Interval-based ranking in noisy evolutionary multi-objective optimization

Hossein Karshenas, Concha Bielza & Pedro Larrañaga

Volume 59, Number 3, Decem...
ISSN: 0926-6003

ONLINE FIRST

COMPUTATIONAL OPTIMIZATION AND APPLICATIONS

An International Journal

Editor-in-Chief:

William W. Hager

Springer

Springer

Springer

Interval-based ranking in noisy evolutionary multi-objective optimization

**Hossein Karshenas · Concha Bielza ·
Pedro Larrañaga**

Abstract As one of the most competitive approaches to multi-objective optimization, evolutionary algorithms have been shown to obtain very good results for many real-world multi-objective problems. One of the issues that can affect the performance of these algorithms is the uncertainty in the quality of the solutions which is usually represented with the noise in the objective values. Therefore, handling noisy objectives in evolutionary multi-objective optimization algorithms becomes very important and is gaining more attention in recent years. In this paper we present α-degree Pareto dominance relation for ordering the solutions in multi-objective optimization when the values of the objective functions are given as intervals. Based on this dominance relation, we propose an adaptation of the non-dominated sorting algorithm for ranking the solutions. This ranking method is then used in a standard multi-objective evolutionary algorithm and a recently proposed novel multi-objective estimation of distribution algorithm based on joint variable-objective probabilistic modeling, and applied to a set of multi-objective problems with different levels of independent noise. The experimental results show that the use of the proposed method for solution ranking allows to approximate Pareto sets which are considerably better than those obtained when using the dominance probability-based ranking method, which is one of the main methods for noise handling in multi-objective optimization.

H. Karshenas (✉) · C. Bielza · P. Larrañaga
Computational Intelligence Group, Facultad de Informática, Universidad Politécnica de Madrid,
Campus de Montegancedo, 28660 Boadilla del Monte, Madrid, Spain
e-mail: hosseinkarshenas@gmail.com

    Ⓐ Springer

1 Introduction

Many of the real world optimization problems are often best characterized by more than one objective. When trying to solve these problems, several, possibly conflicting, criteria should be fulfilled simultaneously. Let $\mathcal{F} = \{f_1, \ldots, f_m\}$ be the set of objective functions for a problem, where each objective function is defined as $f_j : \mathbb{D} \mapsto \mathbb{R}$, and assume that all of the objectives should be minimized. $\mathbb{D}$ determines the set of feasible value-settings for the vector of input variables $X = (X_1, \ldots, X_n)$ and sometimes it is referred to as the decision space. Then an unconstrained continuous multi-objective optimization problem (MOP) is defined as

$$\min_{\boldsymbol{x} \in \mathbb{D} \subseteq \mathbb{R}^n} \boldsymbol{f}(\boldsymbol{x}) = \big(f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})\big). \tag{1}$$

The objective functions of an MOP map the solutions in $\mathbb{D}$ to another space which is called the objective space of the problem.

In general, with more than one objective function, the optimum solution to the MOP in Eq. (1) is not unique anymore, even if the constituting objective functions are not multi-modal. The presence of conflicting objectives implies that trying to improve one of the objectives will result in worse values for some other. Thus, the goal of a multi-objective optimization algorithm is to search for solutions which result in an optimal trade-off between different objectives of the problem. One of the most frequently used approaches is to employ the notion of Pareto optimality [1] and the corresponding Pareto dominance relation [2].

Definition 1 (*Pareto Dominance*) Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two solutions of the MOP defined in Eq. (1) in the decision space $\mathbb{D}$. Then $\boldsymbol{x}$ Pareto dominates $\boldsymbol{y}$, denoted as $\boldsymbol{x} \prec \boldsymbol{y}$, if and only if:

1. $\forall f_j \in \mathcal{F} \quad f_j(\boldsymbol{x}) \leq f_j(\boldsymbol{y})$, and
2. $\exists f_k \in \mathcal{F} \quad f_k(\boldsymbol{x}) < f_k(\boldsymbol{y})$.

Pareto dominance defines a *partial* order between the solutions, i.e. an irreflexive, antisymmetric and transitive relation. Using this relation, the solutions to an MOP can be ranked into a number of disjoint Pareto non-dominated sets (or Pareto sets for short) where every two solutions in the same set either have exactly equal objective values or are *incomparable* in terms of Pareto dominance. The corresponding projections of these Pareto sets onto the objective space are called Pareto non-dominated fronts (similarly Pareto fronts for short). The best Pareto set containing optimal trade-off solutions (no other solution in $\mathbb{D}$ dominates them) is called the Pareto optimal set, and its projection onto the objective space is called the Pareto optimal front. In general, the number of solutions in the Pareto optimal set can be exponentially large or even infinite in the case of continuous domains and thus multi-objective optimization algorithms try to obtain a good approximation of this set.

Because of the complexity of solving MOPs, a promising approach which is gaining an increasing interest is to use evolutionary algorithms (EAs), giving rise to evolutionary multi-objective optimization (EMO) algorithms [3–6], otherwise known as multi-objective EAs. Although these algorithms do not ensure the optimality of the solutions compared with some of the conventional mathematical optimization techniques, they have been successfully applied to many complex real-world MOPs. Beside their simple mechanism which motivates their wide-spread use in different applications, another advantage of these algorithms encouraging their use for multi-objective optimization is their population-based search allowing them to simultaneously explore several regions of the search space, finding a set of trade-off solutions in contrast to a single solution, in an individual run.

Estimation of distribution algorithms (EDAs) [7–10] are a relatively recent class of EAs developed by using probabilistic modeling in the framework of EAs. They have proven to be promising optimization algorithms for many difficult problems with high computational complexity. These algorithms explore the search space by building a probabilistic model from a set of selected candidate solutions. This probabilistic model is then used to sample new candidate solutions in the search space. As the result, these algorithms will provide a model expressing the regularities of the problem structure, as well as the final solutions. EDAs have also been used for multi-objective optimization [11–13], where the model building in EDAs is combined with a multi-objective solution ranking and selection mechanism to solve MOPs.

A feature of many real-world problems is the existence of noise, which can appear as variable or environmental change, or/and as uncertainty in the objective values. Many of the methods that are proposed for solution ranking in EMO algorithms are based on the Pareto dominance relation. When EMO algorithms based on Pareto dominance relation are used for solving these problems, the noise in objective values can mislead the algorithm by selecting inferior solutions that are considered good because of noise, and discarding good solutions that are necessary for directing the search to promising areas. One of the ways to deal with noisy objectives is to assume each objective returns an interval of values for a solution. This interval can be obtained by considering the range of error or amount of noise in the system, or from a set of values obtained from multiple reevaluations of a solution in different conditions. Considering an interval of values instead of a single value allows the EMO algorithm to take into account the extent of noise in the objectives when selecting solutions for recombination.

In this paper we introduce a new solution ranking, called α-degree Pareto dominance, for EMO algorithms applied to noisy problems when independently computed objective functions return intervals. This new relation for ranking the solutions allows the intervals to overlap each other, and determines the dominance among solutions based on the extent to which their corresponding intervals are better than those of other solutions. The method can also be extended to the case where only some of the objective functions of the MOP are noisy. This solution ranking method is then integrated into a standard multi-objective EA and MBN-EDA [14], a multi-objective EDA based on joint variable-objective probabilistic modeling, for finding the solutions of noisy MOPs.

The rest of paper is organized as follows. In the next section we briefly review some of the methods proposed in the literature for noise handling in EMO algo-

rithms. Section 3 introduces the α-degree Pareto dominance relation, discusses some of its properties, and proposes a solution ranking method based on this relation. In Sect. 4, MBN-EDA and its joint probabilistic modeling of variables and objectives are described. The details of the experiments and their results are presented and discussed in Sect. 5. Finally, Sect. 6 concludes the paper and gives some lines of future work.

2 A review on EMO with noise

The topic of EMO under uncertainty and noise has gained a lot of attention in the past few years, and many studies and methods are reported in the literature. In general, as explained by Jin and Branke [15], three different types of noise handling can be identified in EAs. First, the population-based search in these algorithms, by itself, implicitly deals with certain levels of noise in objective values. The average quality of a population or subgroups of the population, presenting certain subspaces, is less susceptible to noise. Therefore, the larger the population is, the better the algorithm can overcome the noise in the quality of solutions for finding the optimal solutions. However, the existence of several objectives in a multi-objective scenario reduces the effectiveness of EAs in noise handling [16]. Second approach is to explicitly reevaluate each solution of the population several times to obtain a better estimation of its objective values. Although this approach greatly increases the computational cost of optimization, it is inevitable for some of the problems where the objective values are obtained as the result of simulations. A third approach is to consider the noise in objective values in the selection step of EAs. This is usually done by modifying the solution ranking method, assuming a level of uncertainty in the objective values.

As explained before, noise can exist in the values of the input variables, the outputs of the objective functions or even both. There are some works on EMO algorithms considering the former type of noise, i.e. noise in the input values, which is sometimes referred to as robust optimization and its aim is to find solutions with the highest stability in their objective values. Soares et al. [17] optimized the worst noisy objective values of the solutions in a min-max formulation using interval analysis. To decrease the amount of uncertainty in the intervals they propose to recursively divide the intervals into halves, resulting in a grid which is placed on the objective space and is used to compute the worst objective values of the solutions. The grid also serves as a niching method, penalizing the solutions that are very close in the objective space. The noise is introduced in the input variables using an uncertainty vector which is incorporated in the definition of the objective functions. Goh et al. [18] offered a classification of the noisy MOPs depending on the effect of noise on the Pareto front, Pareto set and landscape of the problem. Based on this classification, they proposed some guidelines for designing challenging MOPs for robust optimization, and introduced a Gaussian landscape generator for this purpose using a number of basis functions.

Most of the algorithms proposed for EMO in noisy environments consider the second type of noise, i.e. noise in the objective values, since this type of noise is harder to deal with. In this type of MOPs it is usually assumed that the objective values are distorted by an additive noise value. Almost all EMO algorithms reviewed

in this section consider a Gaussian noise model for the objective values of the tested MOPs, unless otherwise mentioned.

One of the earliest and well-known approaches to deal with noisy objectives when comparing two solutions in multi-objective optimization is to use probabilistic dominance, considering the probability that one solution dominates another [19,20]. This approach has been successfully used for optimization in some of the real world problems like designing the shape of acoustic noise attenuation barriers with several receiver points [21]. Since this method is considered as a main reference in many later works on noisy objectives, we explain it here in more detail.

Definition 2 (*Dominance Probability* [19,20]) Let $U = f(x)$ and $V = f(y)$ be two vectors of random variables representing the objective values returned for two solutions x and y of a noisy MOP in decision space $\mathbb{D}$. Then the probability that x dominates y is given by:

$$P(x \prec y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \rho_U(t)\big(1 - \Psi_V(t)\big)dt, \tag{2}$$

where $\rho_U(\cdot)$ denotes the joint probability distribution for the output of objective functions given solution x, and $\Psi_V(\cdot)$ is the cumulative probability distribution for the output of objective functions given solution y.

Based on the probability of dominance, ranking strategies like the mean dominance probability [20] is used to order the solutions with noisy objective values. [19] also tested this approach when the noise is introduced to the input values, showing the ability of this method for handling noise in both input variables and output objective functions. Bui et al. [22] adopted probabilistic dominance in NSGA-II [23] and compared it with a version of this algorithm which is based only on reevaluation of objectives. They also proposed a fitness inheritance method to reduce the complexity of reevaluating the objectives. In a separate work [24], they performed a comparative study of the noise handling capability of the original versions of NSGA-II and SPEA2 [25] in the presence of different noise levels, using several performance measures. [26] proposed a Bayesian method for obtaining an estimation of the Gaussian noise variance when computing the probability of dominance between two solutions. They considered different scenarios like unknown, constant and variable noise for the objective values.

Many of the EMO algorithms proposed for noise handling try to extend the existing solution ranking and selection methods by considering noise in the values of the objectives. Büche et al. [27] proposed a noise-tolerant EA using the Pareto strengths-based solution ranking method [28] assigning a lifetime (number of generations) to the solutions in the archive depending on their strengths. This lifetime is used to determine which solutions are to be reevaluated or to be used for updating the archive in order to limit the influence of noisy objective values and outliers. They claim that elitism does not necessarily result in faster convergence when noise is present in multi-objective optimization. The proposed algorithm is applied to find an optimal flow of fuel in the burner of a gas turbines.

Babbar et al. [29] modified NSGA-II to include neighbors of the non-dominated solutions in the first Pareto front. Neighboring solutions are determined using the mean

and variance of the objective values for each solution, estimated from the reevaluation of the solutions. Solutions that are reevaluated less than a predetermined number of times during the evolution are considered as outliers and removed from the final Pareto set.

Goh and Tan [30] proposed three techniques for noise handling in an EMO algorithm which uses a two-part, discrete-continuous, representation for the solutions. The first technique consists of incorporating the direction of population movement along evolution into the generation of new solutions. The second technique is stretching or shrinking the search domain of each variable depending on the behavior of the algorithm in the current phase of evolution. The third technique is to assume objective values are given as fuzzy numbers and then use two newly proposed dominance relations, called necessity and necessity-possibility, to update the archive of non-dominated solutions maintained by the algorithm.

Eskandari and Geiger [31] considered the expected values of the objective functions and proposed the stochastic dominance relation for ranking the solutions. In the selection process, the solutions are divided into two sets depending on whether they are stochastically dominated. The set of stochastically dominated solutions are further ordered by considering the expected strength of each solution depending on both how many solutions it dominates and by how many solutions it is dominated. To detect the algorithm convergence in noisy environments they proposed to monitor the rate of hypervolume indicator growth.

Bui et al. [32] proposed the use of adaptive non-overlapping hyper-spheres which are locally moved in the search space to reduce the effect of noise. The motivation is that the average objective values of the solutions in a neighborhood of the search space provide a better approximation of the direction of movement during evolution. A particle swarm optimization based algorithm is used to update the center and radius of the hyper-spheres in the search space. The algorithm also deploys an archive of solutions and its corresponding hyper-sphere to represent the global behavior of the population.

