Journal Article

A multi-fidelity surrogate-model-assisted evolutionary algorithm for computationally expensive optimization problems

Liu, B., Koziel, S. and Zhang, Q.

A Multi-Fidelity Surrogate-Model-Assisted Evolutionary Algorithm for Computationally Expensive Optimization Problems

Bo Liu[a], Slawomir Koziel[b], Qingfu Zhang[c]

[a] *Department of Computing, Glyndwr University, UK*
[b] *Engineering Optimization & Modeling Center, Reykjavik University, Iceland.*
[c] *School of Computer Science & Electronic Engineering, University of Essex, UK*

Abstract

Integrating data-driven surrogate models and simulation models of different accuracies (or fidelities) in a single algorithm to address computationally expensive global optimization problems has recently attracted considerable attention. However, handling discrepancies between simulation models with multiple fidelities in global optimization is a major challenge. To address it, the two major contributions of this paper include: (1) development of a new multi-fidelity surrogate-model-based optimization framework, which substantially improves reliability and efficiency of optimization compared to many existing methods, and (2) development of a data mining method to address the discrepancy between the low- and high-fidelity simulation models. A new efficient global optimization method is then proposed, referred to as multi-fidelity Gaussian process and radial basis function-model-assisted memetic differential evolution. Its advantages are verified by mathematical benchmark problems and a real-world antenna design automation problem.

Keywords: multi-fidelity, multilevel, variable fidelity,
surrogate-model-assisted evolutionary algorithm, expensive optimization

Email addresses: b.liu@glyndwr.ac.uk (Bo Liu), koziel@ru.is (Slawomir Koziel), qzhang@essex.ac.uk (Qingfu Zhang)

1. Introduction

Solving many real-world engineering design problems requires both global optimization and expensive computer simulation for evaluating their candidate solutions. For example, computational models utilized in photonics and micro-electromechanical system optimization require well over 1 hour simulation time per design [1],[2],[3]. For these problems, successful surrogate-model-assisted local search methods have been developed [2], but in terms of global optimization, many state-of-the-art surrogate-based optimization techniques are still prohibitively expensive. In the context of global optimization, these tasks are considered as very expensive. Addressing such problems is the objective of this paper.

High cost of evaluating real-world simulation models often results from the necessity of solving complex systems of partial differential equations using numerical techniques or Monte-Carlo analysis. Direct handling of such models is often computationally prohibitive and utilization of cheaper representations (surrogates) of the system at hand might be necessary. Two classes of such replacement models are normally considered. The first type is function approximation model (usually, data-driven surrogates constructed by approximating sampled simulation model data, e.g., radial basis function). Optimization methods making use of such models are often referred to as surrogate / metamodel-based optimization (SBO) methods [4]. The other type is low-fidelity simulation model (e.g., coarse-mesh model in finite element analysis), which exhibits relaxed accuracy but shorter evaluation time. Low-fidelity model is typically used with occasional reference to the high-fidelity model. The methods using such models are often referred to as multi-fidelity / multilevel / variable-fidelity optimization (MFO) methods [5]. For simplicity, two-level modeling is considered in this paper: the coarse model is referred to as the low-fidelity model, whereas the fine model is referred to as the high-fidelity model.

Recently, a trend to combine SBO and MFO in a single algorithm for further speed improvement has been observed; successful examples include [6, 7, 8, 9],

etc. [6, 9] demonstrated an approach which iteratively updates a co-kriging surrogate model [10] using samples from coarse and fine model evaluations accumulated over the entire optimization process. These techniques are mathematically sound and often feature good reliability. However, for problems with more than a few (e.g., 15) decision variables, the computational cost of obtaining sufficient number of samples to build a high-quality co-kriging model may be prohibitive.

In order to alleviate these difficulties, a new hierarchical algorithm structure has been proposed in [7, 8]: It can be considered as an MFO, but SBOs are used for some optimization stages with certain fidelities to replace standard optimization methods without data-driven surrogate models. For example, a coarse model is used for a surrogate model-assisted evolutionary algorithm to explore the space and accurate but expensive fine model evaluations are only used for local search starting from the most promising solutions obtained from space exploration [7, 11]. This kind of methods is scalable if proper SBOs are used, but the reliability of the MFO structure becomes a challenge, which is detailed as follows.

Because both the coarse and the fine model describe the same function (or a physical system), it is reasonable to use the cheaper coarse model for filtering out some non-optimal subregions. However, considering discrepancy between the models of various fidelities, there is a lot of uncertainty regarding the "promising" solutions found using the coarse model. Fig. 1 illustrates this issue using an example of a microwave filter. The problem is to minimize the maximum value of the reflection coefficient, i.e., $max(|S_{11}|)$ for a given frequency range of interest. It can be observed by sweeping one of the nine design variables of the device, that although many non-optimal regions for the coarse model are also non-optimal for the fine model, the two critical challenges appear:

- wasting precious fine model evaluations because some promising locations of the coarse model-based landscape (such as A and its projection A', which may correspond to multiple design variables) may have substantial distance to the desired optimal regions of the fine model-based landscape

(like B');

- making the MFO framework unreliable because the desired optimum such as the point B' is difficult to be reached from points such as A' by exploitation. Note that only one variable is changed in Fig. 1. When considering multiple design variables, the point B may have a low probability to be selected for exploitation because there may be quite a few points with better fitness values according to the coarse model.

Clearly, the lower the fidelity of the coarse model, the higher the efficiency of the space exploration stage of an MFO, but the higher the risk induced by model discrepancy. [8] investigates the discrepancy problem using practical antenna design cases and indicates that, in many cases, a large number of fine model evaluations may be needed which may result in the same or even higher overall design optimization cost than that of direct optimization of the fine model; also, the MFO may simply fail to find a satisfactory design. There are some methods that do not directly use the promising points from the coarse model optimization, include equalization of the computational effort for models of each fidelity [5], space mapping and model correction, where a correction function is applied to reduce misalignment between the coarse and the fine model [8, 12], etc. However, the above two critical challenges still remain.

To address the above two challenges, a new MFO framework is proposed in this paper. Its goal is to make full use of available expensive fine model evaluations and substantially improve the reliability compared to existing MFO frameworks, and thus, addressing the targeted very expensive design optimization problems. Based on this framework, a data mining method is proposed to address the discrepancy between the coarse and the fine model. A new method, referred to as multi-fidelity Gaussian process and radial basis function-model-assisted memetic differential evolution (MGPMDE), is subsequently proposed. Empirical studies on mathematical benchmark problems with different characteristics as well as a real-world antenna design automation problem verify the advantages of MGPMDE.

