XLIII CILAMCE
21–25, November, 2022
FOZ DO IGUAÇU – PR – BRAZIL

Efficient Optimization of Engineering Problems using Multi-Fidelity Models

Leonardo G. Ribeiro[1], Evandro Parente Jr.[1], Antônio M. C. de Melo[1]

[1]*Laboratório de Mecânica Computacional e Visualização (LMCV), Departamento de Engenharia Estrutural e Construção Civil, Universidade Federal do Ceará*
Campus do Pici, Bloco 728, 60440-900, Ceará-Brazil
leonardoribeiro@alu.ufc.br, evandro@ufc.br, macario@ufc.br

Abstract. Optimization methods can be employed to find the optimum design of engineering structures. Due to their ease of implementation and robustness, bio-inspired optimization algorithms have been widely applied to solve complex optimization problems. However, these methods require a large number of expensive function evaluations. For a more efficient process, surrogate models can be used to provide a cheaper estimate of the structural responses. These models are built from a small set of true responses, and their approximated surface assists in the selection of promising trial designs. Efficient Optimization can be performed by iterately improving the model by the addition of new points in regions of interest, thus improving the accuracy of the model near the optimum location. In this work, we study the use of Multi-Fidelity models for the Efficient Optimization of engineering problems. Kriging and Hierarchical Kriging models are employed, and the selection of new points is performed using variations of the Expected Improvement and Probability of Improvement criteria. The obtained results are compared in terms of accuracy, the number of evaluations, and computational efficiency. Results show that Multi-Fidelity approaches are able to find optimal results using fewer high-fidelity evaluations.

Keywords: Surrogate-Based Optimization, Multi-Fidelity Models, Adaptive sampling

1 Introduction

For a long time, engineers were only concerned with conceiving safe and functional designs. However, with advances in computational processing and analysis, structural optimization became a requirement for efficient designs. To that end, one may employ a variety of optimization methods, such as bio-inspired algorithms. Even though these are very popular for engineering applications, they often require hundreds of expensive function evaluations, making the optimization process very time-consuming [1].

For a more efficient optimization process, we may use Surrogate-Based Optimization [1]. The idea behind this approach is to build an approximate model based on a small set of data points and, then, use this model to guide the optimization process. This can be performed using adaptive sampling schemes, where new data points are defined by the optimization of an acquisition function [2]. That being said, when the analysis model is too complex, such approaches may still present a high computational time, especially when a large dataset is required to fit a initial model with sufficient accuracy.

When that is the case, Multi-Fidelity Models can be used to improve the computational efficiency by evaluating some data points using a lower-fidelity source [1, 3, 4]. This approach allows for a better exploration of the design space, while not requiring a large number of expensive analyses.

In this work, we will use Hierarchical Kriging in the optimization of engineering problems. For comparison purposes, the single-fidelity Kriging model will also be used. During the optimization process, the model will be improved by the addition of new data points using proper adaptive sampling techniques, such as Expected Improvement, Probability of Improvement, and Variable-Fidelity versions of such methods [2, 5, 6]. The obtained results will be compared in terms of accuracy and number of evaluations.

The rest of the paper is organized as follows. In Section 2 the surrogate modeling techniques are discussed in more detail, with a focus on multi-fidelity modeling. In Section 3 we introduce the concept of adaptive sampling, and show some popular techniques for variable-fidelity cases. Some applications are presented in Section 4 and, finally, the main conclusions are brought together in Section 5.

2 Surrogate Modeling

Surrogate Modeling is a powerful technique to make the optimization process more efficient by providing an approximate response surface with a much lower evaluation cost. This approximate surface can be fitted based on the true response evaluated at a small set of sampling points. If no prior information about the true function behavior is available, a Design of Experiments (DoE) technique can be employed to pick the initial sampling points. For instance, the Latin Hypercube Sampling (LHS) is a stratified random sampling technique commonly employed for Surrogate-Based Optimization [1].

In this work, we aim at using Multi-Fidelity models, in particular the Hierarchical Kriging, to approximate an expensive structural response. For comparison purposes, results will also be shown for the single-fidelity Kriging model. In the following, we explain how both these models can be fitted based on a set of data points.

