

ZĂVOIANU, A.-C., LACROIX, B. and MCCALL, J. 2020. Comparative run-time performance of evolutionary algorithms on multi-objective interpolated continuous optimisation problems. In *Bäck, T., Preuss, M., Deutz, A., Wang, H., Doerr, C., Emmerich, M. and Trautmann, H. (eds.) Parallel problem solving from nature: PPSN XVI: proceedings of the 16th Parallel problem solving from nature international conference (PPSN 2020), 5-9 September 2020, Leiden, The Netherlands*. Lecture notes in computer science, 12269. Cham; Springer, part 1, pages 287-300. Available from: https://doi.org/10.1007/978-3-030-58112-1_20

Comparative run-time performance of evolutionary algorithms on multi-objective interpolated continuous optimisation problems.

ZĂVOIANU, A.-C., LACROIX, B. and MCCALL, J.

2020

The final authenticated version is available online at https://doi.org/10.1007/978-3-030-58112-1_20. This pre-copyedited version is made available under the Springer terms of reuse for AAMs: <https://www.springer.com/gp/open-access/publication-policies/aam-terms-of-use>

OpenAIR
@RGU

This document was downloaded from
<https://openair.rgu.ac.uk>

SEE TERMS OF USE IN BOX ABOVE

DISTRIBUTED UNDER LICENCE

Comparative Run-Time Performance of Evolutionary Algorithms on Multi-Objective Interpolated Continuous Optimisation Problems

Alexandru-Ciprian Zăvoianu, Benjamin Lacroix, and John McCall

School of Computing Science and Digital Media, Robert Gordon University,
Aberdeen, Scotland, UK

{c.zavoianu, b.m.e.lacroix, j.mccall}@rgu.ac.uk

Abstract. We propose a new class of multi-objective benchmark problems on which we analyse the performance of four well established multi-objective evolutionary algorithms (MOEAs) – each implementing a different search paradigm – by comparing run-time convergence behaviour over a set of 1200 problem instances. The new benchmarks are created by fusing previously proposed single-objective interpolated continuous optimisation problems (ICOPs) via a common set of Pareto non-dominated seeds. They thus inherit the ICOP property of having tunable fitness landscape features. The benchmarks are of intrinsic interest as they derive from interpolation methods and so can approximate general problem instances. This property is revealed to be of particular importance as our extensive set of numerical experiments indicates that choices pertaining to (i) the weighting of the inverse distance interpolation function and (ii) the problem dimension can be used to construct problems that are challenging to all tested multi-objective search paradigms. This in turn means that the new multi-objective ICOPs problems (MO-ICOPs) can be used to construct well-balanced benchmark sets that discriminate well between the run-time convergence behaviour of different solvers.

Keywords: multi-objective continuous optimisation · evolutionary algorithms · performance analysis · large-scale benchmarking.

1 Introduction and Motivation

A multi-objective optimisation problem (MOOP) can be defined as:

$$\text{minimize } F(x) = (f_1(x), \dots, f_m(x))^T, \quad (1)$$

where the search space is multi-dimensional (i.e., $x \in V^d \subset \mathbb{R}^d$) and the $m \in \{2, 3\}$ real-valued objectives of $F(x)$ need to be minimized simultaneously. The conflicting nature of the m objectives means that the general solution of a MOOP is given by a Pareto optimal set (PS) that aggregates all solution candidates $x^* \in V^d$ with the property that they are not fully dominated – i.e., $\nexists y \in V^d : f_i(y) \leq f_i(x^*), \forall i \in \{1, \dots, m\}$ and $F(y) \neq F(x^*)$. The Pareto front (PF) is the objective space projection of the PS.

Because of their ability to discover high-quality PS approximations called Pareto non-dominated sets (PNs) in single runs, multi-objective evolutionary algorithms (MOEAs) have emerged as some of the most successful MOOP solvers [3]. As an increasing number of MOEA practitioners (e.g., mechatronic engineers [19], industrial designers [16], quality assurance analysts [28]) are tackling ever more challenging real-world problems, difficulties stemming from relying on experimentation/simulation-driven $F(x)$ values are being brought to the forefront. A costly evaluation of solution candidate quality (i.e., fitness) greatly reduces the number of fitness evaluations (*nfe*) that can be computed during an optimisation and runs might be stopped prematurely. Furthermore, running multiple optimisation runs often becomes infeasible and only a single solver (with literature recommended parameter settings) is applied despite the well-known implications of the No Free Lunch Theorems for Optimisation (NFL) [22, 4] regarding the benefits of algorithm and/or parameter selection.

To alleviate the aforementioned difficulties of real-world MOEA application, researchers have explored several avenues. Among them, promising results have been delivered by both (i) the integration of surrogate modeling techniques [13, 17] that reduce solver dependency on costly fitness evaluations and (ii) the development of multi-method solvers that can deliver a robust performance over large (benchmark) problem sets with a fixed parameterisation [21, 23]. While surrogate techniques are often demonstrated on MOOP formulations bound to “closed” simulation or experimentation environments, most MOOP benchmark problems share biases like analytically engineered challenges and strong inter-problem correlations [2]. The present work aims to aid both research streams by introducing a new class of MOOPs with tunable fitness landscapes that:

- (i) are intrinsically interesting for benchmark construction as they propose challenges to several state-of-the-art MOEAs, when instantiated randomly;
- (ii) can be used to effortlessly generate easy-to-share, lightweight interpolation-based surrogate formulations, when instantiating with real-world data.

