Scale-Adaptive Surrogate Model Management for Expensive Optimization

Xing-Yu Wang
College of Artificial Intelligence
Nankai University
Tianjin, China

Jian-Yu Li
College of Artificial Intelligence
Nankai University
Tianjin, China
jianyulics@foxmail.com

Zhi-Hui Zhan
College of Artificial Intelligence
Nankai University
Tianjin, China

Jun Zhang
Hanyang University,
Ansan, South Korea

***Abstract*—Data-driven evolutionary algorithms (DDEAs) have been widely used for solving expensive optimization problems by replacing expensive fitness evaluations with surrogate models constructed from limited evaluated data. However, although various surrogate construction and selection strategies have been investigated, the evaluation scale of surrogate-assisted fitness estimation is usually predefined and remains unchanged throughout the evolutionary process. Such a static evaluation scale may be inefficient or even misleading, as the evaluation requirements vary across different evolutionary stages. To alleviate this issue, this paper proposes a scale-adaptive model management (SAMM)-based DDEA, named SAMM-DDEA, together with three contributions. First, a SAMM strategy is proposed to dynamically adjust the evaluation scale of surrogate-assisted fitness estimation according to evolutionary feedback. In addition, a Full-Coverage Sampling (FCS) strategy is proposed to enhance ensemble surrogate diversity under limited data conditions, so as to enhance algorithm performance. Third, the SAMM-DDEA is developed and validated on widely-used expensive optimization problems with 10 to 100 dimensions, demonstrating superior or highly competitive performance compared with several state-of-the-art DDEAs.**

Keywords—Expensive optimization problem, scale-adaptive model management, surrogate modeling, evolutionary computation, model management.

I. INTRODUCTION

There has been a surge in interest from the research community in expensive optimization problems (EOPs) in recent years [1]. Broadly speaking, EOPs are characterized by optimization processes in which the search is driven primarily by data rather than explicit analytical formulations. In many real-world scenarios, analytical objective functions or accurate mathematical models are either unavailable or prohibitively difficult to construct [2]. Under such circumstances, traditional mathematical optimization methods that rely on precise models, as well as conventional evolutionary algorithms (EAs) that depend on repeated expensive evaluations, become inefficient or even impractical.

To overcome the challenges posed by unknown objective models and high evaluation costs, a class of data-driven evolutionary algorithms (DDEAs) has been proposed [3]. These methods exploit previously evaluated samples to construct surrogate models that provide inexpensive approximations of solution quality across the search space [4]. By circumventing the need for explicit analytical modeling and significantly reducing evaluation expenses, DDEAs have demonstrated strong effectiveness in addressing EOPs. As a result, substantial research efforts have been devoted to both the theoretical advancement and practical deployment of DDEAs.

The development of DDEAs has mainly followed two key paths. The first centers on improving the quality and availability of training data, as these factors directly impact surrogate model accuracy and robustness. A particularly critical challenge in this line of research is the limited amount of evaluated data. Empirical studies consistently indicate that surrogate accuracy tends to improve as the volume of training data increases [5]. Consequently, considerable effort has been devoted to expanding offline datasets and fully exploiting existing samples. Among these, representative methods, including local smoothing [7] and artificial data generation [8]-[12], seek to augment the training set without the need for further expensive evaluations.

The second major research direction concentrates on improving surrogate modeling techniques themselves, coupled with the formulation of effective model management strategies (MMS). A variety of surrogate models have been investigated, such as Kriging models [13], radial basis function neural networks (RBFNNs) [14], [15], and artificial neural networks [16]. More recently, ensemble learning has gained prominence as a means of enhancing surrogate performance by combining multiple base learners [17], [18]. By leveraging the complementary strengths of different models, ensemble-based approaches can achieve higher predictive accuracy and improved generalization capability. Through the coordinated management of model utilization and update mechanisms, MMS aims to fully exploit scarce data resources and improve the robustness and overall optimization performance of DDEAs.

Although numerous efforts have been made to MMS, most existing MMS implicitly assume a fixed evaluation scale during

This work was supported in part by the National Natutural Science Foundation of China (NSFC) under Grant 62406152 and Grant U23B2039, in part by the Natural Science Foundation of Tianjin under Grant 24JCQNJC02100, in part by the Natural Science Foudation of Tianjin Key Program uder Grant 25JCZDJC01040, in part by the Fundamental Research Funds for the Central Universities, Nankai University (63261167), in part by the Tianjin Top Scientist Studio Project under Grant 24JRRCRC00030, in part by the Tianjin Belt and Road Joint Laboratory under Grant 24PTLYHZ00250, in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.RS-2025-00555463, and RS-2025-25456394), and in part by the China Higher Education Institution Industry-University-Research Innovation Fund under Grant 2025XJS008. Corresponding authors: Jian-Yu Li and Zhi-Hui Zhan

the evolutionary process. However, this assumption may be inefficient for solving EOPs. Different evolutionary stages demand different fidelity of fitness evaluation (FE): coarse and diverse assessments are often sufficient in early stages to promote exploration and diversity, whereas later stages require finer, more stable evaluations to distinguish high-quality solutions. A fixed scale can lead to inefficient surrogate use, biased guidance, or premature convergence, especially under limited or noisy data. Despite its importance, adaptive regulation of the evaluation scale remains underexplored in DDEAs. Current MMS mainly emphasizes surrogate accuracy or diversity via model construction, selection, or updating, while the adaptive use of surrogates for FE across evolutionary stages has received little attention.