2.1 Kriging

The Kriging model is given by the sum of a global trend g and its autocorrelated localized deviations Z [1]:

$$\hat{y}(\mathbf{x}) = g(\mathbf{x}) + Z(\mathbf{x}) \tag{1}$$

For the Ordinary Kriging, $g(\mathbf{x}) = \mu$ is constant, where μ is the process mean. The localized deviations $Z(\mathbf{x})$ can be assumed as the realization of a stochastic process with mean zero and covariance:

$$\text{cov}(\mathbf{y}, \mathbf{y}) = \sigma^2 \, \mathbf{\Psi} \tag{2}$$

where σ^2 is the process variance and $\mathbf{\Psi}$ is a correlation matrix:

$$\Psi_{ij} = \text{cor}[y(\mathbf{x}_i), y(\mathbf{x}_j)] \tag{3}$$

Correlation between the responses is given by a measure of similarity. Usually, the Gaussian correlation function is employed:

$$\text{cor}\left[y(\mathbf{x}_i), y(\mathbf{x}_j)\right] = \exp\left(-\sum_{k=1}^{m} \theta_k \left| x_{i,k} - x_{j,k} \right|^{p_k} \right) \tag{4}$$

The mean and the variance can be defined using the Maximum Likelihood Estimator (MLE). For a given data set $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, the likelihood of a model with mean μ and variance σ is given by:

$$L(\mathbf{y}|\mu, \sigma) = \frac{1}{(2\,\pi\,\sigma^2)^{n/2}\,|\mathbf{\Psi}|^{1/2}} \, \exp\left[-\frac{(\mathbf{y} - \mathbf{1}\,\mu)^T \, \mathbf{\Psi}^{-1} \, (\mathbf{y} - \mathbf{1}\,\mu)}{2\,\sigma^2} \right] \tag{5}$$

By differentiation, we obtain the MLE for μ and σ as:

$$\hat{\mu} = \frac{\mathbf{1}^T \mathbf{\Psi}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{\Psi}^{-1} \mathbf{1}} \quad \text{and} \quad \hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})^T \mathbf{\Psi}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu})}{n} \tag{6}$$

To define the MLE for θ_k and p_k, however, we need to use an appropriate optimization method. For simplification, most researchers often consider $p_k = 2.0$. Then, optimization is performed considering only the parameter θ_k [1]. Substituting Eq. (6) in Eq. (5), taking the natural logarithm, and removing the constant terms, we end up with the concentrated ln-likelihood function:

$$\ln L \approx -\frac{n}{2} \, \ln(\hat{\sigma}^2) - \frac{1}{2} \, \ln |\mathbf{\Psi}| \tag{7}$$

Then, the definition of θ_k can be performed by solving the optimization problem:

$$\begin{cases} \text{minimize} & -\ln L(\boldsymbol{\theta}) \\ \text{where} & \boldsymbol{\theta}_L \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_U \end{cases} \tag{8}$$

Since this likelihood function is highly multi-modal, in this work we use the Particle Swarm Optimization (PSO) algorithm to solve this problem.

After the model is built, we evaluate the model prediction using the Kriging estimator, given by [1]:

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \boldsymbol{\psi}^T \, \mathbf{\Psi}^{-1} \, (\mathbf{y} - \mathbf{1}\,\hat{\mu}) \tag{9}$$

The Kriging model also allows for the assessment of the posterior variance:

$$\hat{s}^2(\mathbf{x}) = \sigma^2 \left(1 - \boldsymbol{\psi}^T \mathbf{\Psi}^{-1} \boldsymbol{\psi} \right) \tag{10}$$

which can be interpreted as an estimate of the model uncertainty on its own prediction.

2.2 Hierarchical Kriging

The Hierarchical Kriging was proposed by Han and Görtz [4] as a powerful Multi-Fidelity model with a simple formulation. Here, alongside the High-Fidelity (HF) data, given by n_h points (located at $\mathbf{x}_h$) evaluated using a higher fidelity source, the model is able to consider Low-Fidelity (LF) data, given by n_l points (located at $\mathbf{x}_l$) evaluated using a lower fidelity source. This allows for a much better exploration of the design space at a lower cost.

