Author(s): Afsar, Bekir; Miettinen, Kaisa; Ruiz, Francisco

Title: Assessing the Performance of Interactive Multiobjective Optimization Methods : A Survey

Year: 2021

Version: Accepted version (Final draft)

Please cite the original version:

Afsar, B., Miettinen, K., & Ruiz, F. (2021). Assessing the Performance of Interactive Multiobjective Optimization Methods : A Survey. ACM Computing Surveys, 54(4), Article 85. https://doi.org/10.1145/3448301

Assessing the Performance of Interactive Multiobjective Optimization Methods: A Survey

Bekir Afsar[1], Kaisa Miettinen[1], and Francisco Ruiz[2]

[1]University of Jyvaskyla, Faculty of Information Technology, Finland
[2]Department of Applied Economics (Mathematics), Universidad de Málaga, Málaga, Spain

Abstract

Interactive methods are useful decision making tools for multiobjective optimization problems, because they allow a decision maker to provide her/his preference information iteratively in a comfortable way at the same time as (s)he learns about all different aspects of the problem. A wide variety of interactive methods is nowadays available, and they differ from each other in both technical aspects and type of preference information employed. Therefore, assessing the performance of interactive methods can help users to choose the most appropriate one for a given problem. This is a challenging task, which has been tackled from different perspectives in the published literature. We present a bibliographic survey of papers where interactive multiobjective optimization methods have been assessed (either individually, or compared to other methods). Besides other features, we collect information about the type of decision maker involved (utility or value functions, artificial or human decision maker), the type of preference information provided and aspects of interactive methods that were somehow measured. Based on the survey and on our own experiences, we identify a series of desirable properties of interactive methods that we believe should be assessed.

Keywords: Interactive methods; Multiobjective optimization problems; Performance assessment; Decision makers; Preference information.

1 Introduction

Multiobjective optimization problems, where several conflicting objective functions are to be optimized simultaneously, do not typically have any solution where all objectives can reach their individual optima. Instead, there are several so-called Pareto optimal solutions with different tradeoffs. Problems can be solved with different methods and we usually need preference information from a decision maker (DM) having domain expertise to find the best balance between the objectives. Methods can be classified based on the role of the DM in the solution process in no preference methods (where no preference information is available), a priori methods (where solutions satisfying some, a priori stated, DM's preferences are found), a posteriori methods (where a representative set of Pareto optimal solutions is generated for the DM to choose from) and interactive methods (where the DM participates in the solution process iteratively) [29], [42].

Interactive methods have proven to be viable approaches to solve many kinds of multiobjective optimization problems because they enable the DM to learn about the tradeoffs involved, what kind of solutions are available and how feasible her/his preferences are (see, e.g., [4]). Furthermore, they can enhance computational efficiency since only such Pareto optimal solutions need to be generated that reflect the preferences of the DM. Many interactive methods have been proposed in the literature and they differ from each other, for example, in the way the DM expresses preference information, how information is exchanged between the DM and the method, what kind of sub-problems are formulated to get solutions based on the preference information available and what is the stopping criterion (see, e.g., [29], [44], [63], [64]). There are both scalarization-based interactive methods (see, e.g., [42], [44], [48]) and evolutionary methods (see, e.g., [5]).

Assessing the performance of interactive methods and comparing them is important to be able to find the most suitable interactive methods for various needs. First of all, we must define what we understand as "performance" of an interactive method. An interactive method is designed to aid the

DM in finding her/his most preferred solution to a multiobjective optimization problem. Therefore, in a general sense, the "performance" of an interactive method refers to how well it aids the DM in this task. This performance depends on several different aspects, and besides, many of these aspects vary depending on the type of problem solved, and on the type of the DM. Therefore, it is not possible to come up with a single performance measure. Based on the survey carried out in this paper, and on the authors' previous experiences, we have identified some aspects that characterize the performance of interactive methods, which we have named as "desirable properties".

The assessment of methods has many challenges. Because of the intensive role of the DM in interactive solution processes, it is common to say that the process stops when the DM is satisfied and confident with the final solution. However, it is not necessarily clear what this actually means. For example, in some cases, it might be advisable to make sure that the DM has experimented long enough before (s)he decides to stop. Therefore, some mechanisms are needed to prevent DMs from terminating the solution process early [65]. Moreover, the ability to determine whether the obtained solutions are satisfactory is related to the subjective preferences of the DM and the cognitive burden set on him/her. Furthermore, cognitive biases of DMs may affect the final solution [67], and, thus, assessing interactive methods involves also considering how much effect do cognitive biases have on the final solution [66].

Comparing interactive methods is not simple because the DM plays an important role and learns during the solution process and, thus, the order in which different methods are applied affects the results. To compensate this, one would need a large number of DMs to apply methods in different orders. Because the DM must have appropriate domain expertise and feel the responsibility of the final solution, in many real problems, it is not possible to have such a large number of DMs available. Naturally, students can act as DMs but only if the problems to be solved have been formulated so that the students are genuinely responsible for the final solution.

According to [63], we call interactive methods non ad hoc ones, if a utility or value function can play the role of the DM. Such methods can be compared without human DMs. However, this does not necessarily represent all properties relevant to human behaviour like anchoring, cognitive biases or the need to change the preferences thanks to learning. On the other hand, ad hoc methods cannot be compared even if a utility or value function was available.

To avoid the need of having (large numbers of) DMs, artificial DMs have been introduced recently for comparing interactive methods in [2], [28], [51]. However, they do not yet capture all relevant elements and we need to develop new performance metrics to be applied with them.

In this research, we investigate how the performance of interactive methods has been assessed in the published literature. We consider learning and decision phases of interactive solution processes separately [48], since they have different objectives. The DM explores various solutions to find a region of interest in the learning phase and converges to find the most preferred solution in the identified region of interest in the decision phase. Thus, the performance metrics for the different phases should reflect their special properties and objectives.

Despite the practical importance of assessing and comparing interactive methods, this research area has not been studied yet much. In 1992, Olson [52] reviewed published studies where some multiobjective optimization methods were applied with human DMs. The impact of learning on DMs was also considered. In 1996, Aksoy *et al.* [1] provided a survey of the state of the art studies which included comparisons of interactive methods, where both human DMs were involved (6 studies) and utility functions were used to simulate DMs (8 studies). Recently, Xin *et al.* [71] provided a systematic taxonomy to distinguish representative interactive methods of both scalarization based and evolutionary types. To characterize different methods, they identified four factors: *"interaction pattern, preference information, preference model and search engine"*. They also addressed several key issues in interactive methods from different perspectives, such as the DM and interaction of the DM and the algorithm.

In this survey, we concentrate on published experiments (i.e., assessments or comparisons of interactive methods). To this end, we have surveyed the literature published in English over the past 20 years to find papers where interactive methods have been assessed. Since most of these papers focus on assessing a single interactive method (which is proposed in the corresponding paper) and most experiments have been conducted without the involvement of real DMs, we have included some older papers surveyed in [42], where various interactive methods were compared by several human DMs. The total number of papers considered is 45 and the total number of experiments covered is 48.

We have surveyed what has been done relative to assessing the performance of interactive methods in the literature. More specifically, we have identified the aspects that have actually been assessed, and the way they have been assessed, paying attention to the type of experiments used, the type of DM, the preference information required by the methods and other relevant features of the assessments or

comparisons of interactive methods.

As previously mentioned, based on the surveyed studies and on our own experiences on the use of interactive methods, we identify and discuss some desirable properties of an interactive method at the end of the paper. We also show which of these properties have already been assessed and identify others that we believe should be assessed. Once the different properties have been identified and characterized, the next step in future research works will be to formulate indicators to measure the performance and use them in assessing or comparing methods. Examples of aspects to be considered include how well are the preferences reflected in the solutions generated, how many iterations with the DM are needed, can we guarantee Pareto optimality, ability to consider different parts of the Pareto optimal set, DM's confidence in the final solution, insight gained during the solution process, etc.

The rest of this paper is structured as follows. In Section 2, we outline the main concepts to be used. The literature review on assessing and comparing interactive methods is summarized in Section 3, where papers are divided in three classes based on the type of experiments conducted. In Section 4, we discuss desirable properties characterizing good interactive methods, point out which of them have been assessed in the experiments reported, and include some thoughts about experiment setting issues. Finally, we conclude in Section 5 and mention some future research directions.

2 Key Concepts and Terminology

We consider multiobjective optimization problems, where k objective functions are to be optimized simultaneously. They are functions of decision variables in the decision space. We denote the number of continuous variables by n and the number of integer variables by i. The k-dimensional vectors of objective function values corresponding to feasible values of the decision variables are called objective vectors in the objective space. The objective functions and the constraints defining a feasible region for the decision variables can be e.g., convex, concave or differentiable. If this is the case, this can be taken into account in the solution process. Otherwise, we do not make any assumptions of their properties.

As mentioned in the introduction, interactive methods have many desirable properties because of which they have been applied in various real problems. They allow the DM to consider a limited amount of information at a time, which keeps the cognitive burden on a tolerable level. And they limit the computation cost since only solutions that are interesting to the DM are calculated. Because of the iterative nature, the DM can learn about the relationships (tradeoffs) among the different objectives and, thus, gain valuable insight about the phenomena involved. The DM can also change one's preferences based on the learning.

In the presence of multiple conflicting objectives, we typically cannot find a solution where all objectives can attain their individual optimum concurrently. Instead, we have different tradeoffs among the objectives. Most multiobjective optimization methods operate with Pareto optimal solutions. In them, it is impossible to improve any of the objective function values without allowing some impairment in at least one of the others. We sometimes call the set of Pareto optimal objective vectors as a Pareto front (PF).

In interactive methods, the preference information provided by the DM is incorporated in the solution process to generate (Pareto optimal) solutions that reflect the preferences. The DM takes part in the interactive solution process by providing or refining preferences iteratively. Interactive methods differ from each other, e.g., in terms of different types of preference information the DM is expected to provide at each iteration (see, e.g., [39], [42], [48], [63]). Usually, the following types of preference information are considered:

- Comparison of solutions. Two or more Pareto optimal solutions are calculated, and the DM is asked either to choose the best and/or the worst solution, or to rank them.

- Local tradeoffs. When moving from a Pareto optimal solution to another one, we must let at least one objective impair in order to improve another one. This exchange ratio between two objectives (that is, how much one should be impaired in order to allow the other one to improve by one unit, if the rest of the objectives remain constant) is usually known as a tradeoff. These tradeoffs can be defined by quotients of finite increments (finite tradeoffs), or by the corresponding derivatives (partial tradeoffs). In any case, tradeoff based methods can be of two types. First, the DM may be asked to assess certain real tradeoffs from the current iteration (or to compare several ones). Alternatively, other methods ask the DM to give her/his subjective (indifference) tradeoffs from the current iteration to calculate marginal rates of substitution (MRS).

- Aspiration levels. The DM is asked, at each iteration, to give desired levels (usually called aspiration levels or target values) for each of the objectives. Alternatively, instead of aspiration levels forming a reference point, the DM can give desirable search directions, or desirable ranges for the objectives (defined by two reference points consisting of aspiration and reservation levels).

- Classification. The DM is asked to classify, at sight of the current iteration, the objectives into several classes. These classes include objectives that should be improved (as much as possible or up to a certain aspiration level), objectives that are satisfactory at their present values, and objective that can be impaired (freely or till a certain bound). The number of classes varies among methods.

- Weights. If weights are given to different objectives, most of the times, given that the weights are given at sight of the current iteration, they have a local character and therefore, they are equivalent to giving (partial) indifference tradeoffs. But they can also have a global character, or complement some other preference information type, like, for example, giving aspiration levels and weights assessing the importance of reaching each of them.