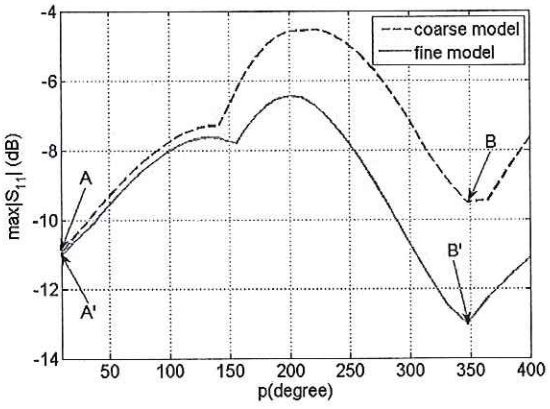

Figure 1: Curves of coarse and fine models

The remainder of this paper is organized as follows. Section II formulates the optimization problem and introduces the basic techniques. Section III describes the new MFO framework, the data mining method and the MGPMDE algorithm. Section IV presents the experimental results of MGPMDE on test problems. Concluding remarks are provided in Section V.

2. Problem Formulation and Basic Techniques

2.1. Problem Formulation

We consider the following problem:

$$\begin{aligned} &min \quad f_f(x) \\ &x \in [\bar{a}, \bar{b}]^d \end{aligned} \tag{1}$$

where $f_f(x)$ is the fine model function, which is expensive but accurate. There is a coarse model function, $f_c(x)$, which is much cheaper than $f_f(x)$, but less accurate than $f_f(x)$, and, consequently, with a distorted landscape. [5] provides an effective method to construct mathematical benchmark problems for MFO, which is as follows.

$$f_f(x) = f_c(s_f \times (x - s_s)) \tag{2}$$

where $f_c(x)$ (also $f_f(x)$) is a periodic function, and there exist minimal and maximal values in each period. s_f is called a frequency shift, which mimics the loss of peaks of $f_c(x)$. s_s is called a spatial shift, which shifts the positions of the optimal points. For example, $f_f(x) = cos(s_f \times (x - s_s))$ and $f_c(x) = cos(x)$. When s_f is set to 1.3, it indicates that 30% of the peaks are lost by $f_c(x)$. The frequency shifts and the spatial shifts often happen for expensive evaluations obtained by solving suitable systems of partial differential equations, where the coarse model is a coarse-mesh model and / or with reduced number of solver iterations. This kind of expensive optimization problem is very (if not the most) popular in computationally expensive engineering design optimization, because most physics simulations (e.g., electromagnetic simulation) are based on solving partial differential equations.

It is worth to determine the focused extent of discrepancy before proposing methods to address it. From the point of view of practical industrial problems [2, 5, 13, 14], we focus on reasonably large discrepancy between computational models of various fidelities in this work. "Reasonably large" discrepancy refers to the fact that the optimal designs in terms of the fine model cannot be obtained by local exploitation based on the optimal points in terms of the coarse model, but the landscape(s) of the coarse model maintains the general shape of that of the fine model. Otherwise, a better coarse model (by increasing the fidelity) should be used. Also based on [2, 5, 13, 14], the focused optimization problems have unimodal or multimodal landscapes and with 5-20 variables.

2.2. Blind Gaussian Process Modeling and Prescreening

In MGPMDE, Gaussian process (GP) regression [15] is used. To model an unknown function $y = f(x), x \in R^d$, GP modeling assumes that $f(x)$ at any point x is a Gaussian random variable $N(\mu, \sigma^2)$, where μ and σ are two constants independent of x. For any x, $f(x)$ is a sample of $\mu+\epsilon(x)$, where $\epsilon(x) \sim N(0, \sigma^2)$. By maximizing the likelihood function to ensure that $f(x) = y^i$ at $x = x^i$ $(i = 1, \ldots, K)$ (where $x^1, \ldots, x^K \in R^d$ and their f-function values $y^1, \ldots, y^K$

are K training data points) and the best linear unbiased prediction:

$$\hat{f}(x) = \hat{\mu} + r^T C^{-1}(y - 1\hat{\mu}) \tag{3}$$

the mean squared error is:

$$s^2(x) = \hat{\sigma}^2[1 - r^T C^{-1} r + \frac{(1 - \mathbf{1}^T C^{-1} r)^2}{\mathbf{1}^T C^{-1} r}] \tag{4}$$

where $r = (c(x, x^1), \ldots, c(x, x^K))^T$. C is a $K \times K$ matrix whose (i, j)-element is $c(x^i, x^j)$. $c(x^i, x^j)$ is the correlation function between x^i and x^j, whose hyper-parameters are estimated by maximization of the likelihood function [15]. $y = (y^1, \ldots, y^K)^T$ and $\mathbf{1}$ is a K-dimensional column vector of ones.

The above surrogate modeling mechanism is called the ordinary GP modeling. In the blind GP modeling [16, 17], the linear combination of some basis functions $\sum_{i=1}^{m} \beta_i b_i(x)$ is used to replace $\hat{\mu}$ so as to capture a portion of the variations which is desirable to represent the general trend of $f(x)$, so as to alleviate the complexity of the ordinary GP modeling, which handles the residuals.

The blind GP modeling consists of the following steps: (1) based on the available training data points, an ordinary GP model is firstly constructed by identifying the hyper-parameter values; (2) given the hyper-parameters and the candidate features, the $b_i(x)$ are ranked based on the estimated $\beta_i (i = 1, \ldots, m)$. The ranking follows a Bayes variable ranking method [16, 17]. For simplicity and efficiency, only linear, quadratic items and two-factor interactions are considered as the candidate features ($b_i(x)$) in this implementation; (3) the most promising features among $b_i(x) (i = 1, \ldots, m)$ are selected and an intermediate GP model with the original hyper-parameters is constructed. Its accuracy is subsequently evaluated by a leave-one-out cross-validation method [16]. This step is repeated until no accuracy improvement can be achieved; (4) given the selected features and the corresponding coefficients, the likelihood function is re-optimized and the final GP model is obtained. The details can be found in [17].

For a minimization problem, given the predictive distribution $N(\hat{f}(x), s^2(x))$

for $f(x)$, a lower confidence bound (LCB) prescreening of $f(x)$ is used:

$$\begin{aligned} f_{lcb}(x) &= \hat{f}(x) - \omega s(x) \\ \omega &\in [0, 3] \end{aligned} \tag{5}$$

where ω is a constant. More details can be found in [18].

2.3. Differential Evolution

The differential evolution (DE) algorithm [19] is used as the global optimization search engine in the MGPMDE method. There are quite a few DE mutation strategies available which lead to various trade-offs between the convergence rate and the population diversity. The properties of different DE strategies in SBO, and more specifically, under the selected SBO framework for global optimization in this work, have been investigated in [20]. Based on [20] and our pilot experiments, DE/current-to-best (6) and DE/best/2 (7) are used in MGPMDE.