For the Hierarchical Kriging, the trend term is given by a Kriging model built based on a set of LF data:

$$\hat{y}(\mathbf{x}) = \beta_0 \, \hat{y}_l(\mathbf{x}) + Z(\mathbf{x}) \tag{11}$$

where β_0 is a constant scaling factor and, again, $Z(\mathbf{x})$ can be assumed to come from a stochastic process with mean zero and covariance given by:

$$\text{cov}(\mathbf{y}, \mathbf{y}) = \sigma^2 \, \mathbf{\Psi} \tag{12}$$

Hierarchical Kriging model building starts by fitting a Kriging model to the LF data $\mathcal{D} = \{(\mathbf{x}_{l,i}, y_{l,i})\}_{l,i=1}^{n}$. Thus, we are able to evaluate $\hat{y}_l$ and $\hat{s}_l$ by Eqs. (9) and (10). Then, building of the Hierarchical Kriging model itself can be performed by maximizing its corresponding ln-likelihood function:

$$\ln L = -\frac{n_h}{2} \, \ln(2\,\pi) - \frac{n_h}{2} \, \ln(\sigma^2) - \frac{1}{2} \, \ln|\mathbf{\Psi}(\mathbf{x}_h, \mathbf{x}_h)| - \frac{(\mathbf{y}_h - \beta_0 \, \mathbf{F})^T \, \mathbf{\Psi}(\mathbf{x}_h, \mathbf{x}_h)^{-1} \, (\mathbf{y}_h - \beta_0 \, \mathbf{F})}{2\,\sigma^2} \tag{13}$$

where:

$$\mathbf{F} = \hat{y}_l(\mathbf{x}_h) \tag{14}$$

By differentiation, MLEs for β_0 and σ^2 can be defined as:

$$\hat{\beta}_0 = \frac{\mathbf{F}^T \, \mathbf{\Psi}^{-1} \, \mathbf{y}_h}{\mathbf{F}^T \, \mathbf{\Psi}^{-1} \, \mathbf{F}} \quad \text{and} \quad \hat{\sigma}^2 = \frac{(\mathbf{y}_h - \beta_0 \, \mathbf{F})^T \mathbf{\Psi}(\mathbf{x}_h, \mathbf{x}_h)^{-1} (\mathbf{y}_h - \beta_0 \, \mathbf{F})}{n_h} \tag{15}$$

Substituting Eqs. (15) in Eq. (13), the concentrated ln-likelihood function for the Hierarchical Kriging model can be derived:

$$\ln L \approx -\frac{n_h}{2} \, \ln(\hat{\sigma}^2) - \frac{1}{2} \, \ln|\mathbf{\Psi}(\mathbf{x}_h, \mathbf{x}_h)| \tag{16}$$

Once again, we adopt $p_k = 2.0$. Then, the rest of hyper-parameters (θ_k) can be set by Eq. (8), but now using Eq. (16) to evaluate the concentrated ln-likelihood. After being built, the Hierarchical Kriging model can be evaluated by:

$$\hat{y}(\mathbf{x}) = \hat{\beta}_0 \, \hat{y}_l(\mathbf{x}) + \boldsymbol{\psi}^T \, \mathbf{\Psi}^{-1} \, (\mathbf{y}_h - \hat{\beta}_0 \, \mathbf{F}) \quad \text{where} \quad \psi_i(\mathbf{x}) = \exp\left(- \sum_{k=1}^{m} \theta_k \, |x_{i,k} - x_k|^{p_k} \right) \tag{17}$$

The estimate of the posterior variance for Hierarchical Kriging is given by:

$$\hat{s}^2(\mathbf{x}) = \hat{\sigma}^2 \left[1 - \boldsymbol{\psi}^T \mathbf{\Psi}^{-1} \boldsymbol{\psi} + \frac{(\hat{y}_l - \mathbf{F}^T \mathbf{\Psi}^{-1} \boldsymbol{\psi})^2}{\mathbf{F}^T \mathbf{\Psi}^{-1} \mathbf{F}} \right] \tag{18}$$

