Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 33(6): 3968–3988.
DOI: 10.3934/era.2025176
Received: 19 March 2025
Revised: 24 May 2025
Accepted: 04 June 2025
Published: 24 June 2025

Research article

Performance metrics evaluation of multi-objective optimization for transportation

Hongyu Zhang[1], Wen Yi[2], Haoran Li[3,*] and Shuaian Wang[1]

[1] Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong 999077, China

[2] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong 999077, China

[3] College of Economics and Management, Shandong University of Science and Technology, Qingdao 266590, China

* **Correspondence:** Email: lihaoransd@sdust.edu.cn.

Abstract: Various multi-objective optimization methods are widely applied in the transportation field, enabling decision-makers to find solutions that balance trade-off objectives. Since comparing the performance of multi-objective optimization methods is generally difficult, many performance metrics are introduced to quantitatively evaluate the performance of multi-objective optimization methods. However, the effectiveness of the performance metrics needs to be further investigated. Thus, we first critically analysed a series of performance metrics, including number of solutions obtained (NOSO), overall nondominated solutions number (ONSN), normalized maximum spread (NMS), error ratio (ER), nearest ideal distance (NID), mean ideal distance (MID), spacing (SP), inverted generational distance (IGD), and hypervolume-based ratio (HR), which were extensively adopted to assess the performance of multi-objective optimization methods. We found that these performance metrics cannot always accurately reflect the quality of solutions obtained and may be misleading. Thereafter, two axioms were proposed to define the criteria for reliable performance metrics. Additionally, whether these performance metrics satisfied the two axioms was rigorously proved. The performance metrics that satisfied both axioms, i.e., NOSO, ONSN, NMS, ER, and HR, were considered reliable. Furthermore, a real-world cargo transportation case was investigated, indicating the unreliability of metrics MID and SP.

Keywords: multi-objective optimization; performance metrics; transportation; Pareto optimality

1. Introduction

As a major source of greenhouse gas (GHG) emissions, the sustainable operation of transportation has received increasing attention from researchers and practitioners [1, 2]. Addressing sustainable transportation problems often requires careful consideration of multiple objectives, such as reducing operational costs, reducing environmental pollution, improving customer satisfaction with transportation services, and enhancing resilience of transportation systems. Optimizing sustainable operational decisions not only improves transportation efficiency, but also mitigates environmental pollution problems, thereby promoting the prosperity and sustainability of the transportation industry. However, since different objectives are often conflicting [3, 4], how to balance the trade-off objectives is a crucial problem both theoretically and practically. In earlier studies, only the economically relevant objective (e.g., reducing operational costs) was usually taken as the sole objective of the optimization model, while other objectives (e.g., reducing GHG emissions and enhancing customer satisfaction with transportation services) were considered as constraints of the optimization model. Such optimization models overemphasize the importance of economically relevant objectives but pay little attention to other objectives [5]. To address this, many researchers have adopted multi-objective optimization models to generate a series of compromise solutions, i.e., Pareto optimal solutions, providing more scientific decision support for transportation systems planning and operations. The Pareto optimal solutions are also called efficient, nondominated, or noninferior solutions [6, 7]. The set of objective values corresponding to all Pareto optimal solutions is the Pareto front, also called the Pareto frontier [8, 9].

Numerous multi-objective optimization methods are designed to find Pareto optimal solutions. Notably, assessing the performance of multi-objective optimization methods is complicated. For single-objective optimization problems, the objective values could be compared to assess the performance of each method. However, since multi-objective optimization problems aim to find solutions that balance trade-off objectives, evaluating the performance of multi-objective optimization methods by directly comparing the objective values of obtained solutions is generally hard [10], except in special cases. For example, it is assumed that two methods, namely, method 1 and method 2, can solve a multi-objective optimization problem. If all the solutions found by method 1 are dominated by solutions found by method 2, it is easy to draw the conclusion that method 2 outperforms method 1. Another case is that for any solution found by method 1 that is not dominated by solutions found by method 2, there must be a solution found by method 2 that has the same objective values as the solution found by method 1 in all dimensions. Additionally, at least one nondominated solution found by method 2 has objective values different from those of solutions found by method 1. In this case, method 2 is definitely better than method 1. However, for most cases where it is difficult to compare multi-objective optimization methods, appropriate performance metrics are of great importance.

Various multi-objective optimization methods can be used to solve multi-objective optimization problems in the transportation field. Selecting appropriate methods for the transportation multi-objective optimization problems is crucial, highlighting the practical significance of performance metrics. The major contributions of this study are as follows. First, we introduce a series of performance metrics that are commonly used to evaluate the performance of multi-objective optimization methods. Whether a performance metric can truly reflect the quality of the obtained solutions determines its effectiveness. Hence, whether larger or smaller values of the performance metrics necessarily imply better performance of the multi-objective optimization methods is critically analyzed. Then, the attributes of reliable performance metrics are defined with two axioms, and whether each performance metric introduced in this study aligns with the two axioms is rigorously proved. On this basis, a real-world multi-objective vehicle routing problem is explored, indicating the unreliability of several metrics.

The remainder of this study is outlined as follows. In Section 2, we provide a thorough review of relevant studies. In Section 3, we elaborate on a series of performance metrics for multi-objective optimization methods and critically analyze these performance metrics. In Section 4, we propose two axioms to evaluate the reliability of the performance metrics and mathematically prove the compliance of performance metrics with two axioms. In Section 5, we explore a real-world multi-objective optimization problem in the transportation field. In Section 6, we summarize the study and suggest directions for future studies.

2. Literature review

Since we evaluate the performance metrics of multi-objective optimization methods, the related studies are reviewed from two key aspects: multi-objective optimization methods and metrics for evaluating the performance of multi-objective optimization methods.

Solving multi-objective optimization problems is more challenging than solving single-objective ones, since multi-objective optimization problems need to balance multiple conflicting objectives [11]. Various multi-objective optimization methods have been proposed in studies, which can be classified into three major categories: classical methods [12], meta-heuristic methods [13], and artificial intelligence (AI) techniques [14]. The weighting method and the ϵ-constraint method are popular classical multi-objective optimization methods. Chen et al. [15] develop an effective group train operation plan by enhancing the grouping scheme, stopping scheme, and running schedule. The weighting method is applied to solve the scheduling problem, aiming to reduce transportation cost and total cargo travel time. Elmi et al. [16] focus on the ship schedule recovery problem, aiming to minimize the total late ship arrivals and total profit loss caused by disruptive events. An ϵ-constraint-based algorithm is designed to generate Pareto optimal solutions. The second category of multi-objective optimization methods is meta-heuristic methods, which progressively approach the theoretical Pareto optimal fronts of the multi-objective optimization problems [17]. Zhao et al. [18] propose a nondominated sorted genetic algorithm-III-differential evolution (NSGA–III–DE) algorithm to address the collaborative optimization of the urban built environment and public transportation structure, improving both the services capacity and emission reduction potential of public transportation. Additionally, AI techniques are also extensively applied in multi-objective optimization. Jia et al. [19] focus on the speed optimization of battery-powered electric trucks while considering the objectives of safety, efficiency, comfort, and battery degradation. To address the complicated multi-objective optimization problem, a twin delayed deep deterministic policy gradient-mixture of experts (TD3-MoE) reinforcement learning (RL) method is designed, which is shown to achieve superior performance. Huang et al. [20] develop a human as AI mentor-based deep RL (HAIM-DRL) framework to optimize traffic flow efficiency while avoiding potential accidents. The experimental results validate the advantages of their methods over traditional methods, such as imitation learning (IL), safe RL, and conventional human-in-the-loop RL.

Performance metrics can quantify the performance of multi-objective optimization methods. The existing metrics are designed to evaluate the quality of solutions generated by multi-objective optimization methods from three perspectives: number of solutions, convergence of solutions, and diversity of solutions [21]. Specifically, the number of solutions determines the number of available choices. Hence, a multi-objective optimization method that can find more solutions is considered better. The convergence of solutions indicates the closeness between the obtained solution set and the theoretical Pareto optimal front. Multi-objective optimization methods whose solution sets are closer to the theoretical Pareto optimal front tend to be preferred. The diversity of solutions, namely, the distribution and spread of solutions, reflects the relative distance between the obtained solutions and the coverage of the solution set [22]. A multi-objective optimization method performs better if solutions are more evenly distributed and the solution set covers a broader range. In addition to quantitatively assessing the performance of multi-objective optimization methods, performance metrics are also applied to the design of metric-based multi-objective evolutionary algorithms [23–25] and the setting of iteration stopping criteria for multi-objective optimization methods [26, 27]. Readers may refer to [28, 29] for more detailed information on performance metrics of multi-objective optimization methods.

Multi-objective optimization methods seek to balance multiple trade-off objectives simultaneously. Although many performance metrics have been designed to assess the performance of multi-objective optimization methods quantitatively, whether these performance metrics can draw true and valid conclusions requires in-depth investigation. Only a few researchers claim that some performance metrics may be misleading since they cannot always accurately reflect the quality of solutions obtained [29, 30]. In addition, the criteria for reliable performance metrics are not defined. To address these issues, we critically analyze a series of performance metrics. On this basis, we propose two axioms to define the criteria for reliable performance metrics and provide mathematical proofs to validate whether these performance metrics comply with the two axioms.