Syberfeldt et al. [33] proposed a method to increase the efficiency of solution reevaluations. When comparing two solutions in the non-dominated sorting algorithm [23], if the confidence in the differences between their objective values are less than a specific level, then one of them is reevaluated more to increase the level of confidence. Different Pareto sets are assigned different values of minimum confidence level, and since during evolution the rank of solutions in the population changes, implicit dynamism is introduced during evolution as each solution is reevaluated. The proposed method is also applied to two real-world problems related to optimization in engine manufacturing lines. Instead of discarding old evaluations of a solution in a noisy problem, Park and Ryu [34] proposed an accumulating approach which combines old evaluations with the new reevaluation of a solution to improve the expected value estimation of objective functions.

Another approach taken by some of the proposed methods is to use decision maker (DM) provided information for EMO in the presence of noise. Mehnen et al. [35] used desirability functions to include both the preferences of DM and to reduce the effect of noise. Instead of performing the search in the objective space, they optimize the expected value of the desirability functions computed from noisy objectives. The

final Pareto front is obtained by either applying the inverse desirability functions on the approximated front or computing the objective values of the final Pareto set. The method is used to find the working parameters of an industrial cutting tool. Woźniak [36] proposed to use a number of reference points provided by DM, each accompanied with a weight vector showing the importance of the objectives, to select fitter solutions in the search process. To maintain the population diversity only one of the solutions in each neighborhood, defined by a predefined neighborhood radius, is selected. The method is then applied for the design of a motor speed controller.

In addition to the noise in objective values, Kaji et al. [37] studied the effect of noise in constraint functions of a constrained MOP, trying to reduce the number of infeasible solutions selected in the Pareto set. For this purpose they form a history of solutions and estimate the value of objective and constraint functions by a locally weighted ridge regression of second order. The weights are defined using a Gaussian kernel. A safety margin is also introduced to the constraint functions and dynamically adjusted depending on the variance of the estimated constraint values.

Basseur and Zitzler [38] proposed an indicator-based EMO algorithm for noisy environments, considering the presence of noise in both the objective functions and the input variables. They proposed methods to estimate the expected value of the epsilon indicator for ranking the solutions. A uniform noise is assumed for both inputs and outputs. Boonma and Suzuki [39] considered different types of distributions like Gaussian, uniform and Poisson for the noise model. Assuming that the quality of each solution is represented with several reevaluations of its objectives, they used a support vector machine to determine the confidence level in these objective values. The coverage metric is used to determine the dominance between two solutions after their values are classified as statistically reliable. The confidence level for accepting the objective values of the solutions is dynamically adjusted during evolution according to the disorder among objective values, which is computed with an entropy-based function.

Table 1 summarizes the reviewed algorithms.

3 α-Degree Pareto dominance

Most of the methods reviewed in the previous section for handling the noise when selecting a subset of solutions, only consider singular objective values. Although these singular values maybe obtained as the result of averaging over several reevaluations, they still do not directly take into account the extent of the noise in the values. We also saw that some of the proposed methods optimize the expected values of the objectives and consider the variance of the objective values, implicitly assuming a type of confidence interval for the objective values. A closely related approach is the possibilistic dominance method [30], where objective values are considered to be fuzzy numbers. In this section we propose a method for directly comparing any kind of intervals, not necessarily confidence intervals, which also takes into account interval widths (representing the amount of noise). Moreover, the proposed method can deal with MOPs consisting of both singular and interval objectives.

Table 1 Summary of the EMO methods for noisy environments

	Method	Noise type	Noise location
Teich [20]	Probabilistic Dominance	Uniform	Output
Hughes [19]	Probabilistic Dominance	Gaussian	Input, Output
Büche et al. [27]	Lifetime-Based Archiving	Gaussian	Output
Babbar et al. [29]	Objective Space Neighborhood	Gaussian	Output
Bui et al. [22]	Fitness Inheritance in Reevaluations	Gaussian	Output
Fieldsend and Everson [26]	Noise Variance Estimation	Gaussian	Output
Basseur and Zitzler [38]	Indicator Value Estimation	Uniform	Input, Output
Goh and Tan [30]	Necessity and Necessity-Possibility Dominance	Gaussian	Output
Mehnen et al. [35]	Desirability Functions Optimization	Gaussian	Output
Woźniak [36]	Weighted Reference Points	Gaussian	Output
Boonma and Suzuki [39]	Confidence-Based Solution Comparison	Gaussian, Uniform, Poisson	Output
Bui et al. [32]	Search Space Neighborhood	Gaussian	Output
Eskandari and Geiger [31]	Stochastic Dominance	Gaussian	Output
Kaji et al. [37]	Value Estimation with Locally Weighted Ridge Regression	Gaussian	Output, Constraints
Soares et al. [17]	Worst-Case Analysis	Gaussian	Output
Syberfeldt et al. [33]	Confidence-Based Reevaluations	Gaussian	Output
Park and Ryu [34]	Reevaluation Accumulation	Gaussian	Output

One way to deal with the noise in an MOP when the objective values are represented with intervals is to extend the Pareto dominance definition (Definition 1). In the following definitions we assume that the set of objectives $\mathcal{F}$ is partitioned into two disjoint subsets of objectives with singular values $\mathcal{F}_S$, and noisy objectives with interval values $\mathcal{F}_I$, such that $\mathcal{F}_S \cup \mathcal{F}_I = \mathcal{F}$ and $\mathcal{F}_S \cap \mathcal{F}_I = \emptyset$.

Definition 3 (*Extended Pareto Dominance*) Let $\lfloor f_j(\boldsymbol{x}) \rfloor$ and $\lceil f_j(\boldsymbol{x}) \rceil$ respectively represent the lower and upper bounds of the interval value returned for solution $\boldsymbol{x} \in \mathbb{D}$, by the noisy objective function $f_j \in \mathcal{F}_I$, i.e. U_j lies in the interval $\left[\lfloor f_j(\boldsymbol{x}) \rfloor, \lceil f_j(\boldsymbol{x}) \rceil \right]$. Then, solution $\boldsymbol{x}$ is said to strictly dominate solution $\boldsymbol{y}$, denoted as $\boldsymbol{x} \prec_e \boldsymbol{y}$, if and only if:

1. $\forall f_j \in \mathcal{F}_S \quad f_j(\boldsymbol{x}) \leq f_j(\boldsymbol{y})$, and
2. $\forall f_j \in \mathcal{F}_I \quad \lceil f_j(\boldsymbol{x}) \rceil \leq \lfloor f_j(\boldsymbol{y}) \rfloor$, and
3. $\left(\exists f_k \in \mathcal{F}_S \quad f_k(\boldsymbol{x}) < f_k(\boldsymbol{y}) \quad or \quad \exists f_k \in \mathcal{F}_I \quad \lceil f_k(\boldsymbol{x}) \rceil < \lfloor f_k(\boldsymbol{y}) \rfloor \right)$.

This definition treats interval values similar to the way it considers singular values and only allows a solution to dominate other solutions if its corresponding interval values, returned by noisy objective functions, are *strictly* better than those of other

solutions. In real-world noisy MOPs, such a requirement is hardly satisfied. A further extension is to relax this strict requirement and allow the intervals to have some degree of overlapping.

Definition 4 (*α-Degree Pareto Dominance*) Assume the same notations as those of Definition 3. Then, solution x is said to dominate another solution y with a degree $\alpha \in (0, 1]$, denoted as $x \prec_\alpha y$, if and only if:

1. $\forall f_j \in \mathcal{F}_S \quad f_j(x) \leq f_j(y)$, and
2. $\forall f_j \in \mathcal{F}_I \quad \deg_j(x, y) \geq \alpha$, and
3. $\left(\exists f_k \in \mathcal{F}_S \quad f_k(x) < f_k(y) \quad \vee \quad \exists f_k \in \mathcal{F}_I \quad \deg_k(x, y) > \alpha \right)$,

where $\deg_j(\cdot, \cdot)$ gives the degree that a solution dominates another with respect to the noisy objective function $f_j \in \mathcal{F}_I$

$$\deg_j(x, y) = \min \left\{ 1, \max \left\{ 0, \frac{\lfloor f_j(y) \rfloor - \lfloor f_j(x) \rfloor}{\lceil f_j(x) \rceil - \lfloor f_j(x) \rfloor} \right\} \right\}. \tag{3}$$

Intuitively, $\deg_j(x, y)$ computes the percentage of the interval obtained for solution x that is not overlapped by the interval obtained for solution y in objective $f_j \in \mathcal{F}_I$ $\left(\text{i.e.} 1 - \frac{\lceil f_j(x) \rceil - \lfloor f_j(y) \rfloor}{\lceil f_j(x) \rceil - \lfloor f_j(x) \rfloor} \right)$. Thus, only the segment of the interval obtained for solution x that is better than the best point in the interval obtained for solution y (i.e. its lower bound, when minimizing objectives) is taken into account. If $\alpha = 1$, then Definition 4 is reduced to Definition 3. Definition 4 allows a solution to dominate other solutions when its corresponding interval values are partially better than those of other solutions.

Figure 1 shows some examples of possible placements of two intervals and the corresponding values of the $\deg_j(\cdot, \cdot)$ function. Figure 1a shows a situation where the interval for solution x is strictly better than the other solution, resulting in $\deg_j(x, y) = 1$. Figure 1b, c show cases when the two intervals partially overlap with $0 < \deg_j(x, y) < 1$. Finally, Fig. 1d shows an example of the situation when intervals completely overlap, i.e. $\deg_j(x, y) = 0$. With higher values of α, a solution can only dominate other solutions if a major segment of its corresponding interval values are better than the best points of the interval values corresponding to other solutions. Thus, higher values of α place a stricter condition for accepting a solution as non-dominated.

3.1 Properties

In this section we show some of the properties of the α-Degree Pareto Dominance (Definition 4) which will allow us to use it for solution ordering.

Proposition 1 *α-Degree Pareto dominance defines a partial relation, i.e. with irreflexivity, antisymmetry and transitivity properties, on the space of candidate solutions.*

Proof Since no $x \in \mathbb{D}$ exists that satisfies $x \prec_\alpha x$ for any $\alpha \in (0, 1]$, the relation is irreflexive. To see the antisymmetry of the relation, assume $x \prec_\alpha y$. If $\exists f_j \in \mathcal{F}_S$

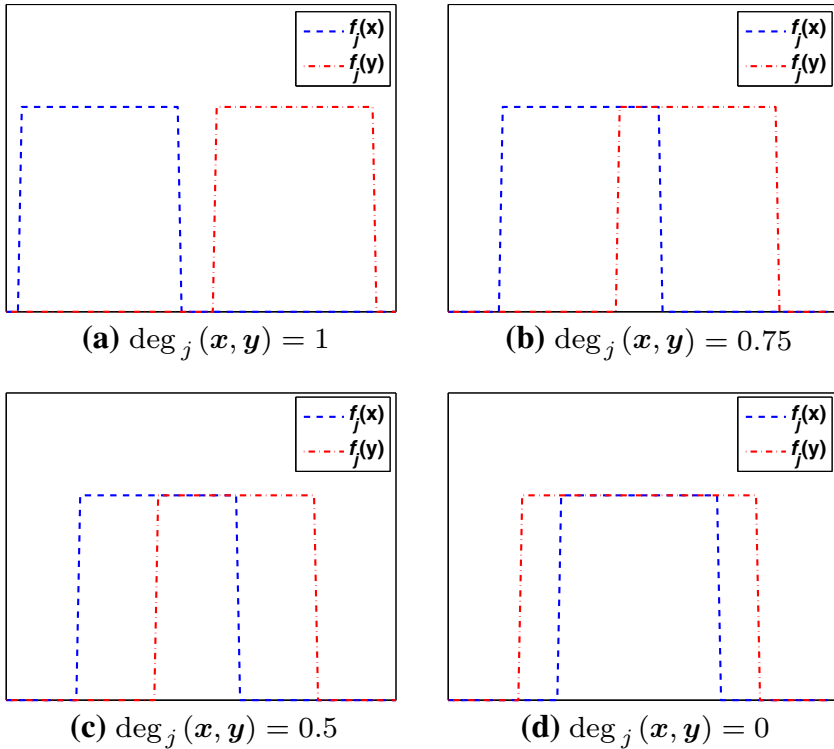**(a)** $\deg_j(\boldsymbol{x}, \boldsymbol{y}) = 1$

(b) $\deg_j(\boldsymbol{x}, \boldsymbol{y}) = 0.75$

(c) $\deg_j(\boldsymbol{x}, \boldsymbol{y}) = 0.5$

(d) $\deg_j(\boldsymbol{x}, \boldsymbol{y}) = 0$

Fig. 1 Examples of interval values and the resulting degrees of dominance. **a** $\deg_j(\boldsymbol{x}, \boldsymbol{y}) = 1$, **b** $\deg_j(\boldsymbol{x}, \boldsymbol{y}) = 0.75$, **c** $\deg_j(\boldsymbol{x}, \boldsymbol{y}) = 0.5$, **d** $\deg_j(\boldsymbol{x}, \boldsymbol{y}) = 0$

such that $f_j(\boldsymbol{x}) < f_j(\boldsymbol{y})$ then $\boldsymbol{y} \not\prec_\alpha \boldsymbol{x}$. If $\exists f_j \in \mathcal{F}_I$ such that $\deg_j(\boldsymbol{x}, \boldsymbol{y}) > \alpha$, then by definition of $\deg_j(\cdot, \cdot)$ in Eq. (3) if $\deg_j(\boldsymbol{x}, \boldsymbol{y}) > 0$ then $\deg_j(\boldsymbol{y}, \boldsymbol{x}) = 0$ and therefore $\boldsymbol{y} \not\prec_\alpha \boldsymbol{x}$ for any $\alpha \in (0, 1]$.

