

GIAA TECHNICAL REPORT GIAA2010E001

# On Current Model–Building Methods for Multi–Objective Estimation of Distribution Algorithms: Shortcomings and Directions for Improvement

Luis Martí, Jesús García, Antonio Berlanga, Carlos A. Coello Coello, and José M. Molina

**Abstract**—There are some issues with multi–objective estimation of distribution algorithms (MOEDAs) that have been undermining their performance when dealing with problems with many objectives. In this paper we examine the model–building issue related to estimation of distribution algorithms (EDAs) and show that some of their, as yet overlooked, characteristics render most current MOEDAs unviable in the presence of many objectives. First, we present model–building as a problem with particular requirements and explain why some current approaches cannot properly deal with some of these conditions. Then, we discuss the strategies proposed for adapting EDAs to this problem. To validate our working hypothesis, we carry out an experimental study comparing different model–building algorithms. In the final part of the paper, we provide an in–depth discussion on viable alternatives to overcome the limitations of current MOEDAs in many–objective optimization.

**Index Terms**—Estimation of distribution algorithms, multi–objective optimization, model–building algorithms, many–objective problems, diversity loss.

## I. INTRODUCTION

THE multi–objective optimization problem (MOP) can be expressed as the problem in which a set of objective functions  $f_1(\mathbf{x}), \dots, f_M(\mathbf{x})$  should be jointly optimized;

$$\min \mathbf{F}(\mathbf{x}) = \langle f_1(\mathbf{x}), \dots, f_M(\mathbf{x}) \rangle, \mathbf{x} \in \mathcal{D}, \quad (1)$$

where  $\mathcal{D}$  is known as the decision space. The image set,  $\mathcal{O}$ , resulting from the projection  $\mathbf{F} : \mathcal{D} \rightarrow \mathcal{O}$  is called the objective space.

In this class of problems the optimizer must find one or more feasible solutions that jointly minimize (or maximize) the objective functions. Therefore, the solution to this type of problem is a set of trade–off points. The adequacy of a solution can be expressed in terms of the Pareto dominance relation [1]. The solution of (1) is the Pareto–optimal set,  $\mathcal{D}^*$ . This is the subset of  $\mathcal{D}$  that contains elements that are not

dominated by other elements of  $\mathcal{D}$ . Its image in the objective space is called the Pareto–optimal front,  $\mathcal{O}^*$ .

A broad range of approaches have been used to address MOPs [2], [3]. Of these, multi–objective evolutionary algorithms (MOEAs) have been found to be a very competitive approach in a wide variety of application domains. Their main advantages are ease of use and lower susceptibility (compared with traditional mathematical programming techniques for multi–objective optimization [2]) to the shape or continuity of the Pareto front.

There is a class of MOPs that are particularly appealing because of their inherent complexity: the so–called many–objective problems [4]. These are problems with a relatively large number of objectives (normally, four or more). Although somewhat counterintuitive and hard to visualize for a human decision maker, these problems are not uncommon in real–life engineering practice, such as, for example, aircraft design [5], land use planning [6], optimization of trackers for air traffic management and surveillance systems [7], [8], bridge design [9] and optical lens design, among others (see [10] for a survey on these problems).

The poor scalability of traditional MOEAs in these problems has triggered a sizeable amount of research, aiming to provide alternative approaches that can properly handle many–objective problems and perform reasonably. Estimation of distribution algorithms (EDAs) are one such approach [11]–[15]. EDAs have been hailed as a paradigm shift in evolutionary computation. They propose an alternative approach that creates a model of the population instead of applying evolutionary operators. This model is then used to synthesize new individuals. Probably because of their success in single–objective optimization, EDAs have been extended to the multi–objective optimization problem domain, leading to the so–called multi–objective EDAs (MOEDAs) [16].

Although MOEDAs have yielded some encouraging results, their introduction has not lived up to *a priori* expectations. This can be attributed to a number of different causes, some of which, although already present in single–objective EDAs, are more obvious in MOEDAs, whereas others are derived from some key components taken from traditional MOEAs. An analysis of this issue led us to distinguish a number of shortcomings, including: weaknesses derived from current multi–

Luis Martí, Jesús García, Antonio Berlanga and José M. Molina are with the Group of Applied Artificial Intelligence, Department of Informatics, Universidad Carlos III de Madrid.

Av. de la Universidad Carlos III, 22. Colmenarejo 28270 Madrid. Spain. <http://www.giaa.inf.uc3m.es>

Carlos A. Coello Coello is with the Department of Computer Science, CINVESTAV-IPN, Av. IPN No. 2508, Col. San Pedro Zacatenco, México, D.F. 07360, Mexico. email: [ccoello@cs.cinvestav.mx](mailto:ccoello@cs.cinvestav.mx)

objective fitness assignment strategies; the incorrect treatment of population outliers; the loss of population diversity, and too much computational effort being spent on finding an optimal population model.

Whereas the first issue is shared with other multi-objective evolutionary approaches, the others are peculiar to MOEDAs. A number of works have dealt with the last three issues listed above, particularly with loss of diversity. Nevertheless, the community has failed to acknowledge that the underlying cause for all those problems could, perhaps, be traced back to the algorithms used for model-building in EDAs.

In this paper we examine the model-building issue of EDAs and show that some its characteristics, which have been disregarded so far, render most current approaches unviable. This analysis includes a theoretical discussion of the issue, as well as an experimental study introducing a candidate solution, as well as some guidelines for addressing this problem.

It is hard to gain a rigorous understanding of the state of the art in MOEDA model building since each model builder is embedded within a different MOEDA framework. In order to comprehend the advantages and shortcomings of each algorithm, then, they should be tested under similar conditions and separated from their corresponding MOEDA. For this reason, we assess, in this paper, some of the main machine learning algorithms currently used or suitable for model-building in a controlled environment and under identical conditions. We propose a general MOEDA framework in which each model-building algorithm will be embedded. This framework guarantees the direct comparison of the algorithms and allows a proper validation of their performance.

The main contributions of this paper can be summarized as follows:

- A presentation of model-building as a problem with particular requirements and an overview of the reasons why some current approaches cannot properly deal with these requirements.
- A discussion of the strategies proposed for adapting current approaches to the problem.
- An experimental study that compares different model-building algorithms aimed at demonstrating our working hypothesis.
- An in-depth discussion of viable alternatives to overcome this issue.

The remainder of this paper is organized as follows. In Section II, we provide an introduction to MOEDAs. Then, in Section III, we deal with the model-building problem, its properties and how it has been approached by the main MOEDAs now in use. In Section IV, we describe the model-building algorithms under analysis and the MOEDA framework we propose for our empirical tests. Then, a set of experiments are performed, using community-accepted, complex and scalable test problems with a gradual increase in the number of objective functions. In Section V we put forward a set of guidelines derived from the previous discussions and experiments that could be used for overcoming the current situation and could lead to the formulation of “second generation” model builders. Note that the results of this research, although focused on multi-objective optimization problems, can be extrapolated to

single-objective EDAs. Finally, in Section VI, we put forward some concluding remarks and lines for future work.

## II. MULTI-OBJECTIVE ESTIMATION OF DISTRIBUTION ALGORITHMS

Estimation of distribution algorithms (EDAs) are population-based optimization algorithms. Instead of applying evolutionary operators to the population like other evolutionary approaches, EDAs build 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 original population following a given substitution policy. Because of this model-building feature, EDAs have also been called probabilistic-model-building genetic algorithms (PMBGAs) [17], [18]. Iterated density estimation evolutionary algorithms (IDEAs) introduced a similar framework to EDAs [19].

The introduction of machine learning techniques implies that these new algorithms lose the straightforward biological inspiration of their predecessors. Nonetheless, they gain the capacity of scalably solving many challenging problems, in some cases significantly outperforming standard EAs and other optimization techniques.

Model-building processes have evolved, too. Early approaches assumed that the different features of the decision variable space were independent. Subsequent methods started to deal with interactions among the decision variables, first in pair-wise fashion and later in a generalized manner, using  $n$ -ary dependencies.

Multi-objective EDAs (MOEDAs) [16] are the extensions of EDAs to the multi-objective domain. Most MOEDAs consist of a modification of existing EDAs whose fitness assignment function is substituted by one taken from an existing MOEA.

Most MOEDAs can be grouped in terms of their model-building algorithm. We will now give a brief description of MOEDAs, as this discussion is essential for our analysis. Note, however, that a comprehensive survey of current MOEDAs is beyond the scope of this paper.

### A. Graphical algorithm MOEDAs

One of the most common foundations for MOEDAs is a set of single-objective EDAs that build the population model using graphical models [20]. Most single-objective EDAs in that class rely on Bayesian networks [21]. This is the case of the Bayesian optimization algorithm (BOA) [22], the estimation of Bayesian network algorithm (EBNA) [23] and the learning factorized distribution algorithm (LFDA) [24]. Of these, BOA was the algorithm extrapolated to the multi-objective domain.

A Bayesian network is a probabilistic graphical model that represents a set of variables and their probabilistic (in)dependencies. They are directed acyclic graphs whose nodes represent variables, and whose arcs encode conditional independencies between the variables. Nodes can represent any kind of variable; either a measured parameter, a latent variable or a hypothesis.

The exhaustive synthesis of a Bayesian network from the algorithm's population is an NP-hard problem. Therefore, the intention behind the former approaches is to provide heuristics for building a network of reasonable computational complexity. BOA uses the so-called K2 metric, based on the Bayesian Dirichlet metric [25], to assess the quality of a network. A simple greedy algorithm is used to add edges in each iteration.

BOA-based MOEDAs combine the Bayesian model-building scheme with an already existing Pareto-based fitness assignment. This is the case of the multi-objective BOA (mBOA) [26] that exploits the fitness assignment used in NSGA-II. Another algorithm based on hierarchical BOA (hBOA) [27]–[29], called mhBOA [30], [31], also uses the same form of fitness assignment but introduces clustering in the objective function space. A similar idea is proposed in [32], [33], where the mixed BOA (mBOA) [34] is combined with the SPEA2 selection scheme to form the multi-objective mBOA (mmBOA).

The multi-objective real BOA (MrBOA) [35] also extends a preexisting EDA, namely, the real BOA (rBOA) [36]. RBOA performs a proper problem decomposition by means of a Bayesian factorization and probabilistic building-block crossover by employing mixture models at the level of sub-problems. MrBOA combines the fitness assignment of NSGA-II with rBOA.

For the following experiments we followed the model-building strategy used by rBOA [36], that is, apply a simple incremental greedy approach to construct the network. It adds edges to an initially fully disconnected graph. Each edge is added in order to improve, at each step, a particular formulation of the Bayesian information criterion (BIC) [37]. Then, the conditional probabilities that take part of the Bayesian factorization are computed for each disconnected subgraph.

Note, finally, that Bayesian networks are not the only graphical model suitable for model-building. Other approaches, in particular Markov random fields [38], have also been applied in single-objective EDAs [39]–[41]. To the best of our knowledge, however, these approaches have not yet been extended to multi-objective problems.

### B. Mixture distribution MOEDAs

Another approach to modeling the subset with the best population elements is to apply a distribution mixture approach. In a series of papers, Bosman and Thierens [42]–[47] proposed several variants of their multi-objective mixture-based iterated density estimation algorithm (MIDEA). They are based on their IDEA framework. Bosman and Thierens proposed a novel Pareto-based and diversity-preserving fitness assignment function. The model construction is inherited from the single-objective version. The proposed MIDEAs considered several types of probabilistic 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.

MIDEAs do not place any constraints on the location of the centers of the distributions. Consequently, the MIDEA clustering mechanism does not provide a specific mechanism to ensure equal coverage of the Pareto-optimal front if the number of representatives in some parts of the front is much larger than the number of representatives in some other parts.

The clustering algorithms applied for this task include the randomized leader algorithm [48], the  $k$ -means algorithm [49] and the expectation maximization algorithm [50].

The leader algorithm [48] is a fast and simple partitioning algorithm that was first used in the EDA context as part of the IDEA framework. Its use is particularly appropriate in situations where the overhead introduced by the clustering algorithm must remain as low as possible. Besides its small computational footprint, this algorithm has the additional advantage of not having to explicitly specify in advance how many partitions should be discovered. On the other hand, the drawbacks of the leader algorithm are that it is very sensitive to the ordering of the samples and that the values of its thresholds must be guessed *a priori* and are problem dependent.

The algorithm goes over the data set exactly once. The distances from each sample to each of the cluster centroids are determined. Then, the cluster whose distance is smallest and below a given distance threshold,  $\rho_{Ld}$ , is selected. If no such cluster can be found, a new one is created, containing just this sample. Once the number of samples in a cluster has exceeded the sample count threshold  $\rho_{Lc}$ , the leader is substituted by the mean of the cluster members. The mean of a cluster changes whenever a sample is added to that cluster. After clustering, a Gaussian mixture is constructed, as described for the naïve MIDEA [47]. This way the model can be sampled in order to produce new elements.

