Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 7 Issue 4, April - 2020

On Metric for Evaluating the Performance of Multi-Objective Algorithms

Salih Duffuaa
Systems Engineering Department
King Fahd University of Petroleum & Minerals
Dhahran, 31261, Saudi Arabia
duffuaa@kfupm.edu.sa

Awsan Mohammed
Construction Engineering and Management
King Fahd University of Petroleum & Minerals
Dhahran, 31261, Saudi Arabia
Awsan.mohammed@kfupm.edu.sa

***Abstract*—In this paper, several new metrics to evaluate the performance of multi-objective optimization algorithms are proposed. A notable metric that is based on the probability of obtaining the best solution is one of the new of metrics. The suggested metrics are concerned with the convergence and diversity characteristics of the solutions. In addition, the relationship among the proposed and existing metrics is investigated. The advantages of the proposed metrics are highlighted. A computational experiment based on a multi-objective supply chain network design problem using three metaheuristic algorithms is conducted to demonstrate the application of the new metrics. The proposed metrics expand and enrich this area of research.**

Keywords—Metrics; Multi-objective; Optimization; Meta-heuristics; Supply chain.

I. Introduction

In recent years, multi-objective optimization has attracted the attention of both academics and practitioners. In this regard, various multi-objective optimization algorithms have been developed. These algorithms are categorized into well-known exact methods and meta-heuristic approaches. The limitations of using exact methods for solving large problems have given rise to the development of a wide range of meta-heuristic optimization approaches [1]. These approaches include particle swarm, ant colony, simulated annealing, and genetic algorithms. The main goal of multi-objective optimization is to obtain a set of solutions for optimizing several conflicting objective functions simultaneously. The general form of the multi-objective model is usually stated as:

Minimize $f\ (x) = [f_1(x), f_2(x), f_3(x), \ldots, f_{n(x)}]$ (1)

S.t. $x \in S$,

Where $S = \{X \in R^m : h(x) = 0\,, g(x) \geq 0\}$,

$n > 1$, S represents the feasible region, and f is a vector valued objective function.

In the case of multi-objective optimization, the optimality conditions are rather different than the one for single objective optimization. The Pareto optimality and Pareto dominance concepts are employed for characterizing a multi-objective solution. The solution is called a Pareto-efficient solution if improving the value of one objective degrades at least one other objective value. Moreover, if there are no other solutions that dominate the solution, it is said to be Pareto optimal.

Many metrics have been developed to evaluate the performance of multi-objective algorithms. However, most of the available metrics in the literature measure only one feature of the solutions for evaluating the performance of multi-objective algorithms. Accordingly, our suggested approaches take into account more than one feature of the solutions. To be more precise, in the comparison of algorithms for multi-objective optimization problems, several solutions of an algorithm are dominated by the solutions of other algorithms. The dominated solutions might have extreme (large or small) distances from the ideal solution. Such solutions are considered as low-quality solutions in terms of the quality of the solutions metric such as percent of domination. Taking these solutions into account will increase or decrease the existing mean ideal distance and spread the non-dominance solutions metrics dramatically. Therefore, it is worthwhile to eliminate such solutions in computing the mean ideal distance and the spread non-dominance solutions metrics for the case of comparing different algorithms. Consequently, to avoid the drawbacks mentioned, new measures are proposed.

In addition, in this paper a novel metric based on the probability of obtaining the best solution is proposed. Moreover, new metrics for measuring the convergence and the diversity of solutions from the grand ideal distance are suggested. Last but not least, the relationship among the existing and proposed metrics is investigated. A computational experiment is conducted to test and demonstrate the use of the proposed metrics. The problems used in the computational experiment deal with the design of supply chain networks in a multi-objective framework.

The rest of this paper is organized as follows: the literature review is provided in section 2, followed by the suggested metrics in section 3. Then, the relationships between the proposed and existing metrics are investigated in section 4. The computational experiment to demonstrate the application of the proposed metrics is provided in section 5. Finally, the conclusions and future work are presented in section 6.

