# On Benchmarking Surrogate-Assisted Evolutionary Algorithms

Vanessa Volz
Queen Mary University of London, UK
v.volz@qmul.ac.uk

Boris Naujoks
TH Köln - University of Applied Sciences, Germany
boris.naujoks@th-koeln.de

## ABSTRACT

In this position paper, we discuss the need for systematic benchmarking of surrogate-assisted evolutionary algorithms and give an overview of existing suitable function suites. Based on the findings, we hope to encourage more comparative studies in this field supported by benchmarks and outline how a concerted effort of the community could create better insight into the various previously proposed algorithms and concepts.

**ACM Reference Format:**
Vanessa Volz and Boris Naujoks. 2019. On Benchmarking Surrogate-Assisted Evolutionary Algorithms. In *Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion), July 13–17, 2019, Prague, Czech Republic.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3319619.3326866

## 1 INTRODUCTION

As a result of their typically exploratory approach, evolutionary algorithms tend to require a relatively large number of function evaluations until convergence or until a suitable solution is found. This becomes an issue when evaluating a solution is computationally / economically expensive, which is common in real-world applications. To alleviate this problem, surrogate-assisted evolutionary algorithms (SAEAs) have been proposed. They are a subclass of evolutionary approaches, where the search is augmented by predictions of fitness from a surrogate model. The intention is to discover solutions of the optimisation problem using fewer evaluations of the fitness function. An overview of such methods is given in section 2.

As demonstrated in the overview, several approaches for combining a surrogate model and an evolutionary algorithm exist. It is, however, unclear which approach is most useful in a given setting. There are several other open questions as discussed in section 3 that can be answered by benchmarking SAEAs systematically. Having thus established the potential of benchmarking SAEAs, we give an overview of suitable benchmarks in section 4. We specifically focus on real-world benchmarks and benchmarks with expensive fitness functions, which is the usecase SAEAs were designed for. We also present a summary of preliminary results on these benchmarks which clearly demonstrate the need for further, larger scale experiments. In our concluding remarks in section 5, we argue the need for more widespread use of comparable benchmarks in research on SAEAs. We further discuss how existing benchmarks could be improved with regards to their suitability and practicality for benchmarking SAEAs in the future.


*GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic*

ACM ISBN 978-1-4503-6748-6/19/07…$15.00
https://doi.org/10.1145/3319619.3326866

## 2 OVERVIEW OF EXISTING SURROGATE-ASSISTED APPROACHES

In a survey of current literature on SAEAs, we have identified two main approaches: iterative sampling (e.g. [6, 7]) and evolution control (survey in [5]). Both are described in more detail below.

### 2.1 Iterative Sampling

Iterative sampling methods seek to improve a surrogate model throughout the runtime of the algorithm. This process is usually guided by a function (also called infill criterion) that expresses both the predicted accuracy of the model after adding the new sample, as well as the estimated progress regarding the original expensive fitness function. If an evolutionary algorithm is used to optimise the infill criterion, iterative sampling methods fall under the umbrella of SAEAs. Such methods usually start from a sample obtained using a space-filling design of experiments. Based on the this first sample, an initial model is build and the evolutionary algorithm starts. For a clear distinction, it is important to note that in iterative sampling approaches, the optimisation algorithm works on the model exclusively and does not trigger evaluations of the true expensive fitness function during its execution.

### 2.2 Evolution Control

In evolution control methods, one or multiple steps in the algorithmic skeleton of an EA are supported by a surrogate model. The method for the integration of surrogate model and evolutionary algorithm is often called model management strategy and can generally be classified as either individual-based, generation-based or population-based [5].

Evolution control approaches can be further characterised by the step of an evolutionary algorithm that is augmented by the surrogate model. For example, the random generation of offspring for a new generation might be biased using information from a model (e.g. pre-screening [3]). More recently, it has also been suggested to use uncertainty information on the predictions in order to introduce *lazy evaluation* into evolutionary algorithms (e.g. GP-DEMO [8] and SAPEO [9, 11]). These algorithms will only evaluate individuals in a population if their comparison based solely on information obtained through the model is too risky according to a given statistical significance level.

