Understanding the Treatment of Outliers in Multi-Objective Estimation of Distribution Algorithms

Luis Martí[1], Nayat Sanchez-Pi[2], and Marley Vellasco[1]

[1] Dept. of Electrical Engineering, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro (RJ) Brazil.
{lmarti; marley}@ele.puc-rio.br
[2] Instituto de Lógica, Filosofia e Teoria da Ciéncia (ILTC), Niterói (RJ) Brazil.
nayat@iltc.br

Abstract. It has been already documented the fact that estimation of distribution algorithms suffer from loss of population diversity and improper treatment of isolated solutions. This situation is particularly severe in the case of multi-objective optimization, as the loss of solution diversity limits the capacity of an algorithm to explore the Pareto-optimal front at full extent.
A set of approaches has been proposed to deal with this problem but —to the best of our knowledge— there has not been a comprehensive comparative study on the outcome of those solutions and at what degree they actually solve the issue.
This paper puts forward such study by comparing how current approaches handle diversity loss when confronted to different multi-objective problems.

Keywords: multi-objective optimization, estimation of distribution algorithms, model building, outlier detection

1 Introduction

Multi-objective optimization problems (MOPs) are problems involving more than one goal that must be simultaneously optimized. Most real-world optimization problems are solved using MOPs. Methods that address this problems In these problems the optimizer must find one or more feasible solutions that conjointly find the extremal values (either maximum or minimum) of two or more functions subject to a set of constraints. Consequently, the solution to that problem is a set of equally valid, trade-off solutions. MOPs have been addressed using evolutionary computation (EC). This fact has prompted the creation of multi–objective optimization evolutionary algorithms (MOEAs) [1].

The inclusion of learning as part of the search process has been pointed out as a relevant alternative to "traditional" MOEAs [2]. Estimation of distribution algorithms (EDAs) [3] are one of those alternatives, as they are capable

of learning the problem structure. EDAs replace the application of evolutionary operators with the creation of a statistical model of the fittest elements of the population in a process known as *model building*. This model is then sampled to produce new elements.

However, multi-objective EDAs (MOEDAs) have not yielded the anticipated results. Most of them have limitations transforming single-objective EDAs into a multi-objective formulation by including an existing multi-objective fitness assignment function.

It can be stated that this straightforward extrapolation might had lead to skip the fact that most current EDAs have some characteristics that interferes their capacity of handling some of the requirements of multi-objective optimization.

These matters have been already pointed out in some previous works [4]. It has been said that some of them are derived from the incorrect treatment of the isolated elements of the model-building dataset (outliers); the loss of population diversity, and that too much computational effort is being spent on model construction.

In particular, previous experimental evidence [5–7] has raised the hypothesis that those model-building algorithms that are more sensitive to outliers frequently yield better results than those who don't. However, this fact must be studied in a more formal manner, paying attention to what is actually happening during the optimization process.

The purpose of this paper is to present an in-depth study regarding outliers and diversity loss issues in MOEDAs from both theoretical and experimental points of view. We carry out an experimental study measuring the diversity of populations that, to the best of our knowledge, has not been yet been carried out. This is a key assessment as its outcome is a key step towards the solution of what constitutes one of the current main problems of multi-objective EDAs.

The rest of the paper is structured as follows. Section 2 briefly discusses the theoretical foundations of the work. After that, the outliers issue is presented in Section 3. Subsequently, in Section 4, an experimental study focused on measuring the diversity of the population is debated. Finally, Section 5 contains some final remarks.

2   Foundations

Multi-objective optimization has received lot of attention by the evolutionary computation community leading to *multi-objective evolutionary algorithms* (MOEAs) (cf. [1]). A *multi-objective optimization problem* (MOP) could be expressed as the problem in which a set of M *objective functions* $f_1(\boldsymbol{x}), \ldots, f_M(\boldsymbol{x})$ with should be jointly optimized;

$$\min \ \boldsymbol{F}(\boldsymbol{x}) = \langle f_1(\boldsymbol{x}), \ldots, f_M(\boldsymbol{x}) \rangle \, ; \ \boldsymbol{x} \in \mathcal{D} \, ; \tag{1}$$

where $\mathcal{D} \subseteq \mathbb{R}^n$ is known as the *feasible set* and could be expressed as a set of restrictions over the *decision set*, that is usually $\mathbb{R}^n$. The image set of $\mathcal{D}$

produced by function vector $\boldsymbol{F}(\cdot)$, $\mathcal{O} \subseteq \mathbb{R}^M$, is called *feasible objective set* or *criterion set*.