Suppose that P is a population and the best individual in P is x^{best}. Let $x = (x_1, \ldots, x_d) \in R^d$ be an individual solution in P. To generate a child solution $u = (u_1, \ldots, u_d)$ for x, DE/current-to-best/1 and DE/best/2 work as follows:

A donor vector is first produced by mutation: **(1)DE/current-to-best/1**

$$v_i = x^i + F \cdot (x^{best} - x^i) + F \cdot (x^{r_1} - x^{r_2}) \tag{6}$$

where x^i is the i^{th} individual in P. x^{r_1} and x^{r_2} are two different solutions randomly selected from P; they are also different from x^{best} and x^i. v_i is the i^{th} mutant vector in the population after mutation. $F \in (0, 2]$ is a control parameter, often referred to as the scaling factor [19].

(2)DE/best/2

$$v_i = x^{best} + F \cdot (x^{r_1} - x^{r_2}) + F \cdot (x^{r_3} - x^{r_4}) \tag{7}$$

where x^{r_3} and x^{r_4} are two different solutions randomly selected from P, and different from x^{best}, x^{r_1} and x^{r_2}.

Having the donor vector, a binomial crossover is applied to produce the child solution with the crossover rate $CR \in [0, 1]$. More details can be found in [19].

2.4. The ORBIT Algorithm

In MGPMDE, the ORBIT algorithm [21] is utilized to perform local search using expensive fine model evaluations. ORBIT is a mathematically sound and successful radial basis function (RBF)-assisted trust-region method and excellent results have been reported in [21]. The main advantage of ORBIT is that it tries to build a unique RBF model based on the sampled points for accurate predictions. Given a set of evaluated data, ORBIT realizes the following procedures: (1) Select points to build (if possible) an RBF model which is fully linear within a neighbourhood of the current trust-region. (2) Include additional points if necessary to ensure that the model parameters remain bounded. (3) Fit the RBF parameters (a cubic RBF is used in this implementation). (4) If the model gradient is not too small, obtain a fully linear RBF model within a *small* neighbourhood. (5) Determine a search step based on the RBF model. (6) Update trust-region parameters. (7) Perform exact evaluation at an additional point if the model is not fully linear within the neighbourhood. More details about ORBIT and fully linear RBF model are in [21].

3. The MGPMDE Algorithm

3.1. The New MFO Framework

The proposed MFO framework is as follows:

Stage 1: Coarse Model-based SBO: Construct the pool of representative candidate solutions based on an efficient and scalable SBO (should have global search ability) using coarse model-based evaluations (CEs).

Stage 2: Data Mining: Generate the initial population for fine model-based optimization by clustering of the representative solution pool and self-development using fine model-based evaluations (FEs).

Stage 3: Fine Model-based Memetic SBO: Carry out memetic SBO from the initial population in Stage 2 using FEs with an appropriate strategy balancing the surrogate-model-assisted global and local search.

Compared to most MFO frameworks, two distinct differences of the new MFO framework are: (1) The initial candidate solutions for FE-based search are generated based on a data mining process, instead of a set of selected "promising" candidates based on CEs (starting FE-based optimization from the latter would lead to wasting FEs and degraded performance of the optimization process). (2) Both global and local search are conducted in FE-based search, instead of only using local exploitation.

The goal of Stage 1 and Stage 2 is to construct an expected good initial population for FE-based optimization in Stage 3, using as few expensive evaluations as possible. Two clarifications are: (1) Despite discrepancies between the models of different fidelities, it is reasonable to assume that the points visited by the CE-based SBO represent the positions of the decision space which are worth to be studied. Many unpromising subregions are naturally filtered out in this process with the support of SBO. (2) Because the qualities of the representative candidates are not known in terms of FE, and the number of FEs should be as few as possible for selecting truly good candidates, clustering techniques are to be used. Because of the model discrepancy and the limited number of FEs in this process, the candidates extracted at this stage may not be of sufficient number and of sufficiently good quality when directly serving as the initial population of Stage 3. Hence, self-development using FEs based on extracted good candidates (seed population) is necessary.

Although the number of FEs used in Stage 2 is small in the above framework, more CEs have to be used than in most MFOs, since a complete global optimization using the coarse model is necessary, instead of just selecting "promising" candidates in the middle of the process. Nevertheless, the cost of CEs can still be reasonable if the following three conditions are satisfied: (1) CE is much cheaper than FE, (2) the SBO used for CE-based optimization is efficient enough (i.e., require fewer exact evaluations), and (3) the stopping criterion is appropriate. More details are in Section III.B.

Only FEs are used in Stage 3. The efficiency improvement comes from an expected initial population with good optimality and diversity from Stage 2

and the memetic SBO. The prerequisites for the memetic SBO are: (1) to have a global optimization ability instead of only performing local exploitation, (2) to exhibit improved efficiency compared to the selected SBO without sacrificing the solution quality, (3) to have sufficient flexibility to find a reasonable trade-off between global exploration and local exploitation for problems with prohibitively expensive FEs where global optimization is not possible.

In the following, an algorithm is proposed according to the above framework and requirements of each stage, called MGPMDE. The surrogate model-aware evolutionary search (SMAS) framework [22] with blind GP surrogate modeling is selected as the SBO for global optimization and the ORBIT algorithm with RBF surrogate modeling is selected as the SBO for local optimization. Empirical studies in Section IV show that the use of SMAS and ORBIT is an appropriate choice for the targeted very expensive problems, but note that it is not the only choice, and other successful SBO frameworks (e.g., [4], etc) can also be applied when satisfying the requirements of each stage of the above framework.

3.2. Construction of the Representative Candidate Solution Pool

This subsection describes operation of Stage 1 of MGPMDE, which works as follows:

Step 1: Use Latin Hypercube Sampling [23] to allocate α solutions within the decision space $[\bar{a}, \bar{b}]^d$, evaluate all these solutions using CEs and form the initial database.

Step 2: If the difference of population diversity P_D is less than δ_1 in 50 consecutive iterations, output the database; otherwise go to Step 3.

Step 3: Select the λ best solutions from the database to form a population P.

Step 4: Apply DE/current-to-best/1 (6) and binomial crossover on P to generate λ child solutions.

Step 5: Select τ nearest evaluated points (based on Euclidean distance) around the centroid of the λ child solutions as training data points [20]. Construct a surrogate model using blind GP with the selected training data points. Prescreen the λ child solutions generated in Step 4.

Step 6: Evaluate $f_c(x)$ (the value of the estimated best child solution from Step 5) using CE. Add this evaluated solution and its function value to the database (i.e., representative candidate solution pool). Go back to Step 2.

