

Article

Handling Multimodality in Pareto Set Estimation via Cluster-Wise Decomposition

Yuki Suzumura ¹, Yoshihiro Ohta ²  and Hiroyuki Sato ^{1,*} ¹ The University of Electro-Communications, Tokyo 182-8585, Japan; y.suzumura@uec.ac.jp² Mitsubishi Electric Corporation, Tokyo 100-8310, Japan; ota.yoshihiro@dw.mitsubishielectric.co.jp

* Correspondence: h.sato@uec.ac.jp

Abstract

Multimodal multi-objective optimization problems often exhibit one-to-many mappings, where multiple distinct variable vectors correspond to the same objective vector. This characteristic makes Pareto set (PS) estimation difficult, as conventional inverse modeling approaches assume a one-to-one correspondence. This study proposes a cluster-wise PS estimation framework in the variable space. Known solutions are partitioned into locally monotonic clusters using oscillation detection with an amplitude threshold, and independent response surface models are constructed for each cluster. By estimating PS solutions from multiple cluster-specific models for a given direction vector, the method preserves multimodal structures that conventional approaches fail to capture. Numerical experiments on the multimodal benchmark problems MMF1–8 and LIRCMOP1–2 demonstrate that the proposed method achieves equal or better HV and IGD values than the conventional method, while improving decision-space approximation as measured by IGDX in most test cases and suppressing the generation of dominated solutions.

Keywords: multi-objective optimization; multimodal problem; Pareto set estimation; clustering; response surface modeling

1. Introduction

Multi-objective optimization problems (MOPs) involve the simultaneous optimization of multiple conflicting objectives [1,2]. Instead of a single optimal solution, a set of trade-off solutions emerges in the variable space, referred to as the Pareto set (PS), whose image in the objective space forms the Pareto front (PF). In many practical applications, objective evaluations are computationally expensive, and thus, the PS must be approximated by a limited number of solutions.

To compensate for limited solutions, Pareto set estimation (PSE) has been studied as a post-optimization technique [3,4]. Rather than performing an additional search, PSE reconstructs the structure of the PS from already evaluated solutions. Existing approaches build response surface models that map direction vectors in the objective space to corresponding variable vectors [3]. This strategy implicitly assumes a one-to-one correspondence between objective and variable vectors.

A fundamental difficulty arises when this assumption does not hold. In multimodal multi-objective optimization problems (MMOPs), multiple distinct variable vectors may correspond to the same or nearly identical objective vector [5,6]. Such one-to-many mappings invalidate conventional inverse modeling and cause multiple decision-space modes to be merged into a single estimate. This difficulty becomes more severe when only a



Academic Editor: Zong Woo Geem

Received: 27 February 2026

Revised: 6 April 2026

Accepted: 7 April 2026

Published: 8 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

limited number of evaluated solutions are available, because sparse samples make the local structure of each mode harder to preserve, increasing the risk that distinct branches are approximated as a single continuous mapping. Although multimodality has been extensively studied in evolutionary optimization [7,8], its impact on Pareto set estimation has received limited attention.

This study proposes a cluster-wise PSE framework to address this limitation on MMOPs. The known solutions are partitioned in the variable space into locally monotonic clusters by detecting oscillatory transitions that indicate potential mode changes. Independent response surface models are constructed for each cluster. For a given direction vector, multiple cluster-specific models are queried, allowing the reconstruction of distinct decision-space modes associated with the same objective vector. This structured decomposition enables PS estimation that preserves multimodal characteristics while remaining compatible with model-based PSE.

The proposed multimodal PSE framework is evaluated on representative multimodal benchmark problems, including the unconstrained MMF1–8 [5] and the constrained LIRCMOP1–2 [6]. The numerical experiments quantitatively assess PF approximation quality using HV [9] and IGD [10–12] and discuss the distribution of the obtained solutions in the objective space. Also, PS approximation quality is evaluated using IGDX [13], together with an analysis of the solution distribution in the variable space.

This paper is an extended version of the conference paper [14] for this invited Special Issue. The main differences from [14] are summarized as follows:

- In the clustering method based on oscillation detection, the previous paper [14] determined oscillation solely by counting local maxima and minima. In this paper, we introduce an amplitude threshold in the variable space. This extension suppresses over-segmentation caused by minor fluctuations and improves clustering stability.
- In [14], the clustering procedure could generate clusters containing only a single solution, which prevents the construction of a response surface model. In this paper, we introduce a reassignment mechanism that avoids single-solution clusters by merging them with the nearest valid cluster, thereby ensuring model constructability and improving robustness.
- In [14], the obtained solution sets were evaluated only by the objective-space metric HV [13]. In this paper, we additionally employ IGD [10–12] in the objective space and IGDX [13] in the variable space, enabling a more comprehensive quantitative evaluation.
- In [14], the clustering parameters were fixed. In this paper, we conduct a detailed sensitivity analysis of these parameters.
- In [14], the distribution of the obtained solutions was discussed only for selected problems. In this paper, we provide a comprehensive analysis for all test problems.

2. Multi-objective Optimization

Let $\mathcal{X} = \prod_{j=1}^D [x_j^{\min}, x_j^{\max}] \subset \mathbb{R}^D$ denote the variable space, which is considered continuous in this study. An optimization problem is defined as follows:

$$\begin{cases} \text{Minimize} & f_i(\mathbf{x}), & (i = 1, 2, \dots, M), \\ \text{subject to} & g_l(\mathbf{x}) \leq 0, & (l = 1, 2, \dots, K), \end{cases} \quad (1)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_D) \in \mathcal{X}$ is the variable vector, and f_i and g_l denote the objective and constraint functions, respectively. When $M \geq 2$, the problem is referred to as a multi-objective optimization problem (MOP), which is the focus of this study. If $K = 0$, the problem is unconstrained; if $K \geq 1$, it is a constrained problem. The constraint violation of a solution $\mathbf{x}$ is defined as $\phi(\mathbf{x}) = \sum_{l=1}^K \max\{0, g_l(\mathbf{x})\}$. Solution $\mathbf{x}$ is feasible if $\phi(\mathbf{x}) = 0$,

and infeasible otherwise. The variable space $\mathcal{X}$ is therefore divided into the feasible set $F = \{x \in \mathcal{X} \mid \phi(x) = 0\}$ and the infeasible set $\bar{F} = \{x \in \mathcal{X} \mid \phi(x) > 0\}$. In unconstrained problems, all solutions are feasible, i.e., $\mathcal{X} = F$.

For two feasible solutions $x, y \in F$, x Pareto dominates y , denoted by $x \prec y$, if

$$\forall i \in \{1, \dots, M\} : f_i(x) \leq f_i(y) \wedge \exists j \in \{1, \dots, M\} : f_j(x) < f_j(y). \quad (2)$$

The Pareto-optimal set is defined as

$$PS = \{x \in F \mid \nexists y \in F : y \prec x\}, \quad (3)$$

and the Pareto front is defined as

$$PF = \{f(x) \mid x \in PS\}. \quad (4)$$

In evolutionary multi-objective optimization [1,2], the primary goal is to approximate the PF by a finite set of non-dominated solutions in the objective space. However, since the PF of an M -objective problem generally forms an $(M - 1)$ -dimensional manifold, accurately approximating it with a finite number of solutions is inherently difficult. This difficulty becomes more pronounced when the evaluation cost of solutions is high. Moreover, in multimodal multi-objective optimization problems (MMOPs), multiple distinct solutions in the variable space may correspond to the same or nearly identical point in the objective space. In such cases, approximating only the PF in the objective space is insufficient. Instead, approximating the PS in the variable space becomes equally important, because multiple distinct solutions may correspond to the same objective-space point.

3. Related Work

3.1. Pareto Front and Pareto Set Estimation

In MOPs, techniques for approximating the PF and PS structures from a limited set of solutions are essential for enhancing decision-making support. These approaches can be broadly classified into PF estimation in the objective space and PS estimation in the variable space.

3.1.1. Pareto Front Estimation in Objective Space

PF estimation (PFE) aims to approximate the geometry of the PF in the objective space. The PAINT method [4,15] is an approach that performs linear interpolation between known non-dominated solutions using a set of simplices. A subsequent extension has enabled high-precision interpolation even for discontinuous PFs [16,17], while Bézier simplex fitting has been introduced to describe PF manifolds with a limited number of solutions [18]. Various attempts have also been made to incorporate geometric shape models. One mechanism for controlling the granularity of PF approximations is the adaptive regulation of the ϵ -dominance concept [19]. Other studies have suggested approximating the PF using a family of curves [20] or utilizing a generalized simplex model to unify the estimation of PFs with diverse curvatures [21]. For PFs with irregular or complex structures, localized modeling approaches through population grouping have demonstrated effectiveness [13,22]. Furthermore, a Kriging-based PFE method has been proposed, where the PF is estimated by using a unit hyperplane in the objective space [23]. Moreover, the directional PF concept has been proposed to visualize trade-off structures even in regions where the PF does not explicitly exist [24]. While these PFE methods enhance the resolution of objective space information, they do not explicitly address the inverse mapping to the corresponding design variables in the variable space.

3.1.2. Pareto Set Estimation in Variable Space

PS estimation (PSE) aims to reconstruct the mapping from the objective space to the corresponding solutions in the variable space. The theoretical foundation of this field lies in the Design and Analysis of Computer Experiments (DACE) model [25,26]. By treating deterministic simulation outputs as realizations of a stochastic process via Kriging, a framework for simultaneous value prediction and uncertainty quantification at unexplored points was established [25,27]. This Kriging-based approach has been integrated into the Supervised Multi-objective Optimization Algorithm (SMOA) to efficiently estimate the PS in the variable space [3]. Recently, research on inverse modeling has accelerated, aiming to learn the direct inverse mapping from objective to variable space. One notable method, IM-MOEA/D, decomposes multivariate inverse functions into a group of univariate models, enabling direct sampling of promising solutions from ideal distributions in the objective space [28]. Additionally, combining support vector regression (SVR) with incremental learning and fuzzy clustering has been proposed to improve model adaptability while maintaining low computational overhead [29].

However, even with these methods, a structured PSE framework capable of simultaneously reconstructing multiple modes in multimodal problems—where the PS is divided into several discrete regions—has yet to be fully established [30].

3.2. Multimodal Multi-Objective Optimization

Solving MMOPs requires maintaining diversity in the variable space alongside convergence in the objective space [7,8]. Many methods utilizing clustering or space partitioning have been developed to address this multimodality. A prominent example is MMOEA/DC [31], which employs dual clustering in both the variable and objective spaces to maintain local Pareto sets within each cluster. The clustering-based Special Crowding Distance (SCD) was introduced in [32] to achieve a balanced population distribution across both spaces. Furthermore, spectral clustering has been applied in MMOCOA-SC [33], successfully separating non-convex PSs by exploiting the graph structure of the data. More recently, a multi-task learning framework [34] has been suggested, which utilizes auxiliary tasks to enhance diversity in the variable space. These trends indicate that structural analysis and space partitioning have become standard paradigms for handling MMOPs [5,35].

3.3. Absence of Multimodal Pareto Set Estimation and Positioning of This Work

PSE methods capable of estimating the structure of multimodal PSs from a limited set of solutions have not been established so far. Existing PSE and inverse modeling approaches [3,28] either assume a single continuous PS manifold or focus on approximating one-to-one relationships, making it difficult to dynamically identify and simultaneously approximate the one-to-many mappings inherent in MMOPs. Research incorporating manifold learning [13,36] also faces limitations, as a single manifold model cannot automatically detect the boundaries of a PS that is fragmented into numerous pieces.

To fill this technical gap, this paper proposes a space-partitioning PSE method using hierarchical clustering integrated with oscillation detection. The core innovation of the method lies in its ability to detect the precursors of one-to-many mappings as oscillations in variables during the model construction phase. By dynamically partitioning the variable space based on these detections, the method assigns independent local models to each mode, thereby enabling the comprehensive and structured estimation of multimodal Pareto sets.

4. Conventional Pareto Set Estimation

4.1. Overview

This work builds upon the conventional PSE method proposed in [3]. PSE estimates the PS from a limited number of known solutions. PSE consists of two processes: the learning process shown in Figure 1a and the estimation process shown in Figure 1b. The left side of the figure represents the decision space, and the right side represents the objective space. Each blue point (x^i, f^i) ($i \in \{1, 2, \dots, N\}$) consists of a variable vector x^i and an objective vector f^i , and they represent corresponding pairs across the decision and objective spaces. Figure 1 shows an example with $N = 9$. The direction vector e^i pointing to each objective vector f^i is illustrated as an arrow. A summary of the notation and abbreviations used throughout this paper is provided in Table A1 and Table A2, respectively.