The solution of (1) is a set of trade-off points. The adequacy of a solution can be expressed in terms of the *Pareto dominance relation*. The solution of (1) is the Pareto-optimal set, $\mathcal{D}^*$; which is the subset of $\mathcal{D}$ that contains all elements of $\mathcal{D}$ that are not dominated by other elements of $\mathcal{D}$. Its image in objective space is called Pareto-optimal front, $\mathcal{O}^*$.

If an MOP has certain characteristics, e. g., linearity or convexity of the objective functions or convexity of $\mathcal{S}$, the efficient set can be determined by mathematical programming approaches [8]. However, in the general case, finding the solution of (1) is an *NP*–complete problem. In this case, heuristic or metaheuristic methods can be applied in order to have solutions of practical value at an admissible computational cost.

Generally, an heuristic algorithm solving an MOP yields a discrete local Pareto-optimal set, $\mathcal{P}^*$, that attempts to represent $\mathcal{S}^*$ as best as possible, although, in the general case, optimality can not be guarantied. The image of $\mathcal{P}^*$ in objective space, $\mathcal{PF}^*$, is known as the local Pareto-optimal front.

2.1 Multi-objective estimation of distribution algorithms

EDAs are population-based optimization algorithms like other "traditional" evolutionary approaches. However, in EDAs, the variation step where the evolutionary operators are applied to the population is substituted by construction of a statistical model of the most promising subset of the population. This model is then sampled to produce new individuals that are merged with the parent population following a given substitution policy. Therefore, it has been stated that an additional benefit of EDAs is that not only do they return a solution to a problem, but a model representing the solutions is presented as well.

Multi-objective EDAs (MOEDAs) [9] are the extensions of EDAs to the multi-objective domain. Most MOEDAs are built by modifying an existing EDAs by substituting it fitness assignment function by one taken from an existing MOEA.

A very popular foundation for MOEDAs is a range of EDAs that builds the population model using a Bayesian network, leading to what has been called Bayesian optimization algorithms (BOA) [9].

Multi-objective real BOA (MrBOA) [10] is a multi-objective EDA based on a variant of single-objective BOA, in this case the real BOA (rBOA) [11]. RBOA performs a proper problem decomposition by means of a Bayesian factorization and probabilistic building-block crossover. To do this, it employs mixture models at the level of subproblems. MrBOA combines the fitness assignment of NSGA–II [1] with rBOA.

Another approach to modeling the subset with the best population elements is to apply a distribution mixture approach. Bosman and Thierens [12] proposed several variants of their multi-objective mixture-based iterated density estimation algorithm (MIDEA). They also proposed a novel Pareto-based and diversity-preserving fitness assignment function. MIDEA considered several types of prob-

abilistic models for both discrete and continuous problems. A mixture of univariate distributions and a mixture of tree distributions were used for discrete variables. A mixture of univariate Gaussian models and a mixture of multivariate Gaussian factorizations were applied for continuous variables. An adaptive clustering method was used to determine the capacity required to model a population.

The MIDEA family has been progressively improved. One of such enhancements is the introduction of the adaptive variance scaling (AVS) and the standard–deviation ratio (SDR) [13]. The AVS and SDR combination helps fight the early reduction of the mixture densities variances and therefore the premature convergence and diversity loss. Another important milestone has been the introduction of the anticipated mean shift (AMS) that takes into account the previous values of the means of the distribution to "push" solutions towards the Pareto-optimal front. AMS has been conjointly used with AVS in the multi-objective adapted maximum-likelihood Gaussian mixture model (MAMaLGaM-X) [14].

3   Outliers and Model-Building in MOEDAs

There are many data analysis or machine learning tasks in which a large number of variables are being recorded or sampled. One of the first steps towards obtaining a coherent analysis is the detection of outlaying observations. Detected data outliers are likely to be treated as abnormal data that may adversely lead to an invalid modelling, biased parameter estimation and incorrect results. Hence, it has always been of fundamental importance for practitioners to identify and handle with them prior to modeling and analysis.

However, an exact definition of an outlier often depends on hidden assumptions regarding the data structure and the applied detection method. Some intuitive definitions can be regarded general enough to cope with various types of data and methods. For example, it can be said that an outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism [15].

In spite of the intangibility of a proper definition of outliers, there are many real-world world problems, like fraud detection, weather prediction, network intrusion detection, anomaly detection, etc., that have prompted the development of different methods for outlier detection (see [16] for further reading). Different scientific communities define outliers differently, and such criteria can include outliers based on distribution, distance, density, etc. Therefore, several approaches have been proposed to deal with them.

The outliers issue is a good example of insufficient comprehension of the nature of the model-building problem. In machine-learning practice, outliers are handled as noisy, inconsistent or irrelevant data. Therefore, outlying data is expected to have little influence on the model or just to be disregarded.

