BBE: Basin-Based Evaluation of Multimodal Multi-objective Optimization Problems

Jonathan Heins[1] , Jeroen Rook[2(✉)] , Lennart Schäpermeier[1] ,
Pascal Kerschke[1] , Jakob Bossek[3] , and Heike Trautmann[2,4]

[1] Big Data Analytics in Transportation, TU Dresden, Dresden, Germany
{jonathan.heins,lennart.schaepermeier,
pascal.kerschke}@tu-dresden.de
[2] Data Management and Biometrics, University of Twente,
Enschede, The Netherlands
j.g.rook@utwente.nl
[3] AI Methodology, RWTH Aachen University, Aachen, Germany
bossek@aim.rwth-aachen.de
[4] Data Science: Statistics and Optimization, University of Münster,
Münster, Germany
trautmann@wi.uni-muenster.de

Abstract. In multimodal multi-objective optimization (MMMOO), the focus is not solely on convergence in objective space, but rather also on explicitly ensuring diversity in decision space. We illustrate why commonly used diversity measures are not entirely appropriate for this task and propose a sophisticated basin-based evaluation (BBE) method. Also, BBE variants are developed, capturing the anytime behavior of algorithms. The set of BBE measures is tested by means of an algorithm configuration study. We show that these new measures also transfer properties of the well-established hypervolume (HV) indicator to the domain of MMMOO, thus also accounting for objective space convergence. Moreover, we advance MMMOO research by providing insights into the multimodal performance of the considered algorithms. Specifically, algorithms exploiting local structures are shown to outperform classical evolutionary multi-objective optimizers regarding the BBE variants and respective trade-off with HV.

Keywords: Multi-objective optimization · Multimodality · Performance metric · Benchmarking · Continuous optimization · Anytime behavior

---

J Heins and J. Rook—Equal contributions.

---

1 Introduction

Multi-objective optimization (MOO), i.e., the simultaneous optimization of multiple (often competing) objectives, is challenging for both research and industrial applications [19]. Despite the practical relevance of MOO and decades of research in this area, multi-objective optimization problems (MOPs) have long been treated as black boxes – probably due to their numerous dimensions in the decision and objective space. This view made it very difficult to study a MOP's properties or the algorithmic behavior on it. As a result, MOPs were often visualized only by their Pareto fronts (i.e., a representation of the non-dominated solutions of the MOP in the objective space), algorithms were designed to converge to this Pareto front as fast as possible, and visualization of this search behavior was often based on point clouds evolving towards the Pareto front.

In related domains, such as single-objective continuous optimization, knowledge of a problem's characteristics has proven to be critical for a better problem understanding and for designing appropriate algorithms. For example, in single-objective optimization (SOO), it is widely accepted that multimodality can pose difficult obstacles [21]. Despite the insights gained in SOO, research in MOO has only recently begun to focus on multimodality [9]. Nevertheless, several visualization methods capable of revealing multimodal structures of MOPs [13,24,25,30], definitions that provide a theoretical description of a MOP's structural characteristics such as locally efficient sets [15,16], a couple of benchmark suites consisting mainly of multimodal MOPs [8,12,17,34], and optimization algorithms with a particular focus on (finding or at least exploiting local structures of) multimodal MOPs [10,18,26,29] have been proposed in recent years.

All these advances ultimately help to gain a better understanding of MOPs and to develop more efficient algorithms. For example, combining visualizations and theoretical definitions helps categorize MOPs into four categories of multimodality: (1) *Unimodal* MOPs consist of a single locally efficient set (i.e., the multi-objective counterpart of a local optimum in single-objective optimization) that naturally maps to the (single) Pareto front of the MOP; (2) *Multiglobal* MOPs contain multiple efficient sets that are all mapped to the same Pareto front; (3) *Multilocal* MOPs that contain multiple locally efficient sets that map to different fronts in the objective space; (4) (Truly) *Multimodal* MOPs, where some efficient sets map to the same (Pareto) front and others map to different fronts. A schematic representation of multiglobal and multilocal MOPs is shown in Fig. 1. Note that due to space limitations, we refrain from showing unimodal and multimodal MOPs as those are special variants of the shown ones.

Multimodal solutions may, e.g., be interesting to consider if the decision space values of the optimal points are not feasible, but the values of only slightly worse non-optimal solutions are. The problem with assessing algorithm performance concerning multimodality is that classical evaluation methods like hypervolume (HV) [36] cannot account for dominated solutions. Therefore, measures that consider decision space diversity like the Solow-Polasky (SP) indicator [28] are used to express to how distributed the solutions are over the efficient sets. A recent study on the two aforementioned indicators showed that these indicators
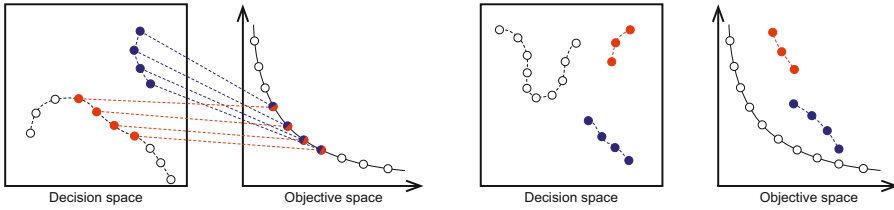
Fig. 1. Schematic differentiation of a multiglobal (left) and a multilocal MOP (right). There are two further specializations: MOPs containing only a single efficient set (and front) are called unimodal, whereas MOPs, which are both multilocal and multiglobal, are called multimodal. This figure is inspired by Figure 1 in [9].

alone are not achieving the desired performance assessment in multimodal problems by MOO algorithms [23]. However, diversity measures, also those specifically considering multimodality [22], can neither capture the properties of the HV nor problem-specific aspects. Thus, there is a need for a new indicator developed in this paper.