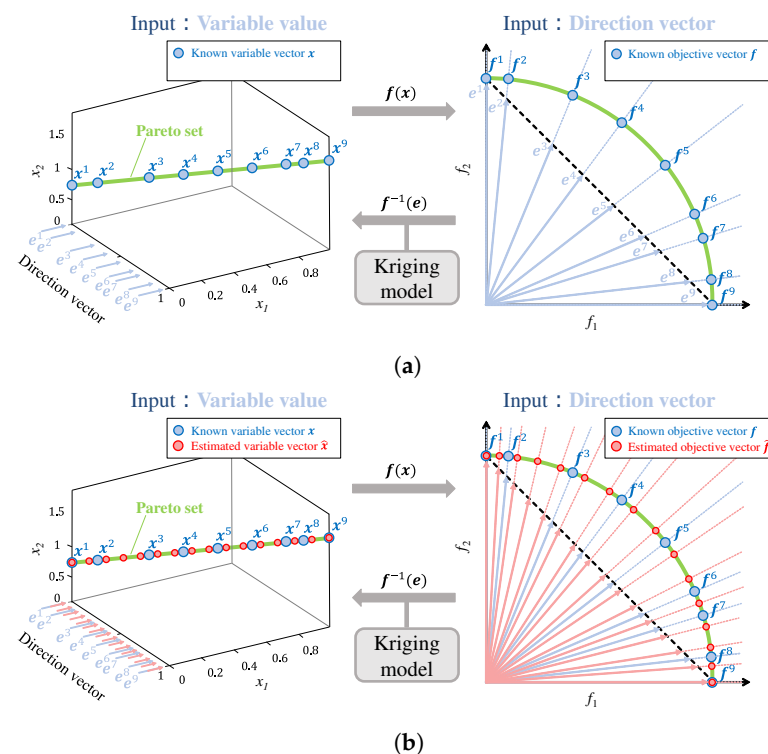


Figure 1. Conventional Pareto set estimation [3]. (a) Learning process. (b) Estimation process

In the learning process shown in Figure 1a, a response surface, referred to as the x -model, is constructed using Kriging [25], based on the set of known solutions $\mathcal{P} = \{(x^1, f^1), (x^2, f^2), \dots, (x^N, f^N)\}$. The input is the set of direction vectors $\mathcal{E} = \{e^1, e^2, \dots, e^N\}$, and the output is the set of variable vectors $\mathcal{X} = \{x^1, x^2, \dots, x^N\}$.

In the estimation process shown in Figure 1b, a large number of direction vectors $\hat{\mathcal{E}} = \{\hat{e}^1, \hat{e}^2, \dots\}$, illustrated as red arrows, are input to the x -model, and the estimated variable vectors $\{\hat{x}^1, \hat{x}^2, \dots\}$ are obtained.

While the objective function f maps a position x^i in the decision space to its corresponding position $f(x^i)$ in the objective space, PSE approximates the inverse mapping of f , associating a direction vector $\hat{e}$ in the objective space with a position $\hat{x}$ in the decision space.

4.2. Method

The pseudocode is shown in Algorithm 1. The input is the set of known solutions $\mathcal{P} = \{(x^1, f^1), (x^2, f^2), \dots, (x^N, f^N)\}$. From the known solution set $\mathcal{P}$, line 1 extracts the set of objective vectors $\mathcal{F}$, and line 2 extracts the set of variable vectors $\mathcal{X}$.

Algorithm 1 Conventional Pareto Set Estimation [3]

Require: Known solution set $\mathcal{P} = \{(x^1, f^1), (x^2, f^2), \dots, (x^N, f^N)\}$, a large direction vector set $\hat{\mathcal{E}} = \{\hat{e}^1, \hat{e}^2, \dots\}$
Ensure: Population $\mathcal{P}'$

(a) Learning Process

- 1: $\mathcal{F} = \{f^1, f^2, \dots, f^N\} \leftarrow$ Extract objective vectors ($\mathcal{P}$)
- 2: $\mathcal{X} = \{x^1, x^2, \dots, x^N\} \leftarrow$ Extract variable vectors ($\mathcal{P}$)
- 3: $\mathcal{E} \leftarrow \emptyset$ ▷ Direction vectors of known solutions
- 4: **for** $i \leftarrow 1, 2, \dots, N$ **do**
- 5: $e^i \leftarrow f^i / \|f^i\|_1$ ▷ Direction vector of known f^i
- 6: $\mathcal{E} \leftarrow \mathcal{E} \cup e^i$
- 7: **end for**
- 8: $x\text{-model} \leftarrow$ Train PS model ($\mathcal{E}, \mathcal{X}$)

(b) Estimation Process

- 9: $\mathcal{Q} \leftarrow \emptyset$ ▷ Solutions generated by PS estimation
- 10: **for each** $\hat{e} \in \hat{\mathcal{E}}$ **do**
- 11: $\hat{x} \leftarrow x\text{-model}(\hat{e})$
- 12: $\hat{f} \leftarrow f(\hat{x})$
- 13: $\mathcal{Q} \leftarrow \mathcal{Q} \cup (\hat{x}, \hat{f})$
- 14: **end for**
- 15: **return** $\mathcal{P}' \leftarrow$ Extract non-dominated ($\mathcal{P} \cup \mathcal{Q}$)

Lines 4–6 calculate the direction vector e^i corresponding to each objective vector f^i of the known solutions and add them to the set of direction vectors $\mathcal{E}$. Line 8 constructs the Kriging response surface model, referred to as the x -model, with the set of direction vectors $\mathcal{E}$ as input and the set of variable vectors $\mathcal{X}$ as output. This corresponds to the learning process shown in Figure 1a, where the variable vectors x (blue circles) in the decision space are learned from the direction vectors (blue arrows) in the objective space.

Line 9 initializes the solution set $\mathcal{Q}$ to store the estimated solutions. Lines 10–14 generate solutions $\hat{x}$ by iterating over each direction vector $\hat{e}$ in the uniformly distributed direction vector set $\hat{\mathcal{E}}$. In line 11, each direction vector $\hat{e}$ is provided to the x -model to obtain the corresponding estimated variable vector $\hat{x}$. In line 12, the estimated variable vector $\hat{x}$ is evaluated by the objective function f to obtain the objective vector $\hat{f}$. Line 13 adds the resulting solution pair $(\hat{x}, \hat{f})$ to the solution set $\mathcal{Q}$. This corresponds to the estimation process shown in Figure 1b.

Finally, line 15 extracts the non-dominated set $\mathcal{P}'$ from the union of the known solution set $\mathcal{P}$ and the estimated solution set $\mathcal{Q}$, and outputs it.

4.3. Limitation

The conventional PSE method encounters a fundamental difficulty when applied to MMOPs. In MMOPs, a single point on the PF in the objective space may correspond to multiple distinct points in the variable space. Equivalently, for a given direction vector $\hat{e}$ in the objective space, several valid variable vectors may exist. An illustrative example is shown in Figure 2a. In this example, the single green curve in the objective space on the right represents the PF, while it corresponds to two separate green lines in the variable space on the left. However, for each direction vector $\hat{e}$, the conventional PSE method predicts only a single variable vector $\hat{x}$, even when multiple corresponding solutions exist in the variable space. As a consequence, multiple modes of the PS are implicitly merged into a single estimation model. When such a model is used to generate new solutions, it may produce variable vectors in regions that deviate from the PS. As illustrated by the yellow region in Figure 2a, the estimated solutions shown in red can appear in areas that are distant from both green PS lines in the variable space. Consequently, their corresponding objective values deteriorate, leading to inferior solutions.

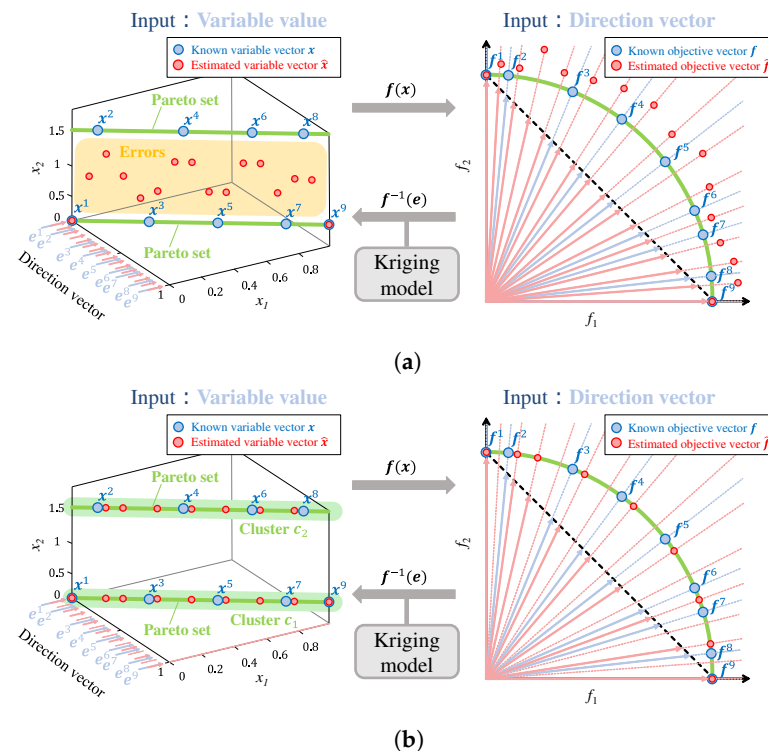


Figure 2. Pareto set estimation of a multimodal problem. (a) Conventional Pareto set estimation [3]. (b) Proposed Pareto set estimation.

To address this issue, as illustrated in Figure 2b, the spatially separated clusters of blue known solutions in the variable space should be partitioned into distinct groups, and independent prediction models should be constructed for each cluster. New solutions shown in red should then be generated separately within each model.

5. Proposed Multimodal Pareto Set Estimation

5.1. Overview

In this paper, we propose a clustering-based PSE method for MMOPs. The concept is illustrated in Figure 2b. In the proposed method, the set of known solutions $\mathcal{P}$ is divided into clusters in the variable space. In the example shown in Figure 2b, the known solution set $\mathcal{P}$ is partitioned into two clusters: $C_1 = \{x^1, x^3, x^5, x^7, x^9\}$ and $C_2 = \{x^2, x^4, x^6, x^8\}$.

This decomposition is introduced because conventional inverse modeling implicitly assumes a one-to-one correspondence between direction vectors and variable vectors. When multiple decision vectors correspond to similar objective vectors, a single global response surface tends to merge distinct local structures and loses multimodal information. By dividing known solutions into locally monotonic clusters, each response surface model can approximate a simpler local mapping, which makes it possible to preserve multiple decision-space modes.

Next, a response surface model is constructed for each cluster. In Figure 2b, the x_{C_1} -model and the x_{C_2} -model are built. Then, a large set of uniformly distributed direction vectors $\hat{\mathcal{E}}$ is provided to both response surface models, and the estimated PS (shown in red) is obtained from each model. By inputting each direction vector $\hat{e} \in \hat{\mathcal{E}}$ into different response surface models, multiple PS modes can be reconstructed, enabling accurate PS estimation even in MMOPs.

5.2. Hierarchical Clustering with Oscillation Detection

We focus on hierarchical clustering [37]. Conventional hierarchical clustering starts with each data point as an independent cluster and repeatedly merges the two closest clusters. In contrast, the proposed hierarchical clustering incorporates an oscillation detection step during the merging process.

The methodological outline is as follows. For each variable dimension j , hierarchical clustering is applied independently to the ordered samples in the (e_1, x_j) space. To determine whether two candidate clusters C_a and C_b can be merged, the merged variable value sequence is examined for oscillatory behavior. Let $v = (v_1, v_2, \dots, v_{|v|})$ denote the sequence of local extrema extracted from the merged variable value sequence with a sliding window of width w , which suppresses the detection of gradual periodic variations as oscillations. The number of significant adjacent extrema is defined as $peak_{sum} = |\{k \in \{1, 2, \dots, |v| - 1\} \mid |v_k - v_{k+1}| \geq amp_j\}|$, where amp_j is the amplitude threshold for variable j , introduced to prevent clusters from being separated by small fluctuations. An oscillation is detected when $peak_{sum} \geq p$, where p specifies the minimum number of significant extrema required to regard the merged sequence as oscillatory, and only cluster pairs that do not satisfy this condition are merged, thereby preserving locally monotonic cluster structure.

