

## Chapter 6 – L1 BENCHMARK DEFINITIONS AND RESULTS

**Laura Mainini**

Politecnico di Torino and  
Massachusetts Institute of Technology  
ITALY

**Andrea Serani, Simone Ficini,  
Riccardo Pellegrini, Matteo Diez**  
National Research Council  
Institute of Marine Engineering  
ITALY

**Andrew S. Thelen, Markus P. Rumpfkeil**  
University of Dayton Research Institute  
UNITED STATES

**Edmondo Minisci**  
University of Strathclyde  
UNITED KINGDOM (Rep. ITALY)

**Domenico Quagliarella**  
Italian Aerospace Research Centre  
ITALY

**Hayriye Pehlivan Solak, Sihmehmet Yildiz,  
Melike Nikbay**  
Istanbul Technical University  
TÜRKİYE

**Francesco Di Fiore**  
Politecnico di Torino  
ITALY

**Dean Bryson, Philip S. Beran**  
Air Force Research Laboratory  
UNITED STATES

### ABSTRACT

*The chapter presents a collection of analytical benchmark problems specifically selected to provide a set of stress tests for the assessment of multi-fidelity optimization methods. These benchmarks, here denoted as L1 problems, address challenges related to the curse of dimensionality and the scalability associated with multi-fidelity methods, handling localized, multimodal, and discontinuous behaviors of the objective functions, and handling the possible presence of noise in the objective functions. The first objective is to provide the community with hard-to-solve but easy-to-implement problems. The second objective is to assess and compare quite a large variety of multi-fidelity methods against benchmark problems. Multi-fidelity methods performance are assessed in terms of both optimization and global accuracy.*

### 6.1 INTRODUCTION

Innovative product requirements are evolving rapidly, reflecting the technological advances in many engineering disciplines. The accelerating nature of this change is accompanied by the growth in product performance, complexity, and cost. To meet emerging requirements, faster design processes are thus required to: thoroughly and accurately explore design spaces of increased size, leverage potentially complex physical interactions for performance benefit, and avoid deleterious interactions that may greatly increase product cost through late defect discovery [1]. Nowadays, there are design benefits by coupling more disciplines at higher levels of fidelity earlier in the development process. But there is no mathematical framework to determine which disciplines, level of coupling, or level of fidelity is required to capture the physics most critical to a particular system's design, where the design space data is best collected, or how to make the best possible design decision with constrained

## L1 BENCHMARK DEFINITIONS AND RESULTS

---

computing resources. Currently, these decisions are based solely on engineering experience. This approach works reasonably well for systems that are similar to previous designs, but can fail for unique and innovative vehicles and technologies.

In this regard, one of the long-term challenges of multidisciplinary design optimization (MDO) is the efficient increase of modeling fidelity, when it is needed, to capture the critical physics that constrain or enable particular product concepts. Relying on low-fidelity models for the analysis throughout the entire design space may lead to designs that are infeasible, or significantly sub-optimal, when the physics is not sufficiently modeled or resolved. Simply replacing these models with higher fidelity models during optimization is often not a practical strategy, because of the higher computational cost associated with these more informative techniques. multi-fidelity methods offer the conceptual framework to efficiently optimize products by judiciously using a limited number of high-fidelity analyses while leveraging the information provided by low-fidelity models. multi-fidelity approaches are considered here to fall into a larger class of methods that coordinate a set of information sources to accelerate the computational task. These information sources quantify the system's response using computational approaches (i.e., a mathematical description and the concomitant numerical analysis) and/or non-computational approaches (e.g., physical experiments, analytical solutions, and expert analysis).

Despite the development of quite a large number of multi-fidelity methods, their capabilities are still under discussion and their potential is still under-explored [2], [3]. Although multi-fidelity methods have shown their potential in achieving optimal design solutions more efficiently than their high- or single-fidelity counterpart (see, e.g., [4], [5]), systematic comparison studies of these methods are still rare in the literature. This motivates the interest in benchmark problems that could support the comparative and rigorous assessment of multi-fidelity methods. Beran et al. [1] propose to classify use cases and test problems into three classes: L1 problems, computationally cheap analytical functions with exact solutions; L2 problems, simplified engineering applications problems that can be executed with a reduced computational expense; and L3 problems, more complex engineering use cases, usually including multiphysics couplings. The NATO AVT-331 research task group on “Goal-Driven, multi-fidelity Approaches for Military Vehicle System-Level Design,” has been conducting a coordinated activity to collect and study benchmarks for these three classes. This paper provides an overview of the L1 benchmarks which are analytical problems with no explicit resemblance to actual engineering problems but support cross-domain investigations. A large number of L1 benchmark problems have been proposed in the literature, mostly in conjunction with the presentation of a novel multi-fidelity method [6]–[26]. However, a comprehensive framework of computationally efficient benchmarks is not yet available.

The chapter proposes and discusses a suite of analytical benchmark problems specifically formulated and selected to stress-test and assess the capabilities of a broad spectrum of multi-fidelity methods. The framework is intended to provide a set of standard problems, recommended experimental setups, and performance assessment metrics to support the rigorous test and comparison of different computational methods. The benchmarks are selected to exemplify mathematical characteristics and behaviors that are often encountered in simulation-based optimization problems and that can challenge the successful search and identification of optimal solutions for real-world engineering applications. Those challenges include: (i) addressing the curse of dimensionality and the scalability associated with multi-fidelity methods; (ii) handling localized, multimodal, and discontinuous behaviors of the objective functions; and (iii) handling the possible presence of noise in the objective functions.

The benchmarks are designed and selected to be of simple implementation while permitting to isolate and investigate different mathematical characteristics to gain insights about the performance of different multi-fidelity approaches to modeling, design, and optimization. multi-fidelity methods performance are

assessed by optimization and global accuracy metrics as described in Chapter 2. Global accuracy metrics provide a measure of the ability to approximate the highest fidelity function, also considered as the ground truth source of information. Optimization accuracy is a goal-oriented metric that measures the efficiency and effectiveness of the method in searching and finding the global optimum.

Finally, a quite large variety of multi-fidelity methods are assessed and compared against the benchmark problems. Multi-fidelity-methods performance are assessed in terms of both optimization and global accuracy and the results are discussed in this chapter.

## 6.2 ANALYTICAL BENCHMARKS FOR MULTI-FIDELITY OPTIMIZATION

The analytical benchmarks proposed would exemplify potential objective functions to be optimized, somehow related to system-level design for complex industrial/military applications, thus solvable by using goal-driven multi-fidelity methods. Specifically, we will consider a box-constrained optimization problem in the form:

$$\min_{\mathbf{x} \in \mathcal{A}} f(\mathbf{x}), \quad \text{with} \quad \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \quad (6-1)$$

where  $\mathbf{x} \in \mathbb{R}^D$  is a design point in the feasible domain  $\mathcal{A}$  bounded by  $\mathbf{l}$  (lower bound) and  $\mathbf{u}$  (upper bound),  $\{x_k\}$  are the elements of  $\mathbf{x}$  ( $k$  is an integer satisfying  $1 \leq k \leq D$  or  $k \in [1, D]$ ),  $D$  is the dimensionality of the parameter space,  $f(\mathbf{x}) \in \mathbb{R}$  is the objective function, and  $\mathbf{x}^*$  is the optimum design point satisfying:

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathcal{A}}{\operatorname{argmin}} f(\mathbf{x}), \quad (6-2)$$

where  $f^* \equiv f(\mathbf{x}^*)$ .

Within the multi-fidelity settings of the benchmarks, the  $f(\mathbf{x})$  to minimize would be the highest fidelity function  $f_1(\mathbf{x})$ , while all the other possible  $L$  representations would be considered as cheaper-to-evaluate approximations of the objective function, thus providing a fidelity spectrum from  $f_1(\mathbf{x})$  up to  $f_L(\mathbf{x})$ , where the latter is the lowest fidelity level available.

The multi-fidelity benchmark problems are selected to capture fundamental mathematical characteristics and properties which also mimic real-world engineering problems. The distinguishing mathematical features of the selected benchmark problems can be listed as: behaviors, scalability, discrepancy type, and noise. Function

**Table 6-1: Analytical Benchmarks Main Features.**

| ID  | Name                         | Behaviors                  | Scalability              | Discrepancy | Noise |
|-----|------------------------------|----------------------------|--------------------------|-------------|-------|
| MF1 | Forrester                    | Local /<br>(Dis)continuous | -                        | (non)linear | no    |
| MF2 | Rosenbrock                   | Local                      | Parametric               | nonlinear   | no    |
| MF3 | Shifted-rotated<br>Rastrigin | Multi-modal                | Parametric /<br>Fidelity | nonlinear   | no    |
| MF4 | Heterogeneous                | Local /<br>Multi-modal     | Parametric               | nonlinear   | no    |
| MF5 | Spring-Mass<br>system        | Multi-modal                | Parametric /<br>Fidelity | nonlinear   | no    |
| MF6 | Paciorek                     | Multi-modal                | Fidelity                 | nonlinear   | yes   |

## L1 BENCHMARK DEFINITIONS AND RESULTS

behaviors include multi-modality, discontinuities, and atypical local behaviors, that cannot be neglected a priori, especially for real-world problems where it could be important to represent the whole variable domain. In particular, multi-modality and discontinuities are challenging from both a modeling and optimization viewpoint. Scalability takes into consideration both the function parameterization and the fidelity spectrum. The former is one of the important criteria for the comprehensive assessment of the performance of the multi-fidelity methods which enables to represent the same parametric function with different input dimensions, whereas the latter is useful to demonstrate how the modeling process can be improved depending on the fidelity level available. This last point is particularly relevant, because of its relation with the discrepancy type, which describes the relation among fidelities. In general, a linear discrepancy is simpler to model than a nonlinear one. For this reason, the discrepancy type allows for a deeper assessment of the multi-fidelity methods, since the number of fidelities available has to be correlated to the associated discrepancy type. Finally, in real-world engineering problems, noise may exist, it is undesired but inescapable in the overall response of a system which may exhibit abrupt changes within the solution domain. It is difficult or even impossible to distinguish the individual impact of the noise in the overall response and eliminate it. Thus, the user has to deal with a function with some embedded noise, and is important to assess the ability of multi-fidelity methods to model a noisy function. Considering all these mathematical characteristics over the benchmark problems will allow to evaluate the performance metrics to assess the strengths and weaknesses of the multi-fidelity methods employed. The mathematical characteristics which are considered in this work are briefly summarized in Table 6-1. The following subsections present each benchmark formulation in turn.

### 6.2.1 Forrester Function

The proposed Forrester [27] multi-fidelity benchmark (MF1.1) is a well-known one-dimensional benchmark for multi-fidelity methods, described by the following equations (from 1 highest fidelity to 4 lowest fidelity level) and shown in Figure 6-1.

$$f_1(x) = (6x - 2)^2 \sin(12x - 4) \quad (6-3)$$

$$f_2(x) = (5.5x - 2.5)^2 \sin(12x - 4) \quad (6-4)$$

$$f_3(x) = 0.75f_1(x) + 5(x - 0.5) - 2 \quad (6-5)$$

$$f_4(x) = 0.5f_1(x) + 10(x - 0.5) - 5 \quad (6-6)$$

The function is defined in the domain  $0 \leq x \leq 1$  and the minimum is located at  $x^* = 0.75724876$  and given by  $f(x^*) = -6.020740$ .

In order to observe the performance of the multi-fidelity methods in problems with discontinuous behavior, the discontinuous Forrester function [28] is also selected as one of the benchmarks. The benchmark function (MF1.2) is derived from the revision of Forrester's function and is also called the Forrester function with jump [29]. The discontinuous Forrester function is described by the following equation

$$f_1(x) = \begin{cases} (6x - 2)^2 \sin(12x - 4), & 0 \leq x \leq 0.5 \\ (6x - 2)^2 \sin(12x - 4) + 10, & 0.5 < x \leq 1 \end{cases} \quad (6-7)$$

$$f_2(x) = \begin{cases} 0.5f_1(x) + 10(x - 0.5) - 5, & 0 \leq x \leq 0.5 \\ 0.5f_1(x) + 10(x - 0.5) - 2, & 0.5 < x \leq 1 \end{cases} \quad (6-8)$$

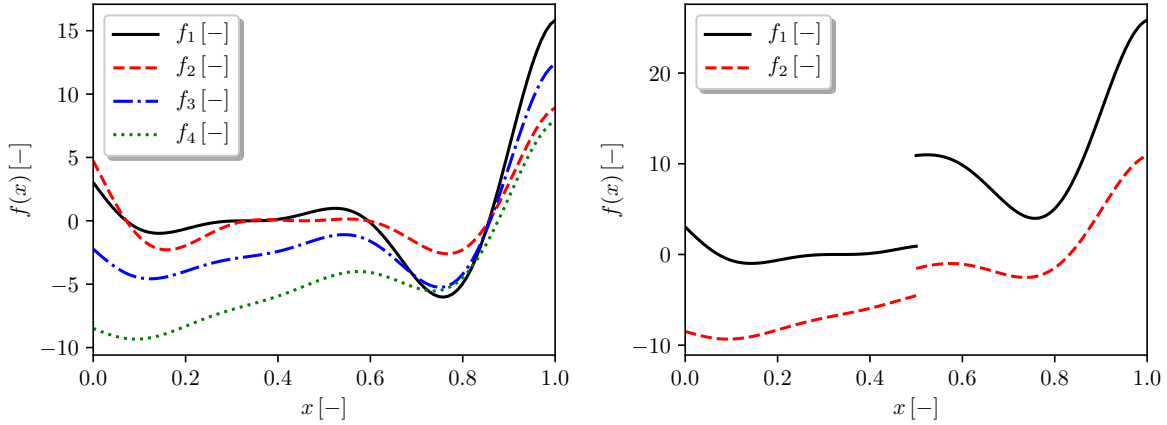


**Figure 6-1: Forrester Function (left) and Discontinuous Variant (right).**

The function is defined in the domain  $0 \leq x \leq 1$  and the minimum is located  $x^* = 0.1426$  and given by  $f(x^*) = -0.9863$ .

### 6.2.2 Rosenbrock Function

The Rosenbrock function, also referred to as the Valley or Banana function, is a well-known D-dimensional optimization benchmark problem described by the following equation

$$f_1(\mathbf{x}) = \sum_{i=1}^{D-1} 100 (x_{i+1} - x_i^2)^2 + (1 - x_i)^2 \quad (6-9)$$

The global minimum is inside a long, narrow, parabolic-shaped flat valley.

The function is unimodal, and the global minimum is located at  $\mathbf{x}^* = \{1, \dots, 1\}^T$  and equal to  $f(\mathbf{x}^*) = 0$ . However, even though this valley is easy to find, convergence to the minimum is difficult. Note that there is also a local minimum at  $\{-1, 1, \dots, 1\}^T$  for  $4 \leq D \leq 7$ . For the present problem, the variable domain is defined as  $-2 \leq x_i \leq 2$  for  $i = 1, \dots, D$ .