However, that behavior is not adequate for model-building. In this case, is known beforehand that all elements in the data set should be take into account as they represent newly discovered or candidate regions of the search space and

therefore must be explored. Therefore, these instances should be at least equally represented by the model and perhaps even reinforced.

As model-building strategies varies from EDA to EDA, it is hard to back the previous statement with a general theoretical support. In order to do so, we must define an individual $\boldsymbol{z}_i$ as the pair representing values in decision and objective sets,

$$\boldsymbol{z}_i = \langle \boldsymbol{x}_i, \boldsymbol{F}(\boldsymbol{x}_i) \rangle \ . \tag{2}$$

In a simplified case, we can state that model building is an unsupervised machine learning problem with learning dataset,

$$\Psi = \{\boldsymbol{x}_i\} \ ; \forall \boldsymbol{z}_i = \langle \boldsymbol{x}_i, \boldsymbol{F}(\boldsymbol{x}_i) \rangle \in \hat{\mathcal{P}}_t \ , \tag{3}$$

where $\hat{\mathcal{P}}_t$ is the model-building dataset which is a subset of the algorithm population at iteration t.

A regular machine learning algorithm tunes the model $\mathfrak{M}(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\phi})$ by adjusting its topology $\boldsymbol{\theta}$ and parameters $\boldsymbol{\phi}$. In error-based learning this process involves the calculation of a set-wise error to which each element-wise error contribute to a different degree,

$$E_{\text{tot}} = \sum_{\boldsymbol{x}_i \in \Psi} E\left(\mathfrak{M}(\boldsymbol{x}_i, \boldsymbol{\theta}, \boldsymbol{\phi})\right) \ . \tag{4}$$

There are many different forms of the set-wise and element-wise errors, E_{tot} and $E(\cdot)$ respectively, but they can be formulated in a more or less similar fashion as above.

If E_{tot} is to be minimized, then $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ will be set in such way that the aggregation of element-wise contributions is as minimal as possible. As outliers, by their own definition, are rare and infrequent, their element-wise contribution to E_{tot} could be left to be relatively large as it is more convenient to focus on those that by being more popular, have a larger contribution to E_{tot}.

Therefore, model $\mathfrak{M}(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\phi})$ would end up representing more accurately elements more densely grouped than those isolated. However, as we already mentioned, in the model-building case, all elements of Ψ are important, and, perhaps, the isolated ones might be even more important than the clustered ones, as they represent locally optimal zones of the objective set that have not been properly explored.

Some EDAs have been proposed with the objective of dealing with the outliers issue and as an outlier-sensitive option. That is the case of the multi-objective neural EDA (MONEDA) [6] which uses a particular model-building growing neural gas (MB-GNG) [5]. MB-GNG relies on GNG, which has been shown to be an outlier-sensitive algorithm.

4 Measuring Outliers in MOEDAs

As already discussed earlier in this work, there are already some indirect experimental evidences that had lead to the conclusion that improper handling of outliers, loss of population diversity and low performance are related [4].

However, a comprehensive experiment is required to assess at what degree different model-building algorithms tend to disregard outliers.

For this task two elements are necessary: a shared EDA framework in that will be used to embed the different model-building algorithms and to define a measure of "outlierness" of population individuals.

4.1   Determining outliers using LOCI

The local correlation integral (LOCI) method [17], and its outlier metric, the multi-granularity deviation factor (MDEF), were proposed with the purpose of correctly dealing with multi-density and multi-granularity. That is why we have chosen to use MDEF as the outlier metric in this paper.

Let the r-neighborhood of an element x of $\mathcal{S}$ be the set of objects within distance r of x,

$$\mathcal{N}(x,r) = \{x' \in \mathcal{S} \,|\, \|x - x'\| \le r\}. \tag{5}$$

Then, having $n(x, \alpha r)$ and $n(x, r)$, that count the number of elements in the αr- and r-neighborhoods of x, denominated local and sampling neighborhoods, respectively, with $\alpha \in [0, 1]$. Relying of $n(\cdot)$ average over all elements x' in the r-neighborhood of x of their $n(x', \alpha r)$ is constructed as

$$\hat{n}(x, \alpha, r) = \frac{\sum_{x' \in \mathcal{N}(x,f)} n(x', \alpha r)}{n(x', r)} \tag{6}$$

The degree of 'outlierness' of a given element x of the dataset is computed as the *multi-granularity deviation factor* (MDEF) at radius $r \in \mathbb{R}$,

$$\mathrm{MDEF}(x, \alpha, r) = \frac{\hat{n}(x, \alpha, r) - n(x, \alpha r)}{\hat{n}(x, \alpha, r)}, \tag{7}$$

given $x \in \mathcal{S}$ and α,

The MDEF at radius r for a x is the relative deviation of its local neighborhood density from the average local neighborhood density in its r-neighborhood. Therefore, an element whose neighborhood density is equals the average local neighborhood density will have an MDEF of 0. On the other hand, outliers will have MDEFs larger than 0.