The population diversity is calculated as [24]:

$$\begin{aligned} P_D &= \frac{1}{\lambda}\sum_{i=1}^{\lambda} min_{j\in\{1,2,\ldots,\lambda\},j\neq i} d_n(P_i,P_j) \\ d_n(P_i,P_j) &= \sqrt{\frac{1}{d}\sum_{k=1}^{d}(\frac{P_{i,k}-P_{j,k}}{b_k-\bar{a}_k})^2} \end{aligned} \tag{8}$$

At this stage, an SMAS-based SBO is used to construct the representative candidate solution pool, consisting of all the solutions which undergo CE. The selection of SMAS is because of its efficiency considering the expensiveness of the targeted problem. The GPEME algorithm [22] based on SMAS and GP modeling ensures comparable results but uses considerably fewer number of exact evaluations compared to SBOs with several popular frameworks, as verified using more than ten benchmark test problems. In an improved SMAS [20], the efficiency is further enhanced along with the exploration ability of the algorithm. For example, the local optimum in the narrow valley of the 20/30-dimensional Rosenbrock function can be escaped by the improved SMAS, which is difficult for many global optimization-focused SBOs. Having in mind the prerequisites for the SBO process formulated in Section III.A, the following two clarifications on efficiency of SMAS and the stopping criterion are provided below.

The efficiency of SMAS comes from high-quality surrogate modeling and the balance between exploration and exploitation. In each iteration, the λ current best candidate solutions form the parent population (it is reasonable to assume that the search focuses on the promising subregion) and the best candidate based on prescreening in the child population is selected to replace the worst one in the parent population. This way, only at most one candidate is altered

in the parent population in each iteration; consequently, the best candidate among the child solutions in several consecutive iterations may not be far away with each other (they will then be evaluated and are used as the training data points). Therefore, the training data points describing the current promising region can be much denser than those generated by a standard EA population updating mechanism (in which the solutions may spread in different subregions of the decision space while there may not be sufficient number of the training data points around the candidate solutions to be prescreened). For this reason SMAS-based SBO is very efficient.

To maintain the exploration ability, an appropriate strategy for generating the child population should be used so that a reasonably high diversity in or around the promising subregions is introduced. Experimental results in [20] show that by combing the promising subregion-emphasised search mechanism and the DE/current-to-best/1 strategy, a good balance of exploration and exploitation is obtained and a good performance in terms of solution quality for 10 to 30-dimensional multimodal benchmark problems is shown. More details of SMAS are in [22, 20].

Similarly as in most SBOs, an intensive search (i.e., fine tuning) may be carried out around the finally obtained optimum when using SMAS. However, considering the discrepancy between the coarse and the fine model, costing CEs in this intensive search process may not be necessary. Therefore, terminating the search based on the population diversity P_D seems to be a more reasonable approach. When P_D is within a very small range for a number of consecutive iterations, a fine tuning is expected to be applied to P, which is the appropriate time to terminate. Note that the population diversity estimation method utilized here takes into account the dimensionality and the search ranges in order to obtain a good generality.

3.3. Population Initialization

This subsection describes Stage 2 (data mining) of MGPMDE, which works as follows:

Step 1: Divide the representative candidate solution pool into G groups based on the $f_c(x)$ values. The $i^{th}(i = 1, 2, \ldots, G)$ group has $f_c(x)$ values of $[min(f_c(x))+\frac{i-1}{G}(max(f_c(x))-min(f_c(x))), min(f_c(x))+\frac{i}{G}(max(f_c(x))-min(f_c(x)))]$.

Step 2: Divide each group of solutions into n_0 clusters (based on Euclidean distance) using the intelligent Kmeans method [25] and obtain $n_0 \times G$ centroids.

Step 3: Find the nearest point to each obtained centroid from the representative candidate solution pool and evaluate them using FEs to obtain $f_f(x)$.

Step 4: Cluster the group with the best $f_f(x)$ into λ_0 clusters using the intelligent Kmeans method to form the seed population P_s.

Step 5: If $||P_s|| = \lambda$, output P_s. Otherwise; go to Step 6.

Step 6: Apply DE/best/2 (7) and the binomial crossover on P_s to generate $||P_s||$ child solutions.

Step 7: Use all the solutions in P_s as the training data points to construct a blind GP model. Prescreen the child solutions generated in Step 6.

Step 8: Evaluate the $f_f(x)$ value of the estimated best child solution from Step 7. Add this evaluated solution and its function value to P_s. Go back to Step 5.

Some clarifications are as follows.

The representative candidate solution pool should not be directly clustered. When preparing the representative solution pool (Section III.B), the search gradually transforms from global exploration to local exploitation. Hence, the earlier visited solutions have much larger distances between each other than the later visited ones. When directly clustering the representative solution pool (based on their distances), the earlier visited solutions will dominate the clustering. Considering the overall similarity of the landscapes of the coarse and fine

models, we do not expect that many promising subregions in terms of FE are located in the vicinity of the candidates visited in the space exploration phase. Therefore, we approximately divide the representative solution pool into groups with comparable (or similar level) distances between each other and then perform clustering in each group. $f_c(x)$ is selected as a reference to approximately reflect different phases of the search, in which the distances between solutions are approximately on the same level. For clustering, the iKmeans clustering [25] is used to prevent the uncertainty of the standard Kmeans clustering.

Because the number of FEs in the data mining process should be as few as possible, only a small number of samples (n_0) can be used to represent each group. Some optimal solutions in terms of FE may thus be missing. However, finding all the optimal candidates is not our goal at this stage. Instead, we aim at constructing an initial population with good optimality and good diversity for the next stage (FE-based optimization). Therefore, a seed population derived from the best group is firstly constructed with expected reasonably good quality and diversity. DE/best/2 is then used to promote both the optimality and the diversity. The population size is continuously increasing until it reaches λ, and in each iteration, all the evaluated solutions are used to generate and prescreen the child population.

3.4. Memetic SMAS

This subsection outlines implementation of Stage 3 of MGPMDE. SMAS is selected as the basic SBO of this stage due to its efficiency and global optimization capability. According to the requirements listed in Section III.A, introducing surrogate-model-assisted local search to SMAS is necessary. The ORBIT algorithm, which is a rigourous and successful method, is selected for this purpose. It has been introduced briefly in Section II and more details of ORBIT are in [21]. A key issue is the method to control the use of SMAS and ORBIT.

When the population of SMAS enters the fine tuning phase (within a small subregion of the search space) and the landscape in this local area is not very

rugged, we can reasonably assume that a surrogate-model-assisted local search with less exploration ability can also obtain an optimal or satisfactory solution using fewer exact evaluations. Due to this, ORBIT is used to replace SMAS when the population diversity becomes small. Note that this is different from many memetic SBOs, where global and local search are iteratively applied in the entire optimization process. Because efficiency is the top priority of MGPMDE considering the targeted problem, local tuning of an intermediate and potentially good solution cannot be afforded. Hence, in MGPMDE, SMAS and ORBIT are separated and ORBIT is applied only once. In addition, a surrogate-model-assisted local search method has ability (to some extent) to avoid getting stuck in local optima because surrogate modeling itself smoothens the landscape. To enhance this ability, a reasonably large initial trust-region radius is used.