The extension to multi-fidelity purposes is described by the following equations, where  $f_2$  can be considered as a medium-fidelity level [30] and  $f_3$  as the lowest fidelity [25].

$$f_2(\mathbf{x}) = \sum_{i=1}^{D-1} 50 (x_{i+1} - x_i^2)^2 + (-2 - x_i)^2 - \sum_{i=1}^D 0.5x_i \quad (6-10)$$

$$f_3(\mathbf{x}) = \frac{f_1(\mathbf{x}) - 4 - \sum_{i=1}^D 0.5x_i}{10 + \sum_{i=1}^D 0.25x_i} \quad (6-11)$$

## L1 BENCHMARK DEFINITIONS AND RESULTS

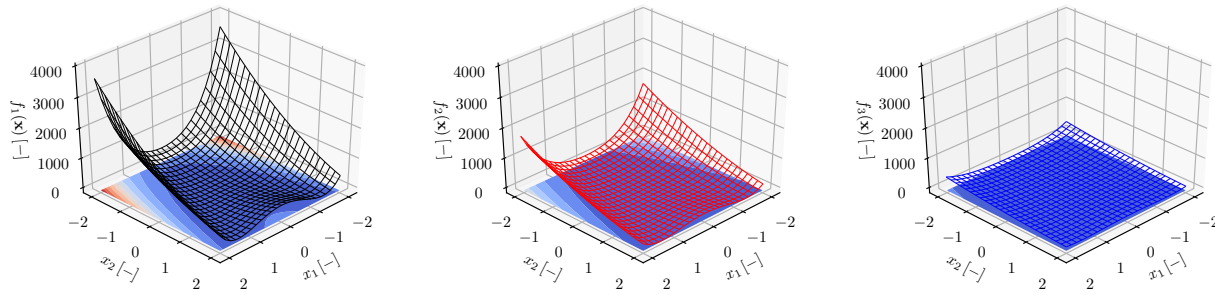


**Figure 6-2: Rosenbrock Function: from Left to Right,  $f_1$  (highest-fidelity),  $f_2$ , and  $f_3$  (lowest-fidelity).**

The three-fidelity levels are shown for two dimensions in Figure 6-2.

### 6.2.3 Shifted-Rotated Rastrigin Function

To address real-world optimization problems, where the landscape of the objective function is usually multi-modal, the Rastrigin function is selected as a benchmark. The function is shifted to change the position of the minimum and rotated to change the properties of the function itself within the variable space. The equation of the shifted-rotated Rastrigin function reads as follows

$$f_1(\mathbf{z}) = \sum_{i=1}^D (z_i^2 + 1 - \cos(10\pi z_i)) \quad (6-12)$$

with

$$\mathbf{z} = R(\theta)(\mathbf{x} - \mathbf{x}^*) \quad \text{with} \quad R(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (6-13)$$

where  $R$  is the rotation matrix in  $2D$  and can be extended to arbitrary dimension by using the Aguilera-Perez algorithm [31].

The variable ranges is defined such as  $-0.1 \leq x_i \leq 0.2$  for  $i = 1, \dots, D$ , rotation angle  $\theta = 0.2$ , and optimum equal to  $f(\mathbf{x}^*) = 0$  at  $\mathbf{x}^* = \{0.1, \dots, 0.1\}^T$ .

The fidelity levels can be defined following the work of Wang et al. [14], where a resolution error is defined as follows

$$e_r(\mathbf{z}, \phi) = \sum_{i=1}^D a(\phi) \cos^2(w(\phi)z_i + b(\phi) + \pi) \quad (6-14)$$

with  $a(\phi) = \Theta(\phi)$ ,  $w(\phi) = 10\pi\Theta(\phi)$ ,  $b(\phi) = 0.5\pi\Theta(\phi)$ , and  $\Theta(\phi) = 1 - 0.0001\phi$ . The fidelity levels are thus described as follows and depicted in Figure 6-3.

$$f_i(\mathbf{z}, \phi_i) = f_1(\mathbf{z}) + e_r(\mathbf{z}, \phi_i) \quad \text{for} \quad i = 1, 2, 3 \quad (6-15)$$

with  $\phi_1 = 10000$  (high-fidelity),  $\phi_2 = 5000$  (medium-fidelity), and  $\phi_3 = 2500$  (low-fidelity).

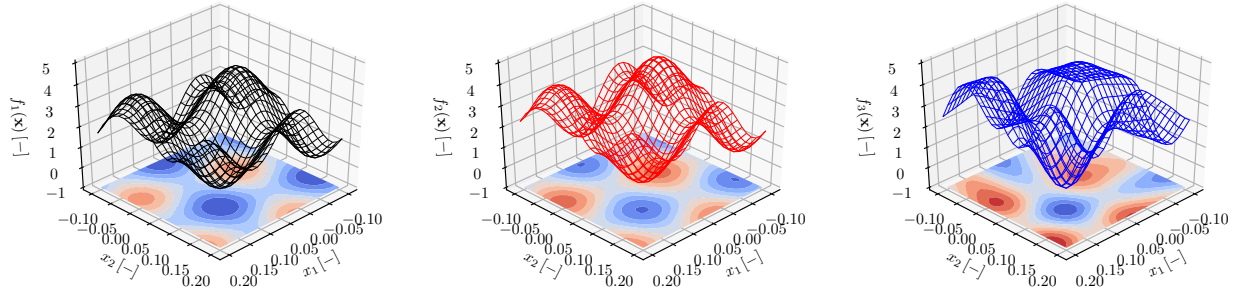


**Figure 6-3: Shifted-rotated Rastrigin Function: from Left to Right,  $f_1$  (highest-fidelity),  $f_2$ , and  $f_3$  (lowest-fidelity).**

## 6.2.4 Heterogeneous Function

This problem employs heterogeneous non-polynomial analytic functions defined on unit hypercubes ( $0 \leq x_i \leq 1$  for  $i = 1, \dots, D$ ) in one, two (taken from [32]), and three dimensions, with low-fidelity functions obtained by using linear additive and multiplicative bridge functions. The benchmark reads

$$\begin{cases} f_1(x) &= \sin[30(x - 0.9)^4] \cos[2(x - 0.9)] + (x - 0.9)/2 \\ f_2(x) &= (f_1(x) - 1.0 + x)/(1.0 + 0.25x) \end{cases} \quad \text{for } D = 1 \quad (6-16)$$

$$\begin{cases} f_1(\mathbf{x}) &= \sin[21(x_1 - 0.9)^4] \cos[2(x_1 - 0.9)] + (x_1 - 0.7)/2 + \sum_{i=2}^D i x_i^i \sin\left(\prod_{j=1}^i x_j\right) \\ f_2(\mathbf{x}) &= (f_1(\mathbf{x}) - 2.0 + \sum_{i=1}^D x_i)/(5.0 + \sum_{i=1}^2 0.25 i x_i - \sum_{D>2} 0.25 i x_i) \end{cases} \quad \text{for } D \geq 2 \quad (6-17)$$

The one- and two-dimensional functions are displayed together with their low-fidelity approximations in Figures 6-4 and 6-5. The optimum for  $D = 1$  is equal to  $f(\mathbf{x}^*) = -0.625$  at  $\mathbf{x}^* = 0.27550$ , while for  $D = 2, 3$  it is equal to  $f(\mathbf{x}^*) = -0.5627123$  at  $\mathbf{x}^* = \{0, \dots, 0\}^T$ .

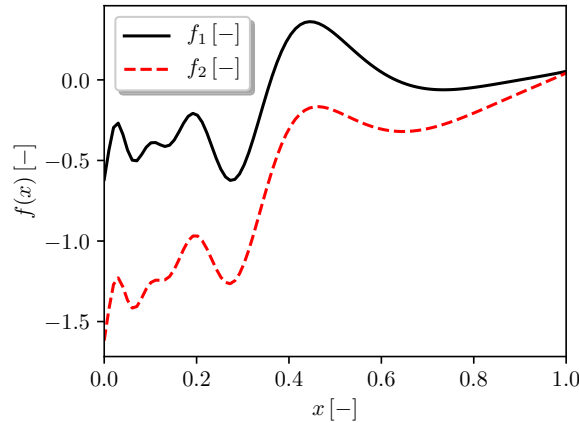


**Figure 6-4: Heterogeneous Function ( $D = 1$ ).**

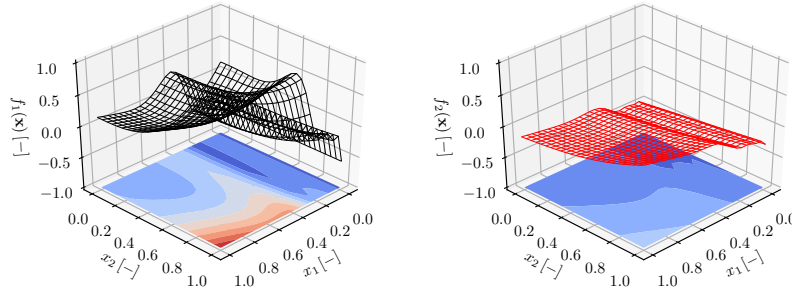


**Figure 6-5: Heterogeneous Function ( $D = 2$ ).**

## 6.2.5 Coupled Spring-Mass System

The proposed problem represents a general coupled spring-mass system. Consider two point masses  $m_1$  and  $m_2$  concentrated at their center of gravity and attached to each other by three springs, that operate according to Hooke's law with constants  $k_1$ ,  $k_2$ , and  $k_3$ . The mass of each spring is negligible and they restore after compression and extension. The masses can slide along a frictionless horizontal surface, while the first and last springs ( $k_1$  and  $k_3$ ) are attached to fixed walls, as shown in Figure 6-6.

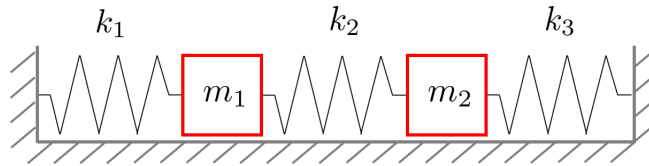


**Figure 6-6: Coupled Spring-masses System.**

Assume that  $\xi_1(t)$  and  $\xi_2(t)$  denote the mass positions along the horizontal surface, measured from their equilibrium positions, positive right and negative left. The equations of motion are given by the following system

$$\begin{cases} m_1 \ddot{\xi}_1(t) &= -k_1 \xi_1(t) + k_2 [\xi_2(t) - \xi_1(t)] \\ m_2 \ddot{\xi}_2(t) &= -k_2 [\xi_2(t) - \xi_1(t)] - k_3 \xi_2(t) \end{cases} \quad (6-18)$$

Eq. (6-18) can be written as a second-order vector-matrix system

$$\mathbf{M} \ddot{\boldsymbol{\xi}}(t) = \mathbf{K} \boldsymbol{\xi}(t) \quad (6-19)$$

where the displacement  $\boldsymbol{\xi}$ , mass matrix  $\mathbf{M}$  and stiffness matrix  $\mathbf{K}$  are defined by the following formulae

$$\boldsymbol{\xi} = \begin{Bmatrix} \xi_1 \\ \xi_2 \end{Bmatrix}, \quad \mathbf{M} = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} -k_1 - k_2 & k_2 \\ k_2 & -k_2 - k_3 \end{bmatrix} \quad (6-20)$$

This is a constant-coefficient homogeneous system of second-order ODEs the solution of which is given by

$$\boldsymbol{\xi}(t) = \sum_{i=1}^2 [a_i \cos(\omega_i t) + b_i \sin(\omega_i t)] \mathbf{z}_i \quad (6-21)$$

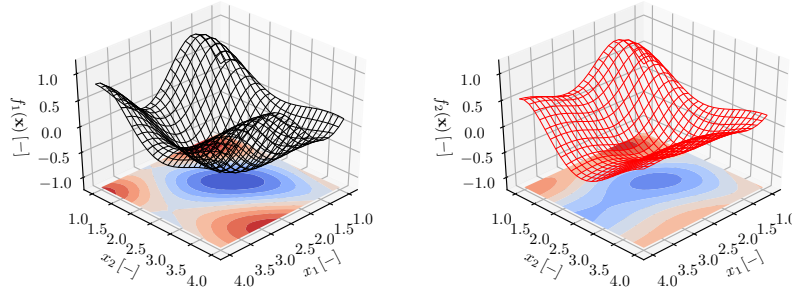


**Figure 6-7: Spring-Mass System ( $D = 2$ , springs only).**

where  $\omega_i = \sqrt{-\lambda_i}$  and  $\lambda_i$  are the eigenvalues of the matrix  $\mathbf{M}^{-1}\mathbf{K}$  and  $\mathbf{z}_i$  are the corresponding eigenvectors. The constants  $a_i$  and  $b_i$  are determined by the initial conditions  $\boldsymbol{\xi}(t = 0) = \boldsymbol{\xi}_0$  and  $\dot{\boldsymbol{\xi}}(t = 0) = \dot{\boldsymbol{\xi}}_0$ .

Converting Eq. 6-19 into a system of first-order ODEs and using the fourth-order accurate Runge-Kutta time-marching method yields a multi-fidelity analysis problem by varying the time-step size  $\Delta t$ . The proposed benchmark uses the initial conditions  $\boldsymbol{\xi}_0 = \{1 \ 0\}^T$  and  $\dot{\boldsymbol{\xi}}_0 = \{0 \ 0\}^T$  with two fidelity levels, defined by the time-step size, specifically equal to  $\Delta t = 0.01$  and  $0.6$ . Two tests are proposed, considering the position of the first mass  $\xi_1$  at time  $t = 6$  as the objective function: (MF5.1) springs  $k_1$  and  $k_2$  are the independent input variables with  $1 \leq (k_1, k_2) \leq 4$  and  $k_3 = k_1$ , while the masses are constant ( $m_1 = m_2 = 1$ ), the optimum is equal to  $f(\mathbf{x}^*) = -1$  at  $\mathbf{x}^* = \{2.467401, 2.193245\}^T$ ; (MF5.2) springs  $k_1$  and  $k_2$  and masses  $m_1$  and  $m_2$  are the independent input variables with  $1 \leq (k_1, k_2, m_1, m_2) \leq 4$  and  $k_3 = k_1$ , the optimum is equal to  $f(\mathbf{x}^*) = -1$  at  $\mathbf{x}^* = \{1.000000, 3.946018, 4.000000, 3.286277\}^T$ . The two-dimensional (springs only) design space is shown in Figure 6-7.

### 6.2.6 Paciorek Function

The Paciorek equation [33], which has localized and multi-modality properties, is considered and a normally distributed random noise parameter is added to the high- and low-fidelity equation to model the noise. The Paciorek function with noise term is defined as follows

$$f_1(\mathbf{x}) = \sin\left(\prod_{i=1}^D x_i\right)^{-1} + \text{rand.normal}(0, \alpha_1) \quad (6-22)$$

$$f_2(\mathbf{x}) = f_1(\mathbf{x}) - 9A^2 \cos\left(\prod_{i=1}^D x_i\right)^{-1} + \text{rand.normal}(0, \alpha_2) \quad (6-23)$$

$A$  is a parameter that models the error among the fidelities and can vary between 0 and 1, when  $A = 0$  the low- and high-fidelity equations are the same, while the change between the low-fidelity and high-fidelity models increases as the value of  $A$  increases. In this study, it is set as  $A = 0.5$ , while  $\alpha_i$  is the coefficient that defines the noise level. The benchmark is defined considering the variable ranges as  $0.3 \leq x_i \leq 1.0$  for  $\forall i$ . For the high-fidelity case, a noise level is added to the equation which corresponds to approximately 5% of the response interval, as  $\alpha_1 = 0.0125$ . Accordingly, for the low-fidelity case, a much higher level of noise is added to the equation corresponding to 10% of the response interval,  $\alpha_2 = 0.075$ . The resulting response surfaces are shown in Fig. 6-8.