In recent years, scale-adaptive fitness evaluation mechanisms have been studied for improving the efficiency of expensive optimization [19], [20]. These approaches mainly adjust the fidelity or frequency of FEs to reduce computational cost. However, they are not designed from a SMM perspective and do not explicitly consider how the evaluation scale of ensemble-based surrogate-assisted FE should be regulated in conjunction with surrogate utilization and update mechanisms. This gap motivates the development of a principled model management strategy that dynamically adjusts the surrogate.

Given these considerations, this paper proposes a scale-adaptive model management-based DDEA, named SAMM-DDEA. The proposed SAMM-DDEA explicitly introduces the concept of evaluation scale into surrogate-assisted fitness estimation and dynamically regulates surrogate utilization in response to evolutionary feedback. By adaptively adjusting the number of surrogate models involved in ensemble-based evaluation, SAMM enables the evolutionary process to employ coarse evaluation scales during exploration and progressively finer evaluation scales as convergence proceeds, thereby better aligning surrogate-assisted evaluation with the evolving requirements of different evolutionary stages. This paper has the following contributions.

First, a scale-adaptive model management strategy is proposed to dynamically regulate the evaluation scale by adjusting the number of surrogate models involved in ensemble-based fitness estimation according to evolutionary feedback, enabling coarse-grained evaluations in early evolutionary stages and progressively finer-grained evaluations as the search converges.

Second, a full-coverage sampling (FCS) strategy is introduced to bolster the diversity and robustness of ensemble surrogate models under limited data conditions. By ensuring that different surrogate models are trained on complementary subsets whose union fully covers the available evaluated data, FCS improves surrogate generalization and mitigates overfitting in data-driven optimization.

Third, with the SAMM and FCS, the SAMM-DDEA is presented, and its effectiveness is validated through extensive experimental studies on benchmark expensive optimization problems. The experimental results demonstrate that SAMM-DDEA attains performance superior to, or at least highly competitive with, several state-of-the-art (SOTA) DDEAs, especially in high-dimensional optimization problems. The code will be provided online if the paper is accepted.

II. Related Work

To date, numerous approaches have been proposed to improve the performance of DDEAs. As this study concentrates on offline DDEAs, this section briefly reviews representative offline DDEA methods, emphasizing their core characteristics and inherent limitations.

Offline DDEAs rely solely on a fixed set of pre-evaluated samples to explore the search space. Since no additional expensive evaluations are allowed, data quality is particularly important. For example, Lin *et al.* [21] analyzed the influence of noise on DDEAs and demonstrated that increasing noise levels can substantially deteriorate optimization performance. To address common data-related issues such as noise, imbalance, and incompleteness, a variety of data preprocessing strategies have been introduced [22]. As a representative application, local regression techniques were adopted in blast furnace optimization to alleviate the adverse effects of noisy observations [4].

In scenarios when available data are insufficient to support reliable surrogate construction, artificial data generation provides an effective alternative. Li *et al.* [12] proposed a localized data generation strategy to enrich sparse datasets, leading to notable improvements in optimization performance. Similarly, data perturbation techniques [8] have been adopted to generate multiple data subsets, thereby facilitating the construction of more robust surrogate models. TT-DDEA [18] further extends this idea by leveraging semi-supervised learning to expand the training set through labeling high-confidence unlabeled samples.

Beyond data-centric approaches, surrogate modeling performance in offline DDEAs can also be enhanced through advanced learning strategies. These techniques aim to improve both predictive accuracy and model robustness. For example, DDEA-SE [17] adopts a bagging framework to construct a large pool of RBFNNs. The fitness approximator is then derived from a selective combination of models within this pool. In contrast, BDDEA-LDG [12] employs a boosting strategy that incrementally refines surrogate accuracy through sequential training. Yu *et al.* [23] proposed a dynamic ensemble approach that adaptively selects the most suitable surrogate model in response to the evolving optimization process.

Other studies emphasize increasing surrogate diversity at the model level. As illustrated by SRK-DDEA [6], one approach involves utilizing multiple RBFNNs that incorporate varied radial basis functions, followed by candidate solution evaluation via a surrogate-assisted ranking mechanism. Additionally, alternative learning paradigms have been explored in recent years. Huang *et al.* [24] studied contrastive learning to capture relational information among individuals, while Sun *et al.* [25] investigated symbolic regression as a surrogate modeling technique within data-driven optimization.