2 Multi-Objective Interpolated Continuous Optimisation Problems

2.1 Interpolated Continuous Optimisation Problems

First proposed in a single-objective context [12], ICOPs are defined by the following elements:

1. **A search space** Ω : a set, whose elements we refer to as *candidate solutions*, that defines the optimisation problem domain. For continuous problems, this will be a (subset of) real space of chosen dimension. In this paper, we have chosen our search spaces to be the d -dimensional cubes: $\Omega = [-5, 5]^d$.
2. **A distance function**, $e(x, y) : \Omega \times \Omega \rightarrow \mathbb{R}$ defining the distance between two solutions x and y . The pair (Ω, e) with these definitions is a *metric space*. A natural choice of distance function for continuous search spaces is the Euclidean distance.

3. **A set of seeds** $S \subset \Omega$: a (generally finite) set of distinct candidate solutions with an assigned fitness. Elements of S and their assigned fitness values will define the entire optimisation problem via interpolation.
4. **An interpolation function** $f_S : \Omega \rightarrow \mathbb{R}$: in this paper, we apply the inverse distance weighting method, originally defined by Shepard [18] for use in spatial analysis. Assuming the seed set S contains N seeds, labelled $S = \{s_1, \dots, s_N\}$, and with the assigned fitnesses $U = \{u_1, \dots, u_N\}$, we define for any solution candidate $x \in \Omega$:

$$f_{S,U}(x) = \begin{cases} \frac{\sum_{j=1}^N \frac{u_j}{e(x,s_j)^k}}{\sum_{j=1}^N \frac{1}{e(x,s_j)^k}} & \text{if } e(x,s_j) \neq 0 \text{ for all } j \\ u_j, & \text{if } e(x,s_j) = 0 \text{ for some } j \end{cases} \quad (2)$$

where k is a positive real number called the *power parameter*. Higher values of k increase the relative influence of nearby seeds on the interpolated value.

5. **An optimisation objective**: e.g., the minimisation of f_S .

2.2 Multi-Objective ICOPs

Multi-Objective ICOPs (MO-ICOPs) can be obtained by associating each single-objective $f_i(x)$, $i \in \{1, \dots, m\}$ from Equation 1 with a distinct ICOP. Given that in a real-world optimisation scenario, we would expect each solution candidate x to be evaluated across all m objectives to be optimised, it would make sense that the individual ICOPs share the same seeds (i.e., solution candidate samples), but differ on the fitness values associated to each seed¹. Hence, a MO-ICOP is defined by the tuple (S, U_1, \dots, U_m) and the value of k and Equation 1 can be rewritten:

$$\text{minimize } F_{(S,U_1,\dots,U_m)}(x) = (f_{S,U_1}(x), \dots, f_{S,U_m}(x))^T, \quad (3)$$

One major caveat is that the PF of a MO-ICOP cannot be computed analytically and must be estimated in an iterative fashion by aggregating the best (Pareto non-dominated) solutions found during multiple optimisation runs. Sampling-based PF discovery was also required for real-world visualisable (i.e., $x \in V^2$) distance-based MOOPs [9], but for this problem type, restrictions can be imposed to generate artificial instances with prescribed PFs [8].

2.3 The k Parameter

The power parameter k defines the influence of each seed on the interpolation of the rest of the search space as illustrated in Figure 1. k increase the importance of distance weighting in the interpolation function. Lower values of k create flat fitness landscape around the average of the seed's value with sharp *peaks* and *pits* around the seeds. Higher values of k create large basins of attraction around each seed. This in turn influences the PF of resulting MO-ICOPs, as illustrated in the third row plots of Figure 1.

¹ A similar strategy of composing MOOPs from single-objective problems was recently used for creating the **bbob-biobj** test suite [1].

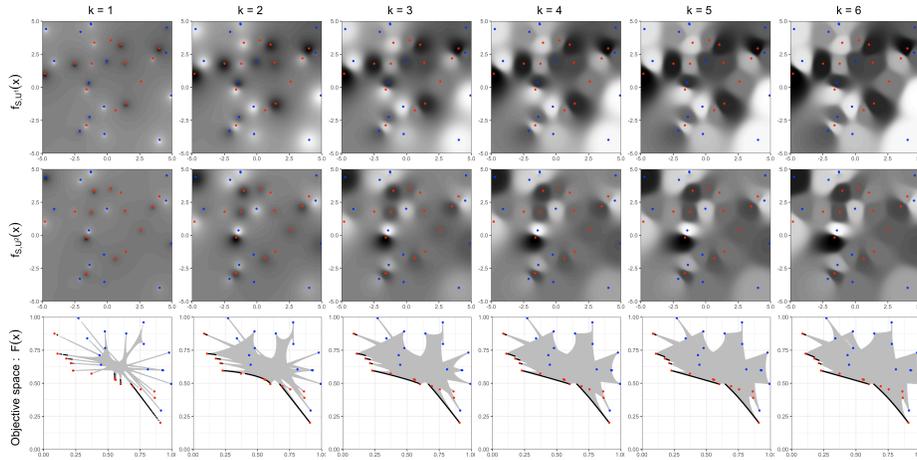


Fig. 1. MO-ICOPs with two objectives for $d = 2$ based on the same tuple (S, U_1, U_2) but different k values. The first two rows show the impact of k on the fitness landscape of each sub-contained ICOP. Non-dominated seeds are marked with red and dominated ones with blue; axes correspond to decision variables x_1 and x_2 ; lighter shades \Rightarrow lower values. The third row shows how this translates to different shapes of the MO-ICOP objective space (gray) and PF (black) when grid sampling 10^6 candidate solutions.