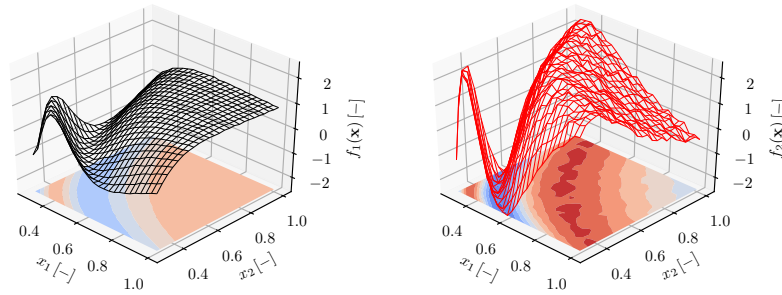


**Figure 6-8: Paciorek Function with Noise:  $f_1$  (left) and  $f_2$  (right).**

## 6.3 SETUP OF THE NUMERICAL EXPERIMENTS

Table 6-2 summarizes the recommended setup for the assessment and comparison of the performance of multi-fidelity methods. The following subsections provide an overview of the rationale and the criteria motivating the recommendations along with additional complementary suggestions. The overview encompasses criteria for the assignment of the evaluation costs to the different fidelity levels as a fraction of the computational expense of the highest fidelity function, criteria for the initialization of the search, and criteria to terminate the search.

**Table 6-2: Experiments Setup Summary.**

| Function                  | Benchmark ID | $D$ | Budget | Fidelity cost |            |            |            |
|---------------------------|--------------|-----|--------|---------------|------------|------------|------------|
|                           |              |     |        | $f_1$         | $f_2$      | $f_3$      | $f_4$      |
| Forrester                 | MF1.1        | 1   | 100    | 1.00000E-0    | 5.00000E-1 | 1.00000E-1 | 5.00000E-2 |
| Jump Forrester            | MF1.2        | 1   | 100    | 1.00000E-0    | 2.00000E-1 | -          | -          |
| Rosenbrock                | MF2.1        | 2   | 200    | 1.00000E-0    | 5.00000E-1 | 1.00000E-1 | -          |
|                           | MF2.2        | 5   | 500    | 1.00000E-0    | 5.00000E-1 | 1.00000E-1 | -          |
|                           | MF2.3        | 10  | 1000   | 1.00000E-0    | 5.00000E-1 | 1.00000E-1 | -          |
| Shifted-rotated Rastrigin | MF3.1        | 2   | 200    | 1.00000E-0    | 6.25000E-2 | 3.90625E-3 | -          |
|                           | MF3.2        | 5   | 500    | 1.00000E-0    | 6.25000E-2 | 3.90625E-3 | -          |
|                           | MF3.3        | 10  | 1000   | 1.00000E-0    | 6.25000E-2 | 3.90625E-3 | -          |
| Heterogeneous             | MF4.1        | 1   | 100    | 1.00000E-0    | 2.00000E-1 | -          | -          |
|                           | MF4.2        | 2   | 200    | 1.00000E-0    | 2.00000E-1 | -          | -          |
|                           | MF4.3        | 3   | 300    | 1.00000E-0    | 2.00000E-1 | -          | -          |
| Springs                   | MF5.1        | 2   | 200    | 1.00000E-0    | 1.66667E-2 | -          | -          |
| Springs-masses            | MF5.2        | 4   | 400    | 1.00000E-0    | 1.66667E-2 | -          | -          |
| Paciorek                  | MF6          | 2   | 200    | 1.00000E-0    | 2.00000E-1 | -          | -          |

### 6.3.1 Fidelity Cost Assignment Criteria

The  $f_1(\mathbf{x})$  functions are given the unitary cost  $\lambda_1 = 1$ , while their lower fidelity representations are assigned costs values as fractions of the  $f_1(\mathbf{x})$  cost. Table 6-3 proposes a set of cost assignments for the different fidelity representations of each of the benchmark problems. The values indicated for the shifted-rotated Rastrigin are

determined according to the non-linear function proposed by Wang et al. [14] for the allocation of cost values to an arbitrary number of fidelities:

$$\lambda_l = \left(1/2^{l-1}\right)^4 \quad \forall l \geq 1 \quad (6-24)$$

where higher values of the integer  $l$  indicate representations of the objective functions at progressively lower levels or fidelity. Differently, the values indicated for the Forrester, the Rosenbrock, the Heterogeneous, and the Spring-Mass problems are driven by the shared experience within the AVT-331 research task group.

### **6.3.2 Initialization Criteria**

To assure a fair comparison across different families of multi-fidelity methods, statistics over a set of different starting samples/points is highly recommended. The cardinality and composition of the initial set are not constrained, but determined for or by the specific multi-fidelity method. This approach is preferred over the definition of specific initial samples to be used. The performance assessment would then consider statistics over the set of all the convergence histories.

### **6.3.3 Termination Criteria**

Real-world design problems are constrained by the limited time and computing resources available to conduct analysis, search, and optimization of the alternative candidate solutions. This motivates the choice to recommend conducting the experiments at given computational budgets assigned to the overall modeling and optimization task, rather than prescribing the maximum number of allowed iterations. The termination condition is reached when no computational budget is left. Table 6-3 indicates the computational budget assigned to each experiment in terms of the equivalent number of evaluations of the high-fidelity representation.

## **6.4 RESULTS**

Table 6-4 summarizes the main characteristics of each method used. Specifically, two surrogate-lean versus seventeen surrogate-rich methods are compared. Among the latter group, the majority (14 out of 17) are GP-based, whereas 12 are online (including an adaptive sampling approach).

**Table 6-3: Summary of Multi-fidelity Benchmark Values for the Evaluation of the Goal-sensitive Metrics.**

| Function                  | Benchmark ID | $D$ | $\mathbf{x}^*$                                   | $f_{\min}$ | $f_{\max}$ |
|---------------------------|--------------|-----|--------------------------------------------------|------------|------------|
| Forrester                 | MF1.1        | 1   | 7.5752E-1                                        | -6.0207E-0 | 1.5830E+1  |
| Jump Forrester            | MF1.2        | 1   | 1.4260E-1                                        | -9.8630E-1 | 2.5830E+1  |
| Rosenbrock                | MF2.1        | 2   | $\{1.0000, 1.0000\}^T$                           | 0.000E-0   | 3.6090E+3  |
|                           | MF2.2        | 5   | $\{1.0000, \dots, 1.0000\}^T$                    | 0.000E-0   | 1.4436E+4  |
|                           | MF2.3        | 10  | $\{1.0000, \dots, 1.0000\}^T$                    | 0.000E-0   | 3.2481E+4  |
| Shifted-rotated Rastrigin | MF3.1        | 2   | $\{0.1000, 0.1000\}^T$                           | 0.000E-0   | 4.0200E-0  |
|                           | MF3.2        | 5   | $\{0.1000, \dots, 0.1000\}^T$                    | 0.000E-0   | 1.0050E+1  |
|                           | MF3.3        | 10  | $\{0.1000, \dots, 0.1000\}^T$                    | 0.000E-0   | 2.0100E+1  |
| Heterogeneous             | MF4.1        | 1   | 2.7550E-1                                        | -6.2500E-1 | 3.6151E-1  |
|                           | MF4.2        | 2   | $\forall x_1 = 0.000E-0$                         | -5.6271E-1 | 1.8350E-0  |
|                           | MF4.3        | 3   | $\forall x_1 = 0.000E-0$                         | -5.6271E-1 | 4.3594E-0  |
| Springs                   | MF5.1        | 2   | $\{2.4674, 2.1932\}^T$                           | -1.0000E-0 | 1.0000E-0  |
| Springs-masses            | MF5.2        | 4   | $\{1.0000, 3.9460, 4.0000, 3.2863\}^T$           | -1.0000E-0 | 1.0000E-0  |
| Paciorek                  | MF6          | 2   | $\forall x_1 x_2 = 2/[(3+j)\pi]$ with $j = 0, 4$ | -1.0000E-0 | 1.0000E-0  |

**Table 6-4: Multi-fidelity Methods Summary.**

| Classification | Acronym      | Surrogate*                                     | Off/Online | Adaptive sampling <sup>†</sup> | DoE <sup>‡</sup> (D/S) | Optimization algorithm <sup>§</sup> (D/S) | Derivative <sup>††</sup> informed/free | Organization | Notes                                                        | Refs.     |
|----------------|--------------|------------------------------------------------|------------|--------------------------------|------------------------|-------------------------------------------|----------------------------------------|--------------|--------------------------------------------------------------|-----------|
| Surrogate lean | BFGS         | –                                              | Online     | –                              | LHS (S)                | BFGS (D)                                  | Informed                               | AFRL         | Highest- and lowest-fidelity only used                       | [34]      |
|                | AMOGA        | –                                              | Online     | –                              | RBP (S)                | AMOGA (S)                                 | Free                                   | CIRA         | First (highest) and second-fidelity levels only used         | [35]      |
| Surrogate rich | GP           | Auto-regressive Gaussian process               | Offline    | –                              | HSOS (S)               | MS_SQP (D)                                | Informed                               | ITU          | Highest- and lowest-fidelity only used                       | [36]      |
|                | GP-PI        | Auto-regressive Gaussian process               | Online     | PI                             | LHS (S)                | GA+SQP (S)                                | Informed                               | PoliTo       | Highest- and lowest-fidelity only used                       | [37]      |
|                | GP-EI        | Auto-regressive Gaussian process               | Online     | EI                             | LHS (S)                | GA+SQP (S)                                | Informed                               | PoliTo       | Highest- and lowest-fidelity only used                       | [38]      |
|                | GP-EIM       | Auto-regressive Gaussian process               | Online     | EIM                            | HSOS (S)               | MS_SQP (D)                                | Informed                               | ITU          | Highest- and lowest-fidelity only used                       | [36]      |
|                | GP-MVES      | Auto-regressive Gaussian process               | Online     | MVES                           | LHS (S)                | GA+SQP (S)                                | Free                                   | PoliTo       | Highest- and lowest-fidelity only used                       | [39]      |
|                | GP-MSES      | Auto-regressive Gaussian process               | Online     | MSES                           | LHS (S)                | GA+SQP (S)                                | Informed                               | EMUKIT       | Highest- and lowest-fidelity only used; tests by ITU, PoliTo | [40]      |
|                | GP-LCB       | Auto-regressive Gaussian process               | Online     | LCB                            | CCF* (D)               | DPSO+LS (D)                               | Free                                   | CNR-INM      | Used with interpolation                                      | [30]      |
|                | NGP          | Nonlinear auto-regressive Gaussian process     | Offline    | –                              | HSOS (S)               | MS_SQP (D)                                | Informed                               | ITU          | Highest- and lowest-fidelity only used                       | [41]      |
|                | KG           | Kriging                                        | Offline    | –                              | HSS (D)                | GA (S)                                    | Free                                   | DAKOTA       | Tests by CNR-INM                                             | [42]      |
|                | KG-GLRD      | Kriging                                        | Online     | GLRD                           | FFD (D)                | –                                         | –                                      | AFRL         | Used for global accuracy (NRMSE)                             | [43]      |
|                | KG-EC        | Kriging                                        | Online     | EC                             | LHS (S)                | GA+SQP (S)                                | Informed                               | UoS          | –                                                            | [44]      |
|                | KG-AEI       | Kriging                                        | Online     | AEI                            | CCF (D)                | L-BFGS-B (D)                              | Informed                               | MARIN        | Highest- and lowest-fidelity only used                       | [45]      |
|                | RCK-EI       | Regressive co-Kriging                          | Online     | EI                             | LHS (S)                | GA (S)                                    | Free                                   | AFRL         | Highest- and lowest-fidelity only used                       | [46]      |
|                | SRBF-LCB     | Stochastic radial basis functions              | Online     | LCB                            | CCF* (D)               | DPSO+LS (D)                               | Free                                   | CNR-INM      | Regressive for MF6 only                                      | [5], [47] |
|                | PCE          | Polynomial chaos expansion                     | Offline    | –                              | SGP (D)                | –                                         | –                                      | AFRL         | Used for global accuracy (NRMSE)                             | [48]      |
|                | SPCE         | Sparse polynomial chaos expansion              | Offline    | –                              | SGP (D)                | –                                         | –                                      | AFRL         | Used for global accuracy (NRMSE)                             | [49]      |
|                | SPCE_KG-GLRD | Sparse polynomial chaos expansion with Kriging | Online     | GLRD                           | FFD (D)                | –                                         | –                                      | AFRL         | Used for global accuracy (NRMSE)                             | [50]      |

Note: D/S = Deterministic/Stochastic behavior; metric medians and standard deviations are given for stochastic methods.  
<sup>†</sup> AEI = augmented expected improvement, EC = evolution control, EI = expected improvement, EIM = expected improvement + function minimum, GLRD = global/local regression discrepancy, LCB = lower confidence bounding, MVES = maximum value entropy search, MSES = multi-source entropy search, PI = probability of improvement.  
<sup>‡</sup> For online methods is the initial DoE: CCF = face-centered central composite design (\* only on the lowest fidelity, center of the domain for the others), FFD = factorial fractional design, HSS = Hammersley sequence sampling, HSOS = Halton sampling with Owen-scramble LHS = Latin hypercube sampling, RBP = random baseline perturbation, SGP = sparse Gauss-Patterson grid.  
<sup>§</sup> Optimization algorithm used to solve Eq. 6-1; for surrogate-rich methods, the algorithm is applied on surrogate prediction.  
<sup>††</sup> Use of derivatives by the optimization algorithm.  
\* None of the surrogates use derivatives to construct the function approximation.

