymbol{v}_L, z_L)$ for an LA data and $(\boldsymbol{v}_H, z_H)$ for an HA point. When no subscript is used for a design value $\boldsymbol{v}$, the point is general and can be low or high-accuracy. We designate the design set of the LA experiments with M_L runs by $D_L = \{\boldsymbol{v}_{L1}, \boldsymbol{v}_{L2}, \cdots, \boldsymbol{v}_{LM_L}\}$ and the corresponding measurement data by $\boldsymbol{z}_L = [z_{L1}, z_{L2}, \cdots, z_{LM_L}]^{\mathsf{T}}$. Our goal is to find an HA design $D_H = \{\boldsymbol{v}_{H1}, \boldsymbol{v}_{H2}, \cdots, \boldsymbol{v}_{HM_H}\}$ with the corresponding HA measurements $\boldsymbol{z}_H = [z_{H1}, z_{H2}, \cdots, z_{HM_H}]^{\mathsf{T}}$; $M_H \ll M_L$ so that when integrated with $\boldsymbol{z}_L$ a more accurate surrogate model is obtained. That is, for an *unexplored* point $\boldsymbol{v}$, the constructed model returns a value that is almost equal to $z_H(\boldsymbol{v})$. Notice that we assume the HA experiment produces results that agree with the true system and therefore the error in the HA experiment data is negligible.

2.1 Fitting a Gaussian process to LA experiment data

The first step of data fusion is to construct a base surrogate model, using the LA data. Mathematically, this can be represented by

$$z_L = f_L(\boldsymbol{v}_L) + \epsilon_L, \tag{1}$$

where ϵ_L accounts for the error and is assumed to follow a normal distribution with mean zero and the variance σ_N^2, i.e., $\epsilon_L \sim \mathcal{N}(0, \sigma_N^2)$, and $f_L(\boldsymbol{v}_L)$ is the core part of the model that captures the features and patterns of the surrogate model. In the case of deterministic LA simulations, the term ϵ_L should be removed to reflect the fact that the data is not noisy. Among all possible models to represent $f_L(\boldsymbol{v}_L)$, a GP model is usually considered due to its flexibility and simplicity

of parameter estimation. A GP model is a random process in which a joint distribution of any $k < \infty$ observations of the process $\{z_{L1}, \cdots, z_{Lk}\}$ follows a multivariate Gaussian distribution. Such a GP model is denoted by $z_L(\boldsymbol{v}) \sim GP(m_L(\boldsymbol{v}_L), k_L(\boldsymbol{v}_L, \boldsymbol{w}_L))$, where $m_L(\boldsymbol{v}_L)$ denotes the mean function of the Gaussian process evaluated at $\boldsymbol{v}_L$ and $k_L(\boldsymbol{v}_L, \boldsymbol{w}_L)$ is the covariance function of the GP. The output of $k_L(\boldsymbol{v}_L, \boldsymbol{w}_L)$ is the covariance between two random variables $z_L(\boldsymbol{v}_L)$ and $z_L(\boldsymbol{w}_L)$ when $\boldsymbol{v}_L \neq \boldsymbol{w}_L$ and the variance of $z_L(\boldsymbol{v}_L)$, otherwise. In many applications, a constant or a linear function is an appropriate selection for the m_L. In this article, we consider a linear function, i.e., we set $m_L(\boldsymbol{v}) = \boldsymbol{\beta}^{\mathrm{T}}[1; \boldsymbol{v}]$, where $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$ is a set of parameters to be estimated. Several functional forms (kernels) have been introduced for covariance functions, such as exponential, squared exponential, and Matï¿œrn covariance functions that are different in terms of smoothness and differentiability (Rasmussen & Williams, 2006). For a stationary GP, the covariance function depends only on the distance between the two points. The distance is usually defined as $r(\boldsymbol{v}, \boldsymbol{w}) = \sum_{i=1}^{p} \theta_i |v_i - w_i|^{d_i}$, where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$ and $\boldsymbol{d} = (d_1, \cdots, d_p)$ are the scale and power parameters. When $d_i = 2; i = 1, \cdots, p$ the distance is Euclidean. Often, a covariance function is written as $k_L(\boldsymbol{v}, \boldsymbol{w}) = \sigma_L^2 h(r(\boldsymbol{v}, \boldsymbol{w}))$, where h is a correlation function, and σ_L^2 is the variance of data at any given point $\boldsymbol{v}$. For example, in the squared exponential kernel function $h(\boldsymbol{v}, \boldsymbol{w}) = \exp(-r(\boldsymbol{v}, \boldsymbol{w}; d_i = 2))$. For more information on the form and properties of the covariance functions see (Rasmussen & Williams, 2006; Lam, 2008). The hyper-parameters $\Theta = \{\boldsymbol{\beta}, \sigma_L, \theta\}$ of the base model are unknown and should be estimated to best fit the observed LA data $\{(\boldsymbol{v}_L, z_L)\}_{i=1}^{M_L}$. The estimation procedure is based on the likelihood maximization of a multivariate normal distribution (see Rasmussen & Williams 2006 and Qian *et al.*, 2006). Once the hyper-parameters of a GP model are estimated given the available data, the model can be used for prediction of z at any unexplored point $\boldsymbol{v}$. The empirical best linear unbiased predictor (BLUP) is usually adopted for this purpose (see Qian *et al.* (2006)) as follows:

$$\hat{z}_L(\boldsymbol{v}) = [1, \boldsymbol{v}]\hat{\boldsymbol{\beta}} + \boldsymbol{k}^{\mathrm{T}}(K + \sigma_N I_{M_L})^{-1}\left(\boldsymbol{z}_L - V\hat{\boldsymbol{\beta}}\right), \tag{2}$$

where $\hat{\boldsymbol{\beta}}$ is the vector of estimated parameters, σ_N is the noise effect due to the noise-contaminated measurements, $\boldsymbol{k}^{\mathrm{T}} = [k(\boldsymbol{v}, \boldsymbol{v}_{L1}), k(\boldsymbol{v}, \boldsymbol{v}_{L2}); \cdots, k(\boldsymbol{v}, \boldsymbol{v}_{LM_L})]$ is a vector whose elements are the covariance between $\boldsymbol{v}$ and all sampled points, $K \in \mathbb{R}^{M_L \times M_L}$ is a covariance matrix whose $(i,j)^{th}$

element is $k(\boldsymbol{v}_{Li}, \boldsymbol{v}_{Lj})$; $i, j = 1, \cdots, M_L$, and $V \in \mathbb{R}^{M_L \times (p+1)}$ is the regressor matrix whose i^{th} row is $[1, \boldsymbol{v}_{Li}]$. Note that when the data is deterministic σ_N is zero. The BLUP smoothly interpolates all the observed points to generate a prediction at a new point.

2.2 Fusion of the LA and HA data using a link model

In the second step, to integrate the LA and HA data and generate the final surrogate model, Qian *et al.* (2006) proposed the following link function:

$$z_H(\boldsymbol{v}) = \rho(\boldsymbol{v})\, \hat{z}_L(\boldsymbol{v}) + \delta(\boldsymbol{v}) + e,$$

where $z_H(\boldsymbol{v})$ is the observed HA value at point $\boldsymbol{v}$, $\hat{z}_L(\boldsymbol{v})$ is the base model prediction, $\delta(\boldsymbol{v})$ is a GP model that captures the bias, $\rho(\boldsymbol{v})$ is a linear function in $\boldsymbol{v}$, i.e., $\rho(\boldsymbol{v}) = \rho_0 + \sum_{i=1}^{p} \rho_i v_i$, and e is a random noise that follows a normal distribution. In the case of deterministic HA simulations, the term e should be removed to reflect the fact that the data is not noisy. Given a set of HA data, the goal is to estimate the parameters of $\delta(\boldsymbol{v})$, i.e., the mean and the covariance function hyper-parameters and $\boldsymbol{\rho} = \{\rho_0, \rho_1, \cdots, \rho_p\}$. The estimates can be computed by maximizing a likelihood function. For a detailed estimation procedure of these parameters refer to (Qian *et al.*, 2006). Next, given $\hat{\rho}_i$; $i = 1, \cdots, p$, one can compute $\boldsymbol{\delta} = [\delta(\boldsymbol{v}_{H1}), \cdots, \delta(\boldsymbol{v}_{HM_H})]$ by

$$\delta(\boldsymbol{v}_{Hi}) = z_H(\boldsymbol{v}_{Hi}) - \hat{\rho}(\boldsymbol{v}_{Hi})\, \hat{z}_L(\boldsymbol{v}_{Hi}).$$