Hierarchical Kriging also allows for the assessment of the LF model y_h variance, which can be performed similar to a regular Kriging model:

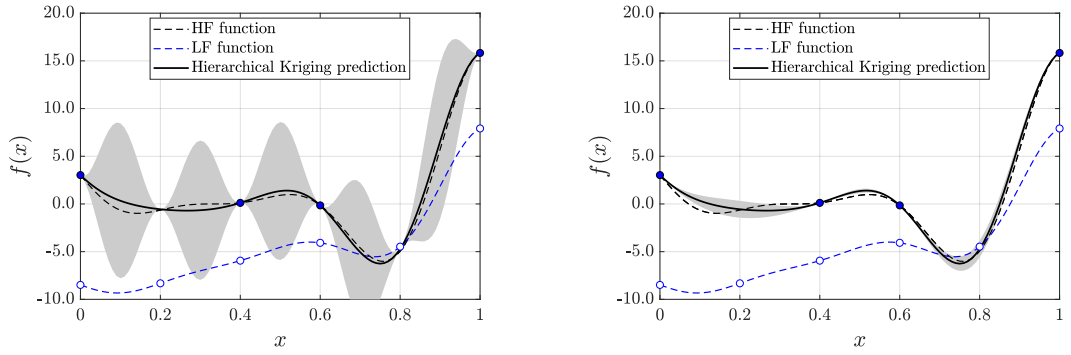
$$\hat{s}_l^2(\mathbf{x}) = \hat{\sigma}_l^2 \left(1 - \boldsymbol{\psi}_l^T \mathbf{\Psi}_l^{-1} \boldsymbol{\psi}_l \right) \tag{19}$$

This way, Zhang et al. [5] proposed the assessment of a Variable-Fidelity (VF) posterior variance, given by:

$$\hat{s}^2(\mathbf{x}, f) = \begin{cases} \beta_0^2 \, \hat{s}_l^2(\mathbf{x}) & \text{, if } f = 1 \\ \hat{s}^2(\mathbf{x}) & \text{, if } f = 2 \end{cases} \tag{20}$$

where $f = 1$ is related to the LF source, and $f = 2$ is related to the HF source. The VF variance is of great importance for the selection and addition of new data points, which will be discussed in the next section.

Figure 1 shows the prediction and variance of a Hierarchical Kriging model. Here, results are shown for the VF variance considering both fidelity sources. Note that, by the consideration of a small number of LF data, the model is able to represent the true function much better.



(a) $\hat{y} \pm \hat{s}^2(\mathbf{x}, 1)$ (b) $\hat{y} \pm \hat{s}^2(\mathbf{x}, 2)$

Figure 1. Prediction and variance for the Hierarchical Kriging model.

3 Adaptive sampling

For a more accurate process, we may use an adaptive sampling technique where new promising designs are added to the model during the optimization. The location of such designs can be found by the optimization of an acquisition function, which often depends on model prediction ($\hat{y}$) and variance ($\hat{s}$).

To that end, Kushner [7] proposed the Probability of Improvement (PI) criterion, which maximizes the probability that a given trial design improves upon the current optimum. The PI can be evaluated by:

$$P[I(\mathbf{x})] = \Phi\left(\frac{y_{min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) \tag{21}$$

where Φ is the Cumulative Distribution Function (CDF) for the Normal distribution.

Alternatively, based on the theory of expected utility, Močkus [8] proposed the Expected Improvement (EI) criterion. This method considers not only the probability that the response is an improvement, but also the magnitude of this improvement. The approach was popularized by Jones et al. [2] after the EI was employed in the well-known Efficient Global Optimization (EGO) algorithm. The EI is given by:

$$E[I(\mathbf{x})] = (y_{min} - \hat{y}(\mathbf{x}))\,\Phi\left(\frac{y_{min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) + \hat{s}(\mathbf{x})\,\phi\left(\frac{y_{min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) \tag{22}$$

where Φ and ϕ are the Cumulative Distribution Function (CDF) and the Probability Density Function (PDF) for the Normal distribution, respectively.