## L1 BENCHMARK DEFINITIONS AND RESULTS

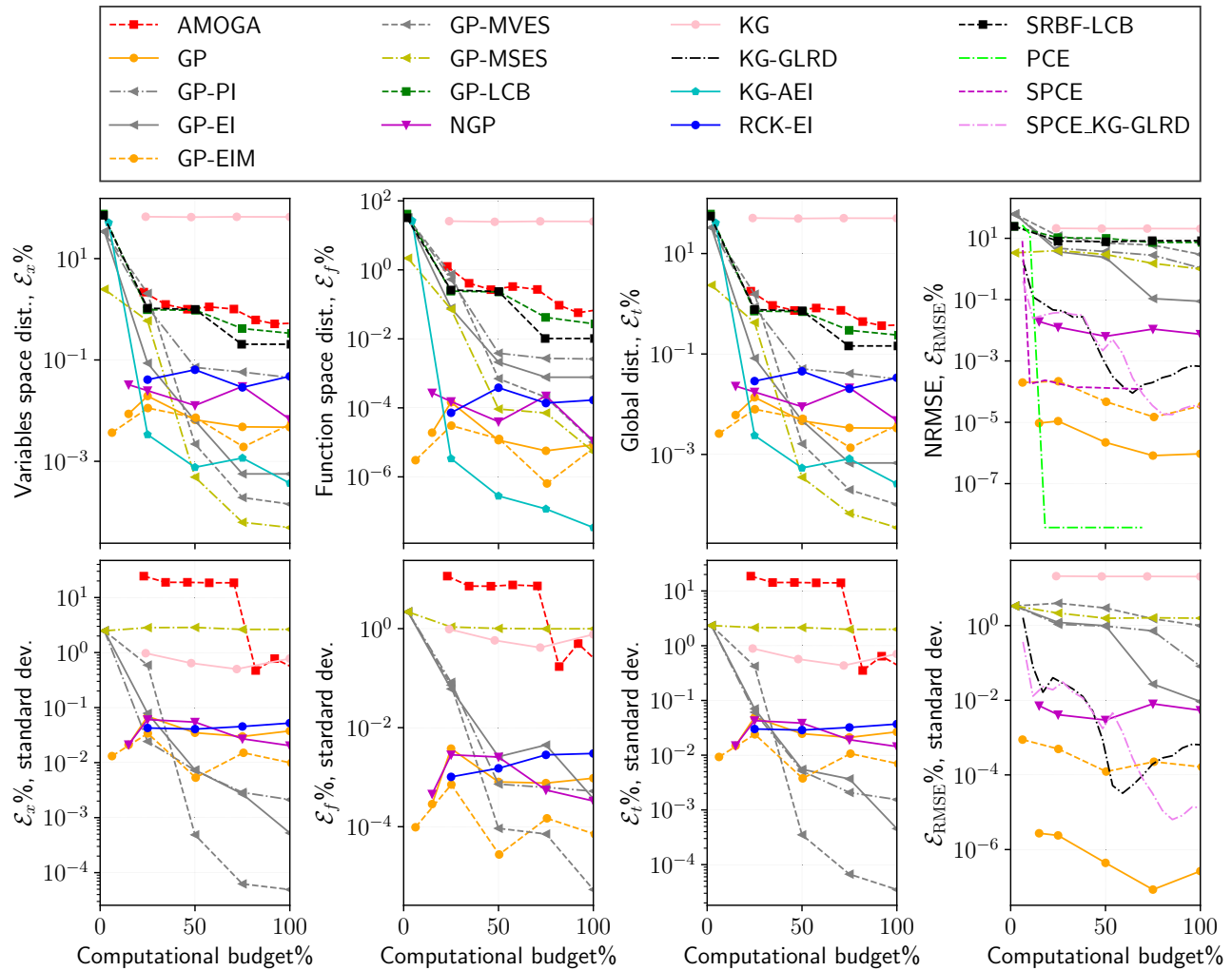


**Figure 6-9: MF1.1 Metrics Convergence.**

While most methods were used to find the global optimum in the design space and the NRMSE is just computed as a byproduct, the KG-GLRD, SPCE-KG-GLRD, PCE, and SPCE methods were only used to build globally accurate surrogates. Thus, a direct comparison in the NRMSE performance is not necessarily fair. Note that all methods are intended to be an MF approach and no comparison with their single-fidelity (high-fidelity) counterpart is given since it is beyond the scope of this work.

In the following, the results for each benchmark problem will be shown and discussed. Note that lines are plotted between markers as a visual aid, but do not necessarily reflect performance at intermediate points. For the statistical assessment of stochastic methods, ten runs are conducted and results are provided in terms of the median (or mean if the median was not provided by a group) and standard deviation.

The x-axis always shows the percentage (0 to 100) of the total computational budget given in Table 6-2. The y-axis shows the percentage error in the four metrics (from left to right: variable, function, and global distances as well as NRMSE) discussed previously. The top row shows the median (or mean) and the bottom row displays the standard deviation. Obviously, no statistics are given for deterministic approaches.

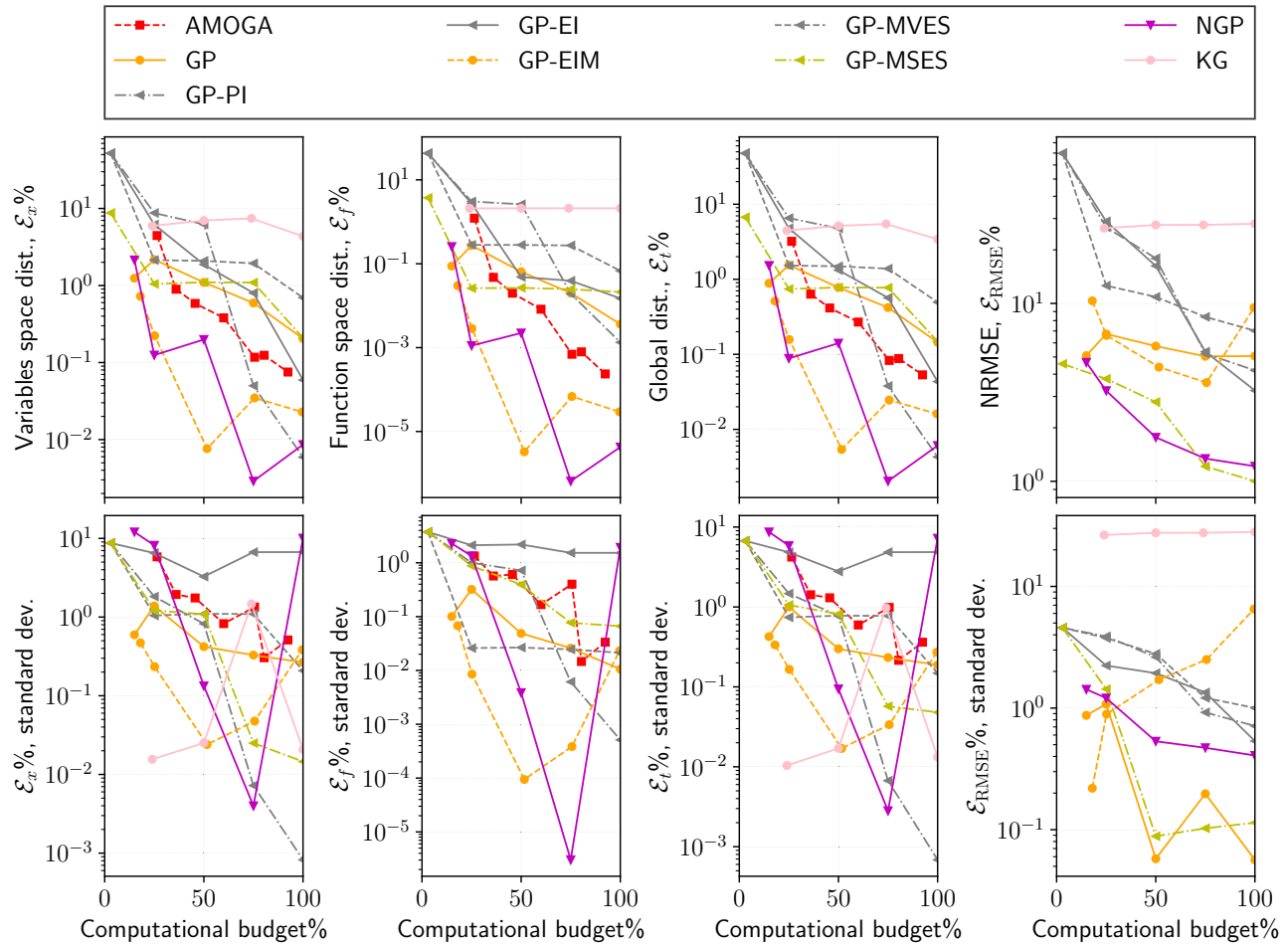


Figure 6-10: MF1.2 Metrics Convergence.

Before discussing benchmark by benchmark, some comments are required to better understand the results of the BFGS. Since this method is gradient-based only local optima can be reliably found which is not a problem if the function is not multi-modal (such as Rosenbrock in 2D). However, when the function is multi-modal (e.g., the Rastrigin function, the spring-mass problem, or the Rosenbrock function in 5D) the acquired optimum depends on the starting point. Here, the center of the domain and nine other latin-hypercubed sampled starting points are employed and the mean and standard deviations of these ten optimization runs are displayed. If, for example, nine of these runs found the global optimum correctly but one of the runs converges to a local optimum (possibly far away in both variable and function space from the global optimum) then the statistics would be dominated by this one run and the convergence history would look fairly flat. In addition, the cost of BFGS is not controlled directly but is rather dictated by the number of required iterations and line-search evaluations in each iteration. Thus, the cost of each iteration varies and whenever a run converges it is not included in the averaging procedure any longer.

The Forrester and Jump Forrester results are shown in Figures 6-9 and 6-10, respectively. Looking at the

## L1 BENCHMARK DEFINITIONS AND RESULTS

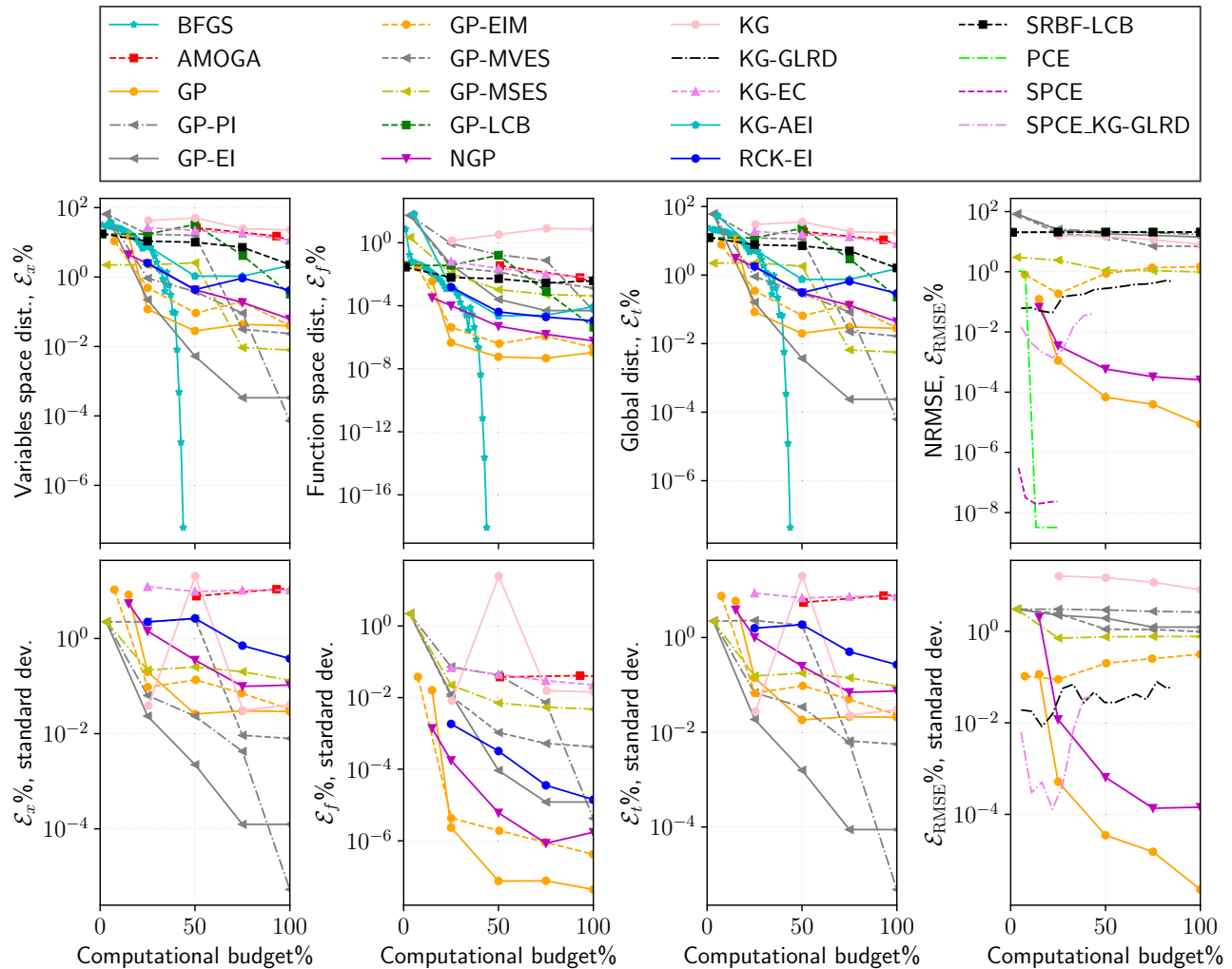


**Figure 6-11: MF2.1 Metrics Convergence.**

optimization of the Forrester function (without jump) RCK-EI and GP-EIM are performing best whereas a globally accurate surrogate of the function is best represented by PCE, GP, and again GP-EIM. For the Jump Forrester function, GP-EIM and NGP are performing relatively well for the optimization. However, none of the methods give a particularly good globally accurate model but NGP and GP-MSES are doing better than the other methods.

The Rosenbrock function results are shown in Figures 6-11 to 6-13 for 2, 5, and 10 dimensions, respectively. This function has only one optimum and thus the statistics for MF BFGS converge well compared to the surrogate-rich methods, especially in higher dimensions. Among the surrogate-rich methods, GP-EIM is performing best over all three dimensions considered here.

For a globally accurate representation, once sufficient training data is gathered to construct a fourth-order polynomial expansion, the PCE and SPCE models should match the truth function exactly with the error dropping accordingly. The full-order PCE surrogate model with HF training data alone requires 30 training points to match

the exact function in two dimensions and 2002 points in ten dimensions due to an oversampling ratio of two. For MF models some of these points may be low-fidelity and thus computational savings can be achieved. This expected behavior can be seen in the figures. In general, sparse PCEs (SPCE) outperform full-order PCEs up to a point when the penalty coefficient,  $\lambda$ , introduces error into the fit and the RMSE cannot drop further as can be seen in the top right-hand corners of Figures 6-11, 6-14, 6-17, 6-18, and 6-20. Among the non-polynomial-based methods, GP and NGP are working the best to represent the function globally as indicated by relatively low NRMSE values.