3. Overview of performance metrics

In this section, we consider a multi-objective optimization problem, where each objective is to be minimized. Let G denote the number of objective dimensions of the multi-objective optimization problem, defined as: $(\min_{\boldsymbol{x}\in X} O_1(\boldsymbol{x}), \ldots, \min_{\boldsymbol{x}\in X} O_G(\boldsymbol{x}))$, where X is a non-empty compact set and $O_1(\boldsymbol{x}), \ldots, O_G(\boldsymbol{x})$ are lower semi-continuous functions over X. If $\boldsymbol{x} \in X$, $\boldsymbol{x}$ is a feasible solution and $(O_1(\boldsymbol{x}), \ldots, O_G(\boldsymbol{x}))$ is a feasible objective vector. A feasible solution $\hat{\boldsymbol{x}}$ is Pareto optimal if there does not exist any solution $\boldsymbol{x} \in X$ such that $O_g(\boldsymbol{x}) \le O_g(\hat{\boldsymbol{x}})$ for each $g \in \{1, \ldots, G\}$ and $O_g(\boldsymbol{x}) < O_g(\hat{\boldsymbol{x}})$ for at least one g [31]. A set K of methods indexed by k can generate solutions. Let S_k denote the set of solutions with nondominated objective vectors obtained by k method. To ensure the rigor of analysis, it is assumed that each S_k contains at least two solutions, and the number of solutions in each S_k is finite. By merging the obtained solution sets of all methods, a union S is generated, namely, $S = S_1 \cup \ldots \cup S_{|K|}$. If multiple solutions with identical objective vectors are found by different methods, only one is retained in S. Let $\boldsymbol{x}$ denote a solution in S_k or S. The maximum and minimum values of the gth objective obtained by all methods in K are represented by $O_g^{\max}$ and $O_g^{\min}$, respectively. Specifically, $O_g^{\max} = \max_{\boldsymbol{x}\in S} O_g(\boldsymbol{x})$, $O_g^{\min} = \min_{\boldsymbol{x}\in S} O_g(\boldsymbol{x})$. Without loss of generality, we assume $O_g^{\max} \neq O_g^{\min}$ for all g. Otherwise, objectives failing to satisfy this condition are excluded from subsequent analysis. Some solutions in S_k may be dominated by other solutions in S. $e_{\boldsymbol{x}}$ is defined as the binary indicator, which equals 1 if and only if solution $\boldsymbol{x}$ is dominated by any other solution in S, and 0 otherwise.

Several metrics for evaluating the performance of multi-objective optimization methods, including number of solutions obtained (NOSO), overall nondominated solutions number (ONSN), normalized maximum spread (NMS), error ratio (ER), nearest ideal distance (NID), mean ideal distance (MID), spacing (SP), inverted generational distance (IGD), and hypervolume-based ratio (HR), are introduced as follows.

3.1. Number of solutions obtained

The value of NOSO(S_k) is the number of obtained solutions in S_k, namely, $|S_k|$. Metric NOSO defined in this study is also referred to as overall nondominated vector generation in studies, such as [32]. It is possible that two nondominated solutions have the same objective vector and then the number of obtained nondominated solutions may be larger than that of nondominated objective vectors. Here, if multiple solutions found by k method have identical objective vectors, only one of them is retained in S_k. Therefore, our definition of NOSO is identical to that of overall nondominated vector generation in existing studies.

An ideal value of metric NOSO is supposed to be as large as possible because a small value of metric NOSO would limit the available options for decision-makers. However, a method that performs excellently on this metric may actually be worse in terms of overall performance compared to other methods.

Example 1. *For a bi-objective optimization problem, Figure 1 shows the objective values of the obtained solutions by method 1 and method 2 ($S = S_1 \cup S_2$). There are 6 obtained solutions in S_1 and 3 obtained solutions in S_2. Hence, NOSO(S_1) = 6, NOSO(S_2) = 3. Although the value of NOSO(S_1) is larger than that of NOSO(S_2), the obtained solutions in S_1 are all dominated by those in S_2. Therefore, the conclusion that method 1 is superior to method 2 cannot be drawn. It is obvious that method 2 outperforms method 1.*

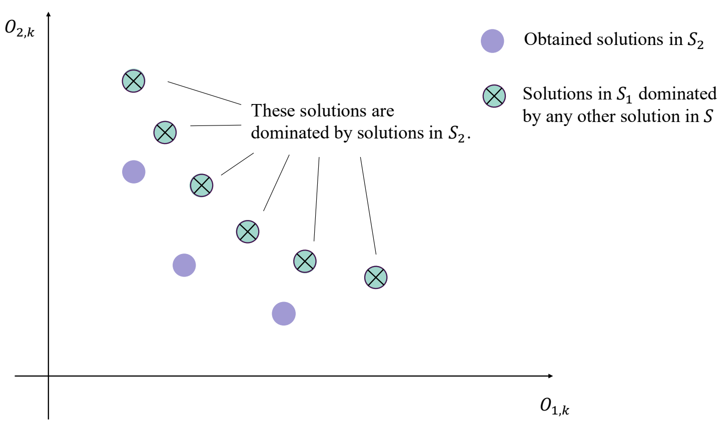


Figure 1. An example illustrating that a larger value of NOSO(S_k) does not always signify better performance.

3.2. *Overall nondominated solutions number*

Metric ONSN is used to evaluate the quality of solutions found by a method compared to other methods. For each k method, the number of solutions in S_k that are not dominated by any other solutions in S is the value of metric ONSN [32], i.e., ONSN(S_k). The value of ONSN(S_k) is defined as follows :

$$\text{ONSN}(S_k) = |S_k| - \sum_{\boldsymbol{x}\in S_k} e_{\boldsymbol{x}}. \tag{1}$$

A larger value of metric ONSN does not necessarily mean that the method performs better, as the value of ONSN(S_k) cannot reflect the distribution range of obtained solutions in S_k.

Example 2. *An optimization problem with two objectives is considered. As shown in Figure 2, method 1 finds 8 solutions, one of which is dominated by solutions in S_2. S_2 contains 6 solutions, with one being dominated by solutions in S_1. Additionally, a solution is in both S_1 and S_2. Hence, ONSN(S_1) = 7, ONSN(S_2) = 5. However, method 1 does not outperform method 2. Although the value of ONSN(S_1) is larger than that of ONSN(S_2), the obtained solutions by method 2 are more evenly distributed.*

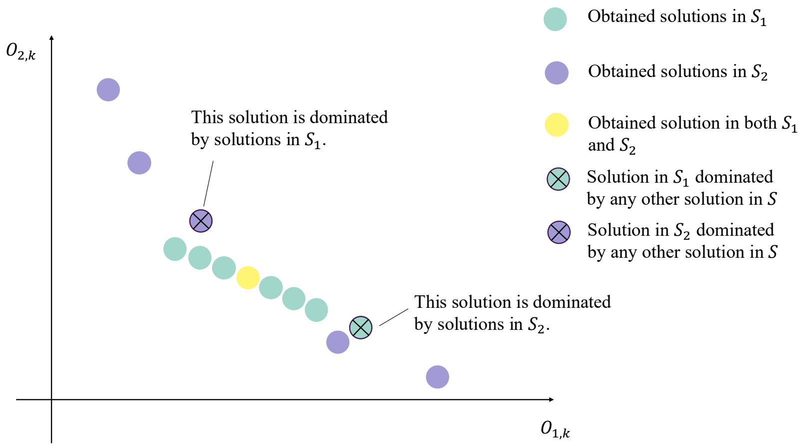


Figure 2. An example illustrating that a larger value of ONSN(S_k), or a smaller value of ER(S_k) does not necessarily indicate better performance.

3.3. *Normalized maximum spread*

NMS is a metric used to evaluate the coverage capability of a solution set across each objective dimension, reflecting the diversity of the solution set. The value of metric NMS can be calculated according to formula (2), which is in line with Amirian and Sahraeian [33]:

$$\text{NMS}(S_k) = \sqrt{\frac{1}{G} \cdot \sum_{g=1}^{G} \left[\frac{\max_{\boldsymbol{x}\in S_k} O_g(\boldsymbol{x}) - \min_{\boldsymbol{x}\in S_k} O_g(\boldsymbol{x})}{O_g^{\max} - O_g^{\min}} \right]^2}. \tag{2}$$

The value of NMS(S_k) is expected to be close to 1, indicating a broad distribution of obtained solutions across all objective dimensions. However, a method with a greater value of metric NMS does not always outperform other methods with smaller values of metric NMS, as illustrated in Example 3.

