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

# Comparison of Kriging-based algorithms for optimization with heterogeneous noise

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.**
Problem modeling through response surfaces, or meta-models, has been a great solution adopted for optimizing problems with high computational cost, especially Kriging-based optimization algorithms. In recent years, algorithms have been proposed which extend the traditional Kriging-based simulation optimization algorithms (assuming deterministic outputs) to problems in the presence of noise or uncertainty. This paper approaching stochastic kriging meta-model in a comparative study of the performance of three Kriging-based algorithms for unconstrained minimization a noisy function. The Minimum Quantile criterion (MQ), stochastic Efficient Global Optimization (sEGO) and Expected Improvement with Reinterpolation (EIR) will be the algorithms compared using an analytical test function. The conclusions and insights obtained may serve as a useful guideline for researchers aiming to deal with optimization problems, especially to apply Kriging-based algorithms to solve engineering problems, and may be useful in the development of future algorithms.

**Keywords:** Optimization problem, stochastic Kriging, heterogeneous noise, algorithms

## 1 Introduction

Recent computational advances have allowed the mathematical models of engineering problems to become highly complex, including a greater amount of data, details and refinements. 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 these procedures 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 level of refinement of the models and the representation of randomness in the analyzes led us to optimization problems that have a high computational cost associated with objective functions of difficult analytical treatment. So, an efficient optimization algorithm must be able to find the best result in the shortest time. An alternative to high cost functions optimization is approach the meta-models based optimization algorithms. The primary motivation for using meta-models in simulation optimization is to reduce the number of expensive fitness evaluations without degrading the quality of the obtained optimal solution.

One of the most popular meta-model is the Kriging, which has a long and successful tradition for modeling and optimizing deterministic computer simulations [2]. Kriging is a Gaussian based meta-model, also known as a Gaussian Process Regression model (GPR), whose purpose is to build a predictor surface of the objective function based on known observations of points in the design domain [3]. The results are now predicted using this predictor without resorting to the use of the primary source (objective function). 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].

The representation of the stochastic problem occurs through objective functions that can be formulated as expected value functions $E[f(\mathbf{x}, \boldsymbol{\theta})]$, where $\mathbf{x} \in \mathcal{R}^n$ is the vector of design variables with dimension $n$ and $\theta \in \mathcal{R}^{n_t}$ is the vector of stochastic parameters of dimension $n_t$. Figure 1 shows an example of Kriging surface based on noisy observations, $f(\mathbf{x}, \boldsymbol{\theta})$, where the bar is the noise amplitude.

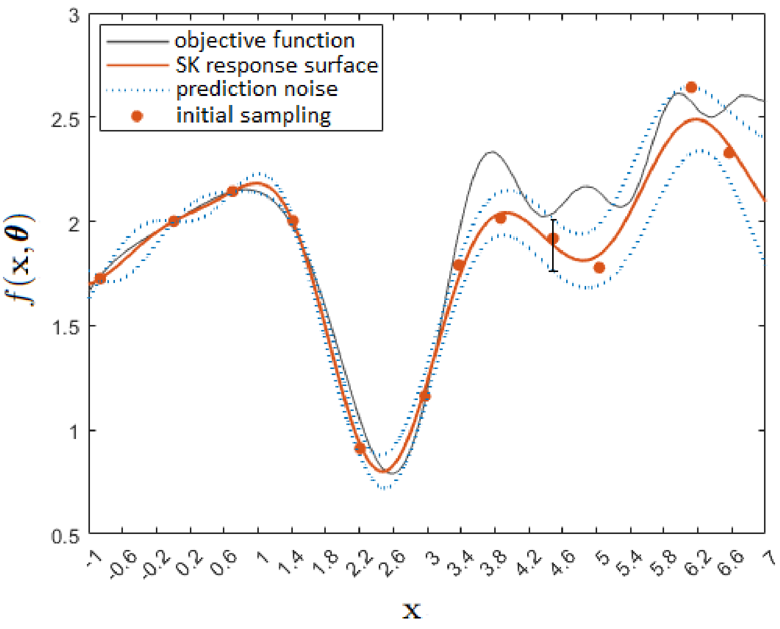


Figure 1. Kriging meta-model on noisy observations

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 (see [7] for references on the use of EGO in constrained and multi-objective optimization problems).