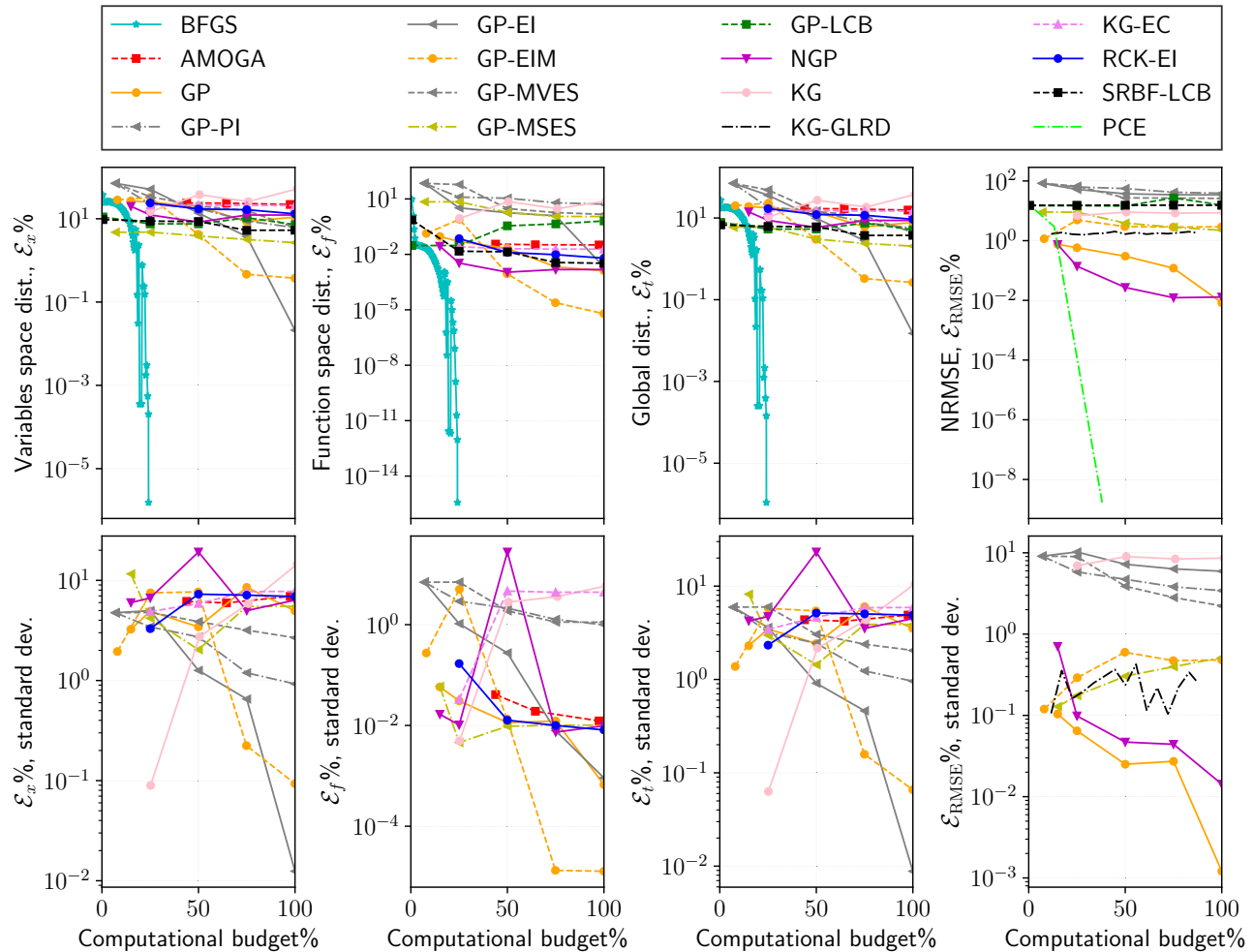


Figure 6-12: MF2.2 Metrics Convergence.

## L1 BENCHMARK DEFINITIONS AND RESULTS

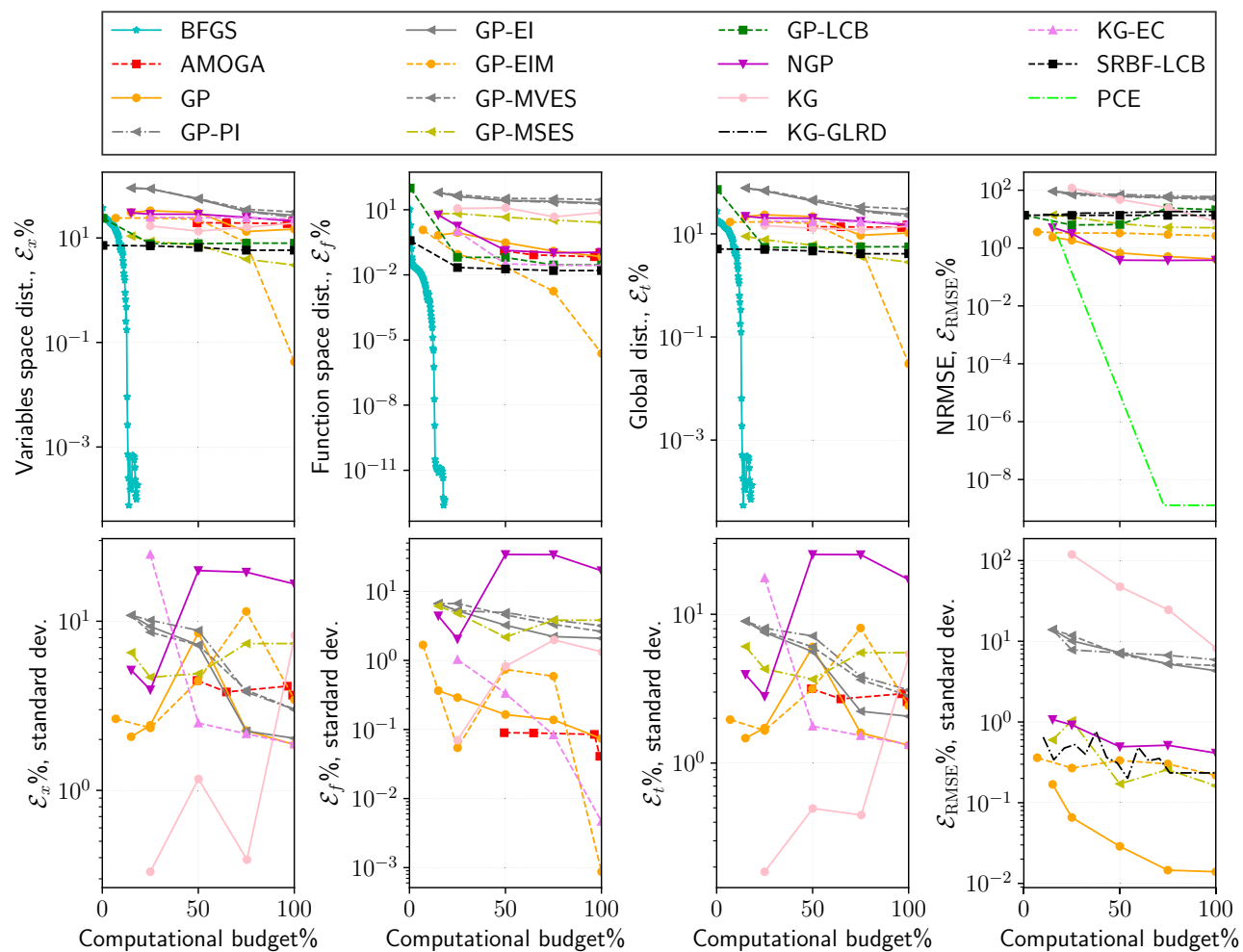


Figure 6-13: MF2.3 Metrics Convergence.

The Rastrigin function results are shown in Figures 6-14 to 6-16 for 2, 5, and 10 dimensions, respectively. For the optimization, BFGS is doing well as are GP-EIM (for all three dimensions) and KG-EC (in 5D and 10D) and RCK-EI (in 2D). For building globally accurate surrogate models the best methods vary with the dimensionality of the problem. In 2D GP, NGP and SPCE are performing best. In 5D NGP, PCE and KG-GLRD seem to work well whereas in 10D NGP and GP are best though in reality most methods are close to each other and all convergence rates are very flat.

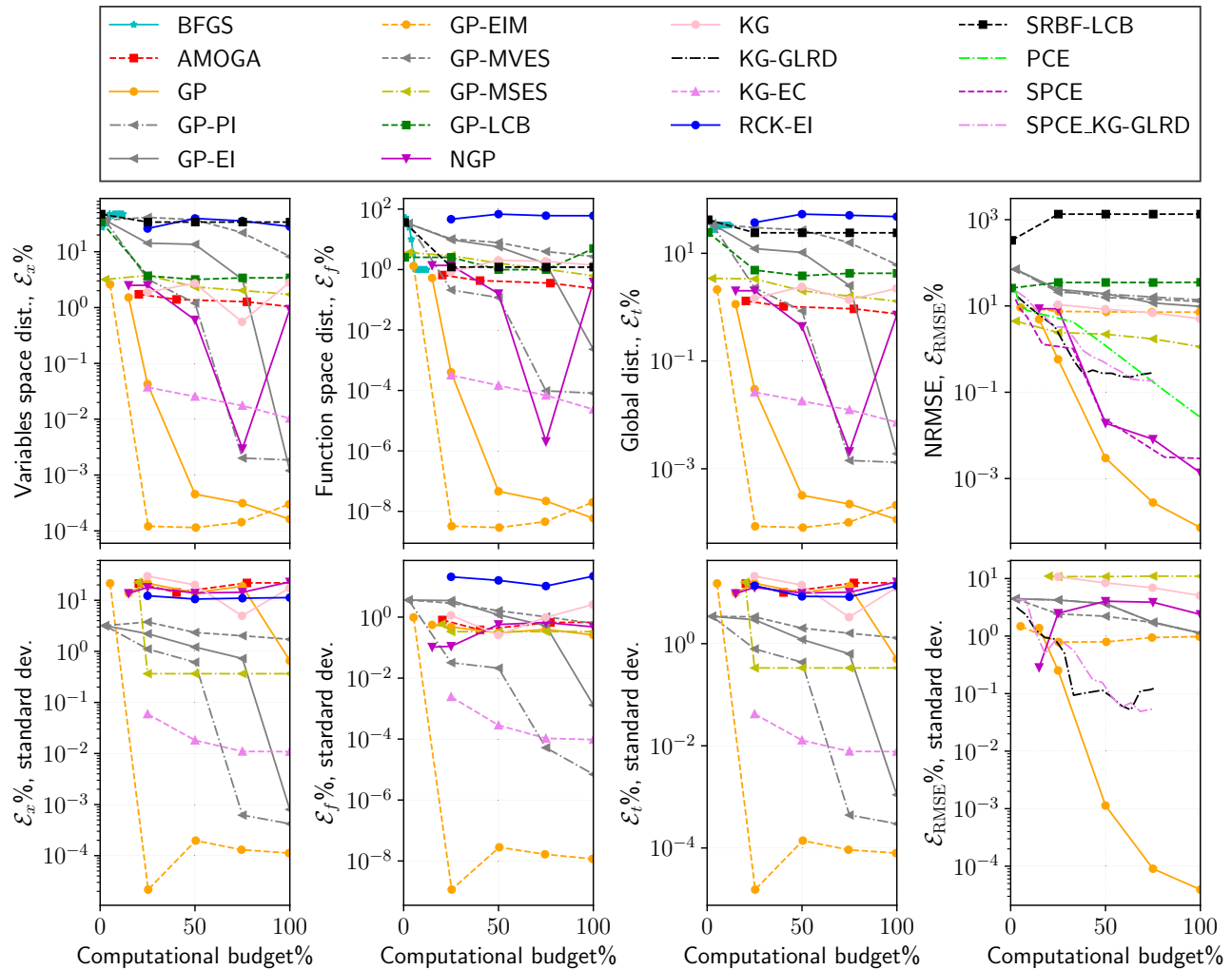


Figure 6-14: MF3.1 Metrics Convergence.

## L1 BENCHMARK DEFINITIONS AND RESULTS

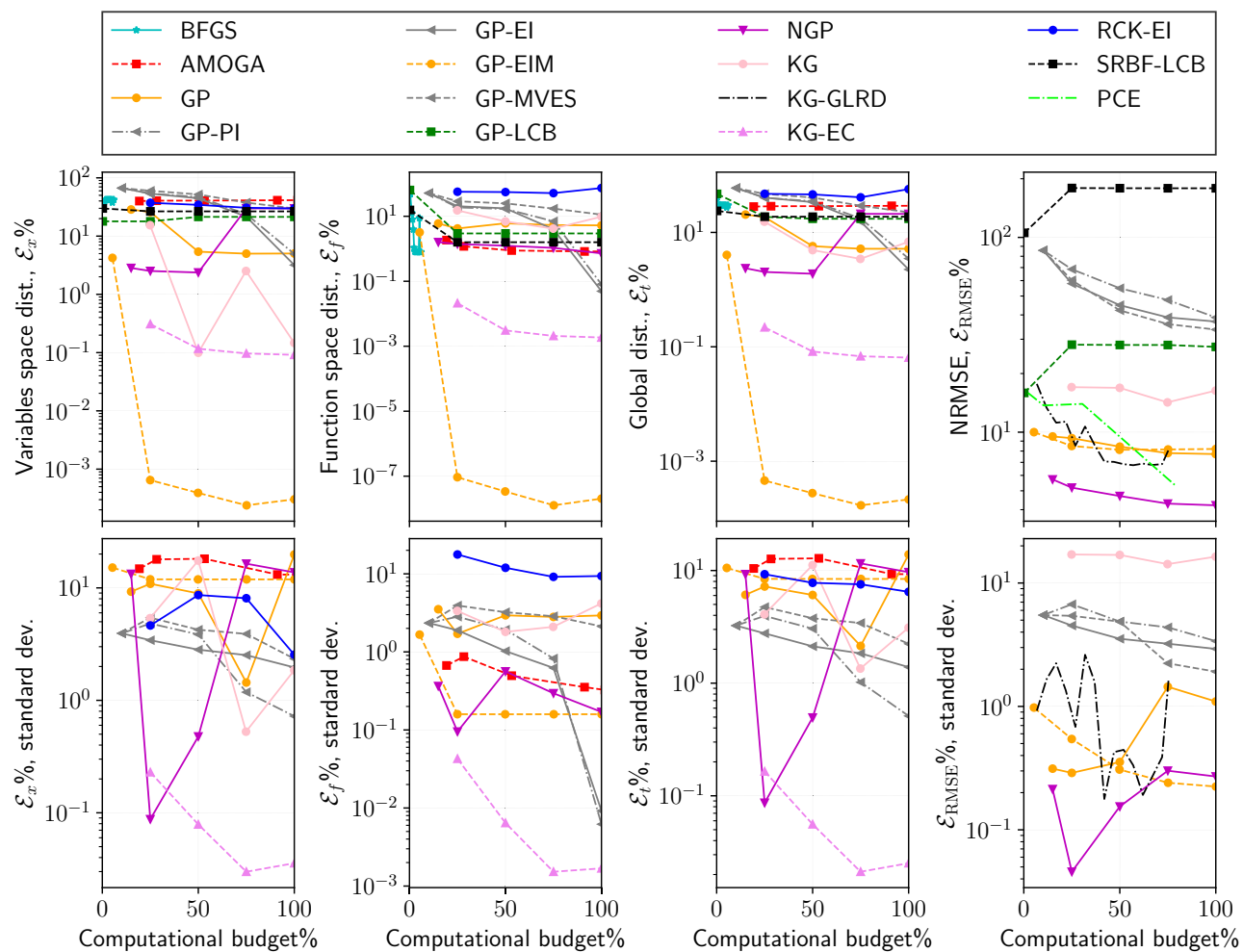


Figure 6-15: MF3.2 Metrics Convergence.