Following the recommendations and results from [17], we have set r as to make $n(x, r) \ge 20$ and $\alpha = 0.05$ in order to have sufficient data as to make MDEF statistically valid.

4.2   Shared EDA framework

In order to assess different model-building algorithms, a general EDA framework is necessary. The model-building algorithms will share this framework. Therefore, it will provide a testing ground common to all approaches and allows us to concentrate only on the topic of interest: measuring the degree at which each model-building algorithm is able to retain or loose outliers.

The shared EDA workflow maintains a population of individuals, $\mathcal{P}_t$, where t is the current iteration. It starts from a random initial population $\mathcal{P}_0$ of $n_{\max}$ individuals. A $\hat{\mathcal{P}}_t$ is constructed by determining the elements of $\mathcal{P}_t$ that produce the larger value of the hypervolume indicator, as in the HypE algorithm [18]. For problems of two and three objectives —like the ones used in our experiments— this task is carried out by calculating it exactly. For cases of more objectives the Monte Carlo sampling alternative is used, as it is more computational cost-effective.

A set $\hat{\mathcal{P}}_t$ containing the best $\lceil \alpha \, |\mathcal{P}_t| \rceil$ elements in terms of hypervolume contribution is extracted from the sorted version of $\mathcal{P}_t$.

The model builder under study is then trained using $\hat{\mathcal{P}}_t$ as training data set. An amount of $\lfloor \omega \, |\mathcal{P}_t| \rfloor$ new individuals are sampled from the model. Each one of these individuals substitute a randomly selected one from $\mathcal{P}_t \setminus \hat{\mathcal{P}}_t$, the section of the population not used for model–building. The set obtained is then united with the best elements, $\hat{\mathcal{P}}_t$, to form the population of the next iteration $\mathcal{P}_{t+1}$.

Iterations are repeated until the given stopping criterion is met. The output of the algorithm is the set of non-dominated solutions obtained from the final iteration, $\mathcal{P}_t^*$.

4.3   Experimental setup

In these experiments will be focusing of a set of problems previously proposed for the CEC 2009 MOP competition [19]. From the set of problems proposed there we selected the unconstrained optimization problems UF1 to UF6. These are two-objective problems that can be configured to have any desired number of variables —60 variables in our experiments. These problems are well-known for the complexity of their Pareto-optimal sets and fronts.

Five model-building algorithm —MrBOA [10], Naive MIDEA [12], SDR-AVS MIDEA [13], MAMaLGaM-X$^+$ [14] and MB-GNG [5]— are tested under this scheme. The HypE [18] hypervolume estimation algorithm for multi-objective optimization algorithm was also included in the analysis to provide a ground for comparison with 'traditional' evolutionary multi-objective algorithms. For the experiments we have used $\alpha = 0.3$ and $\omega = 0.3$ for all cases.

4.4   Results

Figures 1 and 2 summarize the average results of 30 runs for every algorithm and problem combination. From them it is noticeable that there is a substantial difference on MDEF scores between the 'pure' machine learning model-building algorithms, in particular MrBOA. Results from more advanced algorithms like SDR-AVS MIDEA and MAMaLGaM-X$^+$ yield a better performance and also better MDEF scores, something that supports their leitmotif. Finally, and in our opinion most importantly, it is observable how MB-GNG —an algorithm that was devised with the particular objective of solving the outliers issue— is not

Table 1: Results Mann-Whitney-Wilcoxon U tests comparing the MDEF values of HypE (H), MrBOA (M), naive MIDEA (N), SDR-AVS MIDEA (S), MAMaLGaM-X^{+} (A) and MB-GNG (G) as the optimization progressed. When the algorithm in the row yielded larger MDEF than one in the column is marked with a +, if smaller with a −. If there is not a significant difference in results a ∼ is used.