In stochastic simulation where the design variables are discrete, for example, the EGO may not be very appropriate, as it ignores the noise in the observations, assuming that samples were taken with infinite precision [8]. Research has been developed to extend the EGO to stochastic simulation, where most approaches assume homogeneous simulation noise, which means that the noise variance does not depend on the $x$ position. In [2] is compared several algorithms based on Kriging to optimize functions with this kind of noise. In practice, however, the noise is heterogeneous, and in the work of [9] they made a performance comparison between the optimization algorithms based on Kriging that can handle heterogeneous noise, among them the Minimum Quantile criterion (MQ) [10]. Another approach is the stochastic Efficient Global Optimization (sEGO), proposed by [11].

The methods mentioned above are said to be single objective, in this case the optimization process occurs in relation to only one objective and the stochastic noise treatment information is represented directly in its formulation. Another approach is multi-objective optimization algorithms, where the expected improvement in the optimization process is measured against each separate objective, in general, the noise treatment strategy in these methods comprises a second iterative process. An example of a multi-objective optimization algorithm is the Expected Improvement with Reinterpolation (EIR) method approach by [12] for the case of stochastic Kriging with heterogeneous variance.

The goal of this paper is compare the performance of those three Stochastic Kriging-based algorithm optimization - MQ, sEGO and EIR - in a two analytic tests functions. The stochastic kriging metamodel will be construt according to [5], with representation of the heterogeneous noise of the function and considered only box constraints. This article is organized as follows. Section 2 details the Kriging-based optimization algorithms, Section 3 presents the analysis and results of the problems and, lastly, conclusions are present in Section 4.

## 2 Kriging meta-model based optimization

Because the surrogate model, $\hat{y}$, is only an approximation of the true function $f(\mathbf{x}, \boldsymbol{\theta})$ we wish to optimize, enhancing 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 algorithms, 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 [13]. The remainder of this section briefly explains the search and the replication strategy for each algorithm.

### 2.1 Minimum Quantile criterion - MQ

The minimum quantile criterion (MQ) was initially proposed by [14] and consists of carrying out a balance between the global and the local exploration selecting as the next IP the point that minimizes a percentile of the predictor obtained by SK, that is, using a weighted sum between $\hat{y}$ and $s^2$. That said, the quantile of the predicted value is given by Equation 1 [9].

$$MQ(x^+) = \hat{y} + \Phi^{-1}\beta s^2(x^+) \tag{1}$$

where $\hat{y}$ is SK predictor; $s^2$ is the standard deviation obtained by the square root of $\hat{y}$, $\Phi$ is the cumulative probability density of the normal distribution and $\beta = 0.5$. So, the infill point in each iteration is:

$$\mathbf{x} = \arg\min_{x \in \mathcal{X}} MQ(x^+) \tag{2}$$

In this method does not require information on SK variance ($\hat{s}^2$).

### 2.2 stochastic Efficient Global Optimization - sEGO

The sEGO algorithm by [11] chooses the alternative with maximum augmented expected improvement (AEI) as the next infill point:

$$AEI(x^+) = E[max(y_{min} - \hat{y}, 0)] \left(1 - \frac{\hat{\sigma}^2_\epsilon(x^+)}{\sqrt{\hat{s}^2(x^+) + \hat{\sigma}^2_\epsilon(x^+)}}\right) \tag{3}$$

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 $\beta = 0.84$. $\hat{\sigma}^2_\epsilon$ is the variance of the noise intrinsic to the stochastic function and $\hat{s}^2$ is SK variance. The first parcel of the expression could be calculated as:

$$EI(\mathbf{x}^+) = \left(y_{min} - \hat{y}(\mathbf{x}^+)\right) \Phi\left(\frac{y_{min} - \hat{y}(\mathbf{x}^+)}{\hat{s}(\mathbf{x}^+)}\right) + \hat{s}(\mathbf{x}^+)\phi\left(\frac{y_{min} - \hat{y}(\mathbf{x}^+)}{\hat{s}(\mathbf{x}^+)}\right) \tag{4}$$

where $\Phi$ and $\phi$ are the cumulative distribution function and probability density function respectively, and $y_{min}$ is the smallest sampled value of $y$. The next IP is finded maximizing $AEI(x^+)$, i.e., 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.

### 2.3 Expected Improvement with Reinterpolation - EIR

Approached by [12], the method proposes that, instead of modifying the EI for cases stochastic, let's use SK and deterministic Kriging together like noise-handling strategie.

