A comparative study of state-of-the-art multi-objective optimization algorithms

Jiawei Li[1], and *Evgenii* Sopov[1,2*]

[1]Siberian Federal University, Institute of Space and Information Technology, Krasnoyarsk, Russia
[2]Reshetnev Siberian State University of Science and Technology, Institute of Informatics and Telecommunications, Krasnoyarsk, Russia

Abstract. With the development of intelligent algorithms, multi-objective optimization problems are increasingly showing a significant role in various fields. In this paper, we used four multi-objective optimization algorithms and tested them on six ZDT standard test problems. Conducted experiments to analyse the optimization effects of the algorithms and determine the strengths and weaknesses of each. These analyses help to identify the most appropriate optimization algorithm for a given problem.

1 Introduction

Multi-objective optimization (MOO) is a crucial aspect of computational problem-solving, prevalent in both scientific research and practical engineering. MOO challenges involve balancing multiple, often conflicting objectives, a concept central to decision-making in complex systems.

The choice of an optimization algorithm that aligns with the decision maker's preferences is fundamental, as highlighted in the seminal work of Deb [1]. Deb's exploration of evolutionary algorithms for MOO provides a foundational framework for understanding the intricacies and applications of these algorithms.

Among the plethora of algorithms in MOO, NSGA-II, NSGA-III, RVEA, and SMS-EMO stand out for their unique adaptability to various conditions. This paper endeavours to compare these algorithms, focusing on their strengths and limitations in different optimization scenarios. Such comparative analysis is vital for identifying the most suitable algorithm for specific MOO challenges.

The development of MOO algorithms heavily relies on standard test functions, which are instrumental in benchmarking and enhancing these algorithms. Functions like SCH, DEB, ZDT, DTLZ, WFG are pivotal in this context, as they allow for the extrapolation of optimization problems into multi-dimensional spaces. Each of these functions has distinct characteristics that offer diverse challenges, thereby facilitating a thorough evaluation of algorithmic performance.

Evaluating the performance of MOO algorithms generally involves three methods: focusing on convergence indexes, diversity indexes, or a combination of both. This multi-

* Corresponding author: evgenysopov@gmail.com

faceted approach is essential for a comprehensive assessment of an algorithm's effectiveness in navigating the complex landscape of MOO.

Leveraging the PYMOO algorithm framework, this paper designs four MOO algorithms and tests them across six ZDT standard problems. Through this empirical analysis, the paper aims to dissect the advantages and disadvantages of each algorithm, thereby guiding the selection of optimal MOO algorithms for specific problems, in line with the foundational principles laid out by Deb [1].

2 Multi-objective optimization algorithms

In 1967, Rosenberg [2] suggested the use of evolution-based search for multi-objective optimization problems, but there was no concrete implementation. In 1975, Holland [3] proposed a genetic algorithm. Ten years later, Schaffer [4] proposed a vector evaluation genetic algorithm, which was the first implementation of a genetic algorithm with a multi-objective optimization problem. In 1989, Goldberg, in his book "Genetic Algorithms for Search. Optimization, and Machine Learning" [5], Goldberg proposed a new idea of combining Pareto theory in economics with evolutionary algorithms to solve multi-objective optimization problems, which is of great significance for the subsequent research of evolutionary multi-objective optimization algorithms. Subsequently, evolutionary multi-objective optimization algorithms attracted the extensive attention of many scholars and many research results have emerged, and the following are some common algorithms in the field of four evolutionary multi-objective optimization.

2.1 NSGA-II: non-dominated sorting genetic algorithm

NSGA-II, a classic multi-objective optimization genetic algorithm, was proposed by Deb et al. in 2002 [6]. This algorithm is distinguished by its fast non-dominated sorting method and a crowding comparison operation. These features help maintain population diversity and accelerate convergence to the Pareto front. NSGA-II has demonstrated excellent performance across various engineering and scientific problems, establishing it as a benchmark for new algorithm evaluation. In addition to its merits, NSGA-II encounters challenges in handling high-dimensional multi-objective optimization efficiently. The "efficiency problem" here refers to the algorithm's performance in managing computational resources and time, especially in complex, high-dimensional scenarios. This issue is particularly relevant in the context of the improved NSGA-II algorithm discussed by Li and Gou, where optimization of test resource allocation is explored [7].

2.2 NSGA-III

Overall, NSGA-III, while sharing a similar framework with NSGA-II, differentiates itself primarily in its selection mechanism. NSGA-II employs crowding for ranking, a method less effective in high-dimensional objective spaces. In contrast, NSGA-III adapts this mechanism and maintains population diversity through the introduction of widely-distributed reference points [8]. This makes NSGA-III particularly adept at solving super-multi-objective optimization problems, which involve four or more objectives. Geng et al. highlight NSGA-III's specific design for multi-objective optimization problems, emphasizing its enhanced performance in scenarios requiring simultaneous optimization of multiple objectives [9].