- Bounds. The DM be asked to set bounds for certain objectives, while trying to improve the others as much as possible. In some cases, bounds are also combined with some other type of preference information.

Another important aspect of interactive methods is when to stop the solution process. On the one hand, one can incorporate a technical stopping criterion. For example, one can set a prefixed number of iterations or a given number of certain technical issues (like number of function evaluations). Most of the tradeoff based methods have proven mathematical convergence, assuming that the answers of the DM follow an implicit utility or value function. In this case, the methods incorporate convergence tests, basically consisting of estimating the improvement of the utility and stopping when it is small enough. On the other hand, the DM may freely decide to stop, when (s)he is convinced that a sufficiently satisfactory solution has been found. This should imply that the DM has learnt enough about the problem, the tradeoffs among the objectives and about her/his own preferences, to be sure that there are no significantly better solutions in the feasible region. Interestingly, it has been reported (see, e.g., [22]) that the number of iterations carried out in real applications of interactive methods is surprisingly small. This early termination effect may be due to the fact that the DM does not respond in the same way to losses and gains, which causes some decisional stress when trading off in the Pareto optimal set, as discussed in [35].

As mentioned several times, interactive methods allow a DM to provide one's preferences iteratively. This is the main advantage of these methods because of two reasons:

- Preferences have a local character because of the nature of interactive methods. That is, the preferred tradeoffs among the objectives, or their relative importance, or the desired levels for each of them depend on the part of the feasible or Pareto optimal set that is presently being explored, or the direction of the search/movement.

- The DM is expected and allowed to learn during the solution process about the structure of the Pareto optimal set of the problem, the conflict degrees and tradeoffs among the objectives, what kind of solutions are feasible, and the effect of one's preferences on the solutions obtained. In this way, (s)he is able to provide more accurate information as the process goes on. In many methods, the DM is also allowed to change one's preferences (thanks to learning) during the solution process.

For the above-mentioned reasons, and as discussed in the introduction and, e.g., in [48], two different phases can be distinguished when using an interactive method to solve a real decision problem: a learning phase and a decision phase. During the learning phase, the DM explores the set of Pareto optimal solutions, learns about the tradeoffs among the objectives, identifies areas with different conflict degrees, learns about the values of the decision variables that correspond to each area, and finally identifies the area or region of interest, within the Pareto optimal set that seems to best fit her/his preferences. On the other hand, in the decision phase, the DM refines the search within the region of interest identified earlier, until (s)he finds one's most preferred solution (MPS). In other words, the DM converges in the region of interest to the final solution.

For assessing an interactive method or comparing with another one, the necessary preference information should be provided by DMs in experiments. In this survey, we classify different types of experiments based on the type and number of DMs involved. If human DMs are involved in the experiment, there can

be a *single DM*, or *several DMs*. As mentioned in the introduction, some experimental studies have been conducted with *utility functions (UFs)* for testing non ad hoc interactive methods. In these experiments, quantitative evaluations or comparisons can be done by calculating distances between the optimal UF value and the value of the UF at the solutions obtained by the interactive method. On the other hand, UFs cannot be applied directly to replace a DM in ad hoc interactive methods, because the type of preference information that these methods need is not straightforward to get from a UF.

In this paper, we use the terminology that an *artificial DM* is used to replace a human DM in providing preferences for testing ad hoc interactive methods. In [28], the artificial DM generates preference information by a cone concerning the MPSs (which are determined before the simulation), and the angle of the cone delimits the scope of the search area. The learning of DMs is simulated by decreasing the angle during the solution process. Artificial DMs are defined in [2] and [51] by a so-called steady part and a current context. The steady part includes core preferences that do not change in time, and the current context allows changing the preference information according to the current situation. There is a mechanism that generates preference information (such as reference points) based on these two components. For the performance evaluation or comparison of the interactive methods, the distance between the steady preference information and the obtained solution is calculated or compared. Meanwhile, in [14] and [37], the authors simulated human behaviors by using different types of UFs and named their approaches as a virtual DM and a machine DM, respectively.

3 Assessments Available in the Literature

As mentioned in the introduction, we have collected assessments and comparisons of interactive multiobjective optimization methods from 45 papers and we summarize them in this section. They contain a total of 48 experiments. Next, we describe some general aspects found in the literature.

An important characteristic of the experiments reported is whether a single interactive method was tested or different methods were compared. As can be seen in Figure 1, a majority of the papers reviewed (24) concentrate on demonstrating the performance of a single interactive method, while the rest compare an interactive method with non-interactive (a posteriori) ones (3), or compare several interactive methods (21). Thus, we have three classes of experiments.

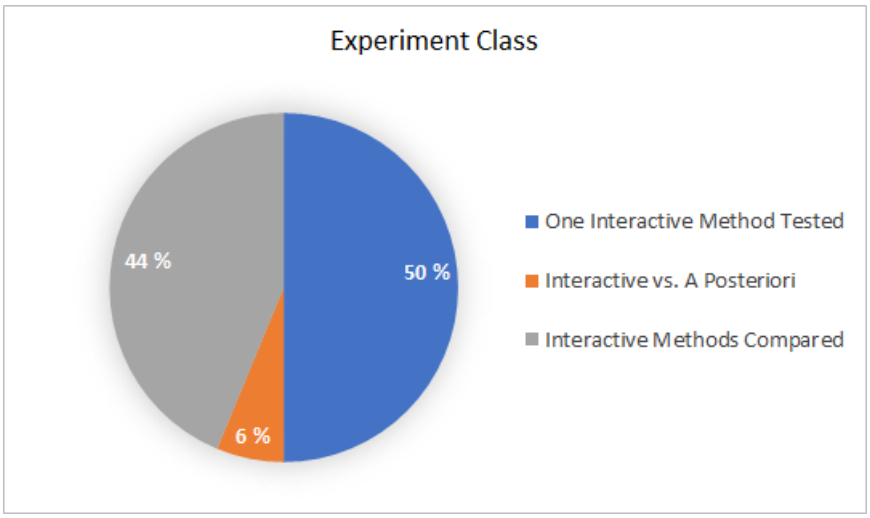

Figure 1: Experiment classes performed in the literature.

In principle, assessing the goodness of an interactive method may seem to be an easy task if it suffices to assess the goodness of the final solution obtained. However, unfortunately, defining goodness is far from trivial. For example, how do we know that a good (or the best possible) solution has been obtained? Basically, the aim of any interactive method is to support the DM in finding the one's MPS, that is, the feasible solution that best meets her/his preferences. Theoretically, we may assume that a utility or a value function perfectly defines these preferences. In such a case, the original multiobjective optimization problem can be converted into a single objective one optimizing the utility/value function, and therefore, any good multiobjective optimization method should find this optimal solution, or at least, one that is close enough to it. This is why a significant number of papers assess the performance of interactive

methods in terms of the closeness of the final solution obtained to the optimal solution of a utility or a value function.

Nevertheless, when it comes to practice, a theoretical utility/value function does not necessarily exist (or even if it does, the DM is not able to specify it). For this reason, a majority of the experiments reviewed (33 out of 48) employ DMs, either artificial (3) or human (30) ones. Out of the 30 experiments with human DMs, 18 had a single one and the remaining 12 involved several DMs, who independently assessed the method(s) considered.

As can be seen in Figure 2, the use of human DMs was the option preferred for all the classes of experiments, except for the three comparisons of interactive methods with a posteriori methods, that used utility functions and artificial DMs. It must be pointed out that in the experiments carried out with a single human DM, it was in many cases the author(s) who simulated the DM's behavior by providing different kinds of preference information. Moreover, when several human DMs were involved, they were often students or staff members of the university (only in four experiments reported, experts in the problem domain participated). Therefore, it seems to be a difficult task to test methods with real DMs who are really concerned about the final solution found.

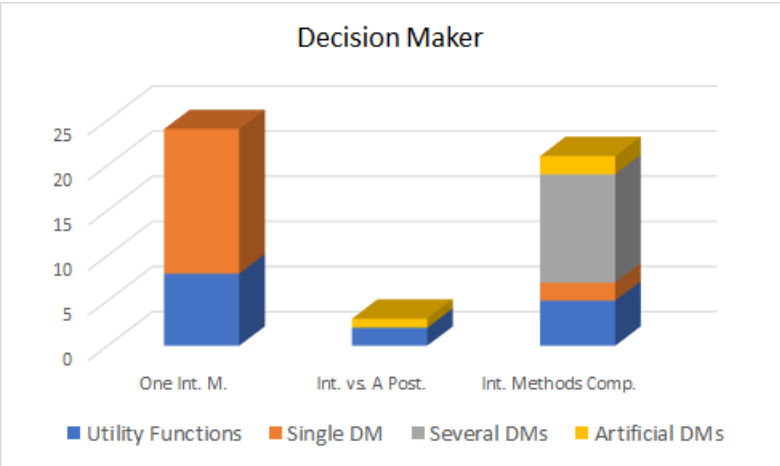

Figure 2: Type of DM.

As previously mentioned, interactive methods differ from each other, e.g., in the type of preference information that the DM provides at each iteration to direct the solution process. Therefore, it is very important to know the type of preference information considered in the experiments reported. As can be seen in Figure 3, the most popular types were:

- Choosing the best (and/or worst) one(s) in a given subset of (Pareto optimal) solutions;
- Ranking a subset of solutions;
- Performing pairwise comparisons of solutions;
- Giving desirable aspiration values (reference points) or directions of improvement for objective functions;
- Classifying objectives (improvement in some objective(s) only possible by allowing impairment in some other(s));
- Giving indifference tradeoff information to derive MRS among objectives;
- Providing weights for objective functions.
- Giving upper or lower bounds on certain objective functions.

In general, the different types of preference information used were quite similarly distributed in the three classes of experiments. One can notice that reference point related preference information (18 experiments) and choosing the best and/or worst solution(s) of a given subset of Pareto optimal solutions (20 experiments) were the most widely used preference types. Nevertheless, while the latter was used both in experiments with utility functions and human DMs, the former was never used in experiments

with utility functions, which clearly shows the difficulty of simulating this type of preference information. This is an important fact, given that these two types of preference information are widely regarded as the ones that place the lowest cognitive burden on the DM [36]. On the other hand, three experiments with artificial DMs used reference points as preference information.

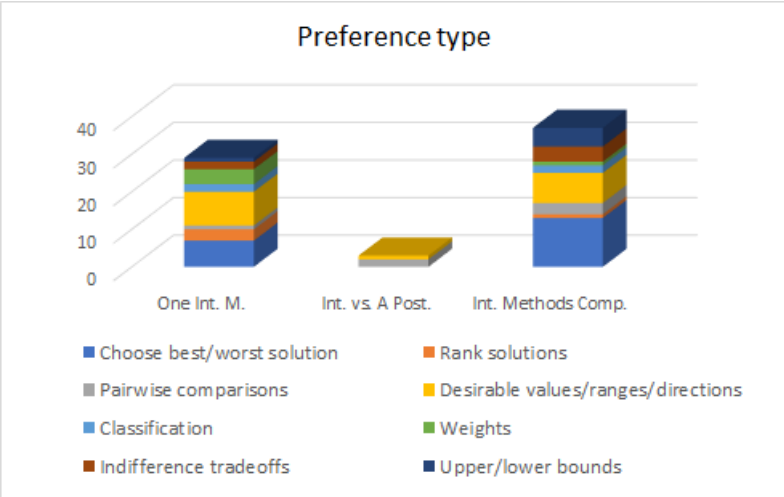

Figure 3: Type of preference information.