In MGPMDE, The memetic SMAS works as follows:

Step 1: Starting from the initial population of Stage 2, iteratively carry out Step 3 to Step 6 in Stage 1 (Section III.B) but with FEs until $P_D \leq \delta_2$ or according to the predetermined computational budget setting.

Step 2: Calculate the Euclidean distances between each individual in the child population and the centroid of the child population. Set the average and the largest distance values as the initial radius and the maximum radius of the trust-region, respectively.

Step 3: Carry out ORBIT until the stopping criterion is met, which can be a certain number of FEs according to the computational budget and / or if the RBF gradient is smaller than a given tolerance.

3.5. Parameter Settings

Each stage of the MGPMDE algorithm has its own parameters. The SMAS parameters and the ORBIT parameters have been well studied and the algorithm performance is not sensitive to their values when following the setup rules [22, 26, 21]. The new parameters that need to be studied are δ_1, δ_2, G, n_0, λ_0. Their settings and our considerations are as follows:

- δ_1 and δ_2 are normalized population diversity values calculated by (8). For a given method, they are not difficult to be obtained by empirical tests. We assume that the search range is $[-1, 1]^d$ (or we can easily adapt δ_1 and δ_2). We suggest $\delta_1 = 0.03$. δ_2 is tuneable to find a reasonable trade-off between exploration and exploitation. If global optimization is expected, δ_2 is suggested to be set to 0.06. For larger values, the local search is more emphasized. Note that in some cases, δ_2 may be replaced by a certain number of FEs according to available computational budget.

- G is the number of groups the pool is divided into based on $f_c(x)$. Its value should not be too small (otherwise the distances in each group are still not at the same level) nor too large (otherwise FEs will be wasted on later steps). We suggest to set it between 4 and 6.

- n_0 is the number of samples to represent each group. As mentioned above, n_0 should be small (Section III.C). We suggest $n_0 = 2$ or 3.

- λ_0 is the size of the seed population. For the sake of reducing the number of FEs and considering the self-development process, λ_0 is suggested to be set around 10.

Although there are 5 parameters, the clear setup rules provided above and their small suggested ranges, prevent parameter setting from being a problem. To verify the robustness, we use the same parameters for various test problems. We use $\delta_1 = 0.03$, $\delta_2 = 0.06$, $G = 5$, $n_0 = 2$ and $\lambda_0 = 10$. For SMAS parameters and ORBIT parameters, we follow [26] and [21]. For simplicity, α and λ are set to 50 for all the tested problems.

4. Experimental Studies

4.1. Mathematical Benchmark Problem Tests

4.1.1. Test problems construction

Six mathematical benchmark problems [27] with different characteristics are selected as the basic functions, as shown in Table I. Their characteristics are

Table 1: Basic Functions Used in the Experimental Studies

Problem	Objective function	Search range	Global optimum
F1	Ellipsoid	$[-30, 30]^{20}$	0
F2	Dixon & Price	$[-30, 30]^{20}$	0
F3	Styblinski-Tang	$[-5, 5]^{5}$	-195.83
F4	Levy	$[-30, 30]^{10}$	0
F5	Ackley	$[-30, 30]^{10}$	0
F6	Ackley	$[-30, 30]^{20}$	0

as follows: (1) The Ellipsoid function is a simple function, which is unimodal and separable. (2) The Dixon & Price function is unimodal but non-separable and with a narrow valley. (3) The Styblinski-Tang function is multimodal and non-separable and the local optima are far away from each other. (4) The Levy function has numerous local optima and the global optimum is located on a plate. It is also non-separable. (5) The Ackley function has numerous local optima. It has a nearly flat outer region and with a narrow peak at the centre. It is also non-separable.

These basic functions serve as either $f_c(x)$ or $f_f(x)$ when constructing the test problems as in [5] (Section II.A). For test problems that contain trigonometric functions, both spatial shifts s_s and frequency shifts s_f are added with increasing complexity. For all the functions, spatial shifts s_s are added, which are randomly generated with up to around 10% of the search ranges. The above discrepancy setting is similar to [5], making sure that the optimal solutions of the FE function are difficult to be obtained by directly exploiting the optimal solutions of the CE function except F1, which is a unimodal problem. More details are provided in the Appendix.

4.1.2. MGPMDE performance study

The statistics of the best function values obtained by MGPMDE based on 30 independent runs are reported in Table II. The number of CEs is determined

Table 2: Statistics of the best function values obtained by MGPMDE for F1-F6

Problem	best	worst	average	std
F1	2.94e-8	2.81e-5	8.75e-6	8.55e-6
F2	0.67	0.94	0.76	0.09
F3	-195.83	-181.69	-194.42	4.31
F4	4.34e-14	3.28	0.52	0.89
F5	3.08e-4	0.91	0.06	0.23
F6	2.8e-4	0.91	0.09	0.28

by δ_1 and the number of FEs is determined by δ_2 and RBF gradient tolerance (1e-4).

It can be inferred from Table II that MGPMDE consistently exhibits *good global optimization capability* and *robustness* when using the same algorithm setup. This is despite of various types and levels of discrepancies between the coarse and fine models, as well as different problem characteristics and complexity. The average values of 30 runs are close to the global optimum.

In terms of reliability and efficiency, Fig. 2 to Fig. 7 show the comparison of MGPMDE and the latest SMAS-based SBO [20], which is an improved version of [22]. The reference method only uses FEs. For comparison purposes, we assume that the coarse model is 10 times faster than the fine model, which is common in industry [2]. The number of CEs used in MGPMDE is thus transformed to "equivalent" FEs for comparisons.