Finally, at an unexplored point $\boldsymbol{v}$, the bias $\delta(\boldsymbol{v})$ can be predicted using a BLUP predictor. In this article, we select a constant value, δ_0, for the mean function. Therefore, the BLUP is constructed as

$$\hat{\delta}(\boldsymbol{v}) = \hat{\delta}_0 + \boldsymbol{k}^{\mathsf{T}} K^{-1} \left(\boldsymbol{\delta} - \hat{\delta}_0 \mathbf{1}_{M_H} \right),$$

where $\mathbf{1}_{M_H}$ is a vector of size M_H whose elements are all one. At a given point $\boldsymbol{v}$ the prediction of the final surrogate model is given by $\hat{z}(\boldsymbol{v}) = \hat{\rho}(\boldsymbol{v})\, \hat{z}_L(\boldsymbol{v}) + \hat{\delta}(\boldsymbol{v})$. Unless it is mentioned otherwise, a Matï¿œrn covariance function with the scale parameter $\nu = \frac{3}{2}$ (see (Rasmussen & Williams, 2006)) is used for $\delta(.)$

Note that the major assumption for this model is that the LA data provides sufficient information

about the overall geometric shape of the true surface, otherwise discarding the LA data and using only HA data may provide a better surrogate model.

3 Adaptive data fusion using a modified EIGF criterion

Instead of assuming that the HA data is already available, this article aims to sample HA points sequentially to be integrated with the available LA data in order to construct a more accurate final surrogate model. For this purpose, we first describe an expected improvement criterion for a global fit, introduced by Lam (2008). Then, we modify this criterion so that it takes the LA data into account when selecting an HA point and ultimately constructing an HA design.

3.1 Expected improvement for a global fit

An expected improvement (EI) for computing the maximum of a black-box function was first proposed by Schonlau (1997). Lam (2008) extends the EI criterion to a global fit setting, where the goal is to find a design that leads to a better global fit rather than the maximum point. Let $Z(.)$ follow a GP model denoted by $GP(m_Z(.), k_Z(.,.))$, i.e., $Z(\boldsymbol{v})$ is a random variable that follows a normal distribution with mean $m_Z(\boldsymbol{v})$ and the variance $Var(Z(v)) = k_Z(\boldsymbol{v}, \boldsymbol{v})$. Then, the improvement of a point $\boldsymbol{v}$ is defined as

$$I(\boldsymbol{v}) = (Z(\boldsymbol{v}) - z(\boldsymbol{v}_c))^2,$$

where $\boldsymbol{v}_c$ is the nearest observed design value to $\boldsymbol{v}$, and $z(\boldsymbol{v}_c)$ is its corresponding data. Note that one may consider $I(\boldsymbol{v}) = \frac{(Z(\boldsymbol{v}) - Z(\boldsymbol{v}_c))^2}{||\boldsymbol{v} - \boldsymbol{v}_c||^2}$ as an improvement criterion, which effectively gives more weight to the points that are closer to $\boldsymbol{v}_c$. But, due to the major assumption of the Gaussian process that the closer the points the more similar they are, we would like to select a point further away from $\boldsymbol{v}_c$ to obtain more information. Therefore, a normalized $I(\boldsymbol{v})$ may not be a good choice. Because $Z(\boldsymbol{v})$ is a random variable, the expected improvement for a global fit (EIGF) should be considered and is defined as

$$EIGF(\boldsymbol{v}) = E\left((Z(\boldsymbol{v}) - z(\boldsymbol{v}_c))^2\right) = Var(Z(\boldsymbol{v})) + [\hat{z}(\boldsymbol{v}) - z(\boldsymbol{v}_c)]^2,$$

where $\hat{z}(\boldsymbol{v})$ is the prediction value of the GP model at point $\boldsymbol{v}$. In the EIGF equation, a large value of $Var(Z(\boldsymbol{v}))$ indicates that the model is uncertain about its prediction at $\boldsymbol{v}$, and therefore including $\boldsymbol{v}$ in the design can significantly improve the model. Similarly, a large value of $[\hat{z}(\boldsymbol{v}) - z(\boldsymbol{v}_c)]^2$ indicates a high local variation around $\boldsymbol{v}$ and therefore selection of $\boldsymbol{v}$ in the design may help the model in capturing the local behaviors. As a result, in theory, the EIGF captures both the global and local behavior of the model based on the variance at each point, and by exploiting the observed neighboring points. To obtain a design D_H, one can use the EIGF criterion sequentially. That is, at each step the maximizer of the EIGF over the design space is sampled until a pre-specified design size is reached.

The main issue with the EIGF criterion is that, with no prior knowledge of the surface and with a small design size, $[\hat{z}(\boldsymbol{v}) - z(\boldsymbol{v}_c)]^2$ may not capture the local variations because the prediction value at a point v is an interpolation of its neighboring points (obtained from the BLUP), and therefore it is likely that $[\hat{z}(\boldsymbol{v}) - z(\boldsymbol{v}_c)]^2$ becomes ineffective. Therefore, given the small size of HA experiments, an adaptive sampling approach based on EIGF that only relies on HA data may result in an inaccurate surrogate model. However, when LA data is available, the information from the base surrogate model can be exploited as the prior knowledge to resolve this problem. Beside this issue, the selection of the EIGF maximizer at each step of the sequential sampling, especially at the early steps when only a few data points are explored, may result in excessively greedy choices. To alleviate this problem, an exploration-exploitation routine can be considered.

In the next section, we introduce a modified version of EIGF that computes the expected improvement of a point by considering not only the explored HA neighboring points, but also the available LA data. Furthermore, we design an exploration and exploitation framework that uses the modified EIGF probabilistically to sample the next HA point.

3.2 Modified EIGF and data fusion

In this section, we propose an approach for sampling HA points, considering the information of the LA data. For this purpose, we modify the EIGF criterion so that it takes the LA data into account when computing the expected improvement of a design value. The underlying assumption of the proposed method is that the base surrogate model, constructed using the LA data, captures some of the local features of the system, and therefore can lead the sampling path to the areas of the

design space with higher local variation. This will help increase the number of samples collected in these local areas and, consequently, results in a more accurate model. In order to achieve this goal, we define the fused expected improvement for a global fit and data fusion (FEIGF) by considering the link function between the HA and LA, $Z_H(\boldsymbol{v}_H) = \rho(\boldsymbol{v}_H)\,\hat{z}_L(\boldsymbol{v}_H) + \delta(\boldsymbol{v}_H)$, as follows:

$$\begin{aligned} FEIGF(\boldsymbol{v}_H) &= E\left((Z_H(\boldsymbol{v}_H) - z_H(\boldsymbol{v}_{Hc}))^2\right) \\ &= E\left((\rho(\boldsymbol{v}_H)\,\hat{z}_L(\boldsymbol{v}_H) + \delta(\boldsymbol{v}_H) - \rho(\boldsymbol{v}_{Hc})\,\hat{z}_L(\boldsymbol{v}_{Hc}) - \delta(\boldsymbol{v}_{Hc}))^2\right) \\ &= (\rho(\boldsymbol{v}_H)\,\hat{z}_L(\boldsymbol{v}_H) - \rho(\boldsymbol{v}_H)\,\hat{z}_L(\boldsymbol{v}_{Hc}))^2 + Var(\delta(\boldsymbol{v}_H)) + \left(\hat{\delta}(\boldsymbol{v}_H) - \delta(\boldsymbol{v}_{Hc})\right)^2, \end{aligned}$$

where $\boldsymbol{v}_H$ is a design value that may be considered for an HA experiment and $\boldsymbol{v}_{Hc}$ is the nearest explored point to $\boldsymbol{v}_H$. In the FEIGF equation, $\rho(.)$ and $\delta(.)$ are assumed to be known or are estimated based on previously observed data (e.g., data from previous sampling steps). Furthermore, $\hat{\delta}(\boldsymbol{v}_H)$ denotes the prediction of the bias GP model, $\delta(\boldsymbol{v}_{Hc})$ represents the bias value computed based on the HA measurement and is assumed to be known, and $\hat{z}_L(.)$ denotes the prediction of the base surrogate model at a given point. The FEIGF consists of three terms: The first term $(\rho(\boldsymbol{v}_H)\,\hat{z}_L(\boldsymbol{v}_H) - \rho(\boldsymbol{v}_H)\,\hat{z}_L(\boldsymbol{v}_{Hc}))^2$, measures the local behavior of the surface as it is revealed by the LA model. The large value of this term at a design value $\boldsymbol{v}_H$ signals high variations around that point and therefore, including it in the design can improve the accuracy of the final model. The second term, $Var(\delta(\boldsymbol{v}_H))$, captures the prediction uncertainty of the bias model, δ, at $\boldsymbol{v}_H$. When $Var(\delta(\boldsymbol{v}_H))$ is high, including $\boldsymbol{v}_H$ in the design may improve the predictions of the bias model. The third term, $\left(\hat{\delta}(\boldsymbol{v}_H) - \delta(\boldsymbol{v}_{Hc})\right)^2$, quantifies the local variations of the bias around $\boldsymbol{v}_H$, and therefore, including this point in the design may result in more accurate bias predictions.