The methodological details are given below. The pseudocode of the proposed hierarchical clustering with oscillation detection is shown in Algorithm 2. Line 1 prepares the set of cluster vectors $\mathcal{Z}$. For each known solution $x^i \in \mathcal{X}$ ($i = 1, 2, \dots, N$), a cluster vector $z^i = (z_1^i, z_2^i, \dots, z_d^i)$ is defined. Here, $z_j^i \in \{1, 2, \dots, k\}$ represents the cluster number when focusing on dimension j of the variable space.

Algorithm 2 Clustering

Require: Direction vectors of known solutions $\mathcal{E}$, variable vectors of known solutions $\mathcal{X}$

Ensure: Cluster labels $l = (l_1, l_2, \dots, l_N) \in \{1, 2, \dots, k\}^N$

```

1:  $\mathcal{Z} = \{z^1, z^2, \dots, z^N\} \leftarrow \{0, 0, \dots, 0\}$  ▷ Cluster vector set
2: for  $j \leftarrow 1, 2, \dots, D$  do
3:    $amp_j \leftarrow \gamma_{amp} \cdot (x_j^{\max} - x_j^{\min})$ 
4:    $l_j \leftarrow \text{OSCILLATION-AWARE LINKAGE}(\mathcal{E}, \mathcal{X}, j, w, p, amp_j)$  ▷ Algorithm 3
5:   for  $i \leftarrow 1, 2, \dots, N$  do
6:      $z_j^i \leftarrow l_{j,i}$ 
7:   end for
8: end for
9:  $l \leftarrow \text{Exact-match clustering of known solutions using } \mathcal{Z}$ 
10: return  $l$ 
```

In lines 2–8, clustering is repeated D times, once for each dimension $j \in \{1, 2, \dots, D\}$. Line 3 defines an amplitude threshold as $amp_j = \gamma_{amp}(x_j^{\max} - x_j^{\min})$, which represents the maximum amplitude allowed within a single cluster. γ_{amp} is an amplitude threshold coefficient and ranges from 0 to 1.

For each variable dimension j , the dataset used for clustering is defined as $\mathcal{EX} = \{(e_1^1, x_j^1), (e_1^2, x_j^2), \dots, (e_1^N, x_j^N)\}$, where e_1^i denotes the first component of the direction vector e^i . Unlike [14], which determined oscillation solely based on the count of local extrema, this paper introduces the amplitude-based criterion to suppress over-segmentation caused by minor fluctuations. Line 4 performs clustering for dimension j with oscillation detection, as described in Algorithm 3.

In Algorithm 3, line 1 starts with N data points as independent clusters. Line 2 constructs the dataset $\mathcal{EX}$ by pairing the first elements of the direction vectors with the corresponding values in the j -th variable dimension, as defined above. Line 3 calculates the distance matrix $\mathcal{D}$ of $\mathcal{EX}$. In this paper, the distance between the two closest data points

in different clusters is used. Lines 4–16 repeat the merging process until only one cluster remains, or until no further clusters can be merged due to oscillation-based divergence.

Algorithm 3 Oscillation-Aware Linkage

Require: 1-th direction vector of known solutions $\mathcal{E}_1 = \{e_1^1, e_1^2, \dots, e_1^N\}$, j -th dimension variable value of known solutions $\mathcal{X}_j = \{x_j^1, x_j^2, \dots, x_j^N\}$, window size w , peak-count threshold p , j -th amplitude threshold amp_j

Ensure: Cluster labels $I \in \{1, \dots, k\}^N$

```

1:  $\mathcal{C} \leftarrow \{\{1\}, \dots, \{N\}\}$  ▷ index per cluster
2:  $\mathcal{E}\mathcal{X} \leftarrow \{(e_1^1, x_j^1), (e_1^2, x_j^2), \dots, (e_1^N, x_j^N)\}$ 
3: Compute pairwise distances  $\mathcal{D} \in \mathbb{R}^{|\mathcal{C}| \times |\mathcal{C}|}$  of  $\mathcal{E}\mathcal{X}$  with  $\mathcal{D}_{ii} \leftarrow \infty$ 
4: while more than one cluster remains do
5:    $a, b \leftarrow \arg \min_{a, b \in \{1, 2, \dots, |\mathcal{C}|\}} \mathcal{D}_{ab}$ 
6:    $\mathcal{E}\mathcal{X}_{C_a}, \mathcal{E}\mathcal{X}_{C_b} \leftarrow$  data points in  $C_a, C_b$ 
7:   if  $\neg \text{Is Oscillating}(\mathcal{E}\mathcal{X}_{C_a}, \mathcal{E}\mathcal{X}_{C_b}, w, p, amp_j)$  then ▷ Algorithm 4
8:      $C_a \leftarrow C_a \cup C_b$ , remove  $C_b$  from  $\mathcal{C}$ 
9:      $\mathcal{D} \leftarrow$  Update distances from  $C_a$ 
10:  else
11:     $\mathcal{D}_{ab}, \mathcal{D}_{ba} \leftarrow \infty$ 
12:  end if
13:  if all  $\mathcal{D}_{ab} = \infty$  then
14:    break
15:  end if
16: end while
17:  $I = (I_1, I_2, \dots, I_N) \leftarrow 0$  ▷ Cluster index vector
18: for each  $C_k \in \mathcal{C}$  do
19:   for each  $i \in C_k$  do
20:      $I_i \leftarrow k$ 
21:   end for
22: end for
23: return  $I$ 

```

Algorithm 4 Is Oscillating

Require: Data $\mathcal{E}\mathcal{X}_{C_a} = \{(e_1^a, x_j^a) \mid a \in C_a\}$ and $\mathcal{E}\mathcal{X}_{C_b} = \{(e_1^b, x_j^b) \mid b \in C_b\}$, window size w , peak-count threshold p , j -th amplitude threshold amp_j

Ensure: Boolean flag indicating vibration

```

1:  $\mathcal{E}\mathcal{X}_{C_{ab}} \leftarrow$  Sort paired elements in  $\mathcal{E}\mathcal{X}_{C_a} \cup \mathcal{E}\mathcal{X}_{C_b}$  in ascending order of  $e_1$ 
2:  $\mathbf{y} = (y_1, y_2, \dots, y_{|C_a|+|C_b|}) \leftarrow$  Extract variable values from sorted  $\mathcal{E}\mathcal{X}_{C_{ab}}$ 
3: for  $i \leftarrow 1$  to  $|C_a| + |C_b| - w + 1$  do
4:    $\mathbf{v} = (v_1, v_2, \dots, v_{|v|}) \leftarrow$  Extract local maxima and minima in sequence  $y_i, y_{i+1}, \dots, y_{i+w-1}$ 
5:    $peak_{sum} \leftarrow |\{k \in \{1, 2, \dots, |v| - 1\} \mid |v_k - v_{k+1}| \geq amp_j\}|$ 
6:   if  $peak_{sum} \geq p$  then
7:     return true
8:   end if
9: end for
10: return false

```

Lines 5 and 6 extract the two closest clusters from the distance matrix $\mathcal{D}$. Based on single linkage, let C_a and C_b denote the closest cluster pair selected as a merging candidate. The samples belonging to these clusters are represented as $\mathcal{E}\mathcal{X}_{C_a}$ and $\mathcal{E}\mathcal{X}_{C_b}$.

Line 7 applies oscillation detection: only if the combined data points exhibit monotonic changes are the clusters merged at line 8, and the distance matrix $\mathcal{D}$ is updated at line 9. The pseudocode for oscillation detection is shown in Algorithm 4.

In Algorithm 4, line 1 sorts the paired elements in the combined data of the closest clusters, $\mathcal{E}\mathcal{X}_{C_a} \cup \mathcal{E}\mathcal{X}_{C_b}$, in ascending order of e_1 and stores them as $\mathcal{E}\mathcal{X}_{C_{ab}}$. Although each element in $\mathcal{E}\mathcal{X}_{C_{ab}}$ is a pair such as (e_1, x_j) , line 2 extracts only the variable values from $\mathcal{E}\mathcal{X}_{C_{ab}}$ as $\mathbf{y} = (y_1, y_2, \dots, y_{|C_a|+|C_b|})$, where $|C_a|$ and $|C_b|$ denote the sizes of clusters C_a and C_b , respectively. For oscillation detection, lines 3–9 employ a sliding window of width w over the value sequence $\mathbf{y}$. Line 4 extracts local maxima and minima in a subsequence of w elements from $\mathbf{y}$ as $\mathbf{v} = (v_1, v_2, \dots, v_{|v|})$, where $|v|$ denotes the number of extrema. Line 5 counts the number of adjacent extrema for which the absolute difference between

successive extrema is greater than or equal to amp_j as $peak_{sum}$. In lines 6–8, if $peak_{sum} \geq p$ in any window, the sequence is judged to be oscillatory and the cluster pair is not merged.

Returning to Algorithm 3, line 11 handles the case where the two clusters are divergent by setting the corresponding distance in the matrix to ∞ so that they will not be merged later. Lines 13–15 handle the case where all elements of the distance matrix become ∞ , indicating that no clusters can be merged and the clustering process terminates. Line 17 prepares the cluster number vector. Lines 18–22 assign the cluster number k to each element x_j^i of the known solutions x^i and record it as l_i .

Returning to Algorithm 2, in lines 5–7, the cluster numbers $l_{j,i}$ ($i = 1, 2, \dots, N$) obtained for dimension j are stored as z_j^i in the cluster vector set $\mathcal{Z}$. After repeating lines 2–8 for all D dimensions, the cluster vector set $\mathcal{Z}$ is completed. Finally, line 9 groups the known solutions x^i ($i = 1, 2, \dots, N$) whose cluster vectors z^i are identical and outputs the final cluster vector $\mathbf{l} = (l_1, l_2, \dots, l_N) \in \{1, 2, \dots, k\}^N$.

If a cluster contains only a single solution, a response surface model cannot be constructed. In [14], such single-solution clusters could be generated. In this paper, a reassignment mechanism is introduced in which the solution is merged into the nearest neighboring cluster, thereby ensuring model constructability and improving robustness. Because this operation is applied only when a cluster contains a single solution and no response surface can otherwise be constructed, it is treated as a minimal implementation-level adjustment rather than a structural modification of the clustering result.

In the current formulation, oscillation-aware linkage uses only the first direction component e_1 as a scalar ordering coordinate, which is naturally suited to bi-objective settings where local ordering is directly induced along a single direction axis. Extension to higher-dimensional objective spaces requires defining multiple scalar orderings or an alternative ordering mechanism derived from the full direction vector.

5.3. Multimodal PSE with Clustering and Oscillation Detection

The methodological outline is as follows. Based on the obtained clusters, a PS estimation model is constructed independently for each cluster C_k . Let the cluster-wise model for C_k be denoted by the x_{C_k} -model. Using the direction vectors $\mathcal{E}_{C_k}$ and variable vectors $\mathcal{X}_{C_k}$ extracted from C_k , the model is defined as

$$x_{C_k}\text{-model} : e \mapsto x, \quad (5)$$

where $e \in \mathcal{E}_{C_k}$ is a direction vector in cluster C_k , and $x \in \mathcal{X}_{C_k}$ is the corresponding variable vector.

For a given direction vector $\hat{e}$, the estimated variable vector obtained from cluster C_k is written as

$$\hat{x} = x_{C_k}\text{-model}(\hat{e}). \quad (6)$$

By applying this estimation independently to all clusters whose direction ranges involve $\hat{e}$, multiple candidate variable vectors can be generated for the same objective-space direction. This cluster-wise independence allows distinct variable vectors to be estimated for the same objective-space direction when multiple local Pareto-set branches coexist.

The methodological details are given below. The pseudocode of the proposed multimodal PSE method is shown in Algorithm 5. Figure 3 presents a flowchart of the proposed multimodal PSE method to facilitate understanding of the overall procedure. To improve traceability, the corresponding line numbers in Algorithm 5 are also indicated in the flowchart. The differences from the conventional PSE method in Algorithm 1 are marked in red. At line 8, the known solution set is divided into clusters $\mathcal{C} = \{C_1, C_2, \dots\}$ in the vari-

able space using hierarchical clustering with oscillation detection, described in Algorithm 2 of Section 5.2. In the example shown in Figure 2b, the set is divided into $\mathcal{C} = \{C_1, C_2\}$.