III. The SAMM-DDEA

A. Scale-Adaptive Model Management Framework

Building on the proven capability of ensemble surrogate models in handling EOPs, SAMM-DDEA employs a multi-

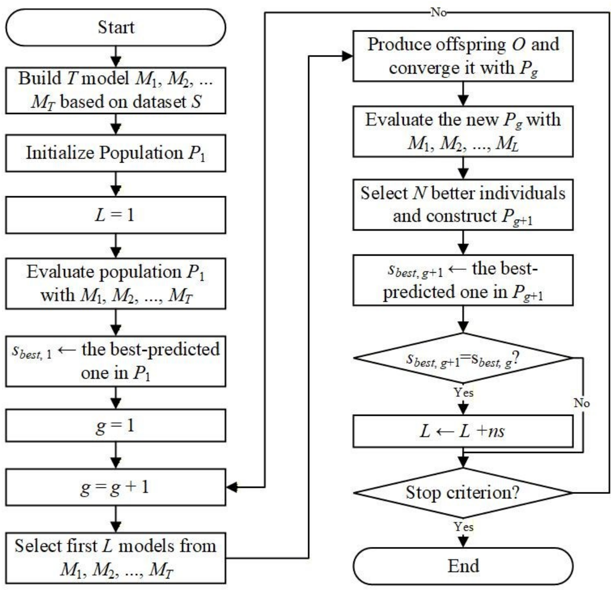

Fig. 1. The framework of SAMM-DDEA.

model ensemble framework to conduct performance prediction. Unlike existing ensemble-based approaches, SAMM-DDEA uses the SAMM mechanism to control model selection, achieving a better balance between accuracy and efficiency. The SAMM framework is illustrated in Fig. 1. Its core idea is to adaptively adjust the evaluation granularity of FEs to meet the requirements of different evolutionary stages. In ensemble model management, using a larger number of surrogate models generally results in more accurate and stable final evaluations. Therefore, SAMM treats the quantity of surrogate models incorporated in the ensemble as a variable for scale adaptation.

Specifically, when a finer evaluation scale is needed in the evolutionary process, SAMM generates L novel base surrogate models and fuses them with the existing ensemble components, thereby renewing the ensemble surrogate model. In this work, a commonly adopted averaging strategy from the literature is employed for model integration. That is, the fitness values predicted by all base models are first computed and then averaged to obtain the final FE of the ensemble surrogate model.

B. Full-Coverage Sampling

FCS is designed to fully exploit the available data and enhance the diversity of base models in the ensemble. Unlike the commonly used random sampling strategies in existing literature, the union of all sampling subsets in FCS is guaranteed to cover the entire dataset.

The pseudocode of FCS is shown in Algorithm 1. Suppose that K base surrogate models are to be trained. When training the $(2k-1)$-th and $2k$-th models, the dataset is first randomly permuted and then partitioned into two equal-sized subsets according to the shuffled order. The first half of the data is allocated for training the $(2k-1)$-th model, whereas the second half is reserved for training the $2k$-th model.

Through this procedure, all data samples are guaranteed to be used, while different models are trained on different data subsets, thereby increasing the diversity of data distributions among the base models and improving the ensemble's robustness.

Algorithm 1: Full-Coverage Sampling

Input: O_o - the evaluated dataset;
T-the number of models to be built.
Output: $M_1, M_2, \ldots, M_T$-the T built models;
Begin
For k =1 to $T/2$ **Do**
$t \leftarrow 2\times k$-1;
Randomly shuffle all the data in O_o;
$O_t \leftarrow$ the first 50% data in O_o;
$O_{t+1} \leftarrow$ the last 50% data in O_o;
Build M_t based on O_t;
Build M_{t+1} based on O_{t+1};
End For
End

Algorithm 2: SAMM-DDEA

Input: O_o - the evaluated dataset;
T - the number of models to be built;
ns -the number of selected data;
Output: $s_{b,g}$ - the best-found solution.
Begin
// Initialization
generation ←1;
Build T models M1, M2, …, MT; //Algorithm 1
Initialize population P;
L=1;
Evaluate population P with $M_1, \ldots, M_L$;
While the stop criterion is false **Do**
generation ←*generation* + 1;
$M_1, \ldots, M_L \leftarrow$ Select first L models from $M_1, M_2, \ldots, M_T$;
// Offspring Production
$O\leftarrow$ Carry out crossover to obtain new individuals from P;
$P \leftarrow P \cup O$;
$O \leftarrow$Conduct mutation to obtain new individuals from P;
$P \leftarrow P \cup O$;
Evaluate individuals in P based on $M_1, \ldots, M_L$;
// Selection
$P \leftarrow$ Select N better individuals from P;
$s_{b,g} \leftarrow$ the best-predicted one in P;
// SAMM
If $s_{b,g}$ is the same as $s_{b,g-1}$ **Do**
$L \leftarrow L + ns$;
End If
End While
End

C. The Complete SAMM-DDEA

Algorithm 2 presents SAMM-DDEA's framework, while Fig. 1 illustrates its flowchart. The algorithm takes as input an evaluated dataset O_o, the number of surrogate models T, and the number of selected models ns, and outputs the best-found solution $s_{b,g}$. At the initialization stage, the generation counter is set to one, and T surrogate models are constructed using Algorithm 1. Then, an initial population P is generated, and the initial ensemble scale is set as L=1. The predicted fitness is the average value of the prediction of L models. Mathematically, the fitness prediction of a solution s, say $fitness(s)$, is computed as:

$$fitness(s) = \frac{1}{L}\sum_{l=1}^{L} M_l(s) \quad (1)$$

where $M_i(s)$ is the output of M_i on x, and $G(s)$ is used as the predicted fitness for comparison and selection.

