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

Comparison of EGO and sEGO optimization algorithm based on Kriging for noisy function

Cibelle D. de C. D. Maia[1], Rafael Holdorf Lopez[2]

[1]*Dept. of Civil Engineering, Federal University of Santa Catarina*
João Pio Duarte Silva, 88040-900, SC/Florianópolis, Brazil
maia.cibelle@gmail.com
[2]*Dept. of Civil Engineering, Federal University of Santa Catarina*
João Pio Duarte Silva, 88040-900, SC/Florianópolis, Brazil
rafaelholdorf@gmail.com

Abstract.
In many engineering optimization problems the number of function evaluations is severely limited by time or computational cost. In addition, the representation of randomness due to noise and uncertainties in the model is essential. One strategy adopted for these cases is solve the problem through response surfaces, or meta-models, especially Kriging model. A traditional Kriging-based algortihm optimization is the Global Efficient Optimization (EGO) method. An most recent algorithm for stochastic problems was sEGO in which it introduces a parcel that reflects the intrinsic noise of the stochastic function in your framework. In this paper these optimization algorithms will be approached through some examples for demonstrate the importance of the variance quantifying approach in the optimization process through Kriging meta-model, highlighting the influence of the noise amplitude in the choice of the optimization strategy. The conclusions obtained may serve as a guideline for choose the best approach for each type of optimization problem.

Keywords: Optimization problem, stochastic Kriging, heterogeneous noise, sEGO

1 Introduction

There are several sources of uncertainty that can be present in the modeling of engineering problems. Practical examples can be observed when analyzing a beam by considering a higher-order beam model, plasticity, damage theories, and other sources of non-linearity, and approximating the solution using a state-of-the-art finite element model. All of this procedure would still be a rough representation of reality if the intrinsic randomness of materials (rock, soil, concrete) and loads (wind, earthquake motion) were disregarded and a deterministic average was used [1].

The representation of the stochastic problem occurs through objective functions that can be formulated as expected value functions (E), being expressed as following:

$$f(\mathbf{x}) = E_{\boldsymbol{\theta}}\left[y(\mathbf{x}, \boldsymbol{\theta})\right], \tag{1}$$

where $\mathbf{x} \in R^{n_x}$ is the design vector, $\boldsymbol{\theta} \in R^{n_{\boldsymbol{\theta}}}$ is the random parameter vector, E is the expected value operator, and $y : R^{n_x} \times R^{n_{\boldsymbol{\Theta}}} \rightarrow R$ is the system performance measure. Since the expected value in practical problems can hardly be evaluated analytically, approximations using simulation, such as Monte Carlo integration (MCI), are often applied.

In order to make an efficient optimization of these problems possible, intelligent optimization strategies successfully coping with noisy evaluations sometimes are required. An alternative to high cost functions optimization is approach the meta-models based optimization algorithms. The basic idea is that the metamodel acts as an interpolating curve or regressor of support points that have information from the objective function and its constraints so that the results can be predicted without resorting to the use of the primary source (objective function) [2].

One of the most popular meta-model is the Kriging, this has a long and successful tradition for modeling and optimizing deterministic computer simulations [3]. The great advantage of this meta-model is that it allows the quantification of the uncertainty of the response surface through the mean square error (MSE) [4]. An extension for the application in noisy problems is Stochastic Kriging (SK) proposed by [5]. Figure 1 shows an example of Kriging surface based on noisy observations, $f(\mathbf{x}, \boldsymbol{\theta})$, where the bar is the noise amplitude.