3 Experimental Setup

3.1 Test Problems

While the properties of MO-ICOPs seeded with real-world data do warrant rigorous examination, in this preliminary study we focus on analysing the run-time convergence behaviour of established MOEA search paradigms with respect to the inherent characteristics of MO-ICOPs by considering a large set of 1200 two-objective random problem instances.

Randomly constructing a MO-ICOP is centred on embedding the seeds in objective space. We define S_{nd} as the set of Pareto non-dominated seeds and S_d as the set of dominated seeds such that $S = S_{nd} \cup S_d$. We start to randomly sample fitness values for each seed in S_{nd} ensuring that no seed in S_{nd} fully dominates another seed in S_{nd} . We then randomly sample fitness values for each seed in S_d , ensuring that every seed in S_d is fully dominated by at least one seed in S_{nd} . The idea behind this construction is to create randomly generated problems that may provide some different PF shapes. Finally, the positions of the seeds in the search space are obtained by random uniform sampling.

For our benchmark set we generated 50 distinct objective-space seed set embeddings and corresponding search space positions for 4 different dimensions $d \in \{5, 10, 20, 30\}$. We obtained the final MO-ICOP instances by associating each of the 200 resulting combinations with 6 values of $k \in \{1, 2, 3, 4, 5, 6\}$. For each problem, the number of non dominated seeds ($|S_{nd}|$) was randomly chosen

between 5 and 20 and the number of dominated seeds ($|S_d|$) was randomly chosen between 1 and 80. For each problem, the PF was estimated using 20 million fitness evaluations spread across 400 independent optimisation runs. This data can be accessed at: https://github.com/czavoiianu/PPSN_2020.

3.2 Solvers and Parameterisation

The four MOEAs used in our tests were chosen because they exemplify different well-proven strategies for solving multi-objective optimisation problems.

NSGA-II [6] is one of the best-known and most widely applied multi-objective solvers. It has popularised alongside SPEA2 [25] a highly elitist multi-objective evolutionary paradigm in which the population of iteration $t + 1$ is obtained by applying a two-tier selection for survival (i.e., filtering) operator on the union between the population of iteration t and all the offspring generated at iteration t . The filtering relies on a primary Pareto non-dominated sorting criterion and a secondary crowding criterion (for tie-breaking situations). The success of NSGA-II has also popularised two genetic operators for real-valued MOOPs: simulated binary crossover (SBX) and polynomial mutation (PM) [5].

The GDE3 [11] solver maintains the two-tier selection for survival operator introduced by NSGA-II, but aims to also exploit the very good performance of the differential evolution (DE) paradigm [20] on continuous optimisation problems by replacing the SBX and PM operators with a *DE/rand/1/bin* strategy.

MOEA/D-DE with Dynamic Resource Allocation [24] is a state-of-the-art solver that achieves highly competitive solutions for a wide-range of MOOPs. MOEA/D-DE refines the multi-objective search paradigm proposed in MOGLS [10] as it decomposes the original MOOP into several single-objective sub-problems that are the result of weighting-based aggregations of the original MOOP objectives. During the run, individuals are evolved via *DE/rand/1/bin* to become the solution to one or more of the sub-problems. Provided a proper choice of weight vectors, the solver population should provide a very good *PS* approximation.

The DECMO2++ [23] solver was designed for rapid convergence across a wide range of MOOPs as it integrates and actively pivots between three different search paradigms implemented via coevolved sub-populations. Specifically, while one sub-population implements Pareto-based elitism via the SPEA2 evolutionary model and the associated SBX and PM operators, the other actively co-evolved sub-population uses the DE-centred GDE3 search strategy. Decomposition is implemented via a largely passive archive based on a uniformly weighted Tschebyscheff distance measure that aims to maintain the best achieved approximation of the *PS* at each stage of the search process.

We used the standard / literature recommended parameterisation for all four MOEAs and allowed a total computational budget of $nfe = 50,000$ for each optimisation run. In the case of NSGA-II, GDE3 and DECMO2++ we used a population/archive size of 200. In the case of MOEA/D-DE DRA we used an archive size of 300 – the recommended setting for MOOPs with two objectives.

3.3 Performance Evaluation

The PN quality measure we track during the run-time is Ind_H - a normalised version of the hypervolume [26] that evaluates a PN by relating the size of the objective space it dominates to that of the objective space dominated by the PF.