As discussed, a stopping criterion is an important issue. The stopping criteria used in the experiments can be seen in Figure 4. While the DM's satisfaction with the current solution was the most widely reported stopping criterion (27 experiments), it was not simulated with utility functions or with artificial DMs. Once again, this shows the difficulty of artificially simulating human behavior. Besides, the "improvement of utility" criterion (that is, being close enough to the optimal value of the utility function) was used in only 4 experiments, while the rest used criteria that could be regarded technical in nature (mostly related to the evolutionary method used) rather than criteria based on DM's preferences. For example, it is hard to believe that in a real application, a DM would be happy if the method stopped just because of a prefixed number of generations, iterations, crossovers etc. has been reached.

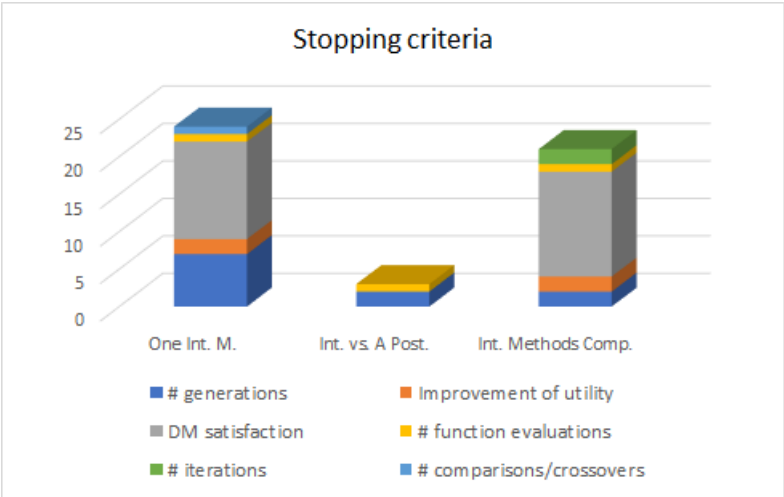

Figure 4: Stopping criteria.

A properly designed user interface is critical for the success of an interactive method in real decision problems, because it enables communication between the DM and the method. Nevertheless, according to Figure 5, nearly half of the experiments (22) did not mention a user interface. It is also interesting to note that even in the experiments that reported the use of a user interface, no assessment was made about its quality and its contribution to the satisfaction of the DM.

As mentioned at the beginning of this section, we divided the assessments considered into three classes depending on whether they demonstrate the performance of a single interactive method, compare inter-

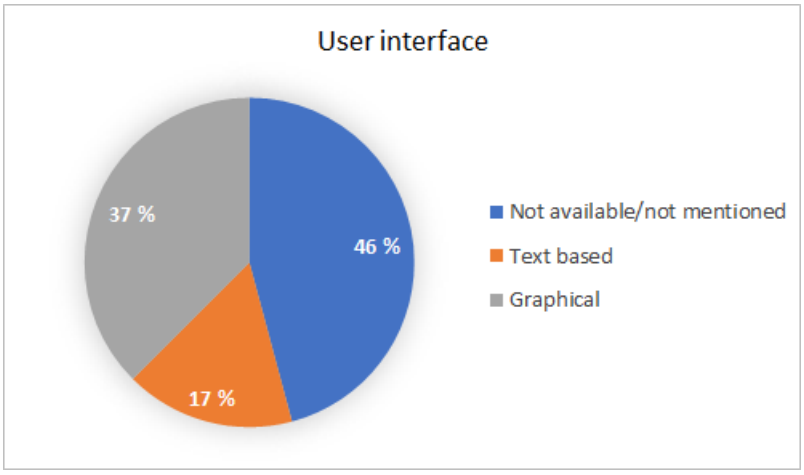

Figure 5: User interface.

active and non-interactive methods or compare several interactive methods. In the following subsections, we give further details of the experiments in each class in Tables 1, 2 and 3, respectively. In the tables, the details of the experiments are listed in the following columns: name(s) of method(s) and references, type of DM, preference type or additional information, what was measured or compared, number of iterations, stopping criterion, test problems considered, type of problems considered, order of methods considered, division between learning and decision phases made and user interface (UI) mentioned.

It is necessary to clarify what is meant by the column names in the tables. For Table 1, only the related method name and its reference are listed since there is no comparison with other methods. On the other hand, all the compared methods and the reference of the related experiment are listed in the first column of Tables 2 and 3. We list how the experiments were done in the *'Decision maker'* column, where the main options are UFs, a single human DM, several human DMs or an artificial DM. We order the rows in the tables so that similar experiments based the DM type are next to each other. Following this, in the *'Preference type'* column, only the preference types that were tested in the experiments are listed. The specific aspects that were measured or compared in the related experiments are summarized in the fourth column of the tables. For example, some statistical results from several runs, the number of iterations or some measurements of practical opinions of DMs like satisfaction, confidence, usability etc., are listed.

Then, the number of iterations carried out in the experiments and whether this number was pre-fixed or not are indicated in the *'Number of iterations(pre-fixed)'* column. If the number of iterations is not pre-fixed, we explain how this value was determined. For instance, *"3 – (No, depends on DM satisfaction)"* means that the number of iterations was 3, this value was not pre-fixed, and it was determined according to the DM's satisfaction. We report how solution processes were terminated in the *'Stopping criterion'* column. In general, DM's satisfaction or technical aspects like a fixed budget of generations or function evaluations were used as stopping criteria. The *'Test problems(# of instances)'* and *'Type of problems'* columns give information about the problems tested in the experiments. Problem names and the number of instances of the problems are given in the *'Test problems(# of instances)'* column. Problem types, number of objectives (k) and decision variables (n for continuous variables and i for integer variables) are given in the *'Type of problems'* column. There is an extra column in Tables 2 and 3 informing whether the order of the methods tested was considered or not in the experiments, in the *'Order considered'* column. Finally, in the last column (*'Learning vs. decision phase/UI'*) of the tables, we comment whether separate learning and decision phases were reported or not and the type of user interfaces. The notation '-' means that the corresponding information was not mentioned in the related reference.

3.1 Demonstrating the performance of a single interactive method

Table 1: Demonstrating the performance of a single interactive method

Method and reference	**Decision maker**	**Preference type (additional information)**	**What was measured**	**Number of iterations (pre-fixed)**	**Stopping criterion**	**Test problems (# of instances)**	**Type of problems**	**Learning vs. decision phase / UI**
I-SIBEA [15]	Human DM	Best and worst solutions of a subset, (max # of iterations, # of solutions to be compared)	-	6 - (Yes)	Max 400 generations of EA	ZDT4 (1 instance)	Convex $k = 2$, $n = 10$	No / Graphical
Interactive RVEA [24]	Human DM	Best and worst solutions of a subset, reference point, preferred ranges	-	5 - (No, depends on DM satisfaction)	DM satisfaction	Multiple-disk clutch design problem (1 instance)	$k = 5$, $n = 5$	No / Graphical
E-NAUTILUS [58]	Human DM	Best solution of a subset, (# of iterations, # of solutions to be compared)	-	5, 7 - (No, depends on DM satisfaction)	DM satisfaction	Auxiliary services of thermal power plants [55] (1 instance)	Discontinuous, Nonconvex $k = 3$ $n = 13, i = 20$	No / Graphical
[38]	Human DM	Reference point	-	3, 4 - (No, depends on DM satisfaction)	DM satisfaction	Aerodynamic airfoil shape optimization problem (3 instances)	$k = 2, 3, 6$ $n = 12$	No / -
WASF-GA [56]	Human DM	Reference point, (# of solutions to be compared)	-	3 - (No, depends on DM satisfaction)	DM satisfaction	DTLZ2 (1 instance)	Concave $k = 5$	No / Graphical
NAUTILUS Navigator [57]	Human DM	Reference point, upper bounds on objectives, speed of the movement	-	3 - (No, depends on DM satisfaction)	DM satisfaction	Auxiliary services of thermal power plants [55] (1 instance)	Discontinuous, Nonconvex $k = 3$ $n = 13, i = 20$	No / Graphical
PIE [61]	Human DM	Reference point, weights, preferred ranges, % distance to PF, (# of solutions to be compared)	-	3 - (No, depends on DM satisfaction)	DM satisfaction	Locating pollution monitoring station [45] (1 instance)	$k = 5$, $n = 2$	No / Text-based
iPICEA-g [69]	Human DM	Reference point, weights	-	(No, depends on DM satisfaction)	DM satisfaction	ZDT1, DTLZ2 (2 instances)	Convex, concave $k = 2, 4$, $n = 30$	No / Graphical

Method and reference	**Decision maker**	**Preference type (additional information)**	**What was measured**	**Number of iterations (pre-fixed)**	**Stopping criterion**	**Test problems (# of instances)**	**Type of problems**	**Learning vs. decision phase / UI**
NAUTILUS [43]	Human DM	Rank of objectives or percentages for how to improve current obj. values, (# of iterations)	-	4 - (Yes)	DM satisfaction	Pollution problem of a river (1 instance)	$k = 4$	No / Graphical
NAUTILUS 2 [47]	Human DM	Direction of improvement, (# of iterations)	-	3 - (Yes)	DM satisfaction	Pollution problem of a river (1 instance)	$k = 4$	No / -
Pareto Navigator [20]	Human DM	Classification of objectives, search direction, specifying starting point	-	4 - (No, depends on DM satisfaction)	DM satisfaction	Sample problem (1 instance)	Nonconvex $k = 3$ $n = 2$	Yes / Graphical
NIMBUS [46]	Human DM	Classification of objectives, (# of intermediate solutions, # of solutions to be compared)	-	(No, depends on DM satisfaction)	DM satisfaction	Sample problem (1 instance)	$k = 6$, $n = 2$	No / Graphical
T-IMO-EA [13]	Human DM	Indifference tradeoffs	-	3 - (No, depends on DM satisfaction)	DM satisfaction	River water quality problem [72][73] (1 instance)	Nonlinear $k = 3$, $n = 3$	No / Text-based
PROJECT [40]	Human DM	Indifference tradeoffs	-	4 - (No, depends on DM satisfaction)	DM satisfaction	DEA problem [70] (1 instance)	$k = 3$, $n = 7$	No / Text-based
FLMOEA [60]	Human DM	Relative importance of objectives	-	3 - (No)	100 generations of EA	Control system for flexible robot arm [18] (1 instance)	$k = 7$	No / Graphical
P-NSGA-II [50]	Human DM	Weights	-	4, 6 - (No)	800, 1425 generations of EA	DTLZ1, welded beam design problem (2 instances)	Linear, nonlinear $k = 2, 3$ $n = 4$	No / Graphical