Example 3. *We consider an optimization problem with two objectives. Figure 3 shows that the obtained solutions in S_2 cover a larger range in both objective dimensions, resulting in the value of NMS(S_2) closer to 1. However, method 2 is not better than method 1, because S_1 contains more obtained solutions and these solutions are more evenly distributed.*

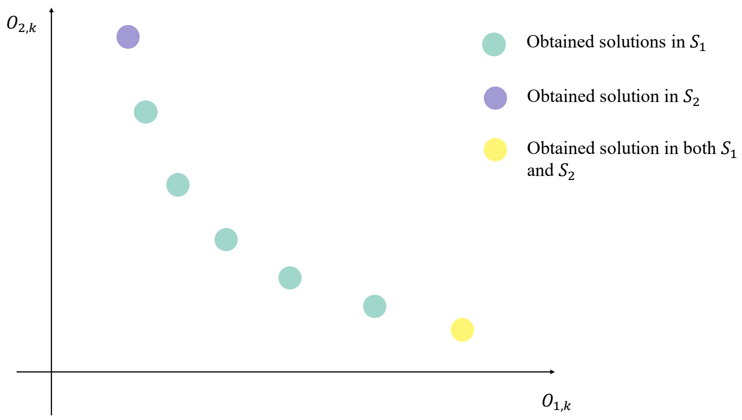


Figure 3. An example showing that a value of NMS(S_k) that is closer to 1 does not necessarily guarantee better performance.

3.4. Error ratio

Metric ER is designed for evaluating the quality of solutions found by a method compared to that of solutions obtained by other methods. Hence, the solutions obtained by other methods have a significant impact on metric ER. The value of ER(S_k) is the average of $e_{\boldsymbol{x}}$ for $\boldsymbol{x} \in S_k$ [33], which is defined as follows:

$$\mathrm{ER}(S_k) = \frac{\sum_{\boldsymbol{x} \in S_k} e_{\boldsymbol{x}}}{|S_k|}. \tag{3}$$

Generally, smaller values of metric ER are preferred, indicating better performance of the obtained solutions. However, a method with a smaller value of metric ER does not always perform better.

Example 4. *As shown in Figure 2, ER(S_1) = $\frac{1}{8}$, ER(S_2) = $\frac{1}{6}$. Although ER(S_1) < ER(S_2), the solutions found by method 2 are more evenly distributed. Therefore, method 1 is not better than method 2.*

3.5. Nearest ideal distance and mean ideal distance

Metrics NID and MID describe the proximity of the obtained solution set to the ideal point, indicating the convergence of the obtained solution set. For some multi-objective optimization problems, such as a multi-objective optimization problem that minimizes the fuel cost and travel time, the ideal point is defined as the zero vector in the G-dimensional space. However, the ideal points of other multi-objective optimization problems may be difficult to define, such as a multi-objective optimization problem that maximizes the profit and minimizes the late deliveries.

Let $C_{\boldsymbol{x}}$ denote the Euclidean objective distance between solution $\boldsymbol{x}$, $\boldsymbol{x} \in S_k$, and the ideal point. That is, $C_{\boldsymbol{x}} = \sqrt{\sum_{g=1}^{G} O_g^2(\boldsymbol{x})}$. The value of NID($S_k$) is the minimum Euclidean objective distance of solutions in S_k to the ideal point, which is defined as follows:

$$\text{NID}(S_k) = \min_{\boldsymbol{x} \in S_k} C_{\boldsymbol{x}}. \tag{4}$$

The value of MID(S_k) is the average Euclidean objective distance between all solutions in S_k and the ideal point [34], which is defined as follows:

$$\text{MID}(S_k) = \frac{\sum_{\boldsymbol{x} \in S_k} C_{\boldsymbol{x}}}{|S_k|}. \tag{5}$$

Smaller values of NID(S_k) and MID(S_k) are preferred. However, if the value of NID(S_k) or MID(S_k) is small, the k method does not necessarily perform well.

Example 5. *A bi-objective optimization problem is considered. As shown in Figure 4, both the minimum Euclidean objective distance and average Euclidean objective distance of obtained solutions in S_1 to the ideal point are smaller than those of solutions in S_2. Hence, NID(S_1) < NID(S_2), MID(S_1) < MID(S_2). However, method 1 does not outperform method 2 because the number of solutions obtained by method 2 is greater than that in S_1, and solutions in S_2 demonstrate a broader distribution.*

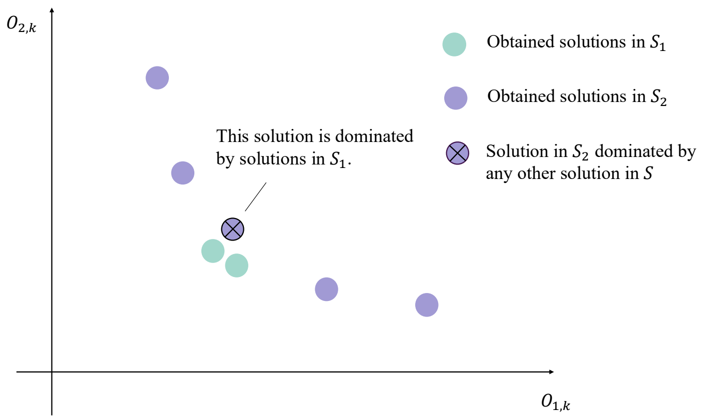


Figure 4. An example illustrating that a smaller value of NID(S_k) or MID(S_k) does not necessarily indicate better solutions.

3.6. Spacing

Metric SP measures the uniformity of the distribution of obtained solutions [35], demonstrating the diversity performance. Let $d_{k,\boldsymbol{x}}$ represent the Manhattan objective distance between solution $\boldsymbol{x}$, $\boldsymbol{x} \in S_k$, and its nearest solution in S_k, which is defined below:

$$d_{k,\boldsymbol{x}} = \min_{\boldsymbol{x}' \in S_k, \boldsymbol{x}' \neq \boldsymbol{x}} \left\{ \sum_{g=1}^{G} |O_g(\boldsymbol{x}) - O_g(\boldsymbol{x}')| \right\} \quad \forall \boldsymbol{x} \in S_k. \tag{6}$$

Let $\bar{d}_k$ denote the average of $d_{k,\boldsymbol{x}}$ for $\boldsymbol{x} \in S_k$. Hence, $\bar{d}_k$ is defined as $\frac{\sum_{\boldsymbol{x} \in S_k} d_{k,\boldsymbol{x}}}{|S_k|}$.

According to Wang et al. [36], the value of SP(S_k) is defined as follows:

$$\text{SP}(S_k) = \sqrt{\frac{1}{|S_k| - 1} \cdot \sum_{\boldsymbol{x} \in S_k} (d_{k,\boldsymbol{x}} - \bar{d}_k)^2}. \tag{7}$$

In general, smaller values of metric SP indicate that the obtained solutions are more uniformly spaced. However, a method with a smaller value of this metric is not necessarily better than other methods.

Example 6. *Figure 5 shows the objective values of a bi-objective optimization problem by method 1 and method 2. The obtained solutions in S_1 are more uniformly spaced, and therefore, SP(S_1) < SP(S_2). However, it cannot be concluded that method 1 outperforms method 2, because method 2 has more obtained solutions with a broader coverage.*

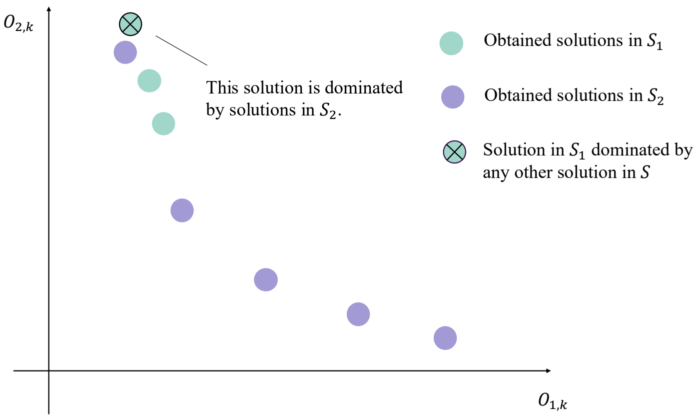


Figure 5. An example illustrating that a smaller value of SP(S_k) does not necessarily indicate better performance.