The  $k$ -means algorithm [49] is a well-known machine learning method. It constructs  $k$  partitions of the input space. To do this, it uses partition centroids. First, the  $k$  centroids are initialized from randomly selected samples. At each iteration, each sample is assigned to the nearest partition based on the distance to the partition centroid. Once all of the points have been assigned, the means of the partitions are updated. The algorithm iterates until the centroids no longer change significantly. An important issue in this algorithm is how to set parameter  $k$  such that partitioning is adequate. Parameter setting requires some experience. In the context of MIDEAs [51] the approach followed is to increment  $k$  and calculate the negative log-likelihood of the mixture probability distribution after estimating a factorized probability distribution in each cluster. If the resulting mixture probability distribution is significantly better than for a smaller value of  $k$ , this value is accepted and the search continues. As in the previous case, after the clusters are determined, a Gaussian mixture is estimated for sampling purposes.

The expectation maximization (EM) algorithm [50] is an iterative approach to computing a maximum likelihood estimate. EM uses the difference in the negative log-likelihood of the estimated probability distribution between subsequent iterations in order to derive the hidden parameters. In a clustering context, EM is used to get an approximation of the maximum likelihood estimation of a mixture probability distribution. The

number of components in the mixture probability distribution is usually chosen beforehand. This choice is similar to the choice of the number of partitions when using a clustering approach to the estimation of a mixture probability distribution from data. In this case a similar approach to the one discussed for  $k$ -means is applied.

MIDEAs are not the only mixture-based algorithms. The multi-objective Parzen EDA (MOPED) [52], [53] puts forward a similar mixture-based approach. MOPED uses the NSGA-II ranking method and the Parzen estimator [54] to approximate the probability density of solutions lying on the Pareto front. The proposed algorithm has been applied to different types of test case problems, and results show a good performance of the overall optimization procedure in terms of the total number of objective function evaluations.

The multi-objective neural EDA (MONEDA) [55] is also a mixture-based MOEDA. It was devised to deal with the model-building issue that will be discussed in Section V. It is based on a modified growing neural gas (GNG) network [56]. GNG networks have been previously presented as good candidates for dealing with the model-building issue [55], [57] because of their known sensitivity to outliers [58].

GNG networks are unsupervised intrinsic self-organizing neural networks based on the neural gas model [59]. The network grows to adapt itself automatically to the complexity of the dataset being modeled. It has a fast convergence to low distortion errors and these errors are better than those yielded by “standard” algorithms, such as  $k$ -means clustering, maximum-entropy clustering and Kohonen’s self-organizing feature maps [59].

We put forward the model-building growing neural gas (MB-GNG) [55] network with the aim of adapting GNG to the model-building task. In particular, MB-GNG incorporates a cluster repulsion term to GNG’s adaptation rule that promotes search and diversity.

### C. Covariance matrix adaptation evolution strategies

Covariance matrix adaptation evolution strategies (CMA-ES) [60], [61] have been shown to yield many outstanding results in comparative studies [62]–[64]. CMA-ES consists of a method for updating the covariance matrix of the multivariate normal mutation distribution used in an evolution strategy [65]. They can be viewed as an EDA, as new individuals are sampled according to the mutation distribution. The covariance matrix describes the pairwise dependencies between the variables in the distribution. Adaptation of the covariance matrix is equivalent to learning a second-order model of the underlying objective function. CMA-ES has been extrapolated to the multi-objective domain [66].

### D. Other approaches

Other MOEDAs have been proposed in order to take advantage of the mathematical properties of the Pareto-optimal front. For example, the regularity model-based multi-objective estimation of distribution algorithm (RM-MEDA) [67], [68] is based on the regularity property derived from the Karush–Kuhn–Tucker condition. This means that, subject to certain

constraints, the Pareto-optimal set,  $\mathcal{D}^*$ , of a continuous multi-objective optimization problem can be induced to be a piecewise continuous  $(M - 1)$ -dimensional manifold, where  $M$  is the number of objectives [2], [69].

At each iteration, RM-MEDA models the promising area of the decision space using a probability distribution whose centroid is a  $(M - 1)$ -dimensional piecewise continuous manifold. The local principal component analysis algorithm [70] is used to build this model. New trial solutions are sampled from the model thus built. Again, this model adopts the fitness assignment mechanism proposed by NSGA-II. The main drawback of this algorithm is its high computational complexity. This is an obstacle to its application in problems with many objective functions.

## III. UNDERSTANDING MODEL-BUILDING IN THE MULTI-OBJECTIVE CASE

Regardless of the many efforts at providing usable model-building methods for EDAs, the nature of the problem itself has received relatively little attention. In spite of the succession of gradually improving results of EDAs, one question hangs over the search for possibilities for further improvement. Would current statistically sound and robust approaches be valid for the problem being addressed? Or, in other words, does the model-building problem have particular demands that can only be met by custom-made algorithms? Machine learning and statistical algorithms, although suitable for their original purpose, might not be that effective in the particular case of model building.

Generally, such algorithms are off-the-shelf machine learning methods that were originally intended for other classes of problems. On the other hand, the model-building problem has particular requirements that the above methods do not meet and may even go against. Furthermore, the consequences of this misunderstanding would be more dramatic when scaling up the number of objectives, since the situation is made worse by the implications of the curse of dimensionality [71].

In this paper we argue that the model-building problem has not been properly identified. For this reason, it has been treated like other previously existing problems overlooking that fact that this problem has particular requirements. This matter did not show up as clearly in single-objective EDAs. Thanks to the extension to the multi-objective domain this issue has become more evident, as we will debate in the remainder of this section.

An analysis of the results yielded by current multi-objective EDAs and their scalability against the number of objectives leads to the identification of some issues that could be preventing MOEDAs from getting substantially better results than other evolutionary approaches. Such issues include:

- 1) drawbacks derived from current MOEA fitness assignment strategies;
- 2) incorrect treatment of data outliers;
- 3) loss of population diversity; and
- 4) excess of computational effort devoted to finding an optimal population model.

The first issue is shared by MOEAs, MOEDAs and other multi-objective evolutionary approaches. It has been shown

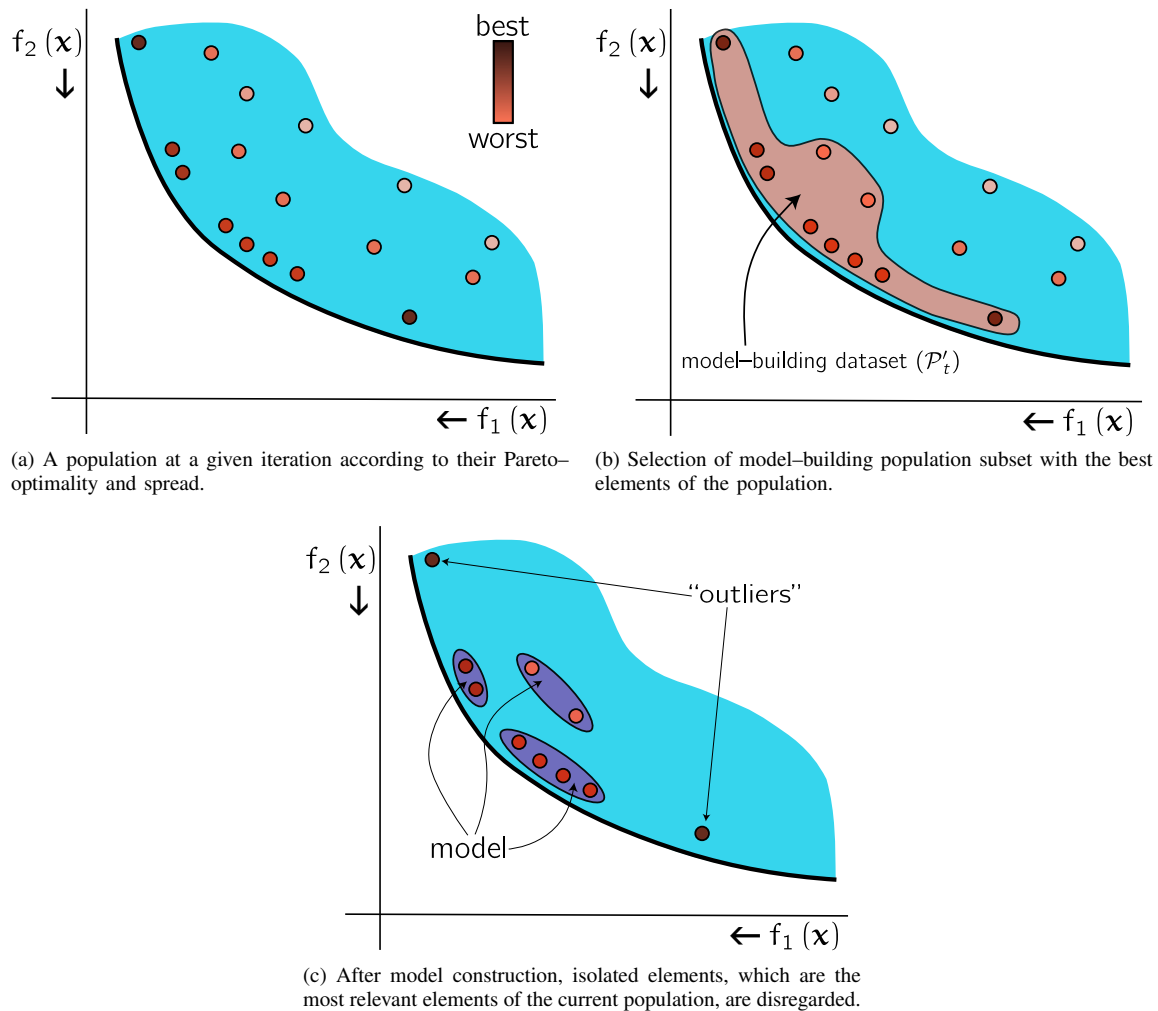


Figure 1: A graphical example of how standard model-building algorithms fail to take into account outliers.

that, as the number of objectives grow, the fitness assignment performance starts to degrade as an exponential increase of the population size is required [4], [72]–[74]. Some alternative approaches have been proposed to deal with this problem, including objective reduction [75]–[78], performance indicator-based fitness assignment [79]–[84], and hybrid methods [85], [86]. This topic is an open research area, which is currently very active within the evolutionary multi-objective optimization community.

The remaining three issues have to do only with EDAs and are the main focus of this work. These issues can be traced back to the single-objective predecessor of most MOEDAs and its respective model-building algorithms. The data outliers issue is a good example of the defective understanding 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 it is just disregarded. However, this behavior is not appropriate for model-building. In this case, it is known beforehand that all elements in the data set should be taken 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. This situation is illustrated in Fig. 1. A model-building algorithm that primes outliers might actually speed up the search process and lower the rate of the exponential dimension–population size dependency.

Another weakness of most MOEDAs (and most EDAs, for that matter) is the loss of population diversity. This is a point that has already been made, and some proposals for addressing the issue have been laid out [87]–[89]. This loss of diversity can be traced back to the above outliers issue of model-building algorithms. The repetitive application of an algorithm that disregards outliers tends to generate more individuals in areas of the search space that are more densely represented. Although there have been some proposals to circumvent this problem, we take the view that the ultimate solution is the use of an adequate algorithm.

The third issue to be dealt with is the computational resources wasted on finding an optimal description for the subpopulation being modeled. In the model-building case, optimal model complexity can be sacrificed in the interests of a faster algorithm. This is because the only constraint

is to have a model that is sufficiently, but not necessarily optimally, complex to correctly represent the data. This is particularly true when dealing with high-dimensional MOPs, as, in these cases, there will be large amounts of data to be repeatedly processed at each iteration. Even so, most current approaches spend considerable effort on finding optimal model complexity, using minimum description length [90], structural risk minimization [91], Bayesian information criterion [37] or other similar heuristics, as explained in the previous section.

In conclusion, we can deduce that an understanding of the nature of the model-building problem and the application of suitable algorithms appear to point the way forward in this area.

#### IV. PROBLEM STATEMENT

To identify the model-building issue debated above, it is helpful to devise a comparative experiment that casts light on the performances of a selected set of model-building algorithms subject to the same conditions to deal with a group of complexity-scaling problems. In particular, we deal with a selection of the Walking Fish Group (WFG) continuous and scalable test problems set [92], [93].

A MOEDA framework is shared by the model-building algorithms involved in the tests in order to ensure the comparison and reproducibility of the results. Two well-known MOEAs, the non-dominated sorting genetic algorithm II (NSGA-II) [94] and the strength Pareto evolutionary algorithm (SPEA2) [95], were also applied as a baseline for the comparison.

The model-building algorithms involved in the tests were:

- Bayesian networks, as used in MrBOA;
- randomized leader algorithm,  $k$ -means algorithm and E-M algorithm, as described for MIDEAS;
- $(1 + \lambda)$ -CMA-ES as described in [66]; and
- GNG and its model-building version, MB-GNG.