For transitivity, assume that $\boldsymbol{x} \prec_\alpha \boldsymbol{y}$ and $\boldsymbol{y} \prec_\alpha \boldsymbol{z}$. Then we have $\forall f_j \in \mathcal{F}_S$, $f_j(\boldsymbol{x}) \leq f_j(\boldsymbol{y})$ and $f_j(\boldsymbol{y}) \leq f_j(\boldsymbol{z})$. Therefore, $f_j(\boldsymbol{x}) \leq f_j(\boldsymbol{z}), \forall f_j \in \mathcal{F}_S$. Similarly, we have $\deg_j(\boldsymbol{x}, \boldsymbol{z}) \geq \alpha, \forall f_j \in \mathcal{F}_I$, because if $\forall f_j \in \mathcal{F}_I$ such that $\lfloor f_j(\boldsymbol{x}) \rfloor < \lfloor f_j(\boldsymbol{y}) \rfloor$ and $\lfloor f_j(\boldsymbol{y}) \rfloor < \lfloor f_j(\boldsymbol{z}) \rfloor$, then we have $\lfloor f_j(\boldsymbol{x}) \rfloor < \lfloor f_j(\boldsymbol{z}) \rfloor, \forall f_j \in \mathcal{F}_I$. Now, given either $\exists f_j \in \mathcal{F}_S$ such that $f_j(\boldsymbol{x}) < f_j(\boldsymbol{y})$ or $\exists f_j \in \mathcal{F}_I$ such that $\deg_j(\boldsymbol{x}, \boldsymbol{y}) > \alpha$ (similarly $\exists f_j \in \mathcal{F}_S$ such that $f_j(\boldsymbol{y}) < f_j(\boldsymbol{z})$ or $\exists f_j \in \mathcal{F}_I$ such that $\deg_j(\boldsymbol{y}, \boldsymbol{z}) > \alpha$) we can respectively conclude either $\exists f_j \in \mathcal{F}_S$ such that $f_j(\boldsymbol{x}) < f_j(\boldsymbol{z})$ or $\exists f_j \in \mathcal{F}_I$ such that $\deg_j(\boldsymbol{x}, \boldsymbol{z}) > \alpha$, which completes the proof of proposition. □

The partial relation defined by α-degree Pareto dominance allows to readily adopt the terms of α-degree Pareto optimal solution, α-degree Pareto optimal set, α-degree Pareto optimal front and α-degree Pareto non-dominated set for noisy MOPs, in the same way they are defined using the conventional Pareto dominance relation. The α-degree Pareto dominance relation also has interesting properties when different confidence levels are considered for the objective values. To see this, lets assume that

the values returned by noisy objective functions for solution x are confidence intervals given in the form of

$$\left(\hat{\mathbb{E}}(f_j(x)) - \varepsilon_\gamma(f_j(x)), \hat{\mathbb{E}}(f_j(x)) + \varepsilon_\gamma(f_j(x))\right),$$

where $\hat{\mathbb{E}}(f_j(x))$ represents an estimation of the expected value of the jth objective function for solution x and

$$\varepsilon_\gamma(f_j(x)) = z_{\frac{1-\gamma}{2}}\hat{\sigma}(f_j(x))$$

is the half-width [31] of the confidence interval computed according to a specific confidence level γ. Here, $z_{(1-\gamma)/2}$ denotes the value for which $\Phi(Z > z_{(1-\gamma)/2}) = \frac{1-\gamma}{2}$, where $\Phi(Z)$ is the cumulative standard Gaussian (or t-student) distribution. $\hat{\sigma}(f_j(x))$ is the estimation for the standard deviation of mean objective value. With confidence intervals, the definition of $\deg_j(x, y)$ for a noisy objective $f_j \in \mathcal{F}_I$ can be rewritten as

$$\deg_j(x, y) = \min\left\{1, \right.$$
$$\left. \max\left\{0, \frac{\left(\hat{\mathbb{E}}(f_j(y)) - \hat{\mathbb{E}}(f_j(x))\right) - \left(\varepsilon_\gamma(f_j(y)) - \varepsilon_\gamma(f_j(x))\right)}{2\varepsilon_\gamma(f_j(x))}\right\}\right\}. \quad (4)$$

According to this definition, the degree that a solution outperforms another solution with respect to a noisy objective $f_j \in \mathcal{F}_I$ is determined by the differences in both expected values and half-widths. The following propositions show how the change in the confidence level γ or dominance degree α affects the α-degree Pareto dominance relation.

Proposition 2 *Any reduction in the confidence level of the interval values given by noisy objective functions in $\mathcal{F}_I$ does not affect $x \prec_\alpha y$, if for every objective function $f_j \in \mathcal{F}_I$ we have $\hat{\mathbb{E}}(f_j(x)) < \hat{\mathbb{E}}(f_j(y))$.*

Proof We only consider noisy objectives in $\mathcal{F}_I$ here because the change in confidence level does not affect singular values. Let $x \prec_\alpha^\gamma y$ denote that solution x dominates solution y with a degree α, when the confidence level in the interval value of all noisy objectives in $\mathcal{F}_I$ is at least γ.

Assuming $x \prec_\alpha^{\gamma_1} y$ implies that

$$\forall f_j \in \mathcal{F}_I, \quad \frac{\left(\hat{\mathbb{E}}(f_j(y)) - \hat{\mathbb{E}}(f_j(x))\right) - \left(\varepsilon_{\gamma_1}(f_j(y)) - \varepsilon_{\gamma_1}(f_j(x))\right)}{2\varepsilon_{\gamma_1}(f_j(x))} \geq \alpha$$

$$\Rightarrow \quad \frac{\hat{\mathbb{E}}(f_j(y)) - \hat{\mathbb{E}}(f_j(x))}{2z_{\frac{1-\gamma_1}{2}}\hat{\sigma}(f_j(x))} - \frac{\hat{\sigma}(f_j(y)) - \hat{\sigma}(f_j(x))}{2\hat{\sigma}(f_j(x))} \geq \alpha.$$

Changing the confidence level does not influence the second quotient above. The numerator of the first quotient is positive because of the proposition assumption and, since $z_{\frac{1-\gamma_1}{2}} > z_{\frac{1-\gamma_2}{2}}, \forall \gamma_2 < \gamma_1$, this quotient will become larger if confidence level decreases to γ_2. Therefore, for all $f_j \in \mathcal{F}_I$

$$\frac{\left(\hat{\mathbb{E}}(f_j(\boldsymbol{y})) - \hat{\mathbb{E}}(f_j(\boldsymbol{x}))\right) - \left(\varepsilon_{\gamma_2}(f_j(\boldsymbol{y})) - \varepsilon_{\gamma_2}(f_j(\boldsymbol{x}))\right)}{2\varepsilon_{\gamma_2}(f_j(\boldsymbol{x}))} >$$

$$\frac{\left(\hat{\mathbb{E}}(f_j(\boldsymbol{y})) - \hat{\mathbb{E}}(f_j(\boldsymbol{x}))\right) - \left(\varepsilon_{\gamma_1}(f_j(\boldsymbol{y})) - \varepsilon_{\gamma_1}(f_j(\boldsymbol{x}))\right)}{2\varepsilon_{\gamma_1}(f_j(\boldsymbol{x}))} \geq \alpha,$$

which means $\boldsymbol{x} \prec_\alpha^{\gamma_2} \boldsymbol{y}$. $\qquad\square$

Now, let $\mathcal{PS}_\alpha^\gamma$ denote the α-degree Pareto optimal set imposed by α-degree Pareto dominance relation $\prec_\alpha^\gamma$, when the confidence level in the interval values returned by the noisy objectives is at least γ.

Corollary 1 *If $\gamma_2 < \gamma_1$, then $\mathcal{PS}_\alpha^{\gamma_2} \subset \mathcal{PS}_\alpha^{\gamma_1}$.*

Proof Assume that there exists a solution $\boldsymbol{y}$ such that $\boldsymbol{y} \in \mathcal{PS}_\alpha^{\gamma_2}$ and $\boldsymbol{y} \notin \mathcal{PS}_\alpha^{\gamma_1}$. This means that there exists a solution $\boldsymbol{x} \in \mathcal{PS}_\alpha^{\gamma_1}$ such that $\boldsymbol{x} \prec_\alpha^{\gamma_1} \boldsymbol{y}$. By Proposition 2, we know that this implies $\boldsymbol{x} \prec_\alpha^{\gamma_2} \boldsymbol{y}$. But the last relation means that $\boldsymbol{y} \notin \mathcal{PS}_\alpha^{\gamma_2}$ which contradicts the hypothesis. $\qquad\square$

Proposition 3 *The α-degree Pareto dominance relation between solutions is preserved when the dominance degree α is decreased.*

Proof It is trivial since if $\alpha_2 < \alpha_1$ then $\deg_j(\boldsymbol{x}, \boldsymbol{y}) \geq \alpha_1$ implies that $\deg_j(\boldsymbol{x}, \boldsymbol{y}) > \alpha_2$, for any two solutions $\boldsymbol{x}, \boldsymbol{y}$ and every objective function $f_j \in \mathcal{F}_I$. $\qquad\square$

Corollary 2 *If $\alpha_2 < \alpha_1$, then $\mathcal{PS}_{\alpha_2}^\gamma \subset \mathcal{PS}_{\alpha_1}^\gamma$.*

Proof The same as the proof for Corollary 1. $\qquad\square$

3.2 α-Degree non-dominated sorting

We use α-degree Pareto dominance relation to develop a version of the well-known non-dominated sorting algorithm [23] that can be applied for solution ranking in noisy environments, when the objectives are given as intervals. The main steps of this ranking method are shown in Algorithm 1.

The algorithm first orders the solutions into a number of α-degree Pareto non-dominated sets, by comparing the solutions of the population with the α-degree Pareto dominance relation. Then, within each α-degree non-dominated set, the solutions are ordered according to their crowding distances in the objective space, which reflects how scattered is each solution in the objective space with respect to the other solutions in the same α-degree non-dominated set. In practice, when this ranking method is used

Inputs:
      A set of solutions P
      A dominance degree α
1 $r \leftarrow 0$
2 **while** there are more solutions in P **do**
3     $r \leftarrow r + 1$
4     $S_r \leftarrow \alpha$-degree Pareto non-dominated solutions of P
5     $P \leftarrow P \setminus S_r$
6 **end while**
7 **for all** $i \in \{1, \ldots, r\}$ **do**
8     $D_i \leftarrow$ Crowding distances of solutions in S_i
9     $S_i \leftarrow$ Reorder the solutions in S_i in decreasing value of D_i
10 **end for**
  Output: $\{S_1, \ldots, S_r\}$

Algorithm 1: The α-Degree Non-Dominated Sorting Algorithm

with techniques like truncation selection, the crowding distance is computed only for the solutions of the α-degree non-dominated set which cannot be added completely to the set of selected solutions.

To compute the crowding distances, the solutions are ordered with respect to each of the objectives individually, and the crowding distance of the solutions with minimum and maximum value for each objective is set to a large number, indicating that these solutions are well scattered with respect to that objective. The crowding distance of other in-between solutions is computed by summing up the normalized distances of each solution to its preceding and succeeding solutions in each objective with respect to the ordering. To order the solutions in each objective dimension, the quick sort algorithm is adapted to work with interval values, comparing two solutions by checking whether $\deg_j(\cdot, \cdot) > 0$. The value of $\deg_j(\cdot, \cdot)$ function is also used as a normalized estimation of the distances between interval values of the solutions in each objective.

4 Multi-objective optimization with joint variable-objective probabilistic modeling

The common practice in most EAs (including EDAs) is to only use the variable values for generating new solutions in the search space. The objective values are used for ranking and selecting a subset of solutions, and apart from that, these algorithms ignore the objective information when generating new solutions. Although this scheme of new solution generation usually offers a relatively good exploration of the search space, the objective information of the selected solutions can be exploited in the new solution generation step of EAs for further improvement, so that the new solutions have better or comparable objective values than their parents.

In the case of EDAs, when objective information is incorporated into probabilistic modeling, the estimated model not only encodes the characteristics of the variable values of the selected solutions, but it also encodes preferences concerning objective values of these solutions. This especially applies to MOPs, where, because of the existence of several objectives, more information about the quality of the solutions is available. EDAs try to represent the problem structure by probabilistically approximat-
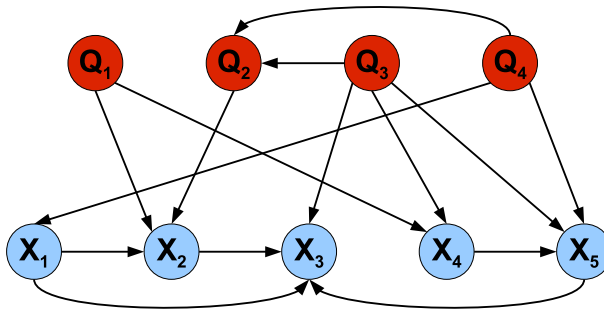
Fig. 2 An example of a multi-dimensional Bayesian network structure

ing the interactions between variables and how their combination of values influence the objective functions. Incorporating the objective values of the solutions in the modeling step of these algorithms will allow them to obtain a probabilistic approximation of the relationships between objectives as well (e.g. based on the expected value of the objectives).

A multi-objective EDA based on this idea is MBN-EDA [14], which uses multi-dimensional Bayesian networks (MBNs), a type of Bayesian networks [40] initially used in multi-dimensional classification [41,42], for joint modeling of variables and objectives. Figure 2 shows an example of an MBN structure. In this type of model, the nodes are organized in two separate layers: the top layer comprises objective nodes and the bottom layer contains variable nodes. The set of arcs in the structure is partitioned into three subsets, resulting in the following subgraphs:

– the class subgraph, containing the objective nodes and the interactions between them,
– the feature subgraph, comprising the variable nodes and their relations, and
– the bridge subgraph, depicting the *top-down* relationships between objective and variable nodes.

The feature subgraph of MBN encodes the relationships between variables like the models learnt by other Bayesian network-based multi-objective EDAs [43–48]. However, the bridge and class subgraphs, encode new types of relationships as the result of joint modeling of variables and objectives. The bridge subgraph shows the interactions between each objective and the variables (selects a subset of relevant variables for each objective), and the class subgraph represents the relationships directly between objectives.