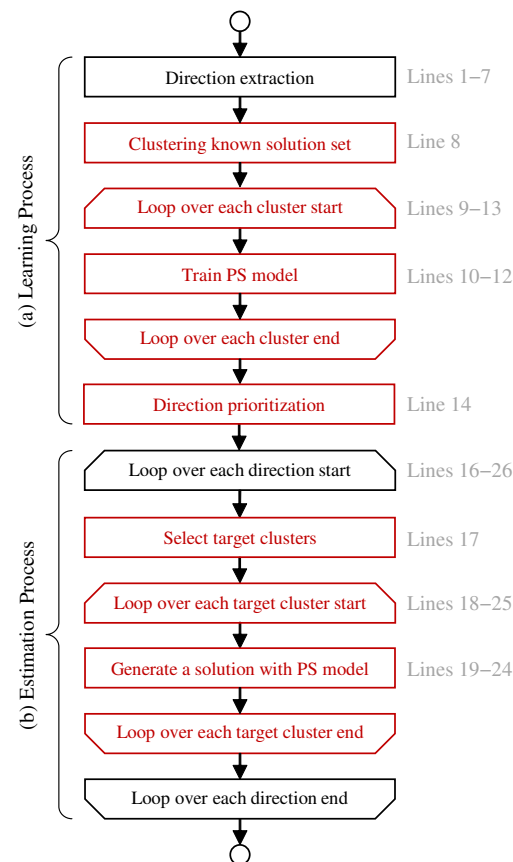


Figure 3. Flowchart of the proposed multimodal Pareto set estimation.

Lines 9–13 build a PS estimation model, denoted as the x_{C_k} -model, for each cluster $C_k \in \mathcal{C}$. In the example, both the x_{C_1} -model and the x_{C_2} -model are constructed.

Line 14 assigns priorities to the direction vectors in the large set $\hat{\mathcal{E}}$ for PS estimation. Specifically, in Algorithm 6, line 1 first selects the edge direction vectors that have 1 as one of their components from $\hat{\mathcal{E}}$, and stores them in $\hat{\mathcal{E}}'$. For example, in Figure 2b, $\hat{\mathcal{E}}' = \{\hat{e}^1 = (0, 1), \hat{e}^{|\hat{\mathcal{E}}|=16} = (1, 0)\}$. Lines 2–5 repeatedly add to $\hat{\mathcal{E}}'$ the direction vector $\hat{e}$ that is farthest from the already selected ones. In the example, $\hat{\mathcal{E}}'$ becomes $\{\hat{e}^1, \hat{e}^{|\hat{\mathcal{E}}|=16}, \hat{e}^8, \hat{e}^{12}, \dots\}$. This ensures that even if the entire set $\hat{\mathcal{E}}$ is not fully explored, the selected direction vectors are approximately uniformly distributed in the objective space.

Returning to Algorithm 5, lines 16–26 perform PS estimation by focusing on each direction vector $\hat{e} \in \hat{\mathcal{E}}'$ in order of priority. Line 17 selects the set of clusters $\mathcal{C}'$ that are compatible with the direction vector $\hat{e}$. Specifically, clusters whose direction-vector ranges include the first component $\hat{e}_1$ of the target direction vector $\hat{e}$ are selected, so that PS estimation is performed only within clusters relevant to the given direction.

Lines 18–25 then compute the estimated variable vector $\hat{x}$ from the response surface x_{C_k} -model of each cluster $C_k \in \mathcal{C}'$ for the given direction vector $\hat{e}$. In other words, for a single direction vector $\hat{e}$, $|\mathcal{C}'|$ estimated variable vectors are generated.

At line 22, the algorithm terminates when $|\hat{\mathcal{E}}|$ variable vectors have been generated. This implies that not all direction vectors in $\hat{\mathcal{E}}$ are necessarily explored. Therefore, at line 14, the direction vectors are ordered so that they are approximately uniformly distributed in the objective space.

Algorithm 5 Proposed Multimodal Pareto Set Estimation

Require: Known solution set $\mathcal{P} = \{(x^1, f^1), (x^2, f^2), \dots, (x^N, f^N)\}$, a large direction vector set $\hat{\mathcal{E}} = \{\hat{e}^1, \hat{e}^2, \dots\}$

Ensure: Population $\mathcal{P}'$

(a) Learning Process

- 1: $\mathcal{F} = \{f^1, f^2, \dots, f^N\} \leftarrow$ Extract objective vectors ($\mathcal{P}$)
- 2: $\mathcal{X} = \{x^1, x^2, \dots, x^N\} \leftarrow$ Extract variable vectors ($\mathcal{P}$)
- 3: $\mathcal{E} \leftarrow \emptyset$ ▷ Direction vectors of known solutions
- 4: **for** $i \leftarrow 1, 2, \dots, N$ **do**
- 5: $e^i \leftarrow f^i / \|f^i\|_1$ ▷ Direction vector of known f^i
- 6: $\mathcal{E} \leftarrow \mathcal{E} \cup e^i$
- 7: **end for**
- 8: $\mathcal{C} = \{C_1, C_2, \dots\} \leftarrow \text{CLUSTERING}(\mathcal{E}, \mathcal{X})$ ▷ Algorithm 2
- 9: **for each** $C_k \in \mathcal{C}$ **do**
- 10: $\mathcal{E}_{C_k} \leftarrow$ Extract direction vectors (C_k)
- 11: $\mathcal{X}_{C_k} \leftarrow$ Extract variable vectors (C_k)
- 12: $x_{C_k}\text{-model} \leftarrow$ Train PS model ($\mathcal{E}_{C_k}, \mathcal{X}_{C_k}$)
- 13: **end for**
- 14: $\hat{\mathcal{E}}' \leftarrow \text{PRIORITIZE DIRECTIONS}(\hat{\mathcal{E}})$ ▷ Algorithm 6

(b) Estimation Process

- 15: $\mathcal{Q} \leftarrow \emptyset$ ▷ Solutions generated by PS estimation
- 16: **for each** $\hat{e} \in \hat{\mathcal{E}}'$ **do**
- 17: $\mathcal{C}' \leftarrow \{C_k \in \mathcal{C} \mid (\hat{e}_1 \geq \min_{i \in C_k} e_1^i) \wedge (\hat{e}_1 \leq \max_{i \in C_k} e_1^i)\}$ ▷ Target cluster selection
- 18: **for each** $C_k \in \mathcal{C}'$ **do**
- 19: $\hat{x} \leftarrow x_{C_k}\text{-model}(\hat{e})$
- 20: $\hat{f} \leftarrow \text{Evaluation}(\hat{x})$
- 21: $\mathcal{Q} \leftarrow \mathcal{Q} \cup (\hat{x}, \hat{f})$
- 22: **if** $|\mathcal{Q}| \geq |\hat{\mathcal{E}}|$ **then**
- 23: **return** $\mathcal{P}' \leftarrow$ Extract non-dominated ($\mathcal{P} \cup \mathcal{Q}$)
- 24: **end if**
- 25: **end for**
- 26: **end for**

Algorithm 6 Prioritize Directions

Require: A large direction vector set $\hat{\mathcal{E}} = \{\hat{e}^1, \hat{e}^2, \dots, \hat{e}^{|\hat{\mathcal{E}}|}\}$

Ensure: Prioritized direction vector set $\hat{\mathcal{E}}'$

- 1: $\hat{\mathcal{E}}' \leftarrow$ Pick extreme vectors with an element of 1 from $\hat{\mathcal{E}}$
- 2: **while** $|\hat{\mathcal{E}}'| < |\hat{\mathcal{E}}|$ **do**
- 3: $\hat{e} \leftarrow$ Pick the farthest vector from $\hat{\mathcal{E}}$
- 4: $\hat{\mathcal{E}}' \leftarrow \hat{\mathcal{E}}' \cup \hat{e}$
- 5: **end while**
- 6: **return** $\hat{\mathcal{E}}'$

5.4. Expected Effect

In the conventional method, a single global PS estimation model is constructed as

$$x\text{-model} : e \mapsto x, \quad (7)$$

which implicitly assumes a one-to-one correspondence between direction vectors and variable vectors. In contrast, the proposed method constructs multiple local models

$$x_{C_k}\text{-model} : e \mapsto x \quad (k = 1, 2, \dots, K), \quad (8)$$

so that multiple distinct variable vectors can be associated with similar direction vectors through different clusters. Because each cluster-wise model is trained on a locally continuous mapping, interpolation is performed within a single branch of the Pareto set, whereas a global model tends to average multiple disconnected branches. As a result, the proposed method can preserve multimodal structures that are otherwise smoothed out in conventional single-model estimation.

By constructing a dedicated x_{C_k} -model for each cluster C_k , the variable space is represented by a set of local response surfaces rather than a single global model. For a given

direction vector $\hat{e}$, each relevant cluster model is queried independently. This allows multiple variable vectors to be associated with the same objective-space direction. As a result, multimodal characteristics of the PS can be preserved, and the estimation accuracy is expected to improve in problems where conventional PSE collapses distinct modes.

6. Experimental Setup

6.1. Method

The proposed multimodal PSE was compared with the conventional PSE [3]. The Kriging model was configured using a quadratic polynomial regression model and an exponential correlation function, where the correlation parameter was initialized at 1 with lower and upper bounds of 0.1 and 1, respectively. The size of the direction vector set was fixed to $|\mathcal{E}| = 1000$. Unless otherwise specified, the window size, the peak-count threshold, and the amplitude threshold coefficient for oscillation detection were set to $w = 10$, $p = 0.4w$, and $\gamma_{\text{amp}} = 0.1$, respectively. The validity of this setting is examined in Sections 7.4 and 7.5.

6.2. Problems

We used the unconstrained MMF1–8 problems [5] with $M = 2$ objectives and $D = 2$ variables, and the constrained LIRCMOP1–2 problems [6] with $M = 2$ objectives, $K = 2$ constraints, and $D = 30$ variables.

The known solution sets $\mathcal{P}$ were obtained using evolutionary algorithms. The common genetic operators were configured as follows: simulated binary crossover with crossover probability 1.0 and distribution index 20, polynomial mutation with mutation probability $1/D$ and distribution index 20, and differential evolution with crossover rate 1.0 and scaling factor 0.5. For MMF1–8, HREA [38] was executed with a population size of 100 for 4000 generations. For HREA, the parent selection probability from the convergence archive was set to $p = 0.5$, and the acceptable local PF gap parameter was set to $\epsilon = 0.3$. For LIRCMOP1, CAEAD [39] was used with 40 individuals for 9900 generations. For CAEAD, the stage-switch threshold was set to 0.01. For LIRCMOP2, URCMO [40] was applied with 40 individuals for 6000 generations. For URCMO, the strategy adaptation control parameter was set to $\alpha = 0.9$, and the auxiliary population update probability was set to $p = 0.1$. The sizes of the non-dominated solution sets were $N = 78, 78, 69, 78, 72, 71, 84, 62, 40$, and 40 for MMF1–8 and LIRCMOP1–2, respectively.

In this study, benchmark problems are used to examine whether the proposed method can preserve multimodal Pareto set structures under the above setting. The experimental setting is intended to provide a comparative assessment under identical sampling conditions rather than to emulate a specific practical evaluation budget.

6.3. Metric

PF approximation performance in the objective space is evaluated using Hypervolume (HV) [9] and Inverted Generational Distance (IGD) [10–12]. HV measures the M -dimensional volume dominated by the solution set $\mathcal{P}$ with respect to a reference point $\mathbf{r}$. A larger HV indicates better convergence and diversity. All objective values are normalized to $[0, 1]$, and $\mathbf{r} = (1.1, 1.1)$ is used. IGD is defined as the average distance from each reference point on the true PF to its nearest solution in $\mathcal{P}$. A smaller IGD indicates better PF approximation.

PS approximation performance in the variable space and Inverted Generational Distance in the decision space (IGDX) [13] are used. IGDX computes the average distance from each reference point on the true PS to its nearest solution in $\mathcal{P}$. A smaller IGDX indicates

better PS approximation. In MMOPs, IGDX is particularly effective for assessing whether separated PS components are uniformly approximated.

For IGD calculation, 10,000 reference points on the mathematically defined PF were used for all benchmark problems. These points were generated by sampling one objective at equal intervals and calculating the corresponding points on the PF. For IGDX calculation, the reference set was constructed from all non-dominated variable vectors obtained during the optimization process conducted in advance to generate the known solutions used in the comparative numerical experiments.