— Problem UF1 —

	t = 200					t = 500					t = end				
	M	N	S	A	G	M	N	S	A	G	M	N	S	A	G
H	+	+	+	+	∼	+	+	+	+	∼	+	+	∼	+	−
M	·	−	−	−	−	·	−	−	−	−	·	−	−	−	−
N		·	−	−	−		·	−	−	−		·	−	−	−
S			·	∼	−			·	+	−			·	+	−
A				·	−				·	−				·	−

— Problem UF2 —

	t = 200					t = 500					t = end				
	M	N	S	A	G	M	N	S	A	G	M	N	S	A	G
H	+	+	+	+	+	+	+	+	+	∼	+	+	+	+	∼
M	·	−	−	−	−	·	−	−	−	−	·	−	−	−	−
N		·	−	−	−		·	−	−	−		·	−	∼	−
S			·	∼	−			·	∼	−			·	∼	−
A				·	−				·	−				·	−

— Problem UF3 —

	t = 200					t = 500					t = end				
	M	N	S	A	G	M	N	S	A	G	M	N	S	A	G
H	+	+	+	+	∼	+	+	+	+	∼	+	+	+	+	∼
M	·	−	−	−	−	·	−	−	−	−	·	−	−	−	−
N		·	∼	∼	−		·	∼	∼	−		·	−	−	−
S			·	∼	−			·	∼	−			·	+	−
A				·	−				·	−				·	−

— Problem UF4 —

	t = 200					t = 500					t = end				
	M	N	S	A	G	M	N	S	A	G	M	N	S	A	G
H	+	+	+	+	∼	+	+	+	+	∼	+	+	+	+	−
M	·	−	−	−	−	·	−	−	−	−	·	−	−	−	−
N		·	−	−	−		·	−	−	−		·	−	−	−
S			·	∼	−			·	∼	−			·	∼	−
A				·	−				·	−				·	−

— Problem UF5 —

	t = 200					t = 500					t = end				
	M	N	S	A	G	M	N	S	A	G	M	N	S	A	G
H	+	+	+	+	∼	+	+	+	+	∼	+	+	+	+	∼
M	·	−	−	−	−	·	−	−	−	−	·	−	−	−	−
N		·	−	∼	−		·	−	∼	−		·	−	−	−
S			·	+	−			·	∼	−			·	+	−
A				·	−				·	−				·	−

— Problem UF6 —

	t = 200					t = 500					t = end				
	M	N	S	A	G	M	N	S	A	G	M	N	S	A	G
H	+	+	+	+	+	+	+	+	+	∼	+	+	+	+	−
M	·	∼	−	−	−	·	−	−	−	−	·	−	−	−	−
N		·	−	−	−		·	−	−	−		·	−	−	−
S			·	∼	−			·	∼	−			·	+	−
A				·	−				·	−				·	−

only capable of producing good results in terms of convergence to the Pareto-optimal front, but also is the algorithm that is able to maintain a higher diversity on its population.

The statistical validity of the judgment of these results calls for the application of statistical hypothesis tests. It has been previously remarked by different authors that the Mann-Whitney-Wilcoxon U test is particularly suited for experiments in the context of multi-objective evolutionary optimization [20]. This test is commonly used as a non-parametric method for testing equality of population medians. In our case we performed pair-wise tests on the significance of the difference of the indicator values yielded by the executions of the algorithms. A significance level, α, of 0.05 was used for all tests.

The results of these tests applied to the MDEF values yielded at the end of each run are summarized in Table 1. That table reflects when an algorithm produced significantly larger or smaller values of MDEF and when the results are indistinguishable from one another. This condensed representation prompts

some interesting and relevant conclusions. First, and most importantly, MG-GNG was the only algorithm capable to maintain a population diversity similar to that of a regular evolutionary algorithm, and correspondingly, larger than the rest of the algorithms. The rest of the algorithms consistently exhibited a lower diversity with regard to the EA and to MB-GNG. It is also very interesting how SDR-AVS MIDEA and MAMaLGaM-X frequently had similar MDEF scores.

5   Final Remarks

In this work we have taken the first steps towards the understanding and verification of the nature of the model-building problem of MOEDAs with an especial emphasis on the outliers issue. We have found that approaches based on traditional machine-learning techniques performed worst than better than less robust ones, a fact that we have shown to be related to their inability to maintain a proper population diversity. Although the outliers and diversity loss issues have been receiving some attention in recent years, this paper is the first study that exposes at what degree its situation is actually taking place, how it might impact the quality of results and how different algorithms behave with regard to it.

However, in order to gain a better comprehension more experiments are necessary. In one hand, different test problems must be addressed to realize if the results obtained here can be generalized. On the other hand, it is also of interest to further scale the problems to more objective functions. The analysis of the behavior of the algorithms in those situations might lead to their adaptation to the problem.

The experiences gained here can be used to sketch the requirements for a new model-building algorithm capable of inducing a quantum leap in the performance of MOEDAs and EDAs, for that matter.