In order to adaptively and sequentially sample an HA data point using the FEIGF criterion, we first consider a dense grid, $\mathcal{G}$, over the design space and at each step select the design value that for example maximizes the FEIGF criterion over such a grid. In other words, at the sampling step t (i.e., when $t-1$ HA points are already collected), we first estimate the GP model of bias (δ_{t-1}) and the scaling parameter (ρ_{t-1}) based on the available LA points and the HA data points sampled up to step $t-1$, using the approach explained in section 2.2. Next, given the GP model of bias and the scaling parameter, we calculate the $FEIGF$ over the grid $\mathcal{G}$. Finally to select a point to be added to the design, we may consider the point that maximizes $FEIGF$ criterion. That is, we

may sample $\boldsymbol{v}_t = \arg\max_{\boldsymbol{v}\in\mathcal{G}} \left(FEIGF\left(\boldsymbol{v}\right)|\delta_{t-1}, \rho_{t-1}\right)$. However, the problem with sampling the point with maximum $FEIGF$ is that at the early stages of HA sampling when the number of HA points is small, relying on the maximum value of FEIGF may result in excessively greedy choices. The reason is that when the GP estimation is based on only few points the prediction variance (i.e., the second term in FEIGF) is large for many points, increasing the influence of the third term in separating and selecting the HA point. The third term, however, is inaccurate when the bias δ is estimated based on only few points, and results in a poor choice of a design value. To alleviate this problem, we adopt a probabilistic sampling approach in collecting the points so that early on in the process it allows selecting points with smaller-than-maximum FEIGF value (exploration) and as more information becomes available, it high-likely samples design values that maximize the FEIGF criterion. For this purpose, we define the following probability distribution parameterized by α_t over the $FEIGF\left(\boldsymbol{v}\right)$:

$$p_{\alpha_t}\left(FEIGF\left(\boldsymbol{v}\right)\right) = \frac{\exp\left(\alpha_t FEIGF\left(\boldsymbol{v}\right)\right)}{\int \exp\left(\alpha_t FEIGF\left(\boldsymbol{v}\right)\right) dv},$$

where α_t is an increasing sequence that tends to zero as $t \to 0$ and tends to ∞ as $t \to \infty$. In an extreme case, when $\alpha_t = 0$, the distribution $p_{\alpha_t}\left(FEIGF\left(\boldsymbol{v}\right)\right)$ is a uniform distribution and as $\alpha_t \to \infty$, the above distribution places all the probability mass over the $\max\left(FEIGF\left(\boldsymbol{v}\right)\right)$. In this paper, we select $\alpha_t = \frac{t}{M_H}$ at the sampling step t. We randomly sample from $p_{\alpha_t}\left(FEIGF\left(\boldsymbol{v}\right)\right)$ at each step t to select the HA data point. Observe that early in the sampling process, we allow selecting points that have smaller values of improvement to explore the design space, and as more information becomes available, we select points with larger FEIGF. Note that, to compute the probability distribution, we approximate the continuous distribution by a discrete distribution in which the probability mass is located at each point of the grid $\mathcal{G}$, where we calculated the FEIGF. Finally note that, one can sample a batch of points, instead of one, at each sampling process step. However, in our experience, when the batch size is small comparing to the total sampling budget, it will not significantly influence the results. Figure 1 illustrates an overview of the proposed adaptive data fusion approach.

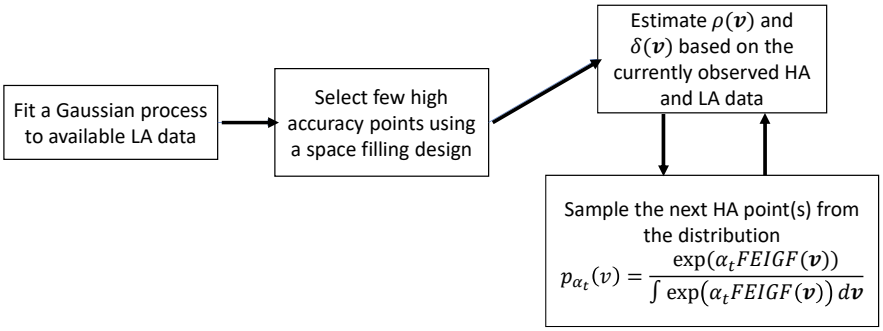


Figure 1: Overview of the proposed adaptive data fusion approach

4 Performance evaluation using simulation study

In this section, we conduct simulations to evaluate the performance of the proposed adaptive data fusion technique. We compare the proposed method, labeled as *adaptive with fusion*, with three other benchmarks: In the first benchmark, we consider a one-step design in which M_H data points from HA experiments are sampled using an LHD and are directly used to construct the final model. We label this benchmark as *one-step* approach. The second benchmark, designated as *one-step-fused*, selects M_H data points from HA experiments using an LHD and fuses them with the available LA data through a link model described in Section 2.2. Finally, the third benchmark adaptively samples the HA data using the EIGF without integrating the sampled data with the available LA data. We label this method *adaptive w/o fusion*. Note that we used the standard LHD which is not based on optimizing any criterion. We compare all the methods in terms of prediction accuracy using standard mean square error, i.e.,

$$SMSE = \frac{\sum_{i=1}^{M} (z_i - \hat{z}_i)^2}{\sum_{i=1}^{M} z_i^2},$$

where z_i is the true system value, $\hat{z}_i$ is the predicted value of the final surrogate model, and M is the total number of points in a grid over which we sample points. In order to estimate the hyper-parameters of a GP model, we use a MATLAB package developed by Rasmussen & Williams (2006).

One-dimensional surface simulation: We first consider a one-dimensional non-stationary func-

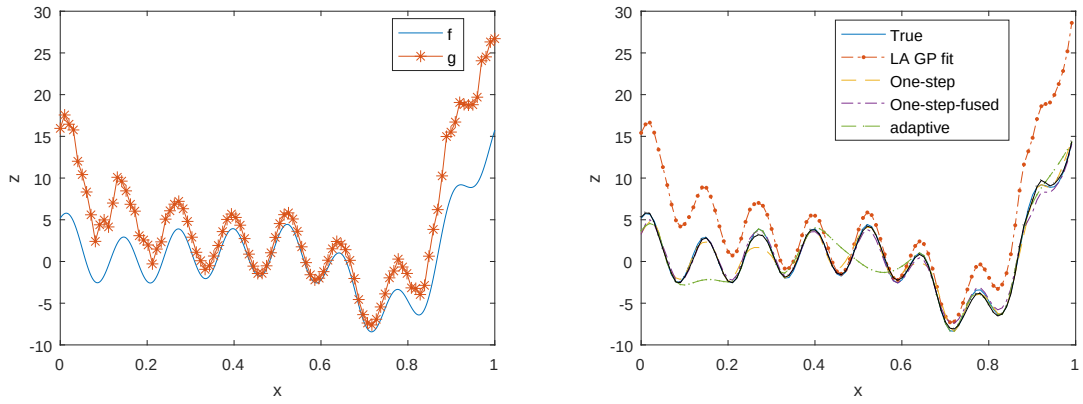
tion

$$z_H = f(v) = (6v-2)^2 \sin(12v-4) + 3\cos(50v-1) + \cos(4v-1)$$

as the true function for $v \in [0,1]$. Next, we generate the LA data from a shifted and scaled version of the function f defined as