Also in the light of obtaining better problem understanding there is a shortage of multi-objective landscape features. These features, however, are needed for a variety of tasks, such as automated algorithm selection [14].

The remainder of this paper is organized as follows. Section 2 describes the considered algorithms and performance indicators from the literature. Subsequently, Sect. 3 presents our proposed measure. Finally, our experimental study is described in Sect. 4 before Sect. 5 concludes our work.

2   Background

2.1   Algorithms

Until recently, the main focus of algorithmic developments was on the approximation of globally optimal solutions of MOPs. Well-known and commonly used respective evolutionary MOO algorithms (EMOAs) are NSGA-II [4] and SMS-EMOA [7]. NSGA-II uses non-dominated sorting in a first step and crowding distance in a second step to focus on global convergence and diversity in objective space. SMS-EMOA implements a $(\mu + 1)$ steady-state approach where NSGA-II's second step is replaced by a procedure which drops the individual with the least contribution to the dominated hypervolume. Again, this algorithm does not consider diversity in decision space. Omni-Optimizer [6] was developed with the idea in mind to be very generic in the sense that it allows for optimization of both single- and multi-objective problems. It operates very much like NGSA-II, but also adopts diversity preservation in decision space. NSGA-II, SMS-EMOA and Omni-Optimizer are all evolutionary optimization algorithms. Recently, Schäpermeier et al. [26] proposed MOLE as a member of the family of gradient-based multi-objective optimizers, which refines upon earlier work on the MOGSA concept [10]. MOLE takes a very different approach to multimodality by actively modeling locally efficient sets and exploiting interactions

between their attraction basins, leading to a sequential exploration of the MO optimization landscape. Note that due to this sequential nature, MOLE does not maintain a "population" and thus points need to be sampled from its return values to enable a fair comparison to the other algorithms, which have fixed-size populations.

2.2 Indicators

Performance assessment of multi-objective optimizers is a non-trivial task. Research came up with a plethora of indicators, i.e., functions which map an approximation set to the real-valued numbers. Usually, an indicator measures either cardinality, convergence, spread/diversity, or a combination of these. Prominent examples are the Inverted Generational Distance [36] or the dominated hypervolume in the objective space or the Solow-Polasky measure in decision space. In this study we focus on the latter two for which we provide more details:

The *hypervolume (HV)* [36] is arguably one of the most often used performance indicators im MOO. The dominated hypervolume can be interpreted as the (hyper-)space enclosed by the approximation set and the reference point. HV rewards both convergence to the Pareto front and diversity and brings along many desirable properties, e.g., Pareto compliance [37].

In 1994, Solow and Polasky introduced their eponymous *Solow-Polasky* (SP) indicator to measure the amount of diversity between species in biology [28]. Its first application in evolutionary computation dates back to work by Ulrich and Thiele [33] in the context of *Evolutionary Diversity Optimization* to guide a single-objective EA towards diversity in (continuous) decision space subject to a minimum quality threshold. Given a set of points $X = \{x_1, \ldots, x_\mu\}$ and pairwise (Euclidean) distances $d(x_i, x_j)$, $1 \leq, i, j \leq \mu$ let M be a $(\mu \times \mu)$ matrix with $M_{ij} = \exp(-\theta \cdot d(x_i, x_j))$. Here, θ serves for normalization of the relation between d and the number of species; its choice is not critical [28]. Now the Solow-Polasky diversity is defined as the sum of all elements of the Moore-Penrose generalized inverse M^{-1} of the matrix M. The measure can be interpreted "as the number of different species in the population" [28]. Note, however, that the measure calculates a real-valued diversity in $[1, \mu]$ and no integer value. As pointed out in [32], SP is maximized if points are aligned in a grid.