Like other types of Bayesian networks, each node of MBN stores a set of parameters representing the conditional probability distributions for the values of each variable given different value-settings of their parents according to the network structure. In continuous domains, assuming a Gaussian distribution for the joint vector of variables and objectives, $(X, Q) = (X_1, \ldots, X_n, Q_1, \ldots, Q_m)$, where Q_j is a random variable taking values from the range of the jth objective function, i.e. $Q_j = f_j(X)$, the joint probability distribution encoded in the probabilistic model of MBN-EDA can be represented as

$$\rho\,(x_1, \ldots, x_n, q_1, \ldots, q_m) = \prod_{i=1}^{n} \rho(x_i | \boldsymbol{pa}_i) \cdot \prod_{j=1}^{m} \rho(q_j | \boldsymbol{pa}'_j), \qquad (5)$$

where $\boldsymbol{Pa}_i \subseteq \big\{ \{X \setminus X_i\} \cup \boldsymbol{Q} \big\}$ and $\boldsymbol{Pa}'_j \subseteq \{ \boldsymbol{Q} \setminus Q_j \}$ are the parents of each variable and objective, respectively, according to the MBN structure, and $\boldsymbol{pa}_i$ and $\boldsymbol{pa}'_j$ represent one of their possible value-settings. $\boldsymbol{q} = (q_1, \ldots, q_m)$ denotes a possible value-setting for the objective variables $\boldsymbol{Q} = (Q_1, \ldots, Q_m)$.

The presence of noise in objective values, represented with intervals, means that in the joint vector $(X_1, \ldots, X_n, Q_1, \ldots, Q_m)$, a subset of objective variables $\boldsymbol{Q}_I$, where I is the set of objective function indices in $\mathcal{F}_I$, take on interval values. Thus, not all of the values in the joint variable-objective dataset provided for probabilistic modeling are scalar values.

There are some studies in the area of imprecise probabilities and credal sets for estimating a probability distribution for a vector of random variables, when some of these variables take non-scalar values (e.g. set of values or intervals). When Bayesian networks are used to encode this kind of probability distributions they are called credal networks [49]. Because of the inherent complexity of this type of models, the methods proposed for their learning and inference are usually very time consuming.

To be able to iteratively perform the joint modeling of variables and objectives in each generation of MBN-EDA, we have used a simple approach for learning a probabilistic model in the presence of noisy objectives with interval values. Before learning the joint model, all of the interval values are replaced by representative scalar values. For example, when the values returned by noisy objective functions are considered to be confidence intervals, a good representative value for each interval is its estimated expected value, $\hat{\mathbb{E}}(Q_j)$. This way of scalarization is also justified when taking into account the fact that in joint probabilistic modeling, objective values are only used to obtain an approximation of the influence of objectives on each other and on the variables.

Once the dataset is scalarized, a relatively fast greedy local structure search algorithm [50] is used for learning an MBN from the dataset. The algorithm starts with an initial structure for the network, which can be generated randomly or given based on some prior knowledge of the problem. At each iteration of the algorithm, all possible edge addition, removal and reversal operations that do not violate the constraint for top-down interactions between variable and objective nodes in the bridge subgraph are considered, and the one resulting in the best improvement of a scoring metric is selected and applied to the structure. This step is repeated until no more operations can be found to further improve the scoring metric. If the maximum number of node score evaluations is not reached yet, the structure search is restarted from a new random structure and otherwise the algorithm stops. At the end, the highest scoring Bayesian network in all these sub-searches is returned. We use the Bayesian information criterion (BIC) [51] as the scoring metric, which computes a penalized maximum likelihood of the Bayesian network with respect to the dataset.

New solutions are generated from the estimated MBN using the probabilistic logic sampling (PLS) method [52]. PLS starts with finding an ancestral or topological ordering of the nodes in MBN. In such an ordering, each node appears after its parent nodes

according to the MBN structure. Due to the restrictions imposed on the bridge sub-graph in the learning process, all objective nodes appear before variable nodes in the ancestral ordering obtained for an MBN. Next, the conditional probability distributions encoded in the nodes of MBN are sampled one-by-one according to their order of appearance in the ancestral ordering. This means that objective nodes are treated the same as variable nodes and new values are generated for both variables and objectives when sampling the estimated model. In this way, the PLS algorithm can take into account the approximated values for objectives, leading to a sampling which is more consistent with the probabilities encoded in the estimated model. To generate a set of solutions (or samples) from the estimated MBN this sampling process is repeated.

5 Experiments

In this section we perform an experimental study to test the proposed solution ranking method based on α-degree Pareto dominance relation and examine the optimization performance of an standard multi-objective EA and MBN-EDA when using this ranking method in noisy MOPs. The problems, noise characteristics and other experimental design features are explained in the following sections.

5.1 Noise model

In Sect. 2, we saw that many of the works in the literature simulate the noise in objective functions of an MOP with an additive zero-mean Gaussian distribution:

$$f_j(\boldsymbol{x}_i) + \mathcal{N}(0, \sigma_n^2), \tag{6}$$

where $f_j(\boldsymbol{x}_i)$ is the true value of the jth objective function for solution $\boldsymbol{x}_i$, and σ_n controls the level of noise introduced to the objective functions of MOP, which often varies in the range [0.01, 1]. Arnold and Beyer [53] have also studied several other noise models and their influence on the performance of EAs.

As explained earlier, the noise model in Eq. (6), which is used in many of the previous works, results in independent singular noisy values for the objective functions. To obtain interval values for noisy objectives, as it is assumed in this paper, we can draw several random values ξ_k from the noise model and use these values to compute a confidence interval, based on a specific confidence level γ, for each objective function f_j and each solution $\boldsymbol{x}_i$:

$$\left(f_j(\boldsymbol{x}_i) + \bar{\xi} \ \pm \ z_{\frac{1-\gamma}{2}} \frac{s(\xi)}{\sqrt{K}} \right), \tag{7}$$

where K is the number of random values drawn from the noise model, and $\bar{\xi}$ and $s(\xi)$ are respectively their mean and standard deviation. However, in practice, using such a method to generate interval values for the noisy objective functions imposes a significant computational overhead to the solution evaluation phase, especially for larger values of K. Moreover, this method can reduce the stochasticity of the interval

values generated for the objective functions, as for example with a Gaussian noise model when K increases, the mean and standard deviation of the sampled random values tend to the corresponding parameters of the noise model.

To have a somewhat similar randomness in the generated interval values as in the singular values obtained from Eq. (6), we draw two random values ξ_m and ξ_s from a Gaussian noise model to compute a confidence interval for each objective function f_j and each solution x_i:

$$\left(f_j(x_i) + \xi_m \ \pm \ z_{\frac{1-\gamma}{2}} \xi_s\right).$$ (8)

These two random values can be generated from two different Gaussian distributions. However, as it is explained in [31], it is more reasonable that the level of noise in ξ_m and ξ_s increase and decrease correspondingly. In this study we use similar Gaussian distributions for sampling these two values.

5.2 WFG test problems

Huband et al. [54] reviewed many of the benchmark MOPs proposed in the literature like ZDT [6], DTLZ [55] and OKA [56], and based on the analysis of these problems, they proposed a new set of MOPs called the walking fish group (WFG) problems. These MOPs have a diverse set of properties found in real-world problems and, therefore, can be a great challenge for any multi-objective optimization algorithm. Each objective function f_j of an MOP in this benchmark is defined as

$$\min_z \ f_j(z) = D \cdot z_m + S_j \cdot h_j(z_1, \ldots, z_{m-1}),$$ (9)

where D and S_j are scaling factors and $h_j(\cdot)$ is a shape function, meaning that it will determine the shape of the Pareto optimal front of an MOP (e.g. concave, convex, etc.) together with the shape functions in the definition of other objective functions of that MOP. $z = (z_1, \ldots, z_m)$ is an m-dimensional vector of parameters obtained by applying a number of transformation functions, like shifting, biasing or reduction, to the n-dimensional input solution $x \in \mathbb{D}$.

We have selected 5 MOPs of this benchmark for our experiments in this paper. Theses problems are WFG1, WFG2, WFG3, WFG7 and WFG9, and cover different shapes for the Pareto optimal front. WFG1 has a mixed convex-concave optimal front shape, WFG2 has a disconnected convex front, and WFG3 has a degenerated linear front. WFG7 and WFG9 problems have a concave Pareto optimal front. Besides different shapes, the objective functions of these MOPs have various properties like inseparability, multi-modality and deception. We consider three objectives and 10 variables for all of the MOPs. Noise is introduced to the output of all three objective functions in each MOP using Eq. (8), resulting in a confidence interval for each solution in the search space.

5.3 Experimental design

In the experiments performed in this section, we study and compare two solution ranking methods when the noisy objective values are given as intervals. The first

method is the proposed α-degree non-dominated sorting algorithm (Algorithm 1), hereafter referred to as degree ranking (DR), and tested with three different dominance degrees: $\alpha \in \{0.1, 0.5, 0.9\}$. As the second method, we adopt probabilistic ranking (PR) [19], based on dominance probability (Definition 2), which is often used as a reference solution ranking method in the studies on multi-objective optimization in noisy environments, and ranks each solution x_i as follows:

$$\text{rank}_{PR}(x_i) = \sum_{k=1}^{N} P(x_k \prec x_i) + \frac{1}{2} \sum_{k=1}^{N} P(x_k \equiv x_i) - \frac{1}{2}, \tag{10}$$

where

$$P(x \equiv y) = 1 - P(x \prec y) - P(x \succ y)$$

represents the probability that neither x nor y dominate each other, and N is the population size. The last term in Eq. (10) is subtracted so that

$$\sum_{k=1}^{N} \text{rank}_{PR}(x_k) = \frac{N(N-1)}{2}.$$

This ranking method defines a total ordering between the solutions of population, allowing to sort the solutions based on the ranks assigned by $\text{rank}_{PR}(\cdot)$.

To simplify dominance probability computation in multi-objective case, which involves multivariate integration, the objective functions are assumed to be statistically independent [19,20], and therefore the dominance probability in Eq. (2) is approximated as:

$$P(x \prec y) = \prod_{j=1}^{m} P(U_j < V_j), \tag{11}$$

where $U_j = f_j(x)$ and $V_j = f_j(y)$ are the random variables representing the jth objective values returned for two solutions x and y. Moreover, when noise is modeled as a Gaussian distribution, Hughes [19] proposed an approximation of the univariate integration required in the computation of $P(U_j < V_j)$ to further reduce the computational complexity of the overall solution ranking:

$$P(U_j < V_j) = \frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{U_j - V_j}{\sigma_V\sqrt{2 + 2(\frac{\sigma_U}{\sigma_V})^2}}\right)$$

$$\approx \frac{1}{2} - \frac{1}{2}\tanh\left(\frac{U_j - V_j}{0.8\sigma_V\sqrt{2 + 2(\frac{\sigma_U}{\sigma_V})^2}}\right), \tag{12}$$

where σ_U and σ_V are the standard deviations of the random variables $U_j = f_j(x)$ and $V_j = f_j(y)$, respectively. When objective values are given as confidence intervals, the estimated expected values and half-widths are used as approximations for U_j and

V_j and their standard deviations in the computation of dominance probability with Eq. (12).

These ranking methods are used in MBN-EDA to rank and select a subset of solutions for offspring generation with probabilistic modeling. A deterministic binary tournament selection with replacement strategy is employed to select 50% of the population solutions.

For better comparison, we have also included a standard EMO algorithm based on NSGA-II [23] in the experimentation which uses simulated binary crossover [57] and polynomial mutation [58] for offspring generation and search space exploration. We refer to this algorithm simply as EA in contrast to MBN-EDA which will be indicated as EDA in the experimental results. Both of the algorithms use an elitist replacement strategy, a population of 50 solutions and stop after 300 generations. The initial population is generated randomly in both algorithms by uniform sampling from the variables domain.

For each of the five MOPs tested in our experiments, we have studied three different levels of noise and confidence: $\sigma_n \in \{0.01, 0.1, 1\}$ and $\gamma \in \{0.90, 0.95, 0.99\}$. Therefore, all together, there are $5 \times 3 \times 3 \times 2 \times 4 = 360$ different possible combinations for the experiments. For each combination, 10 independent runs are performed. To evaluate the results of experiments, we have used 5 different quality measures: hypervolume [28], inverted generational distance (IGD) [59], epsilon indicator [60], relative maximum spread [30] and Schott's spacing [61]. In the evaluation process, the Pareto set approximated by each algorithm is used to obtain a noiseless Pareto front which is then used to compute any of these indicators. For this purpose the objective values (without the influence of the noise model) are computed for the solutions in the approximated Pareto sets, resulting in the corresponding noiseless Pareto fronts. Therefore, the noisy Pareto fronts approximated by the algorithms in the course of evolution might be (and usually are) better than the noiseless Pareto fronts, especially for higher levels of noise.

Given a Pareto set approximation A, its hypervolume indicator value with respect to a set R of reference points in the objective space, denoted as $I_h(A, R)$, is given as

$$I_h(A, R) = \bigcup_{x \in A} H(f(x), R), \tag{13}$$

where $H(a, B)$ indicates the (hyper-)volume between point a and the points of set B in the objective space. A higher value for this indicator implies a better Pareto set approximation. We use a single reference point, often called the nadir point, with a value of 100 for all objectives to compute this indicator.

IGD indicator accounts for both the diversity of the approximated front as well as its convergence to the Pareto optimal front. Given a set of points F^*, representing a well-distributed sampling of the Pareto optimal front of an MOP, the IGD value for a Pareto set approximation A is computed as:

$$IGD_{F^*}(A) = \frac{\sum_{r \in F^*} \min_{\forall x \in A} d(r, f(x))}{|F^*|} \tag{14}$$

where d($\cdot$, $\cdot$) gives the Euclidean distance between the two points in objective space. A smaller value for this indicator means a better approximation. A sampling of 10, 000 points is used for representing the Pareto optimal front of the MOPs when computing the values of this indicator. The description of the other three quality indicators is given in the supplementary Online Resource of the paper.