TABLE I. NOTATIONS

n	The number of optimal Pareto solutions obtained by an algorithm
d	Number of dominated solutions by the solutions of another algorithm
MID	Mean ideal distance
POD	Percent of domination
SNS	Spread of non-dominance solution
MID_{NDS}	Mean ideal distance of solutions that are non-dominated by another algorithm
SNS_{NDS}	Spread of non-dominance solutions that are non-dominated by another algorithm
$MGID$	Mean grand ideal distance
$SGNS$	Spread grand of non-dominance solutions
$\vec{f}_i$	Vector of solution i
$\vec{f}_{ideal}$	Vector of the ideal point which is the best value of objective that can be obtained by the algorithm when the problem is solved as a single objective
$\vec{f}_{grand-ideal}$	Vector of the grand ideal point which is the best solution selected among all the ideal solutions obtained by competing algorithms
c_i	The distance between solution i and the ideal point
RDC	The ratio of the new diversity metric to new convergence metric
Ed_i	The Euclidean distance between each solution obtained and the nearest member of the Pareto-optimal solutions
$\hat{P}_i$	The fraction of times that the obtaining obtains the best solution
Val_i	The current value of the performance metric i
Val_{min}	The minimum value of the metric obtained by the algorithms
Val_{max}	The maximum value of the metric obtained by the algorithms

II. LITERATURE REVIEW

Recently, various performance measures for the evaluation of multi-objective meta-heuristics algorithms have been developed. One of the performance metrics is the hypervolume (HV) or S metric [2]. It measures the space size of the objective for all the obtained solutions. A reference point is utilized to compute the landscape of the objective. HV can measure both the convergence and diversity of solutions. However, the major disadvantage of the HV metric is the computation time required to calculate the objective space. It is also necessary to select the reference point carefully.

Other existing metrics for measuring the convergence include the generational distance metric (GD), inverted generational distance (IGD) [4], the convergence metric [8], and the convergence measure (ϒ) [7]. GD measures the average of the minimum distances of the solutions from the nearest member of the Pareto-optimal solutions. It is developed by Van Veldhuizen and Lamont [3] and is computed as:

$$GD = \sqrt{\frac{\sum_{i=1}^{n} Ed_i^2}{n}} \quad (2)$$

The disadvantage of this metric is that it is essential to know the nearest Pareto-optimal solutions. The complementary metric of generational distance is IGD. This measures the minimum distance of the solutions (instead of average) from the nearest member of the Pareto-optimal solutions. The unavailability of the optimal Pareto front is the major drawback of this metric. The convergence measure (ϒ) is the same as IGD, but it evaluates the mean distance from the optimal Pareto front. This metric cannot be utilized if the optimal Pareto solutions are unavailable.

Moreover, the common convergence metric used for evaluating the performance of algorithms in the multi-objective supply chain area is the mean ideal distance (MID). MID measures the closeness of the Pareto solutions to a calculated ideal solution. It is calculated through the following formula:

$$MID = \frac{\sum_i c_i}{n} \text{ where } c_i = \|\vec{f}_i - \vec{f}_{ideal}\| \quad (3)$$

The drawback of this metric is that it cannot provide more information about the characteristics of solutions.

On the other hand, the existing metrics that measure the domination of solutions include percent of domination (POD) and two set coverage (C-metric) [6]. POD measures the ability of the solutions obtained by an algorithm to dominate the solutions of other algorithms. It is computed as follows, first, all efficient solutions obtained by the algorithms are mixed and the combination of Pareto efficient solutions is constructed. Then, all dominated solutions are deleted and the ratio of the solutions belonging to each algorithm is computed. POD can be computed by the following formula:

$$POD(algi) = \frac{n-d}{n} \qquad (4)$$

The algorithm with the higher POD has better performance.

The C-metric measures the ability of a solution set to dominate another set. It does not provide more information about the domination and its values are not easy to interpret if the two sets of solutions are not comparable.

In contrast, the main existing metrics that measure the diversity of solutions include the spread of non-dominated solutions (SNS) and the spread metric (Δ) [5]. SNS evaluates the standard deviation of the distance of the ideal point from the Pareto solutions. It is calculated by the following formula:

$$SNS = \sqrt{\frac{\sum_{i=1}^{n}(MID-c_i)^2}{n-1}} \qquad (5)$$

The disadvantage of these metrics is that they only measure the diversity of solutions without considering any more information about the solutions. Other traditional metrics that evaluate the diversity are the diversity metric (*DM*) [8], the entropy metric [9], the diversity spacing [10], and the M_3^* metric [11]. The drawback of these metrics is that they only consider one aspect of the solution characteristic.

Based on the literature and the drawbacks of the existing metrics, this paper proposes new metrics for evaluating the performance of multi-objective optimization algorithms. This work is distinguished from previous studies in the following aspects. First, a new metric based on the probability of obtaining the closest solution to the ideal solution is developed. Second, new metrics that measure the average solutions distances and the diversity of solutions from the grand ideal distance are proposed. Third, new measures related to the convergence and the diversity of solutions are proposed. Finally, the relationship between new and existing metrics is investigated.