Hypervolume-ranked performance curves (HRPCs) have been proposed in [27] as a means to quickly estimate the comparative differences in the run-time convergence behaviour of several MOEAs across large benchmark sets. For each MOOP, the strategy is to rank the MOEAs that aim to solve it after every 1000 newly generated individuals (pre-defined comparison stages) using Ind_H values averaged over 100 independent runs. Under a *basic ranking schema*, the solver with the lowest average Ind_H in a set of n_s MOEAs will receive the rank n_s and the best performer will receive the rank 1. A bonus rank of 0 can be given to solvers that are estimated to have fully converged on the problem (i.e., $Ind_H > 0.99$). By averaging for each MOEA at each comparison stage the ranks achieved on individual MOOPs, one can rapidly obtain an overview of the comparative convergence behaviour across the entire benchmark set.

HRPCs are constructed via two-by-two comparisons between the tested solvers in increasing order of Ind_H -indicated performance. As we also wish to illustrate the magnitude of the differences in run-time convergence behaviours, we complement the basic ranking schema with:

- a *pessimistic ranking schema* under which different ranks are awarded in the stage-wise two-by-two comparison only if the difference between the average Ind_H values of the two MOEAs is higher than a $th = 0.01$, $th = 0.05$ or $th = 0.10$ predefined threshold ²;
- a *statistical ranking schema* under which different ranks are awarded only if the difference between the stage-wise average Ind_H values is statistically significant when using a one-sided Mann-Whitney-Wilcoxon test [14] with a considered significance level of 0.025 after a Bonferroni correction [7].

It is noteworthy that in [27] the ranking was based on the average Ind_H assessment of the run-time MOEA populations at each comparison stage. In the present work, motivated by the interactive way in which engineers employ MOEA-based searches in practice [19], the Ind_H averages over independent runs are computed by considering the set of all the Pareto non-dominated individuals that have been discovered by the MOEA till each comparison stage.

4 Results and Analysis

Before going into run-time performance analysis, it worth noting the performance of the four algorithms at $nfe = 50,000$. On the 1200 problems tested, MOEA/D-

² If the difference between the Ind_H -measured qualities of two PNs is larger than $th = 0.05$, the interpretation is that the objective space dominated by the best PN is larger than its counterpart by a size that is equivalent to at least 5% of the size of the objective space dominated by the solution of the MOOP (i.e., the PF).

DE obtained the highest Ind_H 545 times, followed by NSGA-II with 509 end-of-the-run wins. GDE3 and DECMO++ obtained respectively 125 and 21 wins. In the top left plot of Figure 2 we present the run-time Ind_H -averaged performance of the four tested MOEAs across the entire benchmark of 1200 MO-ICOPs. For $nfe > 10,000$, the plot indicates that MOEA/D-DE DRA achieves the best average performance, ahead of NSGA-II and DECMO2++. GDE3 constantly achieved the lowest run-time Ind_H average values.

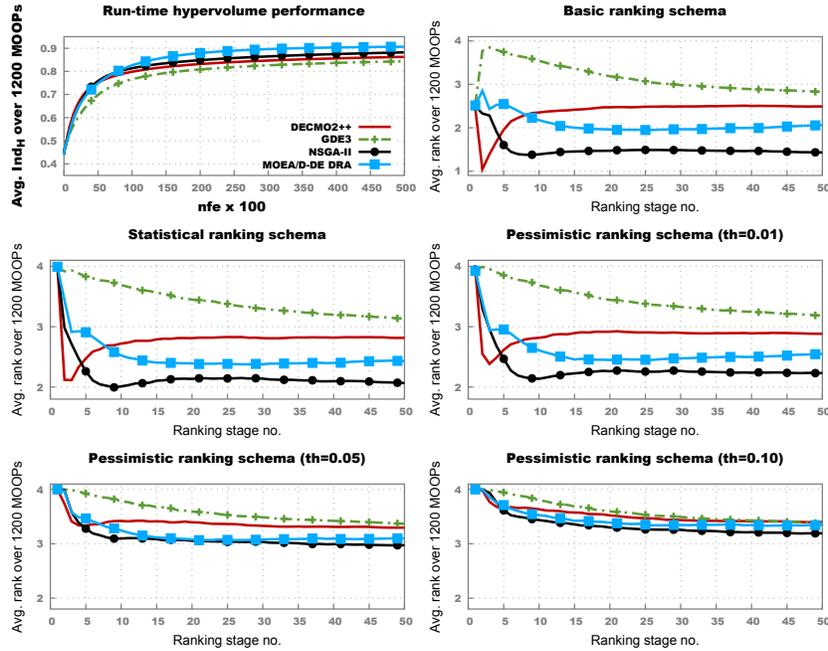


Fig. 2. Results across the entire benchmark of 1200 MO-ICOPs: $d \in \{5, 10, 20, 30\}$ and $k \in \{1, 2, 3, 4, 5, 6\}$. A bonus rank of 0 is awarded for full convergence – i.e., when $Ind_H > 0.99$.