After construction the predictions by SK at the support points, those will be used to build a new model in Kriging, now deterministic, since the predictions will come free of intrinsic error. As this last model is noise free, the classic EI [15] could be used as a metric to obtain new IPs. So, the Kriging prediction for the deterministic case will be rewritten as:

$$\hat{y} = \hat{\mu} + \mathbf{h}^T\Psi^{-1}(\hat{\mathbf{y}}_r - \mathbf{1}\hat{\mu}) \tag{5}$$

$$\hat{\mu} = \frac{\mathbf{1}\Psi^{-1}\hat{\mathbf{y}}_r}{\mathbf{1}\Psi^{-1}\mathbf{1}} \tag{6}$$

where $\hat{\mu}$ is the trend of the deterministic Kriging meta-model obtained with information about the response surface constructed through SK model ($\hat{y}_r$), $\mathbf{h}$ is the correlation vector, $\Psi$ is the covariance matrix of the all support points and $\overline{y}$ is the vector of the approximate mean value of the objective function at each design point.

And, as we can use the SK predictor itself, the re-interpolation will only need the value of the new spatial variance of the correlation between the support points.

$$\hat{\sigma}^2_{ri} = \frac{1}{n}\left[(\overline{y} - 1\hat{\mu}_r)^T \hat{\sum}^{-1} \psi \hat{\sum}^{-1} (\overline{y} - 1\hat{\mu}_r)\right] \quad (7)$$

# 3 Numerical test

## 3.1 One-dimension problem

A one-dimensional problem will be analyzed adapted from [15] for stochastic case. The function $f(\mathbf{x}, \boldsymbol{\theta}) : \mathcal{X} \times \Omega \rightarrow R$ is given by:

$$f(\mathbf{x}, \boldsymbol{\theta}) = (6x - 2)^2 sin(12x - 4).\theta \quad (8)$$

where $\mathbf{x} \in \mathcal{X} = [0, 1]$ is the search domain and $\boldsymbol{\theta} \in \Omega$ is formed by the normally distributed random variables $\theta \sim \mathcal{N}(1, 1)$. The plot of Figure 2 shows the input domain, to view the function's key characteristics. It is possible to observe that the variance of the stochastic parameter increases and oscile as it approaches the upper limit of the search domain. In order to obtain a reliable construction of the function curve in its stochastic forms, 100,000 replications of the function per point were used.

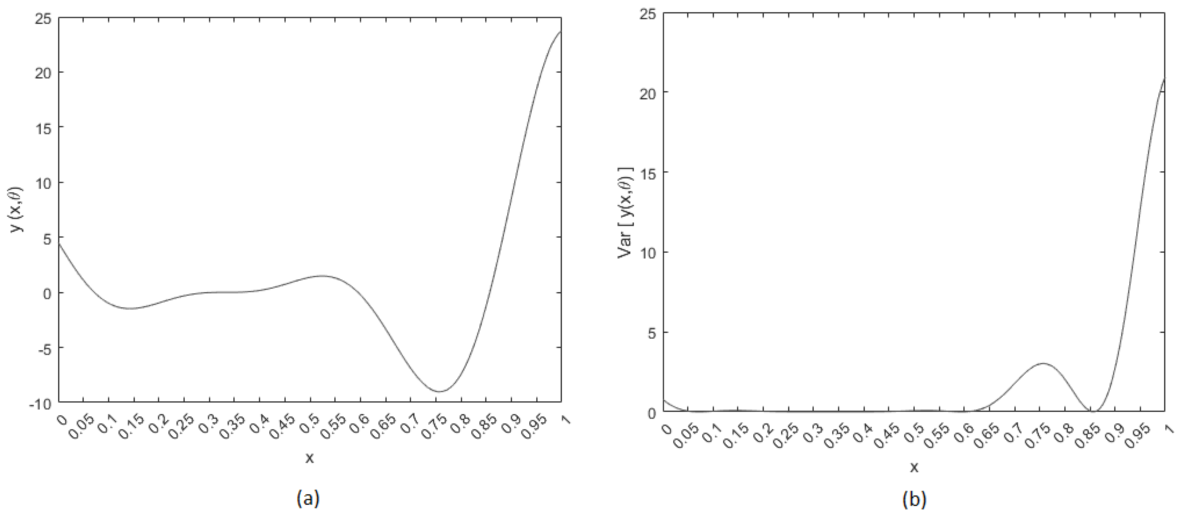


Figure 2. One-dimension problem function (a) expected value; (b) variance

In Figure 3 the results was presented in a statistical way using the box plot technique for 20 simulations from optimization algorithms. It is possible to conclude that the three algorithms obtained minimum values close to the target value. All presented symmetry in the results, being the sEGO and EIR algorithms the ones that obtained greater convergence. And in the MQ the outliers were more significant.

Comparing the average of the values obtained for the twenty simulations, presented in Table 1, we concluded that the algorithm that presented the best result was the sEGO. The basic settings for executing the Kriging-based algorithms are given in Table 3.