2.3 RVEA: reference vector guided evolutionary algorithm

The Reference Vector Guided Evolutionary Algorithm (RVEA), as initially based on Deb et al.'s work [10], employs a reference point method to generate uniformly distributed reference points in the high-dimensional target space. This approach aims to yield a set of uniformly distributed Pareto approximate solutions. Building on this, Cheng et al. introduced the key feature of RVEA: the use of the scalarization method of angle penalized distance (APD). APD dynamically balances convergence and diversity in high-dimensional objective optimization, distinguishing RVEA in the realm of multi-objective evolutionary algorithms [11].

2.4 SMS-EMOA: multi objective selection based on dominated hypervolume

The core of the SMS-EMOA algorithm utilizes the metric of dominating hypervolume to guide the search process. The dominating hypervolume is the volume of the region in the objective space that consists of a set of solutions and a reference point. It is a commonly used quality metric for comparing the results of multi-objective optimization algorithms [12].SMS-EMOA employs a selection mechanism based on the contribution of the hyper-volume, which gives preference to those individuals that contribute more to the hyper-volume. This mechanism encourages individuals in the population to evolve towards the Pareto frontier while maintaining diversity among individuals. SMS-EMOA demonstrates excellent performance on several standard multi-objective optimization test problems, effectively approximating the Pareto frontier and maintaining solution diversity. It joins algorithms such as NSGA-II, SPEA2, MOPSO, and MOEA/D as standard solvers when solving multi-objective optimization problems.

3 The numerical experiments and analysis

In this section, we present the methodology and results of our numerical experiments. The experiments are designed to evaluate and compare the performance of various multi-objective optimization algorithms. We use the ZDT (Zitzler-Deb-Thiele) family of test functions, which are widely recognized benchmark problems in multi-objective optimization.

3.1 Test functions

The ZDT test functions, proposed by Zitzler, Deb, and Thiele, are integral for assessing the performance of multi-objective optimization algorithms. Each function in the ZDT suite is designed to simulate different challenges commonly encountered in multi-objective optimization.

3.1.1 ZDT1

ZDT1 features a continuous, convex Pareto front, ideal for evaluating algorithms' ability to handle problems with smooth, convex trade-off surfaces.

3.1.2 ZDT2

ZDT2 also presents a continuous design space but with a non-convex Pareto front. This function tests algorithms' efficacy in handling problems where the trade-off surface curves inward, away from the origin.

3.1.3 ZDT3

ZDT 3introduces separated, non-continuous Pareto fronts, challenging algorithms to identify and maintain diverse solution sets across disjointed trade-off regions.

3.1.4 ZDT4

ZDT4 is characterized by a non-convex Pareto front and a design space replete with multiple local Pareto fronts. This function assesses how well algorithms can avoid local optima and converge to the global Pareto front.

3.1.5 ZDT5

ZDT5 differs significantly with a discrete design and objective space, testing algorithms' performance in discrete optimization scenarios.

3.1.6 ZDT6

ZDT6 with a non-convex and biased Pareto front, poses a challenge for algorithms in terms of handling biased trade-off surfaces where solutions are not uniformly distributed.

These functions collectively enable a comprehensive evaluation of algorithms across various scenarios, including convex/non-convex Pareto fronts, continuous/discrete design spaces, the presence of multiple local optima, and separated Pareto fronts.

3.2 Experimental environment

The experimental setup comprised a computer equipped with an Intel Core i5-12400F 6-core processor, operating at 2.5GHz, and 16GB of RAM. These experiments were executed on a system running Windows 11. For programming and conducting experimental simulations, Python 3.12.0 was utilized. All development was carried out in the PyCharm 2023.2.3 integrated development environment (IDE).

We employed the PYMOO 0.6.0 framework to implement and evaluate four multi-objective optimization algorithms: NSGA-II, NSGA-III, RVEA, and SMS-EMOA. To ascertain the robustness and consistency of these algorithms, we executed 30 independent runs. Each run was initialized with a unique random seed, ranging from 0 to 29.

In the PYMOO framework, the multi-objective optimization algorithms, namely NSGA-II, NSGA-III, RVEA, and SMS-EMOA, generally utilize a set of common default parameters. These include:

Crossover: Typically, these algorithms adopt the Simulated Binary Crossover (SBX) method. SBX is a widely acclaimed strategy for generating new individuals in a population. The standard crossover probability is usually set around 0.9. Mutation: Polynomial Mutation is the chosen strategy for mutation. This approach plays a pivotal role in preserving the genetic diversity within the population. The mutation probability is commonly set to $1/n\_vars$, with n_vars representing the number of decision variables in the problem.