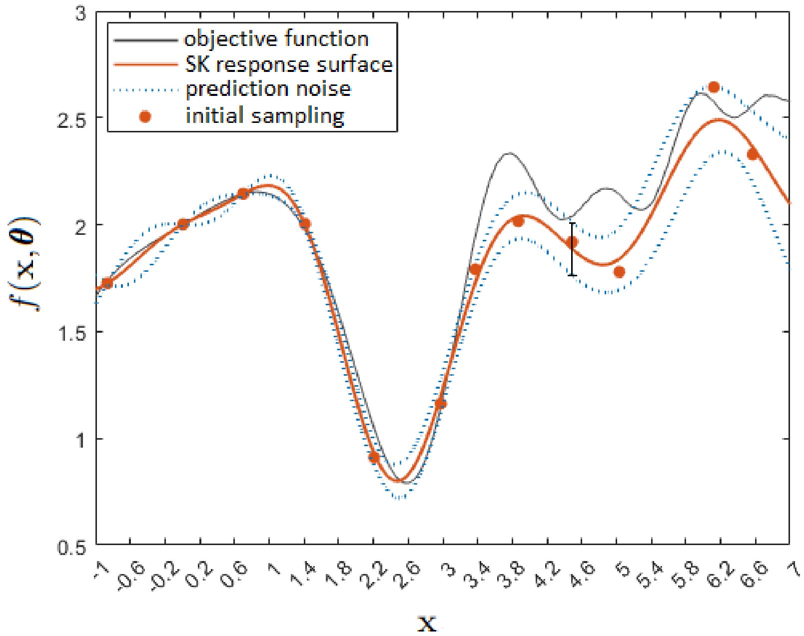

Figure 1. Kriging meta-model on noisy observations

A traditional approach to optimization through Kriging is the Efficient Global Optimization method (EGO) [6], this is one of the most popular algorithms for optimizing noiseless simulation; in this case, the fitted metamodel is the Kriging deterministic model.

In stochastic simulation the EGO may not be very suitable, as it ignores noise in the observations, assuming that samples were taken with infinite precision [7]. Research has been carried out to extend the EGO for stochastic simulation, in the work of [3] he compares several algorithms based on Kriging to optimize functions with homogeneous noise, which means that the variance of the noise does not depend on the x position. And in [8] it compares some algorithms with heterogeneous noise.

In this context of stochastic optimization, [9] developed the stochastic efficient global optimization algorithm (sEGO). sEGO first uses MCI to approximate the objective function, as it provides not only an approximation to the integral, but also the error variance. The error variance is then included in the SK structure, and the filling criterion AEI is used to guide the addition of new points in the stochastic EGO structure.

The goal of this paper is compare the performance of those two Kriging-based algorithm optimization - EGO and sEGO - on a analytic test functions with heterogeneous noise subject only to box constraint. Expected with this demonstrate the importance of the variance quantifying approach in the optimization process through Kriging meta-model, highlighting the influence of the noise amplitude in the choice of the optimization strategy.

The conclusions obtained may serve as a useful for researchers aiming to deal with optimization noisy problems. The insights will help to evaluate when it is advantageous to use kriging based algorithms with deterministic representation of the function, i.e., with objective function approximation for a limited sampling. Or when use variance information in the iterative process will be representative. This article is organized as follows. Section 2 provides a brief explanation of the Kriging meta-model studied, Section 3 details the Kriging-based optimization algorithms (EGO and sEGO), Section 4 presents the analysis and results of the problem and, lastly, conclusions are present in Section 5.

2 Kriging meta-model

2.1 Deterministic Kriging

In the interest of fitting a meta-model for the response $f(x)$ at n_x design points, the deterministic Kriging assumes that the unknown response surface ($\hat{y}$) can be represented as [5]:

$$\hat{y}(\mathbf{x}) = M(\mathbf{x}) + Z(\mathbf{x}). \tag{2}$$

where $M(\mathbf{x})$ is a vector of known trend functions while $Z(\mathbf{x})$ represents the extrinsic uncertainty imposed on the problem due to predictor construction. $Z(\mathbf{x})$ is define as a realization of a Gaussian random field with mean zero and covariance-stationary.

What relates one observation to another is the covariance function, denoted $\sum_z$, also referred to as *kernel*. Multiple covariance functions exist in the field of Gaussian process, the choice depends on prior hypothesis about the unknown functions. One of the most commonly used *kernel* in Kriging literature are the stationary squared exponential like a Gaussian base, Equation 3,focus on this research.

$$\left[\sum\nolimits_z\right]_{ij} = \sigma_z \mathbf{h}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sigma_z exp\left[-\sum_{k=1}^{n_x} c_k |x_k^{(i)} - x_k^{(j)}|\right]^{p_k}. \tag{3}$$

where σ_z is the variance of uncertainty caused by the construction of the surrogate model, $\mathbf{h}$ is the correlation vector, c_k and p_k are hyperparameters from the model.