$$z_L = g(v) = Af(v) + B(v-0.5)^2 + C + \epsilon(v),$$

where, $A = 1.2$, $B = 40$, and $C = 0.5$, and $\epsilon(v)$ is a normally distributed error with mean 0 and standard deviation $\tau(v) = 4(v-0.5)^2$, i.e., $\epsilon(v) \sim \mathcal{N}\left(0, \tau(v)^2\right)$. Notice that z_L is designed to have larger bias and noise near the boundaries as it is the case in many real applications (e.g., in metrology). Figure 2a illustrates the true curve along with the LA data collected over an equidistant grid of size 100. Figure 2b illustrates the true curve, the fitted GP to the LA data and the fitted curves obtained by one-step and one-step-fused benchmarks as well as the adaptive with and without fusion methods. The initial number of points for the adaptive approaches is $M_{H,init} = 5$ and the budget, M_H, is 25 for all methods. As described, the one-step benchmarks sample the 25 points at once using an LHD, whereas the adaptive methods first sample $M_{H,init} = 5$ design values by an LHD, and then sequentially sample points from a grid of size $M = 1000$. As illustrated, the predictions obtained from the proposed method and the one-step-fused benchmark are fairly accurate, indicating the advantage of the data fusion when the LA data captures some of the local and global behaviors of the surface. Between these two methods, the proposed method performs slightly better where local variations exist and the LA data is more biased (e.g., in $[0, 0.1]$ and $[0.9, 1]$ intervals). The adaptive w/o fusion selects several points on the interval $[0.8, 1]$ where the function has a large slope and generates an accurate prediction. However, it fails to capture the local variations in the intervals [0.1, 0.2] and [0.4, 0.6] because $\hat{z}(\boldsymbol{v})$ is an interpolation of its neighbors, which can be located far apart when the sample size is small. As a result, the adaptive without fusion does not perform well in this scenario. The one-step approach performs fairly well but fails to capture some of the local variations due to a lack of prior information. Such information is provided by the LA data when fusion is performed.



(a) (b)

Figure 2: (a) Illustration of the true curve and LA data in one dimensional simulation; (b) Example of the fitted curves in comparison to the true curve

To further investigate and compare the performance of the proposed method with the benchmarks, we calculate the SMSE values after sampling of each point in adaptive approaches and compare them to the SMSEs obtained by one-step methods. Because our approach is probabilistic in nature, we consider 100 replications of the simulation and take the average of SMSE values over these replications. For all these simulations, we take the initial number of points to be 20% of the budget, M_H. Figure 3 illustrates the performance of proposed method in comparison to all benchmarks at different values of M_H. For a better visualization, we illustrate the logarithm of SMSE in these figures. For the one-step methods the SMSE represents the error of the fit when M_H points are all selected at once. Observe that in all cases the adaptive approach outperforms the other methods in terms of SMSE. First, with the same number of samples the proposed method produces more accurate predictions. For instance, when $M_H = 20$, the logarithm of the SMSE of the proposed method is about -5, where it is about -4.2 and -2 for one-step-fused and one-step benchmarks, respectively. Second, it reaches the same level of the prediction error with fewer number of samples than the benchmarks. For example, in Figure 3c and 3d the adaptive approach with fusion obtains the same level of SMSE with about 15 and 20 samples rather than 25 and 30 used by other methods.

In all cases, the SMSE of the adaptive approach drops rapidly at the beginning of the sampling process. The reason is that early samples provide information about the overall behavior of the bias (i.e., they introduce large improvements), but late samples mostly introduce detailed information

and provide small improvements. Therefore, after some sampling steps the improvement in SMSE becomes very slow. In practice, this step can be considered as a stopping criterion. For example, in Figure 3d after sampling 26 points, no significant improvement can be observed.

Two-dimensional surface simulation: As a two-dimensional example, we consider a six-hump camel surface proposed by Branin (1972) as the true surface:

$$z_H = f(v_1, v_2) = 4v_1^2 - 2.1v_1^4 + \frac{1}{3}v_1^6 + v_1v_2 - 4v_2^2 + 4v_2^4$$

with $v_1 \in [-2, 2]$ and $v_2 \in [-1, 1]$. We scale and shift the true function to construct an LA data generator as

$$z_L = g(\boldsymbol{v}) = Af(\boldsymbol{v}) + B_1(v_1 - 0.5) + B_2(v_2 + 0.5) + C + \epsilon,$$

where $A = 0.5$, $B_1 = 1$, $B_2 = 2$, and $C = 2$, and ϵ is a normally distributed error with mean 0 and standard deviation τ, i.e., $\epsilon \sim \mathcal{N}(0, \tau^2)$. Figure 4 illustrates the f and g functions evaluated over an equidistant grid of 50×50 with $\tau = 0.2$. We evaluate the performance of the proposed method in comparison to all benchmarks at different values of M_H based upon the SMSE values calculated over an equidistant grid of 50×50 over the design space. Similar to the previous simulation study, the SMSE of the one-step methods represents the prediction error of the final model generated by M_H points that are selected at once using an LHD. Figure 5 illustrates the SMSE of the proposed method versus the benchmarks. Because our approach is probabilistic in nature, we consider 100 replications of the simulation and take the average of SMSE values over these replications. As illustrated, in all cases the adaptive approach results in more accurate predictions. For instance, when $M_H = 80$, the logarithm of the SMSE is about -5.5 that outperforms the one-step-fused benchmark (-5), and the other two benchmarks (-3.5). Furthermore, the proposed method can achieve the same level of prediction error with fewer number of samples than the benchmarks. For example, in Figure 5d the adaptive approach with fusion obtained the same level of SMSE with about 65 samples rather than 100 that is used by other methods. In addition, the initial SMSE of the proposed method is always smaller than the initial SMSE of the adaptive w/o fusion benchmark. This superiority is due to the information provided by the LA data. Finally, it is worth mentioning the two small jumps in Figures