3 The BBE Measure(s)

For classical MOO, the HV serves as an excellent measure capturing the coverage of the Pareto front. However, in MMMOO, the local efficient sets that, per definition, cannot contribute to the HV are of interest as well. One approach to achieve this is measuring decision space diversity with SP. SP, however, does not focus on the coverage of the local efficient sets but rather on the coverage on the whole decision space. Therefore, we introduce a *basin-based evaluation (BBE)* method in this paper, which focuses on the coverage of the Pareto front as well
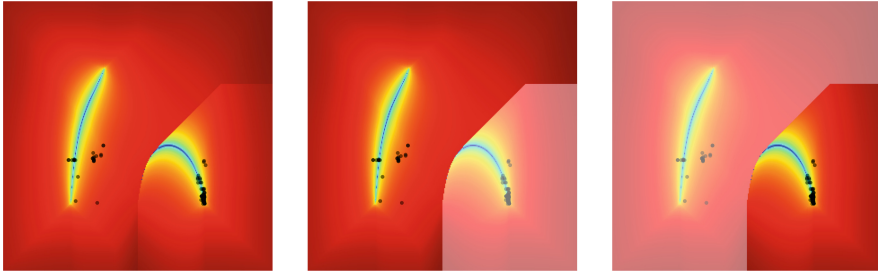
Fig. 2. If a set of solution points (the black points in the left image) is to be evaluated with BBE, first (middle image) only the points in the first basin (on the left hand side) are evaluated and then (right image) only the points in the next basin (on the right hand side). This continues until all basins of interest are evaluated.

as local efficient sets simultaneously. The main idea is to compute the HV per basin and not only globally.

The division of the decision space into basins is done based on the technique for decision space visualization by Schäpermeier et al. [24]. In order to visualize the optimization landscape, they divide the decision space into equal-sized regions arranged in a grid. Every region is represented by the contained point with the lowest value in all decision variables. This enables the computation of multi-objective gradients for all parts of the grid. Then, based on the hull spanned by the gradients, regions which likely contain parts of the efficient sets can be identified. With the gradients and the approximation of the efficient sets, the path from a region to an efficient set can be traced. With this, the accumulated gradient length along the path can be calculated as a measure of distance to the attracting set. Based on this measure of distance, the visualization of the regions is determined. For more detailed information on this procedure we refer the interested reader to [24]. As a by-product, the affiliation of a region to a basin is determined as well. The latter information is used to evaluate a set of returned solution points separately per basin for the proposed measure. For every point of the solution set, the region, and thus, the corresponding basin they are encapsulated in, is identified. This allows to filter out all points that are not contained in a basin and calculate a specific metric only for the points of interest. In the default case, this metric is the HV. See Fig. 2 for a visualization.

As not all basins can be included in the evaluation, only the first k are evaluated. The order of the basins is determined by non-dominated sorting [4] of the regions that approximate the efficient sets. The basin which contributes the most points to the approximated Pareto front thus will be the first in the constructed order. In case of an equal number of regions, the number of regions attributed to the next domination layer is decisive. Basins that are not part of the Pareto front have an equal number of regions (i.e., zero) constituting the Pareto front. In case of many basins, they can also be joined based on the contained
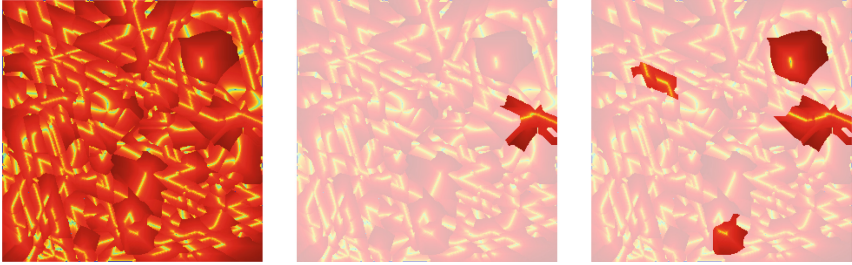
Fig. 3. The shown gradient field heatmap belongs to the same highly multimodal problem instance as the PLOT visualization in Fig. 8: the bi-objective BBOB function with FID 55 and IID 11. If the basins are considered individually the area covered by the first basin is relatively small (image in the middle). To circumvent too many individual basins the efficient sets can be joined, which leads to a potentially distributed area covered by the first basin (right image).

regions attributed to the domination layer with the lowest number. The joint basins are regarded as a single one during evaluation, see Fig. 3.

To aggregate the attained HV from the k basins of interest, the arithmetic mean is taken. Note that this includes a natural weighting between the basins, as a higher HV is attainable in the basins closer to the Pareto front in case the reference point is fixed. To capture an algorithm's anytime behavior, this mean is recorded in every interval of a specific number of function calls needed by the algorithm. Here, one can decide if the accumulation of all the points evaluated by the algorithm up until this point (the solution archive) or only the ones evaluated in the interval should be considered. A visualization of the latter case can be seen in Fig. 4. In order to aggregate those intermediate results, the area under the curve is computed. This value captures the anytime behavior of an algorithm regarding the convergence to k local efficient sets of interest when we focus on multimodal multi-objective optimization.

4   Analysis

To test our basin-based indicators we conduct a benchmark study with four MO optimizers on a set of multimodal problem instances. First, we provide the experimental setup, followed by our experimental results and we end with a discussion and interpretation of these results.