## 3.2 Two-dimension problem

In this section a two-dimensional problem will be analyzed adapted from [16] for stochastic case. The function $f(\mathbf{x}, \boldsymbol{\theta}) : \mathcal{X} \times \Omega \rightarrow R$ is given by:

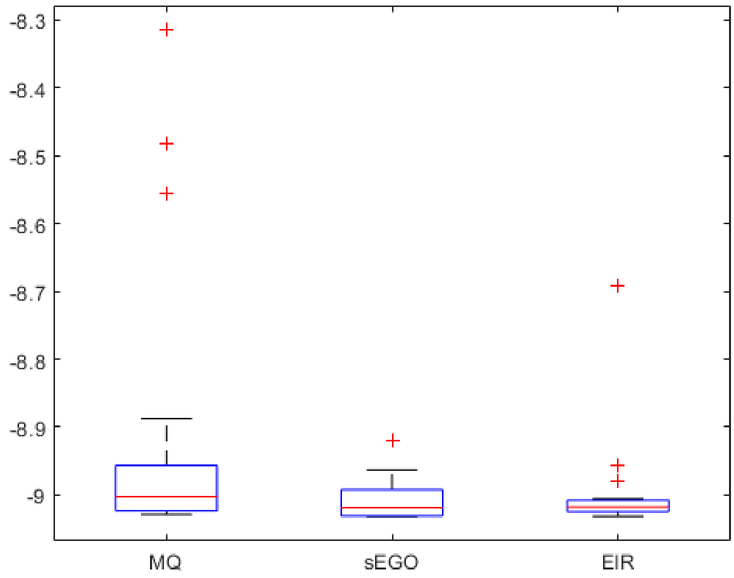


Figure 3. Boxplot for 1D problem result

Table 1. Minimum values for 1D problem.

| Algorithms | $\overline{y}_{min}$ |
|---|---|
| target value | -9,033 |
| MQ | -8,8005 |
| sEGO | -9,0104 |
| EIR | -8,9961 |

$$f(\mathbf{x},\boldsymbol{\theta}) = \left(4 - 2,1x_1^2 + \frac{x_1^4}{3}\right) x_1^2.\theta_1 + x_1.x_2.\theta_2 + \left(-4 + 4x_2^2\right) x_2^2.\theta_3 \quad (9)$$

where $\mathbf{x} \in \mathcal{X} = [-1,1] \times [-2,2]$ is the search domain and $\boldsymbol{\theta} \in \Omega$ is the two-dimensional space formed by the normally distributed random variables $\theta_i \sim \mathcal{N}(1, \sigma_{xi})$. The case will be analyzed for $\sigma_{x1} = 1$, $\sigma_{x2} = 0,8$ and $\sigma_{x2} = 0,5$. The plot in Figure 4 shows the input domain where it's possible to view the function's key characteristics. The function has three local minima, one of which are global.

In Figure 5 the results was presented in a statistical way using the box plot technique for 12 simulations from optimization algorithms. It is possible to conclude that by the MQ algorithm there was a greater dispersion of the results with a greater tendency of the data above the median. None of the methods showed discrepancies in values, i.e., outliers. And the method with the best performance was the sEGO. Now, comparing the average of the values obtained for twelve simulations, presented in Table 2, it's concluded that all algorithms presented similar result. The basic settings for executing the Kriging-based algorithms are given in Table 3.

Table 2. Minimum values for 2D problem.

| Algorithms | $\overline{y}_{min}$ |
|---|---|
| MQ | -1,2827 |
| sEGO | -1,2885 |
| EIR | -1,2854 |