Method and reference	**Decision maker**	**Preference type (additional information)**	**What was measured**	**Number of iterations (pre-fixed)**	**Stopping criterion**	**Test problems (# of instances)**	**Type of problems**	**Learning vs. decision phase / UI**
[21]	Chebyshev, linear UFs	Best and worst solutions of a subset	Average values and deviation from best found solution in several runs	3, 5, 6, 9, 11, 15 - (Yes)	20, 500 generations of EA	0/1 knapsack (6 instances)	Combinatorial, $k = 2, 3, 4$ $i = 100, 200$	No / -
iMOEA/D [23]	4 UFs	Best solution of a subset	Values of 30 runs shown on graphs	4 - (Yes)	500 generations of EA	ZDT1-2, DTLZ1-2, welded beam design problem (5 instances)	Linear, concave convex $k = 2, 3, 4$ $n = 4, 7, 12, 30$	No / -
iTDEA [31]	Chebyshev, linear, quadratic UFs	Best solution of a subset	Mean, standard deviation, absolute deviation, relative deviation of 50 runs	4, 6 - (Yes)	80000-320000 function evaluations	DTLZ1-2, ZDT4 (6 instance)	Linear, convex concave $k = 2, 3$	No / -
PI-EMO-PC [62]	Linear, nonlinear UFs	Best solution of a subset	Best, median, worst values of 21 runs - # of function evals - # of DM calls	(No, depends on stopping criterion)	Small expected improvement	Modified ZDT1, DTLZ2 (3 instances)	Nonconvex, concave $k = 2, 3, 5$ $n = 30$	No / -
BC-EMO [3]	Linear, nonlinear UFs	Ranking solutions of a subset	Median values of 100 runs	1, 2, 3 - (Yes)	500 generations of EA	0/1 knapsack, DTLZ1, 6-7 (36 instances)	Combinatorial, disconnected, linear $k = 2 - 10$, $i = 100$ $n = 2k - 10k$	Yes / -
Machine DM [37] BC-EM0	UFs	Ranking solutions of a subset	Mean values of 10 runs	3 - (Yes)	500 generations of EA	DTLZ1-2, 6-7 (8 instances)	Concave, disconnected, linear $k = 5, 7$, $n = 2k$	No / -
PI-EMO-VF [17]	Linear, nonlinear, stochastic UFs	Ranking solutions of a subset	Best, median, worst values of 21 runs, # of function evals, # of DM calls	(No, depends on stopping criterion)	Small expected improvement	Modified ZDT1, DTLZ2 (3 instances)	Nonconvex, concave $k = 2, 3, 5$ $n = 30$	No / -
IEM [54]	Chebyshev, linear UFs	Pairwise comparison	Average and standard deviation of 50 runs - # of comparisons by DM	(No, based on estimated fitness)	Predetermined # of crossovers/ population convergence	0/1 knapsack, spanning tree problem (9 instances)	Combinatorial $k = 2, 3, 4$ $i = 20, 50, 200$	No / -

3.1.1 Type of DMs

Many interactive methods considered were tested alone, rather than compared with other interactive methods. As mentioned, 24 experiments out of 48 contained no comparisons among different interactive methods. Information of these experiments is summarized in Table 1. As can be seen, UFs or a single human DM were used to provide preference information. In eight experiments, different types of UFs such as Chebyshev, linear, nonlinear or quadratic UFs were used ([3], [17], [21], [23], [31], [37], [54], [62]) (as mentioned in Section 2, we associate machine DMs of [37] to UFs). Furthermore, human DM(s) took part in the remaining 16 experiments. A single human DM participated in [13], [15], [20], [24], [38], [40], [43], [46], [50], [56]–[58], [60], [61], [69], except for one experiment [47], where a group of DMs was involved in a group decision making context. In fact, in all these experiments, the author(s) acted as DMs establishing some assumptions about possible preference information. Interestingly, a separate learning and decision phase was mentioned in only two papers [3], [20]. It should be noted that some methods are particularly suited for one of the phases. For example, navigation based methods like in [20], [27] and the NAUTILUS family ([43], [47], [57], [58]) are more fitted for the learning phase while classification based methods like [46] are well suited for the decision phase.

3.1.2 Type of preferences

Different types of preferences were tested in the experiments in this class. Although some methods had different ways to capture preference information, only one type of preference was tested in each experiment done by using UFs, such as choosing the best or worst solution in a given subset of solutions ([21], [23], [31], [62]), ranking a subset of solutions ([3], [17], [37]) or carrying out pairwise comparisons of solutions [54]. On the other hand, by simulating a human DM's responses, author(s) could sometimes test different preference types in the same experiment. The most widely used and tested preference type was specifying a reference point ([24], [38], [56], [57], [61], [69]). In [24], other types of preference information were tested: selecting preferred solutions, specifying non-preferred solutions and specifying preferred ranges for objective function values in addition to the reference point. Likewise, specifying preferred and non-preferred solutions was tested in [15]. A DM's preferences were simulated in [50] by specifying values of Gaussian functions. In [61] and [69], DMs expressed their preferences by giving both reference points and weights for the objectives. Besides, the DM was asked to give the percentage distance between the current reference point and the corresponding Pareto optimal solution, and the number of solutions to be compared in [61].

Moreover, the DM classified objectives into some classes, and gave aspiration levels and bounds for corresponding objectives in [46]. In [20], the DM gave the starting point of the navigation and specified the search direction during the navigation phase as a reference point or with classification. In [43], [47] the preference information consisted of ranking the relative importance of improving each objective value, while in [60] some grades of the relative importance of the objectives were given. The DM also expressed her/his preferences by providing some percentages indicating how the current objective values should be improved in [43]. The DM chose the best solution from the shown solutions and could decide the number of solutions to be shown and the number of iterations to be taken in [58]. The DM gave upper bounds for some objectives in addition to specifying the reference point and speed of the movement in [57]. The DM provided local indifference tradeoffs that were used to calculate MRS in the experiments in [13] and [40].

3.1.3 What was measured

A key issue of this survey is to find out what was measured in the experiments and how. Unfortunately, not many aspects have been measured in this class of experiments. On the one hand, it is apparent from Table 1 that some statistical results were measured from several runs if the experiments were carried out by using UFs. In this type of experiments, algorithms were run several times to measure the results consistently. These results were presented as best, median, mean, worst values, or standard, absolute, relative deviation over several runs. Besides these statistical results, the number of DM calls (in this case, it is the number of UF calculations) was measured and listed in some experiments ([17], [54], [62]) and the number of iterations was not pre-fixed. The number of overall function evaluations was also shown in the corresponding experiments in [17], [62]. On the other hand, in the rest of the experiments, the author(s) simulated a human DM's responses and assumed the necessary preference information, in order to demonstrate the usage of the proposed interactive methods. They had either graphical or text-based user interfaces, and they presented some screenshots showing the interaction between the DM

and the method. Since the purpose of this kind of experiments is just to demonstrate the functioning of the method and the user experience and not to evaluate the performance, the results were shown on graphs from a single run of the method.

3.1.4 Stopping criteria

Three main stopping criteria were used to terminate interactive methods in this class. The most widely used stopping criterion was DM's satisfaction, that is, the method was terminated when the DM was satisfied with the solution obtained. This criterion was used in most of the experiments involving human DMs ([13], [20], [24], [38], [40], [43], [46], [47], [56]–[58], [61], [69]). The second stopping criterion was very technical and based on an initially fixed number of function evaluations or generations ([3], [15], [21], [23], [31], [37], [50], [60]), while the last stopping criterion was based on expected improvement of the utility of the solutions ([17], [62]). Besides these main types, the method in [54] was terminated when either the maximum number of crossovers was met or 95% of the genes converged, which are very technical criteria as well.

3.1.5 User interface

As was pointed out earlier in this section, the user interface can be an extremely relevant element in the success of applying interactive methods. A total of 14 experiments out of 24 were carried out by using either text-based or graphical user interfaces. Only five papers explicitly mentioned the importance of the user interfaces and the necessary features they should have ([20], [24], [43], [46], [57]). However, none of the listed experiments in this class made any assessment to evaluate the goodness of their user interfaces.

3.2 Comparing interactive and non-interactive methods

Table 2: Comparing interactive and non-interactive methods

Compared methods and reference	**Decision maker**	**Preference type (additional information)**	**What was compared**	**Number of iterations (pre-fixed)**	**Stopping criterion**	**Test problems (# of instances)**	**Type of problems**	**Order considered**	**Learning vs. decision phase / UI**
[28]	Artificial DM	Reference point	Computational time	5 - (Yes)	MPS is reached, max # of function evaluations	Inventory routing problem (1 instance)	$k = 2$, $i = 2$	No	No / -
NEMO-I [6] NSGA-II	Chebyshev, linear UFs	Pairwise comparison	Best and average values of 4 (for NEMO-I), 10 (for NSGA-II) runs	(No, depends on stopping criterion)	200 generations of EA	ZDT1, ZDT2 (2 instances)	Convex, concave $k = 2$, $n = 30$	No	No / -
NEMO-0 [7] NSGA-II	Chebyshev, linear UFs	Pairwise comparison	Best and average values of 100 runs	(No, depends on stopping criterion)	300, 15000 generations of EA	ZDT1-2, 4, DTLZ2, WFG1 (5 instances)	Convex, concave $k = 2, 3, 4, 5$	No	No / -

The second class of assessments involves comparisons of interactive and non-interactive methods. In principle, one can question the meaningfulness of such experiments because methods of different types are compared. As can be seen in Table 2, we have only three such experiments. In [6] and [7], two interactive methods were compared with an a posteriori evolutionary algorithm. Chebyshev and linear UFs were used for pairwise comparisons of solutions. As results of comparisons, some statistical values from several runs were listed. Since UFs were used to simulate the DM's preferences in both experiments, the order of the methods was not considered. The stopping criterion was a prefixed number of generations of the underlying evolutionary algorithm (specified by the authors).

In [28], a reference point-based interactive method was compared with its a posteriori variant on a real-world problem with an artificial DM simulating the DM's responses as described in Section 2. The authors compared the computational effort as the time for finding the MPSs, which were prefixed by the authors. The interactive method was terminated when the MPS was found, objective function values were not improved further, or the maximum number of function evaluations was reached.

3.3 Comparing several interactive methods

Table 3: Comparing several interactive methods

Compared methods and reference	**Decision maker**	**Preference type (additional information)**	**What was compared**	**Number of iterations (pre-fixed)**	**Stopping criterion**	**Test problems (# of instances)**	**Type of problems**	**Order considered**	**Learning vs. decision phase / UI**
Experimental study [2] R-NSGA-II, WASF-GA	Artificial DM	Reference point	Distance of final solution and steady preference information of 31 runs (mean, standard deviation, min), # of iterations	11 - (No, depends on stopping criterion)	Max # of iterations, could not create a new reference point	DTLZ1-7 (21 instances)	Linear, concave, disconnected $k = 3, 5, 7$	No	No / -
Experimental study [51] R-NSGA-II, ASF	Artificial DM	Reference point	Distance of final solution and steady preference information of 10 runs (mean, standard deviation, min), # of iterations	11 - (No, depends on stopping criterion)	Max # of iterations, could not create a new reference point	DTLZ1-4 (24 instances)	Linear, concave $k = 2, 4, 6$	No	No / -
Nonconvex Pareto Navigator [27] PAINT + NIMBUS	Human DM	Classification of objectives, aspiration levels, specifying starting point, bounds on objectives	User experience and learning of DM	(No, depends on DM satisfaction)	DM satisfaction	Wastewater treatment plant operation [26] (1 instance)	Nonconvex $k = 5$	No	Yes / Graphical
r-NSGA-II [59] g-NSGA-II, PBEA, R-NSGA-II	Human DM	Reference point, (# of generations for each iteration, population size)	Objective values shown on graphs, additive binary indicator [74]	4 - (No, depends on DM satisfaction)	DM satisfaction	ZDT1, 3, DTLZ2 (4 instances)	Concave, convex, disconnected $k = 2, 3, 10$ $n = 12, 30$	No	No / -
Experimental study [8] STEM, Steuer	142 DMs (31 professional, 111 students	Best solution of a subset, aspiration levels, changing between methods	Satisfaction, instrumentality, usability, understandability, quality, # of changes	(No, depends on DM satisfaction)	DM satisfaction	Buying a car (6 instances)	Linear $k = 2, 4, 6$ $i = 3, 6$	No	No / -
Experimental study [9] PLANE, CONE, GUESS	24 students	Best solutions of a subset, rating solutions, upper and lower bounds on objectives	Confidence, usability, understandability, ability to capture preferences, # of iterations, elapsed time	(No, depends on DM satisfaction)	DM satisfaction	Production planning problem, river basin problem (2 instances)	Linear $k = 3$	No	No / -
Experimental study [10] GUESS, ZW	84 students (4 groups * 21)	Best solution of a subset, upper and lower bounds on objectives	Anchoring (distance from starting points)	(No, depends on DM satisfaction)	DM satisfaction	Production scheduling problem (1 instance)	Spherical $k = 3$	Yes	No / Graphical