4.1   Experimental Setup

Hardware and Software. All experiments were conducted on PALMA, the high-performance compute cluster of the WWU Münster. Each optimizer run had access to 1 CPU core and 4 GB of memory. In total, all experiments required 20 000 CPU hours. All experimental code for reproducibility is available online[1].
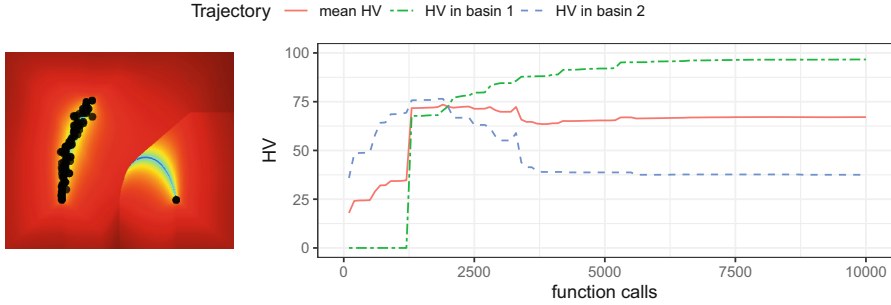

Fig. 4. Shown is a run of NSGA-II on the gradient field heatmap of the Aspar Function [9] together with the corresponding BBE scores. Note that only the most recent points are considered, and therefore, the achieved HV of the points in the second basin eventually declines. (See supplementary material for an animated version of the figure.)

Resources. To run our analysis we use the R implementations of the optimizers SMS-EMOA [1], NSGA-II [1], Omni-Optimizer [3], and MOLE [27]. Furthermore, we compiled a set of 35 well-established, mainly multimodal MOP instances. We selected all instances from ZDT [35], DLTZ [5], MMF [34], with exception of MMF13 and ZDT5, which are provided by `smoof` [2]. Furthermore, we selected 5 problem instances (FID $\in \{46, 47, 50, 10, 55\}$, IID $= 1$) from BiOb-jBBOB [31], which are provided by `moPLOT` [24]. All problem instances have a 2D decision and objective space. For each instance, we approximated the Pareto front and chose the reference point such that it covered the whole reachable objective space with `moPLOT`. These reference points are needed to compute the HV and the BBE indicators. The approximated Pareto front is used to compute the maximum obtainable HV. In turn, the maximum HV is used to normalize the BBE and HV indicators to make them comparable.

Indicators. For computing the BBE measures we use our own R package[2]. We used 3 variants of BBE; 1) the mean HV of the basins with the population returned by the optimizers (BBE(HV)), 2) the mean HV of the basins with the complete archive of function calls the optimizer made ($\text{BBE}^{\text{cum}}(\text{HV})$), and 3) the area under the curve of the convergence of the mean HV across all basins during search ($\text{BBE}^{\text{cum,auc}}(\text{HV})$). For each variant we considered the 5 most important basins (automatically derived by the BBE package with the landscape exploration of `moPLOT`) and we did not merge basins of joined fronts.

Configuration. To maximize optimizer performance on the problem instances w.r.t. the indicators we make use of the automated algorithm configurator SMAC [11] (through Sparkle[3]). Here, each configuration scenario aims to maximize

---

[2] BBE available at: github.com/jonathan-h1/BBE.

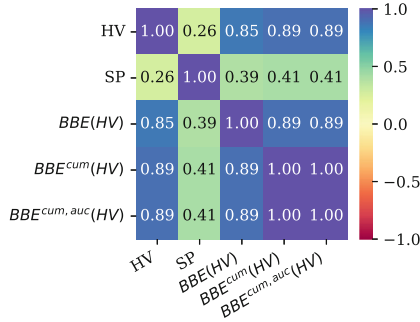[3] Accessible through ada.liacs.nl/projects/sparkle.

Fig. 5. Spearman correlation between performance metrics. Points taken from 25 runs with all algorithms in all configuration scenarios and on all instances.

performance for one indicator in 10 separate configuration runs. Each configuration run had a budget of 250 algorithm calls and the configuration run with the highest performance score on the whole training set is used for validation. We used leave-one-out validation, i.e., we configured on 34 instances to derive the parameters and then validated the performance on the left out instance. Separate configuration experiments were conducted, each aiming to maximize one of the 5 indicator scores: HV, SP, BBE(HV), BBE^{cum}(HV), and $BBE^{cum,auc}$(HV).

Validation. We validated each optimizer configuration on each instance 25 times with fixed random seeds per run. The median score over these runs was used to represent the configuration's performance. Furthermore, each run was given a budget of 25 000 function evaluations. In total, we validated on all 840 pairs that can be made out of the 6 configurations (including the default configuration), 4 algorithms, and 35 problem instances.

4.2 Indicator Similarity

We start our analysis by focusing on the similarity of the measures. We specifically look at the Spearman correlations between the indicators over all conducted optimizer runs, which are visualized in Fig. 5. A high correlation score indicates that the two indicators yield a similar ordering within the underlying optimizer runs.

The correlation matrix shows that the global convergence in objective space measured by hypervolume (HV) is highly correlated with the variants of our basin-based approach. These high correlations indicate that, despite the reduction in focus on obtaining global convergence, the basin-based approaches are still able to measure this property. Another observation is a clear trade-off between

Table 1. The mean indicator score of the best ranked optimizer under different configuration scenarios. Before the mean was taken over the runs, all indicators, except SP, were first normalized against the maximum approximated HV for each instance.