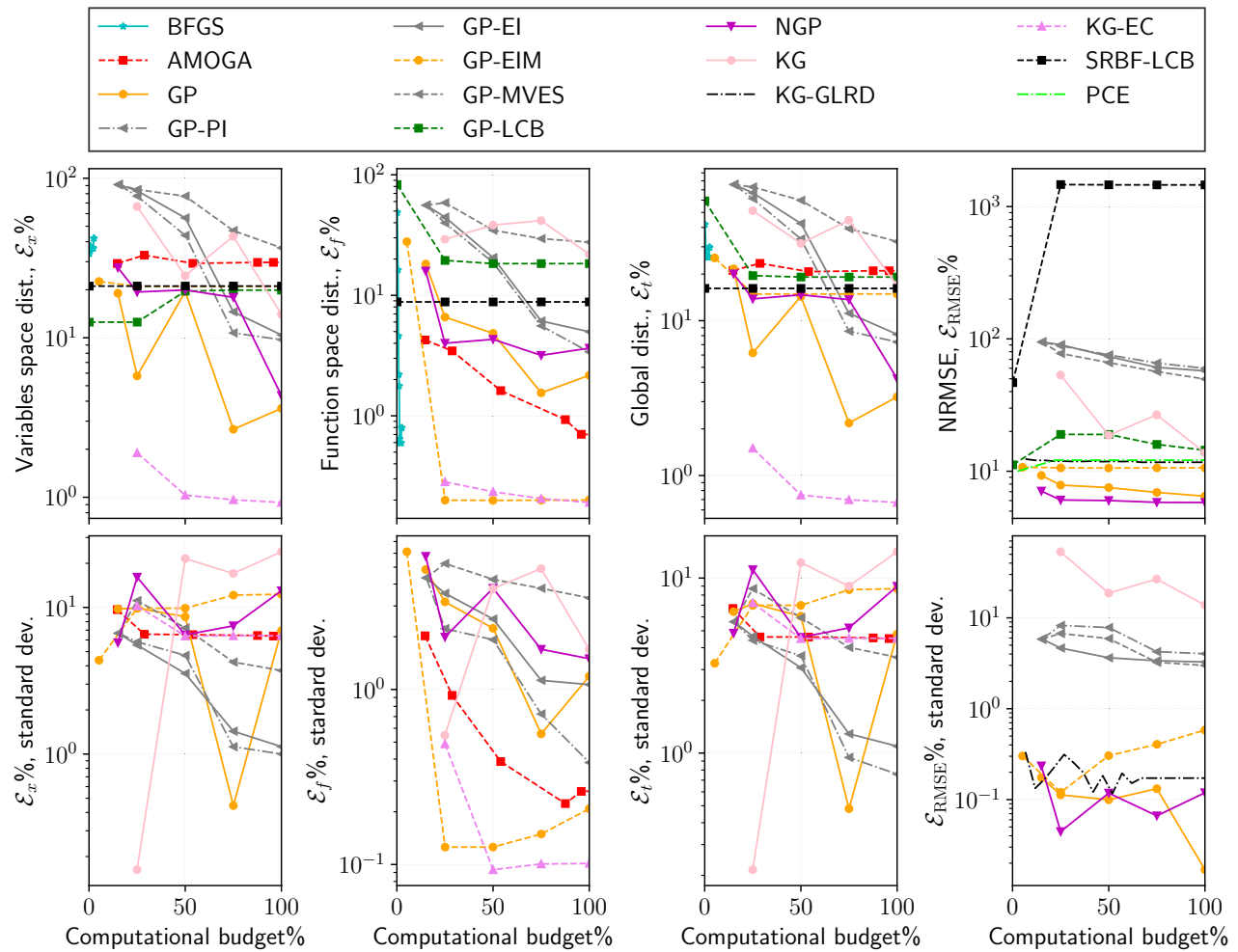


Figure 6-16: MF3.3 Metrics Convergence.

## L1 BENCHMARK DEFINITIONS AND RESULTS

The Heterogeneous function results are shown in Figures 6-17 to 6-19 for 1, 2, and 3 dimensions, respectively. The suitability of the two- and three-dimensional case for the assessment of optimization performances is somewhat questionable due to the fact that the optimum lies on a boundary and is in fact the entire boundary line in 2D. Depending on the initial DOE some methods may find that optimum right away as can be seen in Figure 6-18. For the optimization in 1D GP-EIM is best with RCK-EI and GP-EI somewhat trailing behind. In 3D SRBF-LCB and GP-LCB are performing best.

Looking at the global accuracy via NRMSE, in one dimension GP and GP-EIM are doing well, in two dimensions KG-GLRD is performing well early on, and at the end GP and GP-EIM. In three dimensions SPCE\_KG-GLRD and again GP and GP-EIM are best, however, all method's convergences are fairly flat.

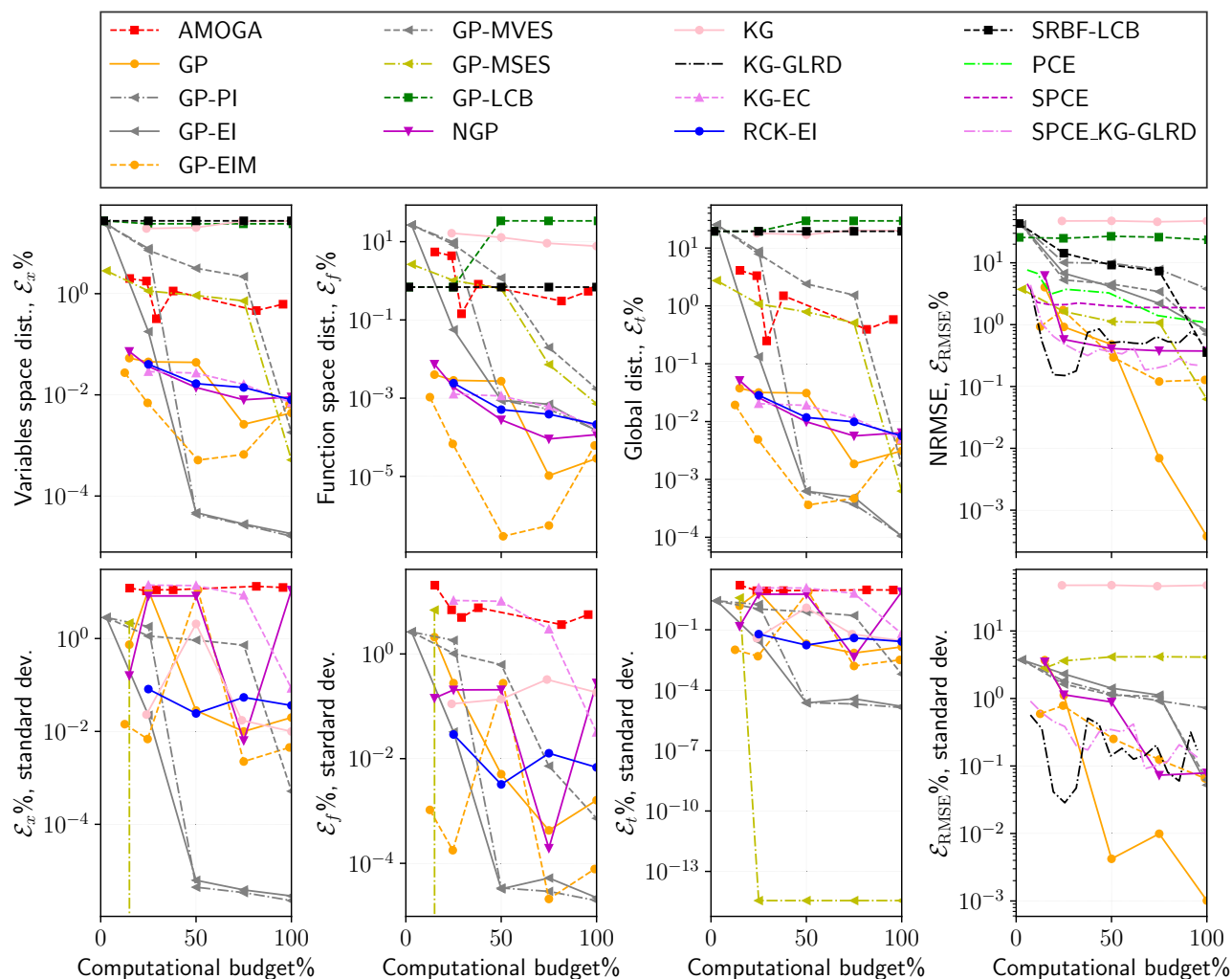


Figure 6-17: MF4.1 Metrics Convergence.

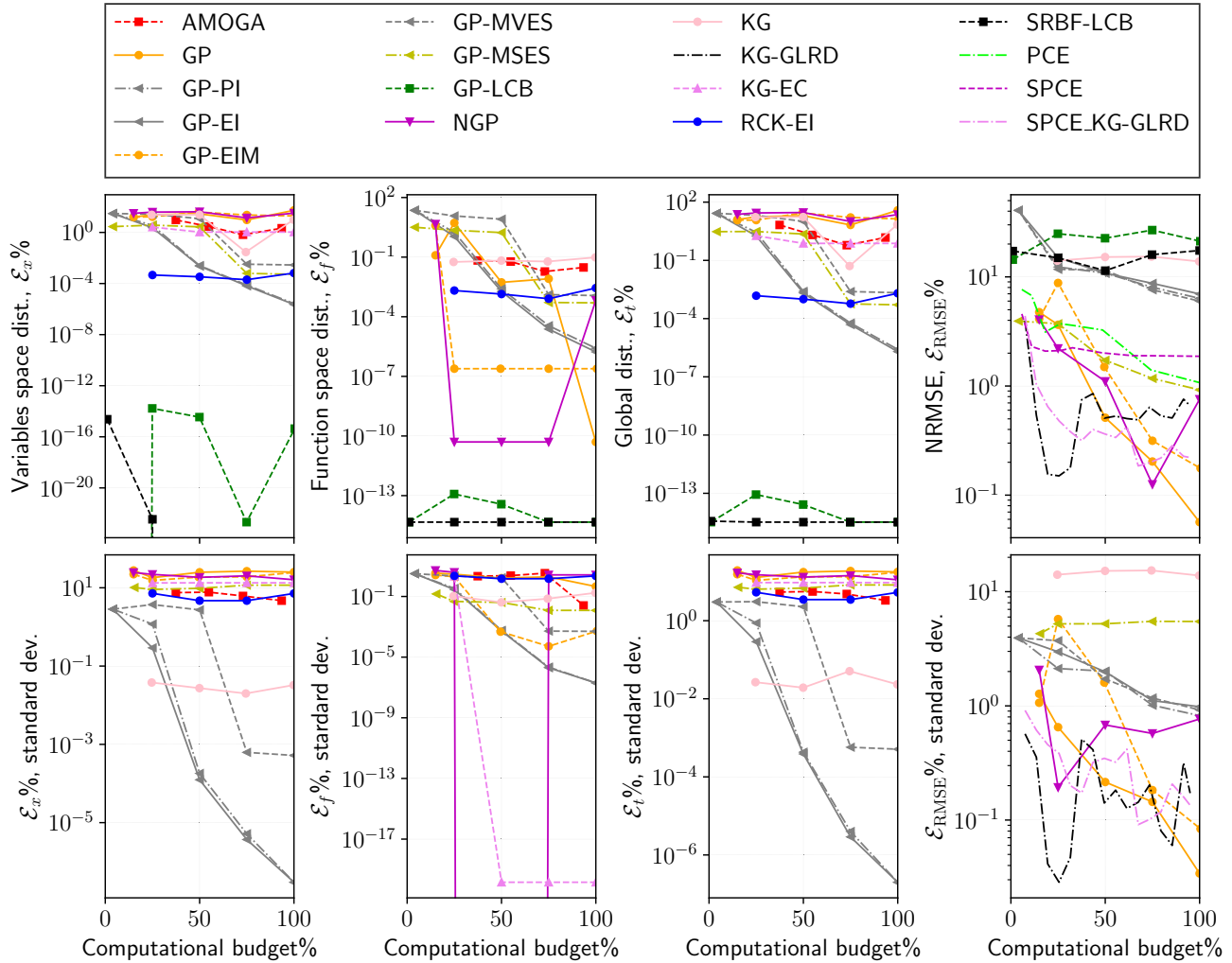


Figure 6-18: MF4.2 Metrics Convergence.

# L1 BENCHMARK DEFINITIONS AND RESULTS

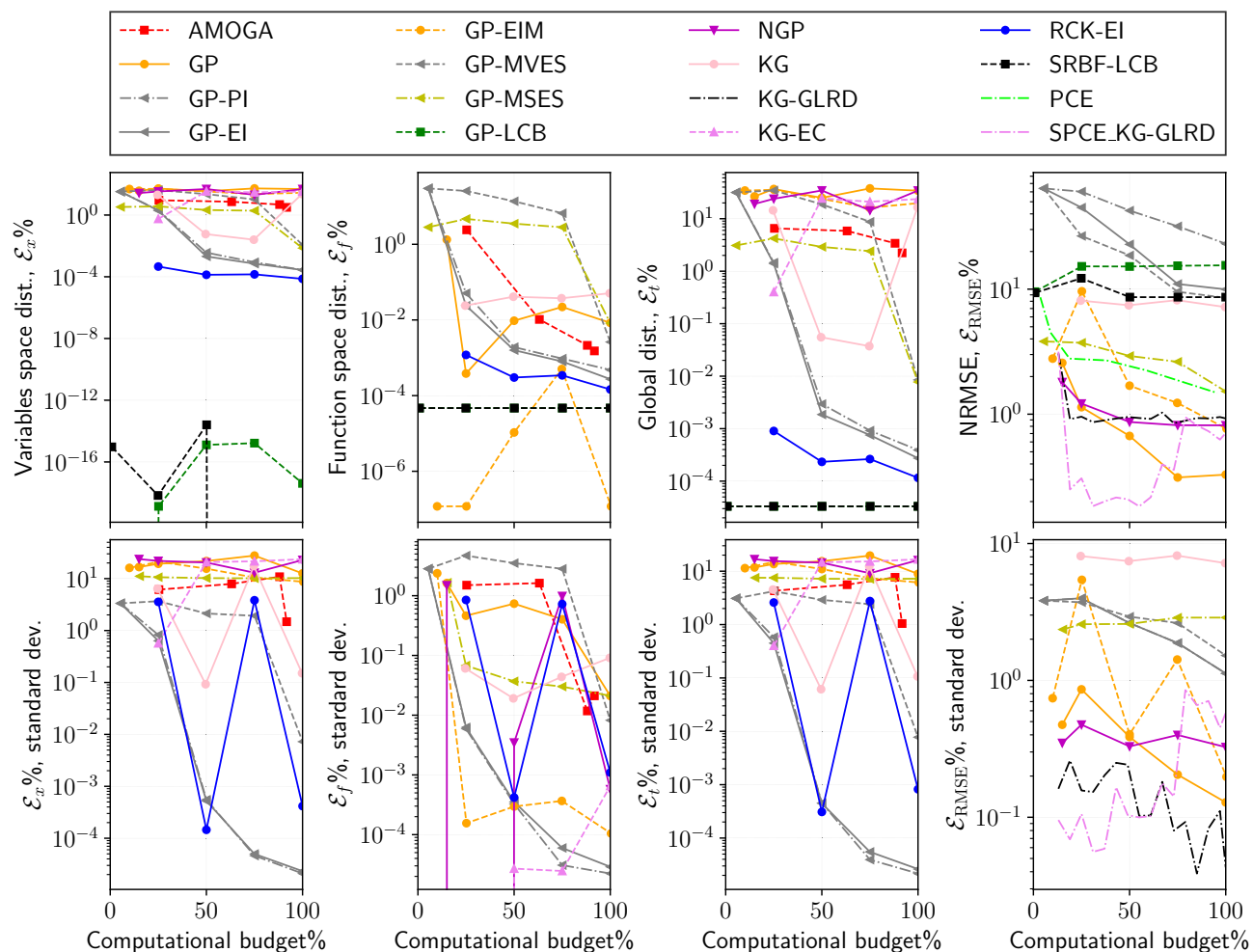


Figure 6-19: MF4.3 Metrics Convergence.