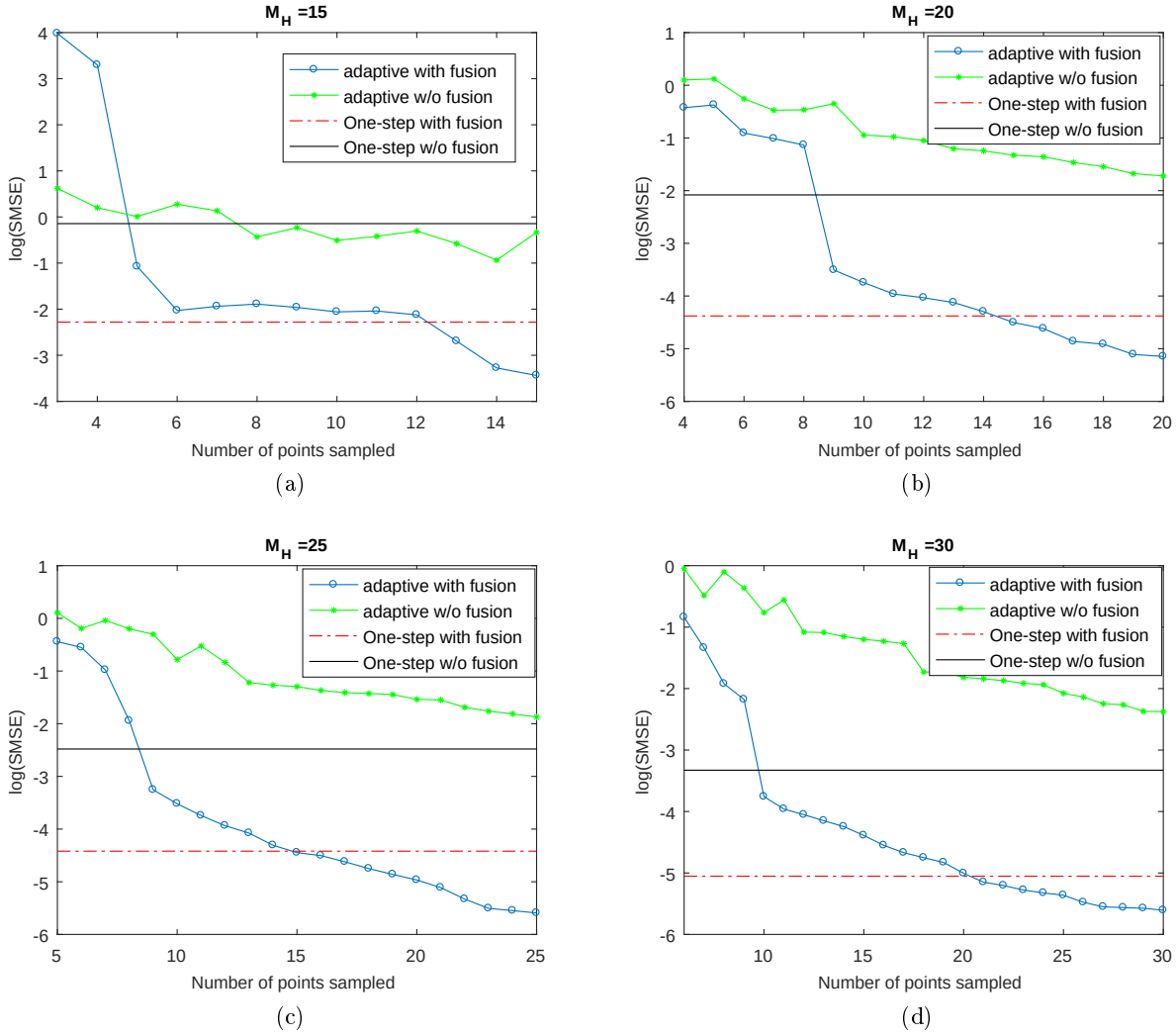


Figure 3: Comparison of the proposed method to benchmarks in simulation study I at different values of budget M_H. Observe that the adaptive approach reaches the same level of error as benchmarks with much fewer samples. For example in (c) the proposed approach reaches the same level of error of one-step-fused with only 15 sampled points.

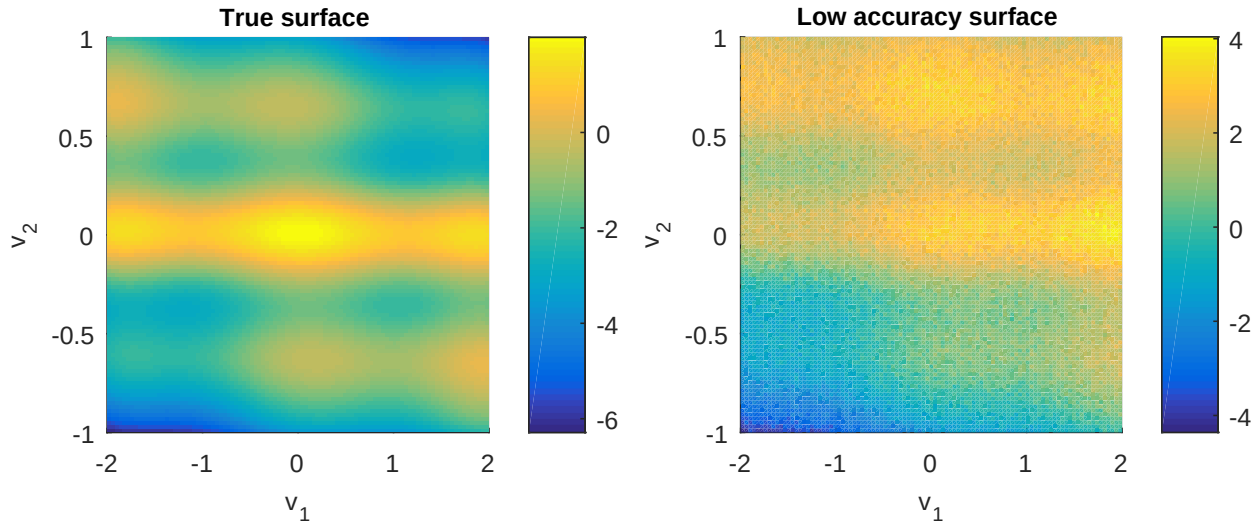


Figure 4: Illustration of the true and biased surfaces used for two dimensional simulation

5a and 5b. The reason for these jumps is the computational instabilities that may occasionally occur when fitting a GP model. Such instabilities are related to the likelihood optimization procedure that may be trapped in a local maxima. To avoid (or alleviate) these situations, we incorporate a simulated annealing heuristic for function optimization (Kirkpatrick *et al.*, 1983) in our procedure.

5 Case Study

In this section, we consider two case studies to evaluate the performance of the proposed method in real-world applications. First, we consider a metrology example in which a set of HA metrology data are collected using a CMM to improve the LA measurements collected by a structured laser (SL) scanner for reconstructing a free-form surface. In the second case study, we are interested in constructing a surrogate model that predicts the air mass of an engine at a specific torque and revolution speed by performing few new HA experiments on the engine and combining it with previously collected data from another similar engine.

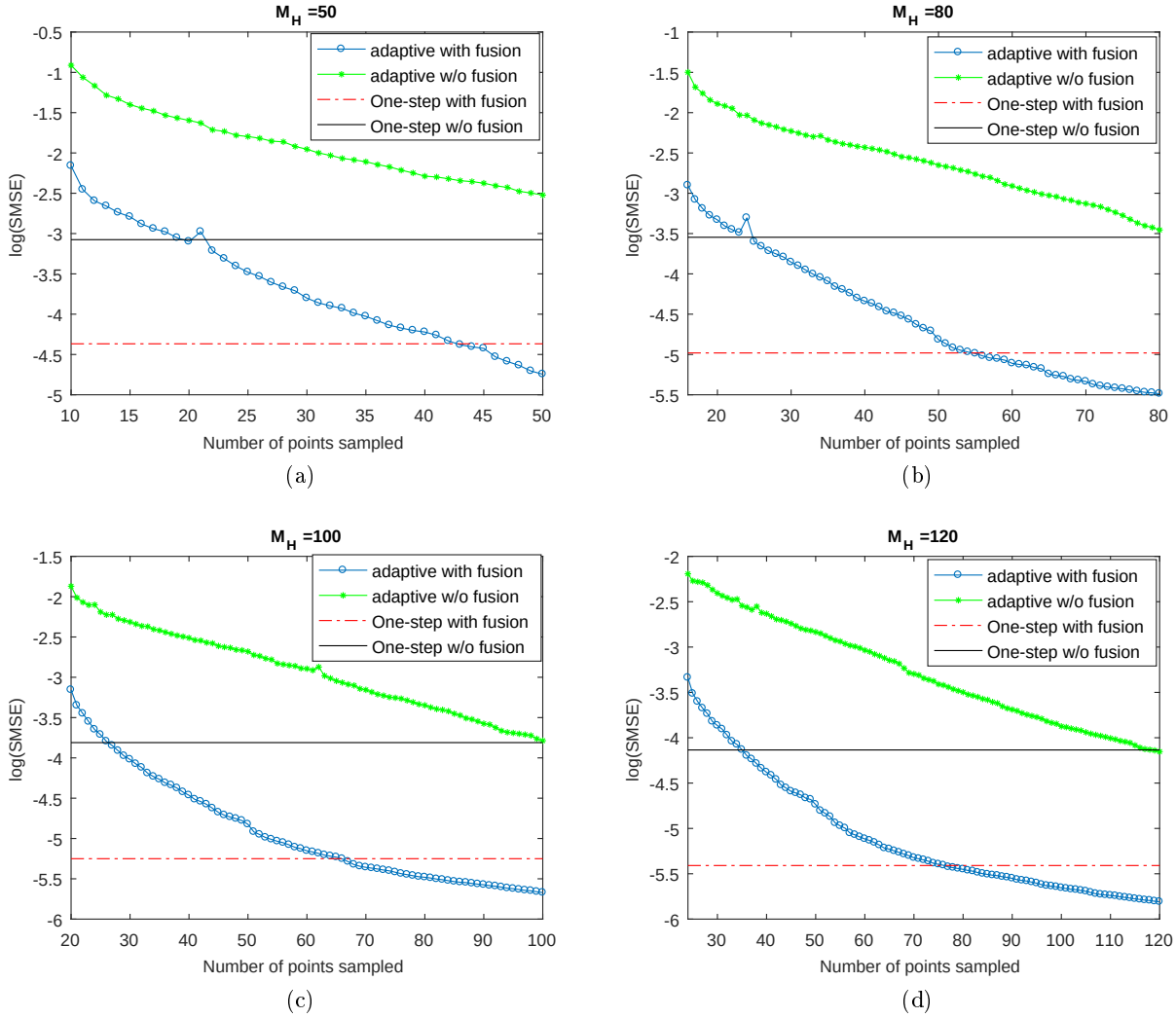


Figure 5: Comparison of the proposed method to benchmarks in simulation study II at different values of budget M_H. Note that the adaptive approach can obtain the same level of error as benchmarks with fewer samples. For example in (c), the proposed approach achieves the same level of error as one-step with fusion benchmark with only 65 sampled points rather than 100.