This assortment of algorithms offers a broad sample of different approaches, ranging from the most statistically rigorous algorithms, such as Bayesian networks, E-M or CMA-ES, to others, like the leader algorithm and MB-GNG, that have some clear shortcomings in the context of their original application scope. Nevertheless, they can also be assumed to deal with outlying elements in a more adequate manner.

##### A. Shared MOEDA framework

A general MOEDA framework must be proposed in order to assess different model-building algorithms, in particular the algorithms described in Section II. The model-building algorithms will share this framework. Therefore, such a framework will provide a testing ground common to all approaches and we will be able to focus solely on the topic of interest.

Our general MOEDA workflow is similar to other previously existent algorithms, as illustrated in Fig. 2. It maintains a population of individuals,  $\mathcal{P}_t$ , where  $t$  is the current iteration. It starts with a random initial population  $\mathcal{P}_0$  of  $n_{\text{pop}}$  individuals. It then proceeds to sort the individuals using the NSGA-II fitness assignment function [94]. This fitness function was chosen because it is in widespread use, although we are aware

that better strategies, such as indicator-based options, would probably yield better results.

The fitness function is used to rank individuals according their Pareto dominance relations. Individuals with the same domination rank are then compared using a local crowding distance. This distance favors individuals that are more isolated than those residing in crowded regions of the Pareto front.

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

$$|\hat{\mathcal{P}}_t| = \lceil \alpha |\mathcal{P}_t| \rceil. \quad (2)$$

Here  $\alpha$  is known as the selection percentile.

The model builder under study is then trained using  $\hat{\mathcal{P}}_t$  as the training data set. A set of  $\lfloor \omega |\mathcal{P}_t| \rfloor$  new individuals, which is regulated by the substitution percentile  $\omega$ , is sampled from the model. Each of these individuals substitutes an individual randomly selected from  $\mathcal{P}_t \setminus \hat{\mathcal{P}}_t$ , which is the section of the population not used for model-building. The output set is then united with the best elements,  $\hat{\mathcal{P}}_t$ , in order 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 from the final iteration,  $\mathcal{P}_t^*$ .

After some exploratory tests with our EDA, we settled for  $\alpha = 0.3$  and  $\omega = 0.3$ .

##### B. Experimental setup

The problems to be addressed are part of the Walking Fish Group problem toolkit (WFG) [96]. This is a toolkit for creating complex synthetic multi-objective test problems that can be devised to exhibit a given set of target features.

Unlike previous test suites where complexity is embedded in the problem, a test problem designer using the WFG toolkit has access to a series of components to control specific test problem features (e.g., separability, modality, etc.). The WFG toolkit was used to construct a suite of test problems that provides a thorough test for optimizers. This set of nine problems, WFG1-WFG9, are formulated in such manner that each poses a different type of challenge to multi-objective optimizers.

The WFG test suite exceeds the functionality of previous existing test suites. In particular, it includes a number of problems that exhibit properties not evident in other commonly used test suites such as the Deb-Thiele-Laumanns-Zitzler (DTLZ) [97] and the Zitzler-Deb-Thiele (ZDT) [98] test suites. These differences include: non-separable problems, deceptive problems, a truly degenerate problem, a mixed shape Pareto front problem, problems scalable by the number of position-related parameters, and problems with dependencies between position- and distance-related parameters. The WFG test suite provides a better form of assessing the performance of optimization algorithms on a wide range of different problems.

From the set of nine problems, the test functions WFG4 to WFG9 were selected because of the simple form of their Pareto-optimal fronts, which lie on the first orthant of a unit hypersphere. For this reason, the progress of the optimization process can be determined without having a sampled version

- 1: **Parameters:**  $n_{\text{pop}}$ ,  $\alpha$  and  $\omega$ .
- 2:  $t \leftarrow 0$ .
- 3: Randomly generate initial population,  $\mathcal{P}_0$ , with  $n_{\text{pop}}$  individuals.
- 4: **repeat**
- 5:   Sort  $\mathcal{P}_t$  individuals with regard to their fitness function values.
- 6:   Extract the first  $\lceil \alpha |\mathcal{P}_t| \rceil$  elements of the sorted  $\mathcal{P}_t$  to  $\hat{\mathcal{P}}_t$ .
- 7:   Build model of  $\hat{\mathcal{P}}_t$ .
- 8:   Sample  $\lfloor \omega |\mathcal{P}_t| \rfloor$  new individuals from the model.
- 9:   Substitute randomly selected individuals of  $\mathcal{P}_t \setminus \hat{\mathcal{P}}_t$  with the new individuals to produce  $\mathcal{P}'_t$ .
- 10:  $\mathcal{P}_{t+1} = \hat{\mathcal{P}}_t \cup \mathcal{P}'_t$ .
- 11:  $t \leftarrow t + 1$ .
- 12: **until** end condition met
- 13: Determine the set of non-dominated individuals of  $\mathcal{P}_t$ ,  $\mathcal{P}_t^*$ .
- 14: **return**  $\mathcal{P}_t^*$  as the algorithm's solution.

Figure 2: Algorithmic representation of the shared MOEDA.

of the Pareto-optimal front. In particular, we measure the similarity of the current non-dominated front,  $\mathcal{PF}_t^*$ , to the Pareto-optimal front as the mean distance of the elements of  $\mathcal{PF}_t^*$  to the origin of coordinates minus one,

$$I_{\text{prog}} = \frac{\sum_{\mathbf{x} \in \mathcal{F}_t^*} \left( \sum_{m=1}^M (f_m(\mathbf{x}) - 1)^2 \right)^{0.5}}{|\mathcal{F}^*|}. \quad (3)$$

For this reason, the local progress of the algorithms can be easily determined as executions taking place without having to turn to more computationally expensive options such as performance indicators.

Even so, assessing the progress of the algorithms in high dimensions is a complicated matter. To do this, we used the MGBM multi-objective optimization cumulative stopping criterion [99], [100]. This criterion combines the measurement of progress across iterations  $I_{\text{prog}}$  with a simplified Kalman filter that is used for the evidence-gathering process. This mechanism is able to gauge the progress of the optimization process at a low computational cost. This makes it suitable for solving complex or many-objective problems.

Performance indicators are required to gauge and compare the quality of the solutions yielded by each algorithm. In these experiments the binary hypervolume indicator [101] was used for performance assessment<sup>1</sup>. This indicator gauges how similar the solution yielded by each algorithm is to the Pareto-optimal front of the problem. Therefore, it requires an explicit sampling of that front, which is not viable in problems with many objectives. To address this issue, we took an approach similar to the method adopted by the purity performance indicator [102], [103]. A combined set  $\mathcal{PF}^+$  is defined as the union of the solutions obtained from the different algorithms across all the experiment executions.  $\tilde{\mathcal{O}}^*$  is then determined by extracting the non-dominated elements,

$$\mathbf{x} \in \tilde{\mathcal{O}}^* \text{ iff } \mathbf{x} \in \mathcal{PF}^+ \text{ and } \nexists \mathbf{y} \in \mathcal{PF}^+ \text{ such that } \mathbf{y} \prec \mathbf{x}. \quad (4)$$

Although this procedure circumvents the problems of performing a direct sampling of the Pareto-optimal front shape

function, special precautions should be taken when interpreting the results. Notice that the algorithm's performance will be measured with regard to the set of overall best solutions and not against the actual Pareto-optimal front. We consider this to be a valid approach, though, since the intention of these experiments is to compare the different model-building algorithms rather than actually solving the problems.

Each problem was configured with 3, 5, 7 and 9 objective functions. For all cases, the decision space dimension was set at 15. The experiments were carried out under the PISA experimental framework [104]. All the algorithms were executed 30 times for each problem/dimension pair.

Statistical hypothesis tests have to be applied to validate the results of different executions. Different frameworks for carrying out this task have been already discussed by other authors (see for example [101], [105], [106]).

In our case, we performed a Kruskal-Wallis test [107] with the indicator values yielded by each algorithm's run for each problem/dimension combination. In context of these experiments, the null hypothesis for the test was that all algorithms were equally capable of solving the problem. If the null hypothesis was rejected, which was the case in all the experimental instances, the Conover-Inman procedure [108, pp.288–290] was applied in a pairwise manner to determine if the results of one algorithm were significantly better than those of the other. A significance level,  $\alpha$ , of 0.05 was used for all the tests. A similar test framework had been previously applied to assess similar experiments [109], [110].

Besides measuring how good the solutions output by the algorithms are, it is also very important to analyze how long it takes the algorithms to reach the solutions. For these experiments we measured two variables: the number of objective function evaluations and the number of floating-point operations carried out by each model-building algorithm. This last measurement assumes that all floating-point operations have to do with the optimization process itself. This requirement can be easily met under experimental conditions. There are a number of profiling tools that are capable of tracking the number of floating-point operations that have taken place as part of a process. For this work, we chose the OProfile program

<sup>1</sup>For the values yielded by other indicators, see the web appendix of this paper at <http://www.giaa.inf.uc3m.es/miembros/lmarti/model-building>

profiling toolkit [111]. As the study also covered NSGA-II and SPEA2 and they do not perform model building, we measured the operations dedicated to the application of the evolutionary operators in their case.

### C. Results

As already explained, the scope of the experiments reported here is to validate or reject the hypotheses stated in Section III. For this reason, the performance of each algorithm is compared in terms of both the quality of the solutions that they generate and their cost in terms of computational resources. Particularly, we are concerned with the number of floating-point operations dedicated to the model-building task and with the number of function evaluations performed.

The first results have to do with the WFG4 problem. WFG4 is a separable and strongly multi-modal problem that, like the other problems, has a concave Pareto-optimal front. This front lies on the first orthant of a hypersphere of radius one located at the origin. The separability property should, in theory, allow Bayesian networks-based approaches to perform well, as already reported in [112].

Fig. 3 summarizes the outcome of the experiments related to this problem. These results show what will be a common characteristic of all the results presented here. In low dimensionality, in particular with  $M = 3$ , none of the models yielded substantially different results as illustrated in Fig. 3a. Better results could possibly be achieved by further tuning the parameters. However, this situation gradually changes as the number of objectives increases (see Figs. 3b–3d). In these cases, the least robust approaches (statistically speaking), such as the leader algorithm, GNG and MB-GNG, outperform the others in terms of approximation to the Pareto-optimal front, with the exception of the 7-objective case, where Bayesian networks outperform the other algorithms. This is a result that could be attributed to the fact that this is a separable problem. This outcome can be verified by looking at the statistical hypothesis test results shown in Fig. 3g.

Another illustrative analysis emerges when analyzing the mean number of floating-point operations and the number of function evaluations shown in Figs. 3e and 3f. Let us draw attention in the first figure to the fact that EM, Bayesian networks and CMA-ES consume far more resources and exhibit poorer scaling properties with regard to the other algorithms even with respect to the standard MOEAs used for a baseline comparison. The fact that such a rise in the computational demand of those algorithms did not lead to an increase in the number of function evaluations is even more interesting. Therefore, this increase in the computational cost was not caused by an increase in the amount of searching done; instead, it can be attributed to just the creation of the data models.

WFG5 is also a separable problem but it has a set of deceptive locally optimal fronts. This feature is meant to evaluate the capacity of the optimizers to avoid getting trapped in local optima. Fig. 4 shows the results for this problem. In spite of the hurdle of the multiple local optima, the results are quite consistent with those obtained for WFG4. The

scenario that differentiates the three-objective problem from the other dimensions is repeated here, save that CMA-ES is the algorithm that yields better solutions in the  $M = 7$  case. In the other two “high” dimensions, 5 and 9, MB-GNG is the algorithm that yields the best results. As in WFG4, if we contrast the floating-point operations and the objective function evaluations, it is clear that EM, Bayesian networks and CMA-ES required much more computational time to perform a similar level of search space exploration.

The next problem, WFG6, is a separable problem without the strong multi-modality of WFG4. Fig. 5 summarizes the comparative performances of the different algorithms when dealing with this problem. In this case, MB-GNG outperforms the other algorithms in terms of Pareto optimality in all the high-dimensional cases. It is also noticeable that Bayesian networks yield similar results to non-statistically rigorous algorithms. This can be attributed to problem separability. The pattern of floating-point operations and function evaluations relations already discussed in the previous problems is also present here.

The remaining three problems have the added difficulty of having a parameter-based bias. WFG7 is uni-modal and separable, like WFG4 and WFG6. Its results are reported in Fig. 6. In this case, GNG and MB-GNG outperform their peers in the problems with 5 and 7 objectives. However, Bayesian networks yielded better average results when tackling the problem with 9 objectives, although this improvement was not deemed as statistically significant.

WFG8 is a non-separable problem and its results are illustrated in Fig. 7. So far, this is the problem where non-rigorous algorithms most obviously outperformed the others with a more solid statistical foundation in the higher dimensionality (in objective function space). In the nine-objective case (Fig. 7d) there seems to be little difference among the results of the leader algorithm, CMA-ES, GNG and MB-GNG. However, the much higher cost of running CMA-ES than the other three approaches is much clearer from the results shown in Fig. 7e.