III. Proposed Metrics

In this section, the proposed metrics for evaluating the performance of algorithms for multi-objective optimization problems are presented.

A. *The probability of best solution (P)*

This metric is defined based on the probability of success in obtaining the best solution in N trials. A trial is a problem of a certain dimension which is called an incidence to be solved. To compute P, first, each algorithm is utilized to solve N trials of the multi-objective problem with a certain dimension. At each trial, each algorithm might be employed for several runs. In this case, the best solution is selected based on the minimum distance of the solution from the ideal point at each run and then the average is calculated. The fraction of times ($\hat{p}_i$) that algorithm *i* obtains the best solution is computed and used as an estimate of the probability of obtaining the best solution or probability of success. ($\hat{p}_i$) is used as a measure of performance and the algorithm with the highest ($\hat{p}_i$) is the best. The probability of success is calculated by the following formula:

$$\hat{P}_i = \frac{\Upsilon_i}{N} \qquad (6)$$

Where Y_i is the number for obtaining the best solutions for algorithm *i*. It is known that $\hat{p}_i$ is an unbiased and a minimum variance for $\hat{p}_i$.

B. *Mean grand-ideal distance*

MGID measures the distance of the solutions obtained by an algorithm from the grand ideal point. In this metric, the grand ideal point is the best solution selected among all the ideal solutions obtained by competing algorithms. The main advantage of MGID is that it computes the mean distance of the solutions considering all the ideal solutions obtained from all of the algorithms. It is calculated as:

$$MGID = \frac{\sum_{i=1}^{n} EC_i}{n} \qquad (7)$$

where $EC_i = \| \vec{f}_i - \vec{f}_{grand-ideal} \|$.

The algorithm with less MGID has better performance.

C. *Spread grand-of non-dominance solutions*

It evaluates the standard deviation of the distance of the solutions obtained by an algorithm from the grand ideal point. It is computed as follows:

$$SGNS = \sqrt{\frac{\sum_{i=1}^{n}(MGID-EC_i)^2}{n-1}} \qquad (8)$$

The higher value of SGNS produces a better solution quality.

D. *Mean ideal distance of non-dominated solutions by other algorithms*

MID_{NDS} measures the distance of nondominated solutions by a competing algorithm from the ideal solution. The main advantage of MID_{NDS} is that it computes the mean distance of non-dominated solutions from the ideal point. It is calculated as:

$$MID_{NDS} = \frac{\sum_{i=1}^{n-d} c_i}{n-d} \qquad (9)$$

The algorithm with less MID_{NDS} has better performance.

E. Spread of non-dominance solutions by other algorithms

It evaluates the standard deviation of the distance of the non-dominated solutions by competing algorithm from the ideal solution. This metric is affected, not only by the spread of solutions, but also by the quality of the solutions as defined by being non-dominated.It is computed as follows:

$$SNS_{NDS} = \sqrt{\frac{\sum_{i=1}^{n-d}(MID_{NDS}-c_i)^2}{n-d-1}} \quad , d < n-1 \quad (10)$$

The higher value of SNS_{NDS} produces a better solution quality.

F. The ratio of the diversity metric to the convergence metric

This measure combines the two main suggested measures in (5) and (6) above. It is calculated using the following formula:

$$RDC = \frac{SNS_{NDS}}{MID_{NDS}} \quad (11)$$

The higher value of RDC produces a better solution quality. The major advantage of RDC is that it considers both the diversity and convergence of non-dominated solutions by other algorithms.

G. The ratio of the quality metrics (RQ)

This metric combines the main performance measures of quality solutions of multi-objective algorithms. It measures the ratio of the suggested mean ideal distance to the percent of non-dominated solutions. It is computed as follows:

$$RQ = \frac{MID_{NDS}}{POD} \quad (12)$$

The algorithm with less RQ has better performance.

IV. THE RELATIONSHIP BETWEEN THE EXISTING AND PROPOSED METRICS

In this section, the relationship between the most common existing metrics and the proposed ones for evaluating the performance of algorithms for multi-objective optimization is investigated. We also try to show that MID_{NDS} measures the mean ideal distance of the solutions and is affected by the quality of the solutions.