Indicator	Default	Configuration target				
		HV	SP	BBE(HV)	BBE^{cum}(HV)	$BBE^{cum,auc}$(HV)
HV	1.031	1.071	1.067	1.048	**1.072**	1.053
SP	2.226	5.195	**5.568**	3.598	5.197	3.572
BBE(HV)	0.417	0.515	0.497	**0.531**	0.517	0.488
BBE^{cum}(HV)	0.613	**0.651**	0.650	0.647	0.650	**0.651**
$BBE^{cum,auc}$(HV)	15 074	16 092	16 069	15 891	16 064	**16 103**

HV and the diversity in decision space (SP). Interestingly, the correlation scores between the basin-based indicators and SP are higher than between HV and SP. This suggests that the basin-based indicators are more considerate of the global decision space diversity than HV is.

We bolster the latter observations by looking at the mean indicator values of the best-ranked optimizers after automatically configuring for a particular indicator, depicted in Table 1. Here, we see that the best indicator score (or within proximity to the best) is obtained when explicitly configuring for that indicator. Additionally, the untargeted indicators tend to improve as well by configuration. Disappointingly, when we configure for the mean basin-based HV, both HV and SP are behind compared to the improvements we see in the other configuration scenarios. Speculatively, this is because the points of the last populations of the optimizer are more distributed over the different basins. Further, we see that the configurations of the cumulative BBE variants have excellent performance across all indicators compared to the BBE variant that focuses only on the points in the last population. This could potentially be explained by the fact that these variants aim to visit at least all the basins during search and not on maintaining a population across the different basins. Thus, by keeping an archive of all visited points, one can easily obtain a good coverage across all basins retrospectively.

4.3   Rankings

We now shift our focus to the rankings between optimizers under varying circumstances. Specifically, we compute the average rankings between the 4 optimizers (Sect. 2.1) for each indicator-configuration target pair, resulting in a total of 30 rankings. Figure 6 plots these rankings where each column indicates the indicator by which the ranking was generated, and the rows refer to the configuration target indicator which was used to tune the optimizers' parameters. These rankings reveal that NSGA-II and SMS-EMOA rank best for default parameters. After configuration, both optimizers are almost always exceeded by MOLE or Omni-Optimizer for all indicators, except for HV. There, SMS-EMOA also remains the best-ranked optimizer after configuration. MOLE and Omni-Optimizer have a larger parameter space compared to SMS-EMOA and NSGA-II, likely making them more configurable. However, these two optimizers are also conceptually dif-
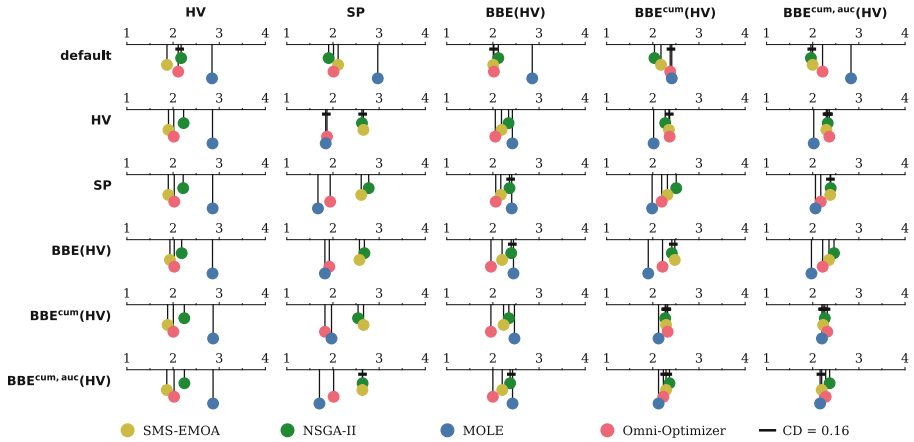
Fig. 6. Average rankings of the optimizers for each measure (columns) and configuration scenario (rows). The confidence distance (CD) to be significantly differently ranked, as determined by a Nemenyi test [20] with $\alpha < 0.1$, is 0.16.

ferent because they exploit structural knowledge of the problem instance during search. Especially for Omni-Optimizer this causes larger ranking improvements when it is configured for the basin-based indicators.