In the early phases of the optimisation runs (i.e., $nfe < 7,000$), both NSGA-II and DECMO2++ achieve slightly better average Ind_H values than MOEA/D. This is confirmed by the HRPC plots from Figure 2 as they indicate a slight advantage in early convergence for NSGA-II and DECMO2++ even when considering a pessimistic ranking with a threshold $th = 0.05$. Surprisingly, the general picture of the five rank-based comparisons indicates that NSGA-II edges MOEA/D-DE DRA as the best performer throughout the run-time. Since this is contrasting with the results of the Ind_H -averaged performance plot, it is worthy

to focus the analysis on NSGA-II and MOEA/D-DE and thus remove the impact of convergence behaviour cliques on the relative rankings.

Furthermore, by not awarding a bonus rank of 0 for full convergence (i.e., when $Ind_H > 0.99$), the restricted NSGA-II vs. MOEA/D-DE analysis highlights more clearly at each ranking stage the difference between the number of problems on which one solver performs better than the other. Thus, since only ranks of 2 and 1 can be awarded, achieving an average value of 2 at a certain ranking stage is a clear indication that a solver is not better than its counterpart (given the considered ranking criterion) across any MOOPs in the chosen benchmark set. Conversely, HRPC values very close to 1 indicate a clear better performance. More formally, in a one-on-one comparison with no bonuses, a rank value of $1.x$ associated with a solver, indicates that the solver outperforms its counterpart on $(1 - 0.x) \times 100\%$ of the problems considered in the benchmark.

In light of strong empirical evidence (please see Figure 1) that the power parameter used in the interpolation function k from Equation 2 has a significant impact on the geometry of the resulting MO-ICOP problem, it is natural to analyse if and how this translates into divergent MOEA run-time behaviours.

Therefore, in Figure 3, we present the comparative convergence behaviour of NSGA-II and MOEA/D-DE when only considering a benchmark subset containing the 600 MO-ICOPs with the lowest weights of the inverse distance interpolation function: i.e., $k \in \{1, 2, 3\}$. The average Ind_H values from the top left plot indicate that MOEA/D-DE generally outperforms NSGA-II in all stages of the optimisation runs. Furthermore, this subset of benchmark MO-ICOPs is more challenging as, for both solvers, the achieved average Ind_H values are lower than those reported in Figure 2 over all 1200 problems.

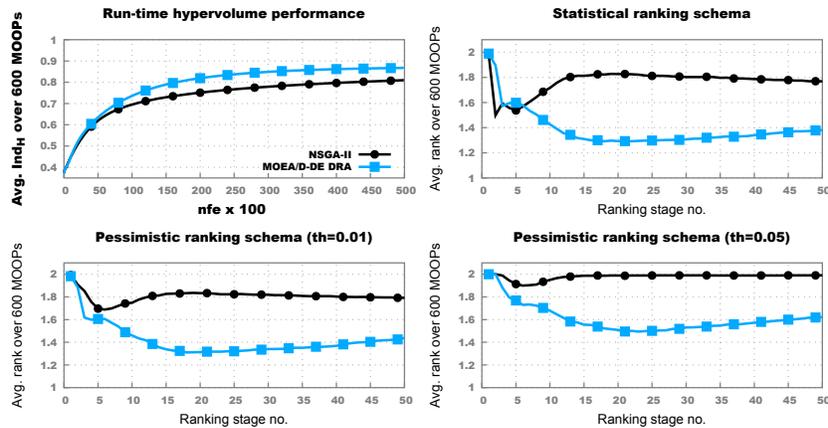


Fig. 3. Results across the benchmark subset with 600 MO-ICOPs with $d \in \{5, 10, 20, 30\}$ and $k \in \{1, 2, 3\}$. No bonus rank is awarded for full convergence.

The different convergence behaviours of the two solvers for low k values is confirmed by all three corresponding HRPC plots. These also highlight that, for $nfe > 15,000$, the Ind_H -measured performance of MOEA/D-DE solutions is better across $\approx 40\%$ of the problems in this benchmark subset when considering a Ind_H threshold $th = 0.05$.

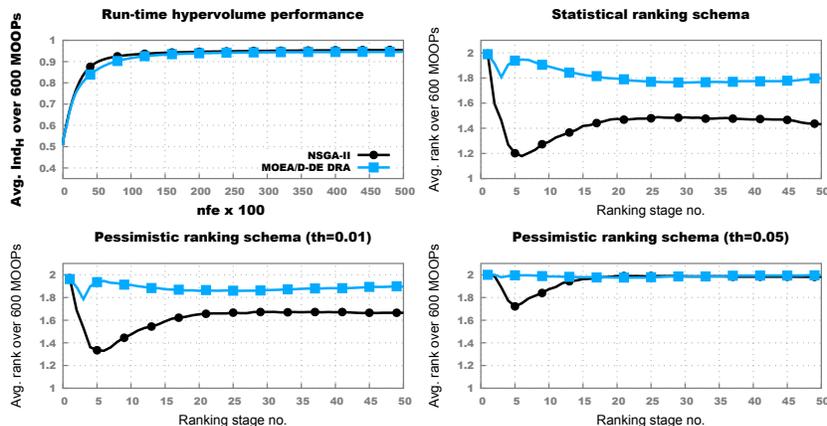


Fig. 4. Results across the benchmark subset with 600 MO-ICOPs with $dim \in \{5, 10, 20, 30\}$ and $k \in \{4, 5, 6\}$. No bonus rank of is awarded for full convergence.