First, the relationship among POD, MID and MID_{NDS} is presented:

$$MID = \frac{\sum_{i=1}^{n} c_i}{n} = \frac{\sum_{i=1}^{d} c_i + \sum_{i=1}^{n-d} c_i}{n}$$

Divide MID by POD, then

$$\frac{MID}{POD} = \frac{\sum_{i=1}^{d} c_i + \sum_{i=1}^{n-d} c_i}{n} \times \frac{n}{n-d} = \frac{\sum_{i=1}^{d} c_i + \sum_{i=1}^{n-d} c_i}{n-d}$$

$$= \frac{\sum_{i=1}^{d} c_i}{n-d} + \frac{\sum_{i=1}^{n-d} c_i}{n-d}$$

$$\frac{MID}{POD} = \frac{\sum_{i=1}^{d} c_i}{n-d} + MID_{NDS} \quad (13)$$

The ratio of $\frac{MID}{POD}$ measures two features of the solutions which are: the mean ideal distance (MID) and the quality of solutions (POD). The algorithm with lower $\frac{MID}{POD}$ is better. From eq.1, the ratio of $\frac{MID}{POD}$ is a function in terms of MID_{NDS}.

To clarify the effect of the domination ratio on the convergence metrics, the following lemmas are provided:

Lemma 1: If there is no domination, then $MID = MID_{NDS}$

This is obvious, since there is no domination, then $d = 0$

$$MID_{NDS} = \frac{\sum_{i=1}^{n-d} c_i}{n-d} = \frac{\sum_{i=1}^{n} c_i}{n} = MID.$$

Lemma 2: If there is no domination, then $RQ = MID = MID_{NDS}$

Since $d = 0$, $POD = 1$ and $RQ = \frac{MID_{NDS}}{1} = MID_{NDS} = MID$

Lemma 3: If $d = n$, then this is an extreme case and the dominated algorithm is not competitive and should be eliminated from the comparison.

Based on the aforementioned lemmas and the formulas of MID and MID_{NDS}, it can be concluded that MID has the same values in all lemmas. This means that MID is not affected by the quality and the characteristics of solutions acquired by the algorithm. On the other hand, the value of MID_{NDS} depends on the distance of the solutions from the ideal point, the number of solutions that are dominated by other algorithms, and the characteristics of quality solutions. Therefore, MID_{NDS} is a more suitable measure.

V. COMPUTATIONAL EXPERIMENT

In this section, a computational experiment is conducted to illustrate the applicability of the proposed multi-criteria measures for evaluating the performance of multi-objective algorithms and to demonstrate the relationship among the metrics. The application used is the design multi-objective, multi-products, five-echelon supply chain network problem. Thirteen problems covering different supply chain network sizes are generated, as shown in Table II.

For solving these problems, three meta-heuristic algorithms are utilized; namely, the tabu search algorithm (TS), genetic algorithm (GA), and simulated annealing (SA). The objective is to evaluate the effectiveness and efficiency of the algorithms for solving multi-objective supply chain design problems using the existing and the proposed metrics as comparison measures.

The first evaluation metric is the probability of the best solution. To assess the applicability of the probability of

TABLE II. TEST PROBLEM INSTANCES

Problem number	Suppliers	Plants	warehouses	DCs	Customers	Products
1	3	3	3	3	10	4
2	5	5	5	5	20	4
3	6	6	6	6	20	4
4	8	8	8	8	28	4
5	10	10	10	10	40	4
6	12	12	12	12	60	4
7	15	15	15	15	70	4
8	18	18	18	18	80	4
9	20	20	20	20	90	4
10	22	22	22	22	100	6
11	25	25	25	25	100	6
12	28	28	28	28	120	6
13	30	30	30	30	130	6

best solution metric, the proper number of trials of the problem (N) is selected. To do this, several trials of a small size-scale are conducted. For each trial, the probability of success is computed, as shown in Table III. Then, the median of the probability of success obtained by each algorithm for all trials is tested. In this regard, we are interested in testing the null hypothesis H_0: median = m_0 against the alternative hypothesis H_1: median ≠ m_0, where m_0 is equal to 0.4, 0.1, and 0.5 for tabu search, genetic algorithm, and simulated annealing, respectively. Since the distribution of the source of the collected data is unknown, a non-parametric test called a sign test is utilized to test this hypothesis.

As a result, the p-value of the test is greater than a 0.05 significance level, as shown in Table IV. Thus, we do not reject the null hypothesis. This means that the probabilities of success for obtaining the best solution are 0.4, 0.1, and 0.5 for the tabu search algorithm, genetic algorithm, and simulated annealing, respectively.

TABLE III THE PROBABILITY OF SUCCESS IN OBTAINING THE SOLUTIONS FOR ALL TRIALS IN A SMALL PROBLEM

Problem trial \ Algorithm	TS	GA	SA
5	0.40	0.00	0.60
8	0.38	0.13	0.50
10	0.40	0.10	0.50
12	0.42	0.08	0.50
15	0.40	0.13	0.47
18	0.39	0.11	0.50
20	0.35	0.10	0.55

Accordingly, the proper number of trials is selected to be 10. Afterward, the metric is used for evaluating the performance of the meta-heuristic algorithms by solving three types of problems. The problems cover different sizes: small size, medium size, and large size. First, each algorithm is utilized to solve 10 trials or incidences of a multi-objective problem with a certain dimension.