Finally, WFG9 is non-separable, multi-modal and has deceptive local-optima. These properties make WFG9 the hardest problem of all the problems chosen for the study. Fig. 8 shows the results obtained with the tested algorithms. As in the previous experiments, MB-GNG manages to yield the best results, in this case, sharing its success with the leader algorithm in the nine-objective case.

Looking at this relatively large set of results, even in the light of the most advantageous representation chosen, they are rather cumbersome. First of all, it is noticeable that there is no clear winner in the three-objective problems, where the different model-building algorithms alternately outperform each other. This changes as the number of objectives is increased. Noticeably, model-building approaches that rely on solid statistical foundations, such as Bayesian networks, EM, or CMA-ES are outperformed by the others without such properties. In terms of computational cost, we find that, while the overall number of function evaluations remained within similar ranges for the different algorithms, the effort expended on model building was far greater for EM, CMA-ES and the



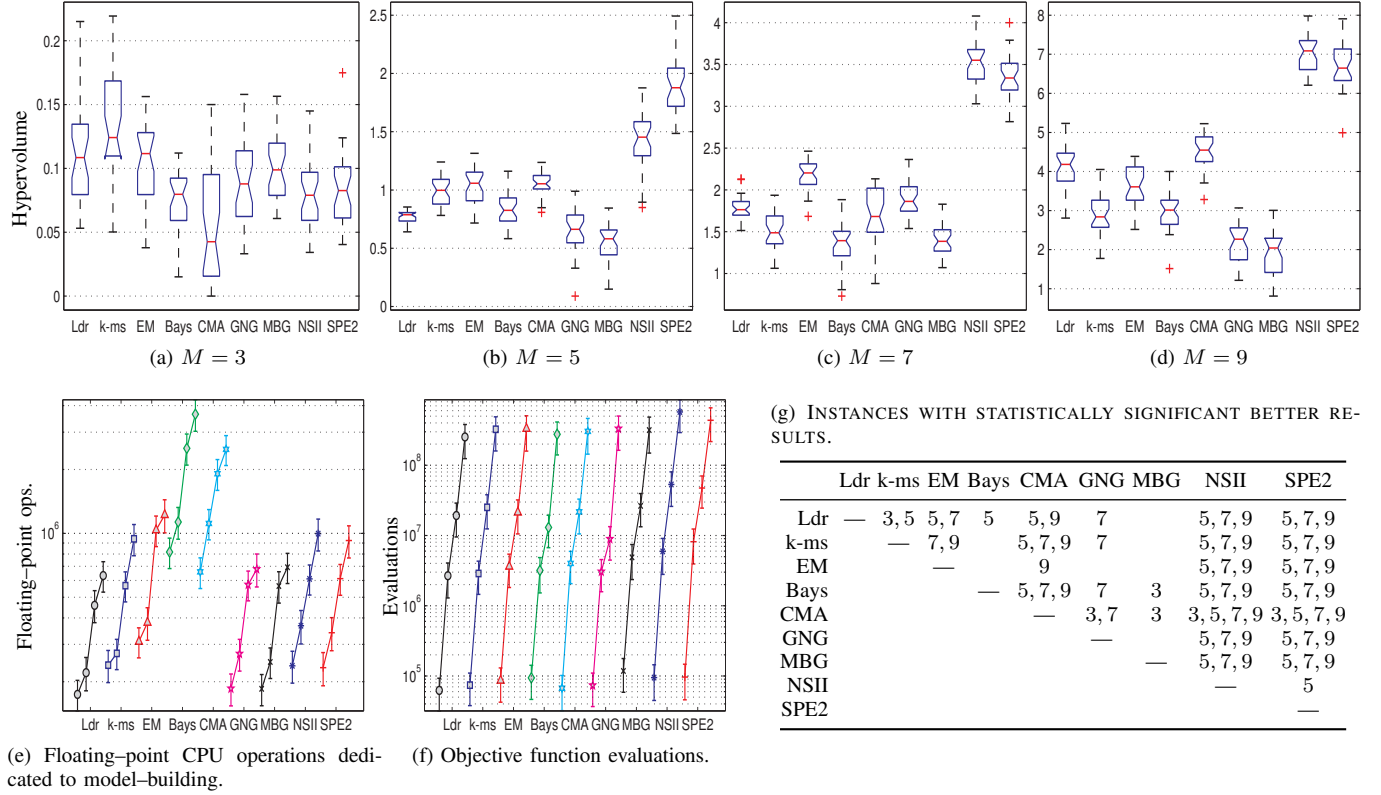


Figure 3: Results for problem WFG4 of applying for model-building the randomized leader algorithm (Ldr), the  $k$ -means (k-ms) algorithm, expectation maximization (EM), Bayesian networks (Bays), covariance matrix adaptation evolutionary strategy (CMA), growing neural gas network (GNG) and the model-building growing neural gas network (MBG). For comparison reasons NSGA-II (NSII) and SPEA2 (SPE2) evolutionary algorithms are also shown. Figs. (a)–(d) summarize the statistical description of the hypervolume values obtained after each experiment as box-plots. Fig. (e) shows the progression across problem dimensions of the floating-point operations used by the model-building algorithms, while Fig. (f) contain a similar representation but for the number of function evaluations. Table (g) summarizes the outcome of performing the statistical hypothesis tests. The numbers shown are the problem dimension where the test detected a statistically significant better indicator values of the algorithm in each row with respect of those in the columns.

Bayesian networks.

#### D. Analyzing the Results

It is not easy to assess these facts, as it implies cross-examining and comparing the results presented separately in Figs. 3–8. For this reason, we decided to adopt a more integrative representation along the lines of the schema proposed in [109], [110].

That is, for a given set of algorithms  $A_1, \dots, A_K$ , a set of  $P$  test problem instances  $\Phi_{1,m}, \dots, \Phi_{P,m}$ , configured with  $m$  objectives, the function  $\delta(\cdot)$  is defined as

$$\delta(A_i, A_j, \Phi_{p,m}) = \begin{cases} 1 & \text{if } A_i \gg A_j \text{ solving } \Phi_{p,m} \\ 0 & \text{in other case} \end{cases}, \quad (5)$$

where the relation  $A_i \gg A_j$  defines whether  $A_i$  is significantly better than  $A_j$  when solving the problem instance  $\Phi_{p,m}$ , as computed by the above statistical tests.

Relying on  $\delta(\cdot)$ , the performance index  $P_{p,m}(A_i)$  of a given

algorithm  $A_i$  when solving  $\Phi_{p,m}$  is then computed as

$$P_{p,m}(A_i) = \sum_{j=1; j \neq i}^K \delta(A_i, A_j, \Phi_{p,m}). \quad (6)$$

This index should summarize the performance of each algorithm with regard to its peers.

Fig. 9 exhibits the results computing the performance indexes. Fig. 9a represents the mean performance indexes yielded by each algorithm when solving each problem in all of its configured objective dimensions,

$$\bar{P}_p(A_i) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} P_{p,m}(A_i). \quad (7)$$

We have not included NSGA-II and SPEA2 in the plots as they were clearly outperformed by the other algorithms, and would, therefore, not be useful for presenting results. Nevertheless, their results were used to compute the performance indexes.

It is worth noticing that GNG and MB-GNG have better overall results than the other algorithms. It is somewhat

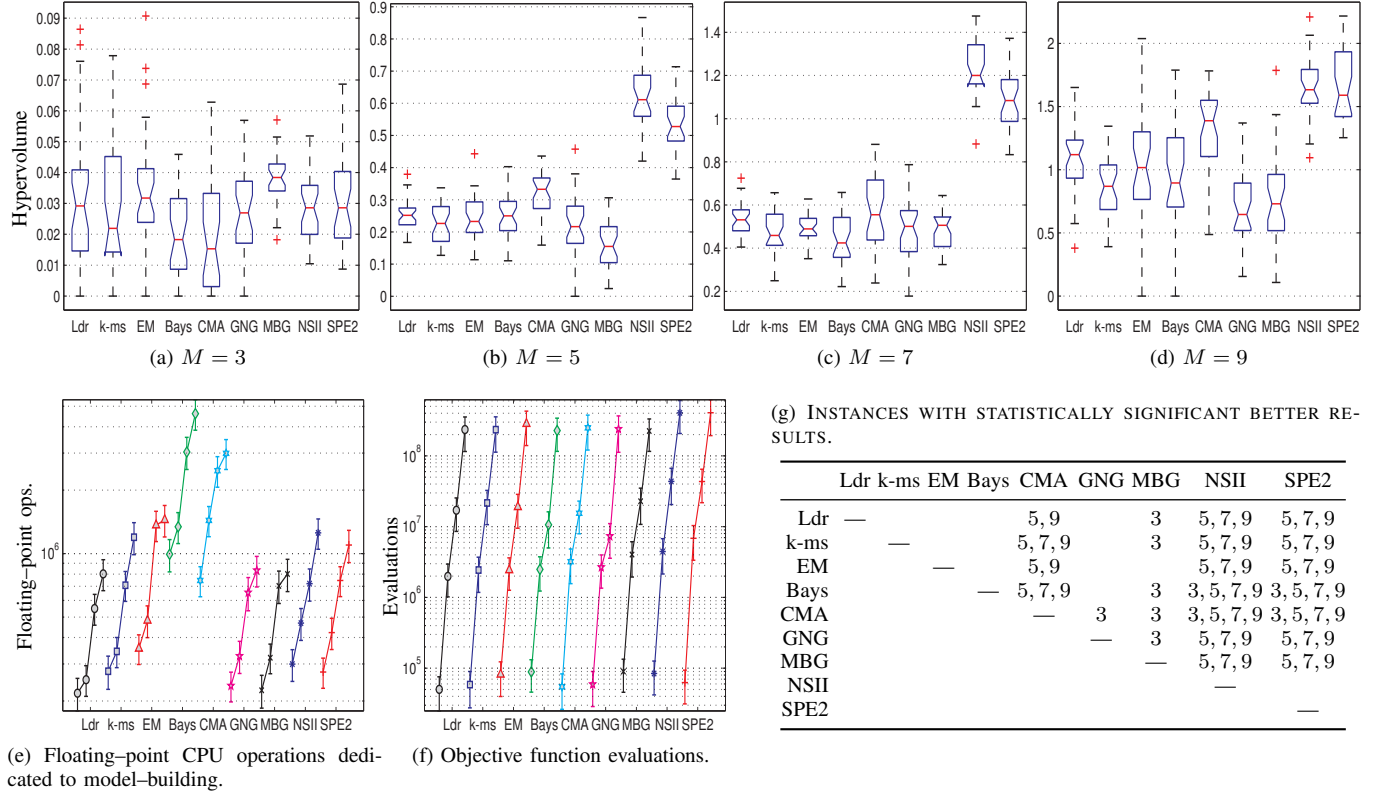


Figure 4: Results when solving the WFG5 problem. See Fig 3 for a description of each subfigure and abbreviations.

unexpected that the randomized leader and the  $k$ -means algorithms do not have a very good overall performance for some problems, like WFG5 and WFG7 for the randomized leader and WFG8 and WFG9  $k$ -means. A possible hypothesis is that these results may be biased by the three-objective problems, where there are sizable differences compared with the results of the other dimensions.

This situation is clarified in Fig. 9b, which presents the mean values of the index computed for each dimension

$$\bar{P}_m(A_i) = \frac{1}{P} \sum_{p=1}^P P_{p,m}(A_i). \quad (8)$$

There is evidence that there is no substantial difference between the results yielded by the different algorithms in the three-objective case, as their index values are more uniform. It is also noticeable that CMA-ES seems to outperform all the other algorithms for all problems in this dimension. This panorama changes when inspecting the results in higher dimensionality (in the objective function space). In those cases the least statistically robust algorithms tend to perform comparatively better, with the exception of Bayesian networks that seem to improve as the number of dimensions increases, but, of course, at the expense of a great computational cost.

It is worthwhile analyzing the performance of MB-GNG. In most cases, MB-GNG outperformed the other algorithms in higher dimensionality. This corroborates the results that we presented elsewhere [55], [113]. This outcome can be attributed to the fact that MB-GNG is the only algorithm

that has so far been devised especially for the model-building problem.

## V. TOWARDS A PARADIGM SHIFT

The above results prompt a series of considerations that we believe could be used as the basis for a possible paradigm shift within MOEDAs. One of the main conclusions is that model-building algorithms without a solid statistical foundation generally outperform the others for problems with a dimensionality greater than three (in the objective function space). These results, therefore, sustain the hypothesis put forward in Section III. It is now more evident that the model-building problem has different characteristics to other existing machine learning problems.

The improvement achieved with the application of MB-GNG is particularly noteworthy. Although it could be argued that the custom-designed MB-GNG yields substantially better results with respect to current alternatives, we find that there is still a lot of room for improvement in this area. Therefore, a more fruitful debate would be around how to create algorithms that are capable of properly dealing with the model-building issue.