3.3 Performance indicators

3.3.1 Hypervolume (HV) performance of different algorithms with different ZDT functions

Hypervolume (HV) is a widely used performance metric in multi-objective optimization. It measures the region in the objective space that is dominated by the non-dominated (or Pareto-optimal) solution set and bounded by a reference point. This reference point is often worse than the worst possible solution in the objective space for all objectives. A higher HV value signifies that the solution set not only covers a larger area in the objective space but also indicates a better approximation of the Pareto front.

Diversity and Convergence: HV captures both the diversity and convergence aspects of the solution set. A diverse set of solutions will cover a more extensive area, and a set that converges well to the Pareto front will extend deeper into the desirable regions of the objective space.

Reference Point Sensitivity: The choice of the reference point is crucial as it can significantly affect the HV value. The reference point should ideally represent the worst-case scenario for each objective.

The following figures represent the Hypervolume (HV) performance of different algorithms with different ZDT functions.

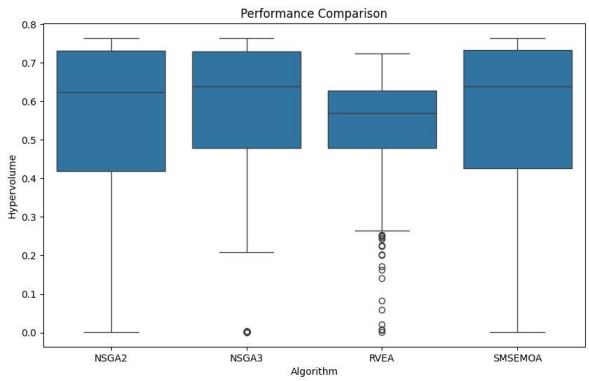

Fig. 1. HV performance of the algorithm in ZDT1.

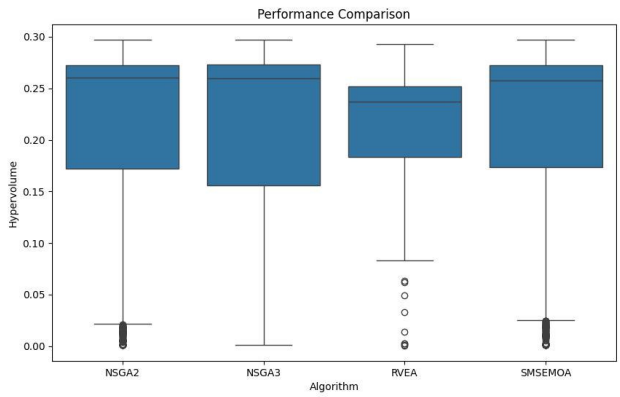

Fig. 2. HV performance of the algorithm in ZDT2.

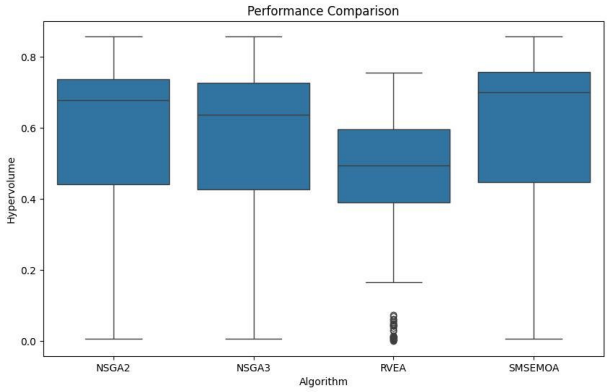

Fig. 3. HV performance of the algorithm in ZDT3.

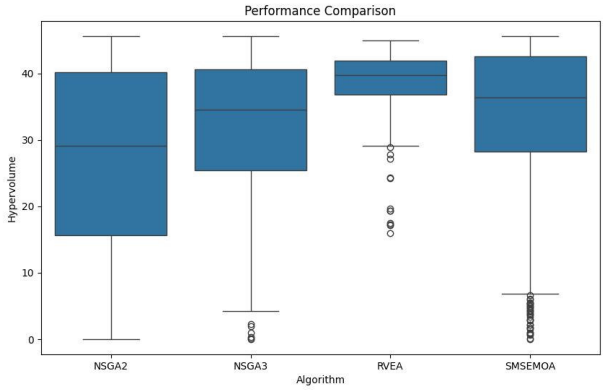

Fig. 4. HV performance of the algorithm in ZDT4.

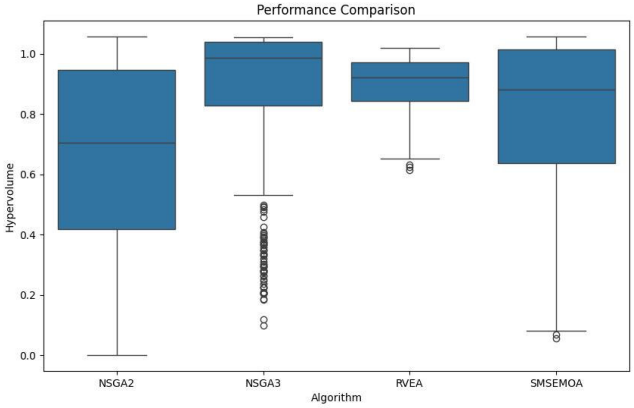

Fig. 5. HV performance of the algorithm in ZDT5.