7. Experimental Results and Discussion

7.1. Clusters of Solution Sets in the Variable Space

The distribution of the known solution set $\mathcal{P}$ in the variable space and the clustering results obtained by the proposed multimodal PSE are shown in Figure 4. As in the left panels of Figures 1 and 2, the axis of the direction-vector component e_1 is added to the variable space for visualization. For LIRCMOP1 and LIRCMOP2, only two dimensions out of the $D = 30$ variables are plotted. The cluster sizes are summarized in Table 1.

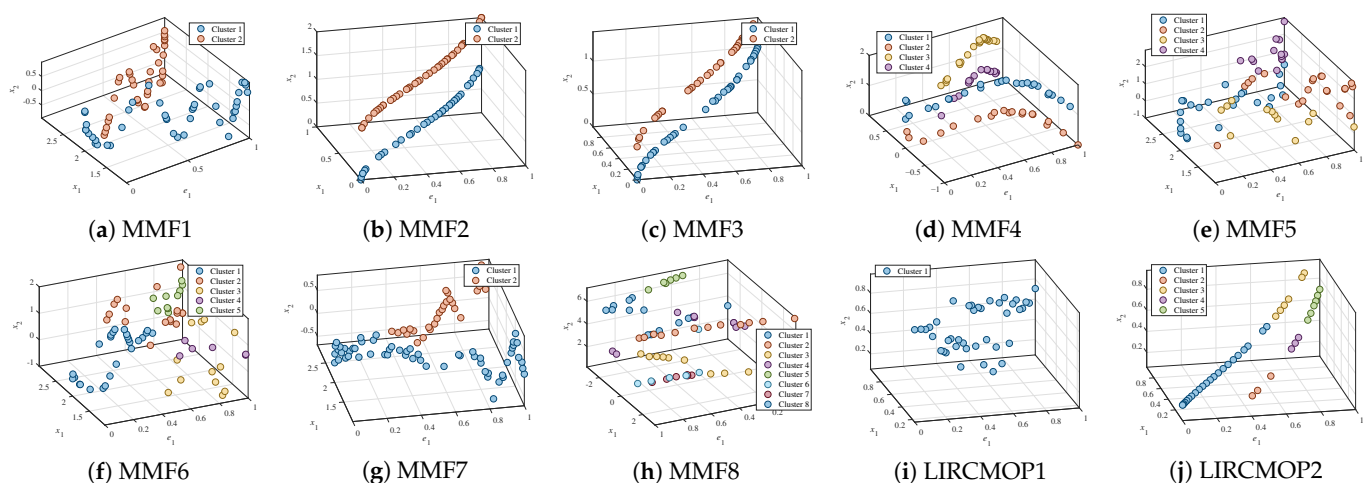


Figure 4. Known solutions $\mathcal{P}$ clustered by the proposed multimodal PSE in the variable space. Each color represents a cluster identified in the decision space, where solutions belonging to the same cluster are expected to share a similar local PS structure. The number of legend entries corresponds to the number of clusters, and the cluster sizes are summarized in Table 1.

Table 1. Obtained cluster sizes.

Cluster Number	MMF1	MMF2	MMF3	MMF4	MMF5	MMF6	MMF7	MMF8	LIRCMOP1	LIRCMOP2
1	43	36	37	25	25	26	57	7	40	22
2	35	42	32	19	21	17	27	12	-	3
3	-	-	-	20	12	12	-	9	-	6
4	-	-	-	14	14	6	-	9	-	3
5	-	-	-	-	-	10	-	6	-	6
6	-	-	-	-	-	-	-	6	-	-
7	-	-	-	-	-	-	-	6	-	-
8	-	-	-	-	-	-	-	7	-	-

The results show that $\mathcal{P}$ is divided into two clusters for MMF1–3 and MMF7, and into four clusters for MMF4 and MMF5. In contrast, only a single cluster is formed for LIRCMOP1, whereas five or more clusters are identified for several other problems such as MMF6, MMF8, and LIRCMOP2.

In these problems, the proposed multimodal PSE successfully separates $\mathcal{P}$ according to differences in its distribution in the variable space. In MMF1, shown in Figure 4a, the x_2 values of the known solutions exhibit periodic behavior. However, since oscillation detec-

tion is based on the number of peaks within a sliding window, smooth periodic variation is not regarded as multimodality. As a result, the solutions are appropriately clustered.

For LIRCMOP1, shown in Figure 4i, only a single cluster is formed despite the high-dimensional distribution complexity. LIRCMOP1 has $D = 30$ variables and a highly complex distribution of known solutions. In the proposed multimodal PSE, clustering is first performed independently for each dimension and then integrated into multidimensional clusters. During this integration process, each solution was treated as an independent singleton cluster, which ultimately resulted in all solutions being merged into a single cluster. This indicates that LIRCMOP1 has distribution characteristics that make cluster separation difficult for the proposed multimodal PSE.

On the other hand, problems such as Figure 4f,h,j are divided into five or more clusters. Although the solution set is divided into multiple clusters due to distribution complexity, the overall structural characteristics of the distribution are appropriately captured.

7.2. Solution Distribution in the Variable Space

We analyze the distribution of the known solution set $\mathcal{P}$ and the estimated solution set $\mathcal{Q}$ in the variable space. The results of the conventional PSE are shown in Figure 5, and those of the proposed multimodal PSE are shown in Figure 6. In the figures, blue points represent feasible known solutions in $\mathcal{P}$, whereas red and green points denote feasible and infeasible solutions in the estimated solution set $\mathcal{Q}$ generated by PSE, respectively. The green infeasible solutions appear only in LIRCMOPs, which include constraints.

As shown in Figure 5, the estimated solutions of the conventional PSE are often distributed far from the regions densely populated by the known solutions. In contrast, Figure 6 shows that the estimated solutions of the proposed multimodal PSE are concentrated around these regions.

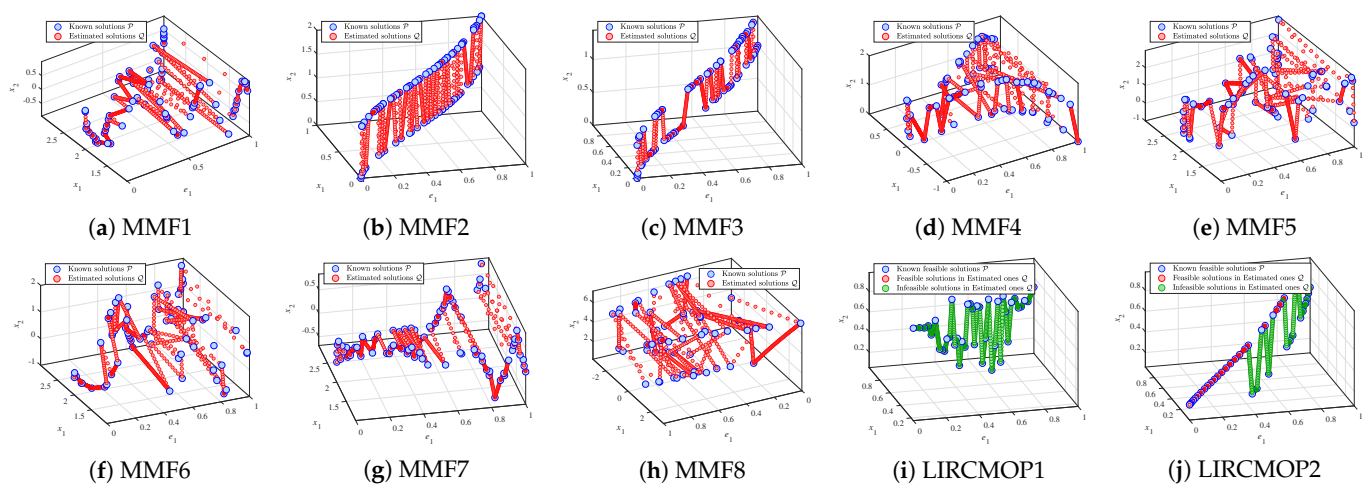


Figure 5. Known solutions and estimated ones by the conventional PSE [3] in the variable space.

For the constrained LIRCMOPs, infeasible solutions (green points) are observed in both methods, whereas their occurrence is reduced by the proposed multimodal PSE in LIRCMOP2. In LIRCMOP1, the conventional and proposed methods exhibit similar behavior. In contrast, in LIRCMOP2, the proposed multimodal PSE more effectively suppresses the generation of infeasible solutions.

However, because the proposed model estimates solutions mainly within clusters, the number of clusters may exceed the true number of variable regions. In such cases, a single variable region can be divided into multiple clusters. If boundary solutions are not included in the initial samples, the estimated region within each cluster may become narrower.

Figure 6j illustrates this phenomenon. Although the true Pareto set consists of two separated regions, the proposed multimodal PSE partitions the solutions into five clusters, as shown in Figure 4j. While estimation within each cluster is accurate, no estimated solutions appear between clusters. As a result, the intermediate regions are not reconstructed, and gaps arise between clusters. This may lead to a loss of diversity in the variable space.

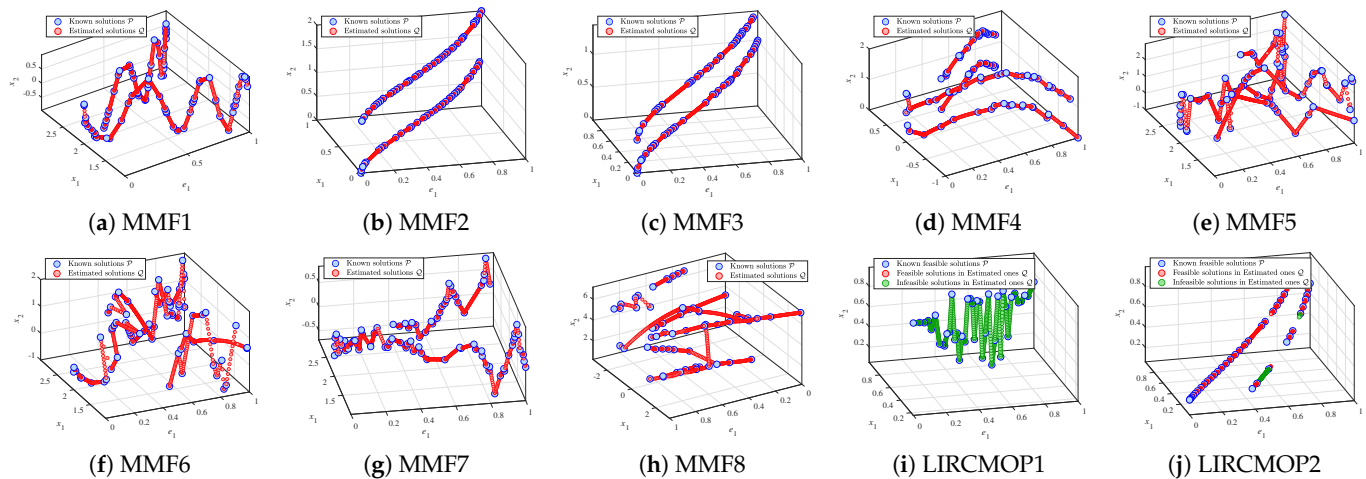


Figure 6. Known solutions and estimated ones by the proposed multimodal PSE in the variable space.

The IGDX results for the solution sets $\mathcal{P}'$ obtained by the conventional and proposed methods are shown in Table 2. The better value for each problem is highlighted in bold.

Table 2. Quantitative evaluation of the obtained solutions $\mathcal{P}'$ in the variable space based on IGDX (smaller is better; best values are shown in bold).

Metric	Method	MMF1	MMF2	MMF3	MMF4	MMF5	MMF6	MMF7	MMF8	LIRCMOP1	LIRCMOP2
IGDX	Conventional PSE [3]	0.0703	0.0180	0.0219	0.0509	0.0992	0.1177	0.0402	0.1653	0.6356	0.2314
	Proposed Multimodal PSE	0.0525	0.0028	0.0082	0.0339	0.0975	0.1145	0.0347	0.1621	0.6356	0.4781

The proposed multimodal PSE achieves lower IGDX values than the conventional PSE for all problems except LIRCMOP2.

For LIRCMOP2, as discussed above, the diversity loss caused by gaps between clusters results in a higher IGDX value for the proposed method.

Such a case is intentionally retained in this study to provide a fair assessment of the proposed method, including situations where cluster-wise decomposition does not necessarily improve recovery performance. Introducing stabilization mechanisms such as minimum cluster size constraints or adaptive merging is an important direction for mitigating this limitation.

7.3. Solution Distribution in the Objective Space

We now examine the distribution of the known solution set $\mathcal{P}$ and the estimated solution set $\mathcal{Q}$ in the objective space.