The plots in Figure 4 indicate that for a higher weight of the inverse distance interpolation function (i.e., $k \in \{4, 5, 6\}$), the Ind_H -measured performance of NSGA-II solutions is:

- consistently better across $\approx 50\%$ of the MO-ICOPs in this benchmark subset when considering the one-sided Mann-Whitney-Wilcoxon statistical significance test;
- consistently better across $\approx 30\%$ of the MO-ICOPs, when considering a Ind_H threshold of 0.01.
- only better by large margins (i.e., $th = 0.05$) across $\approx 10\text{-}25\%$ of the MO-ICOPs in the early stages of the optimisation runs: $5000 < nfe < 10,000$.

It is also noteworthy that on the MO-ICOPs associated with higher values of k , both multi-objective solvers perform well and are able to reach average benchmark-wide Ind_H values higher than 0.9 after 10,000 fitness evaluations.

The plots in Figure 5 illustrate the run-time convergence behaviour of NSGA-II and MOEA/D-DE on problems with a lower dimension: $d \in \{5, 10\}$. The HRPCs show that while solver performance is similar early on, for $nfe > 5,000$, when comparing based on statistical significance, NSGA-II performs better across $\approx 50\text{-}60\%$ of the 600 MO-ICOPs and MOEA/D-DE perform better across $\approx 20\text{-}30\%$ of the problems. Smaller differences that progressively favour NSGA-II as

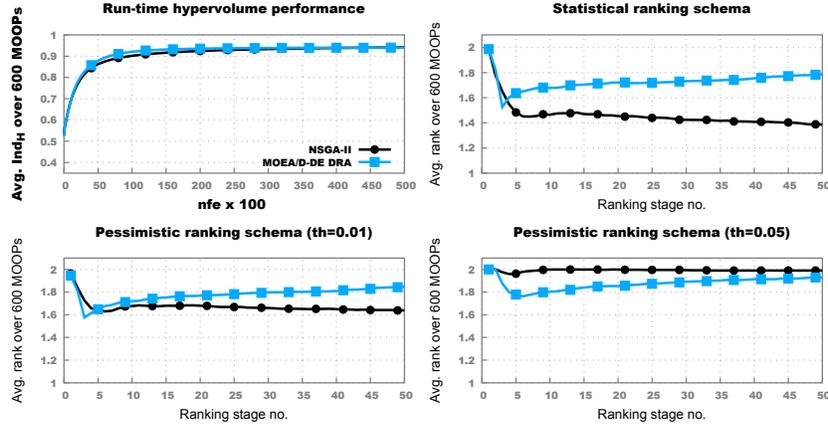


Fig. 5. Results across the benchmark subset with 600 MO-ICOPs with $d \in \{5, 10\}$ and $k \in \{1, 2, 3, 4, 5, 6\}$. No bonus rank is awarded for full convergence.

the nfe increases are also observed when imposing a small Ind_H ranking threshold of $th = 0.01$. However, when looking at the larger thresholds of 0.05, NSGA-II doesn't outperform on any problem in the benchmark while MOEA/D-DE still perform better across ≈ 10 -20% of the problems throughout the entire run-time. The fact that NSGA-II performs better on more problems but with a lower margins (i.e., $th < 0.05$) while MOEA/D performs better on fewer problems but with higher margins (i.e., $th \geq 0.05$) helps to explain the average benchmark-wide Ind_H plot (top-left) associated with this benchmark subset. As a side note, when considering Figures 4 and 5, the HRPCs provide valuable insight that helps to differentiate convergence behaviours captured by largely similar benchmark-wide Ind_H averaging plots.

The plots in Figure 6 show that on the benchmark subset of 600 MO-ICOPs with a higher dimension – i.e., $d \in \{20, 30\}$ – MOEA/D-DE consistently outperforms NSGA-II for $nfe > 15,000$ across $\approx 60\%$ of the problems (based on a statistical significance and pessimistic $th = 0.01$ criteria) and ≈ 30 -40% of the problems (when considering the stricter $th = 0.05$ criterion). It is noteworthy that in the early part of the runs ($nfe < 15,000$), NSGA-II performs notably better than MOEA/D-DE across all the considered comparison criteria.

Finally, in order to better understand the interplay between the inverse distance weighting parameter and the MO-ICOP problem dimension on one side and the comparative solver performance on the other, we can compute a (k, d) preference matrix at a fixed point of interest during the run-time. For example, the top-left plot of Figure 2 indicates that all solvers are past their knee-point in convergence at $nfe = 20,000$ (i.e., ranking stage no. 20). When considering the statistical significance ranking criterion, we can compute the preference for MOEA/D-DE over NSGA-II for the 24 (k, d) combinations by subtracting the