It is indeed true that the curse of dimensionality cannot be avoided in the long term. Similarly, the no-free-lunch theorem in the multi-objective case has shown that there will be no universal multi-objective optimizer that outperforms all the other algorithms in all cases [114]. However, if we analyze the issues debated in this paper and in the light of the experimental results presented here, we can point out different

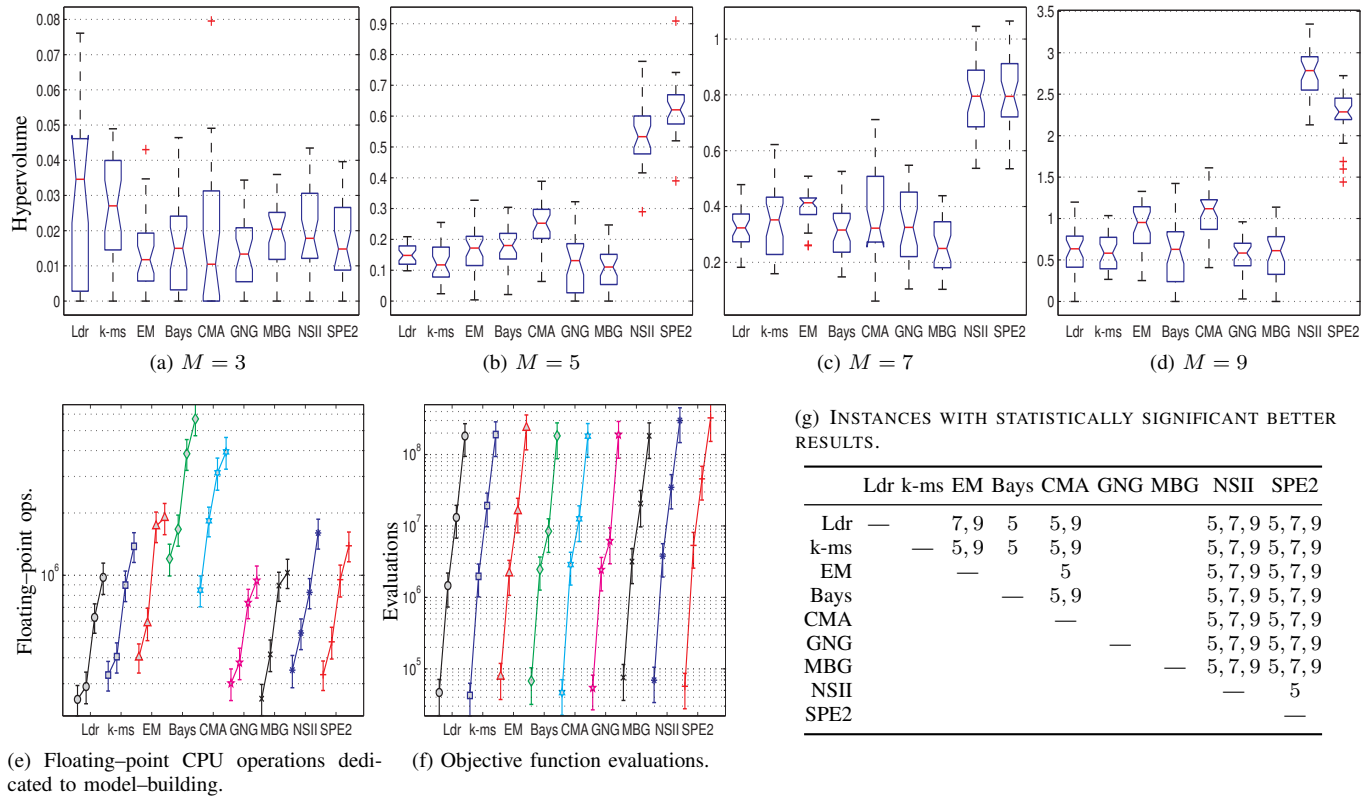


Figure 5: Results when solving the WFG6 problem. See Fig 3 for a description of each subfigure and abbreviations.

directions that may be pursued in order to achieve a substantial improvement in the MOEDA area.

As stated previously, one of the main causes of the current limitations of MOEDAs can, in our opinion, be attributed to their disregard of outliers. In turn, this behavior can be put down to the error-based learning approaches that take place in the underachieving MOEDAs.

Error-based learning is rather common in most machine learning algorithms. It implies that model topology and parameters are tuned in order to minimize a global error measured across the learning data set. This type of learning of isolated data is not taken into account because these data contribute little to the overall error and, therefore, do not take an active part in the learning process.

This behavior makes sense in the context of many problems, as isolated data can be interpreted as being spurious, noisy or invalid. As we argued in Section III, however, this is not the case in model-building. In model-building, all data are equally important, and, furthermore, isolated data might have a greater significance as they represent unexplored regions of the current optimal search space. This assessment is supported by the fact that most of the better-performing approaches do not follow the error-based scheme. For this reason, perhaps another class of learning, such as instance-based learning (IBL) [115], [116] or match-based learning [117] would yield a sizable advantage. As a matter of fact, the leader and  $k$ -means algorithms are good representatives of IBL.

Another strategy of interest is the fusion of the information present in both the decision variable space and objective

function space. Most MOEDAs construct their models by exploiting only the decision variable space information, since the resulting model can be used for sampling new individuals. To the best of our knowledge, the only MOEDA work that has addressed this issue is related to the use of the multi-objective hierarchical BOA (mhBOA) [16], [31]. MhBOA performs a  $k$ -means clustering of the local Pareto front obtained after applying the NSGA-II ranking function. Then, a local model is built for each cluster. It is worth remarking that a simpler approach would be to replace the NSGA-II's ranking function for one based on SPEA2, which has an embedded clustering process. Nevertheless, the underlying idea here is that the model would benefit from taking into account the properties of the individuals in both spaces.

Model reuse across iterations is another important issue. The most popular approaches so far either (i) create and later discard new models in every iteration or (ii) infer some of the most costly properties (such as the network topology in Bayesian networks) beforehand and tune the others in each iteration.

The first solution has the obvious drawback of wasting resources when large parts of the model are likely to be able to be reused across iterations. On the other hand, the other approach does not take into account the evolution of the local Pareto-optimal front and set as the optimization process progresses. To get MOEDAs with better scalability, the model-building algorithms must be able to handle some degree of reusability and, therefore, minimize the amount of computation carried out in each iteration.

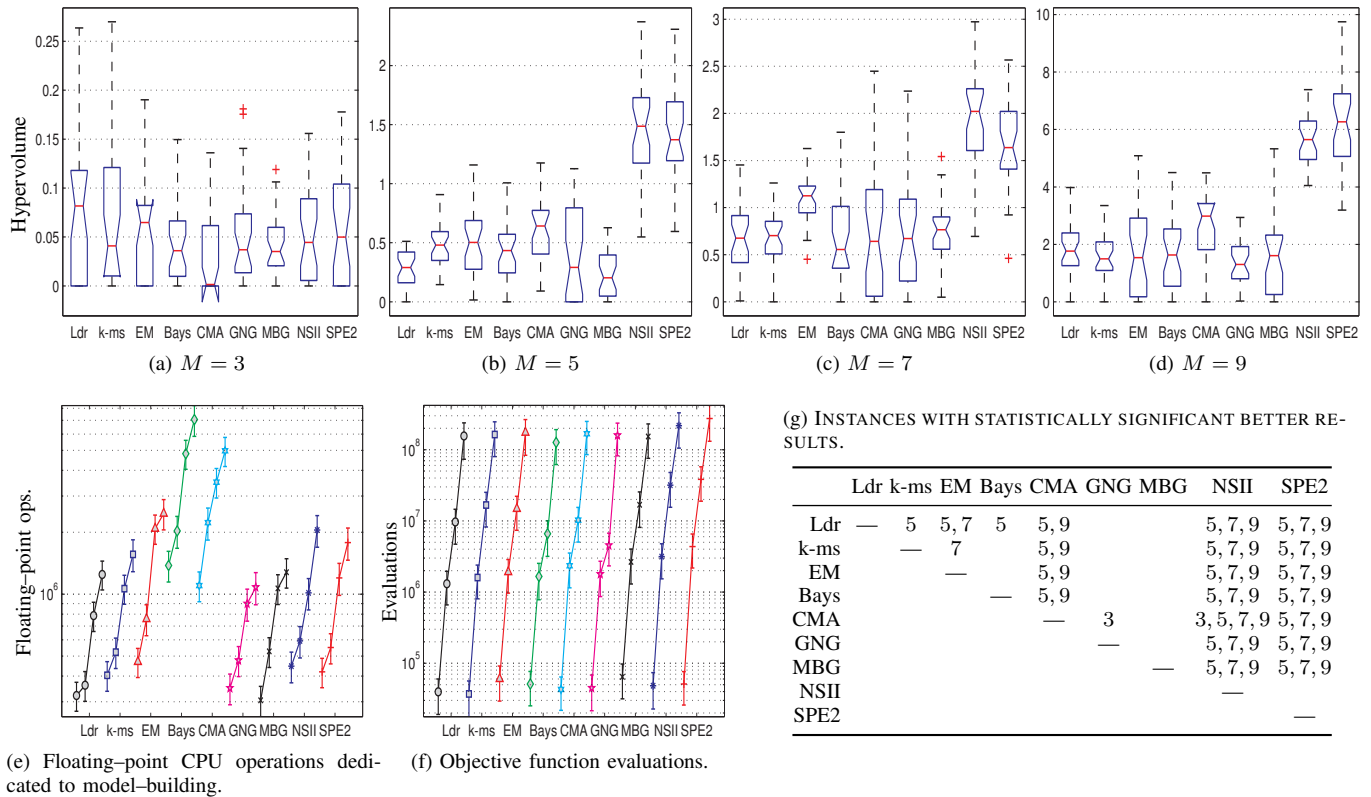


Figure 6: Results when solving the WFG7 problem. See Fig 3 for a description of each subfigure and abbreviations.

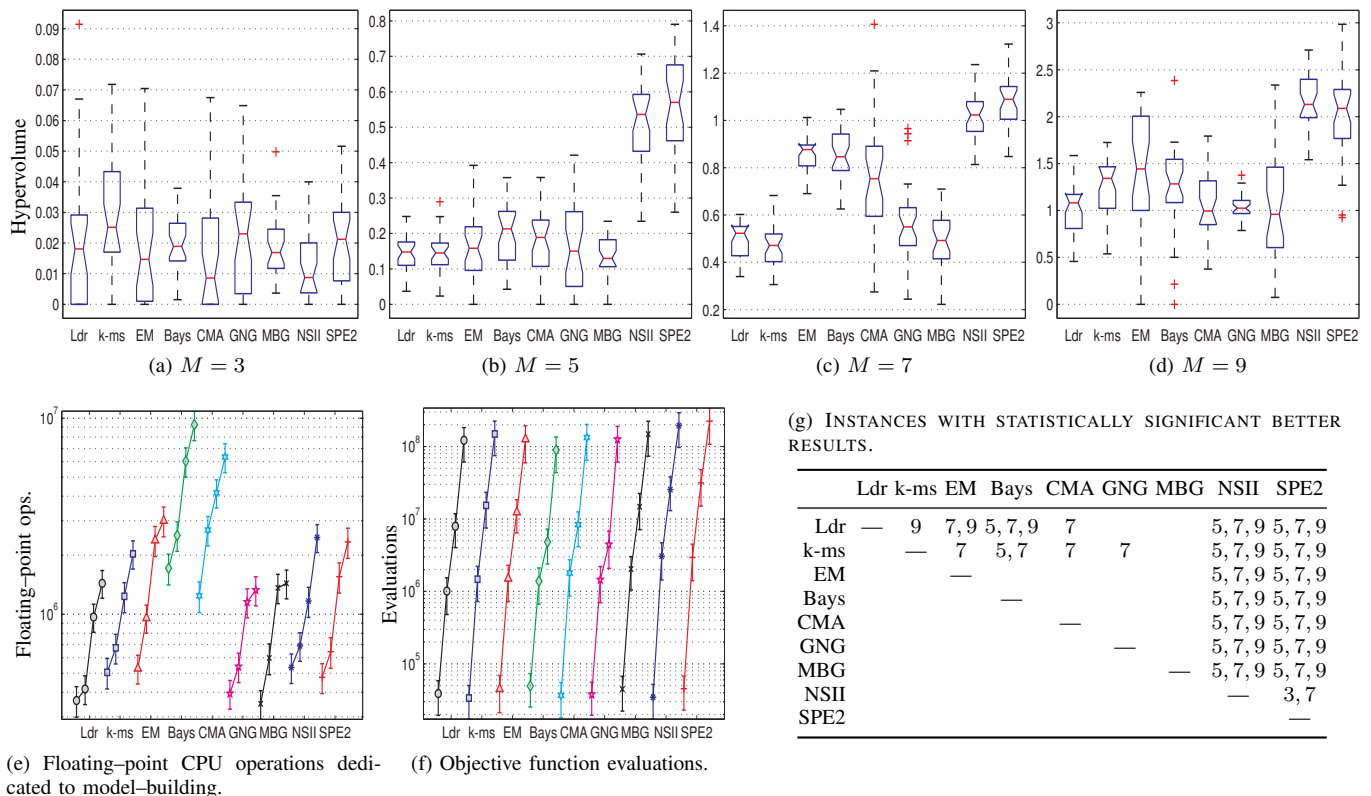


Figure 7: Results when solving the WFG8 problem. See Fig 3 for a description of each subfigure and abbreviations.