Compared methods and reference	**Decision maker**	**Preference type (additional information)**	**What was compared**	**Number of iterations (pre-fixed)**	**Stopping criterion**	**Test problems (# of instances)**	**Type of problems**	**Order considered**	**Learning vs. decision phase / UI**
Experimental study [11] ZW, SWT, STE, GUESS	24 DMs (students and staff)	Best or worst solutions of a subset, upper and lower bounds on objectives	Confidence, usability, understandability, CPU time, elapsed time, most and least preferred method	(No, depends on DM satisfaction)	DM satisfaction	Production scheduling problem (1 instance)	Linear $k = 3$	Yes	No / -
Experimental study [12] GUESS-PM, GUESS-PU	58 students	Best solutions of a subset, reference point	Preferred method	(No, depends on DM satisfaction)	DM satisfaction	Production scheduling problem (1 instance)	Spherical $k = 3$	No	No / Graphical
Experimental study [16] GUESS, ZW, SMART	84 students (4 groups * 21)	Best solution of a subset, upper and lower bounds on objectives	Confidence, elapsed time, usability, understandability, willingness to use methods again	(No, depends on DM satisfaction)	DM satisfaction	Production planning problem (1 instance)	Spherical $k = 3$	Yes	No / Graphical
Experimental study [19] Geoffrion, GUESS	9 students	Best solution of a subset, indifference tradeoffs	Confidence, usability	(No, depends on DM satisfaction)	DM satisfaction	Buying a car (1 instance)	Linear $k = 3$, $i = 3$	No	No / Text-based
Experimental study [30] AHP, IGP, STM, WACM, ZW	5 DMs	Best solution of a subset, pairwise comparison, weights	Computer costs, usability, information load, learning effects	(No, depends on DM satisfaction)	DM satisfaction	Energy planning problem (2 instances)	Linear $k = 5, 9$	No	No / Graphical
Experimental study [34] 5 ways to give reference direction: asp. levels, BPR, AHP, MRS, unit vectors	65, 72 students	Best solution of a subset, reference direction	Satisfaction, confidence, understandability, usability, usefulness of provided information, speed of convergence, # of iterations	(No, depends on DM satisfaction)	DM satisfaction	Time allocation [32], choosing washing machine [33], buying a home [33] (3 instances)	Linear, discrete $k = 3, 5$	Yes	No / Graphical
Experimental study [41] Steuer, ZW	5 DMs	Best solution of a subset	# of iterations	3, 4, 5, 6 - (No, depends on DM satisfaction)	DM satisfaction	Production planning problem (1 instance)	Linear $k = 3$ $i = 42$	No	No / Text-based

Compared methods and reference	**Decision maker**	**Preference type (additional information)**	**What was compared**	**Number of iterations (pre-fixed)**	**Stopping criterion**	**Test problems (# of instances)**	**Type of problems**	**Order considered**	**Learning vs. decision phase / UI**
Experimental study [49] AHP, ZW	28 students	Best solution of a subset	Satisfaction, gained insight, understandability, usability, ability to capture preferences, elapsed time	(No, depends on DM satisfaction)	DM satisfaction	Resource allocation problem (1 instance)	Linear $k = 3$	No	No / Text-based
Experimental study [68] Geoffrion, STEP, unstructured approach (UN)	36 DMs (18 students, 18 managers)	Indifference tradeoffs, solution of step size problem, objective to deteriorate, max amount of relaxation, vector of objective values	Most and least preferred method, confidence, usability, understandability, usefulness of provided information, speed of convergence, CPU time	(No, depends on DM satisfaction)	DM satisfaction	Production planning problem (1 instance)	Linear $k = 3$, $i = 7$	Yes	No / Text-based
Experimental study [14] T-IMO-EA, I-SIBEA	Quadratic UFs	Best and worst solutions of a subset, indifference tradeoffs	Best, mean, worst values of 20 runs, mean running time	2, 3, 4 - (No, depends on stopping criterion)	Change in UF small enough	ZDT1, DTLZ1 (2 instances)	Convex, linear $k = 6, 10$	No	No / -
I-SIBEA [15] W-Hype	Chebyshev UFs	Best and worst solutions of a subset	Mean, standard and absolute deviation, optimal UF values of 10, 30, 50 runs	2, 4, 6, 8 - (Yes)	20000, 40000, 120000 function evals	DTLZ1, DTLZ2, ZDT4 (3 instances)	Linear, concave, convex $k = 2, 3$ $n = 7, 10, 11$	No	No / Graphical
NEMO-0 [7] IEM	Chebyshev UFs	Pairwise comparison	Best and average values of 100 runs	(No, depends on stopping criterion)	300, 15000 generations of EA	ZDT1-2, 4, DTLZ2, WFG1 (5 instances)	Convex, concave $k = 2, 3, 4, 5$	No	No / -
INSPM [53] iTDEA	3 UFs	Pairwise comparison	Average values and average running times of 50 runs	(Yes)	50 generations of EA	ZDT4 (1 instance)	Convex $k = 10$	No	No / -
T-IMO-EA [13] GRIST	Quadratic UFs	Indifference tradeoffs	Best, mean, worst values of 20 runs, mean running time	(No, depends on stopping criterion)	Change in UF small enough	GLT5, DTLZ1 (3 instances)	Linear, nonlinear $k = 3, 5$	No	No / Text-based

3.3.1 Type of DMs

Lastly, we summarize comparisons of several interactive methods. Table 3 provides detailed information about these experiments. As shown, the most preferred experiment type was using several human DMs ([8]–[12], [16], [19], [30], [34], [41], [49], [68]). In general, students were used as DMs and professionals were included together with students in only two experiments ([8], [68]). In [30] and [41], only domain experts were used. Generally, the DMs were divided into groups, which solved problems with interactive methods in different orders to avoid cognitive biases such as anchoring and learning. Interestingly, we can see that this kind of experiments have not been published in recent years. By using human DMs, the authors could compare interactive methods from the practical applicability point of view. On the other hand, in order to compare the computational performances of the methods, UFs were used to simulate DMs in five experiments ([7], [13]–[15], [53]). (As mentioned in Section 2, we associate virtual DMs of [14] to UFs). Artificial DMs, as described in Section 2, were used to simulate the preference information of DMs in two experiments in order to compare reference point based interactive methods ([2], [51]). Finally, the last experiment type in this class was using a single human DM ([27], [59]).

3.3.2 Type of preferences

Let us next consider the types of preference information tested. In a majority of experiments, DMs were asked to choose the best (most preferred) and/or the worst (least preferred) solution from a subset of solutions shown ([8]–[12], [16], [19], [30], [34], [41], [49]). The same preference type was tested by using UFs in [14] and [15]. In addition to this preference type, DMs were also asked to specify upper and lower bounds for objective function values in some experiments ([9]–[11], [16]). DMs were allowed to change from one method to another while solving a problem in [8], and they could also specify aspiration levels. In [10], several solutions representing the set of Pareto optimal solutions were shown to the DMs who were asked to specify a rating of each shown solution using a 20-point cardinal scale.

In [34], five different ways for specifying reference directions were tested (aspiration levels, the boundary point ranking method (BPR), the analytic hierarchy process (AHP), MRS and the use of unit vectors). A tradeoff type of preference information can be found in [68], where DMs were asked to provide indifference tradeoffs to estimate MRSs between objectives for one method, to choose the objective to deteriorate for another method and to enter a vector of objective values for an unstructured approach. In [27], a human DM specified a starting point of the navigation method as well as aspiration levels and bounds of objectives and compared the solution process with a combination of an approximation method and a classification based method. Several reference point-based methods were tested by specifying reference points by artificial DMs ([2], [51]), or by human DMs ([12], [59]). To simulate the preferences of the DMs, some UFs were used for pairwise comparisons in [7], [53], and to calculate MRSs in [13], [14].

3.3.3 What was compared

To answer the question of what was compared, we can say that it highly depended on the experiment type. If UFs were used, some computational performance indicators were compared, such as best, mean or worst values of UF values from several runs or deviations from the best found solution over several solution processes ([7], [13]–[15], [53]). In [13], [14] and [53], average times of solution processes were also compared. We have 2 experiments done by using a human DM. In [27], two solution processes with different interactive methods were compared with a domain expert as a DM. As a result, it was concluded that the DM learned better and gained more insight into the problem by using a navigation method. Preference information such as reference points, the number of generations for each iteration and the population size of the underlying evolutionary algorithm to compare reference point based interactive methods were provided in [59]. The comparison was made by considering graphical representations of the solutions obtained and evaluating the performance indicator called additive binary indicator [74].

Furthermore, several measures related to the practical applicability were compared by using human DMs. In these experiments, DMs were asked to evaluate the methods by rating several aspects on a scale ranging from, e.g., 1 to 10. In fact, DMs were asked to determine some scores for each type of measure after the experiments. For example, a question like *"How much were you satisfied with the final solution?"* was asked and the DM gave a score as a response. Since this kind of measure is directly related with the personality or the background of a DM, several DMs were involved in these experiments. The following features were measured and compared in this type of experiments:

- Satisfaction or confidence in the final solution [8], [9], [11], [34], [49], [68].

- Confidence in the method [16], [19], [34].
- Ease of understanding the method [8], [9], [11], [16], [34], [49], [68].
- Ease of using the method [8], [9], [11], [16], [19], [30], [34], [49], [68].
- Ability to capture preferences [9], [49].
- Usefulness of the provided information to aid the DM [34], [68].
- Most and least preferred method [11], [12], [68].
- Willingness to use the methods again [16].
- Instrumentality of the methods to achieve a solution [8].
- Anchoring (distance from the starting points) [10].
- "Experienced" speed of convergence [34], [68].
- Information load [30].
- Learning effects [30].
- Gained insight of the problem [49].
- CPU time (computation time) [11], [30], [68] or elapsed time (total time for solving the problem) [8], [9], [11], [16].

3.3.4 Stopping criteria

When we examine the experiments in terms of stopping criteria, we observe that 1) if experiments were done with human DMs, methods were terminated if the DM was satisfied with the obtained solution ([8]–[12], [16], [19], [27], [30], [34], [41], [49], [59], [68]), 2) if experiments were done by using UFs, methods were terminated based on a fixed budget of function evaluations or generations ([7], [15], [53]) and 3) if experiments were done by using artificial DMs, methods were terminated if the artificial DM was unable to generate a new reference point or if the maximum number of iterations had been reached ([2], [51]). In addition to these stopping criteria, methods were terminated if the value of UF did not improve along a tradeoff direction, or a change of the UF value was smaller than a threshold in [13], [14].

3.3.5 User interface

As we discussed earlier, designing a good user interface is important for the success of applying interactive methods. As can be seen in Table 3, user interfaces, either graphical ([10], [12], [15], [16], [27], [30], [34]) or text-based ([13], [19], [41], [49], [68]), were mentioned in 12 experiments out of 21. When we analyzed the experiments that included comparisons of several interactive methods, we could not see any assessments of the user interfaces.