The following conclusions can be drawn: (1) Considering model discrepancy, *good reliability* of MGPMDE is shown, because the median performance of MGPMDE is not worse than the reference method (based on only FE and SMAS) in terms of solution quality. (2) The initial population constructed from the first two stages of MGPMDE exhibits good optimality. According to Figs. 2 to 7, the initial fitness values for F1-F6 are much better than those obtained by means of the reference method using the same number of FEs. To verify

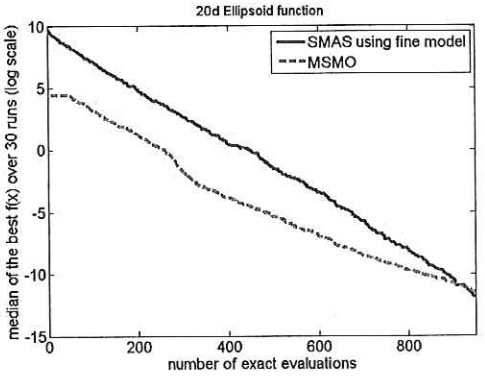

Figure 2: Convergence curve of the objective function for F1

this, Table III lists the best fitness values of the representative solution pool (where CE is used) in terms of FE (f_{rp}), of the seed population (f_{sp}), and of the initial population for Stage 3 (f_{initp}). It can be observed that most f_{sp} are similar to f_{rp}, which verifies the method to generate the seed population (Steps 1-4 in Stage 2). f_{initp} shows clear improvement compared to f_{sp}, which verifies the self-development process in Stage 2 (Steps 5-8) in terms of optimality. (3) Good overall quality (including both optimality and diversity) of the initial population can be observed, because the convergence rate of MGPMDE is similar to the reference method starting from the corresponding fitness values, except for F1 which is a little lower but still similar. Otherwise; the convergence rate should be much lower than that of the reference method or stuck at a local optimum although starting from a population with good optimality. Conclusions (2) and (3) indicate that the goals of the new MFO framework, specifically, obtaining high quality initial population based on CE, data mining and a small number of FEs are met. Conclusion (3) also verifies the overall *good efficiency* of MGPMDE.

The memetic SMAS framework inherits the advantages in terms of computational efficiency of SMAS and provides further improvements by applying ORBIT at suitably selected step of the optimization procedure as explained before. A clear example is shown in Fig. 5, where ORBIT starts at around 680

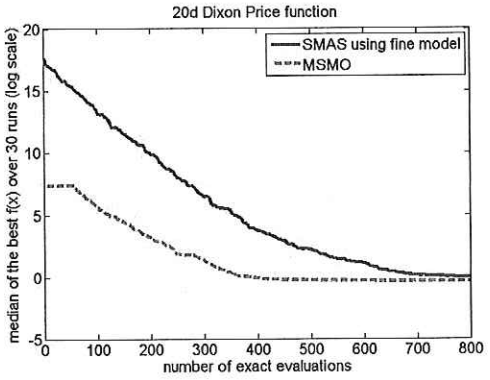

Figure 3: Convergence curve of the objective function for F2

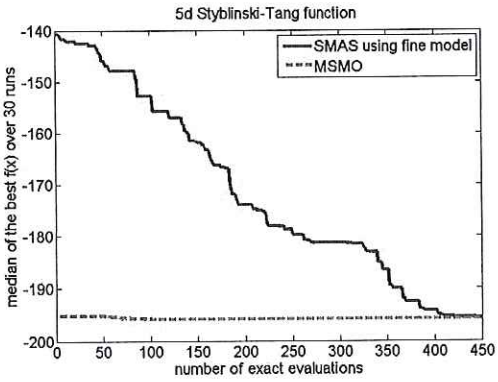

Figure 4: Convergence curve of the objective function for F3

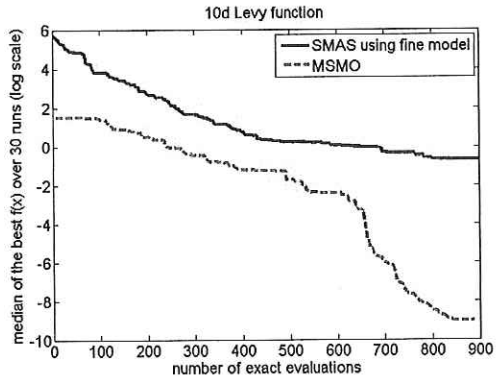

Figure 5: Convergence curve of the objective function for F4

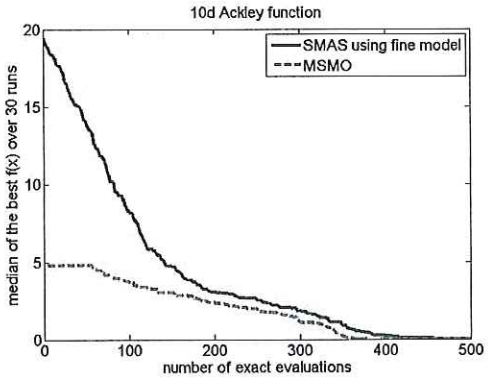

Figure 6: Convergence curve of the objective function for F5

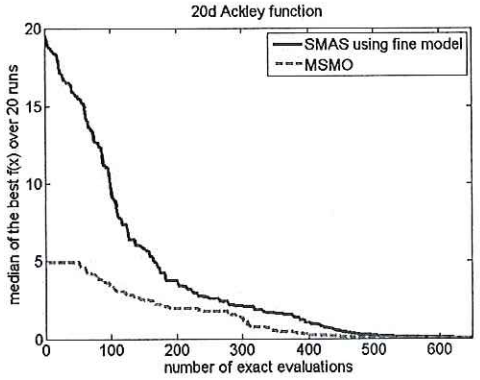

Figure 7: Convergence curve of the objective function for F6

Table 3: Best fitness values for the representative solution pool, seed population and initial population (average over 30 runs)

Problem	f_{rp}	f_{sp}	f_{initp}
F1	121.9	142.1	79.4
F2	3150.5	3823.6	1666.3
F3	-189.2	-186.5	-193.7
F4	4.9	7.6	5.3
F5	4.9	5.4	4.7
F6	5.3	5.8	4.9

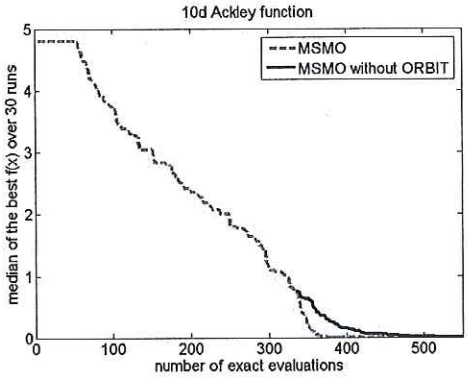

Figure 8: Comparison of MGPMDE with / without ORBIT

iterations. To further demonstrate the benefits of memetic SMAS, a representative example using F5 is shown in Fig. 8. MGPMDE is compared with a method using the same first two stages but with only SMAS (without ORBIT) used at the third stage. It can be seen that about 20% FEs are saved.

4.2. Real-world Problem Tests

MGPMDE is applied to a real-world antenna design automation problem [28]. It is a 10 GHz 16-element microstrip patch antenna implemented on a finite 1.575-mm-thick Rogers RT5880 dielectric substrate, which is shown in Fig. 9. The objective is minimization of the side lobes assuming ±8-degree main beam. The design variables are excitation amplitudes $a_k, k = 1, 2, \ldots, 16$ with a range of $[0, 1]^{16}$. The objective function is as follows:

$$\min \quad SLL(a_k), k = 1, 2, \ldots, 16 \tag{9}$$

where SLL is the sidelobe level, i.e., the maximum relative power for the angles 0 to 82 degrees and 98 to 180 degrees.