4.4 Ranking Changes

The visualized average rankings in Fig. 6 revealed significant changes in the ranking concerning a measure if the algorithms are configured based on different desired properties. However, the relationship of the ranking shifts, and therefore the mutual configuration impact on two measures cannot be assessed with this figure. Thus, we consider the correlation between the ranking shifts with regard to the measures aggregated over all runs per algorithm and problem instance. Here, the ranking shift is the difference between the average ranking if the algorithms are run with default settings and the average ranking if the algorithms are configured for HV or SP. Therefore, a high correlation between two measures means that the shifts in algorithm rankings per problem instance are similar. The corresponding correlation heatmaps can be seen in Fig. 7. Independent of whether the aim is convergence to the Pareto front (configuring for HV) or maximizing diversity (configuring for SP), the BBE measure variants yield a higher correlation of ranking changes with the diversity indicator SP than with HV. The correlation with HV is relatively small in general, which may be caused by the small changes in the ranking w.r.t. HV independent of the configuration (see Fig. 6). Nevertheless, the correlation of the BBE measures with SP indicates that the former are more similar to a diversity measure than HV and can result in more similar ranking changes in relation to diversity.
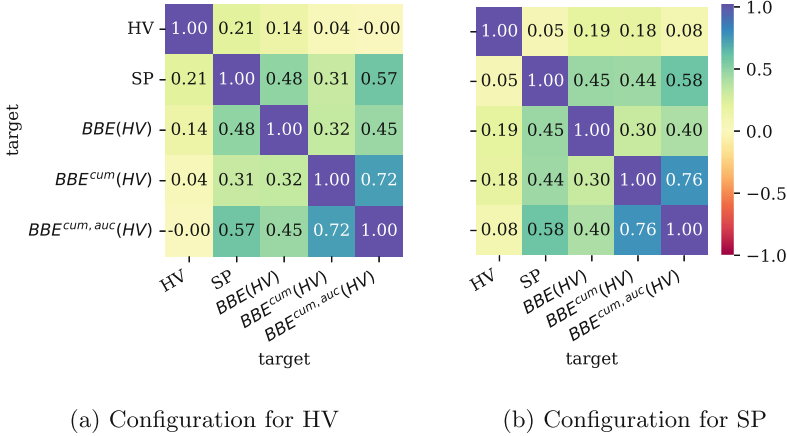
(a) Configuration for HV   (b) Configuration for SP

Fig. 7. The Spearman correlations of the differences between the resulting rankings after configuration for a certain target and the rankings with default parameters. In the left plot the configuration target is the HV and in the right one the target is the SP.

4.5   Discussion

From the presented experimental results, we derive the following insights:

First, the proposed basin-based evaluation captures the main properties of the classical HV evaluation. This is supported by the observed high correlation between HV and the introduced BBE variants. Theoretically, this is explained by the fact that HV is measured per basin. In case of only one basin containing the efficient set which makes up the Pareto front, the found solution points in other basins do not matter for the achieved HV score. In general, the proposed measure variants are generalizations of HV, where HV corresponds to the special case in which the whole decision space is regarded as one basin.

Second, the proposed measure additionally captures diversity aspects that HV cannot capture. The HV, even though a reliable measure for global convergence, does not cover decision space diversity, as shown by the low correlation with the SP. Further, the low correlation of rank shifts, if the algorithms are configured for SP, demonstrates that the HV lacks the ability to score an algorithm run based on the diversity in decision space of the proposed solution set as all points that are dominated by others cannot contribute to the HV. The BBE variants alleviate this issue, as can be seen by the higher correlation with the SP regarding the general scores and the rank shifts.

Third, the actual aim of MMMOO is not to find an algorithm that is performing well w.r.t. diversity over the complete decision space but *conditional diversity* as explained in the following. So far, in MMMOO, a form of general diversity is used to measure how points are distributed in the complete decision space and not just areas in decision space that have objective values along the Pareto front (i.e., multiglobal basins). However, this general diversity can only
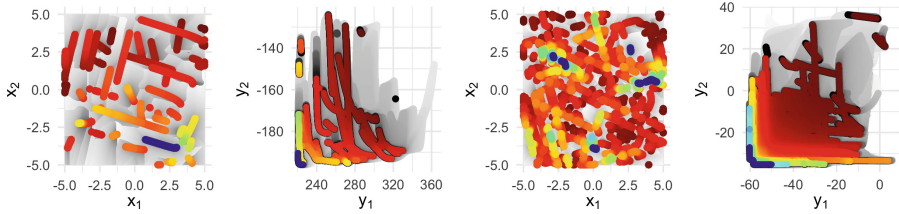
Fig. 8. PLOT visualizations of two functions (left: FID 10, IID 5; right: FID 55, IID 11) from the bi-objective BBOB in decision (x) and objective space (y). The colored points are locally efficient, with the blue points being globally efficient. The gray-scale background illustrates the attraction basins of each locally efficient set, with darker colors indicating more optimal points within a basin. While the left one contains the good-performing locally efficient sets mostly in the lower right corner, the globally efficient regions are much further spread in the right function. (Color figure online)

serve as a proxy of the actual goal to find points other than the global efficient sets. Evenly distributed points may allow choosing the favored combination of decision variables, but the corresponding objectives may be arbitrarily bad. For some problem instances, the interesting space may be a fraction of the complete decision space. As, e.g., shown in Fig. 8, the interesting basins with efficient sets close to the Pareto front are distributed in the decision space but cover only a fraction. Thus, the actual desired diversity is a conditional one. The proposed measure enforces this conditional diversity by only considering the user-defined basins of interest and neglecting the other parts of the decision space.