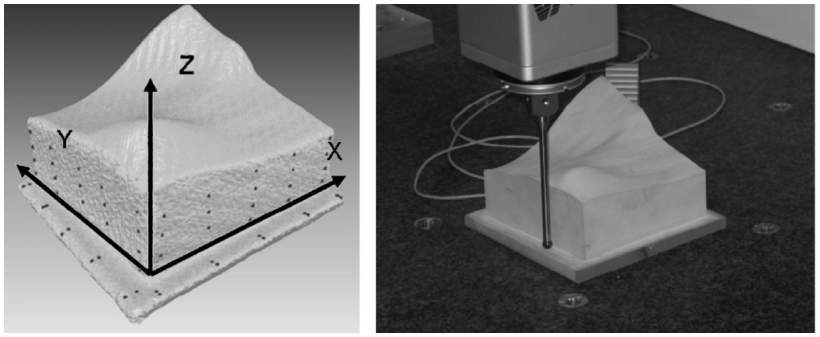
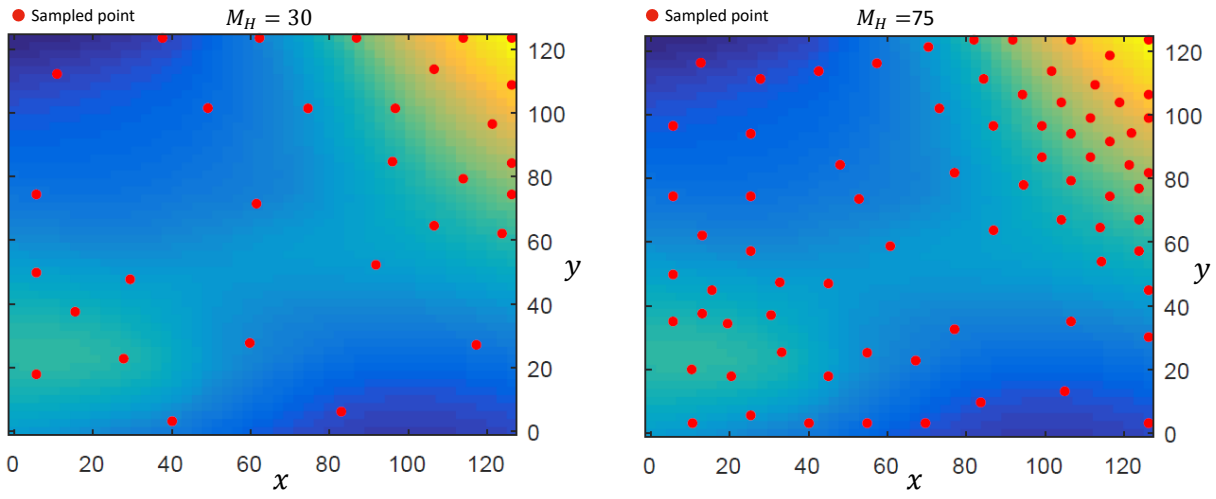


Figure 6: The free-form surface used for the case study (Colosimo *et al.*, 2015). Three orthogonal reference surfaces (left panel). CMM sampling (right panel).

5.1 Freeform surface metrology

The freeform surface, illustrated in Figure 6, has been used for this case study. The freeform surface is machined from a $100\,mm \times 100\,mm \times 100\,mm$ workpiece. The surface is measured, using both SL scanner and a Coordinate Measuring Machine (CMM) Zeiss "Prismo 5 HTG VAST" equipped with an analog probe head with maximum probing error of $MPEP = 2\,\mu m$ (according to ISO 10360-2). For more details on the measurement procedure refer to (Colosimo *et al.*, 2015). The point cloud obtained from each machine (SL scanner and CMM) contains 9635 points. The measurements from the SL scanner is considered as the LA data points, and the available data from the CMM is used to construct an emulator of the CMM by fitting a GP model. The goal is to collect a set of HA data from the CMM so when combined with the LA data points a more accurate final surrogate model is achieved.

We use our proposed method to construct a final surrogate model of the surface, and will compare its performance in reconstructing the freeform surface to other benchmarks, i.e., one-step, one-step-fused, and adaptive w/o fusion at different levels of budget M_H. In the case of adaptive approaches, we choose the initial sample size to be 20% of the budget. In order to estimate the link model parameters, we consider the squared exponential covariance function. Before discussing the prediction performance of surrogate models obtained by the proposed method and the benchmarks, it is interesting to look into the location of the points selected by the proposed adaptive approach. Figure 7 illustrates the location of the selected points for $M_H = 30$ and 75. As illustrated, more points are selected on the upper right corner and at the lower left corner where more local variations exist. For example, several points are sampled around the boundary of the hump on the lower left



(a) (b)

Figure 7: Location of the points selected by the proposed method for (a) $M_H = 30$ and (b) $M_H = 75$

area. This indicates the capability of our approach in selecting points at the locations with higher variability. Figure 8 illustrates the logarithm of SMSE of each method at different levels of budget. As can be seen from the figures and similar to the simulation results, the proposed method can achieve the same level of SMSE as other approaches, particularly the one-step-fused approach, with much smaller sampled points. For example, when $M_H = 70$, the proposed approach obtains same SMSE as the one-step-fused approach with about 60% of the budget. Observe that the initial SMSE of the proposed method is smaller than both the one-step and adaptive w/o fusion benchmarks. The reason is the good accuracy of the LA data that results in the base surrogate model with small errors. The base surrogate model provides significant information about the surface even without any (or with very few) HA data points available.

5.2 ECU calibration

In a modern vehicle, the engine control unit (ECU) ensures the functionality of the vehicle and diagnoses failure for a number of components. The ECU implements surrogate models of complex physical dynamical systems that are constructed based on a large number of tests performed at different levels of engine torque and speed. To reduce the experiment cost, exploiting the performance measurements of another engine (LA data) with similar specifications (e.g., with the same number of cylinder and displacement) may help identify few design values at which the engine should be