3.7. Inverted generational distance

The Euclidean objective distance between solution $\boldsymbol{x}$, $\boldsymbol{x} \in S_k$, and the nearest solution $\boldsymbol{x}'$ in S_k is denoted by $L_{k,\boldsymbol{x}}$, which is defined as follows:

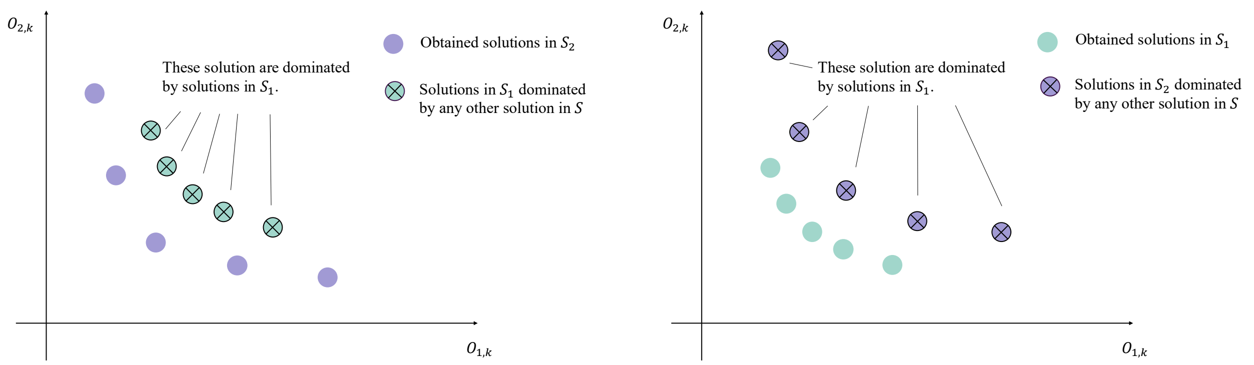
$$L_{k,\boldsymbol{x}} = \min_{\boldsymbol{x}' \in S_k, \boldsymbol{x}' \neq \boldsymbol{x}} \sqrt{\sum_{g=1}^{G} \left(O_g(\boldsymbol{x}) - O_g(\boldsymbol{x}')\right)^2} \quad \forall \boldsymbol{x} \in S_k. \tag{8}$$

Metric IGD is adjusted based on Mirjalili et al. [37], and the value of IGD(S_k) is defined as follows:

$$\mathrm{IGD}(S_k) = \frac{\sqrt{\sum_{\boldsymbol{x} \in S_k} L_{k,\boldsymbol{x}}^2}}{|S_k|}. \tag{9}$$

However, a larger or smaller value of IGD(S_k) does not necessarily indicate better performance of k method. Therefore, this metric is difficult to accurately reflect the performance of solutions in S_k.

Example 7. *Figure 6 shows different results of a bi-objective optimization problem by method 1 and method 2. Both Figure 6(a) and Figure 6(b) show that S_1 and S_2 contain the same number of solutions, but the more clustered distribution of solutions in S_1 leads to IGD(S_1) < IGD(S_2). However, neither of these two figures demonstrates that one method is definitively better than the other.*



(a) An example illustrating that a smaller value of IGD(S_k) does not necessarily indicate better performance.

(b) An example illustrating that a larger value of IGD(S_k) does not necessarily indicate better performance.

Figure 6. An example illustrating that a smaller or a larger value of IGD(S_k) does not necessarily indicate better performance.

3.8. Hypervolume-based ratio

HR is also a metric for assessing the performance of obtained solutions. Let $\boldsymbol{p}$ represent a reference objective vector in the G-dimensional space, with its objective value in the gth dimension equal to $O_g^{\max}$. Let $r_{\boldsymbol{x}}$ represent the hyperrectangle (i.e., G-dimensional interval) defined by $\boldsymbol{p}$ and $\boldsymbol{x}$. The hypervolumes of S_k and S are defined as the hypervolumes of the union of $r_{\boldsymbol{x}}$ for $\boldsymbol{x} \in S_k$ and the union of $r_{\boldsymbol{x}}$ for $\boldsymbol{x} \in S$, denoted by h_k and h, respectively. The metric is enhanced on the basis of Audet et al. [29], and the value of HR(S_k) is defined as follows:

$$\mathrm{HR}(S_k) = \frac{h_k}{h}. \tag{10}$$

Metric HR reflects both convergence and diversity performance of the obtained solutions. In general, a larger value of HR(S_k) is preferred, but larger HR(S_k) does not always show that k method performs better. A deficiency of this metric is that the solution $\boldsymbol{x}$, $\boldsymbol{x} \in S_k$, with its gth objective value $O_g(\boldsymbol{x})$ equal to $O_g^{\max}$ has no contribution to both h_k and h.

Example 8. *We consider a bi-objective optimization problem. In Figure 7, the orange point represents the reference objective vector. h_1 is the hypervolume of the union of the yellow and green regions, and h_2 is the hypervolume of the yellow region. Hence, h_1 is larger than h_2, and $HR(S_1) > HR(S_2)$. However, method 1 does not outperform method 2 because solutions in S_2 have a broader distribution.*

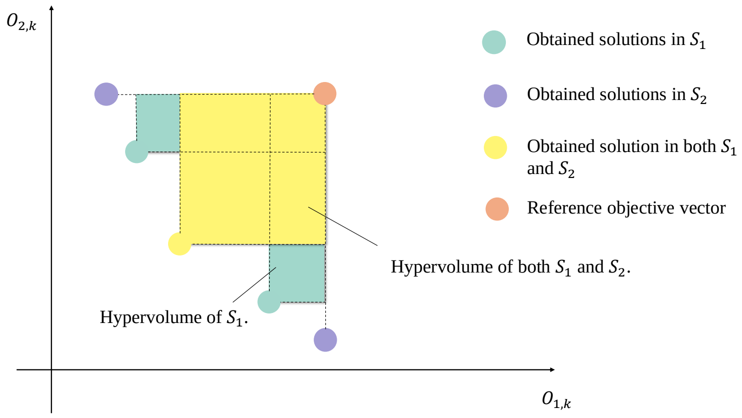


Figure 7. An example illustrating that a larger value of HR(S_k) does not necessarily indicate better performance.

As the number of objective dimensions increases, finding h_k and h becomes time-consuming. To address this, a method for approximately estimating the value of HR(S_k) is designed. In a hyperrectangle defined by $\boldsymbol{p}$ and the origin as vertices, a large number (e.g., 1 million) of uniformly distributed points are randomly generated. The numbers of points that are dominated by at least a solution in S_k and S are denoted by n_k and n, respectively. Then, the value of $\frac{n_k}{n}$ is an approximate value of HR(S_k). Note that the approximation accuracy enhances as the number of points generated in the hyperrectangle increases.

3.9. Summary

In this section, we analyze several performance metrics critically. It is worth noting that all the metrics discussed cannot truly reflect the quality of solutions obtained. To be specific, larger or smaller values of the performance metrics do not necessarily mean that the corresponding multi-objective optimization methods perform better. In addition, as can be seen from the above analysis, adopting a single performance metric to reflect all the properties of a solution set of a multi-objective optimization problem is difficult, but we can analyze the obtained solutions in detail from different perspectives through various performance metrics, thus evaluating the overall performance of multi-objective optimization methods more comprehensively.

4. Criteria for reliable performance metrics

In this section, the attributes of reliable performance metrics are investigated. The quality of a solution is unchanged when expressed in different units, or adjusted for benchmark values. In other words, after the same scale or translation transformations, the quality of the solution sets generated by different multi-objective optimization methods remains unchanged. Therefore, the relative ranking of the solution sets reflected by reliable performance metrics should also remain unchanged.

Two axioms are proposed to provide criteria for reliable performance metrics. First, the obtained solutions take on different values depending on the measurement units adopted, but the quality of these solutions should remain unaffected. For example, one ton of carbon dioxide (CO_2) emissions is equivalent to 1000 kilograms of CO_2 emissions. Based on this idea, Axiom 1 is proposed. Let c_g denote a positive coefficient in the gth dimension.

Axiom 1. ***(Scale-invariant metrics)*** *Substituting the gth objective value for each solution **x** (i.e., $O_g(\boldsymbol{x})$) with $c_g O_g(\boldsymbol{x})$, where $c_g > 0$, the comparison results regarding the performance of various methods remain unchanged.*

Second, adding the same constants to objective values of all solutions in a certain dimension does not affect the performance of these solutions. For instance, for the objective of minimizing annual variable cost, adding the same annual fixed cost to the original objective would not change the relative ranking of all solutions in this dimension. Hence, Axiom 2 is proposed. Let b_g and $\mathbb{R}$ represent a real number in the gth dimension and the set of real numbers, respectively.

Axiom 2. ***(Translation-invariant metrics)*** *Substituting the gth objective value for each solution **x** (i.e., $O_g(\boldsymbol{x})$) with $O_g(\boldsymbol{x}) + b_g$, where $b_g \in \mathbb{R}$, the comparison results regarding the performance of various methods remain unchanged.*

In order to investigate whether the performance metrics introduced in Section 3 comply with these two axioms, the following theorems are proposed.

Theorem 1. *Metric NOSO satisfies both Axiom 1 and Axiom 2.*

Proof. Select any two solutions $\boldsymbol{x}$ and $\boldsymbol{x}'$ from S_k. Since $\boldsymbol{x}$ is not dominated by $\boldsymbol{x}'$, we assume $O_{g_1}(\boldsymbol{x}) < O_{g_1}(\boldsymbol{x}')$, $O_{g_2}(\boldsymbol{x}) > O_{g_2}(\boldsymbol{x}')$, $g_1, g_2 \in \{1, 2, \ldots, G\}, g_1 \neq g_2$. Since $c_{g_1} > 0$ and $c_{g_2} > 0$, it can be deduced that $c_{g_1} O_{g_1}(\boldsymbol{x}) < c_{g_1} O_{g_1}(\boldsymbol{x}')$, $c_{g_2} O_{g_2}(\boldsymbol{x}) > c_{g_2} O_{g_2}(\boldsymbol{x}')$. Additionally, it can be inferred that $O_{g_1}(\boldsymbol{x}) + b_{g_1} < O_{g_1}(\boldsymbol{x}') + b_{g_1}$, $O_{g_2}(\boldsymbol{x}) + b_{g_2} > O_{g_2}(\boldsymbol{x}') + b_{g_2}$, where $b_{g_1}, b_{g_2} \in \mathbb{R}$. After shifting objectives by linear scaling or adding constants, the relative ordering of objective values of $\boldsymbol{x}$ and $\boldsymbol{x}'$ remains unchanged. Hence, $\boldsymbol{x}$ is not dominated by $\boldsymbol{x}'$. It can be concluded that shifting objectives by linear scaling or adding constants does not affect the dominance relationships among solutions in S_k. Therefore, the number of solutions in S_k is unchanged.