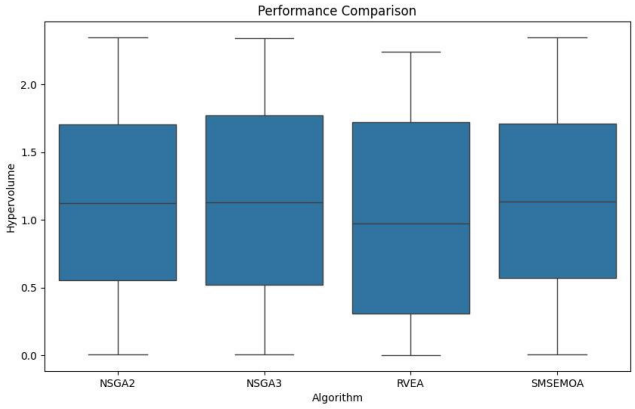

Fig. 6. HV performance of the algorithm in ZDT6.

From the above figures, we can tell that, NSGA-II and NSGA-III show consistent median HV performance across most test functions, suggesting reliable performance. SMS-EMOA also demonstrates consistency but with more variability in HV values, as indicated by the larger interquartile ranges. The spread of the HV values (interquartile range, IQR) and outliers (points outside of the 'whiskers' in the boxplot) are indicators of robustness. A narrow IQR and fewer outliers suggest an algorithm is more robust to changes in test functions. Statistical tests in Figure 2, Figure 3, Figure 5, RVEA appears to have a wider spread and more outliers, indicating less robustness. In terms of the median HV values, which indicate the central tendency of each algorithm's performance, SMS-EMOA and NSGA-II tend to have the highest median values across most ZDT functions, which may suggest better performance in approximating the Pareto front. The presence of outliers, especially those that are significantly lower than the main cluster of data, suggest that there are instances where the algorithm performs poorly compared to its typical performance. Statistical tests in Figure 2 and Figure 3, RVEA shows several outliers, indicating some runs where the performance was significantly worse than average. Each ZDT test function poses different challenges to optimization algorithms. For example, ZDT3 and ZDT6 are known to be particularly challenging due to their disconnected Pareto fronts and non-uniform density of Pareto-optimal solutions, respectively. The performance on these functions can be very telling about the algorithm's ability to handle complex optimization landscapes. Comparing algorithms, NSGA-II and SMS-EMOA seem to have better overall performance in terms of higher median HV and fewer low outliers. NSGA-III appears to perform slightly less well but is still competitive, especially on ZDT6. RVEA has a more varied performance with a significant number of outliers on ZDT2 and ZDT3, which might indicate difficulty with certain types of optimization landscapes.

In conclusion, based on these plots, NSGA-II and SMS-EMOA would generally be preferable choices for their higher and more consistent hypervolume scores across the ZDT benchmark functions. However, the specific choice of algorithm may also depend on the particularities of the problem at hand, such as the landscape of the objective space and the desired balance between exploration and exploitation.

3.3.2 Generational Distance (GD) performance of different algorithms with different ZDT functions

Generational Distance (GD) is a measure of how close the solutions found by an algorithm are to the known Pareto-optimal front. It is computed as the average distance from each solution in the non-dominated set to the nearest point on the true Pareto front.

Convergence Measurement: GD is primarily a convergence measure, assessing how near the algorithm's solution set is to the Pareto front. A smaller GD value indicates better convergence.

Limitation: While GD is effective in measuring convergence, it does not provide information about the spread or diversity of the solutions. Hence, it is often used in conjunction with diversity metrics.

The following figures represent the Generational Distance (GD) performance of different algorithms with different ZDT functions.

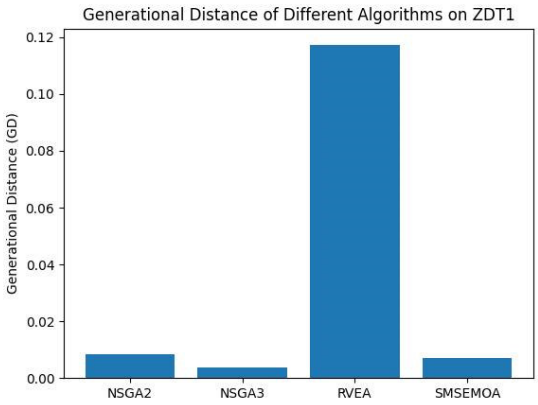

Fig. 7. GD performance of the algorithm in ZDT1.