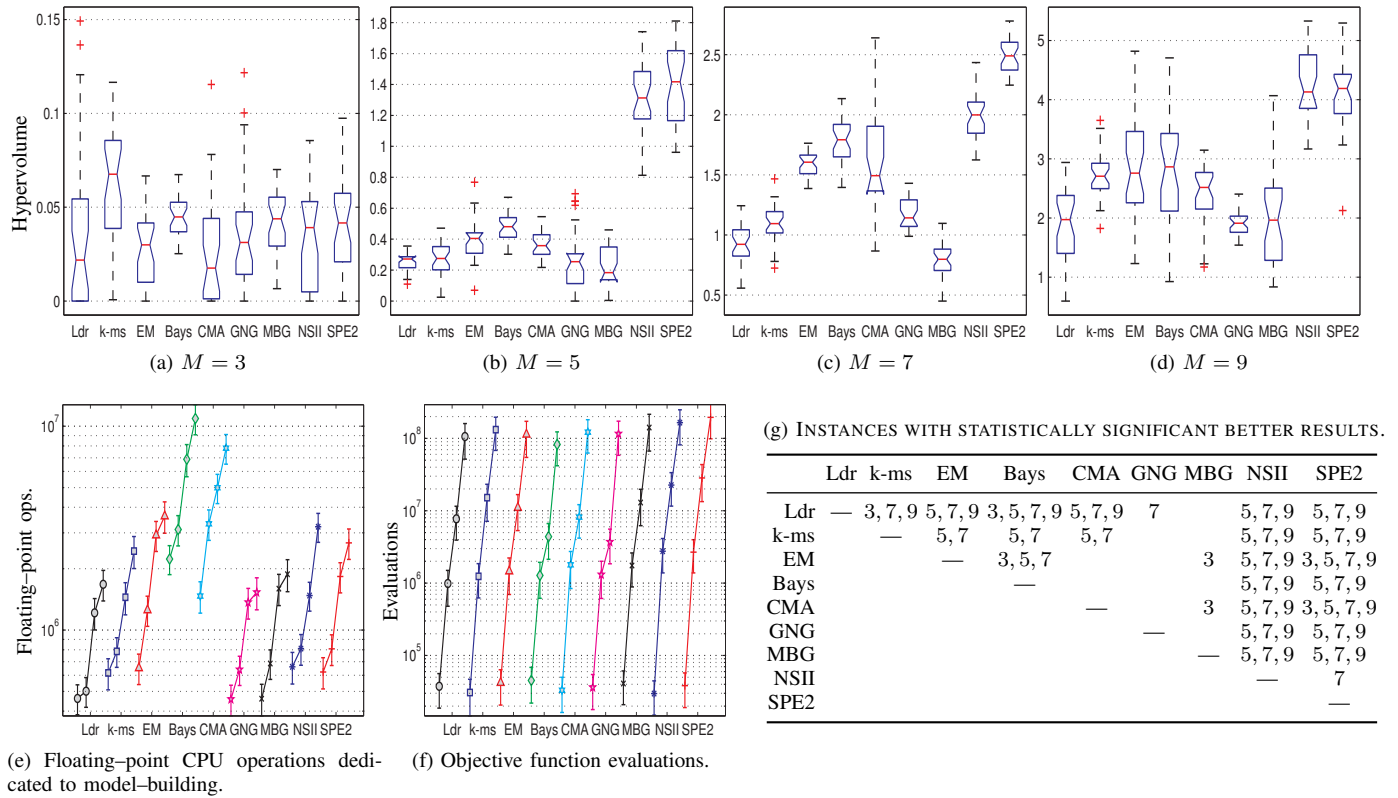


Figure 8: Results when solving the WFG9 problem. See Fig 3 for a description of each subfigure and abbreviations.

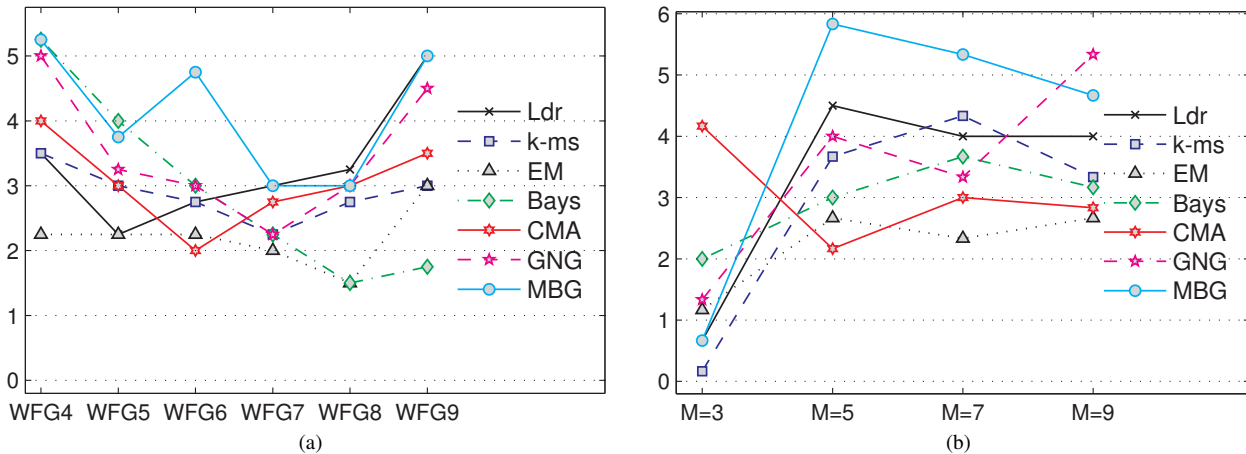


Figure 9: Mean values of the performance index across the different problems,  $\bar{P}_p()$ , (Fig. (a)) and objective space dimensions,  $\bar{P}_m$  (Fig. (b)).

In any case, it is clear from the above discussions and experiments that the model-building problem warrants a different approach that takes into account the particularities of the problem being solved. The ultimate solution to this issue is, perhaps, to create custom-made algorithms that meet the specific requirements of the problem at hand.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have discussed an important issue in current evolutionary multi-objective optimization: how to build algo-

rithms that have better scalability with regard to the number of objectives. In particular, we have focused on one promising set of approaches: estimation of distribution algorithms.

We have argued that most of the current approaches do not take into account the particularities of the model-building problem that they are addressing and that, for this reason, they fail to yield results of substantial quality.

We have also carried out a set of experiments that showed the points being discussed. The experiments illustrated empirically that algorithms that have no statistical groundwork

Table I: Parameters of the algorithms used in the experiments.

Common parameters	
Population size ( $n_{\text{pop}}$ )	$250 \cdot 10^{\frac{M}{3}} - 1$
Shared EDA framework	
Selection percentile ( $\alpha$ )	0.3
Substitution percentile ( $\omega$ )	0.3
GNG and MB-GNG	
Number of initial GNG nodes ( $N_0$ )	2
Maximum edge age ( $\nu_{\text{max}}$ )	40
Best node learning rate ( $\epsilon_b$ )	0.1
Neighbor nodes learning rate ( $\epsilon_v$ )	0.05
Insertion error decrement rate ( $\delta_I$ )	0.1
General error decrement rate ( $\delta_G$ )	0.1
Accumulated error threshold ( $\rho$ )	0.2
$\hat{P}_t$ to $N_{\text{max}}$ ratio ( $\gamma$ )	0.5
Randomized leader algorithm	
Maximum number of clusters $[0.5[\alpha n_{\text{pop}}]]$	
Threshold for the leader algorithm	0.1
$k$ -means algorithm	
Number of clusters $[0.25[\tau n_{\text{pop}}]]$	
Stopping threshold	0.0001
Expectation maximization	
Maximum number of clusters $[0.5[\tau n_{\text{pop}}]]$	
Threshold for the leader algorithm	0.1
Bayesian networks	
Number of parents of a variable	5
Number of mixture components	3
Threshold of leader algorithm	0.1
Covariance matrix adaptation	
Offspring number ( $\lambda$ )	1
Target success probability ( $p_{\text{succ}}^{\text{target}}$ )	$\frac{1}{5 + \sqrt{\frac{\lambda}{2}}}$
Step size damping ( $d$ )	$1 + \frac{n}{2\lambda}$
Success rate averaging parameter ( $c_p$ )	$\frac{\lambda p_{\text{succ}}^{\text{target}}}{2 + p_{\text{succ}}^{\text{target}}}$
Cumulation time horizon parameter ( $c_c$ )	$\frac{n+2}{2}$
Covariance matrix learning rate ( $c_{\text{cov}}$ )	$\frac{n+2}{n^2+6}$
Success rate threshold ( $p_{\text{thresh}}$ )	0.44
NSGA-II	
Crossover probability ( $p_c$ )	0.7
Distribution index for SBX ( $\eta_c$ )	15
Mutation probability ( $p_m$ )	$\frac{1}{n_{\text{pop}}}$
Dist. index for polynomial mut. ( $\eta_m$ )	20
SPEA2	
Crossover probability ( $p_c$ )	0.7
Distribution index for SBX ( $\eta_c$ )	15
Mutation probability ( $p_m$ )	$\frac{1}{n_{\text{pop}}}$
Dist. index for polynomial mut. ( $\eta_m$ )	20
Ratio of pop. to archive sizes	4 : 1

outperformed others that do. According to the hypothesis put forward in this paper, such behavior is caused by the fact that model-building has not yet been recognized as different from typical machine learning problems and, as such, having specific requirements that need to be met. The main aim of this paper is to trigger further studies on this topic and, ultimately, new model-building algorithms.

## ACKNOWLEDGMENTS

This work was supported by projects CICYT TIN2008-06742-C02-02/TSI, CICYT TEC2008-06732-C02-02/TEC, SINPROB, CAM MADRINET S-0505/ TIC/0255 and DPS2008-07029-C02-02. The fourth author acknowledges support from CONACyT project no. 103570.

APPENDIX A  
PARAMETERS

The parameters of the different algorithms involved in the experiments are summarized in Table I.

## REFERENCES

- [1] V. Pareto, *Cours D'Économie Politique*. Lausanne: F. Rouge, 1896.
- [2] K. Miettinen, *Nonlinear Multiobjective Optimization*, ser. International Series in Operations Research & Management Science. Norwell, MA: Kluwer, 1999, vol. 12.
- [3] M. Ehrgott, *Multicriteria Optimization*, ser. Lecture Notes in Economics and Mathematical Systems. Springer, 2005, vol. 491.
- [4] R. C. Purshouse and P. J. Fleming, "On the evolutionary optimization of many conflicting objectives," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 770–784, 2007. [Online]. Available: <http://dx.doi.org/10.1109/TEVC.2007.910138>
- [5] O. Brandt and S. Malinchik, "A Broad and Narrow Approach to Interactive Evolutionary Design - An Aircraft Design Example," in *Genetic and Evolutionary Computation-GECCO 2004. Proceedings of the Genetic and Evolutionary Computation Conference. Part II*, K. D. et al., Ed. Seattle, Washington, USA: Springer-Verlag, Lecture Notes in Computer Science Vol. 3103, June 2004, pp. 883–895.
- [6] T. J. Stewart, R. Janssen, and M. van Herwijnen, "A genetic algorithm approach to multiobjective land use planning," *Computers and Operations Research*, vol. 32, pp. 2293–2313, 2004.
- [7] J. A. Besada, J. García, G. De Miguel, A. Berlanga, J. M. Molina, and J. R. Casar, "Design of IMM filter for radar tracking using evolution strategies," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 3, pp. 1109–1122, July 2005.
- [8] J. García, A. Berlanga, and J. M. Molina, "Effective evolutionary algorithms for many-specifications attainment: Application to air traffic control tracking filters," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 1, pp. 151–168, 2009. [Online]. Available: <http://dx.doi.org/10.1109/TEVC.2008.920677>
- [9] H. Nakayama, S. Kaneshige, S. Takemoto, and Y. Watada, "An application of a multi-objective programming technique to construction accuracy control of cable-stayed bridges," *European Journal of Operational Research*, vol. 87, pp. 731–738, 1995.
- [10] T. J. Stewart, O. Bandt, H. Braun, N. Chakraborti, M. Ehrgott, M. Göbelt, Y. Jin, H. Nakayama, S. Poles, and D. Di Stefano, "Real-world applications of multiobjective optimization," in *Multiobjective Optimization*, ser. Lecture Notes in Computer Science, J. Branke, M. Kaisa, K. Deb, and R. Słowiński, Eds. Berlin/Heidelberg: Springer-Verlag, 2008, vol. 5252, pp. 285–327.
- [11] P. Larrañaga and J. A. Lozano, Eds., *Estimation of Distribution Algorithms. A new tool for Evolutionary Computation*, ser. Genetic Algorithms and Evolutionary Computation. Boston/Dordrecht/London: Kluwer Academic Publishers, 2002.
- [12] J. A. Lozano, P. Larrañaga, I. Inza, and E. Bengoetxea, Eds., *Towards a New Evolutionary Computation: Advances on Estimation of Distribution Algorithms*. Springer-Verlag, 2006.
- [13] M. Pelikan, K. Sastry, and E. Cantú-Paz, Eds., *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications*, ser. Studies in Computational Intelligence. Springer, 2006.
- [14] S. Baluja, "Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning," Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-CS-94-163, 1994.
- [15] H. Mühlenbein and G. Paaß, "From recombination of genes to the estimation of distributions I. Binary parameters," in *Parallel Problem Solving from Nature - PPSN IV*, H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, Eds. Berlin: Springer Verlag, 1996, pp. 178–187, LNCS 1141.