Theorem 2. *Metrics ONSN and ER satisfy both Axiom 1 and Axiom 2.*

Proof. Let $\boldsymbol{x}$ represent a solution in S_k and $\boldsymbol{x}^\#$ represent a solution in S, $\boldsymbol{x} \neq \boldsymbol{x}^\#$. Suppose $\boldsymbol{x}$ is not dominated by $\boldsymbol{x}^\#$, $O_{g_1}(\boldsymbol{x}) < O_{g_1}(\boldsymbol{x}^\#)$, $O_{g_2}(\boldsymbol{x}) > O_{g_2}(\boldsymbol{x}^\#)$, $g_1, g_2 \in \{1, 2, \ldots, G\}, g_1 \neq g_2$. It can be deduced that $c_{g_1} O_{g_1}(\boldsymbol{x}) < c_{g_1} O_{g_1}(\boldsymbol{x}^\#)$, $c_{g_2} O_{g_2}(\boldsymbol{x}) > c_{g_2} O_{g_2}(\boldsymbol{x}^\#)$, $O_{g_1}(\boldsymbol{x}) + b_{g_1} < O_{g_1}(\boldsymbol{x}^\#) + b_{g_1}$, $O_{g_2}(\boldsymbol{x}) + b_{g_2} > O_{g_2}(\boldsymbol{x}^\#) + b_{g_2}$, where $c_{g_1}, c_{g_2} > 0, b_{g_1}, b_{g_2} \in \mathbb{R}$. Suppose $\boldsymbol{x}$ is dominated by $\boldsymbol{x}^\#$, $O_{g_1}(\boldsymbol{x}) < O_{g_1}(\boldsymbol{x}^\#)$, $O_{g_2}(\boldsymbol{x}) \leq O_{g_2}(\boldsymbol{x}^\#)$. It can be inferred that $c_{g_1} O_{g_1}(\boldsymbol{x}) < c_{g_1} O_{g_1}(\boldsymbol{x}^\#)$, $c_{g_2} O_{g_2}(\boldsymbol{x}) \leq c_{g_2} O_{g_2}(\boldsymbol{x}^\#)$, $O_{g_1}(\boldsymbol{x}) + b_{g_1} < O_{g_1}(\boldsymbol{x}^\#) + b_{g_1}$, $O_{g_2}(\boldsymbol{x}) + b_{g_2} \leq O_{g_2}(\boldsymbol{x}^\#) + b_{g_2}$. The relative ordering of objective values of $\boldsymbol{x}$ and $\boldsymbol{x}^\#$ is unchanged. As a result, the dominance relationship between $\boldsymbol{x}$ and any other solution $\boldsymbol{x}^\#$ in S remains unchanged. Recall that Theorem 1 proves that $|S_k|$ is unchanged after linear scaling or adding constants. Therefore, metrics ONSN and ER remain unchanged after shifting objectives by linear scaling or adding constants.

Theorem 3. *Metric NMS satisfies both Axiom 1 and Axiom 2.*

Proof. After linear transformation, the gth objective value for $\boldsymbol{x}$ is scaled from $O_g(\boldsymbol{x})$ to $c_g O_g(\boldsymbol{x})$. Let $\text{NMS}'(S_k)$ denote the value of metric NMS after linear scaling. $\text{NMS}'(S_k)$ can be calculated as follows:

$$\begin{aligned}\text{NMS}'(S_k) &= \sqrt{\frac{1}{G}\cdot\sum_{g=1}^{G}\left[\frac{\max_{\boldsymbol{x}\in S_k} c_g O_g(\boldsymbol{x}) - \min_{\boldsymbol{x}\in S_k} c_g O_g(\boldsymbol{x})}{\max_{\boldsymbol{x}\in S} c_g O_g(\boldsymbol{x}) - \min_{\boldsymbol{x}\in S} c_g O_g(\boldsymbol{x})}\right]^2} \\ &= \sqrt{\frac{1}{G}\cdot\sum_{g=1}^{G}\left[\frac{c_g\cdot\max_{\boldsymbol{x}\in S_k} O_g(\boldsymbol{x}) - c_g\cdot\min_{\boldsymbol{x}\in S_k} O_g(\boldsymbol{x})}{c_g\cdot\max_{\boldsymbol{x}\in S} O_g(\boldsymbol{x}) - c_g\cdot\min_{\boldsymbol{x}\in S} O_g(\boldsymbol{x})}\right]^2} \\ &= \sqrt{\frac{1}{G}\cdot\sum_{g=1}^{G}\left[\frac{\max_{\boldsymbol{x}\in S_k} O_g(\boldsymbol{x}) - \min_{\boldsymbol{x}\in S_k} O_g(\boldsymbol{x})}{O_g^{\max} - O_g^{\min}}\right]^2}.\end{aligned}$$

Hence, metric NMS remains unchanged after linear scaling and satisfies Axiom 1.

After adding constants, the gth objective value for $\boldsymbol{x}$ varies from $O_g(\boldsymbol{x})$ to $O_g(\boldsymbol{x})+b_g$. Let $\text{NMS}''(S_k)$ denote the value of metric NMS after adding constants. $\text{NMS}''(S_k)$ can be calculated as follows:

$$\begin{aligned}\text{NMS}''(S_k) &= \sqrt{\frac{1}{G}\cdot\sum_{g=1}^{G}\left\{\frac{\max_{\boldsymbol{x}\in S_k}\left[O_g(\boldsymbol{x}) + b_g\right] - \min_{\boldsymbol{x}\in S_k}\left[O_g(\boldsymbol{x}) + b_g\right]}{\max_{\boldsymbol{x}\in S}\left[O_g(\boldsymbol{x}) + b_g\right] - \min_{\boldsymbol{x}\in S}\left[O_g(\boldsymbol{x}) + b_g\right]}\right\}^2} \\ &= \sqrt{\frac{1}{G}\cdot\sum_{g=1}^{G}\left[\frac{b_g + \max_{\boldsymbol{x}\in S_k} O_g(\boldsymbol{x}) - b_g - \min_{\boldsymbol{x}\in S_k} O_g(\boldsymbol{x})}{b_g + \max_{\boldsymbol{x}\in S} O_g(\boldsymbol{x}) - b_g - \min_{\boldsymbol{x}\in S} O_g(\boldsymbol{x})}\right]^2} \\ &= \sqrt{\frac{1}{G}\cdot\sum_{g=1}^{G}\left[\frac{\max_{\boldsymbol{x}\in S_k} O_g(\boldsymbol{x}) - \min_{\boldsymbol{x}\in S_k} O_g(\boldsymbol{x})}{O_g^{\max} - O_g^{\min}}\right]^2}.\end{aligned}$$

Therefore, metric NMS remains unchanged after adding constants and satisfies Axiom 2.

Theorem 4. *Metrics NID and MID violate both Axiom 1 and Axiom 2.*

Proof. We consider a bi-objective optimization problem: $\min x_1, \min x_2$ subject to $x_1 + x_2 \geq 10, x_1 \geq 0, x_2 \geq 0$. Let $S_1 = \{(0, 10)\}$, $S_2 = \{(5, 5)\}$. Then, $\text{NID}(S_1) = \text{MID}(S_1) = 10$, $\text{NID}(S_2) = \text{MID}(S_2) = 5\sqrt{2}$. Hence, $\text{NID}(S_1) > \text{NID}(S_2)$, $\text{MID}(S_1) > \text{MID}(S_2)$. Let $\text{NID}'(S_k)$ and $\text{MID}'(S_k)$ denote the values of metrics NID and MID after shifting objectives by linear scaling, respectively. If $c_1 = 10$ and $c_2 = 1$, $\text{NID}'(S_1) = \text{MID}'(S_1) = 10$ and $\text{NID}'(S_2) = \text{MID}'(S_2) = 5\sqrt{101}$. Thus, $\text{NID}'(S_1) < \text{NID}'(S_2)$, $\text{MID}'(S_1) < \text{MID}'(S_2)$.