The hyperparameters of these covariance function are usually estimated using Maximum Likelihood Estimation (MLE) [2]. To deal with functions on the presence of noise or uncertainties [5] proposed the extension of deterministic Kriging for stochastic Kriging (SK). We thus focus on SK for be capable of accommodating noisy evaluations in the optimization framework.

The Kriging prediction and variance for a given point x^+ are, respectively:

$$\hat{y}(x^+) = \hat{\mu} + \mathbf{h}^T \Psi^{-1}(\mathbf{y} - 1\widehat{\mu}) \tag{4}$$

$$s^2(x^+) = \hat{\sigma}_z^2 \left[1 - \mathbf{h}^T \Psi^{-1}\mathbf{h} + \frac{(\mathbf{1} - \mathbf{1}^T \Psi^{-1}\mathbf{h})^2}{\mathbf{1}^T \Psi^{-1}\mathbf{1}}\right]. \tag{5}$$

where $\hat{\mu}$ is the trend of the SK model, $\hat{\sigma}_z^2$ is the variance trend of the SK model, Ψ is the covariance matrix of the all support points of Z and $\overline{y}$ is the vector of the approximate mean value of the objective function at each design point, those parameters are define in [2].

2.2 Stochastic Kriging

To deal with functions on the presence of noise or uncertainties, the extension of deterministic Kriging for stochastic Kriging (SK) was proposed [5]. We thus focus on SK for being capable of accommodating noisy evaluations in the optimization framework.

In the stochastic Kriging (SK) it is added a parcel ϵ for meta-model construction, Equation 6, to account the sampling variability inherent in a stochastic simulation, which represents an intrinsic uncertainty of the problem.

$$\hat{y}(\mathbf{x}) = M(\mathbf{x}) + Z(\mathbf{x}) + \epsilon(\mathbf{x}). \tag{6}$$

where $\epsilon(\mathbf{x})$ has zero mean and is independently and identically distributed across replications.

The SK prediction and variance for a given point x^+ are, respectively:

$$\widehat{y}(x^+) = \hat{\mu} + \hat{\sigma}_Z^2 \mathbf{h}^T \left[\sum Z + \sum \epsilon\right]^{-1} (\overline{\mathbf{y}} - \widehat{\mu}\mathbf{1}). \tag{7}$$

$$\hat{s}^2(x^+) = \hat{\sigma}_Z^2 - (\hat{\sigma}_Z^2)^2 \mathbf{h}^T \left[\sum Z + \sum \epsilon\right]^{-1} \mathbf{h} + \frac{\delta^T \delta}{\mathbf{1}^T \left[\sum Z + \sum \epsilon\right]^{-1} \mathbf{1}}. \tag{8}$$

where $\delta(\mathbf{x}^+) = \mathbf{1} - \mathbf{1}^T \left[\sum Z + \sum \epsilon\right]^{-1} \hat{\sigma}_z^2 \mathbf{h}(\mathbf{x}^+)$, $\mathbf{h}$ is the correlation vector, $\hat{\mu}$ and $\hat{\sigma}_z^2$ are the mean and variance trend of the SK meta-model, respectively, $\sum Z$ is the covariance matrix of all the support points of Z, $\sum \epsilon$ is the covariance matrix of ϵ and $\overline{y}$ is the vector of the approximate mean value of the objective function at each design point, those parameters are defined in [5].

3 Kriging meta-model based optimization

Because the surrogate model, $\hat{y}$, is only an approximation of the true function $f(\mathbf{x})$ we wish to optimize, for enhance the accuracy of the model are made new function calls, define as infill points (IPs), in addition to the initial sampling plan. The use of Kriging meta-model is attractive because, not only can it give good predictions of complex landscapes, it also provides a credible estimate of the possible error in these predictions. So, in Kriging-based optimization algorithm, the error estimates make it possible to make tradeoffs between sampling where the current prediction is good (local exploitation) and sampling where there is high uncertainty in the function predictor value (global exploration), allowing searching the decision space efficiently [6].

Kriging-based optimization algorithms start by simulating a limited set of input combinations (referred to as initial sampling) and iteratively select new input combinations to simulate by evaluating an infill criterion (IC), which reflects information from Kriging. The response surface is then updated sequentially with information obtained from the newly simulated IPs. The procedure is repeated until the desired performance level is reached and the estimated optimum is returned [10]. The remainder of this section briefly explains the search and the replication strategy for each algorithm.