References

1. Coello Coello, C.A., Lamont, G.B., Van Veldhuizen, D.A.: Evolutionary Algorithms for Solving Multi-Objective Problems. Second edn. Genetic and Evolutionary Computation. Springer, New York (2007)
2. Corne, D.W.: Single objective = past, multiobjective = present, ??? = future. In Michalewicz, Z., ed.: 2008 IEEE Conference on Evolutionary Computation (CEC), part of 2008 IEEE World Congress on Computational Intelligence (WCCI 2008), Piscataway, New Jersey, IEEE Press (2008)
3. Lozano, J.A., Larrañaga, P., Inza, I., Bengoetxea, E., eds.: Towards a New Evolutionary Computation: Advances on Estimation of Distribution Algorithms. Springer–Verlag (2006)
4. Martí, L.: Scalable Multi-Objective Optimization. PhD thesis, Departmento de Informática, Universidad Carlos III de Madrid, Colmenarejo, Spain (2011)
5. Martí, L., García, J., Berlanga, A., Coello Coello, C.A., Molina, J.M.: MB-GNG: Addressing drawbacks in multi-objective optimization estimation of distribution algorithms. Operations Research Letters **39**(2) (2011) 150–154

6. Martí, L., García, J., Berlanga, A., Molina, J.M.: Introducing MONEDA: Scalable multiobjective optimization with a neural estimation of distribution algorithm. In: GECCO '08: 10th Annual Conference on Genetic and Evolutionary Computation, New York, NY, USA, ACM Press (2008) 689–696

7. Martí, L., García, J., Berlanga, A., Molina, J.M.: Multi-objective optimization with an adaptive resonance theory-based estimation of distribution algorithm. Annals of Mathematics and Artificial Intelligence **68**(4) (2013) 247–273

8. Branke, J., Miettinen, K., Deb, K., Słowiǹski, R., eds.: Multiobjective Optimization. Volume 5252 of Lecture Notes in Computer Science. Springer–Verlag, Berlin/Heidelberg (2008)

9. Pelikan, M., Sastry, K., Goldberg, D.E.: Multiobjective estimation of distribution algorithms. In Pelikan, M., Sastry, K., Cantú-Paz, E., eds.: Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications. Studies in Computational Intelligence. Springer–Verlag (2006) 223–248

10. Ahn, C.W.: Advances in Evolutionary Algorithms. Theory, Design and Practice. Springer (2006) ISBN 3-540-31758-9.

11. Ahn, C.W., Goldberg, D.E., Ramakrishna, R.S.: Real-coded Bayesian optimization algorithm: Bringing the strength of BOA into the continuous world. In: 2004 Genetic and Evolutionary Computation (GECCO 2004). Volume 3102 of Lecture Notes in Computer Science. Springer (2004) 840–851

12. Bosman, P.A.N., Thierens, D.: The naïve MIDEA: A baseline multi–objective EA. In Coello Coello, C.A., Hernández Aguirre, A., Zitzler, E., eds.: Evolutionary Multi-Criterion Optimization. Third International Conference, EMO 2005, Guanajuato, México, Springer. Lecture Notes in Computer Science Vol. 3410 (March 2005) 428–442

13. Bosman, P.A., Thierens, D.: Adaptive variance scaling in continuous multi-objective estimation-of-distribution algorithms. In: Proceedings of the 9th annual conference on Genetic and evolutionary computation - GECCO '07, New York, New York, USA, ACM Press (2007) 500

14. Bosman, P.A.N.: The anticipated mean shift and cluster registration in mixture-based EDAs for multi-objective optimization. In: Proceedings of the 12th annual conference on Genetic and evolutionary computation - GECCO '10, New York, New York, USA, ACM Press (2010) 351

15. Hawkins, D.: Identification of Outliers. Chapman and Hall (1980)

16. Hodge, V.: A survey of outlier detection methodologies. Artificial Intelligence Review (2004) 1–43

17. Papadimitriou, S., Kitagawa, H., Gibbons, P., Faloutsos, C.: LOCI: Fast outlier detection using the local correlation integral. In: Proceedings 19th International Conference on Data Engineering (ICDE'03), IEEE Press (2003) 315–326

18. Bader, J.: Hypervolume-Based Search for Multiobjective Optimization: Theory and Methods. PhD thesis, ETH Zurich, Switzerland (2010)

19. Zhang, Q., Zhou, A., Zhao, S., Suganthan, P., Liu, W., Tiwari, S.: Multiobjective optimization test instances for the CEC 2009 special session and competition. Technical report, University of Essex, Colchester, UK and Nanyang Technological University, Singapore (2009)