Figure 7 and Figure 8 present the solution sets obtained by the conventional method and proposed method, respectively.

As shown in Figure 7, the conventional PSE generates a considerable number of dominated solutions, whereas Figure 8 shows that the proposed multimodal PSE significantly suppresses their generation.

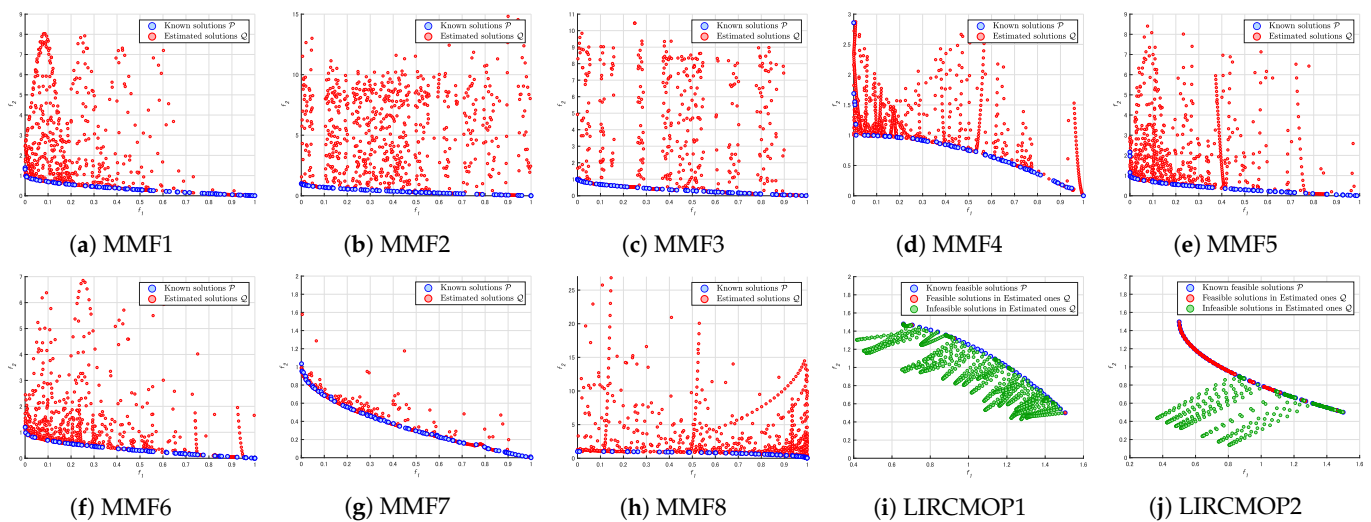


Figure 7. Known solutions and estimated ones by the conventional PSE [3] in the objective space.

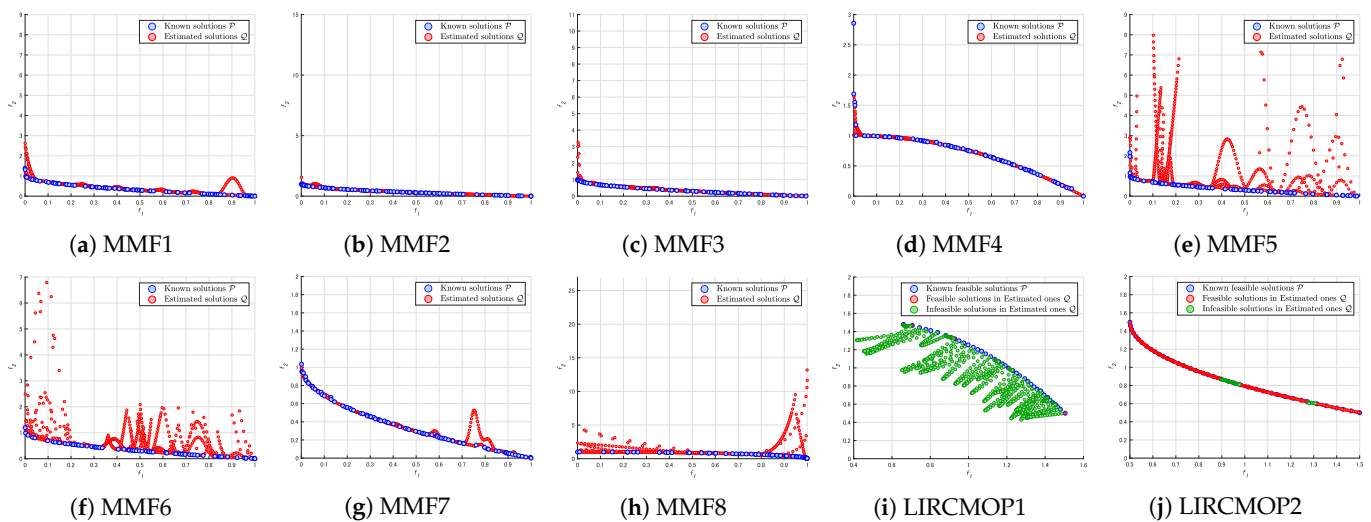


Figure 8. Known solutions and estimated ones by the proposed multimodal PSE in the objective space.

For the constrained LIRCMOPs, infeasible solutions (green points) are observed in both methods, while the proposed multimodal PSE more effectively suppresses their generation in LIRCMOP2. In LIRCMOP1, both methods exhibit similar behavior.

As discussed in Figure 5, this is attributed to the fact that many estimated solutions are located far from the regions densely populated by the known solutions in the variable space. Consequently, these solutions are projected to inferior regions in the objective space. In contrast, because the proposed method estimates solutions near the densely populated regions of known solutions in the variable space (see Figure 6), the resulting objective-space distribution remains closer to the Pareto front.

The HV and IGD results for the solution sets $\mathcal{P}'$ obtained by the conventional and proposed methods are presented in Table 3. The better value for each problem is highlighted in bold.

Table 3. Quantitative evaluation of the obtained solutions $\mathcal{P}'$ in the objective space based on HV (larger is better) and IGD (smaller is better); best values are shown in bold.

Metric	Method	MMF1	MMF2	MMF3	MMF4	MMF5	MMF6	MMF7	MMF8	LIRCMOP1	LIRCMOP2
HV	Conventional PSE [3]	0.7191	0.7169	0.7184	0.4419	0.7180	0.7182	0.7215	0.3439	0.2269	0.3608
	Proposed Multimodal PSE	0.7197	0.7222	0.7216	0.4469	0.7180	0.7182	0.7222	0.3484	0.2269	0.3624
IGD	Conventional PSE [3]	0.0080	0.0156	0.0059	0.0071	0.0094	0.0061	0.0035	0.0111	0.0805	0.0112
	Proposed Multimodal PSE	0.0038	0.0016	0.0020	0.0019	0.0061	0.0054	0.0020	0.0015	0.0805	0.0016

Across all problems, the proposed multimodal PSE achieves performance comparable to or better than the conventional PSE, indicating improved PF approximation accuracy.

For LIRCMOP1, all metrics are identical for both methods. As shown in Figure 4i, the proposed multimodal PSE failed to partition the known solutions into multiple clusters and treated them as a single cluster. As a result, the behavior becomes equivalent to the conventional PSE.

In contrast, for LIRCMOP2, the proposed method attains superior HV and IGD values. While the MMFs are low-dimensional problems ($D = 2$), the LIRCMOPs are high-dimensional ($D = 30$). The results for LIRCMOP2 suggest that even in high-dimensional settings, accurate PF estimation can be achieved if the PS distribution is not excessively complex.

For MMF5 and MMF6, both methods obtain similar HV values, whereas the proposed multimodal PSE achieves lower IGD values. This indicates that, although the overall objective-space coverage is comparable, the proposed method produces solutions that are more uniformly close to the PF.

7.4. Sensitivity Analysis of Window Size w and Peak-Count Threshold p

We analyze the sensitivity of the two main parameters of the proposed multimodal PSE: the window size w and the peak-count threshold p .

The HV, IGD, and IGDX results for different w and p are shown in Table 4. The variation in the performance metrics is more sensitive to changes in p than to changes in w , and all three metrics achieve their best values when $p = 0.4w$.

We next examine five representative parameter settings. The clustering results of the known solution set $\mathcal{P}$ are shown in Figure 9, the distribution of the estimated solution set $\mathcal{Q}$ in the variable space in Figure 10, and their distribution in the objective space in Figure 11.

For (c) $w = 10, p = 0.4w$, the known solutions are appropriately divided into four clusters according to distribution similarity. As a result, the estimated solution set $\mathcal{Q}$ complements $\mathcal{P}$ within each cluster, and the objective space exhibits a distribution with few dominated solutions.

In other parameter settings, a smaller p generally leads to a larger number of clusters. For (a) $w = 5, p = 0.2w$ and (b) $w = 10, p = 0.2w$, the increased number of clusters slightly reduces PS estimation accuracy in the variable space, but suppresses the generation of dominated solutions in the objective space.

In contrast, for larger p , such as (d) $w = 15, p = 0.6w$ and (f) $w = 15, p = 1.0w$, the number of clusters decreases, but solutions farther from the known set $\mathcal{P}$ are more frequently generated, leading to an increase in dominated solutions.

Overall, the peak-count threshold p has a stronger impact on performance than w . Smaller p suppresses dominated solutions but may reduce PS estimation accuracy in the variable space, whereas larger p tends to increase dominated solutions.

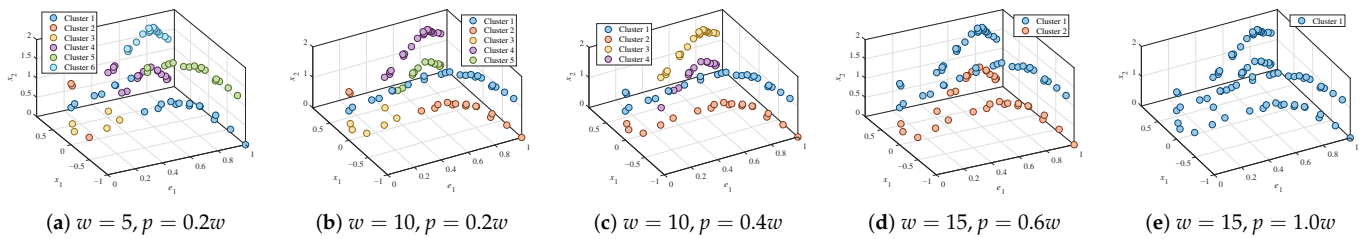


Figure 9. Clustering results by varying window size w and peak-count threshold p .

Table 4. Quantitative evaluation by varying the window size w and the peak-count threshold p based on HV (larger is better), IGD (smaller is better), and IGDX (smaller is better); best values are shown in bold.

(a) HV in the objective space					
		Window size w			
		5	10	15	20
Peak-count threshold p	0.2w	0.4464	0.4469	0.4469	0.4469
	0.4w	0.4469	0.4469	0.4469	0.4469
	0.6w	0.4419	0.4469	0.4420	0.4420
	0.8w	0.4419	0.4419	0.4419	0.4419
	1.0w	0.4419	0.4419	0.4419	0.4419
(b) IGD in the objective space					
		Window size w			
		5	10	15	20
Peak-count threshold p	0.2w	0.0027	0.0020	0.0020	0.0019
	0.4w	0.0019	0.0019	0.0019	0.0019
	0.6w	0.0071	0.0019	0.0062	0.0062
	0.8w	0.0071	0.0071	0.0071	0.0071
	1.0w	0.0071	0.0071	0.0071	0.0071
(c) IGDX in the variable space					
		Window size w			
		5	10	15	20
Peak-count threshold p	0.2w	0.0381	0.0371	0.0371	0.0339
	0.4w	0.0339	0.0339	0.0339	0.0339
	0.6w	0.5090	0.0339	0.0588	0.0588
	0.8w	0.5090	0.5090	0.5090	0.5090
	1.0w	0.5090	0.5090	0.5090	0.5090

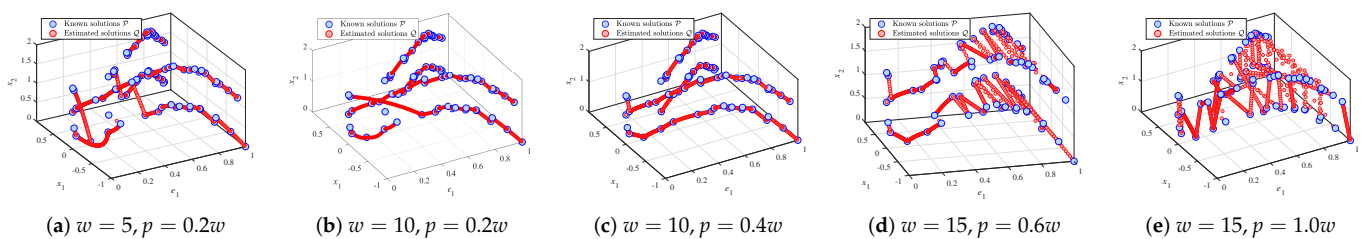


Figure 10. Solutions in the variable space by varying window size w and peak-count threshold p .