4 Discussion

In this section, we discuss several important aspects of the assessments of interactive methods. We first propose in Section 4.1 desirable properties of interactive methods by collecting them from the papers considered in the previous section and using our own experiences. We consider learning and decision phases separately as well as properties that are common for both phases. In Section 4.2, we then indicate which of the properties have already been measured in the papers surveyed. We also discuss some challenges of assessing interactive methods.

4.1 Desirable properties of interactive methods - What should be measured?

To be able to assess the performance of an interactive method or compare several methods, we must characterize what we regard desirable. Thus, we must decide what makes an interactive method good. Because learning and decision phases have somewhat different objectives, they do naturally have their own desirable properties. In addition, we have general desirable properties that are independent of the phase of the solution process.

- General properties (GP)

GP_1 - The method captures the preferences of the DM.

GP_2 - The method sets as low cognitive burden on the DM as possible.

GP_3 - A user interface supports the DM in problem solving.

GP_4 - The DM feels being in control while interacting with the method.

GP_5 - The method prevents premature termination of the overall solution process.

- Learning phase (LP)

LP_1 - The method helps the DM avoid anchoring.

LP_2 - The method allows exploring any part of the Pareto optimal set.

LP_3 - The method easily changes the area explored as a response to a change in the preference information given by the DM.

LP_4 - The method allows the DM to learn about the conflict degree and tradeoffs among the objectives in each part of the Pareto optimal set explored.

LP_5 - The method properly handles uncertainty of the information provided by the DM.

LP_6 - The method allows the DM to find one's region of interest at the end of the learning phase.

- Decision phase (DP)

DP_1 - The method allows the DM to be fully convinced that (s)he has reached the best possible solution at the end of the solution process.

DP_2 - The method reaches the DM's MPS.

DP_3 - The method allows the DM to fine-tune solutions in a reasonable number of iterations and/or reasonable waiting time.

DP_4 - The method does not miss any Pareto optimal solution that is more preferred (with a given tolerance) for the DM than the one chosen.

Each desirable property has a particular purpose. To be clear, we briefly describe the listed desirable properties. The purpose of GP_1 is to determine the ability to capture DM's preferences correctly. Hence, an interactive method should adequately react to DM's preferences. GP_2 is valid for both the learning and the decision phases since the cognitive burden set on the DM may affect the whole solution process and the final solution, correspondingly. The method should not make the DM tired or confused during the interactive solution process. In order to support the DM, the method should provide useful information to the DM via a user interface which is mentioned in GP_3. Besides, with the help of the method, the DM should be aware of what is going on during the whole solution process without feeling to be losing the control of the method (GP_4). Each phase of the solution process should not be terminated before the DM finds the region of interest at the end of the learning phase or is satisfied with the final solution at the end of the decision phase (GP_5). In other words, the method should not make the DM want to stop iterating before (s)he is sure about having obtained a good enough final solution.

As far as the learning phase is concerned, a DM may not have a clear idea about the tradeoffs of the problem at the beginning of the solution process. Therefore, the starting solution may affect the final decision of the DM if (s)he does not learn enough (LP_1). Accordingly, (s)he may want to explore the search area (LP_2) in order to gain insight into the problem in the learning phase. For this purpose, LP_3 deals with the change in the preference information by changing the explored area efficiently. During the solution process, the DM may provide inconsistent or uncertain preference information and this requires special attention which is referred to the property LP_5. Finally, the purpose of the learning phase is to find the region of interest after understanding the problem details (LP_6).

As mentioned earlier, assessing the goodness of an interactive method is quite challenging due to the difficulty of designing an objective performance metric. Therefore, both qualitative and quantitative, objective and subjective assessments are needed to find the most suitable interactive method based on the needs of a DM. Here, DP_1 is about the qualitative and subjective appreciation of the DM about her/his confidence and satisfaction in the final solution. On the other hand, a quantitative goodness is emphasized in DP_2 and can be measured by UFs, artificial DMs and human DMs. Since the DM learned and gained insight in the learning phase, (s)he should be sure about her/his region of interest at the beginning of the decision phase. Therefore, the method should allow the DM to refine the solution within this region of interest in a reasonable time and a number of solutions (DP_3). Of course, it is highly important not to miss any other Pareto optimal solution which would be better than the one selected by the DM (DP_4).

4.2 General considerations of assessing interactive methods

Most of the desirable properties listed in the previous subsection were not measured in the literature reviewed in Section 3. Actually, only GP_1, GP_2, LP_1, LP_4, DP_1, DP_2 and DP_3 were measured (see Table 4).

The anchoring effect was measured in [10] by comparing the Euclidean distances of starting points and final solutions. Conclusively, starting points of the solution processes affected the final solution. Therefore, selecting a starting point of an interactive method is an important issue. Otherwise, the DM can get stuck in local optima or cannot explore the objective space sufficiently in the learning phase. As a consequence, the DM may not be able to learn well enough about the problem and (s)he may terminate the solution process with unsatisfactory solutions. In order to avoid the anchoring effect, an inferior point is proposed as a starting point in the NAUTILUS family where the DM can see improvement of each objective simultaneously at each iteration ([43], [47], [57], [58]).

In [9], [68] the authors asked the DMs to give some points for qualitative questionnaires regarding the method's ability to capture the preferences of the DMs. Capturing preferences sufficiently and providing adequate information to aid the DM are crucial issues that should be assessed for interactive methods. The more useful information for the DM is shown and the better capabilities to capture different types of preferences of the DM there are, the better the method is.

Questionnaires were also used to measure the DM's satisfaction with the obtained solution in [8], [9], [11], [34], [49], [68]. There was no comparison or measurement of the DM's satisfaction for most of the other methods, which use DM's satisfaction as a stopping criterion. Instead of carrying out experiments with human DMs, the authors usually acted as DMs and provided the necessary preference information to demonstrate the usage of their methods. In this type of experiments, the authors claimed a solution as a final one. Measuring the DM's satisfaction or confidence in the final solution is one of the most challenging issue in assessing and comparing interactive methods. There should be more systematic experiments done by several human DMs to understand the confidence or satisfaction level of the obtained solution or the method itself.

It can be easily observed from the experiments that the number of iterations or time spent in the experiments was generally measured and compared since these values can be collected from the experiments effortlessly. However, many real DMs have limited time available to solve optimization problems. Therefore, the number of iterations with the method and the total time spent and/or waiting time between each iteration should be reasonable.

In Section 4.1, we listed desirable properties of interactive methods. We count the number of experiments of our survey where they have been assessed, and suggest whether they can be measured with human DMs or replacing humans by UFs or artificial DMs in Table 4. The properties are given with reminders (in parenthesis) which are compact ways of representing the actual properties. The symbol ('✓') indicates that the property in question was measured or can be measured. We use a question mark ('?') if it is not clear whether the property can be measured by the corresponding type of DMs.

As previously mentioned, non ad hoc interactive methods can be assessed or compared by using UFs. Naturally, experimenting with UFs requires less resources and time than experimenting with human DMs [1]. However, it is not possible to use UFs for comparing methods employing all types of preference information. Therefore, artificial DMs have been proposed in [2], [28], [51] to test reference point based interactive methods, which are ad hoc by nature.

Experimenting with UFs and artificial DMs enables controlling the experimental setting and repeating experiments if needed. Besides, one can avoid problems arising from human nature such as cognitive biases, tiredness and learning transfers [28]. However, the usefulness and the practical applicability of the

Table 4: Suggested ways of measuring desirable properties

Properties	**# of experiments**	**Human DMs**	**UFs**	**Artificial DMs**
GP_1 (Capturing preferences)	2	✓	✓	✓
GP_2 (Cognitive burden)	1	✓		
GP_3 (User interface)	-	✓		
GP_4 (Being in control)	-	✓		
GP_5 (Early termination)	-	✓	?	?
LP_1 (Anchoring)	1	✓	✓	✓
LP_2 (Exploring PO)	-	✓	✓	✓
LP_3 (Changing area)	-	✓	?	✓
LP_4 (Learning)	2	✓		✓
LP_5 (Uncertain preference)	-	✓	✓	✓
LP_6 (Region of interest)	-	✓	✓	✓
DP_1 (Convinced)	6	✓		
DP_2 (MPS)	18	✓	✓	✓
DP_3 (Iterations / waiting time)	8 / 10	✓	✓	✓
DP_4 (Not missing PO)	-	✓	✓	✓

interactive methods can hardly be tested by using UFs. Furthermore, formulating UFs mathematically is not an easy task. Even if we assume that the DM can specify a UF, it may change since the DM might gain more insight of the problem over time [42]. Therefore, in order to compare interactive methods by using UFs, there is a need of constructing more types of UFs that take into account various cognitive biases and behaviors of human DMs. These issues are also valid for artificial DMs. Except for the positive feature that artificial DMs can be programmed to learn and change over time.

As previously discussed, several important issues can only be properly assessed by humans. Of course, a single DM, if really involved with the problem solved, can give her/his opinions about the interactive method but a single opinion, referring to a single problem, is not enough to assess the quality of a method. Moreover, because a human DM is involved in the solution process, the performance of the method depends on the personality of the DM. Human DMs have subjective preferences and their feelings, morale or tiredness may affect the solution process directly. Besides, assessing the interactive method or comparing several ones might not be fair because of the cognitive biases such as learning and anchoring as mentioned before. Therefore, we believe that experiments with several human DMs must be conducted to measure some properties. From the previous experiences with this kind of experiment reported in the literature, we learn that the following aspects should be carefully addressed:

- *Who acts as a DM?* Ideally, DMs should be experts in the problem domain considered in testing methods, but it seems nearly impossible to involve the number of experts that these experiments would need. Instead, some experiments reported in the literature fall back on students as DMs, which seems a more reasonable way to find the appropriate number of persons. Nevertheless, in this case, some effective way needs to be found to assure the students' involvement with the problems solved. That is, the students must have a sufficient knowledge about the problems, and they must be concerned about finding a good enough final solution.

- *Several problems and several methods.* It seems reasonable to think that a single problem may have special features that make certain methods more suitable than others. Therefore, we believe that several problems should be considered in the experiments. On the other hand, several methods are usually compared. This multiplicity of problems and methods brings new decisions, regarding how many problems should each DM solve, and with how many methods. If a single DM solves the same problem with different methods, there is an unavoidable learning effect that may alter the results. In this case, it is important to have other DMs testing the same methods in a different order. On the other hand, it may be hard to have a single student properly involved with several problems. Therefore, these decisions must be properly made depending on the number of methods to be tested, the number of problems used, and the number of students and their knowledge and involvement with the problems.

- *What to ask.* If the above experimental setup can be created, then we need to specify questions that DMs are asked. These questions must be understandable and easily answerable, since the DMs

should feel comfortable to answer them. Separate questions should be asked for different features of interactive methods. Moreover, answering should not take much of DM's time.

4.3 Some aspects on selecting an interactive method

As mentioned in the introduction, supporting the selection of an appropriate interactive method for various needs was one of the objectives of this research. We have identified many aspects that deserve further study. In general, the choice of the method depends at least on the availability of the DM, the type of preference information the DM can provide, desires of the DM in terms of what kind of information should be available during the solution process and characteristics of the problem to be solved. A matter which deserves further study is the possibility of changing the method during the solution process. One can, for example, apply first a method for learning, where a high accuracy of solutions is not necessary and then switch to a different method in the decision phase to fine-tune the solution with accurate calculations.

As said, many aspects must be considered when choosing an adequate interactive method for a given problem. Having access to an accurate assessment of these aspects for different methods becomes, therefore, crucial. In this paper, we have identified how such assessments have been carried out in the published literature, and we have found some issues that have not been measured yet. For this reason, we feel that there is still much to be done in the assessment field, before reliable recommendations of interactive method can be made. Nevertheless, based on the lessons learned in this paper, we can identify some aspects that need to be taken into account when selecting a method.