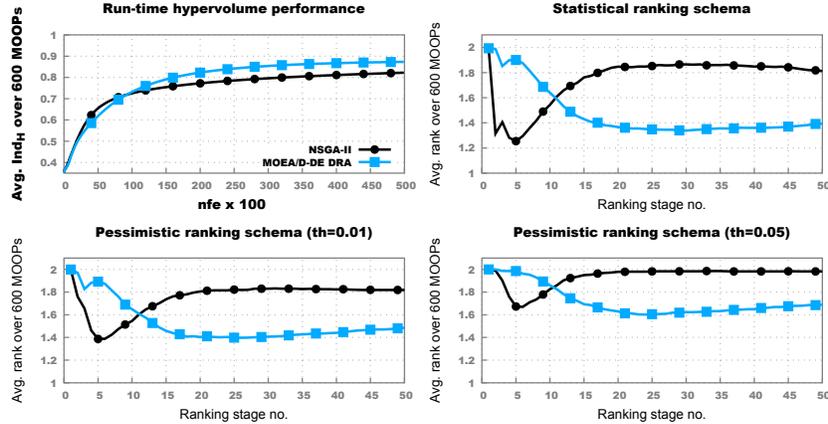


Fig. 6. Results across the benchmark subset with 600 MO-ICOPs with $d \in \{20, 30\}$ and $k \in \{1, 2, 3, 4, 5, 6\}$. No bonus rank is awarded for full convergence.

percentage of problems on which NSGA-II outperforms (at ranking stage no. 20) from the percentage of problems on which MOEA/D-DE performs better. The resulting (k, d) preference matrix shown in Figure 7 indicates that while MOEA/D-DE obtains better results than NSGA-II on problems with low k parameters and larger dimensions, NSGA-II performs better on problems with large k values and lower dimensions. Low k values (in particular $k = 1$) result in very discontinuous point-wise PFs (as illustrated in Figures 1 and 7) that highly favour the directional decomposition search strategy of MOEA/D-DE. Higher k values generate more continuous PFs that are generally easier to converge on for both solvers, but on which the decomposition strategy is at a slight disadvantage (please see Figure 4) as its more rigid exploration mechanism likely generates PNs with an inferior spread.

5 Conclusions and Future Work

In this paper we (i) describe a new class of multi-objective interpolated continuous optimisation problems (MO-ICOPs) constructed using a weighted inverse distance function and we (ii) proceed to analyse the comparative run-time performance of four established MOEAs on a benchmark of 1200 random MO-ICOP instances using multiple criteria.

The optimisation results indicate that MO-ICOPs propose challenges to all tested multi-objective optimisation paradigms. GDE3 consistently achieves the lowest benchmark-wide average hypervolume attainment levels despite obtaining the best approximation of the PF on 10% of the problems tested. DECMO2++ only demonstrates its characteristic fast converging behaviour during the very start of the run (i.e., $nfe < 5,000$) and is outperformed by both MOEA/D-DE and

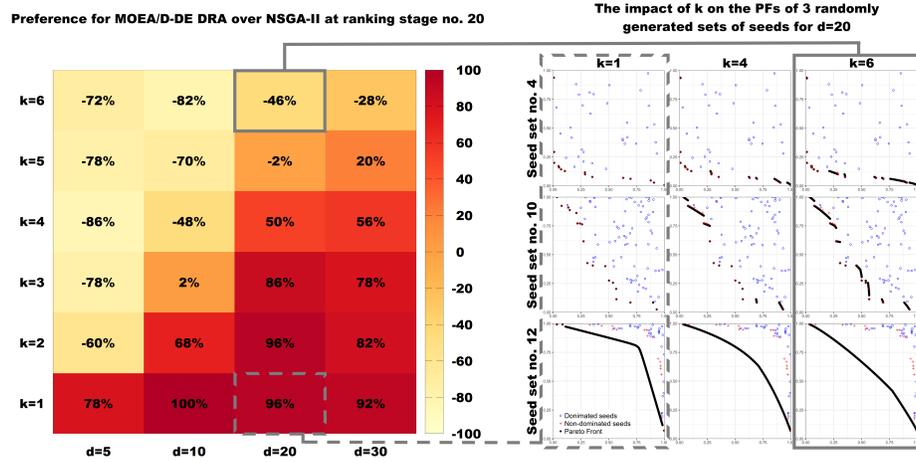


Fig. 7. Preference of MOEA/D-DE over NSGA-II when considering differences in average performance over 100 independent runs confirmed by statistical significance testing (left). Example of the impact of k on the PF shape of 9 MO-ICOPs with $d = 20$ (right).

NSGA-II. The comparative performance of NSGA-II and MOEA/D-DE DRA is strongly influenced by the weighting of the inverse distance function and the dimension of the problem.

Since the observed convergence behaviours of GDE3 and DECMO2++ somewhat contrast with those previously reported on widely used benchmarks [23], moving forward we plan to (i) investigate more closely the causes that impact the general performance of all four algorithms on MO-ICOPs (ii) and complement existing benchmark sets with both random and real-world based MO-ICOPs in order to obtain a well-balanced test rig that can effectively support the discovery of robust MOEAs and/or robust MOEA parameterisations.

Finally, we believe that more comprehensive test sets can provide a better insight on algorithm performance by characterising problems through the development of landscape and objective space features. As it is already the case in single-objective optimisation [15], such advancements could lead to the application of landscape and objective space features for algorithm selection or algorithm performance prediction.

Acknowledgments

This work has been supported by the COMET-K2 “Center for Symbiotic Mechatronics” of the Linz Center of Mechatronics (LCM) funded by the Austrian federal government and the federal state of Upper Austria.