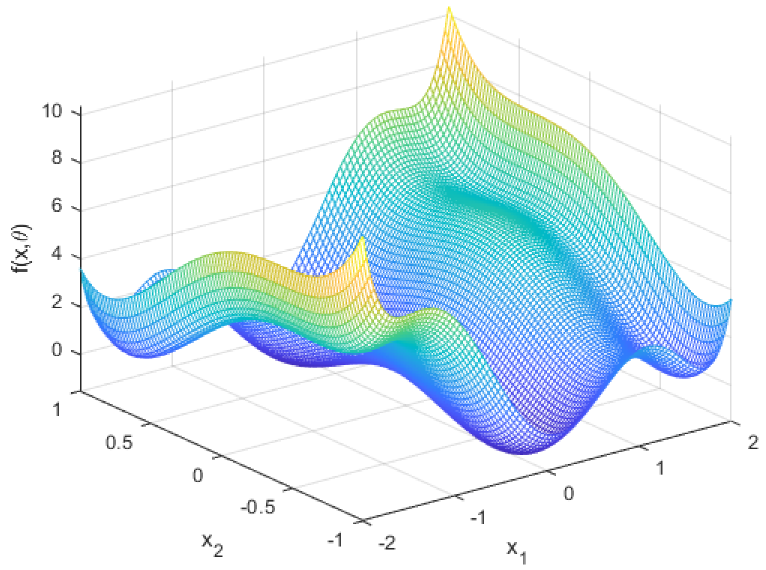


Figure 4. Two-dimension problem function

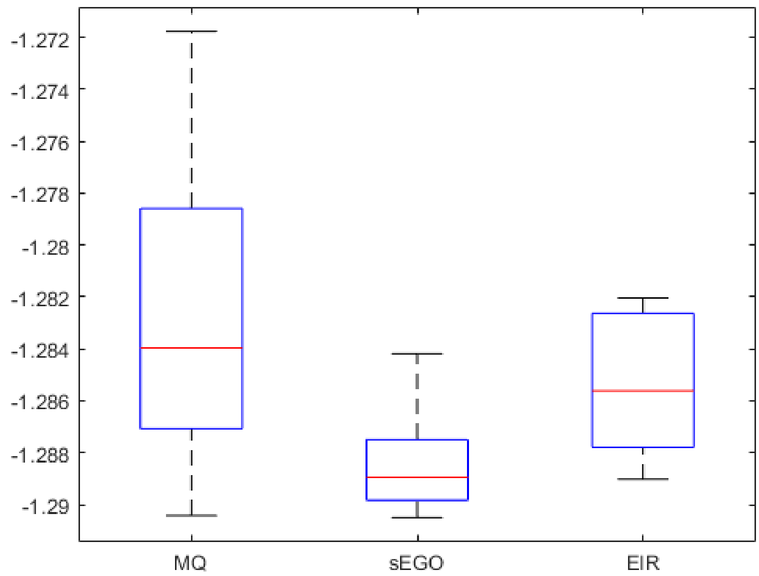


Figure 5. Boxplot for 1D problem result

## 4 Conclusions

In this paper, three algorithms - MQ, sEGO and EIR - based on Kriging for optimization via simulation with heterogeneous noise were compared. The MQ algorithm is the only one that does not address information about the noise variance of the stochastic parameter. That said, it was possible to observe that the quality of the solutions returned by the MQ was strongly affected by its inability to identify good solutions, since it does not address variance information in the iterative process. In general, the sEGO algorithm was the one that presented the best performance with an optimal value closer to the target value and a small variance of results. In summary, in the examples studied, it was essential to combine the algorithms with more intelligent replication strategies that address noise handling of the stochastic function - sEGO and EIR. The use of Kriging-based algorithms for optimizing modeled systems through stochastic simulation, especially with heterogeneous noise, is relatively new and provides an active research area.

Table 3. Valores mínimos da função.

| Parameters Problem 1D | Parameters Problem 2D | Description |
|---|---|---|
| $n_{rep} = 20$ | $n_{rep} = 12$ | The number of simulations, i.e., number of repetitions of the optimization process. |
| $n_0 = 10$ | $n_0 = 20$ | The number of elements of the initial sample space of the meta-model will be adopted $n = 10 \times k$, where k is the dimension of the problem, distributed by the Latin Hypercube [6]. |
| $n_t = 20$ | $n_t = 40$ | The number of elements of the initial sample space of stochastic parameter. |
| $NIPs = 20$ | $NIPs = 40$ | The stop criterion of the iterative process was defined by the maximum amount of infill points. |

**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] 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.
[3] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical science*, vol. 4, n. 4, pp. 409–423, 1989.
[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] J. P. Kleijnen and E. Mehdad. Estimating the variance of the predictor in stochastic kriging. *Simulation Modelling Practice and Theory*, vol. 66, pp. 166–173, 2016.
[8] 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.
[9] 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.
[10] V. Picheny and D. Ginsbourger. Noisy kriging-based optimization methods: a unified implementation within the diceoptim package. *Computational Statistics & Data Analysis*, vol. 71, pp. 1035–1053, 2014.
[11] 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.
[12] F. F. d. S. Nascentes and others. Contribuições à eficiência da otimização global estocástica adaptativa, 2019.
[13] 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.
[14] D. D. Cox and S. John. A statistical method for global optimization. In *[Proceedings] 1992 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1241–1246. IEEE, 1992.
[15] 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.
[16] M. Molga and C. Smutnicki. Test functions for optimization needs. *Test functions for optimization needs*, vol. 101, pp. 48, 2005.