3.1 Expected Improvement - EI

The EGO algorithm of [11] chooses the alternative with maximum expected improvement (EI) as the next infill point:

$$EI(\mathbf{x}^+) = \left(y_{min} - \hat{y}(\mathbf{x}^+)\right)\Phi\left(\frac{y_{min} - \hat{y}(\mathbf{x}^+)}{\hat{s}(\mathbf{x}^+)}\right) + \hat{s}(\mathbf{x}^+)\phi\left(\frac{y_{min} - \hat{y}(\mathbf{x}^+)}{\hat{s}(\mathbf{x}^+)}\right) \tag{9}$$

where Φ and ϕ are the cumulative distribution function and probability density function respectively, and y_{min} is the smallest sampled value of y. Maximizing $AEI(x^+)$leads to the new point x^+ with the highest probability of improvement, either by sampling toward the optimum or improving the approximation of the meta-model.

3.2 Augmented Expected Improvement - AEI

The sEGO algorithm of [12] chooses the alternative with maximum augmented expected improvement (AEI) as the next infill point:

$$AEI(x^+) = E[max(y_{min} - \hat{y}, 0)]\left(1 - \frac{\hat{\sigma}_\epsilon^2(x^+)}{\sqrt{\hat{s}^2(x^+) + \hat{\sigma}_\epsilon^2(x^+)}}\right) \tag{10}$$

where $\hat{y}$ is SK predictor, y_{min} is the Kriging prediction at the current effective best solution, i.e., the point with minimum among the simuleted point, with $\beta \in (0, 0.5]$. $\hat{\sigma}_\epsilon^2$ is the variance of the noise intrinsic to the stochastic function and $\hat{s}^2$ is SK variance. The first parcel of the expression is calculated as Equation 9.

4 Numerical test

In this section, the performance of the EGO and sEGO algorithms are compared in the optimization of noisy functions in relation to the sample size approached to approximate the objective function. The basic setting for executing the Kriging-based algorithms was $n_0 = 10$, that represents the number of elements of the initial sample space of the meta-model - adopted $n = 10 \times k$, where k is the dimension of the problem - distributed by the Latin Hypercube [6] and the number of iterations was define for each case.

4.1 Function 01

A one-dimensional problem will be analyzed from [13]. The function $f(\mathbf{x}, \boldsymbol{\theta}) : \mathcal{X} \times \Omega \rightarrow R$ is given by:

$$f(\mathbf{x}, \boldsymbol{\theta}) = 1 + 0,2\theta x + cos\left(0,3(x\theta)^2\right) \tag{11}$$

where $\mathbf{x} \in \mathcal{X} = [-1, 7]$ is the search domain and $\boldsymbol{\theta} \in \Omega$ is the space formed by the uniform distribution random variables $\theta_i \sim U(1, \sigma_i)$. The case will be analyzed for $\sigma_1 = 1$ and $\sigma_2 = 0,2$.

The optimization problem consists of finding the minimizer and the minimum value of $f(\mathbf{x})$. Figure 2 shows the graph of the function for the cases $\theta_0 \sim U(1,0)$ (deterministic), $\theta_1 \sim U(1,1)$ and $\theta_2 \sim U(1,0.2)$. For a convergence of the stochastic curve, 100 000 replications were necessary.