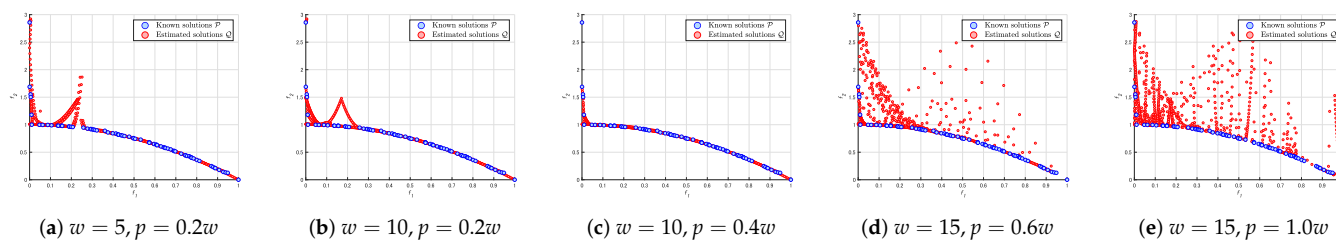


Figure 11. Solutions in the objective space by varying window size w and peak-count threshold p .

7.5. Sensitivity Analysis of the Amplitude Threshold Coefficient γ_{amp}

We analyze the sensitivity of the amplitude threshold coefficient γ_{amp} . Specifically, focusing on MMF4, we evaluated the proposed method with $\gamma_{amp} = 0.0, 0.05, 0.1, 0.15, 0.2, 0.55, 0.85$, and 1.0 and report the corresponding HV, IGD, and IGDX values in Table 5.

Table 5. HV (larger is better), IGD (smaller is better), and IGDX (smaller is better) by varying amplitude threshold coefficient γ_{amp} ; best values are shown in bold.

Metric	Amplitude Threshold Coefficient γ_{amp}							
	0	0.05	0.1	0.15	0.2	0.55	0.85	1
HV	0.4443	0.4469	0.4469	0.4469	0.4469	0.443	0.4419	0.4419
IGD	0.0034	0.0018	0.0019	0.0019	0.0019	0.0055	0.0071	0.0071
IGDX	0.0399	0.0348	0.0339	0.0339	0.0339	0.0425	0.0509	0.0509

Here, $\gamma_{amp} = 0.0$ means that even very small oscillations are regarded as oscillations, whereas $\gamma_{amp} = 1.0$ means that no oscillation is counted, resulting in all solutions being assigned to a single cluster, which corresponds to behavior equivalent to the conventional method.

The results show that HV is maximized when $\gamma_{amp} = 0.05$ – 0.2 , IGD is minimized at $\gamma_{amp} = 0.05$, and IGDX is minimized at $\gamma_{amp} = 0.1$ – 0.2 . Based on these results, $\gamma_{amp} = 0.1$ was selected as an appropriate setting.

In addition, Figure 12 shows the clustering results in the decision space for different γ_{amp} values, Figure 13 shows the distribution of generated solutions in the decision space, and Figure 14 shows the distribution in the objective space.

When γ_{amp} is too large, the number of clusters decreases, and the generated solutions (red points) tend to appear between known solution groups in the decision space, resulting in inferior objective-space distributions. On the other hand, when γ_{amp} is too small, excessive clustering occurs, reducing the number of known solutions within each cluster and degrading the accuracy of the PS model, which again leads to inferior generated solutions.

These observations further support the choice of $\gamma_{amp} = 0.1$.

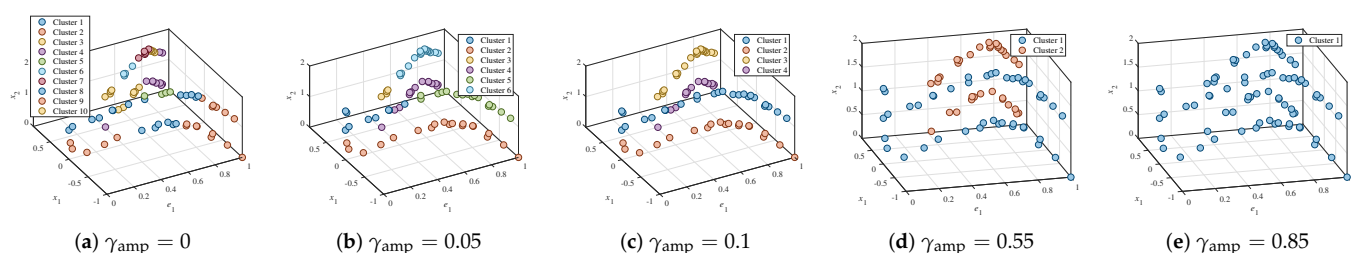


Figure 12. Clustering results by varying amplitude threshold coefficient γ_{amp} .

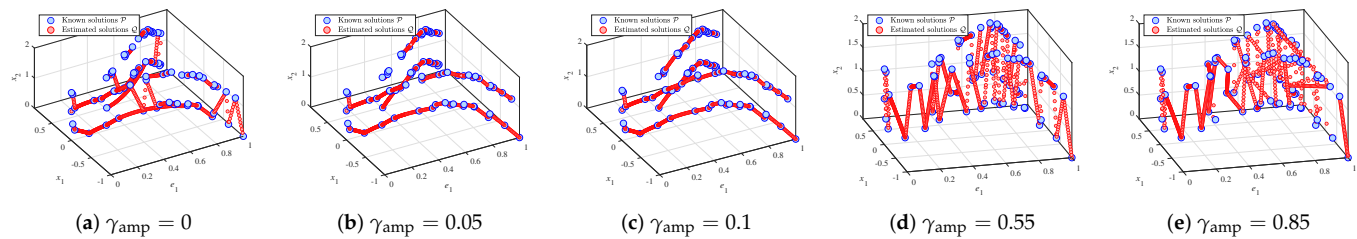


Figure 13. Solutions in the variable space by varying amplitude threshold coefficient γ_{amp} .

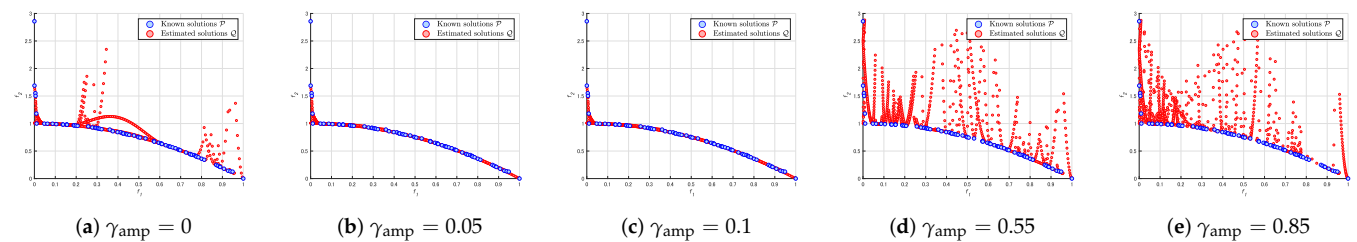


Figure 14. Solutions in the objective space by varying amplitude threshold coefficient γ_{amp} .

7.6. Computational Time

Since the proposed multimodal PSE method additionally performs clustering and constructs PS models for each identified cluster before generating solutions, it is important to evaluate the computational cost of these additional processes.

The execution time of the conventional PSE method and the proposed multimodal PSE method was also measured for all benchmark problems. The experimental environment was as follows: CPU: 13th Gen Intel(R) Core(TM) i9-13900KF, Memory: 128 GB, OS: Windows 11, and programming language: Matlab R2025b. Table 6 shows the average execution time in seconds over 50 independent runs.

The proposed method required a longer execution time than the conventional method for all benchmark problems. This is because it includes additional steps for clustering and constructing PS models for each identified cluster.

However, the average execution time of the proposed method remained below one second for all benchmark problems, indicating that the additional computational cost is limited. Since the proposed method is intended for practical problems where solution evaluation is computationally expensive, this additional overhead is considered sufficiently small.

Table 6. Average computational time of 50 runs in seconds; best values are shown in bold.

Metric	Method	MMF1	MMF2	MMF3	MMF4	MMF5	MMF6	MMF7	MMF8	LIRCMOP1	LIRCMOP2
Time (sec.)	Conventional PSE [3]	0.0067	0.0052	0.0048	0.0057	0.0061	0.0049	0.0054	0.0045	0.0058	0.0059
	Proposed Multimodal PSE	0.1245	0.0959	0.0740	0.0879	0.0886	0.0846	0.1488	0.0496	0.5315	0.6792

8. Conclusions

This paper addressed the limitation of conventional Pareto set estimation (PSE) methods in multimodal multi-objective optimization problems, where multiple distinct decision vectors correspond to the same objective vector. Because conventional PSE implicitly assumes a one-to-one mapping between objective directions and decision vectors, it tends to collapse multiple decision-space modes into a single estimate. To overcome this issue, we proposed a clustering-based PSE framework with oscillation-aware hierarchical clustering. By introducing an amplitude threshold in oscillation detection and constructing cluster-wise response surface models, the proposed method decomposes a multimodal Pareto set into locally unimodal components. This enables multiple decision-space solutions to be estimated for the same objective-space direction. Experimental results on MMF1–8

and LIRCMOP1–2 demonstrated that the proposed method achieves equal or better performance than the conventional method in terms of HV and IGD, while also improving decision-space approximation as measured by IGD_X in most cases. The results indicate that the proposed approach effectively preserves multimodal structures and enhances PF approximation accuracy. The generated clusters play an important role in separating distinct local structures in the variable space before inverse modeling. By approximating each cluster independently, the proposed method avoids merging disconnected solution branches and enables multimodal decision-space structures to be reconstructed more accurately. A sensitivity analysis further showed that the peak-count threshold plays a dominant role in performance and that appropriate parameter settings lead to stable and accurate estimation.

Although the proposed method may struggle when the distribution of solutions is highly complex in high-dimensional spaces, the overall results suggest that cluster-wise inverse modeling is a promising direction for multimodal Pareto set estimation. Future work includes developing more robust clustering strategies for high-dimensional problems and extending the framework to more complex constrained scenarios. Also, it is necessary to investigate how the performance of the proposed method changes when the number of known solutions and search solutions varies, in order to evaluate its robustness under different sampling conditions. In addition, although $\gamma_{\text{amp}} = 0.1$ was effective in the present experiments, its appropriate value may depend on the distribution of known solution groups in the decision space and the number of solutions contained in each cluster. Therefore, adaptive determination of γ_{amp} should also be investigated in future work. Furthermore, visualization methods for higher-dimensional decision spaces, such as PCA or UMAP, will be investigated to improve interpretability for more complex multimodal problems.

Author Contributions: Conceptualization, Y.S., Y.O. and H.S.; methodology, Y.S., Y.O. and H.S.; software, Y.S.; investigation, Y.S.; resources, Y.O.; data curation, Y.S.; writing—original draft preparation, Y.S.; writing—review and editing, Y.O. and H.S.; supervision, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Mitsubishi Electric Corporation through a collaborative research agreement with The University of Electro-Communications.

Data Availability Statement: Data may be available from the corresponding author upon reasonable request and with permission of the collaborating organization due to restrictions related to the collaborative research agreement.

Conflicts of Interest: Yoshihiro Ohta is employed by Mitsubishi Electric Corporation. Mitsubishi Electric Corporation contributed resources and participated in technical discussions under a collaborative research agreement. All other authors have no conflicts of interests to declare.

Appendix A. Notations and Abbreviations

Table A1. Main notations used in this paper.