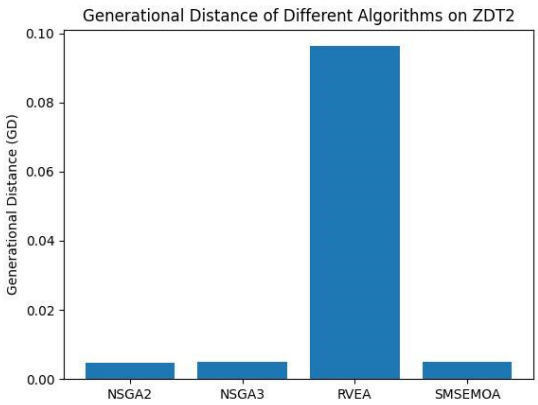

Fig. 8. GD performance of the algorithm in ZDT2.

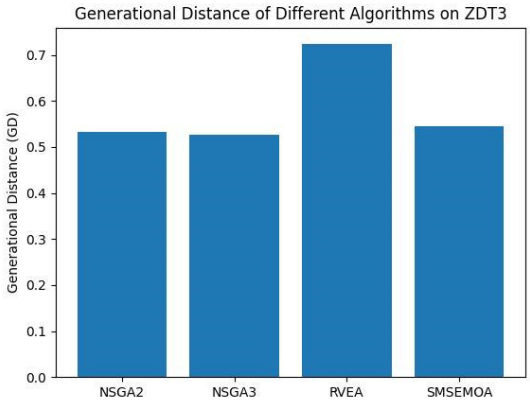

Fig. 9. GD performance of the algorithm in ZDT3.

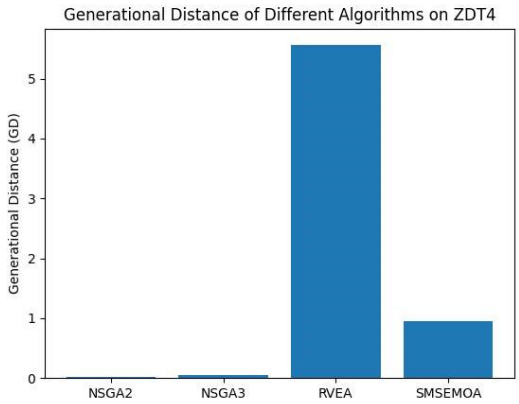

Fig. 10. GD performance of the algorithm in ZDT4.

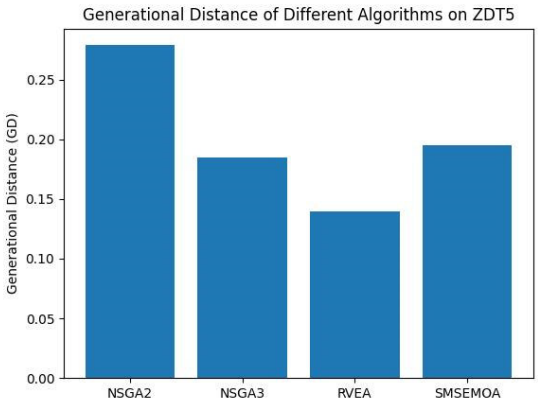

Fig. 11. GD performance of the algorithm in ZDT5.

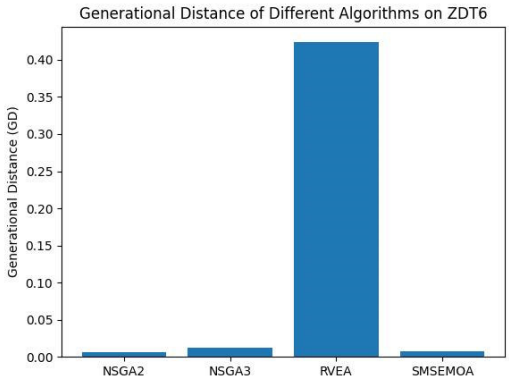

Fig. 12. GD performance of the algorithm in ZDT6.

NSGA-II and NSGA-III are generally consistent in maintaining a low GD across most functions, suggesting they are reliable for approximating the Pareto front. Statistical tests RVEA Figures 8,10,12 exhibits high GD values in several test functions (ZDT2, ZDT4, and ZDT6), which implies that it may not consistently approximate the Pareto front well across different types of problems. SMS-EMOA shows a middle-ground performance with reasonable GD values but not consistently as low as NSGA-II or NSGA-III.

In conclusion, NSGA-II and NSGA-III would likely be the preferred choices for their consistent proximity to the Pareto front across multiple ZDT functions. RVEA seems to struggle with ZDT functions, indicating it may not be the best choice for problems like those test functions. SMS-EMOA represents a reasonable compromise, although not excelling to the level of NSGA-II or NSGA-III.