The Springs and Springs-masses results are shown in Figures 6-20 and 6-21, respectively. For the optimization, BFGS and GP-EI are doing well in both 2D and 4D as are GP-LCB and SRBF-LCB in 2D and RCK-EI and GP-PI in 4D. For global accuracy GP and PCE are looking good in 2D and NGP, GP and KG-GLRD are best in 4D, though the convergence of all methods is somewhat sluggish.

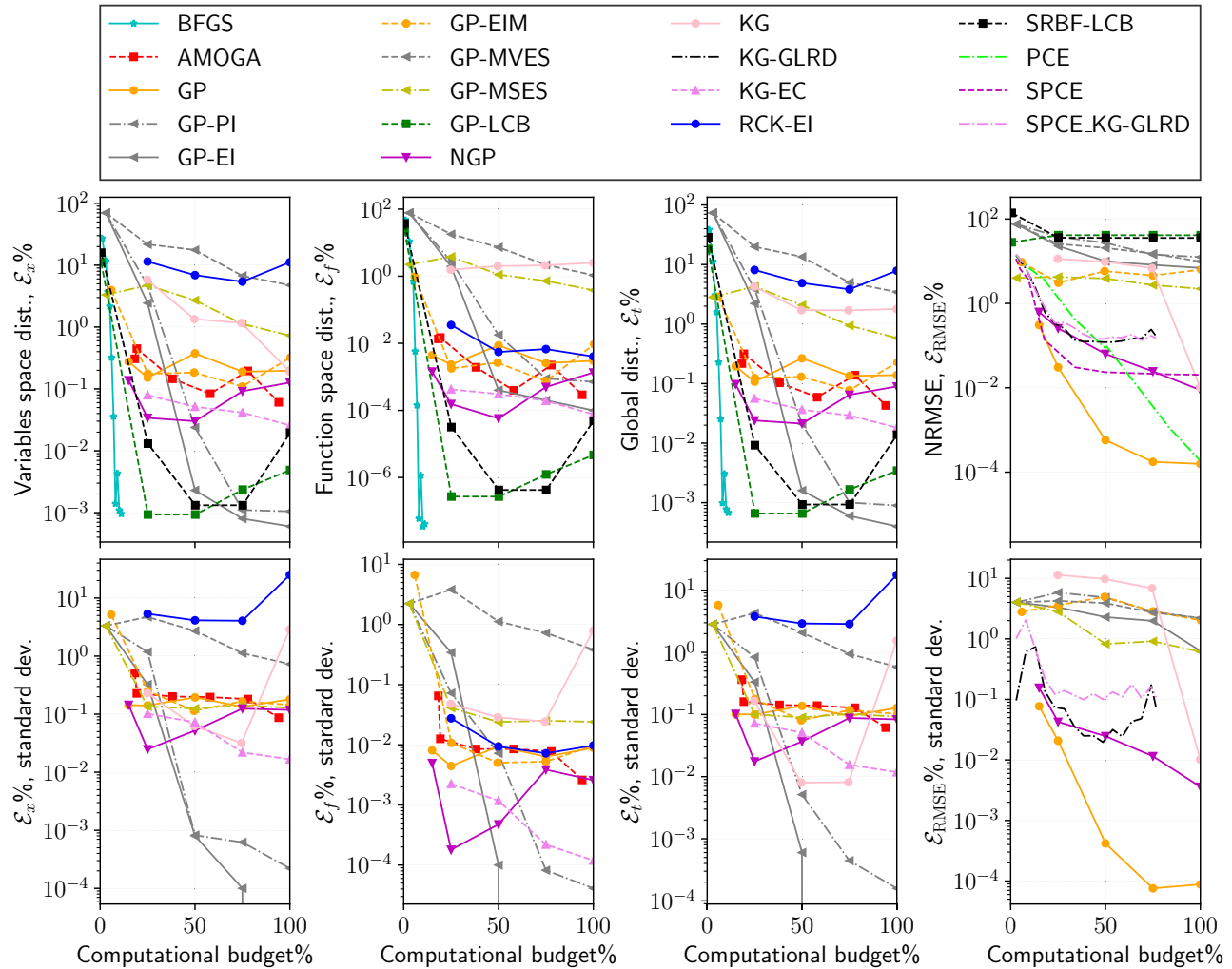


Figure 6-20: MF5.1 Metrics Convergence.

# L1 BENCHMARK DEFINITIONS AND RESULTS

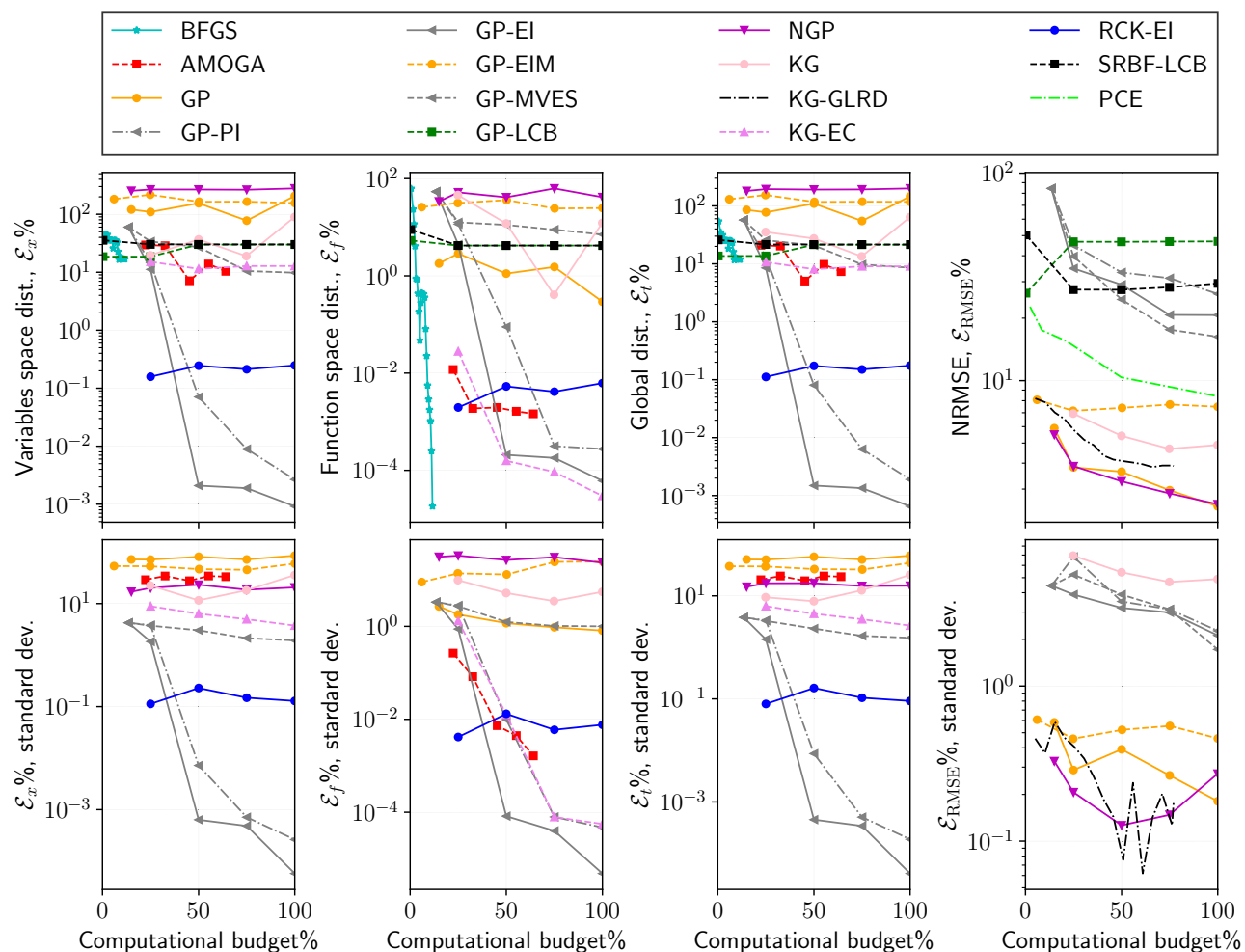


Figure 6-21: MF5.2 Metrics Convergence.

Lastly, the Paciorek results are shown in Figure 6-22. For the optimization GP-PI, GP-EI, and by the end, GP-MSES are performing the best whereas for the global accuracy GP-MSES, GP, and NGP are superior to the other methods.

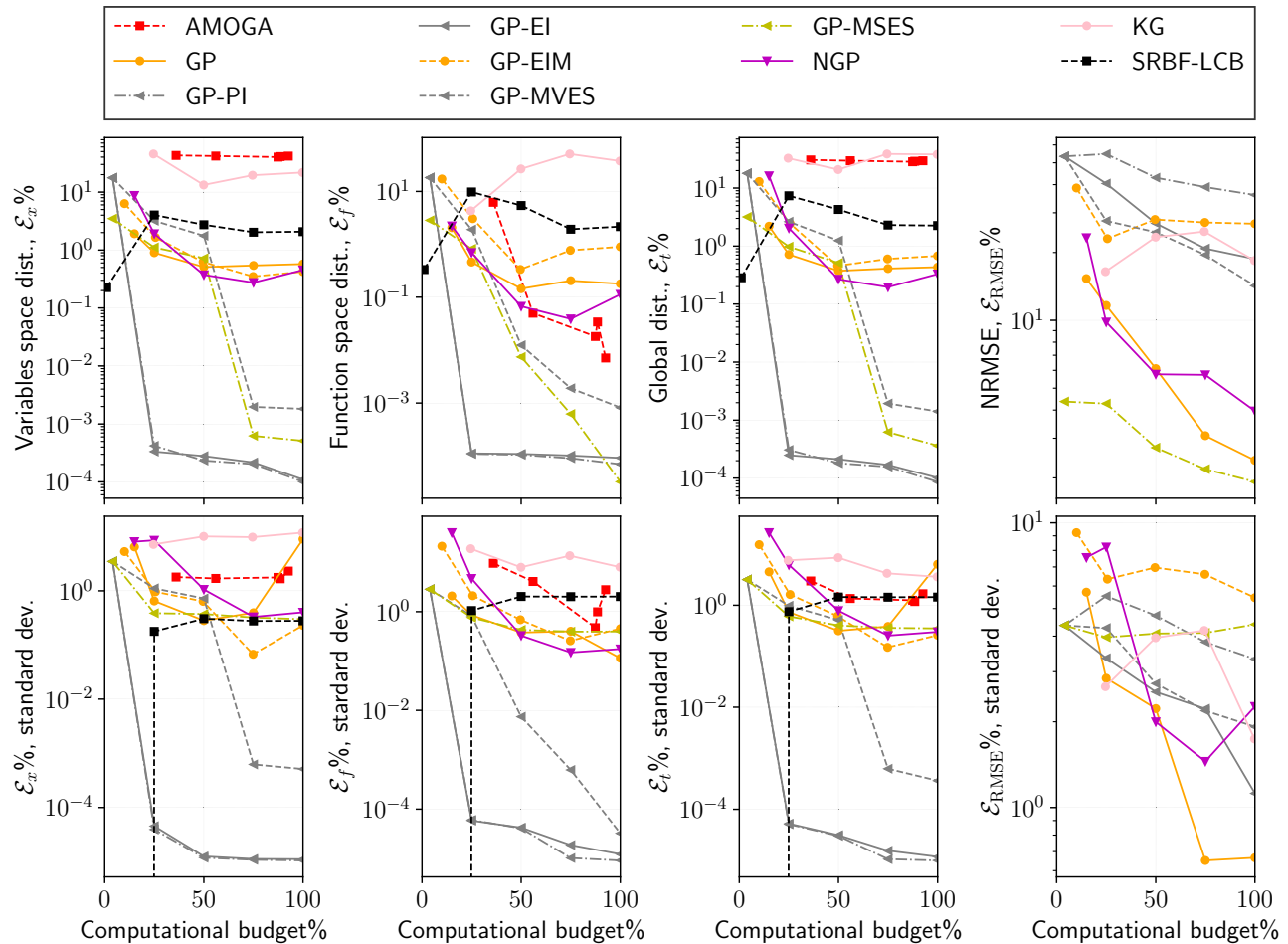


Figure 6-22: MF6 Metrics Convergence.

Table 6-5 gives a comprehensive overview of the results for all multi-fidelity L1 benchmark problems for all methods presented in Table 6-4. Four different budget levels are highlighted (25%, 50%, 75%, and 100% of the total budget) and the percent error is color-coded (orange if greater than 10%, yellow if between 1 and 10%, green if between 0.1 and 1%, light blue if between 0.01 and 0.1% and lastly dark blue if less than 0.01%). This should give the reader a quick visual overview of the various performances.

# L1 BENCHMARK DEFINITIONS AND RESULTS

Table 6-5: Overview of Multi-fidelity Methods Results for L1 Benchmark.

| Method   | $\mathcal{E}_x$ |   |   |   |   |   | $\mathcal{E}_f$ |   |   |   |   |   | $\mathcal{E}_t$ |   |   |   |   |   | $\mathcal{E}_{RMSE}$ |   |   |   |   |   | B% |   |   |   |   |   |   |     |
|----------|-----------------|---|---|---|---|---|-----------------|---|---|---|---|---|-----------------|---|---|---|---|---|----------------------|---|---|---|---|---|----|---|---|---|---|---|---|-----|
| BFGS     | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| AMOGA    | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| GP       | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| GP-PI    | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| GP-EI    | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| GP-EIM   | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| GP-MVES  | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| GP-MSES  | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| GP-LCB   | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| NGP      | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| KG       | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| KG-GLRD  | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| KG-EC    | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| KG-AEI   | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| RCK-EI   | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| SRBF-LCB | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 75  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 100 |
| PCE      | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 25  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . | .               | . | . | . | . | . | .                    | . | . | . | . | . | .  | . | . | . | . | . | . | 50  |
|          | .               | . | . | . | . | . | .               | . | . | . | . | . |                 |   |   |   |   |   |                      |   |   |   |   |   |    |   |   |   |   |   |   |     |

## 6.5 CONCLUSIONS

multi-fidelity methods have shown their potential in achieving optimal design solutions more efficiently than their high- or single-fidelity counterparts [3], [5]. Nevertheless, systematic comparison studies of these methods are still rare in the literature. This motivated the introduction of benchmark problems that could support the comparative and rigorous assessment of multi-fidelity methods.