5   Conclusions

This paper provides different perspectives on multimodality in multi-objective optimization and explicitly contributes to problem characterization and algorithm performance evaluation. We not only introduce a specific method for acquiring comprehensive information on decision-space basins, but also propose variants of a performance indicator BBE which addresses both convergence in objective space as well as decision space diversity. In this regard, not overall diversity but rather conditional diversity adhering to local efficient sets and basin coverage is accounted for. Classical EMOAs and algorithms exploiting local problem structures are experimentally compared and automatically configured. We experimentally show that especially the latter algorithms extremely profit from being configured w.r.t. BBE. By this means, global HV as well as conditional decision space diversity are optimized and the improvements for classical EMOAs lag behind. Moreover, it has to be noted that BBE explicitly needs underlying basin information and is thus not an indicator to be incorporated into an indicator-based EMOA or that can be computed on-the-fly. However, it can be used for optimally configuring EMOAs and for getting an increased understanding on problem hardness w.r.t. multimodality. Also, it can perspectively contribute to deriving multi-objective landscape features. For future work

it would be interesting to see how BBE behaves for different classes of multi-modality (e.g., problems with only multiglobal basins) and for problems with higher dimensionality in both decision or objective space. Furthermore, instead of using HV, also other measures can potentially be used for computing the basin performance (e.g., BBE(SP)).

References

1. Bossek, J.: ECR 2.0: a modular framework for evolutionary computation in R. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO) Companion, pp. 1187–1193. ACM (2017). https://doi.org/10.1145/3067695.3082470
2. Bossek, J.: smoof: Single- and multi-objective optimization test functions. R J. **9**(1), 103–113 (2017). https://doi.org/10.32614/RJ-2017-004
3. Bossek, J., Deb, K.: omnioptr: Omni-Optimizer (2021). , R package version 1.0.0: https://github.com/jakobbossek/omnioptr
4. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. (TEVC) **6**(2), 182–197 (2002). https://doi.org/10.1109/4235.996017
5. Deb, K., Thiele, L., Laumanns, M., Zitzler, E.: Scalable test problems for evolutionary multiobjective optimization. In: Abraham, A., Jain, L., Goldberg, R. (eds.) Evolutionary Multiobjective Optimization. Advanced Information and Knowledge Processing. Springer, London (2005). https://doi.org/10.1007/1-84628-137-7_6
6. Deb, K., Tiwari, S.: Omni-optimizer: a generic evolutionary algorithm for single and multi-objective optimization. Eur. J. Oper. Res. (EJOR) **185**, 1062–1087 (2008). https://doi.org/10.1016/j.ejor.2006.06.042
7. Emmerich, M., Beume, N., Naujoks, B.: An EMO algorithm using the hypervolume measure as selection criterion. In: Coello Coello, C.A., Hernández Aguirre, A., Zitzler, E. (eds.) EMO 2005. LNCS, vol. 3410, pp. 62–76. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-31880-4_5
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 (GECCO), pp. 541–549. ACM (2019). https://doi.org/10.1145/3321707.3321727
9. Grimme, C., et a.: Peeking beyond peaks: challenges and research potentials of continuous multimodal multi-objective optimization. Comput. Oper. Res. (COR) **136**, 105489 (2021). https://doi.org/10.1016/j.cor.2021.105489
10. Grimme, C., Kerschke, P., Trautmann, H.: Multimodality in multi-objective optimization – more boon than bane? In: Deb, K., et al. (eds.) EMO 2019. LNCS, vol. 11411, pp. 126–138. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12598-1_11
11. Hutter, F., Hoos, H.H., Leyton-Brown, K.: Sequential model-based optimization for general algorithm configuration. In: Coello, C.A.C. (ed.) LION 2011. LNCS, vol. 6683, pp. 507–523. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25566-3_40
12. Ishibuchi, H., Peng, Y., Shang, K.: A scalable multimodal multiobjective test problem. In: Proceedings of the IEEE Congress on Evolutionary Computation (CEC), pp. 310–317. IEEE (2019). https://doi.org/10.1109/CEC.2019.8789971

13. Kerschke, P., Grimme, C.: An expedition to multimodal multi-objective optimization landscapes. In: Trautmann, H., et al. (eds.) EMO 2017. LNCS, vol. 10173, pp. 329–343. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54157-0_23

14. Kerschke, P., Hoos, H.H., Neumann, F., Trautmann, H.: Automated algorithm selection: survey and perspectives. Evol. Comput. (ECJ) **27**, 3–45 (2019). https://doi.org/10.1162/evco_a_00242

15. Kerschke, P., et al.: Towards analyzing multimodality of continuous multiobjective landscapes. In: Handl, J., Hart, E., Lewis, P.R., López-Ibáñez, M., Ochoa, G., Paechter, B. (eds.) PPSN 2016. LNCS, vol. 9921, pp. 962–972. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45823-6_90

16. Kerschke, P., et al.: Search dynamics on multimodal multi-objective problems. Evol. Comput. (ECJ) **27**, 577–609 (2019). https://doi.org/10.1162/evco_a_00234

17. Li, X., Engelbrecht, A.P., Epitropakis, M.G.: Benchmark functions for cec'2013 special session and competition on niching methods for multimodal function optimization. Technical report, Evolutionary Computation and Machine Learning Group, RMIT University, Australia (2013). http://goanna.cs.rmit.edu.au/~xiaodong/cec13-niching/competition/