The coarse model is an analytical array factor model assuming ideal isotropic radiators [28], for which each calculation costs about 5×10^{-3}s. The fine model is an electromagnetic (EM) simulation model (no explicit analytical formula is available), for which each simulation costs about 30 minutes. In order to

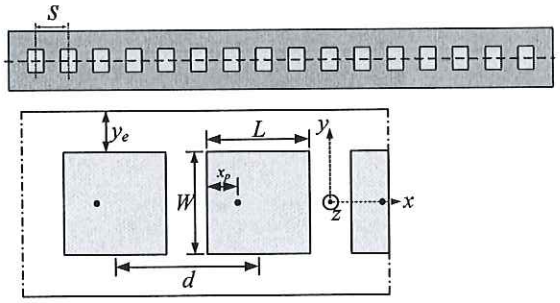

Figure 9: Layout of 16-element microstrip patch array antenna

Table 4: Statistics of the best function values obtained by MGPMDE (150 fine simulations), SMAS (550 fine simulations), standard DE (30,000 fine simulations)

Method	best	worst	average	std
MGPMDE	-22.61	-21.24	-22.16	0.34
SMAS	-22.24	-19.19	-21.43	0.86
Standard DE	-23.14	-23.06	-23.12	0.02

make multiple runs and comparisons possible, the simulation-based superposition model has been created that ensures virtually the same accuracy as the fine EM simulation model but at the fraction of the cost of the latter. The simulation-based superposition model is obtained as superposition of individually simulated far fields of all array elements. Each element is simulated within the array in order to take into account electromagnetic couplings with all other elements.

30 runs of MGPMDE, SMAS and standard DE with the same parameters of Section IV.A are carried out. From practical standpoint, accomplishing the antenna design process within 3 days can be considered as satisfactory. According to the computational budget, 100 fine model simulations are used for Stage 2 and SMAS in Stage 3, and 50 fine model simulations are used for ORBIT.

The results are shown in Table IV. The convergence trends of MGPMDE

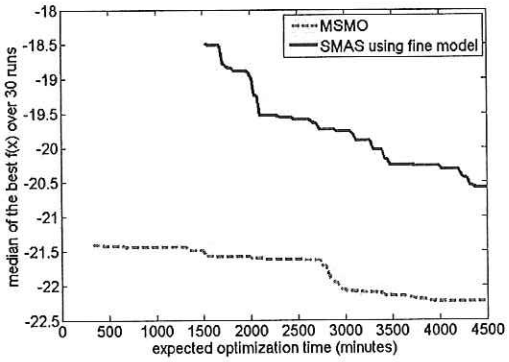

Figure 10: Convergence curve of the objective function for the real-world antenna design example

and SMAS are indicated in Fig. 10. The time expenditure is calculated by multiplying the average cost of each simulation by the number of simulations. The time spent on surrogate model training is less than half an hour (less than 1 FE), and it can be neglected. The following observations can be made: (1) The result of MGPMDE using 150 fine model simulations is of high quality either using antenna theory or the standard DE result as a reference. (2) A large speed improvement of MGPMDE can be observed compared to SMAS with fine model simulations. (3) The average results of MGPMDE and standard DE over 30 runs are compared. To obtain the average result of MGPMDE using 150 fine model simulations, standard DE needs 3800 fine model simulations.

5. Conclusions

In this paper, the MGPMDE method has been proposed. The targeted problems are engineering design optimization tasks with very long simulation time of the relevant computational models, for which even state-of-the-art global SBOs or MFOs may be too expensive. The main contributions of the work are: (1) the development of the new MFO framework which substantially improves reliability of the optimization process and makes comprehensive use of high-fidelity evaluations, (2) the development of the data mining method which provides a

good initial population for fine model-based optimization in terms of both optimality and population diversity, and (3) the development of MGPMDE, showing the combined advantages of high efficiency, high reliability and high optimization quality, as demonstrated by benchmark and antenna design automation problems. The methodology developed under this research also considerably decreases the risk related to usage of lower-fidelity simulation models, as well as further improves the efficiency of the optimization process. Future work will focus on constrained and multi-objective MGPMDE.

Acknowledgement

The authors thank CST AG for making CST Microwave Studio available. The authors would like to thank Dr. Renato Cordeiro de Amorim, Glyndwr University, UK, for valuable discussions.

A. F1: Ellipsoid Problem

$$
\begin{aligned}
& f_c(x) = \textstyle\sum_{i=1}^{d} i \times x_i^2 \\
& f_f(x) = \textstyle\sum_{i=1}^{d} i \times s_{hi} \times (x_i - s_{si})^2 \\
& i = 1, \ldots, d \\
& s_h = [0.3, 0.4, 0.2, 0.6, 1, 0.9, 0.2, 0.8, 0.5, 0.7, \\
& 0.4, 0.3, 0.7, 1, 0.9, 0.6, 0.2, 0.8, 0.2, 0.5] \\
& s_s = [1.8, 0.4, 2, 1.2, 1.4, 0.6, 1.6, 0.2, 0.8, 1, \\
& 1.3, 1.1, 2, 1.4, 0.5, 0.3, 1.6, 0.7, 0.3, 1.9]
\end{aligned}
\tag{.1}
$$

B. F2: Dixon & Price Problem

$$
\begin{aligned}
& f_c(x) = (x_1 - 1)^2 + \textstyle\sum_{i=2}^{n} i(2x_i^2 - x_{i-1})^2 \\
& f_f(x) = f_c(x - s_s) \\
& i = 1, \ldots, d \\
& s_s = [1.8, 0.5, 2, 1.2, 0.4, 0.2, 1.4, 0.3, 1.6, 0.6, \\
& 0.8, 1, 1.3, 1.9, 0.7, 1.6, 0.3, 1.1, 2, 1.4]
\end{aligned}
\tag{.2}
$$

C. F3: Styblinski-Tang Problem

$$
\begin{aligned}
& f_c(x) = 0.5 \times \textstyle\sum_{i=1}^{d}(x_i^4 - 16x_i^2 + 5x_i) \\
& f_f(x) = f_c(x - s_s) \\
& i = 1, \ldots, d \\
& s_s = [0.28, 0.59, 0.47, 0.16, 0.32]
\end{aligned}
\tag{.3}
$$

D. F4: Levy Problem

$$
\begin{aligned}
& f_c(x) = sin^2(s_f \times \pi w_i) + \textstyle\sum_{i=1}^{d-1}[1 + 10sin^2(s_f \times \pi w_i + 1)] \\
& +(w_d - 1)^2[1 + sin^2(2s_f \times \pi w_d)], \\
& w_i = 1 + 0.25(x_i - s_{si} - 1) \\
& f_f(x) = sin^2(\pi w_i) + \textstyle\sum_{i=1}^{d-1}[1 + 10sin^2(\pi w_i + 1)] \\
& +(w_d - 1)^2[1 + sin^2(2\pi w_d)], w_i = 1 + 0.25(x_i - 1) \\
& i = 1, \ldots, d \\
& s_f = 0.8, s_s = [1.2, 0.3, 1, 0.3, 1.6, 0.8, 1.4, 0.7, 2, 1.5]
\end{aligned}
\tag{.4}
$$