- [16] M. Pelikan, K. Sastry, and D. E. Goldberg, "Multiobjective estimation of distribution algorithms," in *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications*, ser. Studies in Computational Intelligence, M. Pelikan, K. Sastry, and E. Cantú-Paz, Eds. Springer-Verlag, 2006, pp. 223–248.
- [17] M. Pelikan, D. E. Goldberg, and F. Lobo, "A survey of optimization by building and using probabilistic models," University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, IL, IliGAL Report No. 99018, 1999.
- [18] M. Pelikan, "Probabilistic model-building genetic algorithms," in *GECCO '08: Proceedings of the 2008 GECCO conference companion on Genetic and evolutionary computation*. New York, NY, USA: ACM, 2008, pp. 2389–2416. [Online]. Available: <http://dx.doi.org/10.1145/1388969.1389060>
- [19] P. A. N. Bosman, "Design and application of iterated density-estimation evolutionary algorithms," Ph.D. dissertation, Universiteit Utrecht, Utrecht, The Netherlands, 2003.
- [20] C. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006, ch. Graphical Models, pp. 359–422.
- [21] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.
- [22] M. Pelikan, D. E. Goldberg, and E. Cantú-Paz, "BOA: The Bayesian optimization algorithm," in *Proceedings of the Genetic and Evolutionary Computation Conference GECCO-1999*, W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela, and R. E. Smith, Eds., vol. I. Orlando, FL: Morgan Kaufmann Publishers, San Francisco, CA, 1999, pp. 525–532.
- [23] R. Etxeberria and P. Larrañaga, "Global optimization using Bayesian networks," in *Proceedings of the Second Symposium on Artificial Intelligence (CIMA-99)*, A. Ochoa, M. R. Soto, and R. Santana, Eds., Habana, Cuba, 1999, pp. 151–173.
- [24] H. Mühlenbein and T. Mahnig, "FDA – a scalable evolutionary algorithm for the optimization of additively decomposed functions," *Evolutionary Computation*, vol. 7, no. 4, pp. 353–376, 1999.
- [25] G. Cooper and E. Herskovits, "A bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309–347, 1992.
- [26] N. Khan, D. E. Goldberg, and M. Pelikan, "Multi-Objective Bayesian Optimization Algorithm," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'2002)*, W. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. Potter, A. Schultz, J. Miller, E. Burke, and N. Jonoska, Eds. San Francisco, California: Morgan Kaufmann Publishers, July 2002, p. 684.
- [27] M. Pelikan, D. E. Goldberg, and E. Cantú-Paz, "Hierarchical problem solving by the Bayesian optimization algorithm," University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, IL, IliGAL Report No. 2000002, 2000.
- [28] M. Pelikan, *Hierarchical Bayesian Optimization Algorithm. Toward a New Generation of Evolutionary Algorithms*, ser. Studies in Fuzziness and Soft Computing. Springer, 2005.
- [29] M. Pelikan and D. E. Goldberg, "Hierarchical bayesian optimization algorithm," in *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications*, ser. Studies in Computational Intelligence, M. Pelikan, K. Sastry, and E. Cantú-Paz, Eds. Springer-Verlag, 2006, pp. 63–90.
- [30] N. Khan, "Bayesian Optimization Algorithms for Multiobjective and Hierarchically Difficult Problems," Master's thesis, Graduate College of the University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, 2003.
- [31] M. Pelikan, K. Sastry, and D. E. Goldberg, "Multiobjective hboa, clustering, and scalability," in *GECCO '05: Proceedings of the 2005 conference on Genetic and evolutionary computation*. New York, NY, USA: ACM Press, 2005, pp. 663–670. [Online]. Available: <http://dx.doi.org/10.1145/1068009.1068122>
- [32] M. Laumanns and J. Ocenasek, "Bayesian Optimization Algorithms for Multi-objective Optimization," in *Parallel Problem Solving from Nature—PPSN VII*, J. J. Merelo Guervós, P. Adamidis, H.-G. Beyer, J.-L. F.-V. nas, and H.-P. Schwefel, Eds. Granada, Spain: Springer-Verlag, Lecture Notes in Computer Science No. 2439, September 2002, pp. 298–307.
- [33] J. Ocenasek, "Parallel Estimation of Distribution Algorithms," Ph.D. dissertation, Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic, November 2002.
- [34] J. Ocenasek and J. Schwarz, "Estimation of distribution algorithm for mixed continuous–discrete optimization problems," in *2nd Euro-International Symposium on Computational Intelligence*, 2002, pp. 227–232.
- [35] C. W. Ahn, *Advances in Evolutionary Algorithms. Theory, Design and Practice*. Springer, 2006, ISBN 3-540-31758-9.
- [36] C. W. Ahn, D. E. Goldberg, and R. S. Ramakrishna, "Real-coded bayesian optimization algorithm: Bringing the strength of boa into the continuous world," in *2004 Genetic and Evolutionary Computation (GECCO 2004)*, ser. Lecture Notes in Computer Science. Springer, 2004, vol. 3102, pp. 840–851.
- [37] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [38] R. Kindermann and J. L. Snell, *Markov Random Fields and Their Applications*, ser. Contemporary Mathematics. Providence, RI: American Mathematical Society, 1980. [Online]. Available: [http://www.ams.org/online\\_bks/conm1/](http://www.ams.org/online_bks/conm1/)
- [39] R. Santana, "A Markov network based factorized distribution algorithm for optimization," in *Machine Learning: ECML 2003*, ser. Lecture Notes in Artificial Intelligence, N. Lavrač, D. Gamberger, H. Blockeel, and L. Todorovski, Eds. Berlin/Heidelberg/New York: Springer, 2003, vol. 2837, pp. 337–348. [Online]. Available: <http://www.springerlink.com/content/957vw4q0gu38q1xa>
- [40] —, "Estimation of distribution algorithms with Kikuchi approximations," *Evolutionary Computation*, vol. 13, no. 1, pp. 67–97, 2005. [Online]. Available: <http://www.mitpressjournals.org/doi/abs/10.1162/1063656053583496>
- [41] S. Shakya and J. McCall, "Optimization by estimation of distribution with DEUM framework based on markov random fields," *International Journal of Automation and Computing*, vol. 4, no. 3, pp. 262–272, July 2007. [Online]. Available: <http://dx.doi.org/10.1007/s11633-007-0262-6>
- [42] D. Thierens and P. A. N. Bosman, "Multi-objective mixture-based iterated density estimation evolutionary algorithms," in *Proceedings of the Genetic and Evolutionary Computation Conference GECCO-2001*, L. Spector, E. Goodman, A. Wu, W. Langdon, H. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. Garzon, and E. Burke, Eds. San Francisco, CA: Morgan Kaufmann Publishers, 2001, pp. 663–670.
- [43] —, "Multi-Objective Optimization with Iterated Density Estimation Evolutionary Algorithms using Mixture Models," in *Proceedings of the Third International Symposium on Adaptive Systems—Evolutionary Computation and Probabilistic Graphical Models*. Havana, Cuba: Institute of Cybernetics, Mathematics and Physics, March 19–23 2001, p. 129–136.
- [44] P. A. N. Bosman and D. Thierens, "Multi-objective optimization with diversity preserving mixture-based iterated density estimation evolutionary algorithms," *International Journal of Approximate Reasoning*, vol. 31, no. 3, pp. 259–289, 2002.
- [45] —, "The Balance Between Proximity and Diversity in Multiobjective Evolutionary Algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 174–188, April 2003.
- [46] D. Thierens, "Convergence Time Analysis for the Multi-objective Counting Ones Problem," in *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. Faro, Portugal: Springer, Lecture Notes in Computer Science. Volume 2632, April 2003, pp. 355–364.
- [47] P. A. N. Bosman and D. Thierens, "The naïve MIDEA: A baseline multi-objective EA," in *Evolutionary Multi-Criterion Optimization. Third International Conference, EMO 2005*, C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, Eds. Guanajuato, México: Springer, Lecture Notes in Computer Science Vol. 3410, March 2005, pp. 428–442.
- [48] J. A. Hartigan, *Clustering Algorithms*, ser. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons, 1975.
- [49] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical*, vol. 1, 1967, pp. 281–297.
- [50] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. B, no. 39, pp. 1–38, 1977.
- [51] P. A. N. Bosman, "Design and application of iterated density-estimation evolutionary algorithms," Ph.D. dissertation, Institute of Information and Computing Sciences, Universiteit Utrecht, Utrecht, The Netherlands, 2003.
- [52] M. Costa and E. Minisci, "MOPEd: A Multi-objective Parzen-Based Estimation of Distribution Algorithm for Continuous Problems," in *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb,