Specifically, a benchmark suite of fourteen (L1) problems is proposed here to assess the efficiency and effectiveness of multi-fidelity optimization methods. The proposed benchmarks are meant to stress multi-fidelity optimization methods, dealing with various challenges typical of complex real-world optimization problems such as handling localized, multimodal, and discontinuous behaviors in the objective functions, as well as handling the possible presence of noise in the objective functions. A distinctive characteristic of the proposed problems is their ease of implementation. Specifically, these problems are based on analytic functions (such as Forrester, Rosenbrock, shifted-rotated Rastrigin, and Paciorek) and the solution of a spring-mass system.

The chapter also provides recommended experimental setups to support the rigorous test and comparison of different multi-fidelity methods. A quite large variety of methods was used to solve the benchmark problems and the results have been reported and discussed. These include surrogate lean and surrogate rich approaches. The former include BFGS and genetic algorithms approaches. The latter includes both offline and online collection of training data. Online approaches use adaptive sampling techniques, where several acquisition functions were used for comparison. Surrogate models include numerous Gaussian processes, radial-basis functions, and polynomial chaos expansions.

The performances of the various methods vary significantly for each problem and with dimensionality. As one can infer from Table 6-5 BFGS, GP-EIM, NGP, and KG-EC seem to overall perform best for the optimization problems, whereas NGP, GP, and PCE seem best suited for the construction of global accurate surrogate models.

## ACKNOWLEDGEMENTS

CNR-INM is supported by the Office of Naval Research (ONR) and ONR Global, Program Officers Drs. Woei-Min Lin and Elena McCarthy, through grants N62909-18-1-2033 and N62909-21-1-2042. The last author acknowledges the support of the US Air Force Office of Scientific Research (Dr. Fariba Fahroo, Computational Mathematics Program Officer). Distribution Statement A: Approved for Public Release; Distribution is Unlimited. AFRL-2023-2664.

## REFERENCES

- [1] P. S. Beran, D. Bryson, A. S. Thelen, M. Diez, and A. Serani, “Comparison of multi-fidelity approaches for military vehicle design,” in *AIAA AVIATION 2020 FORUM*, 2020, p. 3158.
- [2] B. Peherstorfer, K. Willcox, and M. Gunzburger, “Survey of multifidelity methods in uncertainty propagation, inference, and optimization,” *Siam Review*, vol. 60, no. 3, pp. 550–591, 2018.
- [3] M. Giselle Fernández-Godino, C. Park, N. H. Kim, and R. T. Haftka, “Issues in deciding whether to use multifidelity surrogates,” *AIAA Journal*, vol. 57, no. 5, pp. 2039–2054, 2019.

## L1 BENCHMARK DEFINITIONS AND RESULTS

- [4] A. Serani, R. Pellegrini, J. Wackers, C.-E. Jeanson, P. Queutey, M. Visonneau, and M. Diez, “Adaptive multi-fidelity sampling for cfd-based optimisation via radial basis function metamodels,” *International Journal of Computational Fluid Dynamics*, vol. 33, no. 6-7, pp. 237–255, 2019.
- [5] R. Pellegrini, J. Wackers, R. Broglia, A. Serani, M. Visonneau, and M. Diez, “A multi-fidelity active learning method for global design optimization problems with noisy evaluations,” *Engineering with Computers*, pp. 1–24, 2022.
- [6] M. Guo, A. Manzoni, M. Amendt, P. Conti, and J. S. Hesthaven, “Multi-fidelity regression using artificial neural networks: Efficient approximation of parameter-dependent output quantities,” *Computer methods in applied mechanics and engineering*, vol. 389, p. 114 378, 2022.
- [7] J. Liu, H. Dong, and P. Wang, “Multi-fidelity global optimization using a data-mining strategy for computationally intensive black-box problems,” *Knowledge-Based Systems*, vol. 227, p. 107 212, 2021.
- [8] H. B. Moss, D. S. Leslie, J. Gonzalez, and P. Rayson, “GIBBON: General-purpose information-based bayesian optimisation,” *Journal of Machine Learning Research*, vol. 22, no. 235, pp. 1–49, 2021.
- [9] X. Zhang, F. Xie, T. Ji, Z. Zhu, and Y. Zheng, “Multi-fidelity deep neural network surrogate model for aerodynamic shape optimization,” *Computer Methods in Applied Mechanics and Engineering*, vol. 373, p. 113 485, 2021.
- [10] H. Li, Z. Huang, X. Liu, C. Zeng, and P. Zou, “Multi-fidelity meta-optimization for nature inspired optimization algorithms,” *Applied Soft Computing*, vol. 96, p. 106 619, 2020.
- [11] J. Yi, Y. Shen, and C. A. Shoemaker, “A multi-fidelity RBF surrogate-based optimization framework for computationally expensive multi-modal problems with application to capacity planning of manufacturing systems,” *Structural and Multidisciplinary Optimization*, vol. 62, no. 4, pp. 1787–1807, 2020.
- [12] A. Serani, R. Pellegrini, J. Wackers, C.-E. Jeanson, P. Queutey, M. Visonneau, and M. Diez, “Adaptive multi-fidelity sampling for CFD-based optimisation via radial basis function metamodels,” *International Journal of Computational Fluid Dynamics*, vol. 33, no. 6-7, pp. 237–255, 2019.
- [13] J. Song, Y. Qiu, J. Xu, and F. Yang, “Multi-fidelity sampling for efficient simulation-based decision making in manufacturing management,” *IIEE Transactions*, vol. 51, no. 7, pp. 792–805, 2019.
- [14] H. Wang, Y. Jin, and J. Doherty, “A generic test suite for evolutionary multifidelity optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 6, pp. 836–850, 2017.
- [15] E. Hoag and J. R. Doppa, “Bayesian optimization meets search based optimization: A hybrid approach for multi-fidelity optimization,” in *Thirty-second AAAI conference on artificial intelligence*, 2018.
- [16] C. Li, P. Wang, and H. Dong, “Kriging-based multi-fidelity optimization via information fusion with uncertainty,” *Journal of Mechanical Science and Technology*, vol. 32, no. 1, pp. 245–259, 2018.
- [17] C. Durantin, J. Rouxel, J.-A. Désidéri, and A. Glière, “Multifidelity surrogate modeling based on radial basis functions,” *Structural and Multidisciplinary Optimization*, vol. 56, no. 5, pp. 1061–1075, 2017.
- [18] X. Cai, H. Qiu, L. Gao, L. Wei, and X. Shao, “Adaptive radial-basis-function-based multifidelity metamodeling for expensive black-box problems,” *AIAA journal*, vol. 55, no. 7, pp. 2424–2436, 2017.
- [19] B. Liu, S. Koziel, and Q. Zhang, “A multi-fidelity surrogate-model-assisted evolutionary algorithm for computationally expensive optimization problems,” *Journal of computational science*, vol. 12, pp. 28–37, 2016.

- [20] M. P. Rumpfkeil and P. S. Beran, "Multi-fidelity, gradient-enhanced, and locally optimized sparse polynomial chaos and kriging surrogate models applied to benchmark problems," in *AIAA SciTech 2020 Forum*, 2020. DOI: 10.2514/6.2020-0677.
- [21] F. Grassi, G. Manganini, M. Garraffa, and L. Mainini, "Resource aware multifidelity active learning for efficient optimization," in *AIAA Scitech 2021 Forum*, 2021, p. 0894.
- [22] F. Grassi, G. Manganini, M. Garraffa, and L. Mainini, "RAAL: Resource aware active learning for multifidelity efficient optimization," *AIAA Journal*, vol. 61, no. 6, pp. 2744–2753, 2023.
- [23] C. Park, R. T. Haftka, and N. H. Kim, "Remarks on multi-fidelity surrogates," *Structural and Multidisciplinary Optimization*, vol. 55, no. 3, pp. 1029–1050, 2017.
- [24] D. E. Bryson and M. P. Rumpfkeil, "Multifidelity quasi-newton method for design optimization," *AIAA Journal*, vol. 56, no. 10, pp. 4074–4086, 2018.
- [25] D. E. Bryson and M. P. Rumpfkeil, "Variable-fidelity surrogate modeling of lambda wing transonic aerodynamic performance," in *AIAA SCITECH 2016 FORUM*, 2016. DOI: 10.2514/6.2016-0294.
- [26] F. Di Fiore and L. Mainini, "Non-myopic multifidelity bayesian optimization," *arXiv preprint arXiv:2207.06325*, 2022.
- [27] A. I. Forrester, A. Söbester, and A. J. Keane, "Multi-fidelity optimization via surrogate modelling," *Proceedings of the royal society A: mathematical, physical and engineering sciences*, vol. 463, no. 2088, pp. 3251–3269, 2007.
- [28] S. Lee, F. Dietrich, G. E. Karniadakis, and I. G. Kevrekidis, *Linking Gaussian process regression with data-driven manifold embeddings for nonlinear data fusion*, 2019.
- [29] M. Raissi and G. E. Karniadakis, "Deep multi-fidelity gaussian processes," *CoRR*, vol. abs/1604.07484, 2016. arXiv: 1604.07484. [Online]. Available: <http://arxiv.org/abs/1604.07484>.
- [30] S. Ficini, U. Iemma, R. Pellegrini, A. Serani, and M. Diez, "An adaptive n-fidelity metamodel for design optimization via gaussian process regression," in *AIAA AVIATION 2021 Forum*, 2021.
- [31] A. Aguilera and R. Pérez-Aguila, "General n-dimensional rotations," 2004.
- [32] D. L. Clark Jr, H.-R. Bae, K. Gobal, and R. Penmetsa, "Engineering design exploration using locally optimized covariance kriging," *AIAA Journal*, vol. 54, no. 10, pp. 3160–3175, 2016.
- [33] D. J. J. Toal, "Some considerations regarding the use of multi-fidelity kriging in the construction of surrogate models," *Structural and Multidisciplinary Optimization*, vol. 51, no. 6, pp. 1223–1245, Dec. 2014. DOI: 10.1007/s00158-014-1209-5. [Online]. Available: <https://doi.org/10.1007/s00158-014-1209-5>.
- [34] D. E. Bryson and M. P. Rumpfkeil, "Multifidelity quasi-newton method for design optimization," *AIAA Journal*, vol. 56, no. 10, pp. 4074–4086, 2018.
- [35] D. Quagliarella, P. Iannelli, P. L. Vitagliano, and G. Chinnici, "Aerodynamic shape design using hybrid evolutionary computation and fitness approximation," in *AIAA 1st Intelligent Systems Technical Conference*, AIAA Paper 2004-6514, American Institute of Aeronautics and Astronautics (AIAA), Chicago, IL, Sep. 2004. DOI: 10.2514/6.2004-6514.
- [36] S. Yildiz, H. Pehlivan-Solak, M. Diez, O. Goren, and M. Nikbay, "Advanced experiments on gaussian process-based multi-fidelity methods over diverse mathematical characteristics," in *8th European Congress on Computational Methods in Applied Sciences and Engineering*, AIAA Paper 2004-6514, ECCOMAS, Oslo, Norway, Jun. 2022.

- [37] X. Ruan, P. Jiang, Q. Zhou, J. Hu, and L. Shu, "Variable-fidelity probability of improvement method for efficient global optimization of expensive black-box problems," *Structural and Multidisciplinary Optimization*, vol. 62, no. 6, pp. 3021–3052, 2020.
- [38] D. Huang, T. T. Allen, W. I. Notz, and R. A. Miller, "Sequential kriging optimization using multiple-fidelity evaluations," *Structural and Multidisciplinary Optimization*, vol. 32, no. 5, pp. 369–382, 2006.
- [39] S. Takeno, H. Fukuoka, Y. Tsukada, T. Koyama, M. Shiga, I. Takeuchi, and M. Karasuyama, "Multi-fidelity bayesian optimization with max-value entropy search and its parallelization," in *International Conference on Machine Learning*, PMLR, 2020, pp. 9334–9345.
- [40] A. Paleyes, M. Pullin, M. Mahsereci, N. Lawrence, and J. González, "Emulation of physical processes with emukit," in *Second Workshop on Machine Learning and the Physical Sciences, NeurIPS*, 2019.
- [41] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis, "Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 473, no. 2198, p. 20 160 751, 2017.
- [42] B. M. Adams, W. J. Bohnhoff, K. R. Dalbey, M. S. Ebeida, J. P. Eddy, M. S. Eldred, R. W. Hooper, P. D. Hough, K. T. Hu, J. D. Jakeman, M. Khalil, K. A. Maupin, J. A. Monschke, E. M. Ridgway, A. A. Rushdi, D. T. Seidl, J. A. Stephens, L. P. Swiler, and J. G. Winokur, "Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 6.15 user's manual," Sandia, Tech. Rep. SAND2020-12495, Nov. 2021.
- [43] M. P. Rumpfkeil and P. Beran, *Construction of Multi-Fidelity Surrogate Models for Aerodynamic Databases*, Proceedings of the Ninth International Conference on Computational Fluid Dynamics, ICCFD9, Istanbul, Turkey, July 11-15, 2016.
- [44] E. Minisci and M. Vasile, "Robust design of a re-entry unmanned space vehicle by multi-fidelity evolution control," *AIAA Journal*, vol. 51, no. 6, pp. 1284–1295, 2013.
- [45] D. Huang, T. T. Allen, W. I. Notz, and R. A. Miller, "Sequential kriging optimization using multiple-fidelity evaluations," *Structural and Multidisciplinary Optimization*, vol. 32, no. 5, pp. 369–382, 2006. [Online]. Available: <https://www.osti.gov/biblio/1000838>.
- [46] A. Thelen, L. Leifsson, and P. Beran, "Multifidelity flutter prediction using regression cokriging with adaptive sampling," *Journal of Fluids and Structures*, vol. 97, p. 103 081, 2020.
- [47] J. Wackers, R. Pellegrini, A. Serani, M. Visonneau, and M. Diez, "Efficient initialization for multi-fidelity surrogate-based optimization," *Journal of Ocean Engineering and Marine Energy*, pp. 1–17, 2022.
- [48] D. E. Bryson and M. P. Rumpfkeil, "All-at-once approach to multifidelity polynomial chaos expansion surrogate modeling," *Aerospace Science and Technology*, vol. 70, pp. 121–136, 2017.
- [49] M. P. Rumpfkeil and P. Beran, "Multi-Fidelity Sparse Polynomial Chaos Surrogate Models Applied to Flutter Databases," *AIAA Journal*, vol. 58, No.3, pp. 1292–1303, 2020.
- [50] M. Rumpfkeil, D. Bryson, and P. Beran, "Multi-Fidelity Sparse Polynomial Chaos and Kriging Surrogate Models Applied to Analytical Benchmark Problems," *Algorithms*, vol. 15(3), p. 101, 2022.