3.1 Variable-Fidelity methods

The regular PI and EI can also be employed for Multi-Fidelity (MF) models, as long as model prediction and variance are evaluated accordingly. However, in these cases, along with the definition of the new data location, one should also define the new data fidelity. Using the Hierarchical Kriging model, one may employ Variable-Fidelity (VF) methods to assist with this issue. Here, the VF variance, shown in Eq. (20), is used to evaluate the acquisition function. Then, the source which presents the best acquisition function value is used to evaluate the new trial design. These methods were first used by Zhang et al. [5], who proposed the VF-EI:

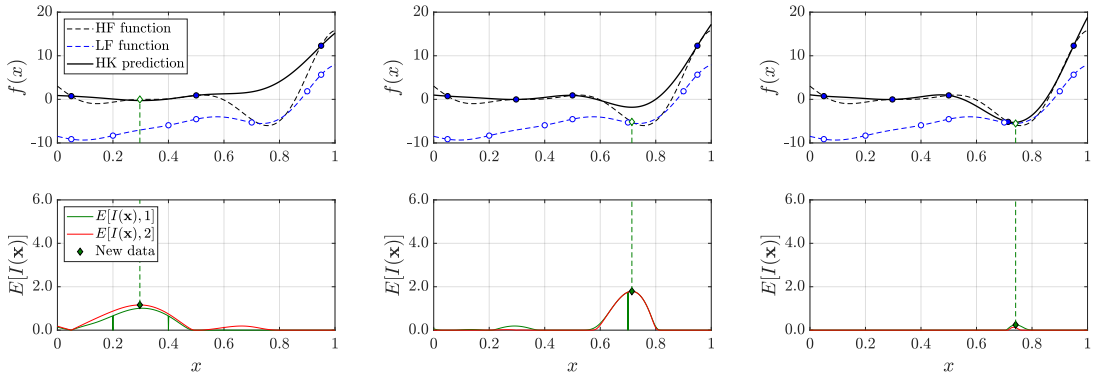
$$E[I(\mathbf{x}), f] = (y_{min} - \hat{y}(\mathbf{x}))\,\Phi\left(\frac{y_{min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x}, f)}\right) + \hat{s}(\mathbf{x}, f)\,\phi\left(\frac{y_{min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x}, f)}\right) \tag{23}$$

where $f = 1$ is related to the LF source, and $f = 2$ is related to the HF source. Figure 2 shows how the VF-EI works for the first three iterations of a test problem. Here, the method is able to find the optimum region in the second iteration. A design very close to the optima is then found in the third iteration.

Later, Ruan et al. [6] also proposed the VF-PI:

$$P[I(\mathbf{x}), f] = \Phi\left(\frac{y_{min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x}, f)}\right) \tag{24}$$

The authors suggest that the method presents good results when correlation between sources is high.



(a) Iteration 1 (b) Iteration 2 (c) Iteration 3

Figure 2. Variable-Fidelity Expected Improvement (VF-EI) using the Hierarchical Kriging model.

4 Applications

In this section, we present some applications using Variable-Fidelity methods for SBO. Two main stopping criteria are considered, one related to the maximum number of HF points evaluated (N_{max}) and other to the maximum number of iterations without improvement (N_{stall}). Unless stated otherwise, this work considers $N_{max} = 100$ and $N_{stall} = 10$. If the algorithm finds the optimum design, it stops the iterative process. The accuracy of each approach will be defined by the NRMSE:

$$\text{NRMSE} = \sqrt{\frac{1}{N_r}\sum_{i=1}^{N_r}\left(\frac{y_{min} - y_{opt}}{y_{opt}}\right)^2} \tag{25}$$

where N_r is the number of runs. Since all methods employed in this work are stochastic, $N_r = 10$ runs will be carried out for each problem and the average of these metrics will be shown. The efficiency will be defined by the number of HF (n_{eh}) and LF (n_{el}) evaluations performed in the entire process and during the adaptive sampling (which will be shown in parenthesis). In this paper, the initial sample is defined using $n_h = 1.5\,m$ and $n_l = 3\,m$, where m is the number of design variables.