As mentioned before, aspects related to the DM should be taken carefully in selecting a method. The availability of the DM has to be checked first. In real applications, DMs may have limited time to be involved in interactive solution processes. In this case, methods that have a reasonable number of iterations and waiting time (DP_3) should be considered. In problems where the function evaluations are time consuming, methods that start with a pre-generated set of solutions are recommendable.

In these methods, time-consuming optimization is mainly conducted before involving the DM in the interactive solution process. Hence, the waiting times are significantly shortened during the solution process.

Another important aspect regarding the DM is the type of preference information depending on the DM's needs and previous experiences. Providing preference information in a preferred way (GP_1) is highly desirable, because the DM can easily understand what is needed from her/him and this enables her/him to better control the interactive process (GP_4). Otherwise, the cognitive burden set on the DM (GP_2) can increase and affect the quality of the solution process and thus, the final solution.

One more important aspect is the experience of the DM regarding the problem to be solved. If the DM is not very familiar with the trade-offs in the problem, selecting a method suitable for the learning phase is needed. Hence, exploring different parts of the Pareto optimal set (LP_2), changing the area explored (LP_3) easily, and learning about the conflict degree and tradeoffs among the objectives (LP_4) helps the DM in gaining more insight about the problem. Furthermore, if the DM does not have deeper knowledge about the problem to be solved, (s)he may not be aware of whether her/his preferences are feasible or not, and therefore, (s)he may provide uncertain or imprecise preference information. In this case, support is needed in dealing with uncertain preferences (LP_5). On the other hand, if the DM has experience with solutions obtained using other methods, a method that can provide support in avoiding anchoring (LP_1) may be useful.

As mentioned, besides the DM, also characteristics of the problem to be solved affect the choice of the method. First of all, the problem may have different types of decision variables. Accordingly, the selected method should incorporate an appropriate solver for the variable type. Second, if the problem has some constraints, the method must be able to handle them efficiently. Third, the number of objectives in the problem influences the computational efficiency. The more objectives in the problem, the higher the computational complexity typically is. As the computational complexity increases, it becomes more difficult and time-consuming to generate solutions for each iteration. During the solution process, waiting for a long time to see some solutions may increase the cognitive burden of the DM. Besides, the DM may have limited time or a deadline to reach, as already mentioned. Therefore, methods that apply different types of surrogate models which are computationally efficient should be selected to decrease the waiting times and the cognitive burden of the DM. Besides, the shape of the Pareto front is also an issue to be considered. Some methods are not able to handle complex shapes (like nonconvex or disconnected ones) that can occur in practice.

Last but not least, one should select a method that has an implementation available. Without a proper

implementation and a user interface that supports the DM and eases the communication/interaction, it is naturally difficult to apply an interactive method to any problem. Naturally, efficient visualizations (GP_3) are needed to decrease the cognitive burden of the DM, especially as the number of objectives of the problem grows. In this, different types of visualizations are important, since coordinated multiple views can complement each other [25] and different DMs may be comfortable with different visualization types.

5 Conclusions

In this paper, we concentrated on assessing and comparing interactive methods and conducted a related literature survey. After discussing challenges involved, we summarized findings from 45 papers covering 48 numerical experiments.

The experiments were classified to those demonstrating a single interactive method, those comparing an interactive and an a posteriori method and those comparing several interactive methods. We collected information about the type of experiments conducted and performance criteria involved, type of DMs involved and nature of problems considered, among others, and analysed the findings.

Finally, we characterized desirable properties of interactive methods to be able to use them in comparisons. We devoted attention to different objectives of learning and decision phases of interactive solution processes. We also suggested types of DMs that can be applied in measuring the properties.

This paper is aimed at supporting the task of assessing and comparing interactive methods. As seen, new performance metrics are needed to assess interactive methods from different perspectives, taking into account the desirable properties of interactive methods.

One must note that the performance of a given interactive method may depend on many factors, including the technical features of the problem to be solved, and also, of course, the type and personality of the DM involved. Different DMs may not only obtain different final solutions, but may also assess the performance of a method in a different way. As a result, giving an absolute measure of a method's performance does not seem realistic and, therefore, being able to say that a given method is better than another one is also, in general, too optimistic. While assessing the performance of an interactive method is a challenging issue, we believe that the research carried out in this paper gives relevant information to researchers in the field, about what has been done in this respect (and how), and what still needs to be done. Nevertheless, the subjective nature of decision making has necessarily to be taken into account in any practical instance.

In summary, we have surveyed what has been done relative to assessing the performance of interactive methods. We answered questions such as; "what has been done about assessing interactive methods?", "what have they measured and how" and "what could be measured?". How to measure the desired properties identified remains an open question, which we intend to undertake as future research. In the long run, well-designed experiments should help in selecting or recommending an appropriate method for different problem and DM types. Then, it can be possible to compare interactive methods in a controlled environment by using artificial DMs with the new performance metrics. In the future, we hope to switch focus to studying solution processes instead of individual methods, where the DM can change the preference type and method as needed.

Acknowledgements

This research is a part of developing an open source software framework DESDEO (desdeo.it.jyu.fi) for interactive multiobjective optimization methods and was partly funded by the Academy of Finland (grants 287496 and 322221). The research is related to the thematic research area DEMO (Decision Analytics utilizing Causal Models and Multiobjective Optimization, jyu.fi/demo) of the University of Jyvaskyla. F. Ruiz's research was partially supported by the Spanish Ministry of Economy and Competitiveness (project ECO2016-76567-C4-4-R) and by the EU ERDF operative program (project UMA18-FEDERJA-065).

References

[1] Y. Aksoy, T. W. Butler, and E. D. Minor III, "Comparative studies in interactive multiple objective mathematical programming", *European Journal of Operational Research*, vol. 89, no. 2, pp. 408–422, 1996.

[2] C. Barba-González, V. Ojalehto, J. Garcıa-Nieto, A. J. Nebro, K. Miettinen, and J. F. Aldana-Montes, "Artificial decision maker driven by PSO: An approach for testing reference point based interactive methods", in *Proceedings of the International Conference on Parallel Problem Solving from Nature, Part I*, A. Auger, C. M. Fonseca, N. Lourenço, P. Machado, L. Paquete, and D. Whitley, Eds., Cham: Springer, 2018, pp. 274–285.

[3] R. Battiti and A. Passerini, "Brain–computer evolutionary multiobjective optimization: A genetic algorithm adapting to the decision maker", *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 5, pp. 671–687, 2010.

[4] V. Belton, J. Branke, P. Eskelinen, S. Greco, J. Molina, F. Ruiz, and R. Słowiński, "Interactive multiobjective optimization from a learning perspective", in *Multiobjective Optimization: Interactive and Evolutionary Approaches*, J. Branke, K. Deb, K. Miettinen, and R. Slowinski, Eds., Berlin, Heidelberg: Springer, 2008, pp. 405–434.

[5] J. Branke, K. Deb, K. Miettinen, and R. Slowiński, Eds., *Multiobjective Optimization: Interactive and Evolutionary Approaches*. Berlin, Heidelberg: Springer, 2008.

[6] J. Branke, S. Greco, R. Słowiński, and P. Zielniewicz, "Interactive evolutionary multiobjective optimization driven by robust ordinal regression", *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 58, no. 3, pp. 347–358, 2010.

[7] J. Branke, S. Greco, R. Słowiński, and P. Zielniewicz, "Learning value functions in interactive evolutionary multiobjective optimization", *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 1, pp. 88–102, 2014.

[8] K. Brockhoff, "Experimental test of MCDM algorithms in a modular approach", *European Journal of Operational Research*, vol. 22, no. 2, pp. 159–166, 1985.

[9] J. T. Buchanan, "An experimental evaluation of interactive MCDM methods and the decision making process", *Journal of the Operational Research Society*, vol. 45, no. 9, pp. 1050–1059, 1994.

[10] J. T. Buchanan and J. Corner, "The effects of anchoring in interactive MCDM solution methods", *Computers & Operations Research*, vol. 24, no. 10, pp. 907–918, 1997.

[11] J. T. Buchanan and H. G. Daellenbach, "A comparative evaluation of interactive solution methods for multiple objective decision models", *European Journal of Operational Research*, vol. 29, no. 3, pp. 353–359, 1987.

[12] J. Buchanan and L. Gardiner, "A comparison of two reference point methods in multiple objective mathematical programming", *European Journal of Operational Research*, vol. 149, no. 1, pp. 17–34, 2003.

[13] L. Chen, B. Xin, and J. Chen, "A tradeoff-based interactive multi-objective optimization method driven by evolutionary algorithms", *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 21, no. 2, pp. 284–292, 2017.

[14] L. Chen, B. Xin, J. Chen, and J. Li, "A virtual-decision-maker library considering personalities and dynamically changing preference structures for interactive multiobjective optimization", in *Proceedings of the 2017 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, 2017, pp. 636–641.

[15] T. Chugh, K. Sindhya, J. Hakanen, and K. Miettinen, "An interactive simple indicator-based evolutionary algorithm (I-SIBEA) for multiobjective optimization problems", in *Proceedings of the 8th International Conference on Evolutionary Multi-Criterion Optimization*, A. Gaspar-Cunha, C. Henggeler Antunes, and C. C. Coello, Eds., Cham: Springer, 2015, pp. 277–291.

[16] J. L. Corner and J. T. Buchanan, "Capturing decision maker preference: Experimental comparison of decision analysis and MCDM techniques", *European Journal of Operational Research*, vol. 98, no. 1, pp. 85–97, 1997.

[17] K. Deb, A. Sinha, P. J. Korhonen, and J. Wallenius, "An interactive evolutionary multiobjective optimization method based on progressively approximated value functions", *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 5, pp. 723–739, 2010.

[18] R. C. Dorf and R. H. Bishop, *Modern Control Systems*, 8th ed. Boston: Addison-Wesley Longman, 1998.

[19] J. S. Dyer, "An empirical investigation of a man-machine interactive approach to the solution of the multiple criteria problem", in *Multi Criteria Decision Making*, J. Cochrane and M. Zeleny, Eds., Columbia: University of South Carolina Press, 1973.

[20] P. Eskelinen, K. Miettinen, K. Klamroth, and J. Hakanen, "Pareto Navigator for interactive nonlinear multiobjective optimization", *OR Spectrum*, vol. 32, no. 1, pp. 211–227, 2010.

[21] J. W. Fowler, E. S. Gel, M. M. Köksalan, P. Korhonen, J. L. Marquis, and J. Wallenius, "Interactive evolutionary multi-objective optimization for quasi-concave preference functions", *European Journal of Operational Research*, vol. 206, no. 2, pp. 417–425, 2010.

[22] L. R. Gardiner and D. Vanderpooten, "Interactive multiple criteria procedures: Some reflections", in *Multicriteria Analysis*, J. Clímaco, Ed., Berlin, Heidelberg: Springer, 1997, pp. 290–301.

[23] M. Gong, F. Liu, W. Zhang, L. Jiao, and Q. Zhang, "Interactive MOEA/D for multi-objective decision making", in *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, New York: ACM, 2011, pp. 721–728.

[24] J. Hakanen, T. Chugh, K. Sindhya, Y. Jin, and K. Miettinen, "Connections of reference vectors and different types of preference information in interactive multiobjective evolutionary algorithms", in *Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 2016, pp. 1–8.

[25] J. Hakanen, K. Miettinen, and K. Matković, "Task-based visual analytics for interactive multiobjective optimization", *Journal of the Operational Research Society*, pp. 1–18, 2020.