18. Maree, S.C., Alderliesten, T., Bosman, P.A.N.: Real-valued evolutionary multimodal multi-objective optimization by Hill-Valley clustering. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pp. 568–576. ACM (2019). https://doi.org/10.1145/3321707.3321759

19. Miettinen, K.: Nonlinear Multiobjective Optimization. International Series in Operation Research and Management Science, vol. 12. Springer, Heidelberg (1998). https://doi.org/10.1007/978-1-4615-5563-6

20. Nemenyi, P.B.: Distribution-free multiple comparisons. Ph.D. thesis, Princeton University (1963)

21. Preuss, M.: Multimodal Optimization by Means of Evolutionary Algorithms. Natural Computing Series (NCS). Springer, Cham (2015). https://doi.org/10.1007/978-3-319-07407-8

22. Preuss, M., Wessing, S.: Measuring multimodal optimization solution sets with a view to multiobjective techniques. In: Emmerich, M. et al. (eds.) EVOLVE - A Bridge Between Probability, Set Oriented Numerics, and Evolutionary Computation IV, pp. 123–137. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-319-01128-8_9

23. Rook, J., Trautmann, H., Bossek, J., Grimme, C.: On the potential of automated algorithm configuration on multi-modal multi-objective optimization problems. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO) Companion. p. tbd. ACM (2022). https://doi.org/10.1145/3520304.3528998, accepted

24. Schäpermeier, L., Grimme, C., Kerschke, P.: One PLOT to show them all: visualization of efficient sets in multi-objective landscapes. In: Bäck, T., et al. (eds.) PPSN 2020. LNCS, vol. 12270, pp. 154–167. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58115-2_11

25. Schäpermeier, L., Grimme, C., Kerschke, P.: To boldly show what no one has seen before: a dashboard for visualizing multi-objective landscapes. In: Proceedings of the International Conference on Evolutionary Multi-criterion Optimization (EMO), pp. 632–644 (2021). https://doi.org/10.1007/978-3-030-72062-9_50

26. Schäpermeier, L., Grimme, C., Kerschke, P.: MOLE: digging tunnels through multimodal multi-objective landscapes. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO). p. tbd. ACM (2022). https://doi.org/10.1145/3512290.3528793, accepted

27. Schäpermeier, L.: An R package implementing the multi-objective landscape explorer (MOLE), February 2022. https://github.com/schaepermeier/moleopt

28. Solow, A.R., Polasky, S.: Measuring biological diversity. Environ. Ecol. Stat. **1**, 95–103 (1994). https://doi.org/10.1007/BF02426650

29. Tanabe, R., Ishibuchi, H.: A Niching indicator-based multi-modal many-objective optimizer. Swarm Evol. Comput. (SWEVO) **49**, 134–146 (2019). https://doi.org/10.1016/j.swevo.2019.06.001

30. Tušar, T., Filipič, B.: Visualization of pareto front approximations in evolutionary multiobjective optimization: a critical review and the prosection method. IEEE Trans. Evol. Comput. (TEVC) **19**(2), 225–245 (2015). https://doi.org/10.1109/TEVC.2014.2313407

31. Tušar, T., Brockhoff, D., Hansen, N., Auger, A.: COCO: the bi-objective black box optimization benchmarking (BBOB-BIOBJ) test suite. arXiv preprint abs/1604.00359 (2016). https://doi.org/10.48550/arXiv.1604.00359

32. Ulrich, T., Bader, J., Thiele, L.: Defining and optimizing indicator-based diversity measures in multiobjective search. In: Schaefer, R., Cotta, C., Kołodziej, J., Rudolph, G. (eds.) PPSN 2010. LNCS, vol. 6238, pp. 707–717. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15844-5_71

33. Ulrich, T., Thiele, L.: Maximizing population diversity in single-objective optimization. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pp. 641–648. ACM (2011). https://doi.org/10.1145/2001576.2001665

34. Yue, C., Qu, B., Yu, K., Liang, J., Li, X.: A novel scalable test problem suite for multimodal multiobjective optimization. Swarm Evol. Comput. **48**, 62–71 (2019). https://doi.org/10.1016/j.swevo.2019.03.011

35. Zitzler, E., Deb, K., Thiele, L.: Comparison of multiobjective evolutionary algorithms: empirical results. Evol. Comput. (ECJ) **8**(2), 173–195 (2000). https://doi.org/10.1162/106365600568202

36. Zitzler, E., Thiele, L.: Multiobjective optimization using evolutionary algorithms - a comparative case study. In: Eiben, A.E., Bäck, T., Schoenauer, M., Schwefel, H.P. (eds.) PPSN 1998. LLNCS, vol. 1498. pp. 292–301. Springer, Heidelberg (1998). https://doi.org/10.1007/bfb0056872

37. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., da Fonseca, V.G.: Performance assessment of multiobjective optimizers: an analysis and review. IEEE Trans. Evol. Comput. (TEVC) **7**(2), 117–132 (2003). https://doi.org/10.1109/TEVC.2003.810758