References

1. Brockhoff, D., Tutar, T., Auger, A., Hansen, N.: Using well-understood single-objective functions in multiobjective black-box optimization test suites. *ArXiv e-prints*, 1604.00359v3 (2019)
2. Christie, L.A., Brownlee, A.E., Woodward, J.R.: Investigating benchmark correlations when comparing algorithms with parameter tuning. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. pp. 209–210 (2018)
3. Coello Coello, C.A., Lamont, G.B.: *Applications of multi-objective evolutionary algorithms*. World Scientific (2004)
4. Corne, D.W., Knowles, J.D.: No free lunch and free leftovers theorems for multiobjective optimisation problems. In: *International Conference on Evolutionary Multi-Criterion Optimization*. pp. 327–341. Springer (2003)
5. Deb, K.: *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons (2001)
6. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* **6**(2), 182–197 (2002)
7. Dunn, O.J.: Multiple comparisons among means. *Journal of the American statistical association* **56**(293), 52–64 (1961)
8. Fieldsend, J.E., Chugh, T., Allmendinger, R., Miettinen, K.: A feature rich distance-based many-objective visualisable test problem generator. In: *Proceedings of the Genetic and Evolutionary Computation Conference*. p. 541–549. GECCO '19, Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3321707.3321727>
9. Ishibuchi, H., Akedo, N., Nojima, Y.: A many-objective test problem for visually examining diversity maintenance behavior in a decision space. In: *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*. p. 649–656. GECCO '11, Association for Computing Machinery, New York, NY, USA (2011). <https://doi.org/10.1145/2001576.2001666>
10. Jaskiewicz, A.: On the performance of multiple-objective genetic local search on the 0/1 knapsack problem - A comparative experiment. *IEEE Transactions on Evolutionary Computation* **6**(4), 402–412 (2002)
11. Kukkonen, S., Lampinen, J.: GDE3: The third evolution step of generalized differential evolution. In: *IEEE Congress on Evolutionary Computation (CEC 2005)*. pp. 443–450. IEEE Press (2005)
12. Lacroix, B., Christie, L.A., McCall, J.A.: Interpolated continuous optimisation problems with tunable landscape features. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. pp. 169–170 (2017)
13. Loshchilov, I., Schoenauer, M., Sebag, M.: A mono surrogate for multiobjective optimization. In: *Proceedings of the 12th annual Conference on Genetic and Evolutionary Computation (GECCO)*. pp. 471–478. ACM (2010)
14. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics* **18**(1), 50–60 (1947)
15. Muñoz, M.A., Sun, Y., Kirley, M., Halgamuge, S.K.: Algorithm selection for black-box continuous optimization problems: A survey on methods and challenges. *Information Sciences* **317**, 224–245 (2015)

16. Oyama, A., Kohira, T., Kemmotsu, H., Tatsukawa, T., Watanabe, T.: Simultaneous structure design optimization of multiple car models using the k computer. In: IEEE Symposium Series on Computational Intelligence (2017)
17. Pilát, M., Neruda, R.: Hypervolume-based local search in multi-objective evolutionary optimization. In: Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation. pp. 637–644. ACM (2014)
18. Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM national conference. pp. 517–524 (1968)
19. Silber, S., Bramerdorfer, G., Dorninger, A., Fohler, A., Gerstmayr, J., Koppelstätter, W., Reischl, D., Weidenholzer, G., Weitzhofer, S.: Coupled optimization in magopt. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering **230**(4), 291–299 (2016). <https://doi.org/10.1177/0959651815593420>
20. Storn, R., Price, K.V.: Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization **11**(4), 341–359 (December 1997)
21. Vrugt, J.A., Robinson, B.A.: Improved evolutionary optimization from genetically adaptive multimethod search. Proceedings of the National Academy of Sciences **104**(3), 708–711 (2007)
22. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation **1**(1), 67–82 (1997)
23. Zăvoianu, A.C., Saminger-Platz, S., Lughofer, E., Amrhein, W.: Two enhancements for improving the convergence speed of a robust multi-objective coevolutionary algorithm. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 793–800. ACM (2018)
24. Zhang, Q., Liu, W., Li, H.: The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances. Tech. rep., School of CS & EE, University of Essex (February 2009)
25. Zitzler, E., Laumanns, M., Thiele, L.: SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization. In: Evolutionary Methods for Design, Optimisation and Control with Application to Industrial Problems (EUROGEN 2001). pp. 95–100. International Center for Numerical Methods in Engineering (CIMNE) (2002)
26. Zitzler, E.: Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications. Ph.D. thesis, Swiss Federal Institute of Technology (1999)
27. Zăvoianu, A.C., Lughofer, E., Bramerdorfer, G., Amrhein, W., Klement, E.P.: DECMO2: a robust hybrid and adaptive multi-objective evolutionary algorithm. Soft Computing **19**(12), 3551–3569 (2014). <https://doi.org/10.1007/s00500-014-1308-7>
28. Zăvoianu, A.C., Lughofer, E., Pollak, R., Meyer-Heye, P., Eitzinger, C., Radauer, T.: Multi-objective knowledge-based strategy for process parameter optimization in micro-fluidic chip production. In: Proceedings of the IEEE SSCI 2017 Conference, pp. 1927–1934. IEEE, Honolulu, USA (2017)