

Investigating Uncertainties with a Game-Benchmark

VANESSA VOLZ¹, BORIS NAUJOKS²

¹Game AI research group, Queen Mary University London, UK

²Institute for Data Science, Engineering, and Analytics, TH Köln, Germany
v.volz@qmul.ac.uk, boris.naujoks@th-koeln.de

Keywords: benchmarking, uncertainty analysis, games, evolutionary algorithms.

ABSTRACT

Most real-world problems cannot be fully specified, and thus contain uncertainties such as noise or modelling bias. As these uncertainties can have a considerable impact on optimisation results, investigating these uncertainties in more depth is becoming increasingly popular in multiple research communities. However, conducting large-scale and in-depth investigations of real-world problems is often difficult, because the corresponding problems are often expensive to evaluate (in terms of costs and computational resources etc.). Furthermore, benchmarks consisting of artificial problems, even when enhanced with simulated uncertainties, have not been shown to possess characteristics of real-world problems. As a result, these benchmarks cannot be used exclusively to investigate algorithm behaviour in the presence of uncertainties.

We identify the following requirements for a benchmark that would be suitable to investigate uncertainties in the context of real-world problem optimisation:

- I: Problem characteristics** Problems should not be artificial in nature. The benchmark should contain a diverse set of fitness functions which are expected to make sense within their real-world context. Fitness functions should be of considerable complexity and involve different types of uncertainties.
- II: Practicality** The execution of the benchmark should still be possible within a reasonable time frame. Therefore, it should be easy to parallelise the benchmark and the evaluation of a single solution should result in practical execution speeds on standard machines.
- III: Analysis of Uncertainties** The benchmark should allow an analysis of all or some of the uncertainties that occur in optimisation problems. It should also include features that allow an analysis of non-symmetric biases.
- IV: Statistical significance** As evolutionary algorithms are stochastic, the statistics obtained via the benchmark should be statistically justified and thus interpretable.
- V: Investigation of scaling behaviour** Functions should be scalable in search space dimension, so that scaling behaviour can be analysed.

The game-benchmark for evolutionary algorithms (GBEA)¹ fulfils all of the above requirements

¹<http://norvig.eecs.qmul.ac.uk/gbea/gamesbench.html>

(see below) and we thus propose it as a means to investigate uncertainty. The GBEA currently features two game-based function suites both taken from research on procedural content generation for well-known games (Top Trumps and Super Mario Bros.). The actual functions in each suites are described in more detail in corresponding publications [2, 3]. The function suites are implemented so that they are compatible with a well-known benchmarking framework².

I: Problem characteristics The GBEA contains more than 80 single-objective problems, and even more multi-objective ones, ranging from non-simulation based functions to ones that depend on playthroughs with different AI agents. Preliminary experiments show that the problems in the GBEA are suitably complex and contain a variety of uncertainties. Similar problems are faced in the game industry regularly.

II: Practicality Game-based problems obviously do not pose safety concerns, but they are also mostly reasonably fast to compute. GBEA functions range between 1-300 seconds, and are usually faster than comparable real-world benchmarks [1]. The benchmark includes a batch mode.

III: Analysis of Uncertainties Games engines are often stochastic, and so are may state-of-the-art game AIs. This often creates vastly differing behaviour in playthroughs for different generated levels, thus resulting in very complex problems and non-symmetric uncertainties. Additional uncertainties typical for search-based procedural content generation are due to modelling errors, as often AI players or other measures are used to approximate human behaviour or perception.

IV: Statistical significance Statistical significance is ensured by the implementation of multiple instances of the same function. In the GBEA, instances are obtained by varying the initiation settings for training processes of the different content generators.

V: Investigation of scaling behaviour In cases where generated content is directly encoded in the solution (see Top Trumps suite), it is possible to scale the search space of the problem by just specifying different targets for the generated content. Otherwise, the representation itself often already contains a corresponding parameter (see Mario suite).

A set of first baseline results has already been obtained using the GBEA. We tested several state-of-the-art evolutionary algorithms (CMA-ES, SMS-EMOA, MO-CMA-ES), as well as surrogate-assisted versions of these optimisers. We were able to uncover multiple weaknesses in the aforementioned algorithms. Results for algorithms designed for noisy and uncertain optimisation (SAPEO) should be produced in the future. Besides that, the obtained data on modelling errors suggests that for some of the analysed algorithms, model validation is absolutely crucial.

We expect results and insights received for uncertainty handling and analysis from the GBEA to be easily transferable to other domains, such as aeronautical engineering.

REFERENCES

- [1] 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)*. Springer Nature, Switzerland, 2018, pp. 296–308.
- [2] V. Volz, G. Rudolph, and B. Naujoks. "Demonstrating the Feasibility of Automatic Game Balancing". In: *Genetic and Evolutionary Computation Conference (GECCO)*. ACM Press, New York, 2016, pp. 269–276.
- [3] V. Volz et al. "Evolving Mario Levels in the Latent Space of a Deep Convolutional Generative Adversarial Network". In: *Genetic and Evolutionary Computation Conference (GECCO)*. ACM Press, New York, 2018, pp. 221–228.

²<http://numbbo.github.io/coco-doc/>