At each trial, each algorithm is also employed for 10 runs. At each run, the best solution is selected based on the minimum distance of the solution from the ideal point. Then, the average of all the best solutions acquired by algorithms is computed for each trial, as shown in Table (V-VI). The number of solutions belonging to each algorithm is counted and the probability of success is calculated, as shown in Table VII. From Table VII, it can be seen that there is a slight statistically significant difference between TA and SA. However, both algorithms perform better than GA in terms of the probability of best solutions metric.

TABLE III . SIGN TEST OUTPUT OF THE PROPOSED ALGORITHMS

Algorithm	N	Below	Equal	Above	P	Median
Tabu search	7	3	3	1	0.63	0.40
Genetic algorithm		2	2	3	1.00	0.10
Simulated annealing		1	4	2	1.00	0.50

Table IX illustrates the performance measures for the three algorithms for solving MOSC problems. From Table IX, it can be seen that the algorithm that obtains the best MID differs from the algorithm that obtains the best MID_{NDS} for many problems. This difference is due to the fact that the mean ideal distance is affected by the number and characteristic of dominated solutions, especially when the dominated solutions have an extreme (large or small) distance. Taking these solutions into account leads to an increase or decrease

TABLE IV. THE AVERAGE OF THE BEST SOLUTIONS OBTAINED BY ALGORITHMS (SMALL SIZE)

Algorithm / Trial number	TS	GA	SA
1	60044014.26	63043999.50	62745626.62
2	104080492.20	105507175.70	104219952.30
3	67755028.65	71700811.75	66643972.70
4	76920176.63	76422701.25	76671439.78
5	57805574.44	61252734.36	58328394.93
6	68969399.88	68332795.57	66604922.80
7	70032896.71	75840102.95	68114772.84
8	36529799.17	38460258.15	35107867.36
9	100255810.20	101478654.90	100384936.20
10	59983691.29	62060561.40	59170514.62

TABLE VI. THE MINIMUM DISTANCE FROM THE IDEAL POINT OBTAINED BY ALGORITHMS (MEDIUM SIZE)

Algorithm / Trial number	TS	GA	SA
1	151353834.20	152020075.90	153678717.40
2	86782255.79	85965490.84	84343976.65
3	142785445.30	138551191.60	135067107.70
4	67816501.81	62720187.69	65018374.03
5	70205572.82	69807104.16	68273858.50
6	35247419.50	37917409.30	35333882.13
7	46335104.45	49210868.09	36226695.14
8	155166204.80	159567511.40	158410570.30
9	90345099.64	92240090.94	94561902.53
10	109896022.70	108764065.70	111764958.70

TABLE VII. THE MINIMUM DISTANCE FROM THE IDEAL POINT OBTAINED BY ALGORITHMS (LARGE SIZE)

Algorithm / Trial number	TS	GA	SA
1	8646811441	8639057104	8643721463
2	8681940768	8682927950	8683039691
3	8565666122	8562987969	8561714433
4	8517262981	8520603121	8523731729
5	8689624542	8683563088	8673179341
6	8581128087	8581632342	8587695102
7	8600058742	8595467975	8590066700
8	8613730753	8604295121	8590757692
9	8604144033	8595662478	8598097619
10	8530176941	8524061770	8522063380

TABLE VII. THE PROBABILITY OF SUCCESS IN OBTAINING THE SOLUTIONS FOR ALL PROBLEMS

Algorithm / Problem type	TS	GA	SA
Small size	0.40	0.10	0.50
Medium size	0.40	0.20	0.40
Large size	0.30	0.20	0.50
Average	0.37	0.17	0.47

TABLE IX. THE OBTAINED METRICS FOR ALGORITHMS' PERFORMANCE (MID, MGID, POD, MID_{NDS}, AND RQ)