5.4 Results

Tables 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11 show the mean and standard deviation of the hypervolume and IGD indicator values for the final Pareto solutions obtained with different combinations of the two algorithmic frameworks and the studied ranking methods, for the tested MOPs. The corresponding figures showing the average values of hypervolume, IGD, epsilon, relative maximum spread and Schott's spacing indicators for the Pareto solutions obtained along the evolution path are given in the supplementary Online Resource of the paper.

With the increase in the noise level, the change in the position of intervals with respect to the original noiseless values of the objective function will be higher, and also the length of intervals increases. Besides, increasing the confidence level results in wider intervals. Therefore, as the level of noise increases the value of objective functions are more distorted. Because of this, we focus the analysis of the results on the common patterns for the average behavior of the algorithms on different instances (with different levels of noise and confidence) of the tested problems.

In the objective functions of WFG1 problem, the input variables are greatly biased, making an even exploration of the search space very difficult, especially in the presence of noise. According to the employed quality indicators, the solution ranking provided by DR method (with different degrees of dominance) allows to obtain better approximations of Pareto set on most of the problem instances. As the noise level increases, the quality of the Pareto sets obtained using each of the two ranking methods become closer. The increase in the noise level also blurs the differences in the performance of the algorithms using different degrees of dominance for DR method. Moreover, the increase in the confidence level has less influence on DR method when using a high dominance degree like $\alpha = 0.9$.

A closer inspection of the Pareto fronts obtained for WFG1 problem shows that the algorithms are not able to obtain a well covering approximation of the Pareto optimal front. This can also be seen by the values obtained for the relative maximum spread indicator (see the supplementary Online Resource). Actually, EA completely minimizes this indicator along the evolution path for smaller noise levels, as a result of the high bias in the objectives. However, the joint modeling adopted in EDA allows a better exploration of this problem's search space, at least for lower noise levels, which is also reflected by the values obtained for hypervolume and IGD indicators.

The results obtained for WFG2 problem (Tables 4, 5) also show that the Pareto sets approximated with DR method are superior to those obtained using PR method. When a higher degree of dominance is used for the DR method, the performance of the algorithms with respect to the different quality indicators are more sensitive to the increase in the length of the intervals, i.e. when increasing the level of noise or

Table 2 Mean hypervolume values of the final Pareto fronts obtained for WFG1 problem with different levels of noise and confidence

	σ_n	γ	EA				EDA			
			PR	DR(0.1)	DR(0.5)	DR(0.9)	PR	DR(0.1)	DR(0.5)	DR(0.9)
Hypervolume	0.01	0.90	9.537E+05 ±1336	9.557E+05 ±485	9.558E+05 ±1112	9.554E+05 ±941	9.567E+05 ±6222	9.569E+05 ±4184	9.604E+05 ±6967	9.561E+05 ±4345
		0.95	9.538E+05 ±1503	9.557E+05 ±534	9.56E+05 ±704	9.554E+05 ±665	9.545E+05 ±647	9.578E+05 ±5778	9.583E+05 ±5494	9.564E+05 ±4344
		0.99	9.533E+05 ±1028	9.558E+05 ±750	9.556E+05 ±612	9.56E+05 ±1400	9.54E+05 ±491	9.564E+05 ±4380	9.605E+05 ±6012	9.577E+05 ±5366
	0.1	0.90	9.539E+05 ±555	9.545E+05 ±885	9.546E+05 ±691	9.541E+05 ±649	9.542E+05 ±501	9.541E+05 ±538	9.541E+05 ±696	9.545E+05 ±490
		0.95	9.538E+05 ±773	9.543E+05 ±823	9.548E+05 ±826	9.553E+05 ±1239	9.543E+05 ±591	9.544E+05 ±724	9.544E+05 ±541	9.543E+05 ±841
		0.99	9.538E+05 ±542	9.542E+05 ±404	9.543E+05 ±459	9.55E+05 ±934	9.541E+05 ±584	9.544E+05 ±537	9.557E+05 ±4464	9.54E+05 ±414
	1	0.90	9.539E+05 ±705	9.542E+05 ±459	9.539E+05 ±404	9.541E+05 ±585	9.542E+05 ±521	9.541E+05 ±514	9.541E+05 ±478	9.539E+05 ±315
		0.95	9.539E+05 ±317	9.541E+05 ±454	9.539E+05 ±407	9.537E+05 ±205	9.538E+05 ±414	9.54E+05 ±533	9.54E+05 ±331	9.541E+05 ±717
		0.99	9.54E+05 ±741	9.541E+05 ±508	9.544E+05 ±1019	9.539E+05 ±550	9.541E+05 ±489	9.539E+05 ±328	9.54E+05 ±495	9.54E+05 ±812

Table 3 Mean IGD values of the final Pareto fronts obtained for WFG1 problem with different levels of noise and confidence

	σ_n	γ	EA				EDA			
			PR	DR(0.1)	DR(0.5)	DR(0.9)	PR	DR(0.1)	DR(0.5)	DR(0.9)
IGD	0.01	0.90	2.761	2.556	2.555	2.572	2.72	2.556	2.494	2.644
			±0.14	±0.06	±0.11	±0.10	±0.19	±0.10	±0.16	±0.14
		0.95	2.762	2.554	2.532	2.573	2.72	2.576	2.51	2.615
			±0.15	±0.05	±0.08	±0.07	±0.07	±0.16	±0.14	±0.11
		0.99	2.805	2.537	2.544	2.525	2.771	2.613	2.428	2.559
			±0.12	±0.08	±0.07	±0.16	±0.07	±0.13	±0.19	±0.16
	0.1	0.90	2.769	2.696	2.656	2.721	2.752	2.737	2.754	2.706
			±0.06	±0.09	±0.09	±0.09	±0.06	±0.06	±0.09	±0.06
		0.95	2.772	2.7	2.645	2.613	2.745	2.734	2.72	2.73
			±0.07	±0.10	±0.10	±0.14	±0.07	±0.07	±0.08	±0.11
		0.99	2.776	2.695	2.704	2.645	2.763	2.713	2.697	2.761
			±0.05	±0.06	±0.06	±0.10	±0.07	±0.06	±0.11	±0.06
	1	0.90	2.757	2.728	2.761	2.739	2.753	2.749	2.776	2.784
			±0.08	±0.05	±0.05	±0.06	±0.05	±0.07	±0.03	±0.03
		0.95	2.759	2.721	2.765	2.777	2.787	2.76	2.78	2.748
			±0.04	±0.06	±0.05	±0.03	±0.04	±0.06	±0.03	±0.08
		0.99	2.746	2.734	2.682	2.756	2.765	2.768	2.769	2.776
			±0.07	±0.05	±0.12	±0.05	±0.06	±0.05	±0.07	±0.09

Table 4 Mean hypervolume values of the final Pareto fronts obtained for WFG2 problem with different levels of noise and confidence

	σ_n	γ	EA				EDA			
			PR	DR(0.1)	DR(0.5)	DR(0.9)	PR	DR(0.1)	DR(0.5)	DR(0.9)
Hypervolume	0.01	0.90	9.726E+05 ±6096	9.77E+05 ±2035	9.788E+05 ±5061	9.792E+05 ±5044	9.752E+05 ±5196	9.771E+05 ±3999	9.771E+05 ±3288	9.802E+05 ±4071
		0.95	9.713E+05 ±5379	9.771E+05 ±2362	9.766E+05 ±2197	9.787E+05 ±4061	9.732E+05 ±4468	9.789E+05 ±927	9.784E+05 ±1750	9.787E+05 ±1967
		0.99	9.703E+05 ±6919	9.771E+05 ±5242	9.767E+05 ±5584	9.771E+05 ±1919	9.735E+05 ±5169	9.783E+05 ±1669	9.798E+05 ±2325	9.789E+05 ±1352
	0.1	0.90	9.706E+05 ±7679	9.786E+05 ±3893	9.805E+05 ±6358	9.755E+05 ±1582	9.709E+05 ±7983	9.799E+05 ±4698	9.794E+05 ±2857	9.781E+05 ±2206
		0.95	9.698E+05 ±7395	9.76E+05 ±4198	9.79E+05 ±6325	9.76E+05 ±6571	9.708E+05 ±7756	9.787E+05 ±1524	9.783E+05 ±1720	9.774E+05 ±1708
		0.99	9.725E+05 ±6276	9.817E+05 ±3775	9.791E+05 ±3749	9.769E+05 ±2178	9.716E+05 ±5136	9.775E+05 ±3715	9.786E+05 ±1613	9.776E+05 ±1665
	1	0.90	9.673E+05 ±7505	9.71E+05 ±4764	9.698E+05 ±5033	9.726E+05 ±5479	9.704E+05 ±8275	9.731E+05 ±5626	9.744E+05 ±3638	9.704E+05 ±4770
		0.95	9.667E+05 ±10151	9.73E+05 ±4939	9.716E+05 ±6456	9.713E+05 ±3962	9.702E+05 ±7664	9.755E+05 ±2360	9.747E+05 ±5058	9.722E+05 ±5786
		0.99	9.647E+05 ±11084	9.733E+05 ±5837	9.713E+05 ±5028	9.717E+05 ±5617	9.698E+05 ±7624	9.748E+05 ±4837	9.764E+05 ±1510	9.727E+05 ±4034

Table 5 Mean IGD values of the final Pareto fronts obtained for WFG2 problem with different levels of noise and confidence

	σ_n	γ	EA				EDA			
			PR	DR(0.1)	DR(0.5)	DR(0.9)	PR	DR(0.1)	DR(0.5)	DR(0.9)
IGD	0.01	0.90	1.314 ±0.32	1.067 ±0.10	1.118 ±0.10	1.047 ±0.11	1.172 ±0.15	1.002 ±0.15	1.001 ±0.14	0.9254 ±0.09
		0.95	1.377 ±0.34	1.049 ±0.11	1.08 ±0.11	1.059 ±0.10	1.141 ±0.23	0.8819 ±0.07	0.9531 ±0.07	0.9233 ±0.09
		0.99	1.341 ±0.36	1.126 ±0.17	1.122 ±0.15	1.098 ±0.12	1.159 ±0.22	0.9545 ±0.08	0.9067 ±0.04	0.8941 ±0.10
	0.1	0.90	1.405 ±0.42	1.027 ±0.08	1.054 ±0.09	1.097 ±0.09	1.286 ±0.39	0.9248 ±0.09	0.9341 ±0.06	0.9241 ±0.10
		0.95	1.358 ±0.36	1.084 ±0.14	1.077 ±0.10	1.096 ±0.24	1.378 ±0.39	0.8931 ±0.06	0.9195 ±0.06	0.9516 ±0.08
		0.99	1.291 ±0.29	1.024 ±0.12	1.018 ±0.10	1.079 ±0.10	1.265 ±0.23	0.9682 ±0.16	0.9453 ±0.04	0.9508 ±0.08
	1	0.90	1.561 ±0.39	1.272 ±0.22	1.273 ±0.22	1.249 ±0.21	1.362 ±0.40	1.183 ±0.27	1.098 ±0.12	1.247 ±0.19
		0.95	1.543 ±0.61	1.207 ±0.19	1.286 ±0.13	1.221 ±0.21	1.35 ±0.43	1.071 ±0.14	1.116 ±0.19	1.183 ±0.30
		0.99	1.578 ±0.68	1.239 ±0.25	1.226 ±0.24	1.181 ±0.21	1.367 ±0.41	1.125 ±0.19	1.054 ±0.10	1.172 ±0.18

Table 6 Mean hypervolume values of the final Pareto fronts obtained for WFG3 problem with different levels of noise and confidence

	σ_n	γ	EA				EDA			
			PR	DR(0.1)	DR(0.5)	DR(0.9)	PR	DR(0.1)	DR(0.5)	DR(0.9)
Hypervolume	0.01	0.90	9.74E+05	9.826E+05	9.812E+05	9.804E+05	9.742E+05	9.791E+05	9.78E+05	9.792E+05
			±4689	±3846	±3554	±3125	±6641	±3135	±2747	±5620
		0.95	9.683E+05	9.818E+05	9.811E+05	9.819E+05	9.709E+05	9.786E+05	9.795E+05	9.79E+05
			±5360	±2405	±4039	±2239	±5220	±2545	±2658	±2556
		0.99	9.673E+05	9.79E+05	9.813E+05	9.802E+05	9.686E+05	9.802E+05	9.806E+05	9.798E+05
			±9163	±3007	±3068	±3623	±9906	±3713	±4499	±3935
	0.1	0.90	9.711E+05	9.812E+05	9.788E+05	9.803E+05	9.728E+05	9.818E+05	9.818E+05	9.788E+05
			±7856	±1898	±2662	±2746	±4481	±2236	±2581	±2386
		0.95	9.666E+05	9.816E+05	9.817E+05	9.8E+05	9.723E+05	9.809E+05	9.841E+05	9.807E+05
			±12485	±3101	±2567	±2622	±8723	±3309	±2765	±3688
		0.99	9.658E+05	9.805E+05	9.817E+05	9.828E+05	9.736E+05	9.822E+05	9.806E+05	9.816E+05
			±14835	±2847	±2541	±2688	±7060	±2628	±2212	±3692
	1	0.90	9.717E+05	9.802E+05	9.744E+05	9.758E+05	9.709E+05	9.742E+05	9.752E+05	9.738E+05
			±6030	±3141	±4287	±4906	±5199	±4061	±2405	±6740
		0.95	9.68E+05	9.75E+05	9.767E+05	9.747E+05	9.713E+05	9.774E+05	9.754E+05	9.751E+05
			±8701	±6746	±4906	±4633	±5905	±5480	±4647	±4594
		0.99	9.673E+05	9.754E+05	9.737E+05	9.743E+05	9.732E+05	9.728E+05	9.715E+05	9.737E+05
			±7144	±4342	±2878	±7510	±4871	±5962	±4465	±3867