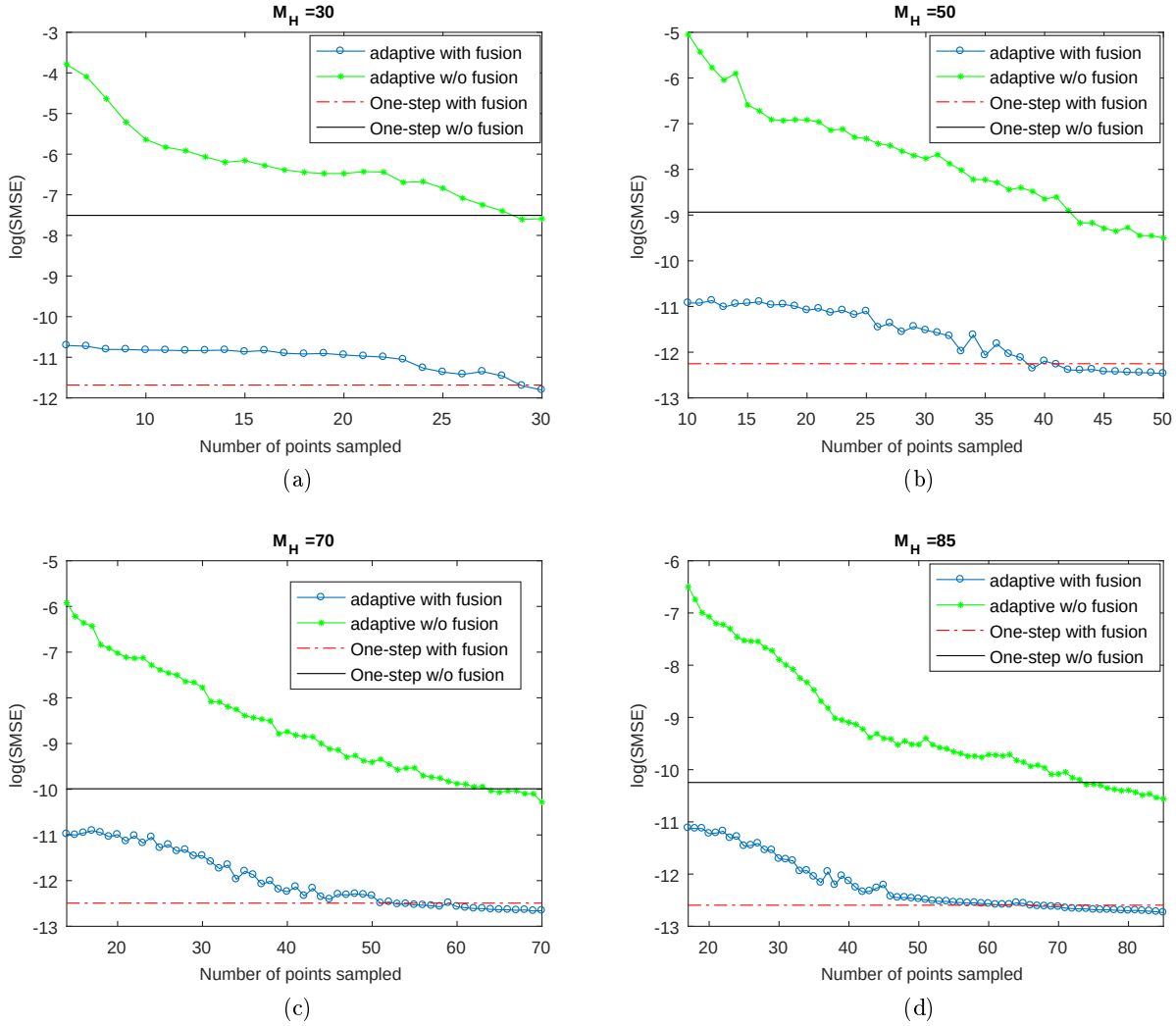


Figure 8: Comparison of the proposed method to benchmarks in case study I at different values of budget M_H. Note that the adaptive approach requires fewer samples to obtain the same level of error as benchmarks. For example, in (b) the proposed approach obtains the same SMSE as the one-step-fused approach with about 75% of the budget.

evaluated. The data collected from these tests (HA data) can then be fused to the LA data to construct a surrogate model. For example, for a new engine, the air mass of the combustion process should be measured at different combinations of engine torque and revolution speed (rpm) to ensure the right performance of the engine. Performing these tests over a large number of rpm-torque combinations is not economical. Historical data obtained from a similar engine may illuminate few points where the air mass should be measured.

For this purpose, we consider two diesel engines, both with the displacement of $2000\,cc$ and four cylinders. In ECU calibration of a diesel engine a typical task is to optimize the air mass of the combustion process due to the different engine states, namely different combinations of engine revolutions (rpm) and torque. We designate the engines by E_1 and E_2. We assume that the data from E_1 has already been collected through a series of experiments and are interested in integrating this data with a few experimental data obtained from E_2 to generate a surrogate model that predicts the performance of E_2 at various design values. Therefore, the data in this case study is the air mass measured over a two dimensional space of rpm and torque.

The available data for both engines E_1 and E_2 contains 256 measurements at different combinations of rpm-torque. In this case study, we first generate an emulator of each engine E_i; $i = 1, 2$ by fitting a GP model with a constant mean and piecewise cubic covariance (PCC) function to the available data. These emulators are then used to replicate the performance of the engines at any given design value. Figure 9 illustrates the predictions of both E_1 and E_2 emulators over a grid of size 50×50. Note that the emulator of E_1 can be considered as the base surrogate model, i.e., a GP model fitted to the LA data, but the emulator of E_2 is used as an HA simulator that produces the result of an experiment on E_2 at any pair of rpm and torque.

We compare the accuracy of the final surrogate models generated by the proposed method and the benchmarks for different budgets. For this purpose we calculate the SMSE over an equidistant grid of size 100×100. Figure 10 illustrates the logarithm of SMSEs of each method at $M_h = 15, 25$. As illustrated, the final surrogate model constructed by the proposed method achieves lower prediction errors. Furthermore, the proposed approach can obtain the same level of accuracy with much fewer sampled points when compared to the benchmarks. For example when $M_H = 25$, the proposed approach obtains the same level of accuracy as the one-step-fused approach with about 80% of the budget.

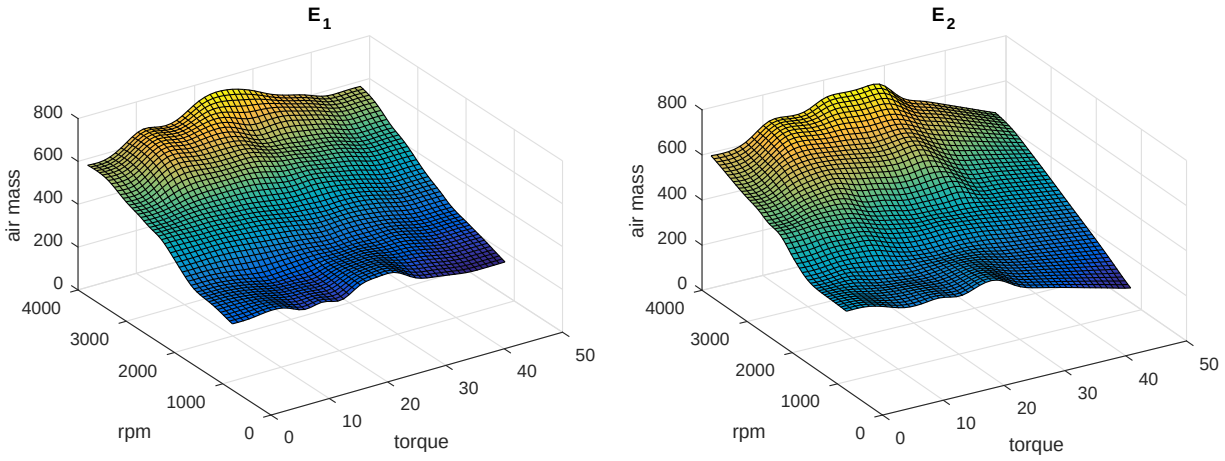
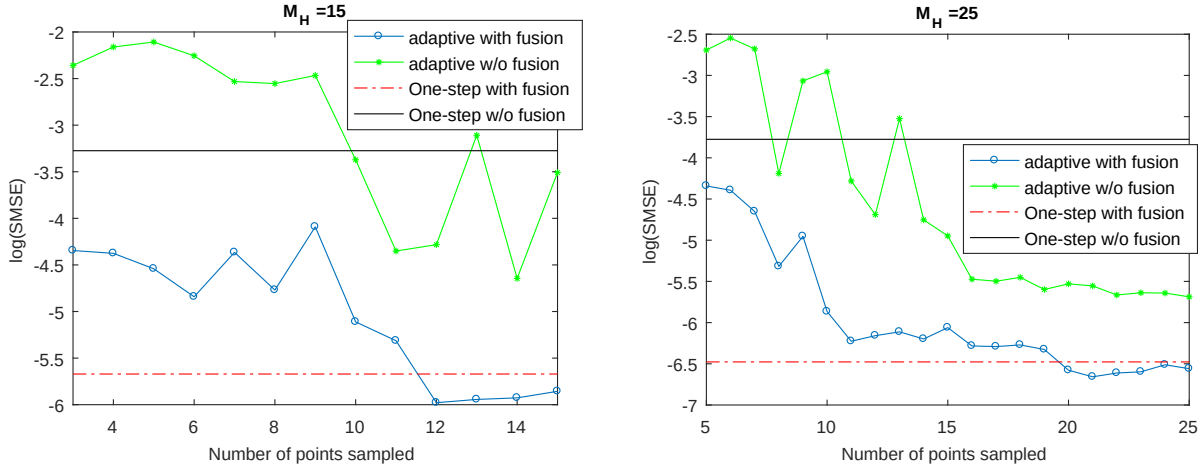


Figure 9: Prediction of the engines air mass at different values of rpm-torque pairs obtained by the emulators.



(a) (b)

Figure 10: Comparison of the proposed method to benchmarks in case study II for different values of budget M_H. Observe that in both situations, the adaptive approach can reach the same level of error as benchmarks with much fewer samples. For example, in (b) the proposed approach obtains the same SMSE as the one-step-fused approach with about 80% of the budget.