Problem / Algorithm	Prob.1	Prob.2	Prob.3	Prob.4	Prob.5	Prob.6	Prob.7	Prob.8	Prob.9	Prob.10	Prob.11	Prob.12	Prob.13
MID													
GA	**11779343**	82277325	40965454	**57433167**	68140599	492307597	107483122.5	1750115527	**334011397**	8653622296	8635511098	2295349550	3756105740
SA	11796552	**81250448**	**39126236**	57670326	69676241	**491994176**	**106951415**	**1745654966**	334061261.4	**8650395541**	**8632347219**	2309156624	3763516129
TS	11827320	82841381	40351160	58658155	**67323108**	495354635	109568762	1747304452	339826148.6	8660517212	8633727169	**2280224199**	**3742371628**
MGID													
GA	12660569	88522812	43450115	61697701	86613527	**556922624**	102157064	1965125538	**351944961**	8737081514	94831926	**104316430**	**3862169133**
SA	**11892476**	86177590	40759527	64936773	79149970	562237410	**101471324**	1974129921	354770022	8740220690	**59311872**	106058974	3867520992
TS	17137609	**85355087**	**38792822**	**61266968**	**71741874**	558118089	107485934	**1817566798**	352543431	**8631550608**	74816392	109298007	3884878809
POD													
GA	0.02	0.53	0.56	0.527	0.636	0.617	0.946	0.562	0.527	0.263	**0.97**	0.21	**0.718**
SA	**0.313**	0.326	0.188	0.575	0.222	0.375	**0.972**	0.02	0.116	0.333	0.718	0.636	0.35
TS	0.02	**0.778**	**0.75**	**0.642**	**0.852**	**0.945**	0.569	**0.98**	**0.98**	**0.77**	0.94	**0.82**	0.67
MID_{NDS}													
GA	2070806	101279053	60082935	79633418	32300434	658091970	113247597	2053532482	418756072	268304551	**8595182253**	2375216519	**3758197936**
SA	12068346	**77910686**	40571654	35824225	5816920	**468145822**	**111256175**	7642002838	498629418	**260038654**	8633136763	2367197830	3789224643
TS	**1001316**	81142955	**23939323**	**30197825**	**24357228**	487131108	234995830	**1821726587**	**413890480**	283729347	8602633286	**2335076183**	3776710362
RQ													
GA	105261350	191092552.8	107290955	151107055	50786845.9	1.067E+09	119712047.6	3653972388	794603552.2	1020169395	**8861012632**	1.1311E+10	**5234258964**
SA	**9803022.36**	238989834.4	215806670	62303000	**26202342.3**	1.248E+09	**114461085.4**	9.21E+10	4298529466	780896858.9	1.2024E+10	3722009167	1.0826E+10
TS	54572200	**104296857.3**	**31919097.3**	**47037110.6**	28588295.8	**515482654**	412997943.8	**1858904681**	**422337224.5**	**368479671.4**	9151737538	**2847653882**	5636881137

TABLE X. AVERAGE OF METRICS FOR ALL TEST PROBLEMS

Algorithm	Performance Metric									
	MID	Rank	MGID	Rank	POD	Rank	MID_{NDS}	Rank	RQ	Rank
GA	0.44	2	0.48	2	0.47	2	0.50	3	0.53	2
SA	0.27	1	0.52	3	0.25	3	0.49	2	0.61	3
TS	0.65	3	0.37	1	0.83	1	0.29	1	0.13	1

TABLE XI. THE OBTAINED METRICS FOR ALGORITHMS'S PERFORMANCE (SNS, SGNS, MID_{NDS}, SNS_{NDS}, AND RDC)

Problem / Algorithm	Prob.1	Prob.2	Prob.3	Prob.4	Prob.5	Prob.6	Prob.7	Prob.8	Prob.9	Prob.10	Prob.11	Prob.12	Prob.13
SNS													
GA	2949080	5255060	4064258	5553359	5720683	5089852.1	5499420.27	7074498.78	6968421.1	7748692.68	9678728.32	16660278.9	16024478.4
SA	**3222456**	**6583532**	**4711693**	6175994	**7300760**	**6362972.6**	**6110407**	**9399708.93**	**8442487.4**	**12708951.8**	14684047.8	**18104638.1**	**16686729.4**
TS	3066016	5199882	3887345	**6436757**	5507736	5365697	4785269	6303348.95	6489076.6	8402942.5	**9658809.69**	12309790.2	14659054
SGNS													
GA	**1897269**	**3772991**	1071396	2139537	12582802	2415298	6388248	1.09E+08	**11062793**	**6889466**	2712383	7661059	**6310466**
SA	327425.8	1569011	**1146277**	**3711296**	**14203988**	**4.08E+08**	4232680	**2.04E+08**	4830288	2949980	**30566858**	**25282018**	4076532
TS	347710.2	2154991	730962.2	1796381	938302.9	2827370	**6550539**	50363811	5397068	1411142	10266719	23493333	3086437
MID_{NDS}													
GA	2070806	101279053	60082935	79633418	32300434	658091970	113247597	2053532482	418756072	268304551	**8595182253**	2375216519	**3758197936**
SA	12068346	**77910686**	40571654	35824225	5816920	**468145822**	**111256175**	7642002838	498629418	**260038654**	8633136763	2367197830	3789224643
TS	**1001316**	81142955	**23939323**	**30197825**	**24357228**	487131108	234995830	**1821726587**	**413890480**	283729347	8602633286	**2335076183**	3776710362
SNS_{NDS}													
GA	0	**69278783.6**	24664232	22031967	41428012	456161120	107530143	**1766465023**	341400240	**375258873**	8801345258	**2827093969**	3731281709
SA	**838062.6**	67357007.9	**32097050**	**25302594**	**42864752**	**458432040**	133804444	0	**365228636**	253577608	**9155804657**	2293654127	**3844773506**
TS	0	65715376.5	14899179	14311054	31918581	451864101	**162721610**	1758860285	328564976	268505269	8600598604	2252614222	3698681192
RDC													
GA	0.000	0.684	0.411	0.277	1.283	0.693	0.950	0.860	**0.815**	**1.399**	**1.024**	**1.190**	0.993
SA	**0.069**	**0.865**	**0.791**	**0.706**	**7.369**	**0.979**	**1.203**	0.000	0.732	0.975	1.061	0.969	**1.015**
TS	0.000	0.810	0.622	0.474	1.310	0.928	0.692	**0.965**	0.794	0.946	1.000	0.965	0.979