Table 7 Mean IGD values of the final Pareto fronts obtained for WFG3 problem with different levels of noise and confidence

		EA				EDA			
σ_n	γ	PR	DR(0.1)	DR(0.5)	DR(0.9)	PR	DR(0.1)	DR(0.5)	DR(0.9)
IGD									
0.01	0.90	1.217 ±0.14	1.027 ±0.06	1.029 ±0.08	1.032 ±0.07	1.202 ±0.30	1.005 ±0.07	1.05 ±0.05	1.049 ±0.16
	0.95	1.408 ±0.34	1.039 ±0.06	1.057 ±0.07	1.074 ±0.06	1.231 ±0.25	1.028 ±0.06	1.017 ±0.08	0.9873 ±0.05
	0.99	1.459 ±0.56	1.026 ±0.06	1.02 ±0.05	1.08 ±0.10	1.504 ±0.61	0.9959 ±0.08	1.009 ±0.08	1.033 ±0.07
0.1	0.90	1.339 ±0.35	1.046 ±0.04	1.099 ±0.06	1.039 ±0.07	1.204 ±0.18	0.9608 ±0.04	0.9696 ±0.04	0.9911 ±0.02
	0.95	1.605 ±0.78	1.023 ±0.10	1.089 ±0.07	1.026 ±0.07	1.263 ±0.36	1.007 ±0.06	0.962 ±0.04	0.9956 ±0.07
	0.99	1.746 ±0.89	1.022 ±0.04	1.059 ±0.08	1.012 ±0.06	1.228 ±0.30	0.9706 ±0.04	0.9816 ±0.04	1 ±0.08
1	0.90	1.239 ±0.20	1.066 ±0.07	1.166 ±0.16	1.103 ±0.13	1.181 ±0.25	1.1 ±0.07	1.051 ±0.08	1.131 ±0.15
	0.95	1.412 ±0.51	1.164 ±0.18	1.142 ±0.13	1.153 ±0.09	1.164 ±0.16	1.051 ±0.05	1.102 ±0.12	1.062 ±0.10
	0.99	1.4 ±0.39	1.083 ±0.07	1.17 ±0.11	1.206 ±0.18	1.131 ±0.17	1.127 ±0.14	1.179 ±0.15	1.075 ±0.12

Table 8 Mean hypervolume values of the final Pareto fronts obtained for WFG7 problem with different levels of noise and confidence

σ_n	γ	EA				EDA			
		PR	DR(0.1)	DR(0.5)	DR(0.9)	PR	DR(0.1)	DR(0.5)	DR(0.9)
Hypervolume 0.01	0.90	9.636E+05 ±15428	9.798E+05 ±4602	9.792E+05 ±5631	9.792E+05 ±3678	9.717E+05 ±13703	9.765E+05 ±6829	9.744E+05 ±6392	9.785E+05 ±5045
	0.95	9.641E+05 ±15292	9.792E+05 ±3844	9.783E+05 ±7057	9.788E+05 ±4661	9.705E+05 ±11399	9.767E+05 ±5674	9.785E+05 ±6544	9.751E+05 ±5206
	0.99	9.657E+05 ±18711	9.79E+05 ±4818	9.814E+05 ±3416	9.776E+05 ±4186	9.7E+05 ±5074	9.777E+05 ±4949	9.758E+05 ±3820	9.776E+05 ±5862
0.1	0.90	9.732E+05 ±5138	9.822E+05 ±2563	9.825E+05 ±2295	9.774E+05 ±7348	9.655E+05 ±10175	9.784E+05 ±4556	9.787E+05 ±4206	9.783E+05 ±4506
	0.95	9.652E+05 ±9926	9.813E+05 ±3304	9.825E+05 ±4559	9.801E+05 ±4727	9.711E+05 ±4542	9.787E+05 ±4043	9.795E+05 ±4872	9.786E+05 ±4436
	0.99	9.626E+05 ±22110	9.8E+05 ±2771	9.812E+05 ±3439	9.793E+05 ±7210	9.693E+05 ±8528	9.805E+05 ±3611	9.804E+05 ±2868	9.797E+05 ±3692
1	0.90	9.607E+05 ±13048	9.76E+05 ±6423	9.767E+05 ±6085	9.752E+05 ±5204	9.681E+05 ±7863	9.725E+05 ±7561	9.736E+05 ±6038	9.724E+05 ±5855
	0.95	9.656E+05 ±12450	9.793E+05 ±3969	9.744E+05 ±6211	9.758E+05 ±6726	9.67E+05 ±11130	9.736E+05 ±3847	9.744E+05 ±4147	9.707E+05 ±7632
	0.99	9.67E+05 ±15446	9.764E+05 ±4134	9.732E+05 ±6255	9.761E+05 ±7195	9.687E+05 ±9159	9.712E+05 ±6473	9.726E+05 ±4400	9.725E+05 ±4589

Table 9 Mean IGD values of the final Pareto fronts obtained for WFG7 problem with different levels of noise and confidence

σ_n	γ	EA				EDA			
		PR	DR(0.1)	DR(0.5)	DR(0.9)	PR	DR(0.1)	DR(0.5)	DR(0.9)
IGD									
0.01	0.90	1.665	1.001	1.043	1.044	1.274	1.075	1.103	0.9745
		±0.83	±0.11	±0.17	±0.12	±0.64	±0.14	±0.22	±0.14
	0.95	1.669	1.056	1.07	1.055	1.318	1.085	1.002	1.039
		±0.75	±0.13	±0.23	±0.10	±0.50	±0.12	±0.12	±0.12
	0.99	1.706	0.9871	1.004	1.066	1.219	1.024	0.9848	1.006
		±0.88	±0.11	±0.08	±0.08	±0.24	±0.11	±0.11	±0.15
0.1	0.90	1.269	0.9771	0.9505	1.055	1.498	0.9425	0.9286	0.9758
		±0.20	±0.09	±0.09	±0.25	±0.47	±0.09	±0.07	±0.15
	0.95	1.61	0.9345	0.9779	0.9985	1.204	0.973	0.9673	0.9522
		±0.43	±0.10	±0.10	±0.13	±0.20	±0.09	±0.10	±0.10
	0.99	1.963	0.9672	0.9256	0.9956	1.356	0.915	0.9986	0.9638
		±1.20	±0.10	±0.12	±0.13	±0.35	±0.05	±0.11	±0.12
1	0.90	1.76	1.144	1.138	1.143	1.356	1.193	1.136	1.15
		±0.67	±0.13	±0.16	±0.10	±0.34	±0.25	±0.13	±0.16
	0.95	1.529	1.102	1.169	1.156	1.398	1.12	1.116	1.169
		±0.69	±0.15	±0.18	±0.22	±0.49	±0.12	±0.14	±0.17
	0.99	1.526	1.09	1.169	1.148	1.267	1.189	1.126	1.116
		±0.70	±0.14	±0.16	±0.22	±0.34	±0.21	±0.10	±0.13

Table 10 Mean hypervolume values of the final Pareto fronts obtained for WFG9 problem with different levels of noise and confidence

σ_n	γ	EA				EDA			
		PR	DR(0.1)	DR(0.5)	DR(0.9)	PR	DR(0.1)	DR(0.5)	DR(0.9)
Hypervolume 0.01	0.90	9.728E+05	9.804E+05	9.781E+05	9.78E+05	9.764E+05	9.759E+05	9.787E+05	9.787E+05
		±3711	±3605	±2578	±3524	±4977	±3137	±3823	±2031
	0.95	9.745E+05	9.779E+05	9.776E+05	9.808E+05	9.742E+05	9.803E+05	9.783E+05	9.783E+05
		±4802	±3047	±2547	±3229	±6516	±3071	±2552	±2133
	0.99	9.724E+05	9.781E+05	9.792E+05	9.789E+05	9.74E+05	9.785E+05	9.793E+05	9.791E+05
		±4195	±3109	±3144	±2658	±6061	±2864	±1130	±3983
0.1	0.90	9.746E+05	9.789E+05	9.794E+05	9.799E+05	9.738E+05	9.805E+05	9.812E+05	9.795E+05
		±5152	±4341	±2835	±3027	±4811	±3052	±5180	±3735
	0.95	9.732E+05	9.785E+05	9.788E+05	9.781E+05	9.725E+05	9.792E+05	9.793E+05	9.783E+05
		±4368	±2574	±2985	±2259	±5480	±3279	±2097	±2445
	0.99	9.742E+05	9.796E+05	9.775E+05	9.777E+05	9.708E+05	9.787E+05	9.79E+05	9.785E+05
		±4222	±2900	±3296	±2198	±7236	±3313	±2771	±2505
1	0.90	9.738E+05	9.751E+05	9.777E+05	9.75E+05	9.757E+05	9.763E+05	9.778E+05	9.756E+05
		±5035	±3059	±2749	±4684	±3350	±3178	±3817	±5029
	0.95	9.745E+05	9.753E+05	9.77E+05	9.739E+05	9.734E+05	9.772E+05	9.755E+05	9.766E+05
		±8366	±3425	±4717	±2935	±6417	±2835	±6227	±4090
	0.99	9.738E+05	9.731E+05	9.758E+05	9.775E+05	9.735E+05	9.759E+05	9.78E+05	9.755E+05
		±5003	±6386	±6772	±4432	±4779	±4445	±3276	±5390

Table 11 Mean IGD values of the final Pareto fronts obtained for WFG9 problem with different levels of noise and confidence

σ_n	γ	EA				EDA			
		PR	DR(0.1)	DR(0.5)	DR(0.9)	PR	DR(0.1)	DR(0.5)	DR(0.9)
IGD									
0.01	0.90	1.258 ±0.20	1.115 ±0.08	1.147 ±0.14	1.166 ±0.16	1.086 ±0.20	1.044 ±0.06	1.108 ±0.14	1.028 ±0.05
	0.95	1.333 ±0.23	1.166 ±0.09	1.175 ±0.07	1.06 ±0.11	1.196 ±0.21	1.016 ±0.04	1.056 ±0.07	1.054 ±0.07
	0.99	1.387 ±0.14	1.144 ±0.10	1.083 ±0.12	1.103 ±0.10	1.191 ±0.20	1.06 ±0.11	1.02 ±0.08	1.025 ±0.10
0.1	0.90	1.28 ±0.19	1.146 ±0.11	1.095 ±0.10	1.055 ±0.05	1.219 ±0.16	0.9553 ±0.08	0.9375 ±0.07	0.9802 ±0.08
	0.95	1.319 ±0.16	1.12 ±0.04	1.107 ±0.13	1.131 ±0.06	1.147 ±0.14	0.951 ±0.08	0.9807 ±0.06	1.023 ±0.09
	0.99	1.319 ±0.18	1.045 ±0.14	1.155 ±0.08	1.116 ±0.06	1.251 ±0.27	1.007 ±0.08	0.9755 ±0.06	0.9956 ±0.09
1	0.90	1.26 ±0.20	1.13 ±0.10	1.108 ±0.12	1.146 ±0.13	1.062 ±0.12	1.057 ±0.08	1.019 ±0.07	1.021 ±0.10
	0.95	1.326 ±0.24	1.18 ±0.09	1.142 ±0.18	1.189 ±0.14	1.171 ±0.14	1.022 ±0.07	1.083 ±0.15	1.081 ±0.09
	0.99	1.368 ±0.21	1.242 ±0.18	1.14 ±0.13	1.122 ±0.10	1.08 ±0.11	1.064 ±0.09	1.036 ±0.09	1.102 ±0.12

confidence. In the last objective of this problem, the variables are considered to be correlated and therefore, employing model learning to find out this relationships helps EDA to obtain better Pareto set approximations under different levels of noise and confidence. Considering the values obtained by hypervolume and IGD indicators on the one hand and relative maximum spread indicator (see the supplementary Online Resource) on the other hand, we can conclude that although the Pareto fronts approximated by EA are better spread, but these fronts are farther from the Pareto optimal front than those obtained by EDA.

Very similar results to those of the previous problem are obtained for WFG3 problem. The difference in the average performance of the algorithms using PR and DR methods are clearer on this problem, and it can be seen that the solution ranking provided by PR method completely misguides the algorithms during evolution for approaching the Pareto optimal set according to the quality indicators, even with small levels of noise and confidence (see Tables 6, 7). The DR method using smaller dominance degrees results in relatively better algorithm performance on different levels of noise and confidence with respect to the quality indicators used. The indicators values also show that EA obtains relatively better results on WFG3 problem, though for larger noise levels the results obtained by EDA are comparable or better than those reached by EA.

The objective functions of WFG7 problem are separable and unimodal, with bias in the optimum values for some of the variables. The results obtained with DR method on this problem outperform those of PR method with a relatively great margin, although the increase in the noise level reduces the effectiveness of this ranking method, especially when using it with EDA. Again, the Pareto fronts approximated with lower dominance degrees for DR method are better. However, it can be seen that when the confidence level increases the performance of the algorithms using a high level of dominance degree is less affected. In general, the search with EA results in relatively better Pareto front approximations comparing with the fronts obtained by EDA-based search according to the quality indicators values. Moreover, with the increase in the level of noise or confidence, which makes the problem harder, the results obtained by the two algorithmic frameworks become comparable.

The objective functions of WFG9 problem in contrast to those of the previous problem are non-separable and multi-modal, with multi-modality being deceptive towards local optima for some of the objectives. As a result introducing noise to the objectives of this problem will make it very difficult to solve. The indicator values computed for the approximated Pareto fronts show that the proposed DR method outperforms solution ranking based on dominance probability, though with the increase in the noise level the difference in the results obtained by the two methods gradually diminishes.

According to the results (Tables 10, 11), lower dominance degrees like $\alpha = 0.1$ and $\alpha = 0.5$ allow a better ranking of solutions in DR method. Moreover, on the contrary to the previous problem, when the confidence level is increased DR method with higher dominance degree ($\alpha = 0.9$) is also greatly affected. As it is shown in the supplementary Online Resource, for this problem both EDA and EA based search with DR method are able to cover a high percentage of the Pareto optimal front (more than 80%). Comparing these two algorithmic frameworks, it is observed that

the Pareto fronts approximated by EDA are comparable or better to those obtained by EA, depending on the quality indicator. This better performance can be explained by the ability of the proposed EDA to capture the relationships between variables of the problem, which is necessary for finding the solutions.