3.3.3 Pareto frontiers for the ZDT family of functions and four algorithms

Pareto solutions are also known as nondominated solutions: when there are multiple objectives, a solution that is best for one objective may be worst for others due to the existence of conflicts and incomparability between the objectives. These solutions that improve any objective function while necessarily weakening at least one other objective function are called non-dominated or Pareto solutions. A set of optimal solutions to an objective function is called a Pareto optimal set. The surface formed in space by the optimal set is called Pareto optimal front.

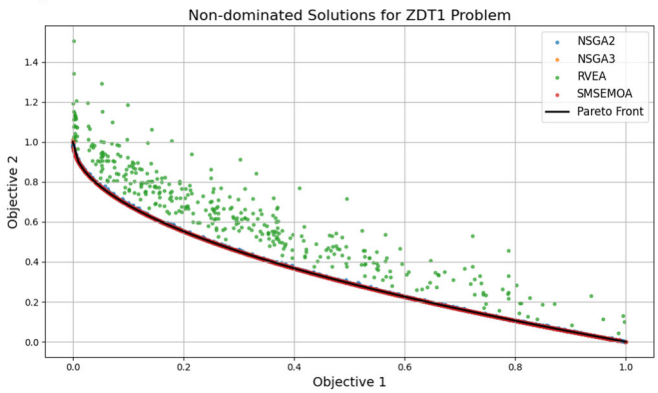

Fig. 13. Pareto frontiers for ZDT1 and other algorithms.

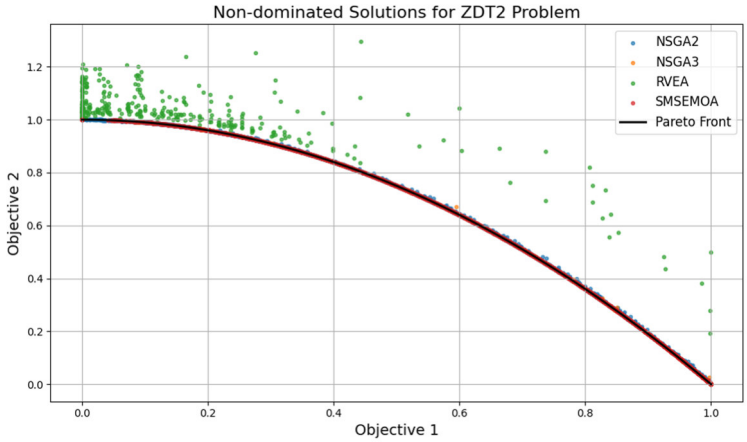

Fig. 14. Pareto frontiers for ZDT2 and other algorithms.

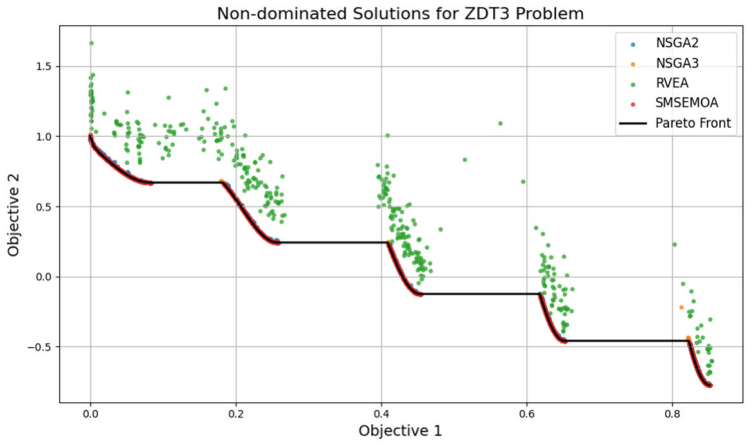

Fig. 15. Pareto frontiers for ZDT3 and other algorithms.

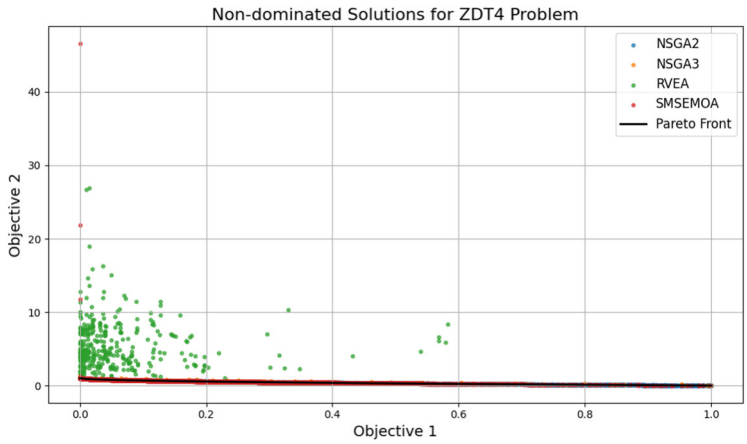

Fig. 16. Pareto frontiers for ZDT4 and other algorithms.