TABLE XII. AVERAGE OF METRICS OBTAINED BY GA, SA, AND TS

Algorithm	Performance Metric											
	MID	Rank	SNS	Rank	SGNS	Rank	MID_{NDS}	Rank	SNS_{NDS}	Rank	RDC	Rank
GA	0.44	2	0.22	2	0.63	1	0.50	3	0.59	2	0.40	2
SA	0.27	1	0.98	1	0.58	2	0.49	2	0.69	1	0.70	1
TS	0.65	3	0.14	3	0.20	3	0.29	1	0.16	3	0.33	3

in the value of MID. Specifically, Table IX shows that the tabu search algorithm obtains the best value of MID_{NDS} for many problems, followed by simulated annealing, and then genetic algorithm. On the other hand, the simulated annealing dominates the other algorithms in terms of MID. The reason behind this is that a few of the solutions obtained by tabu search have large distances. These solutions, which are bad in terms of POD, are dominated by the solutions for other algorithms. Excluding such solutions leads to a reduction in the mean ideal distance.

Moreover, Table IX shows the values of the RQ and MGID metrics for all algorithms. From this metrics perspective, the tabu search algorithm outperforms the other algorithms in many problems. Furthermore, the average of metrics for all problems is computed as shown in Table X. Because the metrics values increase dramatically with problem size, the metrics values are first normalized using the following formula:

$$f_i^{norm}(algo) = \frac{Val_i - Val_{min}}{Val_{max} - Val_{min}}, \; i = 1,2,..,k \qquad (14)$$

Table X illustrates the ranking of algorithms based on the average performance measures. According to the ranking, the simulated annealing algorithm outperforms the other algorithms in terms of MID. In contrast, the tabu search algorithm achieves the highest rank among the proposed algorithms in the POD, MGID, MID_{NDS} , and RQ metrics.

The results of the diversity metrics for evaluating the performance of the proposed algorithms in solving MOSC problems are illustrated in Table XI. The average metrics for all problems is also provided in Table XII. From Table XII, it can be noted that the order of algorithms to obtain SNS is the same order of algorithms for obtaining SNS_{NDS} . This result indicates that the behavior of SNS_{NDS} is the same as the behavior of SNS. With respect to the RDC metric, the order of the algorithms to obtain the best value of RDC is simulated annealing, followed by genetic algorithm, and then the tabu search. The RDC metric measures the ratio of MID_{NDS} and SNS_{NDS} taking the ratio of domination into account. Moreover, the values of SNS_{NDS} and MID_{NDS} are shown in Fig.1. It can be seen that the SNS_{NDS} are smaller than MID_{NDS}. The relationship between SNS and SNS_{NDS} is shown in Fig. 2. It is clear that the SNS is greater than SNS_{NDS} . However, both metrics obtain the same result.