## 3 OPEN QUESTIONS AND ISSUES

Several open questions have not been addressed by the research community on a holistic level. We list some of them below.

*How Expensive?* One central question to answer is at what point an optimisation problem is expensive "enough" to warrant the application of surrogate-assisted methods. Training the surrogate model does require a certain amount of computational effort, especially if global models are used, which can result in longer runtimes for the SAEA, even if fewer function evaluations are used.

*Suitability?* As has been shown in [10], surrogate-assisted algorithms are not guaranteed to use fewer function evaluations than the baseline EA in order to discover solutions of the same fitness. It is thus important to find strategies to identify which types of optimisation problems are suitable for SAEAs. The importance is corroborated by the assumed high cost of the fitness function, which does not allow for extensive experiments to identify suitability.

*Model Assumptions.* As surrogate models are intended to compute reliable predictions based on a small number of samples, most modelling approaches rely on assumptions on the fitness landscape. In many SAEAs, the models are used without due consideration of assumptions, often because characteristics of the landscape are unknown and difficult to determine due to the associated costs. As models are often used without further validation, the likely detrimental effects of violating these assumptions should be evaluated.

*Comparison of Approaches.* Some of the different algorithms proposed in the past are difficult to compare due to the varying budget spent on improving the model versus finding good optimisation solutions. In addition, several parametrisation choices have to be made when implementing a SAEA. This includes the choice of (global or local) surrogate model as well as associated parameters and assumptions as e.g. determined by the choice of the kernel in a Gaussian process model. There is also a large variety of models that is as of yet unexplored in an evolutionary optimisation context. A further question is which evolutionary algorithm is used and how the corresponding variation operators that determine the manner of traversal of the search space interact with the model assumptions on the fitness landscape [1].

## 4 ANSWERS THROUGH BENCHMARKS?

A large number of the questions and issues listed in section 3 could be investigated through large-scale comparisons of different SAEAs and their performance on different problems. In the following, we discuss the practicality of using benchmarks for these purposes.

### 4.1 Existing Benchmarks

*Black-Box-Optimisation Benchmark (BBOB) and related function suites.* The BBOB and its variants (noisy, bi-objective, mixed-integer) are popular in continuous evolutionary optimisation and implemented using the COCO (COmparing Continuous Optimisers) framework [4]. While the suites are designed to cover a wide range of functions, all are artificially designed. It is thus unclear whether they reflect the challenges of fitness landscapes observed in real-world problems. This is especially true for the generation of artificial noise, which often assumes that observations follow specific probability distributions. To benchmark SAEAs, BBOB functions should thus be used in conjunction with benchmarks that resemble real-world problems more.

*Computational Fluid Dynamics (CFD).* CFD simulations are often used in real-world problems, such as the design of aircrafts. A function suite composed of 3 such problems has been proposed recently [2]. The functions in this suite thus resemble real-world problems, albeit only a specific (small) subset of them. However, a much larger suite would probably be impractical to run due to the cost of the evaluations. Still, for more robust performance assessment, benchmarking results from the CFD suite should probably be contrasted with further insights from different benchmarks.

*Game-Benchmark for Evolutionary Algorithms (GBEA).* The GBEA is a collection of recently proposed function suites intended to capture more real-world-like complexities in fitness landscapes using game-related problems. It contains a variety of different functions that are comparably fast to compute. An additional framework described in [10] also features pre-implemented surrogate-assisted evolutionary approaches as well as automatic logging of model predictions and associated errors[1]. Further work is needed to understand the various fitness landscapes and to identify performance patterns.

### 4.2 Preliminary Results

We have obtained some preliminary results on the questions posed in section 3 by running several SAEAs on the BBOB and GBEA benchmarks, which were recently published in [10]. We found that, while there are functions where an improvement can be achieved using surrogate-assisted algorithms, the underlying evolutionary algorithms tend to perform on par or even better. This is especially true for the single-objective versions and for higher budgets. Some SAEAs seem more suitable to identify satisfactory results quickly, while others are more balanced in terms of achievable fitness versus required function evaluations. Performance varies significantly across functions and model validation tends to improve it. Further insights, however, require a much larger scale of experiments.