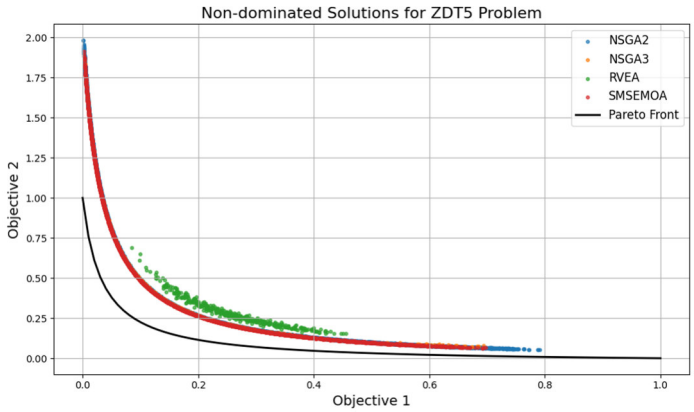

Fig. 17. Pareto frontiers for ZDT5 and other algorithms.

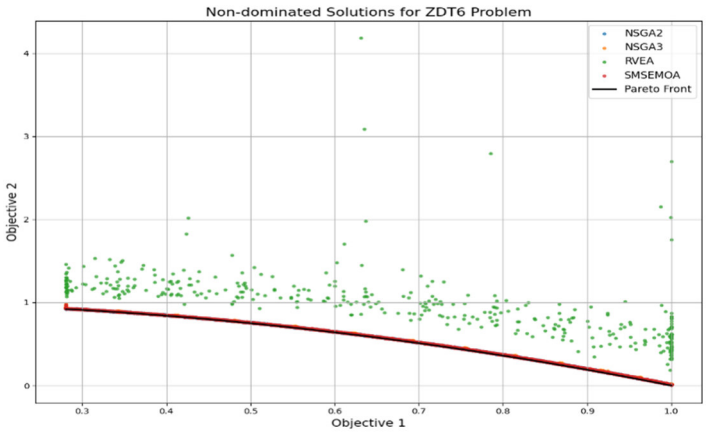

Fig. 18. Pareto frontiers for ZDT6 and other algorithms.

In analysing the ZDT benchmark problems, NSGA-II, NSGA-III, and SMS-EMOA approximate the true Pareto front effectively in ZDT1, with NSGA-III slightly outperforming others in terms of spread, indicative of a better distribution, although RVEA shows commendable coverage at the extremes, its presence near the front is less dense. The performance trends observed in Figure13 persist in Figure 14, where the continuity of the Pareto front is well represented by NSGA-II and NSGA-III, the latter achieving superior diversity; RVEA and SMS-EMOA, while adequate, do not match NSGA-III's distribution density. ZDT3 introduces a discontinuous Pareto front, challenging all algorithms; NSGA-II and NSGA-III adeptly cover these regions, especially in the central segment of objective 1, whereas RVEA and SMS-EMOA's solutions are more dispersed, reflecting the complexity of the problem. In Fig.16, NSGA-III and SMS-EMOA exhibit impressive diversity with a uniform spread of solutions along the Pareto front; conversely, RVEA's solutions tend to cluster, which might signal a limited exploration scope, while NSGA-II, though less dense, shows greater uniformity. For ZDT5, NSGA-II and SMS-EMOA maintain a uniform distribution across the Pareto front, suggesting robust diversity, unlike RVEA, which favours certain sections, potentially reducing solution diversity; NSGA-III, meanwhile, demonstrates a median performance. Lastly, ZDT6 reveals RVEA's solutions are concentrated in specific areas, indicating poor diversity; however, NSGA-II and SMS-EMOA manage a more

balanced distribution, with NSGA-III's solutions being somewhat more dispersed than RVEA's but not as much as the other two algorithms.

3.3.4 Diversity Metric performance of different algorithms with different ZDT functions

Diversity metrics are crucial in evaluating the spread and uniformity of the solutions across the Pareto front.

Spread (Δ): This metric evaluates the extent of spread among the solutions in the non-dominated set. It considers the distance between the extreme solutions in the obtained front and the distances between consecutive solutions. A lower value of Δ indicates a better spread.

Maximum Spread: This metric measures the extent of coverage of the Pareto front by the solutions. It is calculated as the maximum distance in the objective space between the solutions in the non-dominated set.

The following table represent the Diversity Metric performance of different algorithms with different ZDT functions.

Table 1. Diversity Metric performance.