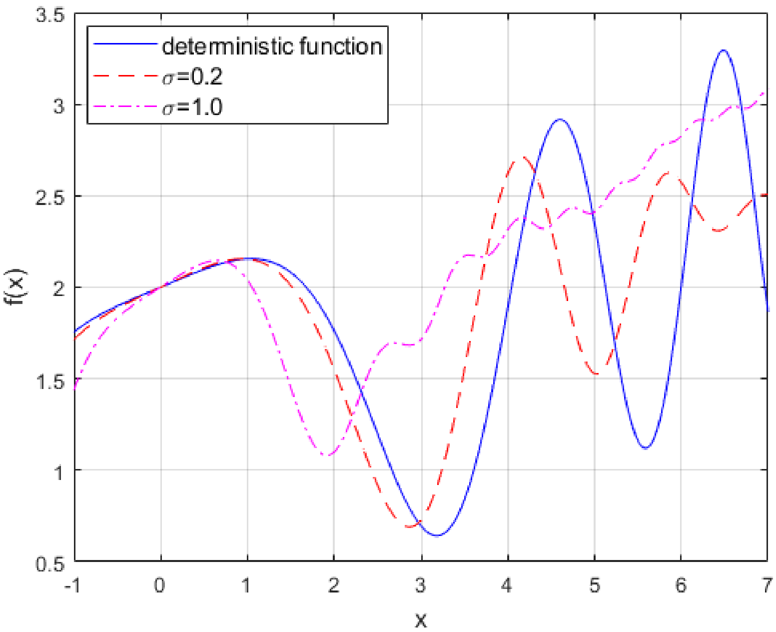

Figure 2. Deterministic and stochastic function plot

The optimal values (f_{min}) and their (x^*) positions found by [13] ware: for $\sigma_1 = 1$, $f_{min} = 1,0840$ and $x_2^* = 1,8999$; for $\sigma_2 = 0.2$, $f_{min} = 0,6889$ and $x_1^* = 2,8709$. In the Figure 4 e Table 1 are presented the minimum value for $\theta \sim U(1,1)$ obtained using the EGO and sEGO algorithm, respectively, for different sample sizes (n_t). The stop criterion of the iterative process was defined by the maximum amount of infill points equal to 20 and it was considered 20 simulations, i.e., 20 repetitions of the optimization process. First, the results was presented in a statistical way using the box plot technique. Since the deterministic Kriging model does not consider noise information from the objective function, the optimization process is performed considering the average of the stochastic parameter sampling.

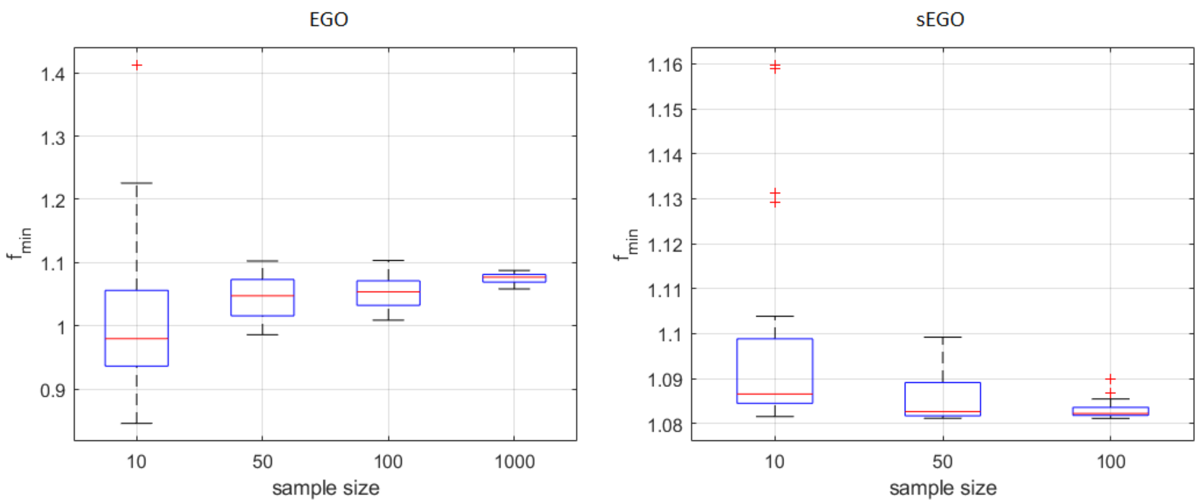

Figure 3. Boxplot for different sample sizes when $\theta \sim U(1,1)$

From Figure 4 it's possible to conclude that by the EGO algorithm there was a greater dispersion of the results presented a convergence with only 1000 samples. With low sample size both methods presented some outliers. Now, compare the average of the values obtained, Table 1, it's possible to observe that the sEGO algorithm presented better performance when reaching the target value with a much smaller sampling. The conclusion is that the sEGO algorithm presented the best performance reaching the target value with just 100 samples, while the EGO needed 100,000.