## 5 CONCLUSIONS

We have discussed some open questions in surrogate-assisted evolutionary algorithms research, and suggested benchmarks as a potential approach for finding answers. We were able to identify several suitable benchmarking suites and obtain preliminary results on some of these questions, pending further experiments. The main issue, however, is that large-scale experiments are required to confidently draw generalisable conclusions on any of these questions.

The point of this paper is thus to draw attention to the potential of benchmarking SAEAs, if done as a concerted effort of the community. Benchmarking should become a standard procedure, ideally through automated and interchangeable interfaces and using diverse real-world-like function suites. Benchmarking results should be publicly available in a standardised format to allow easy numerical comparisons as well as plots of the achieved results. We host a website[2] intended to facilitate sharing and obtaining such results. We also provide automatically generated visualisations and statistical performance tests using the COCO post-processing features if the data is provided in the corresponding format (used by BBOB and GBEA benchmarks).

We hope that by facilitating benchmarking for SAEAs, results can be obtained on a large enough scale to answer the questions described in this paper. Further work also remains to be done on the analysis of existing benchmarks to allow for better interpretations.

[1]Code available at: https://github.com/TheHedgeify/uncertaincoco
[2]http://norvig.eecs.qmul.ac.uk/gbea/results.html

## ACKNOWLEDGMENTS

B. Naujoks acknowledges support by the European Commission's H2020 programme through the UTOPIAE Marie Curie Innovative Training Network, H2020- MSCA-ITN-2016, under Grant Agreement No. 722734.

## REFERENCES

[1] T. Chugh et al. "Towards Better Integration of Surrogate Models in Optimizers". In: *High-Performance Simulation Based Optimization*. Ed. by T. Bartz-Beielstein, B. Filipic, and E. Korosec. Springer, 2019, To appear.

[2] S. J. Daniels et al. "A Suite of Computationally Expensive Shape Optimisation Problems Using Computational Fluid Dynamics". In: *Parallel Problem Solving from Nature (PPSN XV)*. Ed. by A. Auger et al. Springer, Cham, Switzerland, 2018, pp. 296–307.

[3] M. Emmerich, K. Giannakoglou, and B. Naujoks. "Single- and Multi-objective Evolutionary Optimization Assisted by Gaussian Random Field Metamodels". In: *IEEE Transactions on Evolutionary Computation* 10.4 (2006), pp. 421–439.

[4] N. Hansen et al. "COCO: A Platform for Comparing Continuous Optimizers in a Black-Box Setting". In: *CoRR* abs/1603.08785 (2016). arXiv: 1603.08785.

[5] Y. Jin. "Surrogate-assisted evolutionary computation: Recent advances and future challenges". In: *Swarm and Evolutionary Computation* 1.2 (2011), pp. 61–70.

[6] D. Jones, M. Schonlau, and W. Welch. "Efficient global optimization of expensive black-box functions". In: *Journal of Global Optimization* 13.4 (1998), pp. 455–492.

[7] J. Knowles. "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems". In: *IEEE Transactions on Evolutionary Computation* 10.1 (2006), pp. 50–66.

[8] M. Mlakar et al. "GP-DEMO: Differential Evolution for Multiobjective Optimization based on Gaussian Process models". In: *European Journal of Operational Research* 243.2 (2015), pp. 347–361.

[9] V. Volz, G. Rudolph, and B. Naujoks. "Surrogate-Assisted Partial Order-Based Evolutionary Optimisation". In: *Evolutionary Multi-Criterion Optimization (EMO)*. Ed. by H. Trautmann et al. Springer, Berlin, 2017, pp. 639–653.

[10] V. Volz. "Uncertainty Handling in Surrogate Assisted Optimisation of Games". PhD thesis. TU Dortmund University, Germany, 2019.

[11] V. Volz, G. Rudolph, and B. Naujoks. "Investigating Uncertainty Propagation in Surrogate-Assisted Evolutionary Algorithms". In: *Genetic and Evolutionary Computation Conference (GECCO)*. ACM Press, New York, 2017, pp. 881–888.