ZDT	**Diversity Metric**	**NSGA-II**	**NSGA-III**	**RVEA**	**SMS-EMOA**
1	Spread	0.860292	1.024958	0.685378	0.738717
	Maximum Spread	1.423236	1.422899	1.768535	1.414125
2	Spread	0.849992	1.559114	1.004682	0.731784
	Maximum Spread	1.414326	1.431858	1.487819	1.414183
3	Spread	0.793457	1.589563	0.755388	0.853817
	Maximum Spread	1.967593	1.976677	2.509796	1.980187
4	Spread	0.614822	0.997183	0.894445	1.832961
	Maximum Spread	1.472297	1.829654	25.894653	46.603964
5	Spread	0.529074	0.838988	0.644480	0.767804
	Maximum Spread	2.086601	1.917065	0.652805	1.976544
6	Spread	0.491955	0.660177	0.858630	0.480176
	Maximum Spread	1.171665	1.211698	4.061448	1.203924

The Spread measures the extent to which the solutions cover the Pareto front, with a higher value indicating a broader distribution. The Maximum Spread assesses the range of the Pareto front captured by the solutions, with higher values indicating a more comprehensive exploration of the extreme solutions.

Across the ZDT test suite, NSGA-III generally exhibits higher Spread values, suggesting an enhanced diversity in its solution sets. However, through Table 1. it can be appreciated that RVEA shows superior Maximum Spread in certain problems (ZDT3, ZDT4, ZDT6), indicating its potential in capturing the full extent of the Pareto front.

The variance in performance across the test problems suggests that no single algorithm consistently outperforms the others. Instead, each algorithm's strengths emerge differently across various types of objective landscapes. For instance, while RVEA may not always provide the most diverse set of solutions, its ability to capture the extreme points of the Pareto front is notable, especially in problems with complex Pareto fronts (ZDT4, ZDT6).

4 Conclusions

In this analysis, we synthesize evaluations of four evolutionary algorithms—NSGA-II, NSGA-III, RVEA, and SMS-EMOA—across the ZDT benchmark suite. NSGA-II and NSGA-III consistently show reliable performance with median Hypervolume (HV) scores,

indicating robustness in approximating the Pareto front. SMS-EMOA, while consistent, exhibits variability in HV, suggesting fluctuations in its performance. RVEA's wider spread in HV, particularly on ZDT2, ZDT3, and ZDT5, signifies potential vulnerabilities to specific problem landscapes. NSGA-II and NSGA-III demonstrate proximity to the Pareto front with low Generational Distance (GD), emphasizing their dependable performance. RVEA, with high GD values on select functions, may struggle with consistent Pareto front approximation. SMS-EMOA presents a compromise between the extremes, providing reasonable GD values. Regarding the Spread and Maximum Spread indicators, NSGA-III frequently achieves higher Spread values, suggesting diverse solution sets. RVEA stands out with superior Maximum Spread in several test functions, notably capturing the extremes of the Pareto front, a trait valuable for complex optimization problems like ZDT4 and ZDT6. The aggregate analysis indicates that NSGA-II and SMS-EMOA are preferable for their robustness and consistency, as reflected in HV scores. NSGA-III is competitive, particularly on challenging problems like ZDT6. RVEA's varied performance and notable outliers suggest a sensitivity to certain problem types, which may necessitate caution in its application.

In conclusion, our findings suggest a nuanced selection of optimization algorithms based on problem characteristics and desired outcomes. Future research should explore adaptive mechanisms that allow algorithms to tailor their search strategies to specific problem landscapes, thereby enhancing robustness and performance. Additionally, further investigation into the causes of outliers and variability in RVEA's performance could lead to refinements in algorithmic design for improved robustness across diverse optimization scenarios.

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No.075-15-2022-1121).

References

1. K. Deb, *Multi-objective optimisation using evolutionary algorithms: an introduction, Multi-objective evolutionary optimisation for product design and manufacturing* (London, Springer London, 2011)
2. R.S. Rosenberg, *Simulation of genetic populations with biochemical properties* (University of Michigan, 1967)
3. J.H. Holland, Ann Arbor **7**, 390-401 (1975)
4. J.D. Schaffer, *Multiple objective optimizations with vector evaluated genetic algorithms. In Proceedings of the first international conference on genetic algorithms and their applications,* Psychology Press, 93-100, (2014)
5. D.E. Goldberg, Cenetic Algorithms in Search, Optimization, Machine Learning, (1989)
6. K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, IEEE transactions on evolutionary computation **6(2)**, 182-197 (2002)
7. Ch. Li, Y. Gou, Science Consulting **11**, 38, (2017)
8. R. Bhesdadiya, I. Trivedi, P. Jangir, et al., Cogent Engineering **3(1)**, 1269383 (2016)
9. H. Geng, Z. Zhou, J. Shen, et al., J. Entropy **25(1),** 13 (2002)
10. K. Deb, H. Jain, IEEE transactions on evolutionary computation **18(4)**, 577-601 (2013)
11. R. Cheng, Y. Jin, M. Olhofer, et al., IEEE Transactions on Evolutionary Computation **20(5)**, 773-791 (2016)
12. N. Beume, B. Naujoks, M. Emmerich, European Journal of Operational Research **181(3)**, 1653-1669 (2007)