A similar analysis will be made for the case where $\theta \sim U(1, 0.2)$. In the Figure 4 e Table 1 are presented the minimum value obtained using the EGO and sEGO algorithm, respectively, for different sample sizes (n_t). The stop criterion of the iterative process was defined by the maximum amount of infill points equal to 30 and it was considered 30 simulations.

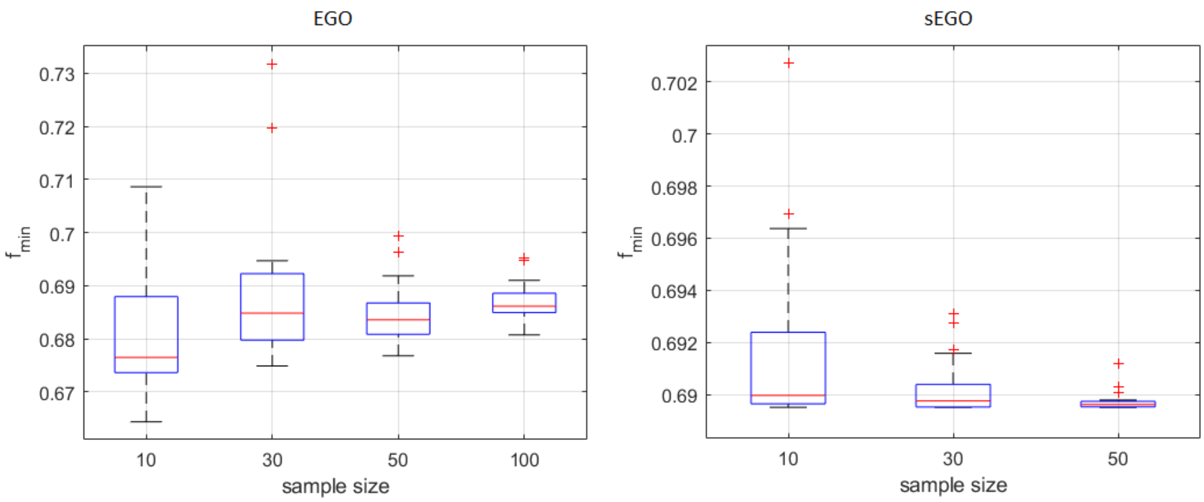

Figure 4. Boxplot for different sample sizes when $\theta \sim U(1, 0.2)$

Table 1. Mean minimum values for Function 01.

	$\theta \sim U(1,1)$		$\theta \sim U(1,0.2)$	
	EGO	sEGO	EGO	sEGO
target value	1,084		0,6889	
sample size				
10	1.0205	1.0989	0.6803	0.6908
50	1.0452	1.0853	0.6846	0.6892
100	1.0539	1.0832	0.6867	-
100,000	1.0831	-	-	-

From Figure 4 it's possible to conclude that by the EGO algorithm there was a greater dispersion of the results and both presented outliers. Now, compare the average of the values obtained, **??**, it's possible to observe that the sEGO algorithm presented better performance, but the both obtained interesting results for small samples.

4.2 Function 02

This second example was adapted from adapted from [14] for stochastic case. The function $f(\mathbf{x}, \boldsymbol{\theta}) : \mathcal{X} \times \Omega \rightarrow R$ is given by:

$$f(\mathbf{x}, \boldsymbol{\theta}) = (6x - 2)^2 sin(12x - 4).\theta \quad (12)$$

where $\mathbf{x} \in \mathcal{X} = [0, 1]$ is the search domain and $\boldsymbol{\theta} \in \Omega$ is the space formed by uniform distribution random variables $\theta_i \sim U(1, \sigma_i)$. The case will be analyzed for $\sigma_1 = 0, 2$.

The plot of Figure 5 shows the input domain, to view the function's key characteristics for the cases $\theta_0 \sim U(1, 0)$ (deterministic), $\theta_1 \sim U(1, 0.2)$ and $\theta_2 \sim U(1, 1)$. For a convergence of the stochastic curve 10 000 replications were necessary.