20. Knowles, J., Thiele, L., Zitzler, E.: A tutorial on the performance assessment of stochastic multiobjective optimizers. TIK Report 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich (2006)
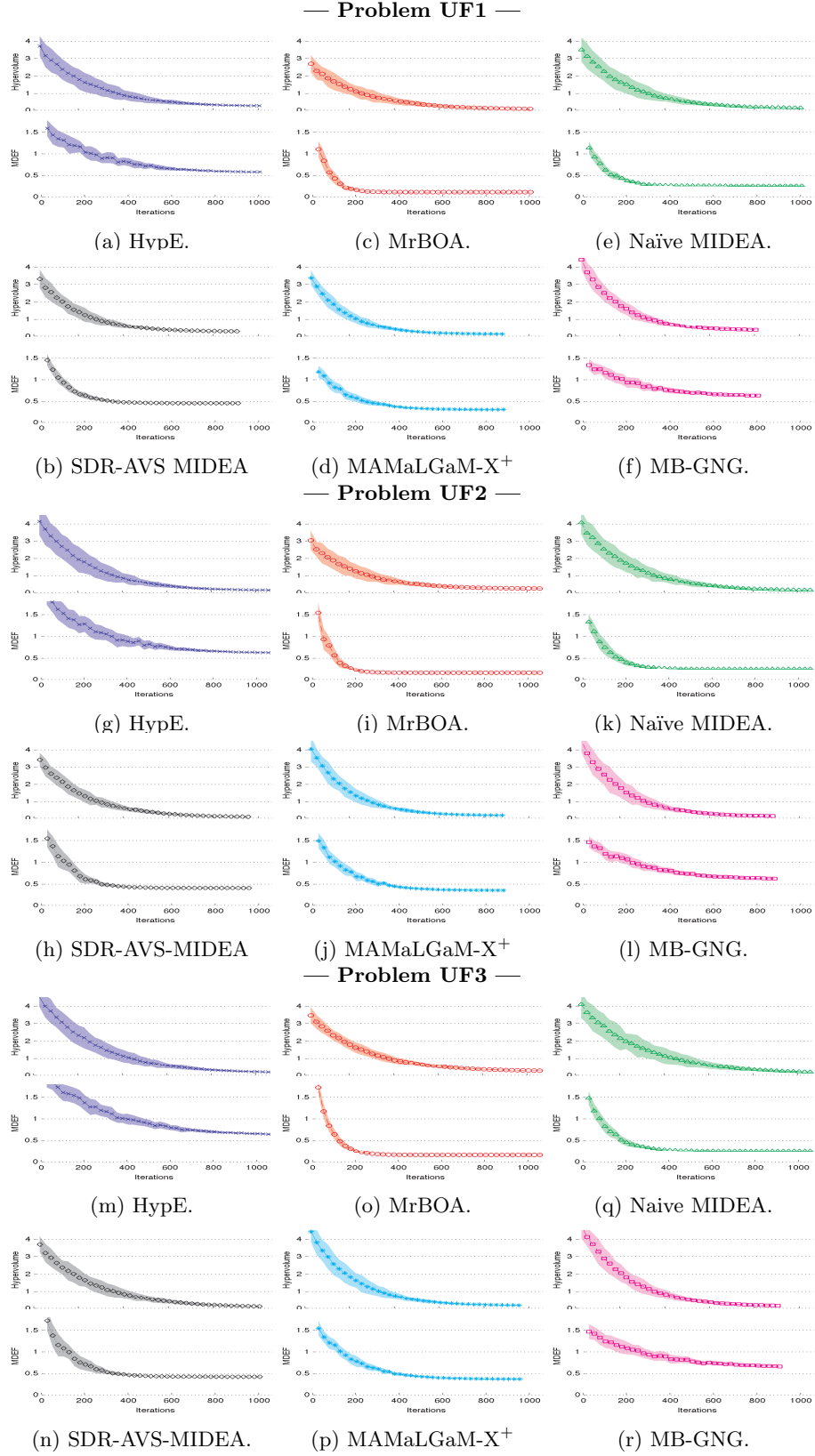
— Problem UF1 —

(a) HypE.

(c) MrBOA.

(e) Naïve MIDEA.

(b) SDR-AVS MIDEA

(d) MAMaLGaM-X$^+$

(f) MB-GNG.

— Problem UF2 —

(g) HypE.

(i) MrBOA.

(k) Naïve MIDEA.

(h) SDR-AVS-MIDEA

(j) MAMaLGaM-X$^+$

(l) MB-GNG.

— Problem UF3 —

(m) HypE.

(o) MrBOA.

(q) Naive MIDEA.

(n) SDR-AVS-MIDEA.

(p) MAMaLGaM-X$^+$

(r) MB-GNG.

Fig. 1: Progress, expressed as the relative hypervolume indicator, and mean out-liers index as calculated by MDEF for the different algorithms under study when solving problems UF1, UF2 and UF3.

— **Problem UF4** —

(a) HypE.

(c) MrBOA.