Let $\text{NID}''(S_k)$ and $\text{MID}''(S_k)$ denote the values of metrics NID and MID after shifting objectives by adding constants, respectively. If $b_1 = 5$ and $b_2 = -5$, $\text{NID}''(S_1) = \text{MID}''(S_1) = 5\sqrt{2}$, $\text{NID}''(S_2) = \text{MID}''(S_2) = 10$. Thus, $\text{NID}''(S_1) < \text{NID}''(S_2)$, $\text{MID}''(S_1) < \text{MID}''(S_2)$. Therefore, Metrics NID and MID violate both Axiom 1 and Axiom 2.

Theorem 5. *Metrics SP and IGD violate Axiom 1 but satisfy Axiom 2.*

Proof. First, we consider a bi-objective optimization problem: $\min x_1$, $\min x_2$ subject to $(x_1, x_2) \in X$, where X is a nonempty compact set. Let $S_1 = \{(0, 10), (2, 6), (10, 0)\}$, $S_2 = \{(0, 10), (6, 1), (10, 0)\}$. It can be calculated that $\mathrm{SP}(S_1) = \frac{8\sqrt{3}}{3}$, $\mathrm{SP}(S_2) = \frac{10\sqrt{3}}{3}$, $\mathrm{IGD}(S_1) = \frac{2\sqrt{35}}{3}$, and $\mathrm{IGD}(S_2) = \frac{\sqrt{151}}{3}$. Let $\mathrm{SP}'(S_k)$ denote the value of metric SP after shifting objectives by linear scaling and let $\mathrm{IGD}'(S_k)$ denote the value of metric IGD after shifting objectives by linear scaling. If $c_1 = 10$ and $c_2 = 1$, it can be calculated that $\mathrm{SP}'(S_1) = \frac{62\sqrt{3}}{3}$, $\mathrm{SP}'(S_2) = \frac{28\sqrt{3}}{3}$, $\mathrm{IGD}'(S_1) = \frac{2\sqrt{1817}}{3}$, and $\mathrm{IGD}'(S_2) = \frac{\sqrt{6883}}{3}$. Since $\mathrm{SP}(S_1) < \mathrm{SP}(S_2)$ but $\mathrm{SP}'(S_1) > \mathrm{SP}'(S_2)$, metric SP violates Axiom 1. Similarly, Since $\mathrm{IGD}(S_1) < \mathrm{IGD}(S_2)$ but $\mathrm{IGD}'(S_1) > \mathrm{IGD}'(S_2)$, metric IGD violates Axiom 1.

Next, metrics SP and IGD are proved to satisfy Axiom 2. Let $\boldsymbol{x}$ and $\boldsymbol{x}'$ denote two different solutions in S_k. After shifting objectives by adding constants, the Manhattan objective distance between solution $\boldsymbol{x}$ and its nearest solution (i.e., $d_{k,\boldsymbol{x}}$) is $\min_{\boldsymbol{x}' \in S_k, \boldsymbol{x}' \neq \boldsymbol{x}} \left\{\sum_{g=1}^{G} |O_g(\boldsymbol{x}) + b_g - O_g(\boldsymbol{x}') - b_g|\right\} = \min_{\boldsymbol{x}' \in S_k, \boldsymbol{x}' \neq \boldsymbol{x}} \left\{\sum_{g=1}^{G} |O_g(\boldsymbol{x}) - O_g(\boldsymbol{x}')|\right\}$. Since $d_{k,\boldsymbol{x}}$ is unchanged, $\bar{d}_k$ and $\mathrm{SP}(S_k)$ also remain unchanged after adding constants. Additionally, after adding constants, the minimum Euclidean objective distance between solution $\boldsymbol{x}$ and any other solution $\boldsymbol{x}'$ (i.e., $L_{k,\boldsymbol{x}}$) is $\min_{\boldsymbol{x}' \in S_k, \boldsymbol{x}' \neq \boldsymbol{x}} \left\{ \sqrt{\sum_{g=1}^{G} \left(O_g(\boldsymbol{x}) + b_g - O_g(\boldsymbol{x}') - b_g\right)^2} \right\} = \min_{\boldsymbol{x}' \in S_k, \boldsymbol{x}' \neq \boldsymbol{x}} \left\{ \sqrt{\sum_{g=1}^{G} \left(O_g(\boldsymbol{x}) - O_g(\boldsymbol{x}')\right)^2} \right\}$. Since both $L_{k,\boldsymbol{x}}$ and $|S_k|$ are unchanged after shifting objectives by adding constants, $\mathrm{IGD}(S_k)$ also remains unchanged. Therefore, both metrics SP and IGD satisfy Axiom 2.

Theorem 6. *Metric HR satisfies both Axiom 1 and Axiom 2.*

Proof. Let $\boldsymbol{x}$ denote a solution in S_k or S. Let h'_k and h' represent the hypervolume of S_k and S after linear scaling, respectively. The hypervolume of $r_{\boldsymbol{x}}$ can be calculated as $\prod_{g=1}^{G}(O_g^{\max} - O_g(\boldsymbol{x}))$. After shifting objectives by linear scaling, the hypervolume of $r_{\boldsymbol{x}}$ becomes $\prod_{g=1}^{G} c_g(O_g^{\max} - O_g(\boldsymbol{x}))$. Hence, $h'_k = h_k \prod_{g=1}^{G} c_g$, $h' = h \prod_{g=1}^{G} c_g$. Therefore, metric HR remains unchanged after linear scaling. In addition, after shifting objectives by adding constants, the hypervolume of $r_{\boldsymbol{x}}$ can be calculated as $\prod_{g=1}^{G}(O_g^{\max} + b_g - O_g(\boldsymbol{x}) - b_g) = \prod_{g=1}^{G}(O_g^{\max} - O_g(\boldsymbol{x}))$. Since the hypervolume of $r_{\boldsymbol{x}}$ is unchanged, the conclusion that h_k and h remain unchanged after adding constants can be drawn. Thus, metric HR remains unchanged after adding constants. Therefore, metric HR satisfies both Axiom 1 and Axiom 2.

Table 1 summarizes whether each performance metric satisfies the two axioms. Reliable performance metrics are supposed to comply with both axioms. That is, a performance metric that violates any of the axioms is not a reliable performance metric. It can be seen from Table 1 that five performance metrics (i.e., NOSO, ONSN, NMS, ER, and HR) comply with both axioms, two performance metrics (i.e., SP and IGD) violate one of them, and two performance metrics (i.e., NID and MID) violate both of them. Thus, from the perspective of satisfying the axioms, it can be concluded that the performance metrics NOSO, ONSN, NMS, ER, and HR are reliable.

Table 1. Summary of the compliance of performance metrics with two axioms.

Performance metric	Abbreviation	Mathematical expression	Axiom 1	Axiom 2
Number of solutions obtained	NOSO	$\mathrm{NOSO}(S_k) = \lvert S_k \rvert$	✓	✓
Overall nondominated solutions number	ONSN	$\mathrm{ONSN}(S_k) = \lvert S_k \rvert - \sum_{\boldsymbol{x} \in S_k} e_{\boldsymbol{x}}$	✓	✓
Normalized maximum spread	NMS	$\mathrm{NMS}(S_k) = \sqrt{\frac{1}{G} \cdot \sum_{g=1}^{G} \left[\frac{\max_{\boldsymbol{x} \in S_k} O_g(\boldsymbol{x}) - \min_{\boldsymbol{x} \in S_k} O_g(\boldsymbol{x})}{O_g^{\max} - O_g^{\min}} \right]^2}$	✓	✓
Error ratio	ER	$\mathrm{ER}(S_k) = \frac{\sum_{\boldsymbol{x} \in S_k} e_{\boldsymbol{x}}}{\lvert S_k \rvert}$	✓	✓
Nearest ideal distance	NID	$\mathrm{NID}(S_k) = \min_{\boldsymbol{x} \in S_k} C_{\boldsymbol{x}}$	×	×
Mean ideal distance	MID	$\mathrm{MID}(S_k) = \frac{\sum_{\boldsymbol{x} \in S_k} C_{\boldsymbol{x}}}{\lvert S_k \rvert}$	×	×
Spacing	SP	$\mathrm{SP}(S_k) = \sqrt{\frac{1}{\lvert S_k \rvert - 1} \cdot \sum_{\boldsymbol{x} \in S_k} (d_{k,\boldsymbol{x}} - \bar{d}_k)^2}$	×	✓
Inverted generational distance	IGD	$\mathrm{IGD}(S_k) = \frac{\sqrt{\sum_{\boldsymbol{x} \in S_k} L_{k,\boldsymbol{x}}^2}}{\lvert S_k \rvert}$	×	✓
Hypervolume-based ratio	HR	$\mathrm{HR}(S_k) = \frac{h_k}{h}$	✓	✓

5. Computational experiments

A real-world case of multi-objective optimization in the field of cargo transportation is investigated. The basic ϵ-constraint method and the weighting method are adopted to address the real-world case. In this section, we first introduce the cargo transportation case in Section 5.1, and then analyze the experimental results in Section 5.2. We carry out computational experiments on a PC (14 cores of CPUs, 2.5 GHz, Memory 64 GB). Both methods are implemented in the commercial solver Gurobi 10.0.0 (Anaconda, Python).