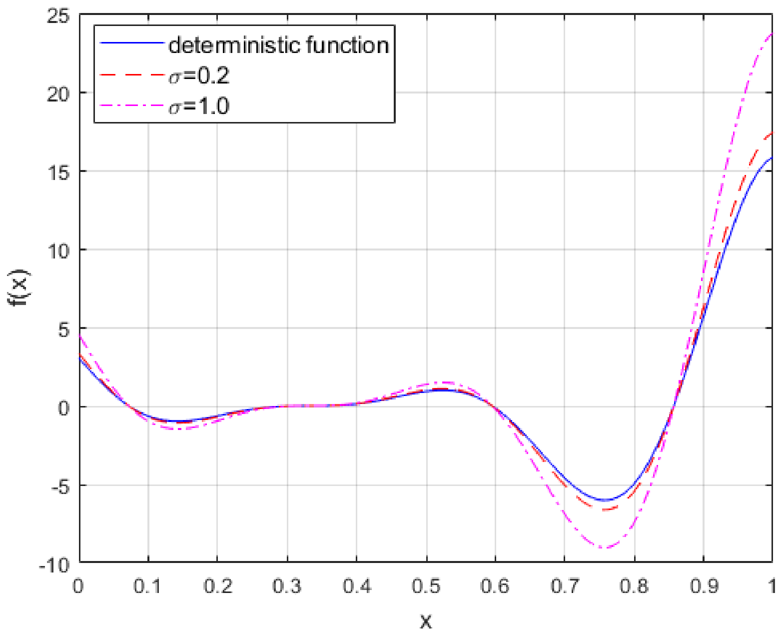

Figure 5. Deterministic and stochastic function plot

The optimal values (f_{min}) were [13]: for $\sigma_1 = 1$, $f_{min} = -9,033$ and for $\sigma_2 = 0.2$, $f_{min} = 0,6889$. In the Figure 6 e Table 2 are presented the minimum value for $\theta \sim U(1, 1)$ obtained using the EGO and sEGO algorithm, respectively, for different sample sizes (n_t). The stop criterion of the iterative process was defined by the maximum amount of infill points equal to 20 and for results was considered 20 simulations, i.e., 20 repetitions of the optimization process. First, the results was presented in a statistical way using the box plot technique. Since the deterministic Kriging model does not consider noise information from the objective function, the optimization process is performed considering the average of the stochastic parameter sampling.

Table 2. Mean minimum values for Function 02.

sample size	EGO	sEGO
target value	-9,033	
20	-9.5408	-9.0160
50	-9.3388	-9.0237
100	-9.2029	-

From Figure 6 it's possible to conclude that by the EGO algorithm there was a greater dispersion of the results. And through sEGO there was greater convergence of the result with some outliners. Now, compare the average of the values obtained, Table 2, it's possible to observe that the sEGO algorithm presented better performance when reaching the target value with a much smaller sampling in contrast with the EGO algorithm that did not achieve accuracy in the result. The conclusion is that the sEGO algorithm presented the best performance reaching the target value.