E. F5,F6: Ackley Problem

$$
\begin{aligned}
& f_c(x) = -20e^{-0.2\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}} - e^{\frac{1}{d}\sum_{i=1}^{d} cos(2\pi x_i)} \\
& f_f(x) = -20e^{-0.2\sqrt{\frac{1}{d}\sum_{i=1}^{d}(x_i - s_{si})^2}} \\
& -e^{\frac{1}{d}\sum_{i=1}^{d} cos(2 \times s_f \times \pi x_i - s_{si})} \\
& i = 1, \ldots, d \\
& F5 : s_f = 1.3, s_s = [1.3, 0.1, 1.4, 0.8, 1.7, 1, 1.5, 0.6, 2, 0.4] \\
& F6 : s_f = 1.3, s_s = [1.2, 0.2, 1.4, 0.8, 1.8, 1, 1.6, 0.6, 2, 0.4, \\
& 1.3, 0.3, 1.5, 0.9, 1.9, 1.1, 1.7, 0.7, 2.1, 0.5]
\end{aligned}
\tag{.5}
$$

References

[1] J. W. Bandler, Q. S. Cheng, S. A. Dakroury, A. S. Mohamed, M. H. Bakr, K. Madsen, J. Sondergaard, Space mapping: the state of the art, IEEE Transactions on Microwave Theory and Techniques 52 (1) (2004) 337–361.

[2] S. Koziel, L. Leifsson, X.-S. Yang, Solving computationally expensive engineering problems: Methods and applications, Springer.

[3] COMSOL, Comsol multiphysics 4.3 user's guide (2012).

[4] J. Müller, C. A. Shoemaker, R. Piché, SO-MI: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems, Computers & Operations Research 40 (5) (2013) 1383–1400.

[5] M. A. El-Beltagy, A. Keane, A comparison of various optimization algorithms on a multilevel problem, Engineering Applications of Artificial Intelligence 12 (5) (1999) 639–654.

[6] A. I. Forrester, A. Sóbester, A. J. Keane, Multi-fidelity optimization via surrogate modelling, in: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, Vol. 463, The Royal Society, 2007, pp. 3251–3269.

[7] I. C. Kampolis, A. Zymaris, V. Asouti, K. C. Giannakoglou, Multilevel optimization strategies based on metamodel-assisted evolutionary algorithms, for computationally expensive problems, in: IEEE Congress on Evolutionary Computation, IEEE, 2007, pp. 4116–4123.

[8] S. Koziel, S. Ogurtsov, Model management for cost-efficient surrogate-based optimisation of antennas using variable-fidelity electromagnetic simulations, IET Microwaves, Antennas & Propagation 6 (15) (2012) 1643–1650.

[9] L. Huang, Z. Gao, D. Zhang, Research on multi-fidelity aerodynamic optimization methods, Chinese Journal of Aeronautics 26 (2) (2013) 279–286.

[10] M. C. Kennedy, A. O'Hagan, Predicting the output from a complex computer code when fast approximations are available, Biometrika 87 (1) (2000) 1–13.

[11] E. Huang, J. Xu, S. Zhang, C.-H. Chen, Multi-fidelity model integration for engineering design, Procedia Computer Science 44 (2015) 336–344.

[12] L. Leifsson, S. Koziel, Multi-fidelity design optimization of transonic airfoils using physics-based surrogate modeling and shape-preserving response prediction, Journal of Computational Science 1 (2) (2010) 98–106.

[13] P. Rocca, G. Oliveri, A. Massa, Differential evolution as applied to electromagnetics, IEEE Antennas and Propagation Magazine 53 (1) (2011) 38–49.

[14] B. Liu, G. Gielen, F. Fernández, Automated design of analog and high-frequency circuits, Springer.

[15] T. J. Santner, B. J. Williams, W. I. Notz, The design and analysis of computer experiments, Springer, 2003.

[16] V. R. Joseph, Y. Hung, A. Sudjianto, Blind kriging: A new method for developing metamodels, Journal of mechanical design 130 (2008) 031102.

[17] I. Couckuyt, A. Forrester, D. Gorissen, F. De Turck, T. Dhaene, Blind kriging: Implementation and performance analysis, Advances in Engineering Software 49 (2012) 1–13.

[18] J. Dennis, V. Torczon, Managing approximation models in optimization, in: Multidisciplinary design optimization: State-of-the-art, Philadelphia, PA: SIAM, 1997, pp. 330–347.

[19] K. Price, R. Storn, J. Lampinen, Differential evolution: a practical approach to global optimization, Springer-Verlag, 2005.

[20] B. Liu, Q. Chen, Q. Zhang, G. Gielen, V. Grout, Behavioral study of the surrogate model-aware evolutionary search framework, in: IEEE Congress on Evolutionary Computation, IEEE, 2014, pp. 715–722.

[21] S. M. Wild, R. G. Regis, C. A. Shoemaker, ORBIT: Optimization by radial basis function interpolation in trust-regions, SIAM Journal on Scientific Computing 30 (6) (2008) 3197–3219.

[22] B. Liu, Q. Zhang, G. Gielen, A gaussian process surrogate model assisted evolutionary algorithm for medium scale expensive optimization problems, IEEE Transactions on Evolutionary Computation 18 (2) (2014) 180–192.

[23] M. Stein, Large sample properties of simulations using latin hypercube sampling, Technometrics 29 (2) (1987) 143–151.

[24] M. Lozano, F. Herrera, J. R. Cano, Replacement strategies to preserve useful diversity in steady-state genetic algorithms, Information Sciences 178 (23) (2008) 4421–4433.

[25] L. Rutkowski, Clustering for data mining: A data recovery approach, Psychometrika 72 (1) (2007) 109–110.

[26] B. Liu, D. Zhao, P. Reynaert, G. G. Gielen, GASPAD: A general and efficient mm-wave integrated circuit synthesis method based on surrogate model assisted evolutionary algorithm, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 33 (2) (2014) 169–182.

[27] M. Jamil, X.-S. Yang, A literature survey of benchmark functions for global optimisation problems, International Journal of Mathematical Modelling and Numerical Optimisation 4 (2) (2013) 150–194.

[28] S. Koziel, S. Ogurtsov, Simulation-based design of microstrip linear antenna arrays using fast radiation response surrogates, IEEE Antennas and Wireless Propagation Letters (2015) 759–762.