5.5 Effect of population size

In addition to the previous experimentations, to make sure that our findings generalize to larger population sizes we have studied the effect of population size on the performance of the algorithms. For this purpose, different combinations of algorithmic frameworks and ranking methods are applied for optimization on a specific problem instance with different population sizes and the final noiseless Pareto fronts are compared. The experiments are conducted using WFG1 problem with noise level $\sigma_n = 0.1$ and confidence level $\gamma = 0.95$ while testing the algorithms with the following population sizes: $N \in \{50, 100, 150, 200, 250, 300\}$. Figures 3 and 4 respectively show the hypervolume and IGD indicators values for final noiseless Pareto fronts approximated with different population sizes.

As the results show, with larger population sizes better indicator values can be reached, although the difference is not significant. The distinction between the quality of the fronts approximated by DR and PR methods becomes clearer when the population size increases, showing that with DR method better results can be obtained in both algorithmic frameworks and with different values of the dominance degree. Indeed with EA, as the population size increases the difference in the quality of the fronts approximated with different values of dominance degree almost fade away while becoming considerably superior to the fronts approximated with PR method, according to the quality indicators used. The results also show that the direct manipulation of solutions in EA-based search is more affected by the increase in population size comparing with the model learning in the EDA-based search. This can be explained by the fact that the model learning method employed in this algorithm (using regularization techniques) is designed to be less sensible to the dimensionality of the training data. Furthermore, it can be seen that in this algorithmic framework strict condition for accepting solutions as non-dominated (i.e. when $\alpha = 0.9$) in DR method results in better Pareto fronts as the population size increases.

5.6 Discussion

The values of the quality indicators for the Pareto sets obtained along the evolution path (see the supplementary Online Resource), show how the population of different algorithm versions evolve. In general, based on the change in the values of these indicators during evolution, we can see that the solution ranking provided by DR method guides the optimization in the correct direction through the search space of the tested MOPs, with some exceptions like WFG1 and WFG3 when a high level of noise exists. On the contrary, when PR method is used for ranking the solutions, the algorithm is misguided in the search space of many of the tested problems. In probabilistic dominance the probability that a solution dominates another solution is
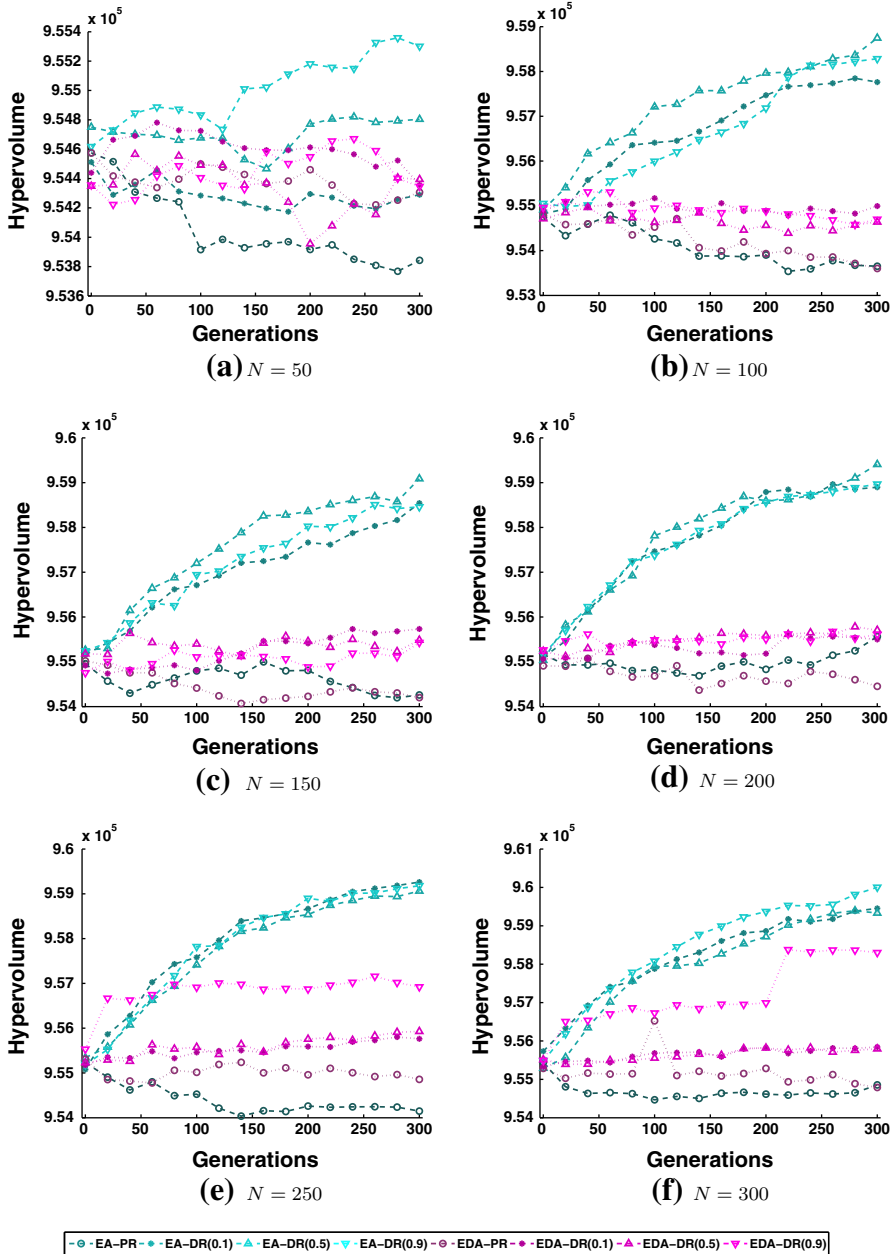
Fig. 3 Average Hypervolume values of the Pareto fronts obtained for WFG1 problem with different population sizes when noise level is $\sigma_n = 0.1$ and confidence level is $\gamma = 0.95$

computed based on the distribution of corresponding intervals of the two solutions. In a multi-objective scenario, when these probabilities are summed up in Eq. (10) and due to the complex combinations that the distributions might have, the ranks generated for
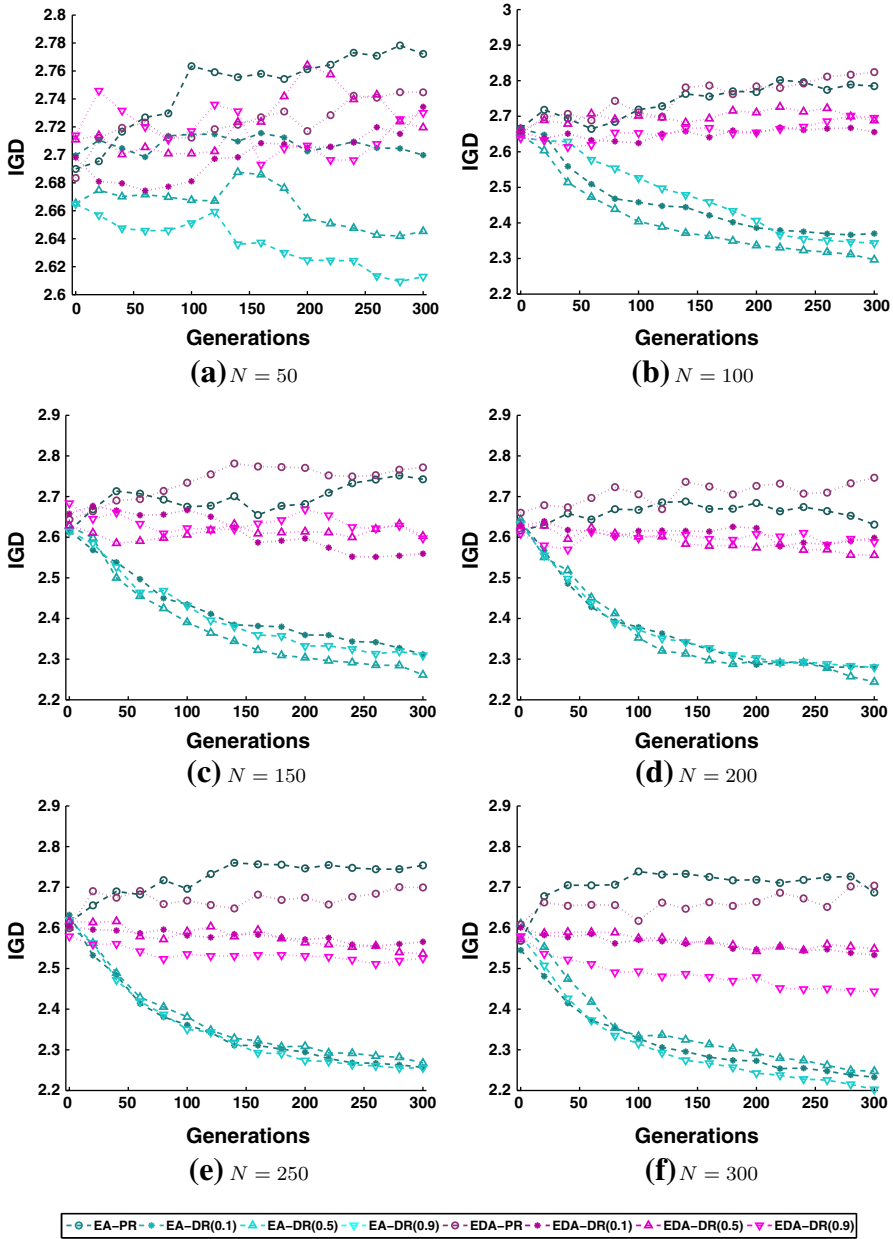
Fig. 4 Average IGD values of the Pareto fronts obtained for WFG1 problem with different population sizes when noise level is $\sigma_n = 0.1$ and confidence level is $\gamma = 0.95$

the solutions can be very close or even equal, increasing the possibility of discarding good solutions in the selection process.

Moreover, the approximation used in Eq. (12) to compute probabilistic dominance because of the high computational complexity of the original formulation naturally

produces small probability values, which in turn can increase the previous effect (generation of similar ranks). It is worth mentioning that even computing the approximation of dominance probability requires considerably more time than deciding dominance between solutions using α-degree Pareto dominance. Furthermore, our experimental results show that since confidence level is not considered in probabilistic dominance, with the increase in confidence level the ranking provided based on dominance probabilities is not able to correctly guide the search.

When using DR method, the Pareto fronts approximated with smaller to medium values of dominance degree α are better on most of the tested problems with different levels of noise and confidence. On the one hand, smaller values of α allow higher degrees of overlapping between intervals, meaning that the solutions can dominate each other easier. With small values of α, according to Corollary 2 the ratio of dominated solutions increases and therefore more solutions become comparable, resulting in a finer ordering of the solutions by the employed α-degree non-dominated sorting algorithm. This increase in the number of comparable solutions seems to compensate for the lack of strict decisions on the quality of the solutions. On the other hand, strict decisions about solution dominance allow only very good solutions to enter the α-degree non-dominated set, while causing many solutions to become incomparable. Based on the experimental results, with a combination of the previous two effects which is achieved when using medium values of dominance degree (e.g. $\alpha = 0.5$), better Pareto set can be approximated on many of the tested MOPs and for different levels of noise and confidence. When the population size increases and more solutions are available in the ranking step, the latter effect (strict decisions) becomes more usable, enhancing the elitism selection in α-degree non-dominated sorting algorithm.

Similarly, as the confidence level of the intervals (i.e. the values of the objective functions) decreases, more and more solutions become dominated (Corollary 1), and thus the solutions will be ordered in a larger number of α-degree non-dominated sets during the first step of Algorithm 1, essentially leading to better solution ranking. Therefore, when DR method is used in the optimization, the Pareto sets approximated for lower confidence levels are usually better, according to the quality indicator values. Taking more strict decisions by using high dominance degrees (e.g. $\alpha = 0.9$) when the confidence level increases, reduces the number of α-degree non-dominated sets found in the first step of Algorithm 1, and thus the ordering of the solutions will depend more on their crowding distance computed in the second step of α-degree non-dominated sorting algorithm. For some of the tested problems (e.g. WFG1 and WFG7), such an ordering with high α values seems to be less affected as the confidence level increases.

Based on our analysis and experimental results obtained with different values of dominance degree, we can say that when working with low levels of noise and confidence that result in smaller interval values and in small populations, smaller values of the α parameter can cause the algorithm to advance better in the noise-affected search space. However, with larger intervals as the values of objective functions and when larger population sizes are available, setting α to larger values can help to faster reach better solutions during the search.

6 Conclusions

In real-world multi-objective optimization, the value of objective functions may involve noise and thus do not correctly represent the quality of solutions. Population-based EAs, which are one of the most successful methods in solving MOPs, have inherent abilities to deal with small levels of noise in objective values. However, with larger noise levels, especially when there are several conflicting objectives to be optimized, specific considerations are needed to perform successful optimization.

We considered noisy objective values given as intervals and proposed α-degree Pareto dominance to deal with this type of values for the objective functions. This relation allowed to determine the dominance between solutions even when there is a specific degree of overlapping between some of the interval values, which is controlled by parameter α. The similarity between this dominance relation and the conventional Pareto dominance allowed related terminology, like Pareto optimal set, to be easily defined. Assuming confidence intervals with a specific confidence level as the values of the noisy objective functions, we studied some of the properties of α-degree Pareto dominance. It was shown that the α-degree dominance relations defined between solutions are unaffected when the confidence level γ or dominance degree α are reduced, and that such a reduction decreases the size of Pareto optimal set.