5.1. Experimental setting

This case is dedicated to designing reasonable routes for a vehicle to reduce its travel time and driving risks. We define a road network as $G = (V, E)$, where V is the node set indexed by i or j, and $E \subseteq \{(i, j) | i, j \in V\}$ is the arc set indexed by (i, j). The vehicle departs from the starting point, denoted by 0, and finally arrives at the destination point, denoted by n. Subject to traffic restrictions, a set E' ($E' \subsetneq E$) of arcs indexed by (i, j) is impassable. Let t_{ij}, r_{ij}, and c_{ij} represent the travel time on arc (i, j), the driving risk on arc (i, j), and the road toll for traveling on arc (i, j), respectively. The available budget for traveling on all arcs is denoted by b. x_{ij} is defined as a binary variable, which equals 1 if and only if the vehicle traverses arc (i, j), and 0 otherwise. The mathematical model is formulated below:

$$[\mathrm{M1}] \quad \min \sum_{(i,j) \in E} t_{ij} x_{ij}, \tag{11}$$

$$\min \sum_{(i,j) \in E} r_{ij} x_{ij}, \tag{12}$$

$$\text{s.t.} \quad \sum_{(i,j) \in E'} x_{ij} = 0, \tag{13}$$

$$\sum_{(i,j) \in E} c_{ij} x_{ij} \leq b, \tag{14}$$

$$\sum_{(0,j) \in E} x_{0j} = 1, \tag{15}$$

$$\sum_{(i,n)\in E} x_{in} = 1, \tag{16}$$

$$\sum_{(i,j)\in E} x_{ij} = \sum_{(j,i)\in E} x_{ji} \quad \forall j \in V \setminus \{0, n\}, \tag{17}$$

$$x_{ij} \in \{0, 1\} \quad \forall (i, j) \in E. \tag{18}$$

Objectives (11) and (12) minimize the total travel time and driving risks, respectively. Constraint (13) guarantees that the vehicle cannot traverse arcs with traffic restrictions. Constraint (14) ensures that the total road tolls cannot exceed the budget. Constraints (15) and (16) ensure that the vehicle starts from the starting point 0 and ends at the destination point n. Constraints (17) guarantee the flow conservation. Constraint (18) defines the decision variable and restricts its range.

A road network with 60 nodes is randomly generated. In the road network, 2% of the arcs are subject to traffic restrictions. The travel time, the driving risk, and the road toll for each arc are set to random integers uniformly distributed within the range of 1 to 20 minutes, within the range of 1 to 50, and within the range of 20 to 40 USD, respectively. The available budget for traveling on all arcs is set to 200 USD.

5.2. *Experimental results*

The basic ϵ-constraint method (with ϵ set to 0.1) and the weighting method (with the weight of the first objective ranging from 0, 0.01, ..., to 1) are used to solve the real-world vehicle routing problem. To investigate the impact of the objective values after scale and translation transformations on the relative ranking of the metric values, the following scale and translation transformations for the first objective values of solutions (i.e., minimal travel time) obtained by solving the model [M1] are performed. In the model [M1], the first objective value is measured in minutes, which could be converted to seconds. Then, the first objective value of each non-dominated solution obtained through both the basic ϵ-constraint method and the weighting method is converted from minutes to seconds by multiplying by 60. Moreover, considering the cargo loading time at the starting point 0 and the cargo unloading time at the destination point n, a constant value of 10 minutes is added to the first objective value of each non-dominated solution obtained through both methods.

The relative ranking of the metric values is explored to compare the performance of the basic ϵ-constraint method and the weighting method. The experimental results in Table 2 show that for model [M1], the values of metrics NOSO, ONSN, MID, and HR of the basic ϵ-constraint method are greater than those of the weighting method, while the values of metrics SP and IGD of the basic ϵ-constraint method are less than those of the weighting method. The values of metrics NMS, ER, and NID of both methods are equal. Additionally, in this case, the relative ranking of metric values remains unchanged after the translation transformation for the first objective values of nondominated solutions. Furthermore, after the scale transformation for the first objective values of nondominated solutions, the relative ranking of the values of metrics MID and SP changes, while the relative ranking of other metric values is unchanged. Hence, metrics MID and SP are verified as unreliable. It is worth noting that in this case, the metrics whose relative ranking remain unchanged after scale and translation transformations are not necessarily reliable, as demonstrated by the theorems in Section 4.

Table 2. Comparison of the metric values.

Performance metric	[M1]		[M1]-scale		[M1]-translation	
	Basic ϵ-constraint method	Weighting method	Basic ϵ-constraint method	Weighting method	Basic ϵ-constraint method	Weighting method
NOSO	11	8	11	8	11	8
ONSN	11	8	11	8	11	8
NMS	1.00	1.00	1.00	1.00	1.00	1.00
ER	0.00	0.00	0.00	0.00	0.00	0.00
NID	40.31	40.31	861.66	861.66	47.80	47.80
MID	87.56	83.39	1532.04	1699.90	92.33	89.08
SP	15.33	16.82	179.05	176.42	15.33	16.82
IGD	6.16	8.76	77.86	110.39	6.16	8.76
HR	1.00	0.99	1.00	0.99	1.00	0.99

6. Conclusions

Real-world transportation problems often involve multiple conflicting objectives [38], such as reducing operational costs, reducing GHG emissions, and enhancing customer satisfaction with transportation services. In situations where various multi-objective optimization methods can solve the same problem, decision-makers often face difficulties in determining which one is better [39]. Hence, performance metrics are adopted to quantitatively assess the performance of multi-objective optimization methods [40]. Here, we focus on the evaluation of performance metrics, exploring whether performance metrics can truly reflect the quality of the solutions obtained, as well as the criteria for reliable performance metrics.

The major contributions of this study are twofold. First, we elaborate on a series of popular performance metrics for multi-objective optimization methods (including NOSO, ONSN, NMS, ER, NID, MID, SP, IGD, and HR) and critically explore whether larger or smaller values of the performance metrics definitely indicate better performance of the multi-objective optimization methods. None of the metrics elaborated in this study can truly reflect the quality of the obtained solutions all the time. In other words, larger or smaller values of the performance metrics do not necessarily indicate better performance of the multi-objective optimization methods. Hence, it is better to adopt multiple metrics, rather than a single metric, to comprehensively assess the performance of multi-objective optimization methods in practical transportation decision-making. Second, we propose two axioms to provide the criteria for reliable metrics, enabling transportation decision-makers to enhance the selection of metrics. Whether each performance metric satisfies the two axioms is mathematically proved. The performance metrics NOSO, ONSN, NMS, ER, and HR that satisfy both axioms are proved as reliable. Moreover, a real-world case is investigated, demonstrating the unreliability of metrics MID and SP. It is worth noting that although the performance metrics and axioms are elaborated based on transportation optimization problems with multiple objectives, these performance metrics and axioms are theoretically applicable to multi-objective optimization problems in other fields, such as the economics and mechanics fields.

Future studies could be devoted to designing more effective performance metrics, assessing the performance of multi-objective optimization methods more appropriately. In addition, the two axioms proposed in this study could be implemented in more studies related to multi-objective optimization, enabling to identify reliable performance metrics, thus guaranteeing the accuracy and validity of the comparison results. Moreover, nonlinear transformation-invariant features of performance metrics may be useful for some specific problems, which could be further explored.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

Shuaian Wang is an editorial board member for [Electronic Research Archive (ERA)] and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

References

1. E. Demir, T. Bektaş, G. Laporte, A review of recent research on green road freight transportation, *Eur. J. Oper. Res.*, **237** (2014), 775–793. https://doi.org/10.1016/j.ejor.2013.12.033
2. Y. J. Hu, L. Yang, H. Cui, H. Wang, C. Li, Developing a balanced strategy: A multi-objective model for emissions reduction and development of civil aviation in China, *Energy*, **307** (2024), 132597. https://doi.org/10.1016/j.energy.2024.132597
3. F. Mastroddi, S. Gemma, Analysis of Pareto frontiers for multidisciplinary design optimization of aircraft, *Aerosp. Sci. Technol.*, **28** (2013), 40–55. https://doi.org/10.1016/j.ast.2012.10.003
4. A. Ala, M. Yazdani, M. Ahmadi, A. Poorianasab, M. Y. N. Attari, An efficient healthcare chain design for resolving the patient scheduling problem: Queuing theory and MILP-ASA optimization approach, *Ann. Oper. Res.*, **328** (2023), 3–33. https://doi.org/10.1007/s10479-023-05287-5
5. Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, L. Hanzo, A survey of multi-objective optimization in wireless sensor networks: Metrics, algorithms, and open problems, *IEEE Commun. Surv. Tutorials*, **19** (2016), 550–586. https://doi.org/10.1109/COMST.2016.2610578
6. W. Deng, X. Zhang, Y. Zhou, Y. Liu, X. Zhou, H. Chen, et al., An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems, *Inf. Sci.*, **585** (2022), 441–453. https://doi.org/10.1016/j.ins.2021.11.052
7. H. Zhang, Q. Ruan, Y. Jin, S. Wang, Bi-objective optimization for transportation: Generating near-optimal subsets of Pareto optimal solutions, *Appl. Sci.*, **15** (2025), 2519. https://doi.org/10.3390/app15052519
8. S. Petchrompo, D. W. Coit, A. Brintrup, A. Wannakrairot, A. K. Parlikad, A review of Pareto pruning methods for multi-objective optimization, *Comput. Ind. Eng.*, **167** (2022), 108022. https://doi.org/10.1016/j.cie.2022.108022