- and L. Thiele, Eds. Faro, Portugal: Springer. Lecture Notes in Computer Science. Volume 2632, April 2003, pp. 282–294.
- [53] M. Costa, E. Minisci, and E. Pasero, “An hybrid neural/genetic approach to continuous multi-objective optimization problems,” in *Italian Workshop on Neural Neural Nets (WIRN)*, ser. Lecture Notes in Computer Science, B. Apolloni, M. Marinaro, and R. Tagliaferri, Eds., vol. 2859. Springer, 2003, pp. 61–69.
- [54] E. Parzen, “On estimation of a probability density function and mode,” *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.
- [55] L. Martí, J. García, A. Berlanga, and J. M. Molina, “Introducing MON-EDA: Scalable multiobjective optimization with a neural estimation of distribution algorithm,” in *GECCO '08: 10th Annual Conference on Genetic and Evolutionary Computation*, D. Thierens, K. Deb, M. Pelikan, H.-G. Beyer, B. Doerr, R. Poli, and M. Bittari, Eds. New York, NY, USA: ACM Press, 2008, pp. 689–696, *EMO Track “Best Paper” Nominee*.
- [56] B. Fritzsche, “A growing neural gas network learns topologies,” in *Advances in Neural Information Processing Systems*, G. Tesauero, D. S. Touretzky, and T. K. Leen, Eds. Cambridge, MA: MIT Press, 1995, vol. 7, pp. 625–632.
- [57] L. Martí, J. García, A. Berlanga, and J. M. Molina, “Scalable continuous multiobjective optimization with a neural network-based estimation of distribution algorithm,” in *Applications of Evolutionary Computing*, ser. Lecture Notes in Computer Science, M. Giacobini, A. Brabazon, S. Cagnoni, G. A. Di Caro, R. Drechsler, A. Ekárt, A. I. Esparcia-Alcázar, M. Farooq, A. Fink, J. McCormack, M. O’Neill, J. Romero, F. Rothlauf, G. Squillero, A. c. Uyar, and S. Yang, Eds. Heidelberg: Springer, 2008, vol. 4974, pp. 535–544. [Online]. Available: [http://dx.doi.org/10.1007/978-3-540-78761-7\\_59](http://dx.doi.org/10.1007/978-3-540-78761-7_59)
- [58] A. K. Qin and P. N. Suganthan, “Robust growing neural gas algorithm with application in cluster analysis,” *Neural Networks*, vol. 17, no. 8–9, pp. 1135–1148, 2004. [Online]. Available: <http://dx.doi.org/10.1016/j.neunet.2004.06.013>
- [59] T. M. Martinetz, S. G. Berkovich, and K. J. Shulten, “Neural-gas network for vector quantization and its application to time-series prediction,” *IEEE Transactions on Neural Networks*, vol. 4, pp. 558–560, 1993.
- [60] N. Hansen and A. Ostermeier, “Completely derandomized self-adaptation in evolution strategies,” *Evolutionary Computation*, vol. 9, no. 2, pp. 159–195, 2001.
- [61] N. Hansen, S. Muller, and P. Koumoutsakos, “Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES),” *Evolutionary Computation*, vol. 11, no. 1, pp. 1–18, 2003.
- [62] A. Auger, “Benchmarking the (1+1)-ES with one-fifth success rule on the BBOB-2009 noisy testbed,” in *GECCO '09: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference*. New York, NY, USA: ACM, 2009, pp. 2453–2458.
- [63] —, “Benchmarking the (1+1) evolution strategy with one-fifth success rule on the bbob-2009 function testbed,” in *GECCO '09: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference*. New York, NY, USA: ACM, 2009, pp. 2447–2452.
- [64] A. Auger and N. Hansen, “Benchmarking the (1+1)-CMA-ES on the BBOB-2009 noisy testbed,” in *GECCO '09: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference*. New York, NY, USA: ACM, 2009, pp. 2467–2472.
- [65] H.-G. Beyer and H.-P. Schwefel, “Evolution strategies — A comprehensive introduction,” *Natural Computing: an international journal*, vol. 1, no. 1, pp. 3–52, 2002.
- [66] C. Igel, N. Hansen, and S. Roth, “Covariance matrix adaptation for multi-objective optimization,” *Evolutionary Computation*, vol. 15, no. 1, pp. 1–28, 2007.
- [67] Q. Zhang, A. Zhou, and Y. Jin, “RM-MEDA: A regularity model-based multiobjective estimation of distribution algorithm,” *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 1, pp. 41–63, 2008. [Online]. Available: <http://dx.doi.org/10.1109/TEVC.2007.894202>
- [68] A. Zhou, Q. Zhang, Y. Jin, E. Tsang, and T. Okabe, “A Model-Based Evolutionary Algorithm for Bi-objective Optimization,” in *2005 IEEE Congress on Evolutionary Computation (CEC'2005)*, vol. 3. Edinburgh, Scotland: IEEE Service Center, September 2005, pp. 2568–2575.
- [69] O. Schütze, S. Mostaghim, M. Dellnitz, and J. Teich, “Covering Pareto Sets by Multilevel Evolutionary Subdivision Techniques,” in *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. Faro, Portugal: Springer. Lecture Notes in Computer Science. Volume 2632, April 2003, pp. 118–132.
- [70] N. Kambhatla and T. K. Leen, “Dimension reduction by local principal component analysis,” *Neural Computation*, vol. 9, no. 7, pp. 1493–1516, 1997.
- [71] R. E. Bellman, *Adaptive Control Processes*. Princeton, NJ: Princeton University Press, 1961.
- [72] V. Khare, X. Yao, and K. Deb, “Performance Scaling of Multi-objective Evolutionary Algorithms,” in *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. Faro, Portugal: Springer. Lecture Notes in Computer Science. Volume 2632, April 2003, pp. 376–390.
- [73] K. Praditwong and X. Yao, “How well do multi-objective evolutionary algorithms scale to large problems,” in *2007 IEEE Congress on Evolutionary Computation (CEC 2007)*. Piscataway, New Jersey: IEEE Press, 2007, pp. 3959–3966. [Online]. Available: <http://dx.doi.org/10.1109/CEC.2007.4424987>
- [74] K. Deb, *Multi-Objective Optimization using Evolutionary Algorithms*. Chichester, UK: John Wiley & Sons, 2001, ISBN 0-471-87339-X.
- [75] K. Deb and D. K. Saxena, “On finding Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems,” KanGAL, Tech. Rep. 2005011, December 2005.
- [76] —, “Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems,” in *2006 IEEE Conference on Evolutionary Computation (CEC'2006)*. Piscataway, New Jersey: IEEE Press, 2006, pp. 3352–3360.
- [77] D. Brockhoff and E. Zitzler, “Dimensionality reduction in multi-objective optimization: The minimum objective subset problem,” in *Operations Research Proceedings 2006*, K. H. Waldmann and U. M. Stocker, Eds. Springer, 2007, pp. 423–429.
- [78] D. Brockhoff, D. K. Saxena, K. Deb, and E. Zitzler, “On handling a large number of objectives a posteriori and during optimization,” in *Multi-Objective Problem Solving from Nature: From Concepts to Applications*, ser. Natural Computing Series, J. Knowles, D. Corne, and K. Deb, Eds. Springer, 2008, pp. 377–403.
- [79] E. Zitzler and S. Künzli, “Indicator-based selection in multiobjective search,” in *Parallel Problem Solving from Nature — PPSN VIII*, ser. Lecture Notes in Computer Science, X. Yao, Ed., vol. 3242. Berlin/Heidelberg: Springer-Verlag, September 2004, pp. 832–842.
- [80] M. Basseur and E. Zitzler, “Handling uncertainty in indicator-based multiobjective optimization,” *International Journal of Computational Intelligence Research*, vol. 2, no. 3, pp. 255–272, 2006.
- [81] —, “A preliminary study on handling uncertainty in indicator-based multiobjective optimization,” in *Applications of Evolutionary Computing. EvoWorkshops 2006: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoINTERACTION, EvoMUSART, and EvoSTOC*, ser. Lecture Notes in Computer Science, F. Rothlauf, J. Branke, S. Cagnoni, E. Costa, C. Cotta, R. Drechsler, E. Lutton, P. Machado, J. H. Moore, J. Romero, G. D. Smith, G. Squillero, and H. Takagi, Eds., vol. 3907. Budapest, Hungary: Springer, April 2006, pp. 727–739.
- [82] D. Brockhoff and E. Zitzler, “Improving hypervolume-based multiobjective evolutionary algorithms by using objective reduction methods,” in *IEEE Congress on Evolutionary Computation (CEC 2007)*. IEEE Press, 2007, pp. 2086–2093.
- [83] J. Bader and E. Zitzler, “HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization,” Computer Engineering and Networks Laboratory (TIK), ETH Zurich, TIK Report 286, 2008.
- [84] J. Bader, K. Deb, and E. Zitzler, “Faster Hypervolume-based Search using Monte Carlo Sampling,” in *Conference on Multiple Criteria Decision Making (MCDM 2008)*. Springer, 2008, to appear.
- [85] H. Li and Q. Zhang, “Multiobjective optimization problems with complicated pareto sets, MOEA/D and NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 284–302, April 2009.
- [86] Q. Zhang and H. Li, “MOEA/D: A multiobjective evolutionary algorithm based on decomposition,” *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, Dec. 2007.
- [87] C. W. Ahn and R. S. Ramakrishna, “Multiobjective real-coded bayesian optimization algorithm revisited: diversity preservation,” in *GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation*. New York, NY, USA: ACM Press, 2007, pp. 593–600.



- [88] J. Shapiro, "Diversity loss in general estimation of distribution algorithms," in *Parallel Problem Solving from Nature - PPSN IX*, 2006, pp. 92–101. [Online]. Available: [http://dx.doi.org/10.1007/11844297\\_10](http://dx.doi.org/10.1007/11844297_10)
- [89] B. Yuan and M. Gallagher, "On the importance of diversity maintenance in estimation of distribution algorithms," in *GECCO '05: Proceedings of the 2005 conference on Genetic and evolutionary computation*. New York, NY, USA: ACM, 2005, pp. 719–726.
- [90] P. D. Grünwald, *The Minimum Description Length Principle (Adaptive Computation and Machine Learning)*. The MIT Press, 2007.
- [91] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, Sep 1999. [Online]. Available: <http://dx.doi.org/10.1109/72.788640>
- [92] S. Huband, L. Barone, L. While, and P. Hingston, "A Scalable Multi-objective Test Problem Toolkit," in *Evolutionary Multi-Criterion Optimization. Third International Conference, EMO 2005*, C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, Eds. Guanajuato, México: Springer. Lecture Notes in Computer Science Vol. 3410, March 2005, pp. 280–295.
- [93] S. Huband, P. Hingston, L. Barone, and L. While, "A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 5, pp. 477–506, October 2006.
- [94] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, April 2002.
- [95] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the Strength Pareto Evolutionary Algorithm," in *EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, K. Giannakoglou, D. Tsahalis, J. Periaux, P. Papailou, and T. Fogarty, Eds., Athens, Greece, 2002, pp. 95–100.
- [96] S. Huband, P. Hingston, L. Barone, and L. While, "A review of multiobjective test problems and a scalable test problem toolkit," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 5, pp. 477–506, 2006.
- [97] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable test problems for evolutionary multiobjective optimization," in *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*, ser. Advanced Information and Knowledge Processing, A. Abraham, L. Jain, and R. Goldberg, Eds. Springer Verlag, 2004, pp. 105–145.
- [98] E. Zitzler, K. Deb, and L. Thiele, "Comparison of Multiobjective Evolutionary Algorithms: Empirical Results," *Evolutionary Computation*, vol. 8, no. 2, pp. 173–195, Summer 2000.
- [99] L. Martí, J. García, A. Berlanga, and J. M. Molina, "A cumulative evidential stopping criterion for multiobjective optimization evolutionary algorithms," in *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (GECCO '07)*, D. Thierens, K. Deb, M. Pelikan, H.-G. Beyer, B. Doerr, R. Poli, and M. Bittari, Eds. New York: ACM Press, 2007, p. 911. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1276958.1277141>
- [100] L. Martí, J. García, A. Berlanga, and J. M. Molina, "An approach to stopping criteria for multi-objective optimization evolutionary algorithms: The MGBM criterion," in *2009 IEEE Conference on Evolutionary Computation (CEC 2009)*. Piscataway, New Jersey: IEEE Press, 2009, pp. 1263–1270.
- [101] J. Knowles, L. Thiele, and E. Zitzler, "A tutorial on the performance assessment of stochastic multiobjective optimizers," *Computer Engineering and Networks Laboratory (TIK), ETH Zurich, TIK Report 214*, 2006.
- [102] S. Bandyopadhyay, S. K. Pal, and B. Aruna, "Multiobjective GAs, Quantitative Indices, and Pattern Classification," *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 34, no. 5, pp. 2088–2099, October 2004.
- [103] H. Ishibuchi and T. Murata, "Multi-Objective Genetic Local Search Algorithm and Its Application to Flowshop Scheduling," *IEEE Transactions on Systems, Man and Cybernetics—Part C: Applications and Reviews*, vol. 28, no. 3, pp. 392–403, August 1998.
- [104] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler, "PISA—A Platform and Programming Language Independent Interface for Search Algorithms," in *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. Faro, Portugal: Springer. Lecture Notes in Computer Science. Volume 2632, April 2003, pp. 494–508.
- [105] C. A. Coello Coello, G. B. Lamont, and D. A. Van Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd ed., ser. Genetic and Evolutionary Computation. New York: Springer, 2007. [Online]. Available: <http://www.springer.com/west/home/computer/foundations?SGWID=4-156-22-173660344-0>
- [106] E. Zitzler, J. Knowles, and L. Thiele, "Quality assessment of pareto set approximations," in *Multiobjective Optimization*, ser. Lecture Notes in Computer Science, J. Branke, K. Miettinen, K. Deb, and R. Słowiński, Eds. Berlin/Heidelberg: Springer-Verlag, 2008, vol. 5252, pp. 373–404.
- [107] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion analysis of variance," *Journal of the American Statistical Association*, vol. 47, pp. 583–621, 1952.
- [108] W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed. New York: John Wiley & Sons, 1999.
- [109] J. Bader, "Hypervolume-Based Search for Multiobjective Optimization: Theory and Methods," Ph.D. dissertation, ETH Zurich, Switzerland, 2010.
- [110] J. Bader and E. Zitzler, "HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization," *Evolutionary Computation*, 2010, to appear.
- [111] J. Levon, *OProfile manual*, Victoria University of Manchester, 2004. [Online]. Available: <http://oprofile.sourceforge.net/>
- [112] M. Pelikan, *Hierarchical Bayesian Optimization Algorithm: Toward a New Generation of Evolutionary Algorithms*, 1st ed., ser. Studies in Fuzziness and Soft Computing. Springer, March 2005.
- [113] L. Martí, J. García, A. Berlanga, and J. M. Molina, "Solving complex high-dimensional problems with the multi-objective neural estimation of distribution algorithm," in *GECCO 2009: 11th Annual Conference on Genetic and Evolutionary Computation*, D. Thierens, K. Deb, M. Pelikan, H.-G. Beyer, B. Doerr, R. Poli, and M. Bittari, Eds. New York, NY, USA: ACM Press, 2009, pp. 619–626. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1569901.1569987>
- [114] D. W. Corne and J. D. Knowles, "No Free Lunch and Free Leftovers Theorems for Multiobjective Optimisation Problems," in *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. Faro, Portugal: Springer. Lecture Notes in Computer Science. Volume 2632, April 2003, pp. 327–341.
- [115] D. Kibler, D. W. Aha, and M. K. Albert, "Instance-based prediction of real-valued attributes," *Computational Intelligence*, vol. 5, no. 2, pp. 51–57, 1989.
- [116] J. R. Quinlan, "Combining instance-based and model-based learning," in *Machine Learning: Proceedings of the Tenth International Conference*. Morgan Kaufmann, 1993, pp. 236–243.
- [117] S. Grossberg, *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control*. Boston: Reidel, 1982.