For reference, the Particle Swarm Optimization (PSO) algorithm used in Ribeiro et al. [9] will also be used here for the optimization of engineering functions. As parameters, we considered $N_{it} = 200$ iterations, a population of $N_p = 100$ particles, inertia $w = 0.72$, the ring topology, and cognitive and social factors $c_1 = c_2 = 1.50$. For the PSO, we will consider $N_{stall} = 50$.

In the first problem, our goal is to minimize the maximum displacement of a simply-supported beam, considering the width b and height h of the cross section as design variables. Here, we consider that the beam length is $L = 5$ m, the distributed load is $q = 10$ kN/m, the Young modulus is $E = 21662$ kPa, and the shear modulus is $G = 8528.35$ kPa. Two constraints will be considered, related to a maximum cross-sectional area $A_{max} = 0.8$ m^2 and a constraint on the cross-section aspect-ratio ($h \leq 3\,b$). Since these constraints are naturally cheap to evaluate, they will be evaluated exactly, as there is no need to build an additional surrogate model to estimate their values.

Considering shear strains, the beam maximum displacement may be obtained using the Timoshenko theory:

$$w_T(\mathbf{x}) = \frac{5\,q\,L^4}{384\,E\,I} + \frac{q\,L^2}{8\,G\,A_s} \text{ where } A_s = \frac{5}{6}\,A \tag{26}$$

Using this function to evaluate the maximum displacement, the optimum is found at $w_T(0.516; 1.549) = 28.1$ mm. Instead of the Timoshenko solution, we may compute the displacement in a simplified way by the Euler-Bernoulli theory, known as the classical beam theory, which disregards shear deformations. Here, the maximum displacement is given by:

$$w_E(\mathbf{x}) = \frac{5\,q\,L^4}{384\,E\,I} \tag{27}$$

Table 1 shows the results obtained using a variety of approaches. It can be noted that the PI-based methods performed worse than their EI counterparts. On the other hand, HKRG behaved exceptionally well, especially using the VF-EI, where, on average, only two extra HF evaluation were required to find the global optima. Note that the conventional PSO algorithm is also able to find the optimum, but at a much higher number of individual evaluations.

Table 1. Averaged results for the simply supported beam problem.

SAO algorithm	w_{opt} (mm)	NRMSE	$n_{ev,h}$	$n_{ev,l}$
PSO	28.10	0.00%	15050	-
KRG-PI	28.31	0.76%	12 (9)	-
KRG-EI	28.10	0.00%	8 (5)	-
HKRG-VF-PI	28.44	1.19%	16 (13)	20 (14)
HKRG-VF-EI	28.10	0.00%	5 (2)	8 (2)

In the second problem, we aim to minimize the maximum displacement of a FG clamped beam with length $L = 10$ m and tip load $P = 40$ kN. The beam has a circular cross-section of radius r, and the material gradation is given by:

$$V_m = \left(\frac{x}{L}\right)^N \text{ and } V_c = 1 - V_m \tag{28}$$

where V_m and V_c refer to the metal and ceramic volume fractions, respectively, and N is a parameter which defines the material gradation. Equivalent material properties are evaluated via the rule of mixtures:

$$E(\mathbf{x}) = V_m(\mathbf{x})\, E_m + V_c(\mathbf{x})\, E_c \tag{29}$$

where $E_m = 90$ GPa and $E_c = 380$ GPa are the metal and ceramic Young's moduli, respectively. Once again, the aim is to minimize the maximum displacement of the beam, considering r and N as design variables. Two constraints will be considered, now related to a maximum cost $C_{max} = 50$ M.U. (Monetary Units), and a minimum percentage of ceramic in the beam $\overline{V}_{c,min} = 50\%$. The cost per unit volume for the metal and ceramic are, respectively, $C_m = 1$ M.U./m^3 and $C_c = 5$ M.U./m^3.