9. J. Zhao, Y. Li, J. Bai, L. Ma, C. Shi, G. Zhang, et al., Multi-objective optimization of marine nuclear power secondary circuit system based on improved multi-objective particle swarm optimization algorithm, *Prog. Nucl. Energy*, **161** (2023), 104740. https://doi.org/10.1016/j.pnucene.2023.104740
10. E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, V. G. da Fonseca, Performance assessment of multiobjective optimizers: An analysis and review, *IEEE Trans. Evol. Comput.*, **7** (2003), 117–132. https://doi.org/10.1109/TEVC.2003.810758
11. E. Demir, T. Bektaş, G. Laporte, The bi-objective pollution-routing problem, *Eur. J. Oper. Res.*, **232** (2014), 464–478. https://doi.org/10.1016/j.ejor.2013.08.002
12. R. H. Stewart, T. S. Palmer, B. DuPont, A survey of multi-objective optimization methods and their applications for nuclear scientists and engineers, *Prog. Nucl. Energy*, **138** (2021), 103830. https://doi.org/10.1016/j.pnucene.2021.103830
13. Z. Wang, A. Ala, Z. Liu, W. Cui, H. Ding, G. Jin, et al., A hybrid equilibrium optimizer based on moth flame optimization algorithm to solve global optimization problems, *J. Artif. Intell. Soft Comput. Res.*, **14** (2024), 207–235. https://doi.org/10.2478/jaiscr-2024-0012
14. A. Ala, A. Goli, Incorporating machine learning and optimization techniques for assigning patients to operating rooms by considering fairness policies, *Eng. Appl. Artif. Intell.*, **136** (2024), 108980. https://doi.org/10.1016/j.engappai.2024.108980
15. W. Chen, Q. Zhuo, L. Zhang, Modeling and heuristically solving group train operation scheduling for heavy-haul railway transportation, *Mathematics*, **11** (2023), 2489. https://doi.org/10.3390/math11112489
16. Z. Elmi, B. Li, B. Liang, Y. Y. Lau, M. Borowska-Stefańska, S. Wiśniewski, et al., An epsilon-constraint-based exact multi-objective optimization approach for the ship schedule recovery problem in liner shipping, *Comput. Ind. Eng.*, **183** (2023), 109472. https://doi.org/10.1016/j.cie.2023.109472
17. C. A. Nallolla, V. P, D. Chittathuru, S. Padmanaban, Multi-objective optimization algorithms for a hybrid AC/DC microgrid using RES: A comprehensive review, *Electronics*, **12** (2023), 1062. https://doi.org/10.3390/electronics12041062
18. C. Zhao, J. Tang, W. Gao, Y. Zeng, Z. Li, Many-objective optimization of multi-mode public transportation under carbon emission reduction, *Energy*, **286** (2024), 129627. https://doi.org/10.1016/j.energy.2023.129627
19. R. Jia, K. Gao, S. Cui, J. Chen, J. Andric, A multi-objective reinforcement learning-based velocity optimization approach for electric trucks considering battery degradation mitigation, *Transp. Res. Part E: Logist. Transp. Rev.*, **194** (2025), 103885. https://doi.org/10.1016/j.tre.2024.103885
20. Z. Huang, Z. Sheng, C. Ma, S. Chen, Human as AI mentor: Enhanced human-in-the-loop reinforcement learning for safe and efficient autonomous driving, *Commun. Transp. Res.*, **4** (2024), 100127. https://doi.org/10.1016/j.commtr.2024.100127
21. N. Riquelme, C. Von Lücken, B. Baran, Performance metrics in multi-objective optimization, in *2015 Latin American Computing Conference (CLEI)*, IEEE, (2015), 1–11. https://doi.org/10.1109/CLEI.2015.7360024

22. Y. Tian, R. Cheng, X. Zhang, M. Li, Y. Jin, Diversity assessment of multi-objective evolutionary algorithms: Performance metric and benchmark problems, *IEEE Comput. Intell. Mag.*, **14** (2019), 61–74. https://doi.org/10.1109/MCI.2019.2919398

23. Q. Fan, W. Wang, X. Yan, Multi-objective differential evolution with performance-metric-based self-adaptive mutation operator for chemical and biochemical dynamic optimization problems, *Appl. Soft Comput.*, **59** (2017), 33–44. https://doi.org/10.1016/j.asoc.2017.05.044

24. K. Lapa, Meta-optimization of multi-objective population-based algorithms using multi-objective performance metrics, *Inf. Sci.*, **489** (2019), 193–204. https://doi.org/10.1016/j.ins.2019.03.054

25. J. G. Falcon-Cardona, C. A. C. Coello, Indicator-based multi-objective evolutionary algorithms: A comprehensive survey, *ACM Comput. Surv.*, **53** (2020), 1–35. https://doi.org/10.1145/3376916

26. T. Wagner, H. Trautmann, Online convergence detection for evolutionary multi-objective algorithms revisited, in *IEEE Congress on Evolutionary Computation*, IEEE, (2010), 1–8. https://doi.org/10.1109/CEC.2010.5586474

27. L. Marti, J. Garcia, A. Berlanga, J. M. Molina, A stopping criterion for multi-objective optimization evolutionary algorithms, *Inf. Sci.*, **367** (2016), 700–718. https://doi.org/10.1016/j.ins.2016.07.025

28. S. Mirjalili, A. Lewis, Novel performance metrics for robust multi-objective optimization algorithms, *Swarm Evol. Comput.*, **21** (2015), 1–23. https://doi.org/10.1016/j.swevo.2014.10.005

29. C. Audet, J. Bigeon, D. Cartier, S. Le Digabel, L. Salomon, Performance indicators in multiobjective optimization, *Eur. J. Oper. Res.*, **292** (2021), 397–422. https://doi.org/10.1016/j.ejor.2020.11.016

30. T. Okabe, Y. Jin, B. Sendhoff, A critical survey of performance indices for multi-objective optimisation, in *2003 Congress on Evolutionary Computation (CEC)*, IEEE, (2003), 878–885. https://doi.org/10.1109/CEC.2003.1299759

31. S. Zajac, S. Huber, Objectives and methods in multi-objective routing problems: A survey and classification scheme, *Eur. J. Oper. Res.*, **290** (2021), 1–25. https://doi.org/10.1016/j.ejor.2020.07.005

32. B. Qian, L. Wang, D. X. Huang, W. L. Wang, X. Wang, An effective hybrid DE-based algorithm for multi-objective flow shop scheduling with limited buffers, *Comput. Oper. Res.*, **36** (2009), 209–233. https://doi.org/10.1016/j.cor.2007.08.007

33. H. Amirian, R. Sahraeian, Augmented ϵ-constraint method in multi-objective flowshop problem with past sequence set-up times and a modified learning effect, *Int. J. Prod. Res.*, **53** (2015), 5962–5976. https://doi.org/10.1080/00207543.2015.1033033

34. M. Rashidnejad, S. Ebrahimnejad, J. Safari, A bi-objective model of preventive maintenance planning in distributed systems considering vehicle routing problem, *Comput. Ind. Eng.*, **120** (2018), 360–381. https://doi.org/10.1016/j.cie.2018.05.001

35. G. G. Yen, Z. He, Performance metric ensemble for multiobjective evolutionary algorithms, *IEEE Trans. Evol. Comput.*, **18** (2014), 131–144. https://doi.org/10.1109/TEVC.2013.2240687

36. J. Wang, X. Hu, E. Demeulemeester, Y. Zhao, A bi-objective robust resource allocation model for the RCPSP considering resource transfer costs, *Int. J. Prod. Res.*, **59** (2019), 367–387. https://doi.org/10.1080/00207543.2019.1695168
37. S. Mirjalili, S. Saremi, S. M. Mirjalili, L. D. S. Coelho, Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization, *Expert Syst. Appl.*, **47** (2016), 106–119. https://doi.org/10.1016/j.eswa.2015.10.039
38. M. Tadaros, A. Migdalas, Bi- and multi-objective location routing problems: Classification and literature review, *Oper. Res.*, **22** (2022), 4641–4683. https://doi.org/10.1007/s12351-022-00734-w
39. J. S. Neufeld, S. Schulz, U. Buscher, A systematic review of multi-objective hybrid flow shop scheduling, *Eur. J. Oper. Res.*, **309** (2023), 1–23. https://doi.org/10.1016/j.ejor.2022.08.009
40. Q. Liu, X. Li, H. Liu, Z. Guo, Multi-objective metaheuristics for discrete optimization problems: A review of the state-of-the-art, *Appl. Soft Comput.*, **93** (2020), 106382. https://doi.org/10.1016/j.asoc.2020.106382