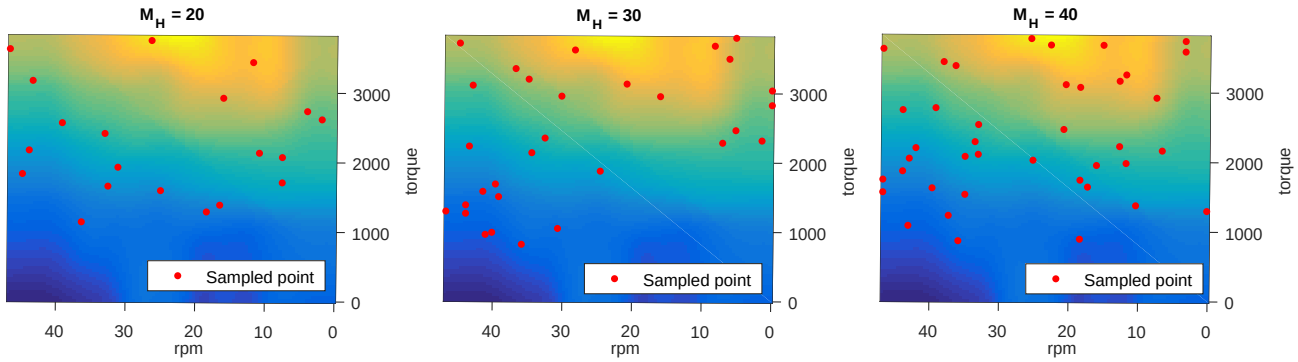


Figure 11: Example of the distribution of points included in the design for different budgets.

Examples of the distribution of the points included in the design for different budgets are illustrated in Figure 11. Note that in all three cases, the points are mostly selected at the locations with higher local variabilities, especially the upper part of the surface. This is mainly because the proposed method can identify the locations with higher local variation and perform exploration at those areas.

6 Discussion

This paper mainly assumes that the LA data is already available and utilizes the information from such data to sample HA data. The available LA data could have been collected using any design of experiment approach such as LHD and criterion-based designs. The design of experiment method by which the LA data is collected may or may not have significant influence on the final accuracy of the model. For example, in the situation, in which obtaining LA data is very inexpensive and a large set of data that effectively covers the sampling space can be collected, any legitimate design of experiment may provide sufficient information about the LA design space. However, when the size of LA data is not very large, a design that better captures the design space, by for example an adaptive approach, can be beneficial.

Given the lowest accuracy data is available, the proposed method may be extended to the situations in which data can be collected from three or more sources with different fidelity levels. For this purpose, our proposed method can be used in a nested or hierarchical fashion to collect data from higher fidelity levels. For example, when three levels of fidelity, low-accuracy (LA), medium-accuracy (MA), and high-accuracy (HA) exists, one can first build a base surrogate model using

the LA data and employ our approach to collect MA data and construct a more accurate model (intermediate model) by fusion of LA and MA data. Then the intermediate model can be used as the base model in our approach to collect HA data and the final model.

7 Conclusion

In many applications, inexpensive high-density but low-accuracy data is fused with high-accuracy data obtained from expensive experiments to improve the model estimation. Current approaches assume that the HA data is already measured and focus on techniques to integrate the LA and HA data. This paper considers the problem of selecting high-accuracy data points to be fused with available LA data. In particular, an adaptive approach that takes advantage of the LA data as well as the previously selected HA points is proposed to sequentially select the next HA point. The proposed approach modifies the concept of expected improvement and combines it with the link model used in data fusion to collect as few HA data points as possible so when combined with the available LA data, a more accurate surrogate model is achieved. The proposed adaptive data fusion approach is compared and contrasted with three different benchmarks in several simulation and case studies. The simulation studies illustrated the strength of our method in modeling a non-stationary surface with a few HA data points when biased LA data of a surface was available. Two case studies illustrated the benefit of our approach in real-world applications. The first case study modeled a freeform surface by collecting (from a CMM) a few HA points and integrating them with a large number of available LA data points measured by a structured laser scanner. The results indicated that our approach selects more points at the locations with larger local variations, which resulted in a more accurate model in terms of prediction errors. In the second case study, the performance measurements of a vehicle engine were exploited to collect HA data from a similar engine to generate an accurate surrogate model for ECU calibration. It was observed that the model generated by the proposed adaptive method produces more accurate predictions compared to the models constructed by the benchmarks.

This paper assumes that LA data is already available and utilizes the information from such data to sample HA data. However, when low-accuracy data is not available, one may or may not consider sampling LA data. Decision to sample LA data to be used for sampling HA data depends

on the relative cost of sampling and the accuracy of LA and HA data and requires further analysis as a future work.

References

Box, George EP, & Draper, Norman R. 1959. A basis for the selection of a response surface design. *Journal of the American Statistical Association*, **54**(287), 622–654.

Bradley, C, & Chan, V. 2001. A complementary sensor approach to reverse engineering. *Journal of Manufacturing Science and Engineering*, **123**(1), 74–82.

Branin, Franklin H. 1972. Widely convergent method for finding multiple solutions of simultaneous nonlinear equations. *IBM Journal of Research and Development*, **16**(5), 504–522.

Colosimo, Bianca Maria, Moroni, Giovanni, & Petrò, Stefano. 2010. A tolerance interval based criterion for optimizing discrete point sampling strategies. *Precision Engineering*, **34**(4), 745–754.

Colosimo, Bianca Maria, Pacella, Massimo, & Senin, Nicola. 2015. Multisensor data fusion via gaussian process models for dimensional and geometric verification. *Precision Engineering*, **40**, 199–213.

Ezzat, Ahmed Aziz, Pourhabib, Arash, & Ding, Yu. 2017. Sequential design for functional calibration of computer models. *Technometrics*.

Fang, Kai-Tai. 1980. The uniform design: application of number-theoretic methods in experimental design. *Acta Mathematicae Applicatae Sinica*, **3**(4), 363–372.

Forrester, Alexander IJ, Sóbester, András, & Keane, Andy J. 2007. Multi-fidelity optimization via surrogate modelling. *Pages 3251–3269 of: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 463. The Royal Society.

Johnson, Mark E, Moore, Leslie M, & Ylvisaker, Donald. 1990. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, **26**(2), 131–148.

Kennedy, Marc C, & O'Hagan, Anthony. 2000. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, **87**(1), 1–13.

Kennedy, Marc C, & O'Hagan, Anthony. 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **63**(3), 425–464.

Kirkpatrick, Scott, Gelatt, C Daniel, & Vecchi, Mario P. 1983. Optimization by simulated annealing. *Science*, **220**(4598), 671–680.

Lam, Chen Quin. 2008. *Sequential adaptive designs in computer experiments for response surface model fit*. Ph.D. thesis, The Ohio State University.

Le Gratiet, Loic, & Garnier, Josselin. 2014. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, **4**(5).

McKay, Michael D, Beckman, Richard J, & Conover, William J. 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**(2), 239–245.

Morris, Max D, & Mitchell, Toby J. 1995. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, **43**(3), 381–402.

Qian, Peter ZG, & Wu, CF Jeff. 2008. Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, **50**(2), 192–204.

Qian, Zhiguang, Seepersad, Carolyn Conner, Joseph, V Roshan, Allen, Janet K, & Wu, CF Jeff. 2006. Building surrogate models based on detailed and approximate simulations. *Journal of Mechanical Design*, **128**(4), 668–677.

Rasmussen, Carl Edward, & Williams, Christopher KI. 2006. *Gaussian processes for machine learning*. Vol. 1. MIT press Cambridge.

Reese, C Shane, Wilson, Alyson G, Hamada, Michael, Martz, Harry F, & Ryan, Kenneth J. 2004. Integrated analysis of computer and physical experiments. *Technometrics*, **46**(2), 153–164.

Sacks, Jerome, Welch, William J, Mitchell, Toby J, & Wynn, Henry P. 1989a. Design and analysis of computer experiments. *Statistical Science*, 409–423.

Sacks, Jerome, Schiller, Susannah B, & Welch, William J. 1989b. Designs for computer experiments. *Technometrics*, **31**(1), 41–47.

Santner, Thomas J, Williams, Brian J, & Notz, William I. 2013. *The design and analysis of computer experiments*. Springer Science & Business Media.

Schonlau, Matthias. 1997. *Computer experiments and global optimization*. Ph.D. thesis, University of Waterloo.

Shewry, Michael C, & Wynn, Henry P. 1987. Maximum entropy sampling. *Journal of Applied Statistics*, **14**(2), 165–170.

Stein, Michael L. 2012. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.

Xia, Haifeng, Ding, Yu, & Mallick, Bani K. 2011. Bayesian hierarchical model for combining misaligned two-resolution metrology data. *IIE Transactions*, **43**(4), 242–258.

Xiong, Shifeng, Qian, Peter ZG, & Wu, CF Jeff. 2013. Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, **55**(1), 37–46.