Using the unit load method and considering the Euler-Bernoulli theory, the tip displacement of the beam can be evaluated as:

$$w(\mathbf{x}) = \int_0^L \frac{M(\mathbf{x})\,\overline{M}(\mathbf{x})}{E(\mathbf{x})\,I}\,dx = \int_0^L \frac{P\,(L-x)^2}{E(\mathbf{x})\,I}\,dx \tag{30}$$

This integral can be solved numerically using a 10-point Gauss quadrature rule. Using this function, the optimum of the problem is found at $w(0.728, 1.0) = 0.1025$ mm.

Alternatively, we may use the Rayleigh-Ritz method to find an approximation for the beam displacement. Assuming a displacement field described by:

$$w(x) = a\,x^2 \tag{31}$$

we may find a considering the principle of stationary total potential energy. Thus, the maximum displacement may be estimated by:

$$w(x) = \frac{P\,L^3\,(N+1)}{4\,I\,(E_c\,N + E_m)} \tag{32}$$

This function can be used as a LF function for the beam displacement.

Table 2 shows the results obtained using each approach. Results found here are very similar to those from the previous problem. Once again, the EI-based approaches outperformed the PI-based approaches. Also, the HKRG model excels at finding the optimum using few evaluations. Finally, even though the conventional PSO algorithm required a much higher number of evaluations, it actually struggled a little to find the optimum.

Table 2. Averaged results for the FG beam problem.

Algorithm	w_{opt} (mm)	NRMSE	$n_{ev,h}$	$n_{ev,l}$
PSO	0.1028	0.36%	14340	-
KRG-PI	0.1033	0.84%	32 (29)	-
KRG-EI	0.1025	0.00%	6 (3)	-
HKRG-VF-PI	0.1034	0.95%	30 (27)	36 (30)
HKRG-VF-EI	0.1025	0.00%	5 (2)	11 (5)

5 Conclusions

In this paper, we performed the efficient optimization of engineering problems. We used two different models, Kriging and Hierarchical Kriging. The latter is a Multi-Fidelity (MF) model, able to consider information from multiple fidelities. Different acquisition functions were considered: Probability of Improvement (PI) and Expected Improvement (EI), as well as Variable-Fidelity (VF) versions of these methods for the Hierarchical Kriging. All results were compared in terms of accuracy and computational efficiency.

It was shown that the Hierarchical Kriging with the VF-EI is able to find very accurate optimal results, achieving outstanding efficiency. In comparison to the regular Kriging model with the EI, the method required fewer HF evaluations. In both cases tested, the algorithm required less than 3 new HF evaluations to find the optimum. We also showed that the PI-based approaches are not able to present similar results, as the method struggles to find the optimum in multi-modal problems. Finally, we compared the SAO-based algorithms with a conventional PSO method. We showed that, in terms of efficiency, SAO-based algorithms require a much fewer number of evaluations. Also, regarding EI-based approaches, accuracy was better than the accuracy from the PSO algorithm itself.

Acknowledgements. The authors gratefully acknowledge CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the financial support.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] A. I. J. Forrester, A. Sobester, and A. J. Keane. *Engineering design via surrogate modelling: a practical guide*. Wiley, 2008.

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

[3] M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, vol. 87, n. 1, pp. 1–13, 2000.

[4] Z.-H. Han and S. Görtz. Hierarchical kriging model for variable-fidelity surrogate modeling. *AIAA Journal*, vol. 50, n. 9, pp. 1885–1896, 2012.

[5] Y. Zhang, Z. H. Han, and K. S. Zhang. Variable-fidelity expected improvement method for efficient global optimization of expensive functions. *Structural and Multidisciplinary Optimization*, vol. 58, n. 4, pp. 1431–1451, 2018.

[6] 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, n. 6, pp. 3021–3052, 2020.

[7] H. J. Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, vol. 86, n. 1, pp. 97–106, 1964.

[8] J. Močkus. On bayesian methods for seeking the extremum. In G. I. Marchuk, ed, *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7*, pp. 400–404, Berlin, Heidelberg. Springer Berlin Heidelberg, 1974.

[9] L. G. Ribeiro, M. A. Maia, E. Parente Jr., and A. M. C. D. Melo. Surrogate based optimization of functionally graded plates using radial basis functions. *Composite Structures*, vol. 252, 2020.