During evolution, SAMM-DDEA generates offspring via crossover and mutation, which are combined with the parent population to form a candidate pool. All individuals in the pool are evaluated using an ensemble surrogate composed of the first

L models. The best N individuals form the next generation, and the individual with the highest predicted fitness is recorded as the incumbent best solution $s_{b,g}$.

To adaptively adjust the evaluation scale, the SAMM mechanism is employed. If the best solution in the previous generation $s_{b,g-1}$ remains unchanged compared with the current one $s_{b,g}$, the number of surrogate models used for ensemble evaluation is increased by ns. This adaptive adjustment enables the algorithm to gradually refine the evaluation granularity as the evolution converges.

The above procedure is repeated until the stopping criterion is satisfied, and the final best-found solution is returned.

IV. Experimental Results

A. Experimental Design

To comprehensively evaluate SAMM DDEA, five widely used benchmark problems, namely Ellipsoid, Rosenbrock, Ackley, Griewank, and Rastrigin (denoted P1 to P5), are tested under varying dimensions $D \in \{10, 30, 50, 100\}$. Following standard practice, an initial dataset of $11 \times D$ samples is uniformly generated using Latin Hypercube Sampling (LHS) [26] and used to construct surrogate models. A key constraint in this data-driven optimization setting is that no DDEA may conduct any real evaluations beyond the initial $11 \times D$ samples, ensuring a rigorous assessment under severe data scarcity.

Each algorithm undergoes 25 independent runs per problem instance, with performance measured by averaged outcomes. Statistical significance is evaluated using the Wilcoxon rank sum test at α=0.05, and comparative results are summarized using the notation "+" (SAMM-DDEA significantly better), "≈" (no significant difference), and "-" (SAMM-DDEA significantly worse), allowing for rapid visual assessment.

To rigorously evaluate the effectiveness of SAMM-DDEA, four SOTA DDEAs, DDEA-PES [8], BDDEA-LDG [12], DDEA-SE [17], and TT-DDEA [18] are selected for comparison. The first three methods share a common foundation of using ensemble surrogate models for prediction, yet are distinguished by their distinct model selection strategies. Consequently, they provide an appropriate comparison for assessing SAMM-DDEA, which introduces a novel SAMM-based ensemble model selection mechanism. Differently, the fourth algorithm is based on tri-training, which help to assess the effectiveness of the SAMM-DDEA.

B. Comparisons with DDEAs

The comparison results between the SAMM-DDEA and widely-used DDEAs are given in Table I.

From Table I, it can be observed that the proposed SAMM-DDEA consistently demonstrates strong competitiveness against four representative DDEAs across different dimensional settings. For low-dimensional problems (D=10), SAMM-DDEA delivers superior or at least comparable performance on the majority of test cases. In particular, it significantly outperforms DDEA-SE, BDDEA-LDG, and DDEA-PES on P1 and P3, while showing comparable results on P2, P4, and P5. Compared with TT-DDEA, SAMM-DDEA achieves two significantly better results, one similar result, and two significantly worse results, indicating a competitive but more stable performance.

At the higher dimension of D=30, the performance edge of SAMM-DDEA is further accentuated. It achieves better results than BDDEA-LDG on all five problems and outperforms DDEA-PES on four of them. Although TT-DDEA performs well on P5, SAMM-DDEA still maintains robust performance on the remaining problems, demonstrating strong advantages in medium-dimensional search spaces.

For the higher-dimensional cases (D=50 and D=100), the advantages of SAMM-DDEA are further amplified. For 50-dimensional problems, SAMM-DDEA markedly surpasses