(e) Naive MIDEA.

(b) SDR-AVS-MIDEA

(d) MAMaLGaM-X$^+$

— **Problem UF5** —

(f) MB-GNG.

(g) HypE.

(i) MrBOA.

(k) Naive MIDEA.

(h) SDR-AVS-MIDEA

(j) MAMaLGaM-X$^+$

— **Problem UF6** —

(l) MB-GNG.

(m) HypE.

(o) MrBOA.

(q) Naive MIDEA.
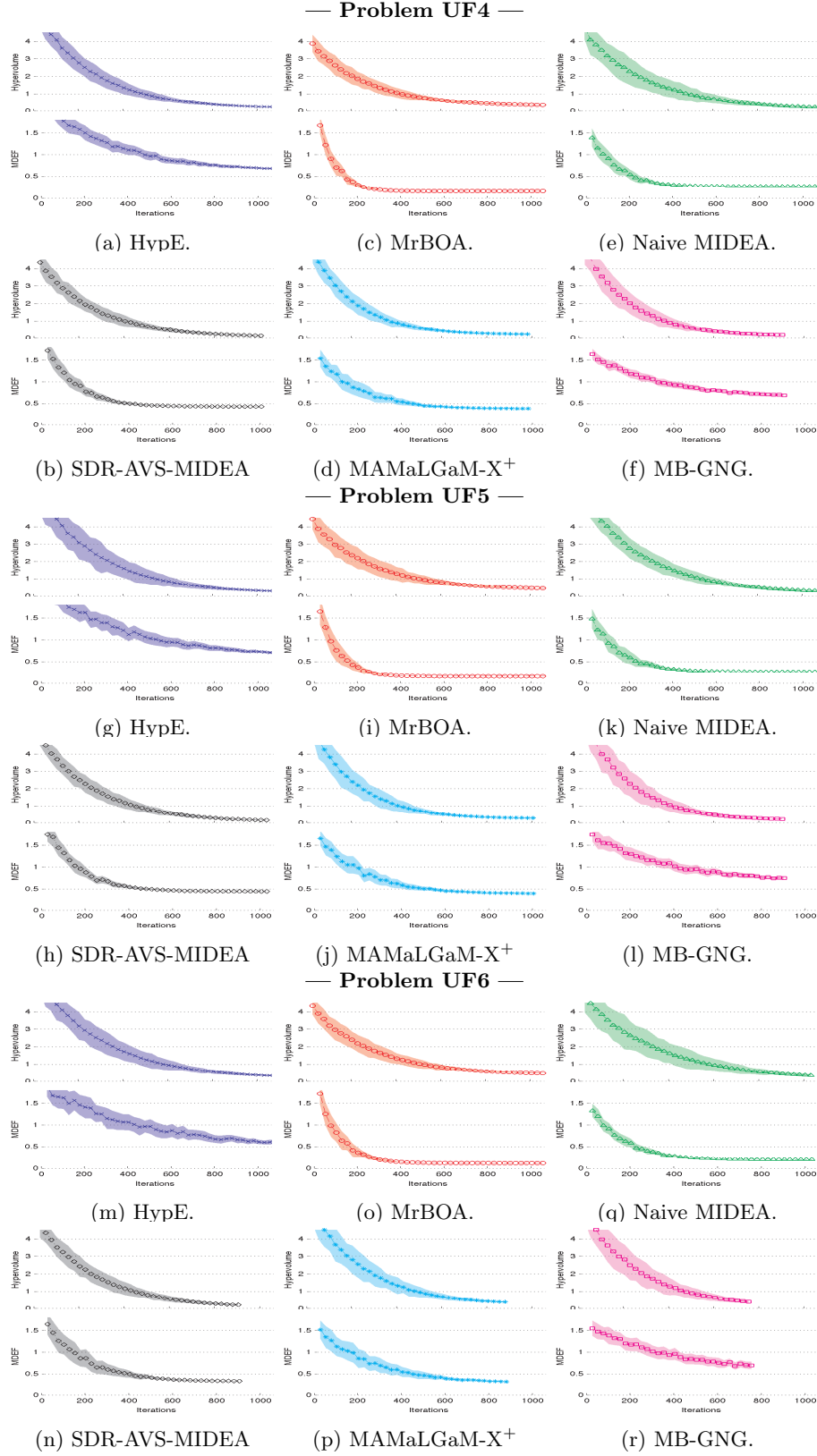
(n) SDR-AVS-MIDEA

(p) MAMaLGaM-X$^+$

(r) MB-GNG.

Fig. 2: Progress, expressed as the relative hypervolume indicator, and mean out-liers index as calculated by MDEF for the different algorithms in study when solving problem UF4, UF5 and UF6.