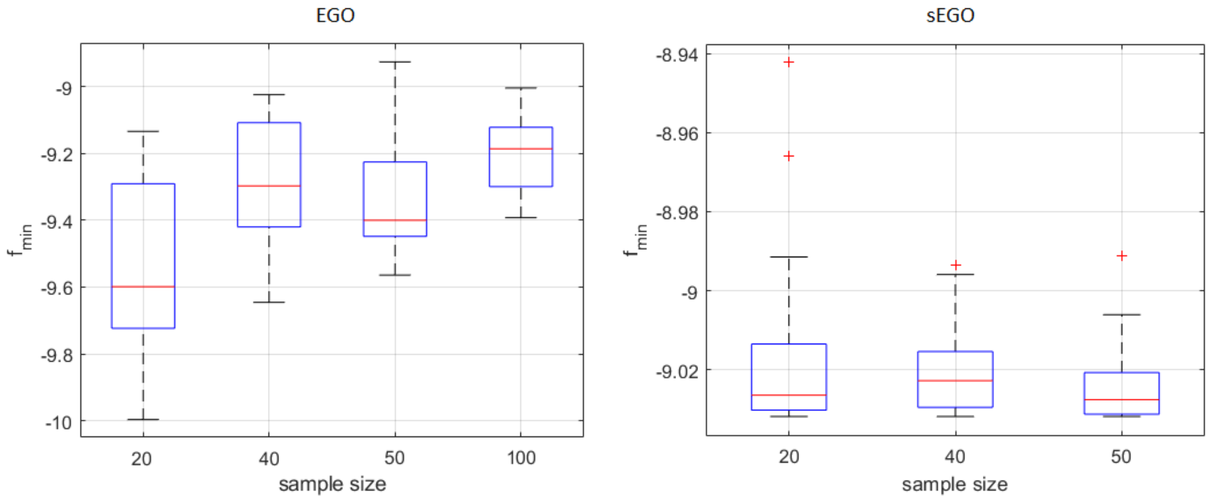

Figure 6. Boxplot for different sample sizes when $\theta \sim U(1, 1)$

5 Conclusions

In this article, the objective was to compare the performance of two optimization algorithms based on the Kriging meta-model applied to noisy objective functions. The first was EGO, a very popular algorithm used in the optimization of deterministic problems. The second was sEGO, developed from the EGO incorporating information from the heterogeneous variance of the noisy objective function. That said, it was possible to observe from the results that the quality of the solutions returned by the EGO when the function presented high noise variance was strongly affected by its inability to identify good solutions for smaller samples, since it does not address variance information in the iterative process. For problems with low variance, incorporating information from the same in the process had low representation, in this case, both algorithms presented good performance for a small sample size. In general, the EGO may not be very appropriate, as it ignores noise in the observations, assuming that samples were taken with infinite precision and the sEGO algorithm presented the best performance with an optimal value closer to the target value for small sample sizes. The use of Kriging-based algorithms for optimizing modeled systems through stochastic simulation, especially with heterogeneous noise, is relatively new and has great research potential.

Acknowledgements. The authors are grateful for the support by CAPES through PROAP (Graduate Support Program).

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] B. Sudret and A. Der Kiureghian. *Stochastic finite element methods and reliability: a state-of-the-art report*. Department of Civil and Environmental Engineering, University of California . . . , 2000.
[2] A. I. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in aerospace sciences*, vol. 45, n. 1-3, pp. 50–79, 2009.
[3] V. Picheny, T. Wagner, and D. Ginsbourger. A benchmark of kriging-based infill criteria for noisy optimization. *Structural and multidisciplinary optimization*, vol. 48, n. 3, pp. 607–626, 2013.
[4] W. C. Van Beers and J. P. Kleijnen. Kriging for interpolation in random simulation. *Journal of the Operational Research Society*, vol. 54, n. 3, pp. 255–262, 2003.
[5] B. Ankenman, B. L. Nelson, and J. Staum. Stochastic kriging for simulation metamodeling. In *2008 Winter Simulation Conference*, pp. 362–370. IEEE, 2008.

[6] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, vol. 13, n. 4, pp. 455–492, 1998.
[7] H. Qu and M. C. Fu. Gradient extrapolated stochastic kriging. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 24, n. 4, pp. 1–25, 2014.
[8] H. Jalali, I. Van Nieuwenhuyse, and V. Picheny. Comparison of kriging-based algorithms for simulation optimization with heterogeneous noise. *European Journal of Operational Research*, vol. 261, n. 1, pp. 279–301, 2017.
[9] F. Carraro, R. H. Lopez, L. F. F. Miguel, and A. J. Torii. Monte carlo integration with adaptive variance selection for improved stochastic efficient global optimization. *Structural and Multidisciplinary Optimization*, vol. 60, n. 1, pp. 245–268, 2019.
[10] S. Rojas-Gonzalez and I. Van Nieuwenhuyse. A survey on kriging-based infill algorithms for multiobjective simulation optimization. *Computers & Operations Research*, vol. 116, pp. 104869, 2020.
[11] D. Huang, T. T. Allen, W. I. Notz, and N. Zeng. Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of global optimization*, vol. 34, n. 3, pp. 441–466, 2006.
[12] F. Carraro and others. A stochastic kriging approach for the minimization of integrals, 2017.
[13] F. F. d. S. Nascentes and others. Contribuições à eficiência da otimização global estocástica adaptativa, 2019.
[14] A. I. Forrester, A. J. Keane, and N. W. Bressloff. Design and analysis of" noisy" computer experiments. *AIAA journal*, vol. 44, n. 10, pp. 2331–2339, 2006.