TABLE I
Comparing the SAMM-DDEA and Some Popular Algorithms

D	ID	SAMM-DDEA	DDEA-SE	BDDEA-LDG	DDEA-PES	TT-DDEA
10	P1	6.96E-01 ± 2.97E-01	9.94E-01 ± 4.62E-01 (+)	1.11E+00 ± 3.98E-01 (+)	1.36E+00 ± 6.43E-01 (+)	1.09E+00 ± 8.97E-01 (≈)
	P2	2.61E+01 ± 5.11E+00	2.80E+01 ± 5.23E+00 (≈)	3.44E+01 ± 7.35E+00 (+)	2.99E+01 ± 7.61E+00 (≈)	4.22E+00 ± 1.82E+00 (-)
	P3	5.16E+00 ± 1.19E+00	5.99E+00 ± 7.80E-01 (+)	6.67E+00 ± 6.35E-01 (+)	6.50E+00 ± 1.42E+00 (+)	8.68E+00 ± 1.61E+00 (+)
	P4	1.29E+00 ± 1.63E-01	1.30E+00 ± 1.37E-01 (≈)	1.26E+00 ± 1.21E-01 (≈)	1.33E+00 ± 1.30E-01 (≈)	6.70E+01 ± 9.61E+00 (+)
	P5	5.59E+01 ± 2.01E+01	6.08E+01 ± 1.80E+01 (≈)	6.94E+01 ± 2.64E+01 (+)	5.87E+01 ± 1.76E+01 (≈)	3.03E+01 ± 5.89E+00 (-)
+/≈/-		NA	2/3/0	4/1/0	2/3/0	2/1/2
30	P1	3.82E+00 ± 1.24E+00	4.09E+00 ± 1.35E+00 (≈)	7.02E+00 ± 2.18E+00 (+)	5.31E+00 ± 1.37E+00 (+)	4.17E+00 ± 1.56E+00 (≈)
	P2	5.73E+01 ± 4.18E+00	5.71E+01 ± 4.72E+00 (≈)	6.96E+01 ± 8.10E+00 (+)	7.00E+01 ± 8.49E+00 (+)	6.57E+01 ± 1.34E+01 (+)
	P3	4.60E+00 ± 3.87E-01	4.85E+00 ± 5.16E-01 (+)	5.65E+00 ± 6.76E-01 (+)	5.48E+00 ± 3.57E-01 (+)	6.30E+01 ± 1.47E+01 (+)
	P4	1.24E+00 ± 6.85E-02	1.26E+00 ± 7.72E-02 (≈)	1.38E+00 ± 1.12E-01 (+)	1.26E+00 ± 9.34E-02 (≈)	1.74E+02 ± 7.57E+00 (+)
	P5	1.32E+02 ± 2.45E+01	1.14E+02 ± 2.68E+01 (-)	1.53E+02 ± 4.03E+01 (+)	1.46E+02 ± 3.05E+01 (+)	4.93E+00 ± 3.38E-01 (-)
+/≈/-		NA	1/3/1	5/0/0	4/1/0	3/1/1
50	P1	8.84E+00 ± 2.46E+00	1.35E+01 ± 4.67E+00 (+)	1.33E+01 ± 3.14E+00 (+)	1.48E+01 ± 4.42E+00 (+)	8.62E+00 ± 1.17E+00 (≈)
	P2	8.34E+01 ± 3.80E+00	8.24E+01 ± 4.03E+00 (≈)	9.88E+01 ± 9.74E+00 (+)	1.09E+02 ± 1.10E+01 (+)	5.71E+01 ± 8.72E+00 (-)
	P3	4.56E+00 ± 2.10E-01	4.90E+00 ± 2.98E-01 (+)	4.79E+00 ± 3.45E-01 (+)	4.90E+00 ± 4.11E-01 (+)	6.64E+00 ± 1.40E+00 (+)
	P4	1.24E+00 ± 4.28E-02	1.89E+00 ± 2.08E-01 (+)	1.42E+00 ± 8.00E-02 (+)	1.34E+00 ± 9.47E-02 (+)	5.02E+00 ± 2.79E-01 (+)
	P5	1.60E+02 ± 3.40E+01	1.78E+02 ± 3.13E+01 (+)	1.98E+02 ± 3.02E+01 (+)	2.38E+02 ± 4.93E+01 (+)	1.59E+00 ± 1.33E-01 (-)
+/≈/-		NA	4/1/0	5/0/0	5/0/0	2/1/2
100	P1	7.56E+01 ± 4.98E+01	2.84E+02 ± 7.76E+01 (+)	5.57E+01 ± 1.23E+01 (≈)	4.17E+02 ± 3.49E+02 (+)	7.19E+01 ± 1.57E+01 (-)
	P2	1.88E+02 ± 3.73E+01	2.41E+02 ± 2.56E+01 (+)	1.94E+02 ± 1.98E+01 (+)	4.99E+02 ± 1.90E+02 (+)	1.73E+02 ± 8.43E+00 (≈)
	P3	4.38E+00 ± 3.32E-01	7.16E+00 ± 6.55E-01 (+)	4.69E+00 ± 2.55E-01 (+)	5.17E+00 ± 6.12E-01 (+)	4.88E+00 ± 1.89E-01 (+)
	P4	1.82E+00 ± 6.11E-01	1.84E+01 ± 1.80E+00 (+)	1.85E+00 ± 2.33E-01 (≈)	3.83E+00 ± 1.98E+00 (+)	4.84E+00 ± 6.60E-01 (+)
	P5	5.00E+02 ± 1.59E+02	7.79E+02 ± 8.46E+01 (+)	4.30E+02 ± 1.43E+02 (-)	9.15E+02 ± 1.27E+02 (+)	2.99E+02 ± 5.13E+01 (-)
+/≈/-		NA	5/0/0	2/2/1	5/0/0	2/1/2
Total of +/≈/-		NA	12/7/1	16/3/1	16/4/0	9/4/7

DDEA-SE, BDDEA-LDG, and DDEA-PES across the vast majority of test functions, achieving five significantly better results against both BDDEA-LDG and DDEA-PES. Similar trends are observed for D=100, where SAMM-DDEA achieves five significantly better results against DDEA-SE and DDEA-PES, and maintains competitive performance against BDDEA-LDG and TT-DDEA. Notably, on P3 and P4, SAMM-DDEA shows substantially lower mean errors and tighter variances, indicating superior surrogate accuracy and more reliable evolutionary guidance.

Overall, the total statistics further confirm the effectiveness of SAMM-DDEA, with 12/7/1 against DDEA-SE, 16/3/1 against BDDEA-LDG, 16/4/0 against DDEA-PES, and 9/4/7 against TT-DDEA. These results demonstrate that SAMM strategy can significantly enhance optimization performance, particularly in high-dimensional and data-scarce scenarios.