Symbol	Description
$\mathcal{P}$	Set of known solutions used for model construction
N	Size of the known solution set $\mathcal{P}$
$\mathcal{Q}$	Set of solutions generated by PS estimation
$\mathcal{P}'$	Final estimated population obtained from $\mathcal{P} \cup \mathcal{Q}$
$\mathcal{C}$	Set of clusters obtained by hierarchical clustering
C_k	k -th cluster in the decision space

Table A1. *Cont.*

Symbol	Description
$\mathcal{E}$	Direction vectors calculated from known solutions
$\hat{\mathcal{E}}$	Large set of uniformly generated direction vectors
$\hat{\mathcal{E}}'$	Prioritized direction vectors used for estimation
x_{C_k} -model	Response surface model trained for cluster C_k
l_i	Cluster label assigned to solution i
w	Window size for oscillation detection
p	Peak-count threshold for oscillation detection
amp_j	Amplitude threshold for the j -th variable
γ_{amp}	Coefficient determining amp_j relative to variable range

Table A2. Main abbreviations used in this paper.

Abbreviation	Full Term
MOP	Multi-objective optimization problem
MMOP	Multimodal multi-objective optimization problem
PS	Pareto set
PF	Pareto front
PSE	Pareto set estimation
HV	Hypervolume
IGD	Inverted generational distance
IGDX	Inverted generational distance in the decision space
PFE	Pareto front estimation
DACE	Design and analysis of computer experiments
SMOA	Supervised multi-objective optimization algorithm
SVR	Support vector regression
PCA	Principal component analysis
UMAP	Uniform manifold approximation and projection
CEC	Congress on Evolutionary Computation
MMF	Multimodal multi-objective test function
LIRCMOP	Large infeasible region constrained multi-objective problem

References

1. Deb, K. *Multi-Objective Optimization Using Evolutionary Algorithms*; Wiley: Chichester Weinheim, UK, 2004.
2. Coello, C.A.C.; Lamont, G.B.; Veldhuizen, D.A.V. *Evolutionary Algorithms For Solving Multi-Objective Problems*; Springer: Boston, MA, USA, 2007.
3. Takagi, T.; Takadama, K.; Sato, H. Supervised Multi-Objective Optimization Algorithm Using Estimation. In *Proceedings of the 2022 IEEE Congress on Evolutionary Computation (CEC)*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–8.
4. Hartikainen, M.; Miettinen, K.; Wiecek, M.M. PAINT: Pareto Front Interpolation For Nonlinear Multiobjective Optimization. *Comput. Optim. Appl.* **2012**, *52*, 845–867. [[CrossRef](#)]
5. Yue, C.; Qu, B.; Liang, J. A Multiobjective Particle Swarm Optimizer Using Ring Topology for Solving Multimodal Multiobjective Problems. *IEEE Trans. Evol. Comput.* **2018**, *22*, 805–817. [[CrossRef](#)]
6. Fan, Z.; Li, W.; Cai, X.; Huang, H.; Fang, Y.; You, Y.; Mo, J.; Wei, C.; Goodman, E. An Improved Epsilon Constraint-Handling Method in MOEA/D for CMOPs with Large Infeasible Regions. *Soft Comput.* **2019**, *23*, 12491–12510. [[CrossRef](#)]
7. Liang, J.J.; Yue, C.T.; Qu, B.Y. Multimodal Multi-Objective Optimization: A Preliminary Study. In *Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC)*; IEEE: Piscataway, NJ, USA, 2016; pp. 2454–2461.
8. Liang, J.J.; Qu, B.Y.; Gong, D.W.; Yue, C.T. *Problem Definitions and Evaluation Criteria for the CEC 2019 Special Session on Multimodal Multiobjective Optimization*; Computational Intelligence Laboratory, Zhengzhou University: Zhengzhou, China, 2019.
9. Zitzler, E.; Thiele, L. Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach. *IEEE Trans. Evol. Comput.* **1999**, *3*, 257–271. [[CrossRef](#)]
10. Czyżżak, P.; Jaszkiewicz, A. Pareto Simulated Annealing—A Metaheuristic Technique for Multiple-Objective Combinatorial Optimization. *J. Multi-Criteria Decis. Anal.* **1998**, *7*, 34–47. [[CrossRef](#)]
11. Sato, H.; Aguirre, H.E.; Tanaka, K. Local Dominance Using Polar Coordinates to Enhance Multiobjective Evolutionary Algorithms. In *Proceedings of the 2004 IEEE Congress on Evolutionary Computation (CEC)*; IEEE: Piscataway, NJ, USA, 2004; Volume 1, pp. 188–195.

12. Coello Coello, C.A.; Cruz Cortés, N. Solving Multiobjective Optimization Problems Using an Artificial Immune System. *Genet. Program. Evolvable Mach.* **2005**, *6*, 163–190. [[CrossRef](#)]
13. Zhou, A.; Zhang, Q.; Jin, Y. Approximating the Set of Pareto-Optimal Solutions in Both the Decision and Objective Spaces by an Estimation of Distribution Algorithm. *IEEE Trans. Evol. Comput.* **2009**, *13*, 1167–1189. [[CrossRef](#)]
14. Suzumura, Y.; Ohta, Y.; Sato, H. Clustering-Based Pareto Set Estimation for Multimodal Multi-Objective Optimization. In Proceedings of the The 26th International Symposium on Advanced Intelligent Systems (ISIS2025), Cheongju, Republic of Korea, 6–9 November 2025; pp. 354–359.
15. Hartikainen, M.; Miettinen, K.; Wiecek, M.M. Decision Making on Pareto Front Approximations with Inherent Nondominance. In *New State of MCDM in the 21st Century*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 648, pp. 35–45.
16. Bhattacharjee, K.S.; Singh, H.K.; Ray, T. An Approach to Generate Comprehensive Piecewise Linear Interpolation of Pareto Outcomes to Aid Decision Making. *J. Glob. Optim.* **2017**, *68*, 71–93. [[CrossRef](#)]
17. Bhattacharjee, K.S.; Singh, H.K.; Ray, T. Enhanced Pareto Interpolation Method to Aid Decision Making for Discontinuous Pareto Optimal Fronts. In *AI 2017: Advances in Artificial Intelligence*; Springer International Publishing: Cham, Switzerland, 2017; Volume 10400, pp. 93–105.
18. Kobayashi, K.; Hamada, N.; Sannai, A.; Tanaka, A.; Bannai, K.; Sugiyama, M. Bézier Simplex Fitting: Describing Pareto Fronts of Simplicial Problems with Small Samples in Multi-Objective Optimization. *arXiv* **2018**, arXiv:1812.05222v1. [[CrossRef](#)]
19. Hernández-Díaz, A.G.; Santana-Quintero, L.V.; Coello Coello, C.A.; Molina, J. Pareto-Adaptive ϵ -Dominance. *Evol. Comput.* **2007**, *15*, 493–517. [[CrossRef](#)]
20. Zapotecas Martínez, S.; Sosa Hernández, V.A.; Aguirre, H.; Tanaka, K.; Coello Coello, C.A. Using A Family of Curves to Approximate the Pareto Front of A Multi-Objective Optimization Problem. In *Parallel Problem Solving from Nature—PPSN XIII*; Springer International Publishing: Cham, Switzerland, 2014; Volume 8672, pp. 682–691.
21. Tian, Y.; Zhang, X.; Cheng, R.; He, C.; Jin, Y. Guiding Evolutionary Multiobjective Optimization with Generic Front Modeling. *IEEE Trans. Cybern.* **2020**, *50*, 1106–1119. [[CrossRef](#)]
22. Tian, Y.; Si, L.; Zhang, X.; Tan, K.C.; Jin, Y. Local Model-Based Pareto Front Estimation for Multiobjective Optimization. *IEEE Trans. Syst. Man, Cybern. Syst.* **2023**, *53*, 623–634. [[CrossRef](#)]
23. Takagi, T.; Takadama, K.; Sato, H. Pareto Front Estimation Using Unit Hyperplane. In *Evolutionary Multi-Criterion Optimization*; Springer International Publishing: Cham, Switzerland, 2021; Volume 12654, pp. 126–138.
24. Takagi, T.; Takadama, K.; Sato, H. Directional Pareto Front and Its Estimation to Encourage Multi-Objective Decision-Making. *IEEE Access* **2023**, *11*, 20619–20634. [[CrossRef](#)]
25. Sacks, J.; Welch, W.J.; Mitchell, T.J.; Wynn, H.P. Design and Analysis of Computer Experiments. *Stat. Sci.* **1989**, *4*, 409–423. [[CrossRef](#)]
26. Sacks, J.; Schiller, S.B.; Welch, W.J. Designs for Computer Experiments. *Technometrics* **1989**, *31*, 41–47. [[CrossRef](#)]
27. Chen, H.; Loepky, J.L.; Welch, W.J. Flexible Correlation Structure for Accurate Prediction and Uncertainty Quantification in Bayesian Gaussian Process Emulation of A Computer Model. *SIAM/ASA J. Uncertain. Quantif.* **2017**, *5*, 598–620. [[CrossRef](#)]
28. Farias, L.R.C.; Araujo, A.F.R. IM-MOEA/D: An Inverse Modeling Multi-Objective Evolutionary Algorithm Based on Decomposition. In Proceedings of the 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, 17–20 October 2021; pp. 462–467.
29. Elaziz, G.M.A.; Abouelseoud, Y.; Kamel, S.H. An Inverse Modeling Multi-Objective Optimization Technique Based on Incremental Learning And Fuzzy Clustering. *IEEE ACCESS* **2025**, *13*, 128337–128359. [[CrossRef](#)]
30. Takagi, T.; Takadama, K.; Sato, H. Pareto Front Upconvert by Iterative Estimation Modeling and Solution Sampling. In *Evolutionary Multi-Criterion Optimization*; Springer Nature Switzerland: Cham, Switzerland, 2023; Volume 13970, pp. 218–230.
31. Lin, Q.; Lin, W.; Zhu, Z.; Gong, M.; Li, J.; Coello, C.A.C. Multimodal Multiobjective Evolutionary Optimization with Dual Clustering in Decision and Objective Spaces. *IEEE Trans. Evol. Comput.* **2021**, *25*, 130–144. [[CrossRef](#)]
32. Liang, J.; Qiao, K.; Yue, C.; Yu, K.; Qu, B.; Xu, R.; Li, Z.; Hu, Y. A Clustering-Based Differential Evolution Algorithm for Solving Multimodal Multi-Objective Optimization Problems. *Swarm Evol. Comput.* **2021**, *60*, 100788. [[CrossRef](#)]
33. Deng, W.; Mo, Y.; Deng, L. A Multimodal Multi-Objective Coati Optimization Algorithm Based on Spectral Clustering. *Symmetry* **2024**, *16*, 1474. [[CrossRef](#)]
34. Wu, X.; Ming, F.; Gong, W. A Multi-Task Framework for Solving Multimodal Multiobjective Optimization Problems. In *Neural Information Processing*; Springer Nature: Singapore, 2024; Volume 14449, pp. 300–313.
35. Tanabe, R.; Ishibuchi, H. A Review of Evolutionary Multimodal Multiobjective Optimization. *IEEE Trans. Evol. Comput.* **2020**, *24*, 193–200. [[CrossRef](#)]
36. Pan, L.; Li, L.; Cheng, R.; He, C.; Tan, K.C. Manifold Learning-Inspired Mating Restriction for Evolutionary Multiobjective Optimization with Complicated Pareto Sets. *IEEE Trans. Cybern.* **2021**, *51*, 3325–3337. [[CrossRef](#)] [[PubMed](#)]
37. Ran, X.; Xi, Y.; Lu, Y.; Wang, X.; Lu, Z. Comprehensive Survey on Hierarchical Clustering Algorithms and the Recent Developments. *Artif. Intell. Rev.* **2023**, *56*, 8219–8264. [[CrossRef](#)]

38. Wenhua Li.; Xingyi Yao.; Tao Zhang.; Rui Wang.; Ling Wang. Hierarchy Ranking Method for Multimodal Multiobjective Optimization with Local Pareto Fronts. *IEEE Trans. Evol. Comput.* **2023**, *27*, 98–110. [[CrossRef](#)]
39. Zou, J.; Sun, R.; Yang, S.; Zheng, J. A Dual-Population Algorithm Based on Alternative Evolution and Degeneration for Solving Constrained Multi-Objective Optimization Problems. *Inf. Sci.* **2021**, *579*, 89–102. [[CrossRef](#)]
40. Liang, J.; Qiao, K.; Yu, K.; Qu, B.; Yue, C.; Guo, W.; Wang, L. Utilizing the Relationship Between Unconstrained and Constrained Pareto Fronts for Constrained Multiobjective Optimization. *IEEE Trans. Cybern.* **2023**, *53*, 3873–3886. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.