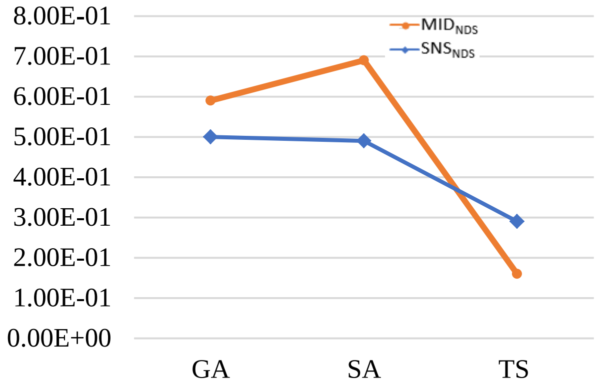

Fig.1 Measures comparison for the average of proposed metric

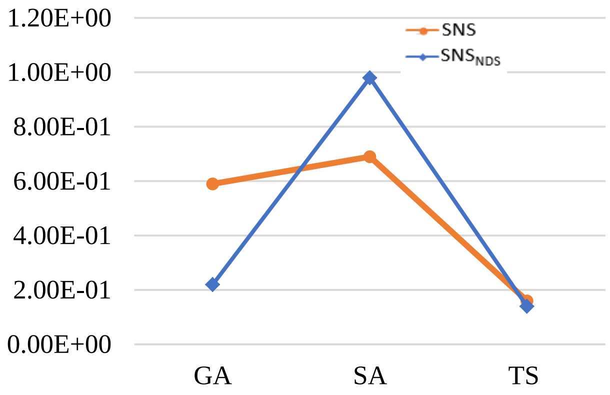

Fig.2 Measures comparison for the average of diversity metrics

VI. CONCLUSION AND FUTURE WORK

In recent years, various metrics have been proposed to evaluate the performance of algorithms for multi-objective optimization problems. In this paper, seven new metrics for measuring the performance of multi-objective algorithms are proposed. The new metrics evaluate the features of multi-objective problems relating to the convergence and diversity of solutions. In addition, a new metric based on the probability of success is developed. Furthermore, the relationship among the existing metrics and the proposed metrics is investigated. A computational experiment is conducted to test the applicability of the proposed metrics. The results show the powerful usage of the proposed metrics for evaluating the performance of the algorithms. The proposed measures are also more suitable for the evaluation of the algorithms performance. They improve and enrich the multi-criteria area.

In this paper, we select the most common metrics that measure convergence and diversity in the multi-objective supply chain area. For future work, other common metrics in different areas could be selected. In this study, the best solution obtained by the algorithm is selected based on the minimum distance from the ideal point. The probability of success metric may use other multi-criteria techniques for the selection of the best solution. This is an area for future research.

References

[1] C., Coello Coello, G.B., Lamont, D.A.,Van Veldhuizen, Evolutionary Algorithms for Solving Multi-Objective Problems, 2nd edn. Springer, New York (2007)

[2] E. Zitzler and L. Thiele, “Multi-objective evolutionary algorithms: a comparative case study and the strength Pareto approach,” evolutionary computation, IEEE Transactions on evolutionary computation, vol. 3, no. 4, pp. 257–271, (1999).

[3] D. A. Van Veldhuizen, “Multi-objective evolutionary algorithms: classifications, analyses, and new innovations,” DTIC Document, Tech. Rep., (1999)

[4] C. A. C. Coello, D. A. Van Veldhuizen, and G. B. Lamont, Evolutionary algorithms for solving multi-objective problems. Springer, (2002), vol. 242.

[5] K. Deb, Multi-objective optimization using evolutionary algorithms. John Wiley & Sons, (2001), vol. 16.

[6] E. Zitzler and L. Thiele, “Multi-objective optimization using evolutionary algorithms a comparative case study,” in Parallel problem solving from nature PPSN V. Springer, (1998), pp. 292–301.

[7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multi-objective genetic algorithm: NSGA-II,” Evolutionary Computation, IEEE Transactions on evolutionary computation, vol. 6, no. 2, (2002), pp. 182–197.

[8] K. Deb and S. Jain, “Running performance metrics for evolutionary multi-objective optimizations,” Proceedings of the Fourth Asia-Pacific Conference on Simulated Evolution and Learning (SEAL’02), (2002), pp. 13–20.

[9] A. Farhang-Mehr and S. Azarm, “Diversity assessment of Pareto optimal solution sets: an entropy approach,” in Computational Intelligence, Proceedings of the World on Congress on, vol. 1. IEEE, (2002), pp. 723–728.

[10] J. R. Schott, “Fault tolerant design using single and multicriteria genetic algorithm optimization.” DTIC Document, Tech. Rep., (1995).

[11] E. Zitzler, K. Deb, and L. Thiele, “Comparison of multi-objective evolutionary algorithms: Empirical results,” Evolutionary Computation, vol. 8, no. 2, (2000), pp. 173–195,