C. *The Effect of SAMM and Parameter Analysis*

To study the SAMM effectiveness, this section presents a comparative analysis between SAMM-DDEA and its variants, evaluated under different model scales L. In particular, the SAMM-DDEA (L=5) is compared with variants using L=1, 2, 3, and 4. The results are plotted in Fig. 2.

The results in Fig. 2 clearly demonstrate that the parameter L has a notable impact on the performance of SAMM-DDEA. Overall, SAMM-DDEA with L=5 consistently achieves superior or highly competitive performance across all test problems compared with its variants using smaller L values. When L is small (e.g., L=1 or L=2), the algorithm tends to exhibit inferior performance, which can be attributed to insufficient surrogate model management and a limited ability to capture the complex fitness landscape. In such cases, the surrogate ensemble lacks diversity and robustness, leading to less accurate fitness estimation and suboptimal evolutionary guidance.

As L increases from 1 to 4, a clear performance improvement can be observed, indicating that incorporating more surrogate models and finer evaluation scales helps enhance prediction accuracy and evolutionary efficiency. However, the performance gains gradually saturate as L approaches 5, suggesting that further increasing L may result in diminishing returns while incurring additional computational overhead. This observation confirms that L=5 strikes a favorable trade-off between the predictive fidelity of the surrogate and the associated computational overhead.

In summary, the experimental comparisons confirm both the effectiveness of the SAMM strategy and the rationale for setting L=5 as the default in SAMM-DDEA. This setting enables robust surrogate-assisted evolution and ensures stable performance across diverse problem instances.

TABLE II

COMPARING THE SAMM-DDEA AND ITS VARIANTS WITHOUT FCS

D	ID	SAMM-DDEA	SAMM-DDEA-w/o-F
10	P1	6.96E-01 ± 2.97E-01	1.24E+00 ± 4.74E-01 (+)
	P2	2.61E+01 ± 5.11E+00	3.36E+01 ± 6.51E+00 (+)
	P3	5.16E+00 ± 1.19E+00	6.34E+00 ± 8.03E-01 (+)
	P4	1.29E+00 ± 1.63E-01	1.38E+00 ± 1.25E-01 (+)
	P5	5.59E+01 ± 2.01E+01	7.46E+01 ± 1.02E+01 (+)
+/≈/-		NA	5/0/0
30	P1	3.82E+00 ± 1.24E+00	8.65E+00 ± 2.42E+00 (+)
	P2	5.73E+01 ± 4.18E+00	6.56E+01 ± 7.01E+00 (+)
	P3	4.60E+00 ± 3.87E-01	5.20E+00 ± 4.38E-01 (+)
	P4	1.24E+00 ± 6.85E-02	1.43E+00 ± 1.10E-01 (+)
	P5	1.32E+02 ± 2.45E+01	1.44E+02 ± 4.16E+01 (≈)
+/≈/-		NA	4/1/0
50	P1	8.84E+00 ± 2.46E+00	1.63E+01 ± 3.05E+00 (+)
	P2	8.34E+01 ± 3.80E+00	1.09E+02 ± 9.39E+00 (+)
	P3	4.56E+00 ± 2.10E-01	4.80E+00 ± 2.74E-01 (+)
	P4	1.24E+00 ± 4.28E-02	1.65E+00 ± 3.14E-01 (+)
	P5	1.60E+02 ± 3.40E+01	2.63E+02 ± 4.81E+01 (+)
+/≈/-		NA	5/0/0
100	P1	7.56E+01 ± 4.98E+01	4.91E+02 ± 4.42E+02 (+)
	P2	1.88E+02 ± 3.73E+01	5.50E+02 ± 2.47E+02 (+)
	P3	4.38E+00 ± 3.32E-01	5.17E+00 ± 4.11E-01 (+)
	P4	1.82E+00 ± 6.11E-01	1.70E+01 ± 1.91E+01 (+)
	P5	5.00E+02 ± 1.59E+02	8.59E+02 ± 1.99E+02 (+)
+/≈/-		NA	5/0/0

D. *Effectiveness of FCS*

To study the effectiveness of the proposed FCS in enhancing SAMM-DDEA, this paper compares the SAMM-DDEA with its variant without FCS. The variant is denoted as SAMM-DDEA-w/o-FCS. In particular, the SAMM-DDEA-w/o-FCS uses all the data rather than the FCS to build every single model, so as to better show the influence of FCS on SAMM-DDEA. The comparative results are summarized in Table II.

The results in Table II clearly show the effectiveness of the proposed FCS in enhancing the performance of SAMM-DDEA. Across all dimensional settings, SAMM-DDEA consistently outperforms its variant without FCS, indicating that selectively filtering training data plays a critical role in improving surrogate quality and evolutionary guidance.