[26] J. Hakanen, K. Sahlstedt, and K. Miettinen, "Wastewater treatment plant design and operation under multiple conflicting objective functions", *Environmental Modelling & Software*, vol. 46, pp. 240–249, 2013.

[27] M. Hartikainen, K. Miettinen, and K. Klamroth, "Interactive Nonconvex Pareto Navigator for multiobjective optimization", *European Journal of Operational Research*, vol. 275, no. 1, pp. 238–251, 2019.

[28] S. Huber, M. J. Geiger, and M. Sevaux, "Simulation of preference information in an interactive reference point-based method for the bi-objective inventory routing problem", *Journal of Multi-Criteria Decision Analysis*, vol. 22, no. 1-2, pp. 17–35, 2015.

[29] C.-L. Hwang and A. S. M. Masud, *Multiple Objective Decision Making—Methods and Applications: a State-of-the-Art Survey*. Berlin, Heidelberg: Springer, 1979, vol. 164.

[30] M. Kok, "The interface with decision makers and some experimental results in interactive multiple objective programming methods", *European Journal of Operational Research*, vol. 26, no. 1, pp. 96–107, 1986.

[31] M. Koksalan and I. Karahan, "An interactive territory defining evolutionary algorithm: iTDEA", *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 5, pp. 702–722, 2010.

[32] P. Korhonen and O. Lantto, *An experimental comparison of some reference direction techniques for MCDM-problems*. Helsingin kauppakorkeakoulu, 1986.

[33] P. Korhonen, H. Moskowitz, and J. Wallenius, "Choice behavior in interactive multiple-criteria decision making", *Annals of Operations Research*, vol. 23, no. 1, pp. 161–179, 1990.

[34] P. Korhonen and J. Wallenius, "Observations regarding choice behaviour in interactive multiple criteria decision-making environments: An experimental investigation", in *Methodology and Software for Interactive Decision Support*, A. Lewandowski and I. Stanchev, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1989, pp. 163–170.

[35] ——, "Behavioural issues in MCDM: Neglected research questions", *Journal of Multicriteria Decision Analysis*, vol. 5, no. 3, pp. 178–182, 1996.

[36] O. I. Larichev, "Cognitive validity in design of decision-aiding techniques", *Journal of Multi-Criteria Decision Analysis*, vol. 1, no. 3, pp. 127–138, 1992.

[37] M. López-Ibáñez and J. Knowles, "Machine decision makers as a laboratory for interactive EMO", in *Proceedings of the 8th International Conference on Evolutionary Multi-Criterion Optimization*, A. Gaspar-Cunha, C. Henggeler Antunes, and C. C. Coello, Eds., Cham: Springer, 2015, pp. 295–309.

[38] A. López-Jaimes and C. A. C. Coello, "Including preferences into a multiobjective evolutionary algorithm to deal with many-objective engineering optimization problems", *Information Sciences*, vol. 277, pp. 1–20, 2014.

[39] M. Luque, R. Caballero, J. Molina, and F. Ruiz, "Equivalent information for multiobjective interactive procedures", *Management Science*, vol. 53, pp. 125–134, 2007.

[40] M. Luque, J.-B. Yang, and B. Y. H. Wong, "PROJECT method for multiobjective optimization based on gradient projection and reference points", *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 39, no. 4, pp. 864–879, 2009.

[41] W. Michalowski, "Evaluation of a multiple criteria interactive programming approach: An experiment", *INFOR: Information Systems and Operational Research*, vol. 25, no. 2, pp. 165–173, 1987.

[42] K. Miettinen, *Nonlinear Multiobjective Optimization*. Boston: Kluwer Academic Publishers, 1999.

[43] K. Miettinen, P. Eskelinen, F. Ruiz, and M. Luque, "NAUTILUS method: An interactive technique in multiobjective optimization based on the nadir point", *European Journal of Operational Research*, vol. 206, no. 2, pp. 426–434, 2010.

[44] K. Miettinen, J. Hakanen, and D. Podkopaev, "Interactive nonlinear multiobjective optimization methods", in *Multiple Criteria Decision Analysis*, S. Greco, M. Ehrgott, and J. R. Figueira, Eds., 2nd ed., New York: Springer, 2016, pp. 927–976.

[45] K. Miettinen, A. V. Lotov, G. K. Kamenev, and V. E. Berezkin, "Integration of two multiobjective optimization methods for nonlinear problems", *Optimization Methods and Software*, vol. 18, no. 1, pp. 63–80, 2003.

[46] K. Miettinen and M. M. Mäkelä, "Synchronous approach in interactive multiobjective optimization", *European Journal of Operational Research*, vol. 170, no. 3, pp. 909–922, 2006.

[47] K. Miettinen, D. Podkopaev, F. Ruiz, and M. Luque, "A new preference handling technique for interactive multiobjective optimization without trading-off", *Journal of Global Optimization*, vol. 63, no. 4, pp. 633–652, 2015.

[48] K. Miettinen, F. Ruiz, and A. P. Wierzbicki, "Introduction to multiobjective optimization: Interactive approaches", in *Multiobjective Optimization: Interactive and Evolutionary Approaches*, J. Branke, K. Deb, K. Miettinen, and R. Slowinski, Eds., Berlin, Heidelberg: Springer, 2008, pp. 27–57.

[49] R. Narasimhan and S. K. Vickery, "An experimental evaluation of articulation of preferences in multiple criterion decision-making (MCDM) methods", *Decision Sciences*, vol. 19, no. 4, pp. 880–888, 1988.

[50] K. Narukawa, Y. Setoguchi, Y. Tanigaki, M. Olhofer, B. Sendhoff, and H. Ishibuchi, "Preference representation using Gaussian functions on a hyperplane in evolutionary multi-objective optimization", *Soft Computing*, vol. 20, no. 7, pp. 2733–2757, 2016.

[51] V. Ojalehto, D. Podkopaev, and K. Miettinen, "Towards automatic testing of reference point based interactive methods", in *Proceedings of the International Conference on Parallel Problem Solving from Nature*, J. Handl, E. Hart, P. R. Lewis, M. López-Ibáñez, G. Ochoa, and B. Paechter, Eds., Cham: Springer, 2016, pp. 483–492.

[52] D. L. Olson, "Review of empirical studies in multiobjective mathematical programming: Subject reflection of nonlinear utility and learning", *Decision Sciences*, vol. 23, no. 1, pp. 1–20, 1992.

[53] L. R. Pedro and R. H. Takahashi, "INSPM: An interactive evolutionary multi-objective algorithm with preference model", *Information Sciences*, vol. 268, pp. 202–219, 2014.

[54] S. Phelps and M. Köksalan, "An interactive evolutionary metaheuristic for multiobjective combinatorial optimization", *Management Science*, vol. 49, no. 12, pp. 1726–1738, 2003.

[55] A. B. Ruiz, J. M. Cabello, C. A. Platero, and F. Blázquez, "Multicriteria optimization of the investment in the auxiliary services of thermal power plants: A case study", *Energy Conversion and Management*, vol. 93, pp. 339–348, 2015.

[56] A. B. Ruiz, M. Luque, K. Miettinen, and R. Saborido, "An interactive evolutionary multiobjective optimization method: Interactive WASF-GA", in *Proceedings of the 8th International Conference on Evolutionary Multi-Criterion Optimization*, A. Gaspar-Cunha, C. Henggeler Antunes, and C. C. Coello, Eds., Cham: Springer, 2015, pp. 249–263.

[57] A. B. Ruiz, F. Ruiz, K. Miettinen, L. Delgado-Antequera, and V. Ojalehto, "NAUTILUS Navigator: Free search interactive multiobjective optimization without trading-off", *Journal of Global Optimization*, vol. 74, no. 2, pp. 213–231, 2019.

[58] A. B. Ruiz, K. Sindhya, K. Miettinen, F. Ruiz, and M. Luque, "E-NAUTILUS: A decision support system for complex multiobjective optimization problems based on the NAUTILUS method", *European Journal of Operational Research*, vol. 246, no. 1, pp. 218–231, 2015.

[59] L. B. Said, S. Bechikh, and K. Ghédira, "The r-dominance: A new dominance relation for interactive evolutionary multicriteria decision making", *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 5, pp. 801–818, 2010.

[60] X. Shen, Y. Guo, Q. Chen, and W. Hu, "A multi-objective optimization evolutionary algorithm incorporating preference information based on fuzzy logic", *Computational Optimization and Applications*, vol. 46, no. 1, pp. 159–188, 2010.

[61] K. Sindhya, A. B. Ruiz, and K. Miettinen, "A preference based interactive evolutionary algorithm for multi-objective optimization: PIE", in *Proceedings of the 6th International Conference on Evolutionary Multi-Criterion Optimization*, R. H. C. Takahashi, K. Deb, E. F. Wanner, and S. Greco, Eds., Berlin, Heidelberg: Springer, 2011, pp. 212–225.

[62] A. Sinha, P. Korhonen, J. Wallenius, and K. Deb, "An interactive evolutionary multi-objective optimization method based on polyhedral cones", in *Proceedings of the International Conference on Learning and Intelligent Optimization*, C. Blum and R. Battiti, Eds., Berlin, Heidelberg: Springer, 2010, pp. 318–332.

[63] R. E. Steuer, *Multiple Criteria Optimization: Theory, Computation, and Application.* Wiley New York, 1986, vol. 233.

[64] T. J. Stewart, "Concepts of interactive programming", in *Multicriteria Decision Making*, T. Gal, T. Stewart, and T. Hanne, Eds., Springer, 1999, pp. 277–304.

[65] T. J. Stewart, "Evaluation and refinement of aspiration-based methods in MCDM", *European Journal of Operational Research*, vol. 113, no. 3, pp. 643–652, 1999.

[66] T. J. Stewart, "Goal programming and cognitive biases in decision-making", *Journal of the Operational Research Society*, vol. 56, no. 10, pp. 1166–1175, 2005.

[67] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases", *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974.

[68] J. Wallenius, "Comparative evaluation of some interactive approaches to multicriterion optimization", *Management Science*, vol. 21, no. 12, pp. 1387–1396, 1975.

[69] R. Wang, R. C. Purshouse, and P. J. Fleming, "'whatever works best for you'-a new method for a priori and progressive multi-objective optimisation", in *Proceedings of the 7th International Conference on Evolutionary Multi-Criterion Optimization*, R. C. Purshouse, P. J. Fleming, C. M. Fonseca, S. Greco, and J. Shaw, Eds., Berlin, Heidelberg: Springer, 2013, pp. 337–351.

[70] B. Y. Wong, M. Luque, and J.-B. Yang, "Using interactive multiobjective methods to solve DEA problems with value judgements", *Computers & Operations Research*, vol. 36, no. 2, pp. 623–636, 2009.

[71] B. Xin, L. Chen, J. Chen, H. Ishibuchi, K. Hirota, and B. Liu, "Interactive multiobjective optimization: A review of the state-of-the-art", *IEEE Access*, vol. 6, pp. 41 256–41 279, 2018.

[72] J.-B. Yang, "Gradient projection and local region search for multiobjective optimisation", *European Journal of Operational Research*, vol. 112, no. 2, pp. 432–459, 1999.

[73] J.-B. Yang and D. Li, "Normal vector identification and interactive tradeoff analysis using minimax formulation in multiobjective optimization", *IEEE Transactions on Systems, Man, And Cybernetics-Part A: Systems and Humans*, vol. 32, no. 3, pp. 305–319, 2002.

[74] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. Da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review", *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 117–132, 2003.