25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

Evaluating multiobjective evolutionary algorithms: A real-world benchmarking framework

J. Zatarain Salazar [a], **D. Hadka** [b], **P.M. Reed** [c], **H. Seada** [d] **and K. Deb** [e]

[a] *Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands*
[b] *Microsoft, Lincoln, Nebraska, United States*
[c] *Civil and Environmental Engineering, Cornell University, Ithaca, New York, United States*
[d] *Ford Motor Company, Dearborn, Michigan, United States*
[e] *Michigan State University: East Lansing, Michigan, United States*
Email: J.ZatarainSalazar@tudelft.nl

Abstract: Multiobjective evolutionary algorithms (MOEAs) have shown significant progress in addressing well-defined test problems, but their effectiveness in real-world applications remains uncertain. To bridge this gap, we provide a comprehensive benchmarking framework designed to rigorously evaluate state-of-the-art MOEAs in real-world scenarios. Our framework comprises a suite of statistical evaluation metrics, for a collection of diverse real-world benchmark applications representing various mathematical problem difficulties.

In this study, we carefully selected four benchmark applications with 3 to 10 objectives, capturing challenging characteristics such as stochastic objectives, severe constraints, strong nonlinearity, and complex Pareto frontiers. We evaluated the performance of five popular MOEAs, including NSGA-II, NSGA-III, RVEA, MOEA/D, and the Borg MOEA, using our benchmarking framework. Multiobjective evolutionary algorithms (MOEAs) have shown significant progress in addressing well-defined test problems, but their effectiveness in real-world applications remains uncertain. To bridge this gap, we provide a comprehensive benchmarking framework designed to rigorously evaluate state-of-the-art MOEAs in real-world scenarios. Our framework comprises a suite of statistical evaluation metrics, for a collection of diverse real-world benchmark applications representing various mathematical problem difficulties.

The results revealed distinct differences in the performance of the evaluated MOEAs across the real-world applications. Surprisingly, MOEAs that excelled on standard test problems struggled when confronted with the complexities inherent in real-world applications. These findings underscore the need to enhance the adaptability and ease-of-use of MOEAs, considering the often ill-defined nature of real-world problem solving. Furthermore, our study provides insights into successful algorithmic design choices for MOEAs. Optimal selection strategies and archive mechanisms are crucial to prevent deterioration, maintain diversity, and provide adequate selection pressure throughout the optimization process. Additionally, the choice of stable and flexible operators plays a vital role in reliably driving the search towards the Pareto front. Recent advancements in hyper-heuristics and multi-operator MOEAs offer promising automated approaches for tackling these challenges.

We found that epsilon non-dominated sorting effectively maintains diversity and selection pressure for problems with up to ten objectives when the entire Pareto front is desired. Moreover, auto-adaptive search operators demonstrate their efficacy in adapting to the search landscape of diverse real-world applications. However, the performance of reference point/vector methods deteriorated at higher dimensions, indicating the need for further investigation. Our study highlights the inadequacy of existing test benchmarks in differentiating MOEAs based on real-world performance. While considerable efforts have focused on optimizing algorithms for test problems, the subpar performance of MOEAs in real-world settings persists. The benchmarking framework and results presented here aim to foster collaborative efforts, encouraging the development of a diverse suite of real-world benchmarking problems. The flexibility of our framework also allows for exploration of new or hybrid algorithm architectures in benchmarking studies.

Keywords: *Many-objective evolutionary algorithm, optimization, benchmarking, diagnostics*