In the low-dimensional case (D=10), SAMM-DDEA achieves significantly better performance on all five problems, resulting in a clean win record of 5/0/0. The improvements are particularly notable on P1, P2, and P5, where the mean errors of SAMM-DDEA-w/o-FCS are substantially larger. This suggests that even in relatively simple search spaces, using all historical

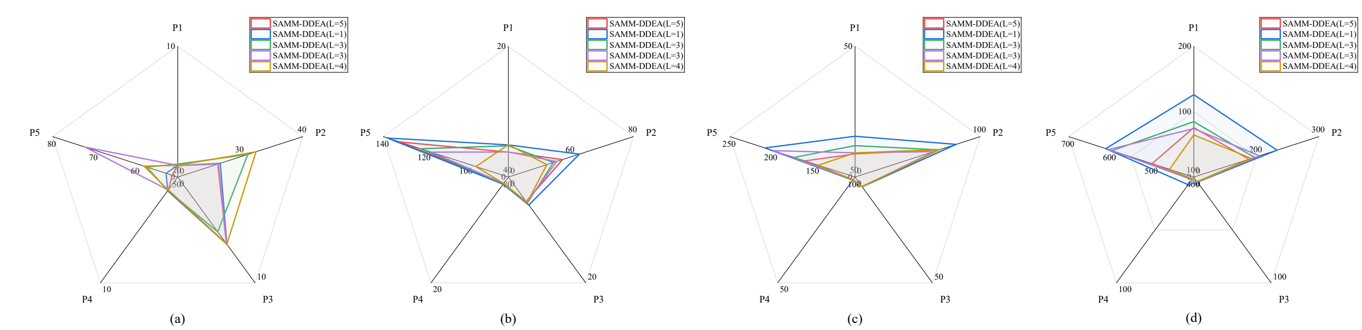

Fig. 2. The performance of SAMM-DDEA variants with different L.

data indiscriminately may introduce noisy or misleading samples, whereas FCS effectively preserves informative data for surrogate construction.

When D=30, SAMM-DDEA again outperforms the variant without FCS on four problems and achieves comparable performance on one problem. The performance gap on P1 and P2 is especially large, indicating that FCS becomes increasingly beneficial as the search space grows. By dynamically selecting samples, FCS enables the surrogate models to better capture local fitness landscapes under limited evaluation budgets.

For higher-dimensional cases (D=50 and D=100), the advantages of FCS are further amplified. SAMM-DDEA achieves significant improvements on all five problems in both settings. Without FCS, the surrogate models suffer from severe degradation, reflected by extremely large mean errors and variances, especially on P1, P2, and P4. In contrast, SAMM-DDEA maintains stable convergence and lower errors, demonstrating strong robustness and scalability.

Overall, the consistently significantly better performance records across all dimensions confirm that FCS is a key component of SAMM-DDEA, which effectively alleviates model bias and overfitting, leading to more accurate surrogate-assisted evaluations and substantially improved optimization performance, particularly in high-dimensional scenarios.

V. Conclusion

This paper proposes a SAMM strategy for expensive evolutionary optimization that explicitly regulates the evaluation scale of surrogate-assisted fitness estimation based on evolutionary feedback. By dynamically adjusting the number of surrogate models involved in ensemble evaluation, SAMM enables coarse fitness assessment in early evolutionary stages and progressively finer evaluation as the search converges. To support effective scale adaptation under limited data, an FCS strategy was further introduced to enhance the diversity and robustness of ensemble surrogate models.

Based on these components, the SAMM-DDEA was developed and evaluated on expensive optimization problems. Experimental results demonstrate that SAMM-DDEA attains performance superior to, or at least highly competitive with, several SOTA DDEAs, particularly in high-dimensional scenarios, indicating that adaptive evaluation scale management is an effective model management mechanism for improving data-driven evolutionary optimization.

References

[1] J. Y. Li, Z. H. Zhan, and J. Zhang, "Evolutionary computation for expensive optimization: A survey," *Mach. Intell. Res.*, vol. 19, pp. 3-23, Jan. 2022.

[2] J. Y. Li, Z. H. Zhan, J. Xu, S. Kwong and J. Zhang, "Surrogate-assisted hybrid-model estimation of distribution algorithm for mixed-variable hyperparameters optimization in convolutional neural networks," *IEEE Trans. Neur. Netw. Learn. Syst.*, vol. 34, no. 5, pp. 2338–2352, May 2023.

[3] X. Y. Wang, J. Y. Li, Z. H. Zhan, and J. Zhang, "A roulette wheel selection-based data-driven evolutionary algorithm for expensive optimization problems," in *Proc. Int. Conf. Inf. Sci. Technol.*, Dec. 2025, pp. 66–71.

[4] H. Li *et al.*, "Privacy-enhanced offline data-driven evolutionary optimization based on cloud server," *IEEE Trans. Evol. Comput.*, doi: 10.1109/TEVC.2025.3584882.

[5] Z. Azarhoosh and M. Ilchi Ghazaan, "A review of recent advances in surrogate models for uncertainty quantification of high-dimensional engineering applications," *Comput. Methods Appl. Mech. Eng.*, vol. 433, Art. no. 117508, Jan. 2025.

[6] P. Huang, H. Wang, and W. Ma, "Stochastic ranking for offline data-driven evolutionary optimization using radial basis function networks with multiple kernels," in *Proc. IEEE Symp. Ser. Comput. Intell.*, Dec. 2019, pp. 2050–2057.