Based on this relation, α-degree Pareto non-dominated sorting algorithm, an adaptation of the well-known non-dominated sorting algorithm in noisy domains with interval values, was proposed and integrated into a multi-objective EDA based on joint probabilistic modeling of variables and objectives to find the solutions of noisy MOPs. The algorithm was tested on a set of MOPs where noisy objective values are given as intervals, and its solution ranking and search space exploration were respectively compared with a reference ranking method based on dominance probability and standard EA operators for continuous optimization.

The analysis of the approximated Pareto sets with five different quality indicators based on the corresponding noiseless Pareto fronts showed that the proposed solution ranking method based on α-degree Pareto dominance relation allows the algorithms to achieve considerably better results comparing with the well-known probabilistic ranking method on the tested problems and for increasing levels of noise and confidence. We discussed how the change in the dominance degree and confidence level changes comparability of the solutions using α-degree Pareto dominance relation and saw that small to medium degrees of dominance allow better ranking of the solutions on most of the tested instances. Moreover, depending on the specific properties of the MOPs, the quality indicator values showed that the joint probabilistic modeling of variables and objectives allowed MBN-EDA to find Pareto set approximations for some of the tested problems that are superior to those approximated by the standard multi-objective EA in different levels of noise.

As future works, other forms of degree function $\deg_j(\cdot, \cdot)$ can be studied for determining the overlapping between intervals by for example taking into account the distribution of the values in the intervals. Moreover, it would be worth considering the correlations between objectives and their noise values in the ranking procedure. A starting point for this line of research is to use the model learning similar to the one used in MBN-EDA. Dynamic adjustment of the dominance control parameter, α, is

also an interesting future research. We also intend to apply the developed methodology on a real-world problem. Specifically, we are considering the application of the proposed approach to the feature subset selection problem in machine learning and data mining tasks. The objectives of this problem are uncertain and can very naturally be represented as intervals. This would also require adapting the joint probabilistic modeling to domains with discrete variables and objective values given as intervals.

References

1. Pareto, V.: The new theories of economics. J. Polit. Econ. **5**(4), 485–502 (1897)
2. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning, 1st edn. Addison-Wesley Longman Publishing Co., Inc., Boston (1989)
3. Abraham, A., Jain, L., Goldberg, R. (eds.): Evolutionary Multiobjective Optimization: Theoretical Advances and Applications. Springer, London (2005)
4. Coello Coello, C.A., Lamont, G.B., Van Veldhuizen, D.A.: Evolutionary Algorithms for Solving Multiobjective Problems, 2nd edn. Springer, Berlin (2007)
5. Deb, K.: Multi-objective Optimization Using Evolutionary Algorithms. Wiley, London (2001)
6. Zitzler, E., Deb, K., Thiele, L.: Comparison of multiobjective evolutionary algorithms: empirical results. Evol. Comput. **8**(2), 173–195 (2000)
7. Larrañaga, P., Lozano, J.A. (eds.): Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation. Kluwer Academic Publishers, Norwell (2001)
8. Lozano, J.A., Larrañaga, P., Inza, I., Bengoetxea, E. (eds.): Towards a New Evolutionary Computation: Advances on Estimation of Distribution Algorithms, vol. 192. Studies in Fuzziness and Soft Computing. Springer (2006)
9. Mühlenbein, H., Paaß, G.: From recombination of genes to the estimation of distributions I. Binary parameters. In: Fourth International Conference on Parallel Problem Solving from Nature (PPSN IV), vol. 1141. Lecture Notes in Computer Science, pp. 178–187. Springer (1996)
10. Pelikan, M.: Hierarchical Bayesian optimization algorithm. Toward a new generation of evolutionary algorithms, vol. 170. In: Studies in Fuzziness and Soft Computing, 1st ed. Springer (2005)
11. Martí, L., Garcia, J., Berlanga, A., Coello Coello, C.A., Molina, J.M.: On current model-building methods for multi-objective estimation of distribution algorithms: shortcommings and directions for improvement. Tech. Rep. GIAA2010E001, Department of Informatics, Universidad Carlos III de Madrid (2010)
12. Pelikan, M., Sastry, K., Goldberg, D.: Multiobjective estimation of distribution algorithms. In: Scalable Optimization via Probabilistic Modeling, vol. 33. Studies in Computational Intelligence, pp. 223–248. Springer (2006)
13. Thierens, D., Bosman, P.A.N.: Multi-objective mixture-based iterated density estimation evolutionary algorithms. In: Conference on Genetic and Evolutionary Computation (GECCO '01), pp. 663–670. Morgan Kaufmann (2001)
14. Karshenas, H., Santana, R., Bielza, C., Larrañaga, P.: Multi-objective estimation of distribution algorithm based on joint modeling of objectives and variables. IEEE Trans. Evol. Comput. **18**(4), 519–542 (2014)
15. Jin, Y., Branke, J.: Evolutionary optimization in uncertain environments—a survey. IEEE Trans. Evol. Comput. **9**(3), 303–317 (2005)
16. Tan, K., Goh, C.: Handling uncertainties in evolutionary multi-objective optimization. Computational Intelligence: Research Frontiers, vol. 5050. Lecture Notes in Computer Science, pp. 262–292. Springer (2008)
17. Soares, G., Guimaraes, F., Maia, C., Vasconcelos, J., Jaulin, L.: Interval robust multi-objective evolutionary algorithm. In: IEEE Congress on Evolutionary Computation (CEC'09), pp. 1637–1643 (2009)
18. Goh, C., Tan, K., Cheong, C., Ong, Y.: An investigation on noise-induced features in robust evolutionary multi-objective optimization. Expert Syst. Appl. **37**(8), 5960–5980 (2010)

19. Hughes, E.: Evolutionary multi-objective ranking with uncertainty and noise. In: Evolutionary Multi-criterion Optimization (EMO'01), vol. 1993. Lecture Notes in Computer Science, pp. 329–343. Springer (2001)

20. Teich, J.: Pareto-front exploration with uncertain objectives. In: Evolutionary Multi-Criterion Optimization, vol. 1993. Lecture Notes in Computer Science, pp. 314–328. Springer (2001)

21. Greiner, D., Galván, B., Aznárez, J., Maeso, O., Winter, G.: Robust design of noise attenuation barriers with evolutionary multiobjective algorithms and the boundary element method. In: Evolutionary Multi-criterion Optimization, vol. 5467. Lecture Notes in Computer Science, pp. 261–274. Springer (2009)

22. Bui, L.T., Abbass, H.A., Essam, D.: Fitness inheritance for noisy evolutionary multi-objective optimization. In: Conference on Genetic and Evolutionary Computation (GECCO'05), pp. 779–785. ACM (2005)

23. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)

24. Bui, L.T., Essam, D., Abbass, H.A., Green, D.: Performance analysis of evolutionary multi-objective optimization methods in noisy environments. Complex. Int. **11**, 29–39 (2005)

25. Zitzler, E., Laumanns, M., Thiele, L.: SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization. In: Evolutionary Methods for Design Optimization and Control with Applications to Industrial Problems (EUROGEN '01), pp. 95–100. International Center for Numerical Methods in Engineering (2001)

26. Fieldsend, J., Everson, R.: Multi-objective optimisation in the presence of uncertainty. IEEE Congr. Evol. Comput. (CEC'05) **1**, 243–250 (2005)

27. Büche, D., Stoll, P., Dornberger, R., Koumoutsakos, P.: Multiobjective evolutionary algorithm for the optimization of noisy combustion processes. IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev. **32**(4), 460–473 (2002)

28. Zitzler, E., Thiele, L.: Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Trans. Evol. Comput. **3**(4), 257–271 (1999)

29. Babbar, M., Lakshmikantha, A., Goldberg, D.E.: A modified NSGA-II to solve noisy multiobjective problems. In: Genetic and Evolutionary Computation Conference (GECCO'03), pp. 21–27 (2003)

30. Goh, C., Tan, K.: An investigation on noisy environments in evolutionary multiobjective optimization. IEEE Trans. Evol. Comput. **11**(3), 354–381 (2007)

31. Eskandari, H., Geiger, C.: Evolutionary multiobjective optimization in noisy problem environments. J. Heuristics **15**(6), 559–595 (2009)

32. Bui, L.T., Abbass, H.A., Essam, D.: Localization for solving noisy multi-objective optimization problems. Evol. Comput. **17**(3), 379–409 (2009)

33. Syberfeldt, A., Ng, A., John, R.I., Moore, P.: Evolutionary optimisation of noisy multi-objective problems using confidence-based dynamic resampling. Eur. J. Oper. Res. **204**(3), 533–544 (2010)

34. Park, T., Ryu, K.R.: Accumulative sampling for noisy evolutionary multi-objective optimization. In: 13th Annual Conference on Genetic and Evolutionary Computation (GECCO'11), pp. 793–800. ACM (2011)

35. Mehnen, J., Trautmann, H., Tiwari, A.: Introducing user preference using desirability functions in multi-objective evolutionary optimisation of noisy processes. In: IEEE Congress on Evolutionary Computation (CEC'07), pp. 2687–2694 (2007)

36. Woźniak, P.: Preferences in evolutionary multi-objective optimisation with noisy fitness functions: hardware in the loop study. In: International Multiconference on Computer Science and Information Technology, pp. 337–346 (2007)

37. Kaji, H., Ikeda, K., Kita, H.: Uncertainty of constraint function in evolutionary multi-objective optimization. In: IEEE Congress on Evolutionary Computation (CEC'09), pp. 1621–1628 (2009)

38. Basseur, M., Zitzler, E.: Handling uncertainty in indicator-based multiobjective optimization. Int. J. Comput. Intell. Res. **2**(3), 255–272 (2006)

39. Boonma, P., Suzuki, J.: A confidence-based dominance operator in evolutionary algorithms for noisy multiobjective optimization problems. In: 21st International Conference on Tools with Artificial Intelligence (ICTAI'09), pp. 387–394 (2009)

40. Pearl, J.: Bayesian networks: a model of self-activated memory for evidential reasoning. In: 7th Conference of the Cognitive Science Society, pp. 329–334 (1985)

41. Bielza, C., Li, G., Larrañaga, P.: Multi-dimensional classification with Bayesian networks. Int. J. Approximate Reasoning **52**(6), 705–727 (2011)

42. de Waal, P., van der Gaag, L.: Inference and learning in multi-dimensional Bayesian network classifiers. In: Symbolic and Quantitative Approaches to Reasoning with Uncertainty, pp. 501–511 (2007)

43. Ahn, C.W., Ramakrishna, R.S.: Multiobjective real-coded Bayesian optimization algorithm revisited: diversity preservation. In: 9th Annual Conference on Genetic and Evolutionary Computation (GECCO '07), pp. 593–600. ACM (2007)

44. Katsumata, Y., Terano, T.: Bayesian optimization algorithm for multi-objective solutions: application to electric equipment configuration problems in a power plant. IEEE Congr. Evol. Comput. (CEC '03) **2**, 1101–1107 (2003)

45. Khan, N., Goldberg, D.E., Pelikan, M.: Multiple-objective Bayesian optimization algorithm. In: Conference on Genetic and Evolutionary Computation (GECCO '02), p. 684. Morgan Kaufmann (2002)

46. Laumanns, M., Očenášek, J.: Bayesian optimization algorithms for multi-objective optimization. In: 7th International Conference on Parallel Problem Solving from Nature (PPSN VII), vol. 2439. Lecture Notes in Computer Science, pp. 298–307. Springer (2002)

47. Pelikan, M., Sastry, K., Goldberg, D.E.: Multiobjective hBOA, clustering, and scalability. In: Conference on Genetic and Evolutionary Computation (GECCO '05), pp. 663–670. ACM (2005)

48. Schwarz, J., Očenášek, J.: Multiobjective Bayesian optimization algorithm for combinatorial problems: theory and practice. Neural Netw. World **11**(5), 423–442 (2001)

49. Corani, G., Antonucci, A., Zaffalon, M.: Bayesian networks with imprecise probabilities: theory and application to classification. Tech. Rep. IDSIA-02-10, Dalle Molle Institute for Artificial Intelligence (2010)

50. Buntine, W.: Theory refinement on Bayesian networks. In: 7th Annual Conference on Uncertainty in Artificial Intelligence (UAI '91), pp. 52–60. Morgan Kaufmann (1991)

51. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**(2), 461–464 (1978)

52. Henrion, M.: Propagating uncertainty in Bayesian networks by probabilistic logic sampling. In: Second Annual Conference on Uncertainty in Artificial Intelligence (UAI '86), vol. 2, pp. 149–163. Elsevier (1986)

53. Arnold, D., Beyer, H.G.: A general noise model and its effects on evolution strategy performance. IEEE Trans. Evol. Comput. **10**(4), 380–391 (2006)

54. Huband, S., Hingston, P., Barone, L., While, L.: A review of multiobjective test problems and a scalable test problem toolkit. IEEE Trans. Evol. Comput. **10**(5), 477–506 (2006)

55. Deb, K., Thiele, L., Laumanns, M., Zitzler, E.: Scalable multi-objective optimization test problems. IEEE Congr. Evol. Comput. (CEC '02) **1**, 825–830 (2002)

56. Okabe, T., Jin, Y., Olhofer, M., Sendhoff, B.: On test functions for evolutionary multi-objective optimization. In: 8th International Conference on Parallel Problem Solving from Nature (PPSN VIII), vol. 3242. Lecture Notes in Computer Science, pp. 792–802. Springer (2004)

57. Deb, K., Agrawal, B.: Simulated binary crossover for continuous search space. Complex Syst. **9**(2), 115–148 (1995)

58. Deb, K., Goyal, M.: A combined genetic adaptive search (GeneAS) for engineering design. Comput. Sci. Inf. **26**(4), 30–45 (1996)

59. Coello Coello, C.A., Cortés, N.C.: Solving multiobjective optimization problems using an artificial immune system. Genet. Program. Evol. Mach. **6**, 163–190 (2005)

60. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C., da Fonseca, V.: Performance assessment of multi-objective optimizers: an analysis and review. IEEE Trans. Evol. Comput. **7**(2), 117–132 (2003)

61. Schott, J.R.: Fault Tolerant Design Using Single and Multicriteria Genetic Algorithm Optimization. Master's thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics (1995)