[7] T. Chugh, Y. Jin, K. Miettinen, J. Hakanen, and K. Sindhya, "A surrogate-assisted reference vector guided evolutionary algorithm for computationally expensive many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 22, no. 1, pp. 129–142, Feb. 2018.

[8] J. Y. Li, Z. H. Zhan, H. Wang, and J. Zhang, "Data-driven evolutionary algorithm with perturbation-based ensemble surrogates," *IEEE Trans. Cybern.*, vol. 51, no. 8, pp. 3925–3937, Aug. 2021.

[9] L. Zheng, F. Gu, and S. Chen, "Data-driven evolutionary optimization with clustering aided ensemble learning and taylor polynomial-based data generation," *Complex Intell. Syst.*, vol. 12, no. 1, p. 43, Jan. 2026.

[10] Z. Guo, S. Lin, R. Suo, and X. Zhang, "An offline weighted-bagging data-driven evolutionary algorithm with data generation based on clustering," *Mathematics*, vol. 11, no. 2, p. 431, Jan. 2023.

[11] R. Zhao, J. Kang, L. Fu, C. Liu, C. Jiang, and Y. Shi, "Data-driven evolutionary algorithm with model calibration strategy and data perturbation strategy for electricity resource economic scheduling optimization," *Ann. Oper. Res.*, Jan. 2026, doi: 10.1007/s10479-025-07007-7.

[12] J. Y. Li, Z. H. Zhan, C. Wang, H. Jin, and J. Zhang, "Boosting data-driven evolutionary algorithm with localized data generation," *IEEE Trans. Evol. Comput.*, vol. 24, no. 5, pp. 923–937, Oct. 2020.

[13] Q. T. Yang, J. Y. Li, Z. H. Zhan, Y. Jiang, Y. Jin and J. Zhang, "A hierarchical and ensemble surrogate-assisted evolutionary algorithm with model reduction for expensive many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 29, no. 6, pp. 2491-2505, Dec. 2025.

[14] S. Wang et al., "Probability selection-based surrogate-assisted evolutionary algorithm for expensive optimization," *Appl. Sci.*, vol. 15, no. 21, p. 11404, Oct. 2025.

[15] J. Yuan et al., "A gumbel-based selection data-driven evolutionary algorithm and its application to chinese text-based cheating official accounts mining," *Algorithms*, vol. 18, no. 10, p. 643, Oct. 2025.

[16] M. Zhang, H. Li, S. Pan, J. Lyu, S. Ling, and S. Su, "Convolutional neural networks-based lung nodule classification: A surrogate-assisted evolutionary algorithm for hyperparameter optimization," *IEEE Trans. Evol. Comput.*, vol. 25, no. 5, pp. 869–882, Oct. 2021.

[17] H. Wang, Y. Jin, C. Sun, and J. Doherty, "Offline data-driven evolutionary optimization using selective surrogate ensembles," *IEEE Trans. Evol. Comput.*, vol. 23, no. 2, pp. 203-216, Apr. 2019.

[18] P. Huang, H. Wang, and Y. Jin, "Offline data-driven evolutionary optimization based on tri-training," *Swarm Evol. Comput.*, vol. 60, no. 100800, pp. 1–12, Feb. 2021.

[19] S. H. Wu, Z. H. Zhan and J. Zhang, "SAFE: Scale-adaptive fitness evaluation method for expensive optimization problems," *IEEE Trans. Evol. Comput.*, vol. 25, no. 3, pp. 478-491, Jun. 2021.

[20] Y. Wang, J. Y. Li, C. Chen, J. Zhang, and Z. H. Zhan, "Scale adaptive fitness evaluation-based particle swarm optimisation for hyperparameter and architecture optimisation in neural networks and deep learning," *CAAI Trans. Intell. Technol.*, vol. 8, no. 3, pp. 849–862, Sep. 2023.

[21] D. Lin, H. Huang, X. Li, and Y. Gong, "Empirical study of data-driven evolutionary algorithms in noisy environments," *Mathematics*, vol. 10, no. 6, pp. 1–26, Mar. 2022.

[22] S. Wang, L. L. Minku, and X. Yao, "Resampling-based ensemble methods for online class imbalance learning," *IEEE Trans. Knowl. Data Eng.,* vol. 27, no. 5, pp. 1356–1368, May. 2015.

[23] M. Yu, X. Li, and J. Liang, "A dynamic surrogate-assisted evolutionary algorithm framework for expensive structural optimization," *Structural and Multidisciplinary Optimization*, vol. 61, no. 2, pp. 711–729, 2020.

[24] H. Huang and Y. Gong, "Contrastive learning: An alternative surrogate for offline data-driven evolutionary computation," *IEEE Trans. Evol. Comput.*, vol. 27, no. 2, pp. 370–384, Apr. 2023.

[25] Y. H. Sun, T. Huang, J. H. Zhong, J. Zhang, and Y. J. Gong, "Symbolic regression-assisted offline data-driven evolutionary computation," *IEEE Trans. Evol. Comput.*, vol. 29, no. 5, pp. 2158–2172, Oct. 2025.

[26] M. Stein, "Large sample properties of simulations using Latin hypercube sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, May 1987.